# DNA variation at the *rp49* gene region in *Drosophila madeirensis* and *D. subobscura* from Madeira: inferences about the origin of an insular endemic species

M. KHADEM,* J. ROZAS,† C. SEGARRA† & M. AGUADÉ†

**Departamento da Biologia, Centro de Investigação em Ciências Agrárias, Universidade da Madeira, Funchal, Portugal*
*†Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain*

*Keywords:*

DNA polymorphism;
*Drosophila madeirensis;*
*Drosophila subobscura;*
*rp49* gene;
speciation.

## Abstract

An ~1.6-kb fragment spanning the *rp49* gene was sequenced in 16 lines of *Drosophila subobscura* from Madeira and in 22 lines of the endemic species *D. madeirensis*. Nucleotide diversity in *D. subobscura* from Madeira ($\pi = 0.0081$) was similar to that in lines from Spain carrying the $O_{3+4}$ chromosomal arrangement ($\pi = 0.0080$). No significant genetic differentiation was detected between insular and continental $O_{3+4}$ lines of *D. subobscura*. These results are compatible both with a rather recent and massive colonization, and with multiple colonization events from the continent. Nucleotide diversity in *D. madeirensis* ($\pi = 0.0076$) was similar to that in *D. subobscura*, which deviates from the expectation, under strict neutrality, of a lower level of variation in an insular species with a small population size. The observed numbers of shared polymorphisms and of fixed differences between *D. madeirensis* and *D. subobscura* are compatible with the isolation model of speciation, where shared polymorphisms are due to common ancestry.

## Introduction

*Drosophila madeirensis* Monclús and *D. subobscura* Collins are closely related species of the obscura group that coexist in Madeira. The former species is endemic to this island, while *D. subobscura* has a wider distribution area that encompasses most of Europe, northern Africa, three archipelagos of the Macaronesia (Azores, Madeira and Canary Islands) and Asia Minor (Krimbas, 1992). Madeira, like the other two archipelagos, has a volcanic origin. It is a rather small island (730 $km^2$) that arose 5–6 Myr ago (Galopin de Carvalho & Brandão, 1991). Its present fauna and flora has been therefore shaped by the set of species that have colonized the archipelago and by their subsequent evolution. Species endemism in the flora of the Madeiran archipelago is about 10% (Press & Short, 1994). This percentage is even higher for the terrestial fauna of Madeira and reaches approximately 15% for Diptera (Baez, 1993). Therefore, despite the rather close proximity of Madeira both to the continent and to the other archipelagos, it seems rather well isolated. Possible founder events during the colonization and differential selection in the island would have caused the genetic differentiation of the colonizing populations and in some cases it would have led to the origin of new species. Natural populations of *D. madeirensis* and *D. subobscura* in the islands therefore offer the opportunity of studying the colonization process and the effects of factors such as founder events and reduced migration, likely associated with insularity, on extant levels of nucleotide variation in these populations.

Populations of *D. subobscura* from Madeira have been characterized at the chromosomal and allozyme levels (Prevosti, 1972; Prevosti, 1974; Larruga *et al*., 1983; Pinto *et al*., 1997). Restriction map variation in the mtDNA and in the nuclear gene *rp49* has also been surveyed (Afonso *et al*., 1990; Pinto *et al*., 1997; Khadem *et al*., 1998). At the chromosomal level, populations from Madeira are well differentiated from continental populations but are rather similar to populations from the Canary Islands (Prevosti, 1972, 1974). However, for both allozyme and

*Correspondence:* Montserrat Aguadé, Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Diagonal 645, 08071 Barcelona, Spain.
Tel.: +34 93 402 1493; fax: +34 93 4110969; e-mail: aguade@bio.ub.es

mtDNA variation they are differentiated from the Canary Islands populations but not from continental populations (Larruga *et al*., 1983; Afonso *et al*., 1990; Pinto *et al*., 1997). In contrast, restriction fragment length polymorphism analysis of the *rp49* gene region of a *D. subobscura* population from Madeira revealed some genetic differentiation between insular and continental populations. This study also revealed less variation in the insular than in continental populations (Rozas *et al*., 1995; Khadem *et al*., 1998).

*D. madeirensis* and *D. subobscura* diverged rather recently, about 0.6–1.0 Myr ago, according to nucleotide divergence at the *rp49* gene region (Ramos-Onsins *et al*., 1998). They are rather similar morphologically (Monclús, 1984) and their reproductive isolation is not complete, as fertile and viable hybrids are obtained in some interspecific crosses (Khadem & Krimbas, 1991; Papaceit *et al*., 1991). Unlike in *D. subobscura*, only a few studies on intraspecific genetic variation have been conducted with *D. madeirensis* (González *et al*., 1983, 1990). Other studies have shown that the ancestors both of *D. madeirensis* and of extant populations of *D. subobscura* in Madeira are probably the result of independent colonization events from the continent (Khadem *et al*., 1998).

In the present study, we have sequenced the *rp49* gene region in a *D. subobscura* population from Madeira and in a natural population of *D. madeirensis*. The *rp49* gene (named *RpL32* in Flybase) encodes the ribosomal protein 49 (or ribosomal protein L32). In *D. subobscura*, this gene is located in the O chromosome at band 91C, close to the proximal breakpoint of inversion 3 (Aguadé, 1988). Two alternative chromosomal arrangements for this genomic region, $O_{3+4}$ and $O_{st}$, segregate in natural populations of *D. subobscura*. Each of these arrangements originated from the ancestral $O_3$ arrangement by a single inversion: $O_{3+4}$ by inversion 4, and $O_{st}$ by inversion 3. While *D. madeirensis* is monomorphic for $O_3$, this arrangement is no longer present in extant populations of *D. subobscura*. Unlike continental populations of *D. subobscura*, the population from Madeira is nearly monomorphic for the $O_{3+4}$ arrangement. Present data indicate that, within this arrangement, insular and continental populations of *D. subobscura* are not genetically differentiated. In contrast, *D. madeirensis* and *D. subobscura* populations are highly differentiated despite their similar level of nucleotide variation. Finally, the distribution of nucleotide variation within and between species is consistent with the isolation model of speciation, where a single ancestral population splits into two descendent populations that, subsequently, remain isolated.

## Materials and methods

### Fly stocks

Sixteen *Drosophila subobscura* and 22 *D. madeirensis* lines were used in the present study. Flies from both species were collected in Ribeiro Frio (Madeira, Portugal) in 1997; isofemale lines were established upon arrival in the laboratory. Highly inbred lines were subsequently obtained by 12 generations of sibmating. The gene arrangement for the O chromosome was determined for each line by observation of polytene chromosomes of salivary glands of third-instar larvae.

### DNA extraction and sequencing

A modification of protocol 48 in Ashburner (1989) was used to extract genomic DNA. Despite the fact that the lines were highly inbred, single flies were used for the DNA extraction to reduce the chance of heterozygosity for the studied region. If for a particular inbred line sequencing revealed that an individual was heterozygous, additional single flies from the same inbred line were used until a homozygous individual was found; only the sequence of this individual was used in the analyses. The complete *rp49* region (~1.6 kb) was amplified by the polymerase chain reaction (PCR; Saiki *et al*., 1988) using 20-mer oligonucleotides. The PCR products were purified with Qiaquick columns (Qiagen). Internal primers, designed at intervals of ~300 nucleotides, were used for sequencing. Both strands were cycle sequenced using the rhodamine sequencing chemistry (Perkin-Elmer) and analysed on a Perkin Elmer ABI PRISM 377 automated DNA sequencer. The newly reported sequences are deposited in the EMBL nucleotide sequence database library under accession numbers ΔJ310269–ΔJ310306.

### DNA analysis

The sequences were multiply aligned using the CLUSTAL W program (Thompson *et al*., 1994) and edited with the MacClade program version 3.0.6 (Maddison & Maddison, 1992). The alignment was further optimized manually.

The DnaSP version 3.5 software (Rozas & Rozas, 1999) was used for the analysis of polymorphism and genetic differentiation. This program was also used to perform Tajima's test (Tajima, 1989) and Fu and Li's tests (Fu & Li, 1993). The significance of the corresponding test statistics was established by computer simulation using the coalescent algorithm without recombination.

The level of polymorphism was estimated as the number of polymorphic sites ($S$), the average number of pairwise nucleotide differences ($k$), nucleotide diversity ($\pi$; Nei, 1987) and expected heterozygosity per site or Watterson's estimator ($\theta$; Watterson, 1975). Genetic differentiation between species, populations or gene arrangements was estimated as the average number of nucleotide substitutions per site between groups ($d_{xy}$). The statistical significance of genetic differentiation between groups, as estimated by $K_S^*$, was established by the permutation test (Hudson *et al*., 1992a). The proportion of nucleotide diversity attributable to variation between populations, $F_{st}$, was calculated according to

Hudson *et al*. (1992b). $F_{st}$ was used to estimate the migration parameter *Nm*, under the island model of population structure and assuming migration-drift equilibrium (Wright, 1951; Hudson *et al*., 1992b).

Recombination events were identified by the four-gamete test (Hudson & Kaplan, 1985). Linkage disequilibrium was analysed between parsimony informative sites and the statistical significance of pairwise associations was obtained by the $\chi^2$ test. The Bonferroni procedure was used to correct for multiple tests (Weir, 1996).

The neighbour-joining method (Saitou & Nei, 1987), as implemented in the PAUP program (Swofford, 1998), was used for phylogenetic reconstruction using genetic distances corrected for multiple hits (Jukes & Cantor, 1969). Bootstrap values were obtained from 500 replicates.

### DNA divergence models

We applied the method developed by Wakeley (1996a,b) to test whether the pattern of nucleotide variation fits to that expected under the isolation model of divergence. In this model, a single ancestral population splits into two descendent populations that remain isolated from each other for some period of time. One of the alternative models is the so-called two-population equilibrium migration model, where gene flow between populations can occur and therefore populations are not completely isolated. The isolation model (Wakeley, 1996a,b; Wakeley & Hey, 1997) can be described by four parameters: $\theta_1$, $\theta_2$, $\theta_A$ and $\tau$ ($\tau = 2\mu t$). $\theta_1$, $\theta_2$, $\theta_A$ are the $\theta$ values per sequence for populations 1, 2 and for the ancestral population, respectively; $\mu$ is the mutation rate per sequence per generation; and $t$ is the time of separation measured in generations. $T$, the isolation time in $2N$ generations (where $N$ is the effective population size), can be easily obtained from $T = \tau/\theta_1$. These parameters can be estimated from four categories of sites: the number of shared polymorphic sites ($S_S$), the number of exclusive polymorphic sites in each species ($S_{X1}$ and $S_{X2}$), and the number of fixed differences between species ($S_F$). For the analysis, we refer to *D. madeirensis* as species 1, and to *D. subobscura* as species 2.

## Results

### Nucleotide variation in *Drosophila subobscura* from Madeira

The *rp49* region was sequenced in 16 lines of *D. subobscura* from Madeira. All lines were homozygous for the $O_{3+4}$ chromosomal arrangement. The multiple alignment included 1508 sites after excluding all sites with alignment gaps. A summary of polymorphism is given in Table 1 and Fig. 1. There were 14 different haplotypes and haplotype diversity was 0.983. A total of 7 length and 46 nucleotide polymorphisms were detected. All length polymorphisms were in noncoding regions (intron and flanking regions). Two length polymorphisms could be considered complex mutational events, and the rest were microsatellites. All nucleotide polymorphisms were silent. In 17 of these polymorphisms (37%), the rarest variant was a singleton, while 29 were parsimony informative sites (28 with two variants and one with three variants). The estimated nucleotide diversity ($\pi$) was lower than the estimated heterozygosity per site ($\theta$) (Table 1), reflecting the relatively high percentage of sites with singletons in the sample.

**Table 1** Nucleotide polymorphism in *Drosophila subobscura* and *D. madeirensis*.

| | *D. subobscura* | | | |
|---|---|---|---|---|
| | Madeira | Galicia | | *D. madeirensis* |
| Chromosomal class | $O_{3+4}$M | $O_{3+4}$G | $O_{st}$ | $O_3$ |
| Sample size | 16 | 18 | 16 | 22 |
| *S* (singletons) | 46 (17) | 54 (27) | 46 (29) | 55 (22) |
| *k* | 12.183 | 11.895 | 9.425 | 11.446 |
| $\pi$ | 0.0081 | 0.0080 | 0.0064 | 0.0076 |
| $\theta$ | 0.0092 | 0.0106 | 0.0093 | 0.0101 |
| $\pi_s$ | 0.0102 | 0.0101 | 0.0080 | 0.0096 |
| $\theta_s$ | 0.0119 | 0.0138 | 0.0120 | 0.0128 |
| Tajima's *D* | −0.5909 | −1.114 | −1.414 | −1.010 |
| Fu-Li's *D* | −0.4034 | −1.103 | −1.684 | −1.040 |
| Fu-Li's *F* | −0.5689 | −1.358 | −1.958 | −1.262 |

*S*, number of segregating sites; *k*, average number of pairwise nucleotide differences; $\pi$, nucleotide diversity; $\pi_s$, silent nucleotide diversity; $\theta$, heterozygosity per site; $\theta_s$, silent heterozygosity per site; M, Madeira; G, Galicia. *D. guanche* was used as the outgroup in Fu and Li tests.

Different tests of neutrality (Tajima, 1989; Fu & Li, 1993) were performed (Table 1). None of the tests detected a significant departure from neutral expectations in a stationary population. The test statistics were in all cases negative, which is an indication of the observed excess of low-frequency variants.

The minimum number of recombination events, as estimated by the four-gamete test (Hudson & Kaplan, 1985), was 7. In the analysis of linkage disequilibrium that included only parsimony informative sites, 54 out of 378 pairwise comparisons (14%) showed a significant association by the $\chi^2$ test ($P < 0.05$). Using the Bonferroni correction for multiple tests (Weir, 1996), this number dropped to 10 (3%). There was a negative relationship between the degree of linkage disequilibrium, as estimated by $r^2$ (Hill & Robertson, 1968), and physical distance between sites.

### Genetic differentiation between *D. subobscura* populations from Madeira and Spain

Nucleotide variation in the *D. subobscura* sample from Madeira that was monomorphic for the $O_{3+4}$ chromosomal

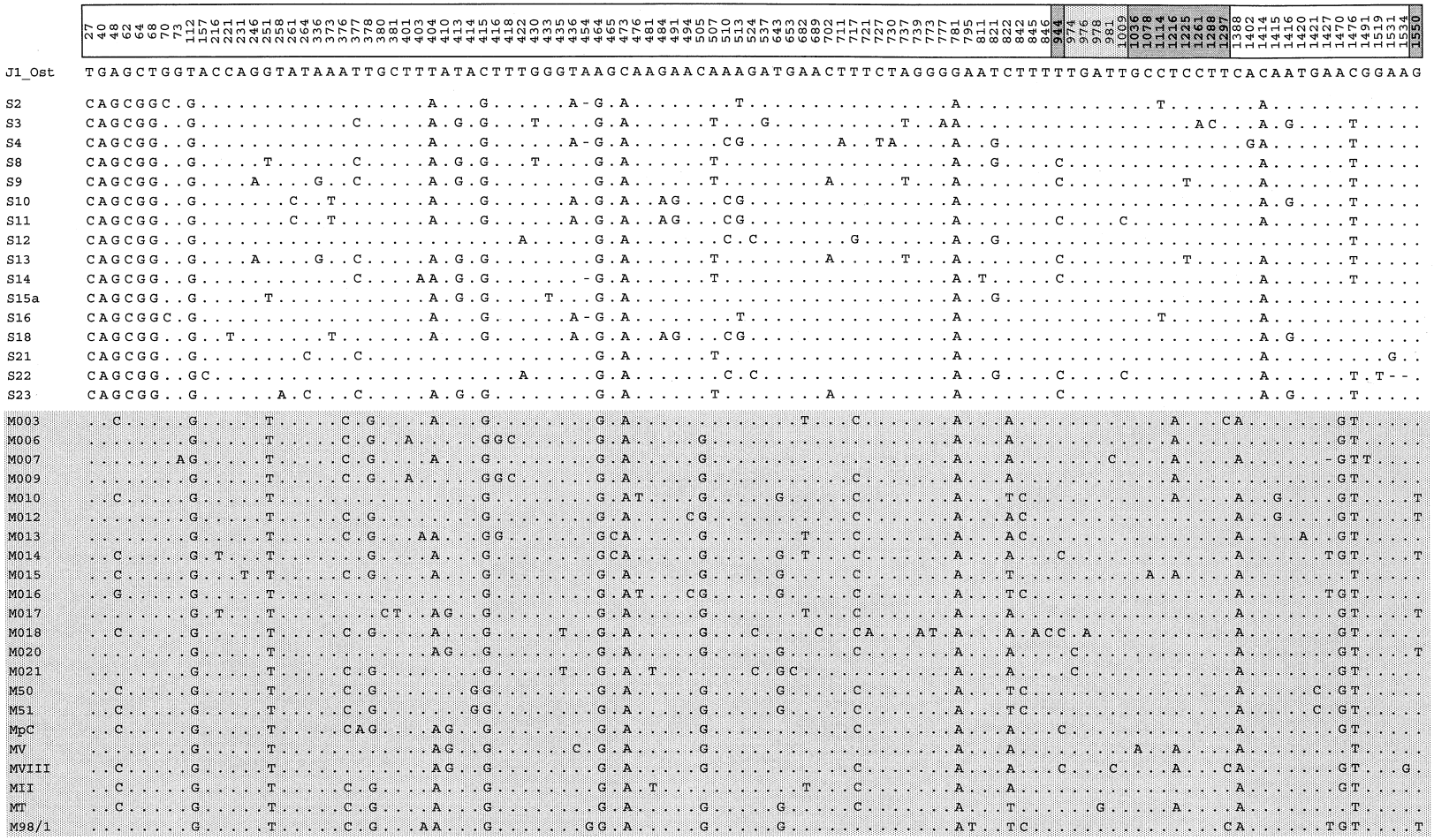


**Fig. 1** Nucleotide polymorphism at the *rp49* gene region in *Drosophila subobscura* and *D. madeirensis*. Nucleotide numbering is according to Aguadé (1988). The first sequence corresponds to line J1 in Rozas & Aguadé (1994) that presents the $O_{st}$ chromosomal arrangement. Lines of *D. subobscura* from Madeira are designated as S and those of *D. madeirensis* as M. Nucleotides identical to the first sequence are indicated by a dot. Length polymorphisms are not shown. A deleted nucleotide in a given polymorphic site is indicated by a gap. Polymorphic sites 27–846 correspond to the 5′ flanking region, site 944 to exon 1, sites 974–1009 to the intron, 1036–1297 to exon 2, 1388–1534 to the 3′ flanking region, and polymorphic site 1550 to the *serendipity* coding region.

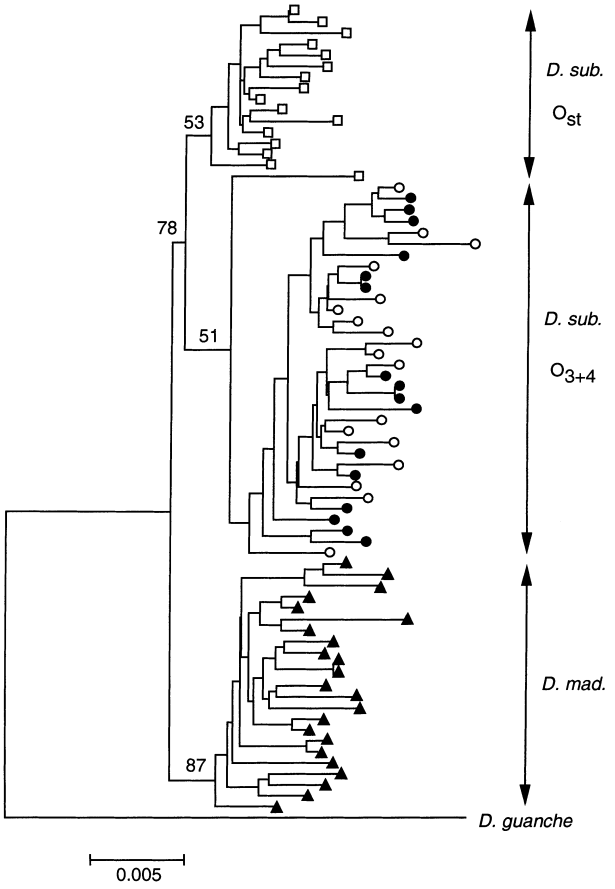


**Fig. 2** Neighbour-joining tree of the *rp49* gene region sequences of *Drosophila subobscura* and *D. madeirensis* using *D. guanche* as the outgroup. Bootstrap values based on 500 replicates are given on the main nodes. The scale bar represents 0.005 nucleotide substitutions per site. $O_{3+4}$ *D. subobscura* lines from Madeira and Galicia are indicated by black and white circles, respectively.

arrangement was compared to variation in lines with the same gene arrangement of a population from Galicia (Spain). A summary of a previously published analysis of polymorphism in that population (Rozas & Aguadé, 1994; Rozas *et al.*, 1999) is also given in Table 1.

The level of nucleotide diversity in the lines from Madeira ($O_{3+4}$ M in Table 1) was similar to that in the $O_{3+4}$ lines from Spain ($O_{3+4}$ G in Table 1), but higher than in the Spanish $O_{st}$ lines. Despite the similar level of nucleotide diversity in the $O_{3+4}$ lines from Madeira and from Galicia, the percentage of singletons was higher in this latter sample (50%). Haplotype diversity in the sample from Madeira (0.983) was only slightly lower than in both the $O_{3+4}$ (1.0) and $O_{st}$ (1.0) lines from Spain.

Genetic differentiation between $O_{3+4}$ lines from Madeira and Galicia was very low as measured by $d_{xy}$ (0.0081). In fact, there was a high proportion of shared polymorphisms, or sites segregating for the same two nucleotides, between these lines (32 out of 68 polymorphic sites in the combined sample) and no fixed differences between populations. No significant genetic differentiation between $O_{3+4}$ lines from both locations was detected by the permutation test (Hudson *et al.*, 1992a). In contrast, the sample from Madeira, like the sample of $O_{3+4}$ lines from Galicia (Rozas & Aguadé, 1994; Rozas *et al.*, 1999), was genetically differentiated from the $O_{st}$ lines from Galicia ($P < 0.001$ for the $K_S^*$ estimator).

The lack of genetic differentiation between $O_{3+4}$ lines is also reflected in the neighbour-joining tree built using the *D. guanche* sequence (Ramos-Onsins *et al.*, 1998) as the outgroup (Fig. 2). In fact, $O_{3+4}$ lines from both Madeira and Galicia are interspersed in the tree. On the other hand, $O_{st}$ lines from Galicia form a separate cluster except for line J11, for which there is evidence of gene conversion from $O_{3+4}$ (see Rozas & Aguadé, 1994).

### Nucleotide variation in *Drosophila madeirensis*

An ~1.6-kb fragment spanning the *rp49* region was sequenced in 22 lines of *D. madeirensis*. All lines were homozygous for the $O_3$ chromosomal arrangement. There were 21 different haplotypes and the haplotype diversity was 0.996. A total of 4 length and 55 nucleotide polymorphisms were detected (Table 1 and Fig. 1). Like in *D. subobscura*, all length polymorphisms were in noncoding regions and all nucleotide polymorphisms were silent. Two length polymorphisms could be considered complex mutational events and the other two were microsatellites (Fig. 1). In 22 of the 55 polymorphic sites detected (40%), the rarest variant was a singleton, and the other 33 were parsimony informative sites (one of them with three variants).

Different neutrality tests based on polymorphism data (Tajima, 1989; Fu & Li, 1993) were performed. Like in *D. subobscura*, none of the tests showed a significant deviation from neutral expectations, although in all cases the test statistics were negative (Table 1).

A minimum of 13 recombination events was inferred by the four-gamete test (Hudson & Kaplan, 1985) in the history of the sample studied. In 51 out of 496 pairwise comparisons between polymorphic sites (10%), the variants were significantly associated or in linkage disequilibrium. Using the Bonferroni correction for multiple tests (Weir, 1996), this number dropped to 6 (1%).

### Differentiation between *D. subobscura* and *D. madeirensis*

The level of polymorphism, measured as nucleotide diversity or $\pi$, was comparable in the samples of *D. madeirensis* and of *D. subobscura* from Madeira (Table 1). Estimated nucleotide diversity in *D. madeirensis* was consequently similar to that in the $O_{3+4}$ lines from Galicia and higher than that in the $O_{st}$ lines from that population. The percentage of singletons was rather similar in the samples of *D. subobscura* from Madeira and

of *D. madeirensis* (37 and 40%, respectively), and lower than in the $O_{3+4}$ and $O_{st}$ lines from Galicia.

The degree of genetic differentiation, as estimated by $d_{xy}$, between *D. madeirensis* and any of the three samples from *D. subobscura* ($O_{3+4}$ M, $O_{3+4}$ G and $O_{st}$) was rather similar (Table 2). The estimated $F_{st}$ values for the three comparisons between *D. madeirensis* and *D. subobscura* were rather high. The permutation test (Hudson *et al*., 1992a) revealed a significant genetic differentiation between both species. The number of fixed differences was, however, rather low (Table 2). Most fixed differences between species were due to mutations in a short stretch of the 5′ flanking region (between sites 27 and 68 in Fig. 1). When this part was excluded from the analyses, the number of fixed differences dropped to 1 in all three comparisons.

Despite the presence of shared polymorphisms between *D. madeirensis* and *D. subobscura* (Table 2), all *D. madeirensis* sequences clustered in the neighbour-joining tree obtained with all the *D. subobscura* and *D. madeirensis* sequences and using *D. guanche* as the outgroup (Fig. 2). This cluster was supported by a much higher bootstrap value (87%) than the $O_{3+4}$ and $O_{st}$ clusters (51% and 53%, respectively).

### Speciation models

The $\psi$-test statistic, proposed by Wakeley (1996b), was computed to distinguish between the isolation and migration models. Present data for *D. madeirensis* and *D. subobscura* do not allow rejection of the isolation model ($\psi = 0.22$; $P[\psi \geqslant 0.22] = 1.0$). We applied the Wakeley & Hey (1997) formulas to obtain estimates of $\theta_1$, $\theta_2$, $\theta_A$ and $T$ (Table 3). This analysis is based on the infinite-sites model. Under that assumption, shared polymorphisms represent polymorphisms present in the ancestral population. However, shared polymorphisms could also have arisen by the independent accumulation of mutations in each species (i.e. by parallel mutation). Therefore, we first tested whether the observed number of shared polymorphisms could be explained by parallel mutation. Assuming that the number of shared polymorphisms follows a hypergeometric distribution (Rozas & Aguadé, 1994), and that each nucleotide can mutate to three alternative states, we rejected the null hypothesis that shared polymorphisms were the result of parallel mutations ($P[X \geqslant 4] = 0.0012$). The similar estimates obtained for $\theta_1$, $\theta_2$ and $\theta_A$ (Table 3) would indicate that the ancestral population size was similar to that of the two extant descendent populations (*D. madeirensis* and *D. subobscura*, respectively). In addition, the estimated time of split indicates that these species have been isolated for a rather long time (Table 3).

**Table 2** Genetic differentiation between *Drosophila subobscura* and *D. madeirensis*.

| | Dmad vs. Dsub $O_{st}$ | Dmad vs. Dsub $O_{3+4}$M | Dmad vs. Dsub $O_{3+4}$G | Dmad vs. Dsub $O_{3+4}$T |
|---|---|---|---|---|
| No. of mutations | 97 | 103 | 108 | 121 |
| Shared mutations | 4 | 4 | 5 | 5 |
| Fixed differences | 1 | 6 | 7 | 6 |
| $d_{xy}$ | 0.0125 | 0.0170 | 0.0160 | 0.0171 |
| *P* value* | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| $F_{st}$ | 0.4526 | 0.5422 | 0.5493 | 0.5504 |
| *Nm* | 0.30 | 0.21 | 0.21 | 0.20 |

Dmad, *D. madeirensis*; Dsub, *D. subobscura*; M, Madeira; G, Galicia; T, total (Madeira and Galicia). **P* value for $K_S^*$ test.

**Table 3** Estimates of population parameters from the isolation model.

| | | | | |
|---|---|---|---|---|
| Segregating/fixed sites | $S_{X1} = 50$ | $S_{X2} = 43$ | $S_S = 4$ | $S_F = 6$ |
| Population parameters | $\theta_1 = 14.28$ | $\theta_2 = 13.43$ | $\theta_A = 16.43$ | $\tau = 15.89$* |

For notation see Materials and Methods. Subindexes 1 and 2 correspond to *D. madeirensis* and *D. subobscura*, respectively. *The split time is $T = 1.11$ measured in units of $2N_1$ generations.

## Discussion

### *Drosophila subobscura* from Madeira

Populations of *D. subobscura* from Madeira, like those from the Canary Islands, are nearly monomorphic for the $O_{3+4}$ chromosomal arrangement (Prevosti, 1971; Larruga *et al*., 1983). This is in contrast with the rich chromosomal polymorphism that most other populations harbour for this chromosome (Krimbas, 1992). Furthermore, the absence in the islands of arrangements present in high frequency in nearby continental populations (e.g. $O_{3+4+8}$ in North-western Africa and $O_{3+4+7}$ in the Atlantic coast of the Iberian peninsula) does not seem to favour the hypothesis of recent migration from the continent. Otherwise, selection against the establishment of these other arrangements in the islands should be very strong given their high frequency in those continental populations (see Khadem *et al*., 1998).

According to present results, the level of nucleotide variation in the *rp49* gene region is similar in $O_{3+4}$ lines from Madeira and the continent. Additionally, $O_{3+4}$ lines from these locations are not genetically differentiated for that region. Some differentiation had been, however, detected in a previous restriction-map survey (Khadem *et al*., 1998). The discrepancy between the results of these studies might be partly due to the relatively low number of nucleotides sampled in the previous study and therefore to the high stochastic variance associated with nucleotide variation estimates.

Available data for chromosomal, mtDNA and *rp49* variation in *D. subobscura* populations from Madeira and the continent do not support the hypothesis that extant

populations from Madeira are the descendents of a single colonization event occurring soon after the origin of $O_{3+4}$ (Khadem *et al.*, 1998). Two different scenarios would be compatible with available data: (i) a rather recent and massive colonization of Madeira by continental *D. subobscura*, and (ii) multiple colonization events from the continent. In both scenarios, selection would have precluded the establishment in Madeira of chromosomal arrangements other than $O_{3+4}$ (see Khadem *et al.*, 1998). Although we cannot ascertain which scenario most likely reflects the origin of extant *D. subobscura* populations in Madeira, the similar level of nucleotide variation detected within $O_{3+4}$ in insular and continental populations allows us to assert that the colonization of Madeira was not associated with a strong founder event.

### Comparison of DNA variation between *D. madeirensis* and *D. subobscura*

The *rp49* region is the first nuclear region whose variation has been analysed in a natural population from the endemic species *D. madeirensis*. 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. Comparison of nucleotide variation at the *rp49* gene region between *D. madeirensis* and *D. subobscura* is not in agreement with that prediction. In fact, the estimated nucleotide diversity in *D. madeirensis*, which is monomorphic for the $O_3$ arrangement, was similar to that estimated for $O_{3+4}$ and slightly higher than for the $O_{st}$ chromosomal arrangement (Table 1). The high level of nucleotide variation in *D. madeirensis* might be explained if ancestral populations of this species were much larger than current populations and the species had suffered a reduction in population size. This reduction could be associated with the colonization and expansion of *D. subobscura* in Madeira after the origin of *D. madeirensis*; it could also be associated with destruction of the natural habitat of *D. madeirensis* (laurisilva forest) occurred during the last 400 years (Doria, 1945; Frutuoso, 1979). In any case, ancestral populations of this insular species would probably be not as large as continental populations of *D. subobscura*. Nevertheless, the level of variation in *D. madeirensis* clearly indicates that the origin of this species was not associated with a strong founder event. In fact, *D. madeirensis* and *D. subobscura* exhibit a comparable level of nucleotide variation in the *rp49* region and they also present a similar frequency spectrum, as measured by Tajima's *D* and Fu and Li's *D* and *F* statistics.

Present data on nucleotide variation in *D. madeirensis* and *D. subobscura* have been compared to those available for other closely related *Drosophila* species pairs with one insular representative. Two such pairs can be found in the triad formed by the cosmopolitan species *D. simulans* and the endemic species *D. sechellia* and *D. mauritiana* (Hey & Kliman, 1993; Kliman & Hey, 1993). These authors analysed nucleotide variation at the *period*, *zeste* and *yolk protein 2* gene regions and found that *D. simulans* and *D. mauritiana* showed comparable levels of variation. In contrast, *D. sechellia* exhibited a much lower nucleotide variation than *D. simulans*. Our results showing an *a priori* unexpected high level of variation in the insular species (*D. madeirensis*) are quite similar to those reported for the *D. simulans* and *D. mauritiana* pair. Nevertheless, the expected positive correlation between heterozygosity and effective population size has been, and still is, a controversial issue in population genetics (Lewontin, 1974; Maynard Smith & Haigh, 1974; Gillespie, 1999, 2000). According to the strict neutral mutation model, the expected heterozygosity is a function of the effective population size (Kimura, 1983). Under several selection models, however, heterozygosity is nearly independent of the population size (Gillespie, 1999). For example, under some deleterious mutation models, heterozygosity is rather insensitive to the population size, and is thus mainly a function of the mutation rate (Gillespie, 1999). This insensitivity is also predicted by the pseudohitchhiking model (Gillespie, 2000) that considers the effect of advantageous mutations on the dynamics of neutral variation at a closely linked locus. Under this model, heterozygosity can even decrease with increasing population size.

Some shared polymorphisms between species were detected both in the sequence comparison of the *period* locus between *D. simulans* and *D. mauritiana*, and of the *rp49* region between *D. madeirensis* and *D. subobscura*. In both cases, the data were compatible with the isolation model, where shared polymorphisms are due to common ancestry. However, shared polymorphisms between closely related species could also be due to the introgression resulting from rare hybridization between species. In fact, it has been recently proposed that, for regions not involved in reproductive isolation, introgressive hybridization might be more important than previously thought (Ting *et al.*, 2000). If that were the case for the *rp49* region, some of the observed shared variants could have originated by mutation in one of the descendent lineages after the split of the species, and they would have entered the second species gene pool by introgression. Although some hybrids between *D. madeirensis* and *D. subobscura* have been detected in recent collections (N. Khadem, unpublished results), this observation is not a proof of introgression. In addition, for a neutrally evolving gene, the relatively high number of fixed differences observed between species seems unlikely under the introgression scenario.

As in the species studied, the *rp49* gene is located in a region affected by chromosomal inversions, there is an

additional difficulty for explaining the presence of shared polymorphisms both if they are due to common ancestry or to introgressive hybridization. In fact, the unique origin of an inversion would *a priori* preclude the existence of shared polymorphisms between the ancestral and derived chromosomal arrangements. In our case, *D. madeirensis* is monomorphic for the ancestral $O_3$ arrangement that went extinct in the *D. subobscura* lineage, while extant populations of *D. subobscura* segregate for the $O_{3+4}$ and $O_{st}$ arrangements that originated from $O_3$. Therefore, gene transfer between arrangements, either by double crossing over or gene conversion, would be required to explain the presence of shared polymorphisms between species (Rozas & Aguadé, 1994; Rozas *et al*., 1999). In this case, the extent of shared polymorphism might be lower at the *rp49* region than in other regions not affected by chromosomal inversions. Only analysis of multiple regions will allow establishing the effect of chromosomal polymorphism on the number of shared polymorphisms between *D. madeirensis* and *D. subobscura*. Additionally, the comparative analysis of nucleotide variation in multiple regions of *D. subobscura* and *D. madeirensis* will improve our understanding of the speciation process in Madeira.

## Acknowledgments

We thank John Wakeley for sharing computer programs, Gema Blasco and David Salguero for technical support, and Serveis Científico-Tècnics, Universitat de Barcelona, for automated sequencing facilities. This work was supported by grants PB97-0918 from Comisión Interdepartamental de Ciencia y Tecnología, Spain, and 1999SGR-25 from Comissió Interdepartamental de Recerca i Innovació Tecnològica, Catalonia, Spain, to M.A., and by CITMA (Centro de Ciências e Tecnologia da Madeira) to M.K.

## References

Afonso, J.M., Volz, A., Hernández, M., Ruttkay, H., González, A.M., Larruga, J.M., Cabrera, V.M. & Sperlich, D. 1990. Mitochondrial DNA variation and genetic structure in Old-World populations of *Drosophila subobscura*. *Mol. Biol. Evol.* **7**: 123–142.

Aguadé, M. 1988. Nucleotide sequence comparison of the *rp49* gene region between *Drosophila subobscura* and *D. melanogaster*. *Mol. Biol. Evol.* **5**: 433–441.

Ashburner, M. 1989. *Drosophila: a Laboratory Handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Baez, M. 1993. Origin and affinities of the fauna of Madeira. *Boletim Do Museu Municipal Do Funchal. Supplement* **2**: 9–40.

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

Frutuoso, G. 1979. *Livro segundo das saudades da Terra*. Ponto delgada, Portugal.

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, Universidade de Lisboa.

Gillespie, J. H. 1999. The role of population size in molecular evolution. *Theor. Popul. Biol.* **55**: 145–156.

Gillespie, J. H. 2000. Genetic drift in an infinite population: the pseudohitchhicking model. *Genetics* **155**: 909–919.

González, A.M., Cabrera, V.M., Larruga, J.M. & Gullón, A. 1983. Molecular variation in insular endemic *Drosophila* species of the Macaronesian archipelagos. *Evolution* **37**: 1128–1140.

González, A.M., Hernández, M., Volz, A., Pestano, J., Larruga, J.M., Sperlich, D. & Cabrera, V.M. 1990. Mitochondrial DNA evolution in the *obscura* species subgroup of *Drosophila*. *J. Mol. Evol.* **31**: 122–131.

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

Hill, W.G. & Robertson, A. 1968. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **38**: 226–231.

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. & 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., 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. W. Munro, ed.), pp. 21–120. Academic Press, New York.

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

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

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

Kliman, R.M. & Hey, J. 1993. DNA sequence variation at the *period* locus within and among species of the *Drosophila melanogaster* complex. *Genetics* **133**: 375–387.

Krimbas, C.B. 1992. The inversion polymorphism of *Drosophila subobscura*. In: *Drosophila Inversion Polymorphism* (C. B. Krimbas & J. R. Powell, eds), pp. 127–220. CRC Press, Boca Raton.

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.

Lewontin, R.C. 1974. *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York.

Maddison, W.P. & Maddison, D.R. 1992. *Macclade: Analysis of Phylogeny and Character Evolution*, Version 3.0. Sinauer, Sunderland, Mass.

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

Monclús, M. 1984. Drosophilidae of Madeira, with the description of *Drosophila madeirensis* n. sp. *Z. f. Zool. System. u. Evol.* **22**: 94–103.

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

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.

Pinto, F.M., Brehm, A., Hernández, M., Larruga, J.M., González, A.M. & Cabrera, V.M. 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.

Press, J.R. & Short, M.J. 1994. *Flora of Madeira*. HMSO, London.

Prevosti, A. 1971. Chromosomal polymorphism in *Drosophila subobscura* Coll. populations from the Canary Islands. *Genét. Ibér.* **23**: 69–84.

Prevosti, A. 1972. Chromosomal polymorphism in *Drosophila subobscura* populations from the Madeira island. *Genét. Ibér.* **24**: 11–21.

Prevosti, A. 1974. Chromosomal inversion polymorphism in the southwestern range of *Drosophila subobscura* distribution area. *Genetica* **45**: 111–124.

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.

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. & Rozas, R. 1999. DnaSP, Version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**: 174–175.

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.

Rozas, J., Segarra, C., Zapata, C., Alvarez, G. & Aguadé, M. 1995. Nucleotide polymorphism at the *rp49* region of *Drosphila subobscura*: Lack of geographic subdivision within chromosomal arrangements in Europe. *J. Evol. Biol.* **8**: 355–367.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. & Erlich, H.A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable polymerase. *Science* **239**: 487–491.

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

Swofford, D.L. 1998. *Paup: Phylogenetic Analysis Using Parsimony*, Version 4.0. Sinauer Associates, Inc., Sunderland, MA.

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

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. *Nucl. 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.

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

Weir, B.S. 1996. *Genetic Data Analysis II*. Sinauer Associates, Inc., Sunderland, MA.

Wright, S. 1951. The genetical structure of populations. *Ann. Eugen.* **15**: 323–354.

*Received 29 January 2001; accepted 19 March 2001*