Short Communication Phylogeography of Drosophila subobscura from north Atlantic islands inferred from mtDNA A + T rich region sequences

Article · July 2001

7 authors, including:

António Brehm
Universidade da Madeira
189 PUBLICATIONS   **3,125** CITATIONS

SEE PROFILE

David James Harris
University of Porto
576 PUBLICATIONS   **9,894** CITATIONS

SEE PROFILE

Mariano Hernandez Ferrer
Universidad de La Laguna
124 PUBLICATIONS   **2,362** CITATIONS

SEE PROFILE

Francisco Pinto
Scientific Research Centre "Isla de la Cartuja"
99 PUBLICATIONS   **3,402** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   PATHOSPILL: using an integrative approach to investigate the consequences of pathogen spill-over between wild and farmed fish species in a key estuarine habitat
View project

Project   Modelling terrestrial biodiversity in the Cape Verde islands: combining ecogeographical information with molecular markers View project

MOLECULAR
PHYLOGENETICS
AND
EVOLUTION

Short Communication

Phylogeography of *Drosophila subobscura* from north Atlantic islands inferred from mtDNA A + T rich region sequences

António Brehm,[a,*] D. James Harris,[b] Mariano Hernández,[c] José A. Perez,[c] José M. Larruga,[c] Francisco M. Pinto,[c] and Ana M. González[c]

[a] *Centre of Macaronesian Studies, University of Madeira, Campus of Penteada, 9000-390 Funchal, Portugal*
[b] *CIBIO, ICETA-U.P., Campus Agrario de Vairão, 4485-661 Vila do Conde, Portugal*
[c] *Department of Genetics, University of La Laguna, Tenerife 38271, Spain*

Received 20 February 2003; revised 29 September 2003

1. Introduction

Drosophila subobscura has been used extensively in the past, in numerous studies focused on the colonization history of natural populations (Afonso et al., 1990; Brehm and Krimbas, 1987; Khadem et al., 1998; Krimbas, 1993; Latorre et al., 1992; Pascual et al., 2001; Pinto et al., 1997; Prevosti, 1974; Prevosti et al., 1984). A native Paleartic with circum-Mediterranean distribution, the species reaches the southern limits of its distribution in Morocco and is also present in the North Atlantic Archipelagos of Azores, Canaries and Madeira. In the two later archipelagos an ancestral lineage gave rise to the endemics *Drosophila guanche* and *Drosophila madeirensis*, respectively (González et al., 1983; Gonzalez et al., 1990; Krimbas and Loukas, 1984; Monclús, 1976, 1984) all together forming the *subobscura* cluster. A recent combined analysis of nucleotide sequences indicated that *D. madeirensis* is the sister taxa of *Drosophila subobscura* (O'Grady, 1999) and this was supported by mtDNA A + T rich region sequences (Brehm et al., 2001). The vast geographical distribution of *D. subobscura* makes it an excellent species for phylogeographic studies assessing past colonization events. It is known that *D. subobscura* from Crete resembles much more populations from Israel and Tunisia than those from mainland Greece, probably due to the last European glaciation which did not extend into the Middle East and North Africa (Krimbas, 1993). Extant North Atlantic island populations are supposed to be long isolated from those of the mainland and were not greatly affected by the last glaciation periods acting as refuge for many plant and animal species. Inversion chromosomal polymorphism of *D. subobscura*, suggest that the Iberian Peninsula (and also the Balkans) may have had an important role in repopulating Europe (Krimbas, 1993; Menozzi and Krimbas, 1991). Based on chromosome inversion polymorphism frequencies the populations from Azores and Madeira were found to differ from the Canary ones but clustered together, and were highly differentiated from Iberian Peninsula populations (Brehm and Krimbas, 1987). Alternatively, autosomal allozyme loci did not discriminate Madeira from the mainland populations (Larruga et al., 1993; Pinto et al., 1997), while Canary Islands populations appeared as a distinct lineage. Using combined chromosomal inversion data and sequences of the *rp49* gene, Khadem et al. (1998) hypothesized that the *D. subobscura* group ancestral lineage invaded twice, but separately, the Canary Islands and Madeira. The first colonizers differentiated in the islands, giving rise to the endemics. MtDNA-RFLP analysis suggested a different scenario, with two Canary Islands (La Palma and El Hierro) more related to continental populations, but the other Canary Islands populations were generally distinct (Pinto et al., 1997). Madeira populations could not be differentiated from Continental ones.

In order to gain additional insight into the colonization and phylogeography of the Atlantic Islands by *D. subobscura* populations, we have sequenced the mtDNA A + T rich region of individuals belonging to populations from Morocco, the Iberian Peninsula and the Atlantic Islands of Azores, Madeira, and the Canaries. This region has proven to be a good marker for intraspecific studies in *D. subobscura* (Brehm et al., 2001),

* Corresponding author. Fax: +351-291-705399.
E-mail address: brehm@uma.pt (A. Brehm).

and can then be compared to the previously published chromosomal and nuclear markers.

2. Materials and methods

2.1. Samples

Thirty-three individuals of *D. subobscura* belonging to 12 populations from the Iberia Peninsula and North Atlantic islands have been sequenced for their entire mtDNA A + T rich region (Fig. 1). In order to gain additional information on haplotype variability, we obtained an additional 103 partial sequences belonging to sub-regions R7 and R8 of the A + T rich region from *D. subobscura* (see below) corresponding to a fragment of 469 bp long. In this case we included individuals sampled in two more islands from the Azores archipelago. Population codes and number of individuals sequenced for the entire A + T rich region are described in Fig. 1 and Table 1. Partial sequences are summarized in an Annex provided as complementary material in www.uma.pt/bioarticles. We included as outgroups the two other species from the *subobscura* cluster: *D. madeirensis* and *D. guanche* for which the complete A + T rich region sequences were previously published (Brehm et al., 2001).

2.2. Sequencing of mtDNA

Total DNA was extracted from individual adult flies by the CTAB method (Towner, 1991) with minor modifications. Amplification and sequencing of this region is complicated by the extremely high A + T content (ca. 93%). It was therefore necessary to amplify and sequence the region using 6 pairs of primers (R1, R11, R2, R10, R7, and R8) corresponding to six consecutive regions named after the primers names, including the control and adjacent mtDNA regions. Description of the primers and their localization in the mtDNA control region of *D. subobscura* (A + T rich region) as well as PCR conditions and methodology has been described elsewhere (Brehm et al., 2001). The complete sequences included in the present study have been deposited under Accession Nos. AY442192–AY442224.

2.3. Data analysis

Multiple sequence alignments were performed with CLUSTAL X (Thompson et al., 1997) and later maximized for sequence similarity by visual inspection. All sequences were compared to the subR complete A + T rich region already published (AJ132900, Brehm et al., 2001). In those cases where indels were found but were non-existent in the published *subobscura* consensus sequence, they were given a small case annotation (e.g., 586a) next to the closest nucleotide in the published sequence. We used the statistical parsimony algorithm (Templeton et al., 1992) performed in TCS (Clement et al., 2000) to estimate the maximum number of differences among haplotypes as a result of a single substitution with a 95% confidence level, and also the most probable ancestral haplotype. Two median-joining networks (Bandelt et al., 1995, 2000) were constructed to represent the phylogeographic relationships among haplotypes (Fig. 2). The first network was restricted to the full sequences and included both species from the *subobscura* cluster (*D. madeirensis* and *D. guanche*) as outgroups. Because of the extensive mutations characterizing these two species in relation to *D. subobscura*, we limited the comparison to those sites that were variable in *D. subobscura*, thus excluding mutated sites typical of the outgroups. The second network was based on partial sequences from the Annex. We also performed phylogenetic analyses, using maximum likelihood and Bayesian inference, following the methodology described in detail in Brehm et al. (2003).

3. Results and discussion

The complete sequence of a 977 bp segment that includes 10 bp of the tRNA^{leu}, the complete control region and 14 bp of the srRNA of the mtDNA of *D. subobscura*, corresponding to base pairs 14929 to 10 in *D. yakuba* (AC NC001322, Clary and Wolstenholme, 1985), was obtained for a total of 33 individuals. Fifty-one polymorphisites were identified in the region studied (Table 1) of which 19 are indels. Positions 52 and 607 showed both transitions and transversions, and positions 588–597 showed two different types of insertions.
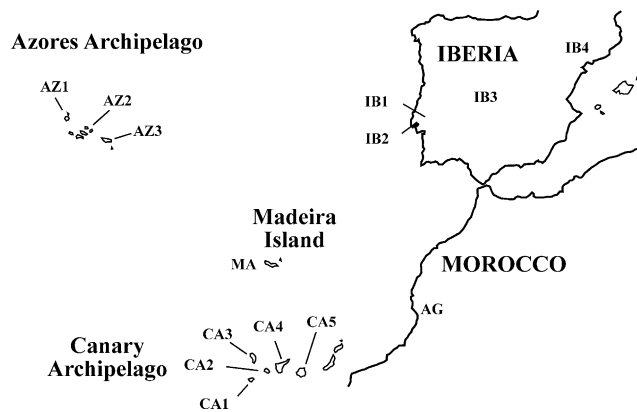

Fig. 1. Geographic localization of the *Drosophila subobscura* samples used in this study. Population codes are as follows: AZ1—Flores Isl., AZ2—Terceira Isl., AZ3—São Miguel Isl., MA—Madeira Island, CA1—El Hierro Isl., CA2—La Gomera Isl., CA3—La Palma Isl., CA4—Tenerife Isl., CA5—Gran Canaria Isl., IB1—Boquilobo, IB2—Lisboa, IB3—San Lorenzo del Escorial, IB4—Barcelona, and AG—Agadir (Morocco).

Table 1
A + T rich region haplotypes found in 33 complete sequences of *D. subobscura* specimens from 12 localities

Nucleotide position headers (read vertically by region):
- R1: 1 1 / 5 8
- R11: 1 1 1 / 5 5 5 7 8 2 4 5 / 0 1 2 0 7 8 5 3
- R2: 3 3 / 1 2 / 0 9
- R10: 4 4 4 4 4 4 5 / 3 3 3 5 5 9 0 / 4 4 7 5 6 1 9 (a b)
- R7: 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 / 0 1 7 7 7 8 8 8 8 8 8 8 8 8 8 8 9 9 0 1 4 4 5 6 / 7 3 4 8 9 0 4 6 6 6 6 6 6 6 6 6 9 5 7 2 2 2 6 2 8 (a b c d e f g h i j)
- R8: 7 8 8 9 9 9 / 7 5 7 0 0 4 / 1 9 0 1 5 0

Haplotype	R1	R11	R2	R10	R7	R8	AZ	MA	CA1	CA2	CA3	CA4	CA5	AG	IB1	IB2	IB3	IB4	TOTAL
Hap1	G A	A T T T T A T T	T T	A A A - - C G	T T A T A A A - - - - - - - - - - - A A T A A T T - A	A A T T T T	2	4			1			1	1		1	1	10
Hap2	. G	. . . . . . A .	. .	. . . .	. . . .	. . A . . .	1												1
Hap3	. .	. . . .	. .	. . . .	. . . . T	. . . . . .	1												1
Hap4	. .	. A	C .	- - G T A T A	. A . . . . T C . . . C . . T	G G . C . C						1	1						2
Hap5	. .	. . . .	C .	. . . .	. A . . . . C . . . C . . T	G G . C . C							1						1
Hap6	. .	. . . .	C .	. . . .	. . . . T . . . . T	G G . C . C							1						1
Hap7	. .	G A A . A . . G	C .	- - G T A T A	. A . . . . T T . . . C . . T	G G . C . C						1							1
Hap8	. G	. A A	C .	. . . .	. A . . . . C . . . C . . T	G G . C . C				1									1
Hap9	. .	. A	C .	- - G T A T A	. A . . . . C . . . C . . T	G G . C . C			1										1
Hap10	. .	. . . .	. .	- - G T A T .	. . T . . . . - . C . . T	. . . . . .	1												1
Hap11	. .	. . . . . . G .	. .	. . . .	A . . - - - - . . . . . . T A T A . . . . - . C . . T	. . . C . C	1												1
Hap12	. .	. . . .	. .	. . . .	. . . . T A T A . . . . - . C . . T	. . . C C C			1										1
Hap13	. .	. . . .	. .	- - G T A T .	. . . .	. . . . . .			1								1	1	3
Hap14	. .	. A	. .	- - G T A T A	. . . . A	. . . . . .												1	1
Hap15	A .	. . . A	. .	- - G T A T A	. . . .	. . . . . .										1			1
Hap16	. .	. . . A	. .	. . . .	. . . . C . A	. . . . . .										1			1
Hap17	A .	. . . .	. C	- - G T A T .	. . . . C . . T	. . . C . .									1				1
Hap18	A .	. . . .	. .	- - G T A T .	. A . . . . C . . T	G . . C . C									1				1
Hap19	A .	. . . .	. C	- - G T A T .	. . . . . . T T A T A T A T A T A . . . - . C . . T	. . . C . .									1				1
Hap20	. .	. A	. .	- - G T A T A	. A . . . . C . . . C . C	G . . C . C								1					1
Hap21	. .	. . G . A . . G	. .	. . . .	. . . . A	. . . . . .									1				1

The codes for 1 the regions are as in Fig. 1. Nucleotide numbers and sub- regions R1–R8 refer to the complete A + T rich region sequence of *D. subobscura* 2 and have been described elsewhere (AC AJ132899–132900, Brehm et al., 2001). A dot indicates the same base as in haplotype 1. Indels are 3 indicated by a dash.
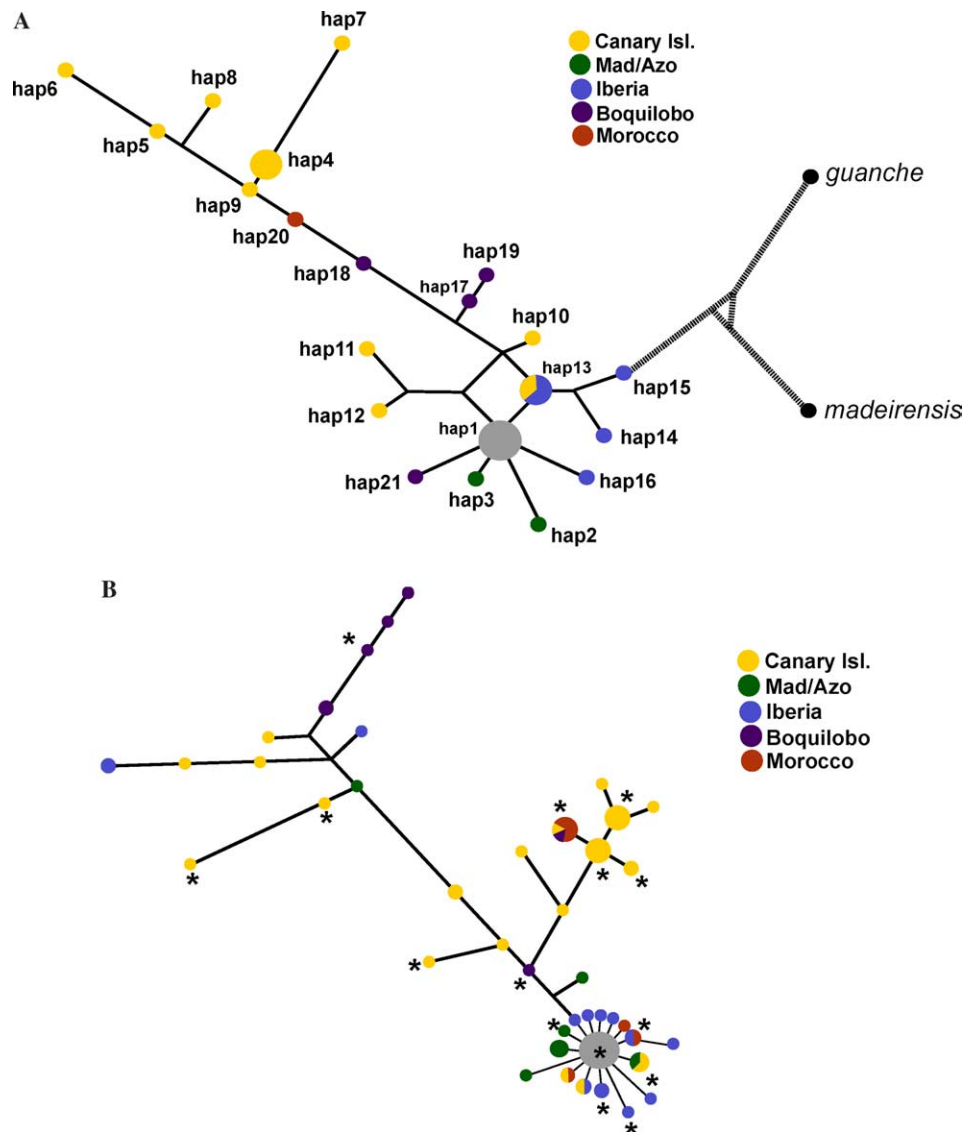
Fig. 2. (A) Median-joining network of 21 haplotypes of the 33 complete A + T rich region sequences included in the present work. *Drosophila guanche* and *D. madeirensis* were also included but only in relation to variable positions common to *D. subobscura*. Indels were disregarded. Thick branches to the outgroups are not in proportion to mutational steps. (B) Network constructed with all partial sequences included in the annex (segments R7 and R8). Individuals with complete sequences are indicated with an asterisk. The grey circle is out of scale and includes individuals belonging to all regions sampled. This circle is estimated to be the ancestral haplotype using TCS.

Two microsatellites occur within this region. Altogether the 33 individuals yielded 21 haplotypes. Haplotype 1 is the most common, represented by 10 individuals and present in all five regions analysed (the three Archipelagos, Morocco, and Iberia). The other 20 haplotypes are region-specific. Except haplotype 1, only two haplotypes are found in more than one locality (haplotypes 4 and 13). The most parsimonious network of haplotypes is shown in Fig. 2A. Using modeltest we estimated that the most appropriate model of evolution for this data was the HKY85 model (Ts/Tv ratio 1.68) with an estimate of invariable sites (0.63) and a discrete approximation of the gamma distribution (0.39). The estimate of relationships derived from the Bayesian analysis was very

similar, with no nodes with even moderate support (50%) in conflict.

Partial segments from fragments R7 and R8 were also obtained adding additional individuals belonging to 14 populations. These segments uncovered additional variable positions than the previously detected with the complete sequences. The data is summarized in the Annex and contains the sample number and the number of fragments sequenced for each of these 3 sub-regions. Sub-region R7 uncovered a total of 34 haplotypes with 34 variable sites excluding indels and R8 showed 26 different haplotypes based on 19 variable positions.

Our results, when compared to the previously published chromosomal inversions and allozyme data sets,

show how different markers can indicate extremely different colonization patterns. Using autosomal loci the populations from the Canary Islands form a single well supported (99% bootstrap) monophyletic unit. The same is true when chromosomal rearrangements are analysed (Pinto et al., 1997). However RFLP analysis of mtDNA clustered populations from two of the Canary Islands, El Hierro, and La Palma, with continental populations. Similarly our data set groups the samples from El Hierro and La Palma (CA1, 3) among continental haplotypes, while sequences from the other island (CA2, 4, 5) are more genetically distinct. Based solely on the mtDNA evidence for the Canary Islands, we would suggest that they had been colonized by two different waves, with El Hierro and La Palma being colonized only more recently, and Tenerife, La Gomera, and Gran Canaria much earlier. However, with the evidence from the nuclear markers it seems more likely that the islands were all colonized during the first wave, and that later gene flow with mainland forms has introduced the continental type mtDNA into the populations of El Hierro and La Palma. Although we only included four individuals from El Hierro and La Palma, approximately 70 individuals were included for the RFLP analysis (Pinto et al., 1997), and all had the same patterns. Thus it seems likely that the continental type mtDNA is now fixed, or at least widespread in these islands.

Populations from Madeira island are much more similar to the continental forms based on nuclear markers, and also have similar A + T rich mtDNA haplotypes. In this case all the markers indicate a relatively recent colonization of the island. However, there are some differences, both in allele frequencies and variation at the rp49 gene (Khadem et al., 1998), thus an anthropogenic introduction is unlikely. The two individuals from the Azores both share the commonest continental haplotype. This could mean that these populations were introduced. Similarly, nuclear enzymatic polymorphisms are not highly differentiated between the Azores and Morocco (Larruga et al., 1993). The fact that there is no separation of Iberian from Moroccan haplotypes in the network weakens the hypothesis that the Strait of Gibraltar was a strong barrier to gene flow in this species as deduced from chromosomal data (Prevosti et al., 1975). Furthermore, the mainland haplotypes most closely related to older Canary Islands haplogroup are from Morocco and from Boquilobo. Chromosome rearrangement data clearly relate the Canary Islands to North Africa rather than to the Iberian Peninsula. It would be very interesting to know the chromosomal polymorphism in Boquilobo, as it is in an area of Portugal that could have acted as glacial refugia, in order to reconcile these discrepant results. The highest level of variation within the Canary Islands (excluding the more recent haplotypes on El

Hierro and La Palma) is almost 2%. Assuming a *Drosophila* molecular clock estimated at 1.7% per million years (Caccone et al., 1988a,b), this would imply that the original radiation within the Canary Islands was at least 1.1 MYA, probably more since more divergent haplotypes many not have been sampled or may have gone extinct. This estimate falls within the range of between 0.7 and 2.8 MYA that was estimated by Khadem et al. (1998). To conclude, the Canary Islands have become a model region for studying colonization patterns (Juan et al., 2000). However most phylogenetic analyses have relied on a single marker, mtDNA. At least in the case of *D. subobscura*, colonization patterns derived from this marker are not completely congruent with those derived from nuclear markers. By examining mtDNA and nuclear markers it is clear that the colonization patterns are even more complex than previously proposed. The same may well be true for other organisms.

References

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

Bandelt, H.-J., Forster, P., Sykes, B.C., Richards, M.B., 1995. Mitochondrial portraits of human populations using median networks. Genetics 141, 743–753.

Bandelt, H.-J., Macaulay, V., Richards, M.B., 2000. Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human *mt*DNA. Mol. Phylogenet. Evol. 16, 8–28.

Brehm, A., Krimbas, C., 1987. The inversion polymorphism of *Drosophila subobscura* natural populations from Portugal. Genét. Ibér. 39, 235–248.

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

Brehm, A., Jesus, J., Spínola, H., Alves, C., Vicente, L., Harris, D.J., 2003. Phylogeography of the Madeiran endemic lizard *Lacerta dugesii* inferred from mtDNA sequences. Mol. Phylogenet. Evol. 26, 222–230.

Caccone, A., DeSalle, R., Powell, J., 1988a. Calibration of the change in thermal stability of DNA duplexes and degree of base pair mismatch. J. Mol. Evol. 27, 212–216.

Caccone, A., Amato, G., Powell, J., 1988b. Rates and patterns of scnDNA and mtDNA divergence within the *Drosophila melanogaster* subgroup. Genetics 118, 671–683.

Clary, D.O., Wolstenholme, D.R., 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization and genetic code. J. Mol. Evol. 22, 252–271.

Clement, M., Posada, D., Crandall, K.A., 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9, 1657–1660.

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 (6), 1128–1140.

Gonzalez, A., Hernandez, M., Volz, A., Pestano, J., Larruga, J., Sperlich, D., Cabrera, V., 1990. Mitochondrial DNA evolution in the *obscura* species subgroup of *Drosophila*. J. Mol. Evol. 3, 122–131.

Juan, C., Emerson, B., Oromi, I., Hewitt, G., 2000. Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. Trends Ecol. Evol. 15, 104–109.

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 *rp*49 gene region. J. Evol. Biol. 11, 439–452.

Krimbas, C., 1993. in: Dr. Kovac (Ed.), *Drosophila subobscura*: Biology, Genetics and Inversion Polymorphism. Verlag, Hamburg.

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

Larruga, J.M., Rozas, J., Hernández, M., González, A.M., Cabrera, V.M., 1993. Latitudinal differences in sex chromosome inversions, sex linked allozymes, and mitochondrial DNA variation in *Drosophila subobscura*. Genetica 92, 67–74.

Latorre, A., Hernandez, C., Martinez, D., Castro, J.A., Ramón, M., Moya, A., 1992. Population structure and mitochondrial DNA gene flow in Old World populations of *Drosophila subobscura*. Heredity 68, 15–24.

Menozzi, P., Krimbas, C., 1991. The inversion polymorphism of *Drosophila subobscura* revisited: synthetic maps of gene arrangement frequencies and their interpretation. J. Evol. Biol. 5, 625–641.

Monclús, M., 1976. Distribución y ecología de drosofílidos en España. II. Especies de *Drosophila* de las Islas Canarias, con la descripción de una nueva especie. Bol. R. Soc. Española Hist. Nat. (Biol.) 74, 197–213.

Monclús, M., 1984. Drosophilidae of Madeira, with the description of *Drosophila madeirensis* n. sp. Sonderdruck aus Z. F. Zool. Systematik u. Evolutionsforschung 22, 94–103.

O'Grady, P., 1999. Reevaluation of phylogeny in the *Drosophila obscura* species group based on combined analysis of nucleotide sequences. Mol. Phylogenet. Evol. 12, 124–139.

Pascual, M., Aquadro, C., Soto, V., Serra, L., 2001. Microsatellite variation in colonizing and Paleartic populations of *Drosophila subobscura*. Mol. Biol. Evol. 18, 731–740.

Pinto, F., Brehm, A., Hernandez, M., Larruga, J.M., Gonzalez, A.M., Carera, V., 1997. Population genetic structure and colonization sequence of *Drosophila subobscura* in the Canaries and Madeira Atlantic Islands as inferred by autossomal, Sex-linked and mtDNA traits. J. Hered. 88, 108–114.

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

Prevosti, A., Ocaña, J., Alonso, G., 1975. Distances between populations of *Drosophila subobscura*, based on chromosome arrangements frequecies. Theor. Appl. Genet. 45, 231–241.

Prevosti, A., de Frutos, R., Alonso, G., Latorre, A., Monclús, M., Martinez, M.J., 1984. Genetic differentiation between natural populations of *Drosophila subobscura* in the western mediterranean area with respect to chromosomal variation. Genet. Select. Evol. 16, 143–156.

Templeton, A.R., Crandall, K.A., Sing, C.F., 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III Cladogram estimation. Genetics 132, 619–633.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmongin, F., Higgins, D.G., 1997. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24, 4876–4882.

Towner, P., 1991. Purification of DNA. In: Brown, T.A. (Ed.), Essential Molecular Biology. A Practical Approach, vol. I. IRL Press, Oxford.