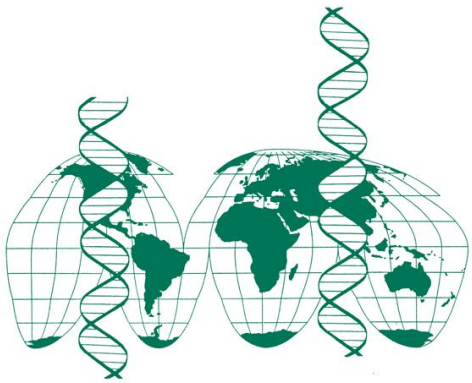
ISSN 0016-6707, Volume 138, Number 8

CODEN GENEA3 Vol. 138 No. 8 August 2010 ISSN 0016 6707

Genetica

An International Journal
of Genetics and Evolution



Springer

Genetica (2010) 138:819–829
DOI 10.1007/s10709-010-9464-x

ORIGINAL RESEARCH

Nucleotide diversity of a ND5 fragment confirms that population expansion is the most suitable explanation for the mtDNA haplotype polymorphism of *Drosophila subobscura*

José A. Castro · Eladio Barrio · Ana González · Antònia Picornell · Maria Misericòrdia Ramon · Andrés Moya

Received: 24 June 2009 / Accepted: 17 May 2010 / Published online: 18 June 2010

Abstract Results from mitochondria (mt) DNA restriction site analyses (RSAs) have revealed that wild populations of *Drosophila subobscura* are formed by two common (I and II) and some rare, often endemic, low-frequency haplotypes. In the study reported here, we analysed nucleotide diversity in a 942-bp fragment of the mtDNA ND5 gene in 48 *D. subobscura* individuals captured from three populations that showed haplotypes I, II or the less common ones, as well as in one additional individual belonging to *D. guanche* that was taken as an outgroup. RSAs and sequencing results were compared. The two approaches yielded similar nucleotide variability parameters, suggesting a consistency in the results obtained from mtDNA dynamics in natural populations of *D. subobscura*. Patterns of polymorphism at ND5 are most consistent with the hypothesis of population expansion after a bottleneck that may have occurred since the last glaciation or which may occur seasonally after the summer and winter. However, we cannot rule out that selection has a role in maintaining the two major haplotypes at intermediate frequencies in worldwide populations of *D. subobscura*.

Keywords ND5 · Nucleotide diversity · Restriction site analysis · mtDNA haplotypes · Population expansion · *D. subobscura* · *D. guanche*

J. A. Castro (✉) · A. Picornell · M. M. Ramon
Laboratori de Genètica, Departament de Biologia, Edifici Guillem Colom, Facultat de Ciències, Universitat de les Illes Balears, Campus de la UIB, Palma de Mallorca 07122, Balearic Islands, Spain
e-mail: jose.castro@uib.es

E. Barrio · A. González · A. Moya
Institut "Cavanilles" de Biodiversitat i Biologia Evolutiva, Universitat de València, Apartado de Correos 22085, 46071 Valencia, Spain

A. Moya
Centre Superior d' Investigació en Salut Pública (CSISP), Av. de Catalunya 21, 46020 Valencia, Spain

A. Moya
CIBER en Epidemiología y Salud Pública (CIBEResp), Valencia, Spain

Introduction

Drosophila subobscura is a Palearctic species of the obscura subgroup of *Drosophila*. It is distributed over most of Europe, the Middle East, northern Africa and the Atlantic islands of the Azores, Madeira and the Canaries and has recently colonized the American continent (Ayala et al. 1989; Rozas et al. 1990). Patterns of polymorphism at mitochondrial restriction digest sites reveal widespread geographical homogeneity in the Old and New World populations of *D. subobscura* with a high prevalence of two main, co-existing mitochondrial DNA (mtDNA) haplotypes, denoted I and II, respectively, and lower frequencies of less common mtDNA haplotypes derived from the two main ones (Latorre et al. 1986; Rozas et al. 1990; Latorre et al. 1992; García-Martínez et al. 1998; Castro et al. 1999; Oliver et al. 2002, 2005). The only exception to date has been detected in the Canary Islands, where an endemic haplotype is predominant on some islands (Tenerife, La Gomera and Gran Canaria), while the main one on El Hierro and La Palma is haplotype II (Pinto et al. 1997).

We have conducted a number of studies on the population dynamics of these haplotypes in *D. subobscura* in an attempt to answer the question as to why the two main

haplotypes are in such quasi equi-frequent proportions in nature and what evolutionary forces are responsible for this phenomenon. To answer this question, we studied the involvement of the nuclear genome at two different levels: nuclear enzymes (Latorre et al. 1992; Castro et al. 1999) and chromosomal inversions (Oliver et al. 2002, 2005). The analyses of the corresponding cytonuclear disequilibria were not conclusive because when such disequilibria were detected, they were transient. We then measured the fitness components under experimental conditions and studied fitness differences between haplotypes I and II in terms of mating behaviour (Castro et al. 2003), when *D. subobscura* were competing in experimental population cages (García-Martínez et al. 1998; Oliver et al. 2005), or for a number of life history traits (Christie et al. 2004). Although we were able to determine that haplotype II showed a higher fitness than haplotype I, in most cases this difference was not statistically supported. In addition, we could not conclude from such observations whether haplotype II was fitter than I under natural conditions.

Nucleotide diversity can also provide some clues about the forces governing the evolution of mtDNA in *D. subobscura*. Moya et al. (1993) studied nucleotide diversity in six functional mtDNA regions (2,377 bp, 15% of the mtDNA genome) in haplotypes I and II and found only three nucleotide differences. These differences proved to be silent changes at the protein level, one of which corresponded to the *Hae*III restriction site that distinguishes haplotypes I and II, which is located on the ND5 gene. On the basis of these results, the authors cautiously proposed that mtDNA haplotypes I and II are selectively neutral and that the corresponding individuals are phenotypically equivalent.

The aim of the study reported here was to analyse patterns of mitochondrial polymorphism at ND5 within the two main *D. subobscura* haplotype groups. The results of this analysis are consistent with those of our previous studies and support a model of population expansion after a bottleneck.

Materials and methods

Drosophila samples

A total of 48 *D. subobscura* isofemale lines were derived from individuals captured in the Balearic Islands of Majorca (17 individuals) and Minorca (25 individuals) and La Canyada (Valencia, 6 individuals). A *Drosophila guanche* female captured in Anaga, Tenerife (Canary Islands) was used to found an isofemale line and was taken as outgroup.

Extraction and digestion of mtDNA

An enriched fraction of mtDNA was obtained following the methodology of Latorre et al. (1986). This mtDNA was digested with five restriction enzymes, of which three (*Eco*RI, *Eco*RV and *Hin*dIII) recognize 6-bp sequences and two (*Hpa*II and *Hae*III) recognize 4-bp sequences. These enzymes were selected for their capability to detect mtDNA polymorphisms (Afonso et al. 1990; Latorre et al. 1992; Castro et al. 1999). The fragments obtained by digestion were separated on horizontal 0.8–2.0% agarose gels, and λ DNA digested with *Hin*dIII–*Eco*RI was used as a size standard to determine fragment size. The gels were stained with 0.1 mg mL^{-1} ethidium bromide and visualized with a 260-nm UV light transilluminator. The restriction map was obtained by means of all possible single and double digestions of the mtDNA, and the different restriction patterns obtained, using a given enzyme and the haplotypes, were named according to the notation of Latorre et al. (1986; 1992).

Sequencing

A 942-bp fragment corresponding to the ND5 region of mtDNA was PCR amplified and sequenced in the 48 *D. subobscura* individuals. The primers designed were the following:

ND5SU5A: 5′-GCTATAGCTAGCCCCTACAC
ND5SU3A: 5′-TGACCAGCTAGCTATTCTGAT

The PCR cycling conditions consisted of an initial denaturing at 95°C for 2–3 min, 33 cycles at 95°C for 1 min (denaturing), 52°C for 1 min (annealing) and 72°C for 1 min (extension), with a final extension at 72°C for 7 min. The QIAquick PCR Purification kit (QIAGEN, Hilden, Germany) was used. Both heavy and light strands were sequenced on an automated ABI PRISM 3100 sequencer using the ABI Prism Terminator BigDye Cycle Sequencing Reaction kit (Applied Biosystems, Foster City, CA) v.1.1, following the manufacturer's instructions. The sequences were aligned using the software BioEdit v.7.0.5.2 (Hall 1999).

Presence of *Wolbachia*

To exclude an incompatibility system in *D. subobscura* promoted by the presence of *Wolbachia*, a PCR assay using 16S rDNA *Wolbachia*-specific primers was carried out according to García-Martínez et al. (1998). We analysed several isofemale lines from Majorca, Minorca and La Canyada populations, and all were negative. To date, no kind of cytoplasmic incompatibility in the populations, such as embryo mortality or production of all-male progeny in crosses with different isofemale lines, has every

been detected. Consequently, we can reasonably accept that the flies used in the experiments do not have this endosymbiont.

Statistical analyses

Estimation of gene genealogies from mtDNA sequences was obtained using TCS software (Clement et al. 2000), where sequences were collapsed into haplotypes in the cladogram estimation (statistical parsimony). The demographic history of each population and of the total was examined using two different approaches. First, to test whether the frequency spectrum of mutations conformed to the expectations of the standard neutral model, we calculated the values of four statistic tests: Tajima's D (Tajima 1989), Fu and Li's D–F (Fu and Li 1993) and Fay and Wu's H (Fay and Wu 2000). Second, the distributions of the observed and expected pairwise differences (mismatch distributions) were calculated to examine changes in population size (Rogers and Harpending 1992). This latter method is based on two parameters: (1) $\theta = 2Nu$ is the expected pairwise differences; (2) $\tau = 2ut$ is the date of the growth measured in units of mutational time. These parameters, as well as those of genetic diversity, were estimated with DnaSP v.5.10 software (Librado and Rozas 2009). The differentiation based on the restriction sites within and between populations (Vw and Vb, respectively) and the degree of population subdivision (N_{ST}) were estimated according to Lynch and Crease (1990).

The test for the neutral mutation hypothesis proposed by McDonald and Kreitman (1991) was also applied. This method contrasts the patterns of within-species polymorphism and between-species divergence at synonymous (silent) and nonsynonymous (replacement) sites in the coding region of a gene.

Results

Table 1 shows the mtDNA haplotypes of the study populations that were obtained with the restriction enzymes used. As *Eco*RI was found to be monomorphic in these samples, it was not included in the analysis. The network connecting the 18 haplotypes is shown in Fig. 1 (Top). In accordance to what has been reported in earlier published studies, we found the two major haplotypes, I and II, to be differentiated by a mutation in ND5 at the *Hae*III restriction site (e3) and the less common haplotypes derived from haplotypes I and II to be differentiated through one or two mutational steps.

Table 1 *Drosophila subobscura* mitochondrial (mt) DNA haplotypes derived from restriction site analyses

Haplotypes	*Eco*RV	*Hae*III							*Hin*dIII				*Hpa*II						Majorca	Minorca	La Canyada	Total
	k1	e1	e2	e3	e4	e5	e6	e7	f2	f3	f5	f6	h2	h3	h4	h5	h7	h8				
I	−	+	−	+	−	−	−	−	−	−	−	−	−	+	−	−	+	−	4	7	1	12
II	−	+	−	−	−	−	−	−	−	−	−	−	−	+	−	−	+	−	5	10	1	16
III	−	+	−	+	−	−	−	−	−	−	−	−	−	+	−	+	+	−	1	0	0	1
IV	−	+	−	−	−	−	−	−	−	−	−	−	−	+	−	+	+	−	1	2	1	4
V	−	+	−	−	−	−	−	−	−	−	+	−	−	+	−	−	+	−	1	0	0	1
VI	−	+	−	+	−	−	−	−	−	−	−	+	−	+	−	−	+	−	1	0	0	1
VII	+	+	−	−	−	−	−	−	−	−	−	−	−	+	−	−	+	−	1	0	0	1
VIII	−	+	−	−	−	−	−	−	+	−	−	−	−	+	−	−	+	−	1	0	0	1
IX	−	+	−	−	−	−	−	−	−	−	−	−	−	+	−	−	+	−	1	0	0	1
X	−	+	−	−	+	−	−	−	−	−	−	−	−	+	−	−	+	−	1	0	0	1
XI	−	+	−	−	−	−	−	−	−	+	−	−	−	+	−	−	+	−	0	1	0	1
XII	−	+	−	+	−	+	−	−	−	−	−	−	−	+	−	−	+	−	0	1	0	1
XIII	−	+	+	−	−	−	−	−	−	−	−	−	−	+	−	−	+	−	0	2	0	2
XIV	−	+	−	+	−	−	+	−	−	−	−	−	−	+	−	−	+	−	0	1	0	1
XV	−	+	−	−	−	−	−	−	−	−	−	−	−	−	+	−	+	−	0	1	0	1
XVI	−	+	−	+	−	−	−	−	−	−	−	−	+	+	−	−	+	−	0	0	1	1
XVII	−	+	−	−	−	−	−	+	−	−	−	−	−	+	−	−	+	−	0	0	1	1
XVIII	−	+	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	+	0	0	1	1
Total																			17	25	6	48

Each haplotype is defined by the polymorphic sites, and the number of individuals showing a given haplotype in the populations from Majorca, Minorca and La Canyada is also shown

Only the polymorphic sites are indicated

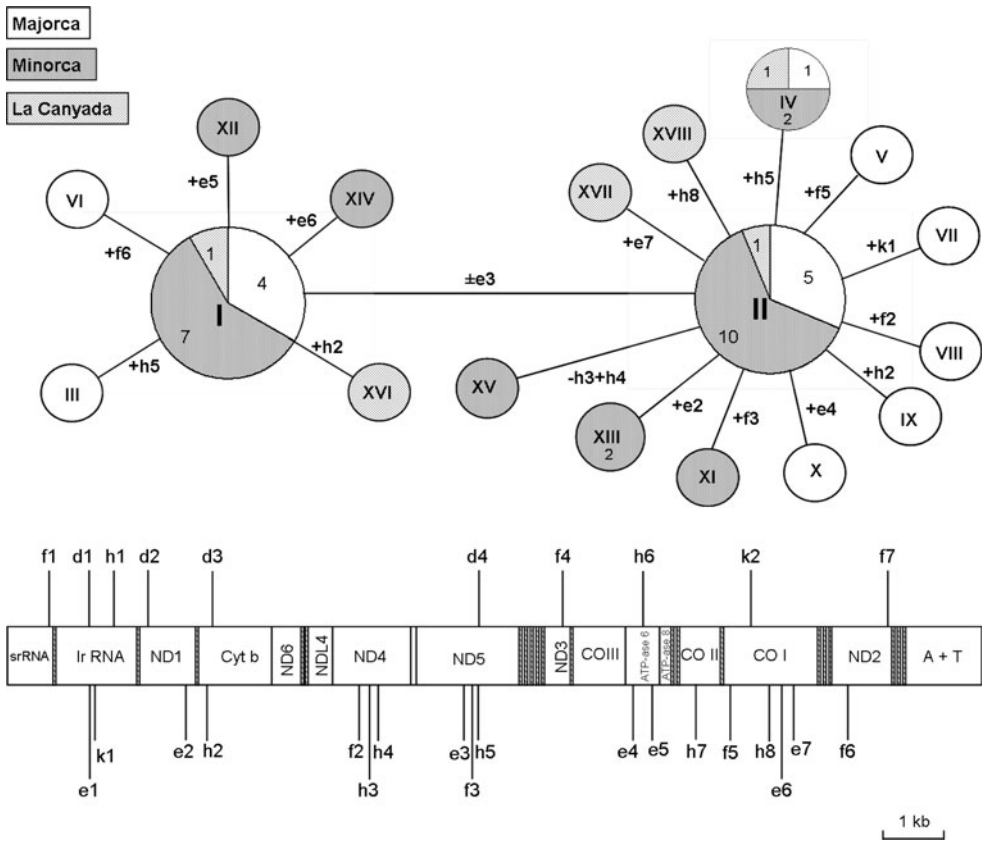


Fig. 1 *Top* Network of the 18 restriction site haplotypes of the Majorca, Minorca and La Canyada *Drosophila subobscura* samples. The haplotypes are connected in a way that minimizes the total number of site changes. *Numbers in large circles marked haplotype I and II* Number of individuals for each population, *small circles* less common haplotypes [only one individual per circle, with the exception of XIII (2) and IV (4)]. *Bottom D. subobscura* mtDNA organization based on the genetic map of *D. yakuba* given by Clary and Wolstenholme (1985). Conserved sites are shown *above* and polymorphic sites are shown *below* the map. *srRNA*, *lrRNA* small and large subunits of ribosomal RNA, respectively, *ND1-6* subunits of the NADH dehydrogenase complex, *Cytb* cytochrome b, *COI-III* subunits of cytochrome oxidase, *A + T* regulatory non-coding region. *d Eco*RI, *k Eco*RV, *e Hae*III, *h Hpa*II, *f Hin*dIII

Table 2 shows the 26 polymorphic sites found in the 942-bp mtDNA fragment of the ND5 gene in the 48 individuals sequenced; 24 of these are synonymous substitutions. With respect to the polymorphic restriction sites, site 96 corresponds to the *Hae*III target that distinguishes haplotype I from haplotype II by a C–T change; site 282, with a G–A change, corresponds to the *Hin*dIII target that separates haplotype XI from II; site 303, with a T–C change that corresponds to the *Hpa*II site, separates haplotypes III from I and IV from II. These three sites correspond to sites e3, f3 and h5 in the tree shown in Fig. 1. The sequence network shown in Fig. 2 indicates that only three sequences were derived from sequence 1 (S1), with all three belonging to the Majorca populations. The remaining sequences were derived from S2, most of them connected by one or two mutational steps, although in some cases haplotypes are missing between them, namely, S17, S18 and S19. As expected, *D. guanche* is well separated from the *D. subobscura* cluster; in this case by 58 mutational steps. The fragment of the ND5 gene sequenced includes the *Hae*III site that differentiates the two major haplotypes I and II. As can be observed, the haplotype and sequence networks are topologically quite similar.

Note that the restriction site analysis (RSA) and the nucleotide sequence analysis are at different scales. The RSA was used for examining entire mtDNA haplotypes, while the nucleotide sequence analysis was used to examine all polymorphic sites within the ND5 locus.

We also created a maximum parsimony tree with ND5 sequences, including *D. guanche* (results not shown), using MEGA v.4 software (Tamura et al. 2007) and bootstrap values based on 1,000 random samples. The tree obtained was found to reproduce the network. As expected, S1 and its derivatives are separated from S2 and its derivatives (bootstrap value of 50%), with the cluster of S17, S18 and S19 being strongly separated from the rest (bootstrap value of 93%).

Table 3 shows a matrix with the ND5 distinct sequences and restriction site haplotypes. Despite the lack of a perfect correspondence, all of the individuals with S1 sequences

Table 2 Sequences of the Majorca, Minorca and La Canyada populations

Sequences	Polymorphic sites	Majorca	Minorca	La Canyada	Total
	111122223333567777888899				
	902780278057976068812682 4				
	962144783238266622076575 42				
	SSSSSSSSSSSSSSNSSSSNSSSSSS				
S1	TCGTATGCTGTTGTTGAGGGCTTGAA	3	9	2	14
S2	.T..........................	5	6	3	14
S3	........................C...	1			1
S4	................A..........	1			1
S5	......A................G...	1			1
S6	.T....A.....................	1			1
S7	CT..........................	1			1
S8	.T...C......................	1			1
S9	.T.C........................	1			1
S10	.T............A..........G	1			1
S11	.T..G.....C.................	1			1
S12	.T.........C................		1		1
S13	.T..................T....T.		1		1
S14	.T..................T.....		1		1
S15	.T.....T....................		1		1
S16	.T.....T..C.................		1		1
S17	.TA....T..........A..C.A..		1		1
S18	.T.....T.A......GT...C.A..		1		1
S19	.T.....TC...AC..G....C.A..		1		1
S20	.T.........C................		2	1	3
Total		17	25	6	48

The 96 site corresponds to the *Hae*III target that distinguishes haplotypes I and II with RSAs

S synonymous substitution, *N* non-synonymous

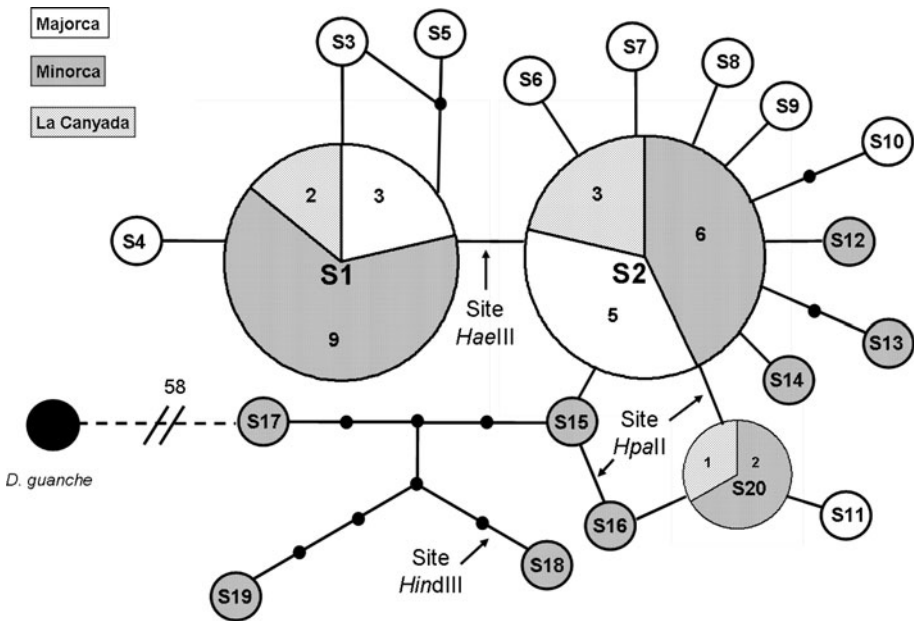


Fig. 2 The different molecular haplotypes connected by means of the sequence network. The sites that coincide with the polymorphic restriction sites are also indicated. *Numbers in large circles marked S1 and S2* Number of individuals of each population, *small circles* less common sequences [only one individual per circle, with the exception of S20 (3)]

and their derivatives (S3, S4, S5) can be seen to correspond to individuals with haplotype I and its derivatives (III, VI, XII, XIV, XVI). In the same way, flies with S2 sequences and their derivatives (S6–S20) correspond to flies with haplotype II and its derivatives. For example, the nine S1 flies from Minorca correspond to haplotypes I (seven individuals), XII (1) and XIV (1), with the latter two haplotypes derived from I. S20 and its derivatives (S11, S16) are related to haplotype IV (four individuals) and XIII (1), with the latter two haplotypes derived from II. S18 was

Table 3 Correspondence between *D. subobscura* mtDNA ND5 sequences and restriction site haplotypes

	Sequences																			
Haplotypes	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20
I	MA		MA	MA	MA															
	MI (7)																			
	LC																			
II		MA (2)					MA	MA		MA			MI	MI	MI				MI	
		LC																		
		MI (6)																		
III	MA																			
IV											MA					MI				LC/MI
V		MA																		
VI	MA																			
VII		MA																		
VIII		MA																		
IX						MA														
X									MA											
XI																		MI		
XII	MI																			
XIII												MI								MI
XIV	MI																			
XV																	MI			
XVI	LC																			
XVII		LC																		
XVIII		LC																		

MA, Majorca; MI, Minorca; LC, la Canyada

In general, S1 sequences and their derivatives (S3, S4 and S5) correspond to haplotype I and its derivatives (III, VI, XII, XIV and XVI). In the same way, S2 sequences and their derivatives (S6–S20) correspond to haplotype II and its derivatives (see text for more details). When there is more than one individual, the number of individuals is indicated in parenthesis

found to have a very different sequence from the rest (i.e., six changes with respect to S2), but it is related to XI, which is derived from haplotype II. The S17 sequence is also very different from the rest (i.e., five steps from S2), but it is also related to haplotype XV, which is derived from haplotype II with two differences. S19, with seven steps from S2, corresponds to haplotype II.

We also carried out genetic differentiation analyses among populations and tests of genetic neutrality for the two different datasets. Table 4 shows the differentiation within and between populations for the RSA datasets (Lynch and Crease 1990). Most of the detected variation was found to be present within populations. Majorca and Minorca showed a similar intra-population variability that is lower than that observed in La Canyada where each individual had a different haplotype. The *average* number of substitutions per nucleotide site for random pairs of haplotypes from the same population was 0.01271 ± 0.00633 and that between populations was -0.00061 ± 0.00355. The fraction of the nucleotide variation between populations, N_{ST}, was -0.051 ± 0.334 (not significant). Consequently, no differentiation was found between the

Table 4 Differentiation of restriction site analysis within and between populations of *D. subobscura*

Population	Majorca	Minorca	La Canyada
Majorca	0.01014 (0.00587)	−0.00021 (0.00100)	−0.00104 (0.00165)
Minorca		0.00871 (0.00532)	−0.00059 (0.00203)
La Canyada			0.01930 (0.00852)

Standard error (SE) is given in parenthesis

Values on the main diagonal are the within-population differentiation (Vw); values above the diagonal represent the between-population differentiation (Vb)

three populations. We also calculated Tajima's D for the three populations and found these to be negative and not significant: Majorca: −1.43, Minorca: −1.40 and La Canyada: −0.93. The total was also negative, but significant: −2.07 ($P < 0.05$).

Tables 5 and 6 show the nucleotide variation analyses of the ND5 sequence dataset. As in the case of RSAs for the entire mtDNA, most of the nucleotide variation was observed within populations. Majorca and Minorca showed similar levels of nucleotide diversity but, in contrast to the RSA dataset, La Canyada showed less variation because the six individuals only belong to three sequence types (S1, S2 and S20). Due to the major divergence of sequences S17, S18 and S19 from Minorca, we formed one subgroup (M2) with these sequences and the rest of that population (M1). In the Majorca populations, the Tajima's D was almost significantly negative ($P < 0.1$). The Fu and Li's D–F are significant, but not the Fay and Wu's H. Negative values of these tests could indicate directional selection and/or population expansion. At the same time, H is not so sensitive to population expansion as the other tests. It therefore appears that in Majorca, the haplotypes could follow a dynamics mainly based on population expansion. In La Canyada, the parameters were not statistically significant, likely due to the low number of sequences. In Minorca, H was highly significant, probably influenced by the M2 samples. When the analyses were performed without these sequences, H was lower but remained

Table 5 Nucleotide diversity analysis in *D. subobscura* populations

Population	Number of individuals (*N*)	Segregating sites[a]	π (standard deviation)[b]	D[c]	D, F[d]	H[e]
Majorca	17	11	0.00199 (0.00034)	−1.572	−2.641**, −2.972**	−0.691
La Canyada	6	2	0.00092 (0.00026)	−0.050	−1.133, −1.257	0.267
Minorca	25	17	0.00266 (0.00065)	−1.575	−0.736, −1.322	−11.020**
Minorca without S17-S18-S19	22	7	0.00137 (0.00023)	−1.050	−1.067, −1.460	−2.909*
Only S17–S18–S19[f]	3	8	0.00566 (0.00156)	–	–	–
Total without S17–S18–S19	45	16	0.00152 (0.00019)	−1.917*	−3.202**, −3.464**	−4.909*
Total	48	26	0.00221 (0.00041)	−2.128*	−2.645**, −3.071**	−13.888**

* $P < 0.05$; ** $P < 0.02$; Tajima's D in Majorca and Minorca were almost significant ($P < 0.1$)

[a] Number of sites are 942

[b] Nei (1987)

[c] Tajima (1989)

[d] Fu and Li (1993)

[e] Fay and Wu (2000)

[d] and [e] calculated with *D. guanche* as outgroup

[f] The neutrality tests were not possible because the number of sequences do not reach 4

Table 6 Nucleotide divergences among *D. subobscura* samples

Population	Number of individuals (*N*)	Nucleotide divergence[a, b]					N_{ST}[b]			
		MA	LC	MI	M1	M2	MA	LC	MI	M1
Majorca (MA)	17	0.00199	−0.00007	0.00004	−0.00001	0.00367	–			
La Canyada (LC)	6	0.00138	0.00092	−0.00003	−0.00008	0.00361	−0.0514	–		
Minorca (MI)	25	0.00237	0.00176	0.00266	–0.00002	0.00249	0.0180	–0.0160	–	
Minorca without S17–S18–S19 (M1)	22	0.00167	0.00106	0.00200	0.00137	0.00353	−0.0066	−0.0784	−0.0097	–
Only S17-S18-S19 (M2)	3	0.00749	0.00690	0.00665	0.00704	0.00566	0.4904*	0.5237	0.3751	0.5017*

* $P < 0.05$

Diagonal, Intra-population divergence (π), above diagonal (Da), net inter-population divergence; below diagonal (Dxy), average interpopulation divergence

[a] Jukes and Cantor (1969)

[b] Lynch and Crease (1990)

significant. In this case, the action of a directional selection is more plausible than a population expansion. Finally, the analyses with the total population gave high significant and negative parameters in all tests (with and without the M2 subgroup). In the inter-population comparisons, all of the samples showed similar and low divergences, with the exception of the M2 subgroup, which contains the most divergent ND5 haplotypes. These results were also confirmed with the N_{ST} values, in which the M2 subgroup was the only subgroup that showed the highest and, in some cases, significant divergence from the other populations.

Figure 3 shows the representation of the distribution of the observed and expected pairwise nucleotide site differences in the three populations, the total and the two subgroups from Minorca. Under the infinite-sites model, mismatch distributions are relatively smooth and unimodal under population expansion but ragged and generally multimodal for stationary populations. As can be seen in Fig. 3, Majorcan populations show an accurate fit to the model that suggests a recent expansion; this is also the case for the La Canyada population, although the fit is not so accurate. The behaviour of the Minorca population is worth noting, since it takes the form of a bi-modal curve; however, the bi-modality disappeared when the M2 sequences were excluded, indicating that these latter sequences belong to a different population. The large number of pairwise differences between these three haplotypes (M2; $\tau = 5.333$) indicates that they are more ancient than the

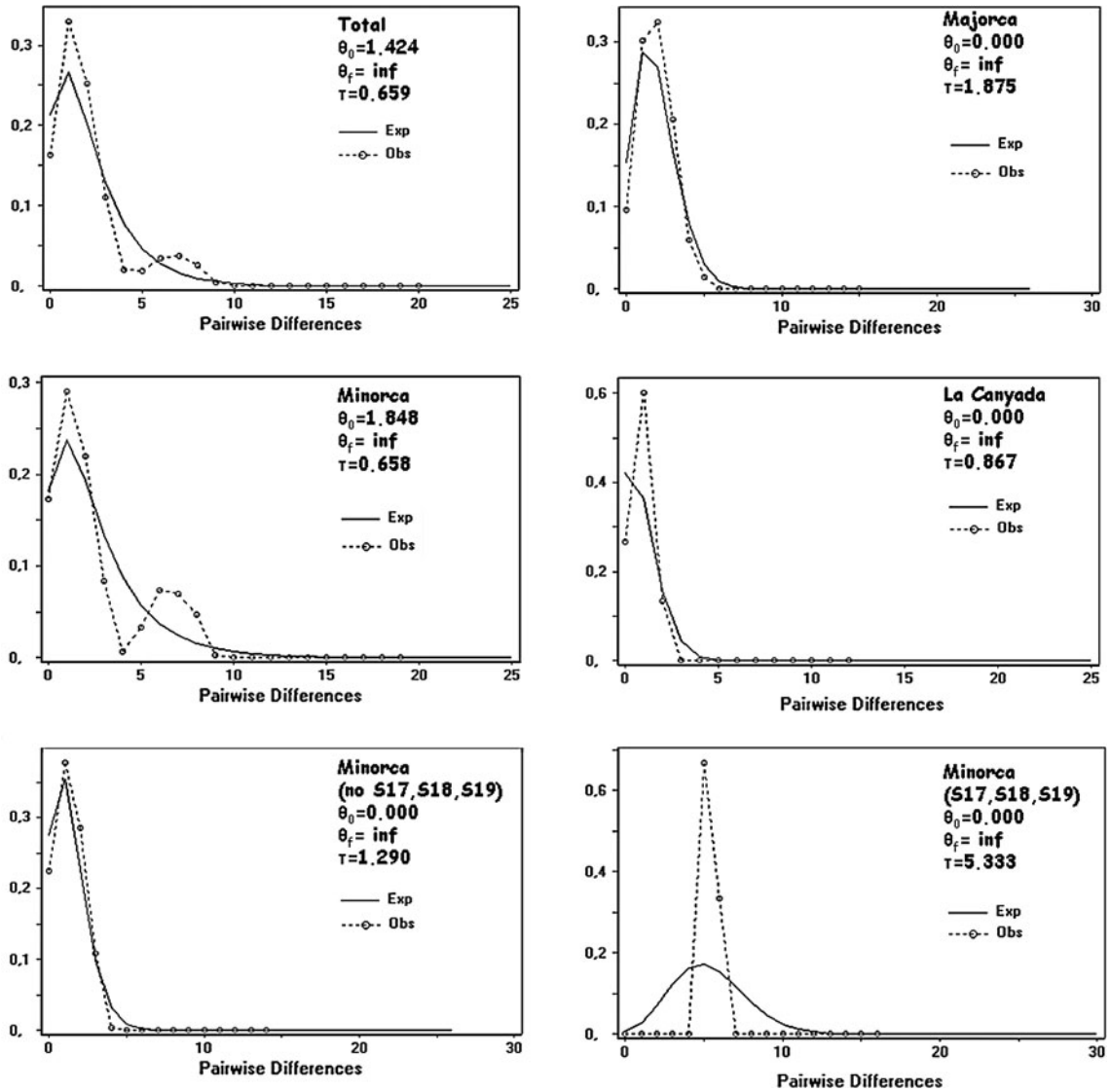


Fig. 3 Representation of frequencies for the observed and expected pairwise differences in the samples. All the samples fitted to a population expansion. $\theta = 2Nu$ is the expected pairwise differences where N is the effective population size of females, and u is the mutation rate per site and generation. θ_0, θ_f are the θ initial and final before and after the population growth, respectively, $\tau = 2ut$ (where t measures time in generations) is the date of the growth measured in units of mutational time. The estimation of the θ_0 and τ was made by allowing θ_f to be infinite. Exp Expected values, Obs Observed values

Table 7 McDonald–Kreitman test between all *D. subobscura* sequences and *D. guanche*

	Polymorphic sites in *D. subobscura*	Changes fixed between species
Non-synonymous	2	10
Synonymous	25	48

Fisher exact test: $P = 0.323$ (non-significant)

G (Williams correction): 1.543, $P = 0.214$ (non-significant)

Neutrality index (NI): 0.38

others. The total *D. subobscura* sample shows an accurate fit, but the double modality is the result of the influence of the Minorcan population.

Table 7 shows the interspecific comparison between the sequences of *D. subobscura* and *D. guanche* (Appendix I) using the McDonald–Kreitman test. All of the *D. subobscura* combinations were tested with *D. guanche*: Majorca, Minorca, La Canyada, Minorca without M2 and only M2, total and the total without M2. As all the tests were not significant, Table 7 only shows the analysis with all of the *D. subobscura* sequences. Despite the lack of significance of the statistical tests, the neutrality index (NI) was only 0.38, indicating that the ratio of replacement to silent variation is greater between species than within species (Rand and Kann 1996; Nachman 1998). However, when the analysis was calculated with the sequences of *D. subobscura* without M2, NI increased to 0.71, which is closer to the complete neutrality (NI = 1).

Discussion

The work reported here consisted of a study of the mtDNA molecular population genetics of *D. subobscura*; the aim was to reinforce or to refute previous results on the neutrality of haplotypes I and II of this species (Moya et al. 1993). To this end, we tested the results of our previous studies by RSA of both the full genome and the ND5 gene. A secondary aim was to provide some explanation for the high and quasi equi-frequency of the two major haplotypes (and the less common ones derived from them) in wild populations of *D. subobscura*.

It should be noted that the two network topologies, based on restriction sites and sequencing, respectively, were similar. We found a correspondence between haplotype I and its derivatives and sequence S1, as well as with the sequences derived from it. A similar situation was observed for haplotype II with respect to S2. Moreover, genetic variability among populations and N_{ST} yielded similar results for both the RSAs and the ND5 datasets. For comparison purposes, with the restriction sites we obtained an overview of the variability of the mtDNA, while with sequences, we only obtained an overview of the gene fragment.

In terms of the restriction sites, the neutrality tests based on the Tajima's D showed negative values that were not statistically significant in the three populations, although it was in the total analysis. These results are quite similar to previous ones obtained in studies on Majorca and Minorca populations (Castro et al. 1999). The conclusion drawn in the previous study (Castro et al. 1999) was that the Minorcan population (with significant negative D) had not reached equilibrium, likely as a consequence of population expansion after a bottleneck, while the Majorcan population (non-significant D) seemed to be more stable. The significance of the total population would be the result of the influence of the Minorcan population. In the present study, we extended our knowledge by sequencing the ND5 gene.

The tests of the frequency spectrum carried out to reject the hypothesis of neutrality within individual populations were, in general, negative and in some cases significant. In the total population, all test results were significant. Given the lack of population differentiation, it would appear that these combined negative tests do support an excess of singletons in *D. subobscura* populations. An excess of singletons could arise after recovery from a selective sweep, whereby a single allele replaces all other alleles in the population. However, this seems to be an unlikely scenario given the presence of two equally frequent haplotypes in *D. subobscura*. Population expansion after a bottleneck may also generate an excess of singletons relative to the neutral expectation. This scenario is consistent with our analyses of the mismatch distributions of pairwise differences between ND5 alleles. Based on our findings, we suggest that the population as a whole has not completely reached equilibrium as a consequence of the periodic seasonal bottlenecks (in summer and winter) that reduce the effective population size, followed by expansions, resulting in the existence of an excess of the less common haplotype polymorphisms. This scenario would be particularly prevalent in the case of mtDNA (small genome, with low or no recombinant and maternal inheritance) as it is more affected by the effective population size than the nuclear genome; in addition, random drift of the stationary bottlenecks is stronger in mtDNA than in nuclear genes (Lynch 2006).

The results of the McDonald–Kreitman test support the neutral evolution of ND5 alleles. A significant difference between the two ratios could be used to reject the neutrality, but we did not detect any such difference, although the neutrality index (NI = 0.38) was rather low (neutrality is total when NI = 1). Nonetheless, this index increased up to 0.71 when the M2 sequences (ancient alleles) were excluded. This value is similar to those reported in other *Drosophila* species, although it is very different from those

reported for mammals such as primates (*Homo versus Pan*) or between *Mus* species (Nachman 1998).

The results from earlier published studies on the genetic population dynamics of mtDNA have often, although not always, shown the rejection of neutral patterns, mainly in humans and their associated commensal taxa (flies and mice; Gerber et al. 2001). More recently, Ballard et al. (2007) provided evidence strongly supporting the hypothesis that selection is operating on *Drosophila simulans* mtDNA, although, due to the small sample size, the possibility that nuclear genes were responsible for the observed effects (at least in part) cannot be ruled out. In the same way, Bazin et al. (2006) and Meiklejohn et al. (2007) have shown positive and negative selection on mtDNA. For these reasons, we cannot reject the possibility that in our study selective forces that could have been acting directly or by cytonuclear disequilibrium with chromosomal rearrangements to maintain these two haplotypes in nature, as has been shown in other studies of our group (Oliver et al. 2002, 2005). As such, the action of some kind of epistatic selection could be invoked, although it could disappear and reappear because cytonuclear disequilibrium could be transient (García-Martínez et al. 1998; Oliver et al. 2005). Within the framework of our ongoing research project, our aim is to study the role of selection on the dynamics of the less common haplotypes, in addition to population expansion.

It is worth noting that three rare haplotypes in the Minorcan population (M2) appeared to be older that the rest. We cannot discard the effect of glaciations in Europe as a possible explanation for the distribution of haplotypes, both the main (that we can consider to be relatively young because they differ at only one site) and the less common ones, in the same way as Menozzi and Krimbas (1992) explained the inversion polymorphisms of *D. subobscura* in Europe. The last major ice ages, namely the Riss (200,000 years ago) and Würm (25,000 yeas ago), could have forced *Drosophila* (like most living species) to the south of Europe and to the north of Africa. Thereafter, the flies re-colonized. Some ancient relicts, such as some rare haplotypes in Minorcan populations (for instance, M2) and the endemic haplotype in the Canary Islands, may have survived to date. This explanation has also been used to explain the inter-specific diversification of the *D. subobscura* group (Latorre et al. 1988; Barrio et al. 1992).

Acknowledgments This work was supported by grants PB96-0793 and BOS2000-1000 from the Dirección General de Enseñanza Superior (Ministerio de Educación y Cultura, Spain).

Appendix I

See Table 8.

Table 8 Polymorphic nucleotide sites in *Drosophila subobscura* and *D. guanche* in the analysed region of the ND5 gene

	1111111111222222223333333334444445555556666667777777788888888888999999999
	1249900224778990122477800355778945679902247702466702678888012222366880011123 34
Site:	9872692613714425765830323661828467051281289361456932324037936014594725052564692
Change:	SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSNSSSSSSNSSSSSNSSSSSSNSNSNNNNNSSSSSNSSNSSSS
Difference:	PFFFPFPFPFFFPPPFPFFPFFPFPPPFFFPPFFPFFFFFFFFFFFPFFPFFPFPFPFPFPFFPFFPFPFFFFPFFP
Sequences	
S1	TATTCGGCTAGGATTGGAACGTTGTAAGTGGTTAAATTTCTGAATCAAGTAAGGTGTGAACTTATCGTTGTGAAGACAA
S2	....T.....
S3	.....C.....
S4	.....A.....
S5	.....A.....G.....
S6	....T.....A.....
S7	C...T.....
S8	....T.....C.....
S9	....T...C.....
S10	....T.....A.....G
S11	....T.....G.....C.....
S12	....T.....C.....
S13	....T.....T.....T...
S14	....T.....T.....
S15	....T.....T.....
S16	....T.....T....C.....
S17	....T.A.....T.....A.....C....A.....
S18	....T.....T...A.....G.T.....C....A.....
S19	....T.....T..C.....A..C.....G.....C....A.....
S20	....T.....C.....
D. guanche	.TACAAAT.GAA..AA.GGTAC.A.TGA.AAG.TGGAAATCAGT.TGGTATCAAGAC.GG.AATATA.CAAAGGATTGG

S, synonymous substitution; N, non-synonymous; F, substitution fixed between species; P, polymorphic substitution in *D. subobscura*
D. guanche was captured in Anaga, Tenerife (Canary Islands)

References

Afonso JM, Volz A, Hernández M, Ruttkay H, González AM, Larruga JM, Cabrera VM (1990) Mitochondrial DNA variation and genetic structure in Old-World populations of *Drosophila subobscura*. Mol Biol Evol 7:123–142

Ayala FJ, Serra LL, Prevosti A (1989) A grand experiment in evolution: the *Drosophila subobscura* colonization of the Americas. Genome 31:246–255

Ballard JWO, Melvin RG, Katewa SD, Maas K (2007) Mitochondrial DNA variation is associated with measurable differences in life-history traits and mitochondrial metabolism in *Drosophila simulans*. Evolution 61:1735–1747

Barrio E, Latorre A, Moya A, Ayala FJ (1992) Phylogenetic reconstruction of the *Drosophila obscura* group, on the basis of mitochondrial DNA. Mol Biol Evol 9:621–635

Bazin E, Glémin S, Galtier N (2006) Population size does not influence mitochondrial genetic diversity in animals. Science 312:570–572

Castro JA, Ramon M, Picornell A, Moya A (1999) The genetic structure of *Drosophila suboscura* populations from the islands of Majorca and Minorca (Balearic Islands, Spain) based on allozymes and mitochondrial DNA. Heredity 83:271–279

Castro JA, Oliver P, Christie JS, Picornell A, Ramon M, Moya A (2003) Assortative mating and fertility in two *Drosophila subobscura* strains with different mitochondrial DNA haplotypes. Genetica 119:295–301

Christie JS, Castro JA, Oliver P, Picornell A, Ramon MM, Moya A (2004) Fitness and life-history traits of the two major mitochondrial DNA haplotypes of *Drosophila subobscura*. Heredity 93:371–378

Clary DO, Wolstenholme DR (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 KA (2000) TCS: a computer program to estimate gene genealogies. Mol Ecol 9:1657–1660

Fay JC, Wu CI (2000) Hitchhiking under positive darwinian selection. Genetics 155:1405–1413

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709

García-Martínez J, Castro JA, Ramón M, Latorre A, Moya A (1998) Mitochondrial DNA haplotype frequencies in natural and experimental populations of *Drosophila subobscura*. Genetics 149:1377–1382

Gerber AS, Loggins R, Kumar S, Dowling TE (2001) Does non-neutral evolution shape observed patterns of DNA variation in animal mitochondrial genomes? Annu Rev Genet 35:539–566

Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Sym Ser 41:95–98

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York, pp 21–132

Latorre A, Moya A, Ayala FJ (1986) Evolution of mitochondrial DNA in *Drosophila subobscura*. Proc Natl Acad Sci USA 83:8649–8653

Latorre A, Barrio E, Moya A, Ayala FJ (1988) Mitochondrial DNA evolution in the *Drosophila obscura* group. Mol Biol Evol 5:717–728

Latorre A, Hernández C, Martínez D, Castro JA, Ramón M, Moya A (1992) Population structure and mitochondrial DNA gene flow in Old World populations of *Drosophila suboscura*. Heredity 68:15–24

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

Lynch M (2006) Streamlining and simplification of microbial genome architecture. Annu Rev Microbiol 60:327–349

Lynch M, Crease TJ (1990) The analysis of population survey data on DNA sequence variation. Mol Biol Evol 7:377–394

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature 351:652–654

Meiklejohn C, Montooth KKL, Rand DM (2007) Positive and negative selection on the mitochondrial genome. Trends Genet 23:259–263

Menozzi P, Krimbas CB (1992) The inversion polymorphism of *D. subobscura* revisited. Synthetic maps of gene arrangement frequencies and their interpretation. J Evol Biol 5:625–641

Moya A, Barrio E, Martínez D, Latorre A, González-Candelas F, Ramón M, Castro JA (1993) Molecular characterization and cytonuclear disequilibria of two *Drosophila subobscura* mitochondrial haplotypes. Genome 36:890–898

Nachman MW (1998) Deleterious mutations in animal mitochondrial DNA. Genetica 102(103):61–69

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Oliver P, Castro JA, Picornell A, Ramon MM, Solé E, Balanyà J, Serra L, Latorre A, Moya A (2002) Linkage disequilibria between mtDNA haplotypes and chromosomal arrangements in a natural population of *Drosophila subobscura*. Heredity 89:133–138

Oliver P, Balanyà J, Ramon MM, Picornell A, Ll Serra, Moya A, Castro JA (2005) Population dynamics of the two majority mitochondrial DNA haplotypes in experimental populations of *Drosophila subobscura*. Genome 48:1010–1018

Pinto FM, Brehm A, Hernandez M, Larruga JM, González AM, Cabrera VM (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

Rand DM, Kann LM (1996) Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. Mol Biol Evol 13:735–748

Rogers A, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol 9:552–569

Rozas JM, Hernandez M, Cabrera VM, Prevosti A (1990) Colonization of America by *Drosophila subobscura*: effect of the founder event on the mitochondrial DNA polymorphism. Mol Biol Evol 7:103–109

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