

1 **Gene flow and gene flux shape evolutionary**
2 **patterns of variation in *Drosophila***
3 ***subobscura***

4
5 **Cinta Pegueroles¹, Charles F. Aquadro², Francesc Mestres¹ and Marta**
6 **Pascual¹**

7
8 ¹Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Barcelona
9 08028, Spain.

10 ²Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York
11 14853.

12
13 Corresponding author: Cinta Pegueroles, Departament de Genètica, Facultat de
14 Biologia, Universitat de Barcelona, Avd. Diagonal 645, 08028-Barcelona, Spain.

15 Phone: +34 93 4035801, fax: +34 93 4034420, e-mail: cintapq@gmail.com

16

17 Running title: Gene exchange shaping variation

18 Word count (main text): 5709

19

20 Gene flow (defined as allele exchange between populations) and gene flux (defined as
21 allele exchange during meiosis in heterokaryotypic females) are important factors
22 decreasing genetic differentiation between populations and inversions. Many
23 chromosomal inversions are under strong selection and their role in recombination
24 reduction enhances the maintenance of their genetic distinctness. Here we analyze levels
25 and patterns of nucleotide diversity, selection and demographic history using 37
26 individuals of *Drosophila subobscura* from Mount Parnes (Greece) and Barcelona
27 (Spain). Our sampling focused on two frequent O-chromosome arrangements that differ
28 by two overlapping inversions (O_{ST} and O_{3+4}), which are differentially adapted to the
29 environment as observed by their opposing latitudinal clines in inversion frequencies.
30 The six analyzed genes (*Pif1A*, *Abi*, *Sqd*, *Yrt*, *Atpa* and *Fmr1*) were selected for their
31 location across the O-chromosome and their implication in thermal adaptation. Despite
32 the extensive gene flux detected outside the inverted region, significant genetic
33 differentiation between both arrangements was found inside it. However, high levels of
34 gene flow were detected for all six genes when comparing the same arrangement among
35 populations. These results suggest that the adaptive value of inversions is maintained
36 regardless the lack of genetic differentiation within arrangements from different
37 populations, and thus favors the Local Adaptation hypothesis over the Coadapted
38 Genome hypothesis as the basis of the selection acting on inversions in these
39 populations.

40

41 **Keywords:** *Drosophila subobscura*; nuclear genes; gene flow; gene flux; inversion
42 dating

43 **INTRODUCTION**

44 Chromosomal inversions are a major mechanism shaping the level and distribution of
45 genomic diversity within and between species. Inversion polymorphism has been
46 described in several *Drosophila* species (Krimbas and Powell, 1992) and has been
47 observed to vary seasonally, and with altitude and latitude (Dobzhansky, 1943;
48 Krimbas, 1993; Kennington *et al.*, 2006). The rich polymorphism in paracentric
49 inversions found in *Drosophila subobscura* (Krimbas and Powell, 1992; Krimbas,
50 1993) has played an important role in demonstrating the adaptive role of inversions, as
51 shown by the coinciding latitudinal clines on inversion frequencies found in Palearctic
52 and also in colonizing populations (Krimbas and Loukas, 1980; Prevosti *et al.*, 1988;
53 Balanyà *et al.*, 2003). Moreover, changes in chromosomal polymorphism matching
54 global warming (Solé *et al.*, 2002; Balanyà *et al.*, 2004, 2006) also suggest the adaptive
55 value of inversions in *D. subobscura*. Despite the evidence for an adaptive role of
56 inversions, the genetic and mechanistic basis underlying their role in adaptation remains
57 unknown.

58 Several hypotheses have been suggested to explain the maintenance of inversion
59 polymorphism in populations (reviewed in Hoffmann and Rieseberg, 2008). The
60 coadaptation hypothesis (Dobzhansky, 1950) suggests that natural selection maintains
61 favourable combinations of alleles that interact epistatically within and between
62 arrangements in a certain population. The local adaptation hypothesis predicts that
63 inversions will spread in a population at migration-selection balance when they capture
64 at least two advantageous alleles that are individually adapted to local conditions
65 (Kirkpatrick and Barton, 2006). Particularly, it explains that the inversion will establish
66 a cline when interacting populations are distributed along an environmental gradient. In
67 both hypotheses the reduced recombination between inversion heterokaryotypes avoids

68 gene exchange with other genetic backgrounds, but epistasis is only a requirement in the
69 former one.

70 The low gene transfer between inversions (gene flux) for genes located inside
71 the inverted region observed in some *Drosophila* species is in agreement with both
72 hypotheses (Laayouni *et al.*, 2003; Schaeffer *et al.*, 2003; Munté *et al.*, 2005; Hoffmann
73 and Rieseberg, 2008). However, despite the fact that Dobzhansky detected a lower
74 fitness of heterozygous individuals from different populations of *D. pseudoobscura* in
75 laboratory experiments (Dobzhansky, 1950), molecular studies failed to detect genetic
76 differentiation within inversions sampled from different populations (Schaeffer *et al.*,
77 2003). In *D. subobscura*, high genetic differentiation between European populations
78 was detected when chromosomal arrangements were used as markers, since their
79 frequency widely varies between populations (Krimbas, 1993). However, given that
80 these chromosomal arrangements are under strong selection (Prevosti *et al.*, 1988;
81 Balanyà *et al.*, 2006), gene flow between populations would likely be underestimated
82 using the inversions themselves as markers. Interestingly, low levels of genetic
83 differentiation between European populations of *D. subobscura* were observed using
84 molecular markers such as RFLPs (Rozas *et al.*, 1995) and microsatellite *loci* (Pascual
85 *et al.*, 2001). Consequently, gene flow and gene flux could be changing the genetic
86 content of inversions from widely separated populations. Thus the analysis of candidate
87 genes undergoing selection in *D. subobscura* could allow contrasting different
88 hypotheses explaining the maintenance of inversion polymorphism in populations.

89 In the present study we analyze DNA sequence variation and differentiation at
90 six genes across the O chromosome within and between two populations of *D.*
91 *subobscura* from the western and eastern Mediterranean: Barcelona (Spain) and Mt.
92 Parnes (Greece). Despite being located at approximately the same latitude, these

93 populations differ substantially in inversion frequencies (Krimbas and Loukas, 1980;
94 Krimbas, 1993; Araúz *et al.*, 2009). Our sampling focused on homokaryotypic
95 individuals carrying either of two arrangements, O_{ST} and O₃₊₄, selected because they are
96 the more common arrangements in both populations and yet show significant latitudinal
97 clines of opposite sign (Prevosti *et al.*, 1988). Furthermore, differential basal expression
98 of *Hsp70* gene, candidate for thermal adaptation, was detected between carriers of these
99 two arrangements from the same population (Calabria *et al.*, 2012). Thus, the six
100 nuclear genes studied in the present work (Table 1) were chosen because they are
101 candidates to thermal adaptation (Laayouni *et al.*, 2007) and because of their cytological
102 location across the O-chromosome (Figure 1).

103

104 **MATERIALS AND METHODS**

105 **Fly samples, DNA isolation and sequencing**

106 *Drosophila subobscura* isochromosomal lines were established from two distant
107 European populations (Barcelona, Spain and Mt. Parnes, Greece) following the pattern
108 of genetic crosses described in Mestres *et al.* (1998). Chromosomal arrangements for
109 each line were determined by analysis of polytene chromosomes of third instar larvae
110 (Figure 2). A total of 25 O₃₊₄ lines (13 from Barcelona and 12 from Mt. Parnes) and 12
111 O_{ST} lines (5 from Barcelona and 7 from Mt. Parnes) were non-lethal and sequenced in
112 the present study. The structure of O_{ST} and O₃₊₄ arrangements is represented in Figure 1.
113 It is worth considering that O₃ and O₄ inversions are never found alone in natural
114 populations. *D. madeirensis* and *D. pseudoobscura* were used for interspecific analyses.
115 *D. madeirensis* sequences were obtained from a laboratory strain (Supplementary Table
116 S1) and *D. pseudoobscura* sequences were downloaded from Flybase
117 (<http://flybase.org>).

118 Six genes distributed across the O chromosome were studied (Figure 1): *Pif1A*
119 (PFTAIRE-interacting factor 1A), *Abi* (Abelson interacting protein), *Sqd* (Squid), *Yrt*
120 (Yurt), *Atpa* (Na pump α subunit), *Fmr1* (Fragile X mental retardation). They were
121 selected because they showed differential expression between laboratory populations
122 maintained at different temperature regimes (Laayouni *et al.*, 2007). Although their
123 cytological location had been previously reported, new *in situ* hybridizations have been
124 carried out in the present study, following standard protocols (Laayouni *et al.*, 2000).
125 After thoroughly inspection, four out of six genes proved to be previously mislocalized
126 and the correct locations are given in Table 1. Primers used for amplification and
127 sequencing reaction are also listed in Table 1. Some primers were reported in Laayouni
128 *et al.* (2007) and others were designed in the present study using Primer Designer v1.01
129 (1990 Scientific and Educational Software). Single fly genomic DNA was extracted
130 using Genra Puregene Cell Kit (Qiagen).

131 PCR amplifications were carried out in a 25 μ l total volume with 3.5 μ l buffer
132 10X, 3 μ l dNTPs (1 mM), 0.2 μ l Taq DNA Polymerase (Qiagen), 0.5 μ l forward primer
133 (10 μ M), 0.5 μ l reverse primer (10 μ M), 16.3 μ l water and 1 μ l DNA. Amplification
134 conditions were 4 min at 94°C of initial denaturation, and 35 cycles with 30 sec at 94°C,
135 30 sec at the required annealing temperature (56°C for all genes with the exception of
136 *Abi* and *Sqd* at 50°C) and 3 min at 72°C, and 5 min at 72°C of final extension. PCR
137 amplification products were purified with ExoI-SAP (BioLabs-Promega). Sequencing
138 reactions were carried out using the ABI Prism BigDye Terminators 3.1 cycle
139 sequencing kit (Applied Biosystems), and sequences read on an ABI 3730 sequencer at
140 the Biotechnology Resource Center of Cornell University. Follow-up PCR and
141 sequencing of some genes was done at the University of Barcelona as follows. PCR
142 amplification reactions were carried out in a 15 μ l total volume with 3 μ l buffer 10X, 3

143 μl dNTPs (1 mM), 0.45 μl DMSO, 0.3 μl Phire Hot Start DNA Polymerase
144 (Finnzymes), 0.75 μl forward primer (10 μM), 0.75 μl reverse primer (10 μM), 5.75 μl
145 water and 1 μl DNA. Amplification conditions were 30 sec at 98°C of initial
146 denaturation, and 35 cycles with 5 sec at 98°C denaturation, 5 sec at the required
147 annealing temperature (56°C for all genes with the exception of *Abi* and *Sqd* at 50°C)
148 and 90 sec at 72°C, and 7 min at 72°C for final extension. PCR amplification products
149 were purified and sequenced as before, but at the Serveis Científico-Tècnics at
150 Universitat de Barcelona. Sequences were assembled with SeqMan II (DNASTAR) and
151 aligned with Clustal W (Thompson *et al.*, 1994) implemented in BioEdit v7 (Hall
152 1999).

153 **Data analysis**

154 The six analyzed genes can be divided into two groups according to their cytological
155 location (Figure 1). *Fmr1* gene is located within inversion O_4 in region SI, which
156 includes the chromosome segment covered by the O_{3+4} arrangement. The other genes
157 (*Pif1A*, *Abi*, *Sqd*, *Yrt* and *Atpa*) are located in region SII which includes the co-linear
158 chromosome segment that can probably freely recombine between arrangements O_{ST}
159 and O_{3+4} (Pegueroles *et al.*, 2010a). Distance in basepairs of each gene to the nearest
160 inversion breakpoints was estimated assuming that all cytological bands contain the
161 same amount of DNA and that the length of the O chromosome of *D. subobscura* is
162 equivalent to chromosome 2 of *D. pseudoobscura* (Pegueroles *et al.*, 2010b). The
163 exonic and intronic regions of each gene sequence were determined by blasting against
164 the genome of *D. pseudoobscura* and the resulting expected proteins were compared to
165 the corresponding *D. pseudoobscura* proteins to ensure the correct assignment of the
166 exons. Exonic and intronic composition of each gene is shown in Figure 1.

167 Analyses were carried out separately for each gene and also for the concatenated
168 data set. *Abi* gene was not included in the final concatenate data set due to the
169 insufficient number of individuals. For this gene, sequencing of some individuals failed
170 due to the presence of a gene duplication (data not shown). Only individuals with
171 sequences for the five remaining genes were used in the concatenation. Gene
172 concatenation was carried out using Concatenator v1 software (Pina-Martins and Paulo,
173 2008). DnaSP v5 (Librado and Rozas, 2009) was used to estimate the standard
174 parameters for each population and arrangement.

175 We estimated the number of haplotypes (h), number of polymorphic sites (S),
176 the number of singletons, nucleotide diversity (π) (Nei, 1987), nucleotide diversity in
177 synonymous sites and non-coding positions (π_{sil}) (Nei and Gojobori, 1986) and
178 heterozygosity in silent sites (θ_{sil}) (Watterson, 1975). Comparisons in diversity values
179 were assessed by Wilcoxon matched pairs tests using Statistica v6. Divergence per
180 silent site (K_{sil}) between *D. subobscura* and *D. madeirensis* or *D. pseudoobscura* (Nei
181 and Gojobori, 1986) was also obtained. Genetic differentiation was assessed using Dxy
182 (Nei, 1987), F_{ST} (Hudson *et al.*, 1992) and S_{nn} (Hudson, 2000), and its significance
183 estimated after 10000 replicates. These analyses were carried out excluding insertion-
184 deletion events (indels).

185 Tajima's D (Tajima, 1989) and Fu and Li's D (Fu and Li, 1993) were used to
186 test for neutrality. Fu and Li's D was computed using *D. pseudoobscura* as outgroup
187 given the similar θ_{sil} and K_{sil} values considering *D. madeirensis* as outgroup (see
188 results). Population size changes were determined by the statistic R_2 (Ramos-Onsins
189 and Rozas, 2002) and its significance was assessed by coalescent simulations using
190 1000 replicates considering either no recombination, free recombination, or level of
191 recombination (ρ) determined from the observed sequence diversity using the composite

192 likelihood method of Hudson (Hudson, 2001) implemented in LDhat v2.1
193 (<http://www.stats.ox.ac.uk/~mcvean/LDhat>). Gene genealogies were constructed using
194 the Neighbour-Joining method with the maximum composite likelihood model
195 implemented in the Mega v4 program (Tamura *et al.*, 2007). Bootstrap values were
196 obtained from 1000 replicates.

197 The age of inversions was estimated with the *Fmr1* gene, since it is located
198 inside the inversions, using two different approaches: using the silent nucleotide
199 diversity of the two most divergent individuals (Rozas and Aguadé, 1994), and using
200 mean silent nucleotide diversity of all individuals (Rozas *et al.*, 1999), except those
201 identified as recombinants. The number of substitutions per site and year was estimated
202 using the divergence per silent site between *D. subobscura* and *D. pseudoobscura*
203 assuming that the two species diverged 17.7 Myr ago, based on a large multilocus data
204 set (Tamura *et al.*, 2004) and 8 Myr ago, based on only one gene (Ramos-Onsins *et al.*,
205 1998) but used for comparison with previous studies.

206

207 **RESULTS**

208 **Nucleotide variation**

209 Genetic variability levels of the six genes within each arrangement (O_{3+4} and O_{ST}) and
210 for each population are presented in Table 2. In all cases, the number of haplotypes
211 detected approached the number of sequences analyzed. No significant differences
212 between populations were obtained neither in nucleotide diversity (π : $Z = 0.58$, $P =$
213 0.56) nor in the number of polymorphic sites (S : $Z = 0.04$, $P = 0.97$). When comparing
214 the two arrangements within populations, no significant differences were detected for π
215 ($Z = 0.31$, $P = 0.75$ for BC population, $Z = 1.36$, $P = 0.17$ for MP population). The
216 number of polymorphic sites within each arrangement was similar for all genes, except

217 for *Pif1A*. This gene presented the highest nucleotide diversity probably due to its larger
218 proportion of intronic regions (Figure 1). Thus, nucleotide diversity, heterozygosity per
219 site and divergence were subsequently analyzed using only synonymous sites and non-
220 coding positions. The mean highest π_{sil} and K_{sil} values found in the *Yrt* gene ($\pi_{\text{sil}} =$
221 0.036, $K_{\text{sil-mad}} = 0.078$, $K_{\text{sil-pseu}} = 0.325$) suggest that this gene might have a higher
222 substitution rate. Genetic divergence from *D. subobscura* was always higher for *D.*
223 *pseudoobscura* than for *D. madeirensis*, as expected from their phylogenetic
224 relationships (Table 2).

225

226 **Genetic differentiation between populations and arrangements**

227 When comparing populations for the same arrangement, S_{nn} values were always not
228 significantly different from zero (Table 3). Between arrangements, significant
229 differences and fixed mutations were only found for *Fmr1* (Supplementary Table S2).
230 This result is consistent with the cytological location of *Fmr1* within inversion O_4
231 (Figure 1). F_{ST} values were small for all the comparisons with the exception of *Fmr1*
232 when comparing the two arrangements. Moreover, there were significant differences
233 between arrangements for the concatenated data set considering all genes, but the
234 analysis of the concatenated data set excluding *Fmr1* showed small and non significant
235 values (Table 3). The highest D_{xy} values were detected for *Pif1A*, (Table 3) which
236 could be explained by its large portion of intronic content resulting in higher variability
237 levels.

238 Distance trees for the concatenated data set or only for *Fmr1* were carried out
239 using *D. madeirensis* and *D. pseudoobscura* as outgroups (Figure 3). In both
240 phylogenetic trees, sequences from different populations randomly clustered. In the tree
241 of the concatenated data set, both populations and arrangements were randomly

242 distributed (Figure 3A) and, although some sequences from the same arrangements
243 were grouped together their bootstrap values were below 30%. Nonetheless, it is worth
244 noting that, in the *Fmr1* genealogy (Figure 3B), the O_{3+4} sequences grouped with a high
245 bootstrap value (97%), though the clustering of the remaining sequences (all belonging
246 to the O_{ST} arrangement) was not statistically supported. By visual inspection of the
247 *Fmr1* sequences (Supplementary Table S1) we detected four recombinant sequences for
248 the O_{ST} arrangement (asterisks show the recombinant individuals in Figure 3B), but
249 none was detected for the O_{3+4} arrangement. The regions affected by recombination
250 (more likely gene conversion due to their short size) were approximately 161 bp for
251 MP19 and FBC58, and 10 bp for MP28 and BC55 individuals. When excluding these
252 recombinant individuals, nine fixed differences were observed between both
253 arrangements combining sequences from both populations (Supplementary Table S1).
254 Furthermore, the phylogenetic reconstruction after excluding these recombinants
255 showed monophyly for the two arrangements with more than 89% bootstrap support.

256

257 **Population demography and selection**

258 Tajima's D and Fu and Li's D test presented negative although non significant values
259 for the concatenated data set as well as for each gene separately, with the exception of
260 the *Atpa* gene in Mt. Parnes (Table 4). Negative D statistic values correspond to an
261 excess of low frequency polymorphism, which could be due to either demography or
262 selection. To further elucidate this issue, the R_2 statistic (Ramos-Onsins and Rozas,
263 2002), which is a test for recent population expansion, was estimated because it is more
264 suitable for small sample sizes presenting recombination. R_2 estimates ranged between
265 0.09 and 0.18 and significance was assayed considering three alternative recombination
266 scenarios, since rejection of the null hypothesis of constant population size depends on

267 the level of recombination implemented (Table 4). Most of the comparisons were non
268 significant assuming no recombination, although they were significant when
269 incorporating for each gene the rate of recombination (ρ) estimated using the composite
270 likelihood method of Hudson (Hudson 2001). Very similar results were obtained
271 assuming free recombination. Since the fraction of coding region varies among genes in
272 order to compare estimates, previously described tests were also calculated using only
273 silent sites (Supplementary Table S3). In general, the same patterns were detected when
274 comparing different recombination levels with the number of significant R_2 estimates
275 being always higher in Mt. Parnes than in Barcelona. Finally, when considering
276 population growth within arrangements for gene *Fmr1* gene, which is the only gene
277 showing signals of significant genetic differentiation, the null hypothesis of constant
278 size was only rejected for the O_{3+4} arrangement (Table 4).

279

280 **Age of inversions**

281 Sequences from *Fmr1* were used for estimating the age of the inversions, since it is the
282 only gene analyzed which is located in section SI, in which both arrangements differ
283 (Figure 1). Recombinant individuals were excluded because only variation originated by
284 mutation is useful for this analysis (Figure 3). The ages of inversions were estimated for
285 each population separately and combining them since their origin should be unique
286 (Table 5). Since older divergence times lead to estimate smaller mutation rates, the ages
287 of the inversions estimated using Tamura's divergence time are sensitively older than
288 using Ramos-Onsins' divergence time (Table 5). In addition, older estimates were
289 obtained using the information from the two more distant individuals, in relation to the
290 values obtained using the mean silent nucleotide diversity. Within each approach, the
291 age estimates for O_{3+4} and O_{ST} were quite similar. Nonetheless, older coalescent times

292 were detected for O_{3+4} when using the most divergent individuals (1.01 and 2.24 Myr
293 for both divergence times when mixing populations and, 0.87 and 1.93 for Barcelona
294 population; see Table 5), while older coalescent time were detected for O_{ST} with the
295 mean nucleotide diversity excluding recombinants (0.49 and 1.09 Myr for both
296 divergence times when mixing populations and, 0.48 and 1.06 for Barcelona population;
297 see Table 5).

298

299 **DISCUSSION**

300 **Patterns of nucleotide variation and demographic effects**

301 The present study focuses on the analysis of six genes in the context of two
302 different arrangements: the O_{ST} and the O_{3+4} . Five of these genes are located outside the
303 inverted region and only one (*Fmr1*) was located inside the O_4 inversion and
304 approximately 1.3 Mb from the nearest breakpoint. In terms of nucleotide variability,
305 lower nucleotide diversity values are expected in markers located close to the inversion
306 breakpoint (Andolfatto *et al.*, 2001, Stevison *et al.*, 2011). In *Drosophila buzzatii* a
307 correlation between nucleotide diversity and distance to the breakpoint was observed
308 (Laayouni *et al.*, 2003). However, *Fmr1* does not present the lowest levels of π_{sil} despite
309 being located inside and close to the inversion breakpoint. Similarly, no reduction of the
310 variation was observed in *D. subobscura* for genes located close to the inversion
311 breakpoints (Munté *et al.*, 2005). Mutation and recombination (including gene
312 conversion) could both have contributed to recover variability levels within the inverted
313 regions. In fact, the variability for *Fmr1* gene in O_{ST} was increased due to the presence
314 of four putative recombinant individuals, despite this gene is closely located to O_4
315 inversion breakpoint. The presence of recombinants only among O_{ST} lines could arise
316 from differences in the frequency of both arrangements in the two populations.

317 According to data from Araúz *et al.* (2009), in Barcelona and Mt. Parnes it is more
318 likely to find O_{3+4+X} homozygous (35% and 74% respectively) than O_{ST}/O_{3+4+X}
319 heterozygous (13% and 18% respectively) or O_{ST} homozygous (1.3% and 1.1%
320 respectively), where X includes non overlapped inversions that can freely recombine in
321 region SI with O_{3+4} (Figure 1). Therefore, the probability of sampling a recombinant
322 O_{ST} chromosome is higher than that for a recombinant O_{3+4+X} . Similarly, a previous
323 study concluded that the frequency of heterozygote hybrids between *D. pseudoobscura*
324 and *D. persimilis* determines the frequency of genetic exchange in the population
325 (Stevison *et al.*, 2011).

326 It has been previously suggested that recombination should be higher in central
327 parts of the inversion because genes located in this region could be influenced by
328 double crossovers and gene conversion, while genes close to inversions breakpoints
329 should be mainly influenced by gene conversion (Navarro *et al.*, 1997). Nevertheless
330 double recombinants have been observed close to the breakpoints when analyzing the
331 offspring of heterokaryotypic females indicating that recombination close to the
332 breakpoint can occur although in very low frequency (Pegueroles *et al.*, 2010a). Since
333 recombination between heterokaryotypes is reduced inside inverted regions (Navarro *et*
334 *al.*, 1997; Stump *et al.*, 2007, Pegueroles *et al.*, 2010a), higher diversity levels are
335 expected outside inverted regions than inside them. In the region where genes can freely
336 recombine between both arrangements (SII region, Figure 1), estimates of π_{sil} ranged
337 from 0.004 to 0.024 for the O_{ST} arrangement and from 0.006 to 0.023 for O_{3+4} , after
338 excluding *Yrt* since it could act as an outlier for the nucleotide variability (see Results
339 section). The π_{sil} values for nine genes located inside the inverted region between the
340 two assayed arrangements (SI region, Figure 1), combining results of a previous study
341 (Munté *et al.*, 2005) and the present work, ranged from 0.004 to 0.013 for the O_{ST}

342 arrangement, and from 0.008 to 0.013 for the O_{3+4} arrangement, after excluding the
343 *Acph1* gene which also acted as a mutation outlier. The ranges of nucleotide variability
344 mostly overlap for genes located inside and outside the inverted regions, and non
345 significant differences were detected when performing a Mann-Whitney U Test ($O_{ST} P$
346 = 0.142, $O_{3+4} P = 0.457$), although the highest values are found outside the inversions.
347 Reduced diversity within inversions has been found in the *D. persimilis* lineage,
348 indicative of a recent fixation process despite inversions may have arisen long ago
349 (McGaugh and Noor, 2012). Our results suggest that despite the studied inversions are
350 maintained by selection (Prevosti *et al.*, 1988, Balanyà *et al.*, 2006), recombination
351 could also contribute in some extent to recover variability inside reasonably old
352 polymorphic inversions.

353 The trend in the excess of rare alleles detected by the Tajima's D in Barcelona
354 and Mt. Parnes populations is similar to that observed for genes located inside the
355 inverted regions in previous studies (Rozas *et al.*, 1999; Munté *et al.*, 2005). The
356 hypothesis of a recent population expansion was further supported by the R_2 statistic.
357 The number of significant R_2 estimates is higher in Mt. Parnes than in Barcelona,
358 suggesting that Mt. Parnes had a larger population expansion, in accordance with its
359 estimated larger effective population size (Araúz *et al.*, 2009, 2011). When considering
360 expansion within arrangements, the null hypothesis of constant size was only rejected
361 for O_{3+4} arrangement. Since this arrangement is considered to be warm adapted
362 (Balanyà *et al.*, 2004), the inferred expansion in the O_{3+4} arrangement could be the
363 result of its increase in frequency due to selection after the Pleistocenic glaciations.
364 Interestingly, this arrangement shows increased frequency worldwide matching recent
365 global warming (Balanyà *et al.*, 2006).

366

367 **Divergence time of inversions**

368 The age of inversions can be estimated considering their likely unique origin (Powell,
369 1997). These estimates can be inferred from the time of coalescence of the sequences,
370 taking into account that variability accumulated in the sequences is proportional to its
371 origin. Divergence was initially estimated using two species of the *obscura* group as
372 outgroups (*D. madeirensis* and *D. pseudoobscura*). On one hand, *D. madeirensis* is
373 more closely related to *D. subobscura*, since their divergence time was estimated at
374 about 0.63 Mya by Ramos-Onsins *et al.*, (1998). This species is endemic of Madeira
375 Island and has a small effective population size. However, its nucleotide diversity is
376 similar to *D. subobscura*, and therefore not showing the lower levels of variation
377 typically found in an island (Ramos-Onsins *et al.*, 1998; Khadem *et al.*, 2001). On the
378 other hand, *D. pseudoobscura* divergence time with respect to *D. subobscura* was
379 estimated to be 17.7 Myr (Tamura *et al.*, 2004) or around 8 Myr (Ramos-Onsins *et al.*,
380 1998). *D. pseudoobscura* is a native species from North America and its effective size
381 could be quite similar to *D. subobscura* (Pascual *et al.*, 2000). For *Abi* and *Atpa* genes,
382 values of silent nucleotide diversity were quite similar to the divergence between *D.*
383 *subobscura* and *D. madeirensis*. Thus, we have used the more divergent outgroup, *D.*
384 *pseudoobscura*, to estimate the age of the inversions.

385 Ages were calculated using two divergence times between *D. subobscura* and *D.*
386 *pseudoobscura*. The divergence time obtained by Tamura *et al.* (2004) is more reliable
387 since it is based on a large multilocus dataset; nonetheless the time estimate by Ramos-
388 Onsins *et al.* (1998), although only based on the *rp49* gene, allows the comparison with
389 previous studies (see below). The age of inversions estimated using the two most
390 divergent sequences seems more sensitive to differences in sample size or to differences
391 in the genetic content between populations, as observed by the higher fluctuation of the
392 estimated ages when mixing populations (1.01-2.24 Myr) or estimating ages for each

393 population separately (0.77-1.70 and 0.87-1.93 Myr for Barcelona and Mt. Parnes
394 respectively, see Table 5). To avoid this potential bias we suggest using the mean π_{sil}
395 and considering individuals from a unique population (bold values in Table 5). The age
396 estimates obtained in the present study (0.48 Myr for O_{ST} and 0.41 Myr for $O_{\underline{3+4}}$) are
397 slightly older than those based on genes *Acph-1* (0.26 Myr for O_{ST} and 0.31 Myr for
398 $O_{\underline{3+4}}$ in Navarro-Sabaté *et al.*, 1999) and *rp49* (0.24 Myr for O_{ST} and 0.33 Myr for $O_{\underline{3+4}}$
399 in Rozas *et al.*, 1999) despite using the same divergence time (Ramos-Onsins *et al.*,
400 1998). Differences could be due to intrinsic characteristics of the different genes and to
401 their distinct genetic location. Genes *Acph-1* and *rp49* are closely located inside
402 inversion O_3 near its distal breakpoint (Munté *et al.*, 2005) while *Fmr1* gene is located
403 far away from them, inside inversion O_4 near the distal breakpoint. However, in the
404 three studies the age distance between the two arrangements is reduced, differing in
405 0.05 Myr in (Navarro-Sabaté *et al.*, 1999), in 0.09 Myr in (Rozas *et al.*, 1999), and in
406 0.07 Myr in the present study. Since genes can differ in their selective pressure,
407 mutation rates, or rate of recombination, each of which can affect estimates of their
408 coalescence time, a multilocus approach as in Tamura *et al.* (2004) is preferred to more
409 precisely infer the age of inversions. Thus considering all data available, we can
410 conclude that the two arrangements arose approximately at the same time.

411

412 **Gene flux is not homogeneously distributed across the chromosomal arrangement**

413 Gene flux is defined as the probability of allele exchange during meiosis in
414 heterokaryotypic females, including both crossing-over and gene conversion (Navarro *et*
415 *al.*, 1997). For the five genes located outside the inversions, no significant genetic
416 differentiation between the O_{ST} and $O_{\underline{3+4}}$ chromosomal arrangements was observed in
417 the present study, suggesting extensive gene flux between them. High levels of gene

418 flux between arrangements were also found in other *Drosophila* species, such as *D.*
419 *pseudoobscura* (Schaeffer and Anderson, 2005) and *D. buzzatii* (Laayouni *et al.*, 2003).

420 Despite the extensive exchange found outside the inverted region, significant
421 genetic differentiation was found within the inverted region. For the *Fmr1* gene, F_{ST}
422 values were similar to those obtained in previous studies using other genes located
423 within the studied inverted region (Rozas *et al.*, 1999; Munté *et al.*, 2005). The adaptive
424 value of inversions in *D. subobscura* has been supported by many observations, such as
425 the latitudinal clines for some chromosomal arrangements (Prevosti *et al.*, 1988;
426 Balanyà *et al.*, 2006) or their seasonal fluctuations (Rodríguez-Trelles *et al.*, 1996). In
427 particular, O_{ST} and O_{3+4} present opposite latitudinal clines, with O_{ST} being more
428 frequent in northern Europe and O_{3+4} being more frequent in the south (Solé *et al.*,
429 2002; Balanyà *et al.*, 2004). However, none of the nine genes located across the O_{3+4}
430 arrangement studied so far show any non-synonymous changes differentiating the two
431 arrangements (Munté *et al.*, 2005 and present study), despite the *Fmr1* gene was a
432 candidate gene to be involved in thermal adaptation. Knowing that selective pressure
433 could focus on regulatory regions instead of coding regions (Torgerson *et al.*, 2009), the
434 differential basal expression for the thermal candidate gene *Hsp70* among O_{3+4} and O_{ST}
435 arrangements (Calabria *et al.*, 2012) may be due to changes in regulatory sequences.
436 Transcriptome variation in *Drosophila* has been shown to be driven by both cis- and
437 trans- regulatory elements, (Genissel *et al.*, 2008), thus future studies should focus on
438 the sequencing of the regulatory regions of *Fmr1* or other thermal candidate genes in SI
439 region and study their pattern of expression among different chromosomal
440 arrangements.

441

442 **High gene flow between populations despite the adaptive value of the inversions**

443 The two populations selected for the present study had been previously observed to
444 differ significantly in frequency of chromosomal inversions (Araúz *et al.*, 2009). For
445 instance, the O_{3+4} arrangement constituted 28% of the Barcelona population, but its
446 frequency was 52% in Mt. Parnes. Furthermore, inversions that did not appear or had
447 very low frequency in one population, showed a moderate frequency in the other (15%
448 O_{3+4+7} , 24% O_{3+4+8} and 4% O_{3+4+1} in Barcelona vs 0% O_{3+4+7} , 3% O_{3+4+8} and 14% O_{3+4+1}
449 in Mt. Parnes). These differences could reflect historical processes with these two
450 populations representing separate Pleistocene refugia (Taberlet *et al.*, 1998; Hewitt,
451 1999) with subsequent low levels of gene flow and/or differential selection to local
452 environments (Kovacevic and Schaeffer, 2000).

453 Low levels of gene flow can be ruled out since in the present work no significant
454 DNA sequence differentiation was detected comparing the same arrangement between
455 populations independently of the gene location. Similarly, in *Drosophila pseudoobscura*
456 (Schaeffer *et al.*, 2003), no genetic differentiation was found between populations
457 within arrangements using genes located in the inverted region. In *D. subobscura* no
458 significant DNA sequence differentiation was found among three European populations
459 (one from Holland and two from Spain) when analysing restriction length
460 polymorphism in O_{ST} and O_{3+4} arrangements (Rozas *et al.*, 1995). No significant DNA
461 sequence differentiation was found between El Pedroso (Spain) and Bizerte (Tunisia)
462 for the $O_{[3+4]}$ group (including O_{3+4+7} and O_{3+4+8}), despite in these group different
463 overlapping inversions were included (Sánchez-Gracia *et al.*, 2011). Furthermore, the
464 same sequence associated to the O_{3+4+1} chromosomal arrangement was found on the
465 *Odh* gene in Barcelona and Mt. Parnes populations (Araúz *et al.* 2011). An independent
466 origin of this association in both populations seems unlikely and the more parsimonious
467 explanation is gene flow between them. Overall, these results suggest that gene flow

468 between populations of *D. subobscura* is high, and agree with several studies detecting
469 non-negligible gene flow between natural European populations of this species (Latorre
470 *et al.*, 1992; Pascual *et al.*, 2001; Zivanovic *et al.*, 2007). High gene flow between
471 distant populations could be attributed to migration both by passive transportation
472 associated to human activities (Pascual *et al.*, 2007) and to the active dispersal
473 capabilities of *D. subobscura* (Serra *et al.*, 1987). Thus, in European populations, if
474 gene flow is extensive, the clinal frequencies of some inversions are likely to be
475 maintained by strong selection (Prevosti *et al.*, 1988; Balanyà *et al.*, 2006). Nonetheless,
476 comparisons between introduced (American) and ancestral (European) populations
477 would lead to genetic differentiation within arrangement due to the founder effect, as
478 observed for the *Odh* gene (Mestres *et al.*, 2004; Gómez-Baldó *et al.*, 2008; Araúz *et*
479 *al.*, 2011).

480 In summary, despite using candidate genes for thermal adaptation (which are
481 suspected to be selected in the chromosome inversions where they are located) we
482 observed no differences between populations within arrangement, although it would be
483 useful to include other thermal candidate genes located in SI region. Moreover, no
484 fitness differences were observed when comparing heterokaryotypes carrying both
485 chromosomes belonging to the same or to different populations (Pegueroles *et al.*,
486 2010a). Thus, we conclude that the adaptive value of inversions can be maintained
487 regardless the lack of genetic differentiation within arrangements from different
488 populations. These results do not agree with the expectation of the coadaptation model
489 that predicts genetic differentiation between populations, but do support the local
490 adaptation hypothesis of Kirkpatrick and Barton (2006).

491

492 **DATA ARCHIVING**

493 Sequence data have been submitted to GenBank: accession numbers JN882376-
494 JN882575.

495

496 **CONFLICT OF INTEREST**

497 The authors declare no conflict of interest.

498

499 **ACKNOWLEDGEMENTS**

500 We thank Ferran Palero for his valuable comments on the manuscript. We also thank
501 Joan Balanyà for the analyses of the polytene chromosomes, Pedro A. Araúz for the
502 collection of the flies and the collaboration with the crosses to obtain the
503 homokaryotypic lines, Gemma Calabria and Olga Dolgova for their collaboration in the
504 *in situ* hybridizations, and Carla Rego for *D. madeirensis* samples. Furthermore, we
505 want to thank Vanessa Bauer DuMont for her technical support. Research was funded
506 by projects CGL2006-13423-C02-02, BFU2009-07564 and CTM2010-22218-C02-02
507 from the Ministerio de Ciencia y Tecnología (MCYT, Spain). The authors are part of
508 the research group 2009SGR-636 from the Generalitat de Catalunya.

509

510 Supplementary information is available at Heredity's website

511

512 **REFERENCES**

513 Andolfatto P, Depaulis F, Navarro A (2001). Inversion polymorphisms and nucleotide
514 variability in *Drosophila*. *Genet Res* **77**: 1-8.

515 Araúz PA, Mestres F, Pegueroles C, Arenas C, Tzannidakis G, Krimbas CB *et al.*

516 (2009). Tracking the origin of the American colonization by *Drosophila*

517 *subobscura*: genetic comparison between Eastern and Western Mediterranean
518 populations. *J Zool Syst Evol Res* **47**: 25-34.

519 Araúz PA, Peris-Bondia F, Latorre A, Serra L, Mestres F (2011). Molecular evidence to
520 suggest the origin of a colonization: *Drosophila subobscura* in America.
521 *Genetica* **139**:1477–1486.

522 Balanyà J, Oller JM, Huey RB, Gilchrist GW, Serra L (2006). Global genetic change
523 tracks global climate warming in *Drosophila subobscura*. *Science* **313**: 1773-
524 1775.

525 Balanyà J, Serra L, Gilchrist GW, Huey RB, Pascual M, Mestres F *et al.* (2003).
526 Evolutionary pace of chromosomal polymorphism in colonizing populations of
527 *Drosophila subobscura*: an evolutionary time series. *Evolution* **57**: 1837-1845.

528 Balanyà J, Solé E, Oller J, Sperlich D, Serra L (2004). Long-term changes in the
529 chromosomal inversion polymorphism of *Drosophila subobscura*. II. European
530 populations. *J Zool Syst Evol Res* **42**: 191-201.

531 Calabria G, Dolgova O, Rego C, Castañeda LE, Rezende EL, Balanyà J *et al.* (2012).
532 Hsp70 protein levels and thermotolerance in *Drosophila subobscura*: a
533 reassessment of the thermal co-adaptation hypothesis. *J Evol Biol* **25**: 691-700.

534 Dobzhansky T (1943). Temporal changes in the composition of populations of *D.*
535 *pseudoobscura*. *Genetics* **28**: 162-186.

536 Dobzhansky T (1950). Genetics of natural populations. XIX. Origin of heterosis
537 through natural selection in populations of *Drosophila pseudoobscura*. *Genetics*
538 **35**: 288-302.

539 Fu YX, Li WH (1993). Statistical tests of neutrality of mutations. *Genetics* **133**: 693-
540 709.

541 Genissel A, McIntyre LM, Wayne ML, Nuzhdin SV (2008). Cis and trans regulatory
542 effects contribute to natural variation in transcriptome of *Drosophila*
543 *melanogaster*. *Mol Biol Evol* **25**: 101–110.

544 Gómez-Baldó L, Latorre A, Serra L, Mestres F (2008). Molecular variation in the *Odh*
545 gene in Chilean natural populations of *Drosophila subobscura*. *Hereditas* **145**:
546 154–162.

547 Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and
548 analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41**: 95-98.

549 Hewitt GM (1999). Post-glacial re-colonization of European biota. *Biol J Linn Soc* **68**:
550 87-112.

551 Hoffmann AA, Rieseberg LH (2008). Revisiting the impact of inversions in evolution:
552 from population genetic markers to drivers of adaptive shifts and speciation?
553 *Annu Rev Ecol Evol Syst* **39**: 21-42.

554 Hudson RR (2000). A new statistic for detecting genetic differentiation. *Genetics* **155**:
555 2011-2014.

556 Hudson RR (2001). Two-locus sampling distributions and their application. *Genetics*
557 **159**: 1805-1817.

558 Hudson RR, Slatkin M, Maddison WP (1992). Estimation of levels of gene flow from
559 DNA-sequence data. *Genetics* **132**: 583-589.

560 Kennington WJ, Partridge L, Hoffmann AA (2006). Patterns of diversity and linkage
561 disequilibrium within the cosmopolitan inversion *In(3R)Payne* in *Drosophila*
562 *melanogaster* are indicative of coadaptation. *Genetics* **172**: 1655-1663.

563 Khadem M, Rozas J, Segarra C, Aguadé M (2001). DNA variation at the *rp49* gene
564 region in *Drosophila madeirensis* and *Drosophila subobscura* from Madeira:

565 inferences about the origin of an insular endemic species. *J Evol Biol* **14**: 379-
566 387.

567 Kirkpatrick M, Barton N (2006). Chromosome inversions, local adaptation and
568 speciation. *Genetics* **173**: 419-434.

569 Kovacevic M, Schaeffer SW (2000). Molecular population genetics of X-linked genes
570 in *Drosophila pseudoobscura*. *Genetics* **156**: 155-172.

571 Krimbas CB (1993). *Drosophila subobscura: Biology, Genetics and Inversion*
572 *polymorphism*, Verlag Dr. Kovac, Hamburg.

573 Krimbas CB, Loukas M (1980). The inversion polymorphism of *Drosophila*
574 *subobscura*. *Evol Biol* **12**: 163-234.

575 Krimbas CB, Powell JF (1992). *Drosophila Inversion Polymorphism*. CRC Press, Boca
576 Raton, Florida.

577 Laayouni H, Garcia-Franco F, Chavez-Sandoval BE, Trotta V, Beltran S, Coromines M
578 *et al.* (2007). Thermal evolution of gene expression profiles in *Drosophila*
579 *subobscura*. *BMC Evol Biol* **7**: 42.

580 Laayouni H, Hasson E, Santos M, Fontdevila A (2003). The evolutionary history of
581 *Drosophila buzzatii*. XXXV. Inversion polymorphism and nucleotide variability
582 in different regions of the second chromosome. *Mol Biol Evol* **20**: 931-944.

583 Laayouni H, Santos M, Fontdevila A (2000). Toward a physical map of *Drosophila*
584 *buzzatii*. Use of randomly amplified polymorphic dna polymorphisms and
585 sequence-tagged site landmarks. *Genetics* **156**: 1797-1816.

586 Latorre A, Hernandez C, Martinez D, Castro JA, Ramón M, Moya A (1992). Population
587 structure and mitochondrial DNA gene flow in Old World populations of
588 *Drosophila subobscura*. *Heredity* **68**: 15-24.

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

591 McGaugh SE, Noor MAF (2012). Genomic impacts of chromosomal inversions in
592 parapatric *Drosophila* species. *Phil Trans R Soc B* **367**: 422-429.

593

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

596 Mestres F, Abad L, Sabater-Muñoz B, Latorre A, Serra L (2004). Colonization of
597 America by *Drosophila subobscura*: Association between *Odh* gene haplotypes,
598 lethal genes and chromosomal arrangements. *Genes Genet. Syst.* **79**: 233–244.

599 Munté A, Rozas J, Aguadé M, Segarra C (2005). Chromosomal inversion
600 polymorphism leads to extensive genetic structure: a multilocus survey in
601 *Drosophila subobscura*. *Genetics* **169**: 1573-1581.

602 Navarro-Sabaté A, Aguadé M, Segarra C (1999). The relationship between allozyme
603 and chromosomal polymorphism inferred from nucleotide variation at the *AcpH*-
604 *I* gene region of *Drosophila subobscura*. *Genetics* **153**: 871-889.

605 Navarro A, Betran E, Barbadilla A, Ruiz A (1997). Recombination and gene flux
606 caused by gene conversion and crossing over in inversion heterokaryotypes.
607 *Genetics* **146**: 695-709.

608 Nei M (1987). *Molecular Evolutionary Genetics*. Columbia University Press, New
609 York.

610 Nei M, Gojobori T (1986). Simple methods for estimating the numbers of synonymous
611 and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**: 418-426.

612 Pascual M, Aquadro CF, Soto V, Serra L (2001). Microsatellite variation in colonizing
613 and palearctic populations of *Drosophila subobscura*. *Mol Biol Evol* **18**: 731-
614 740.

615 Pascual M, Chapuis MP, Mestres F, Balanyà J, Huey RB, Gilchrist W *et al.* (2007).
616 Introduction history of *Drosophila subobscura* in the New World: a
617 microsatellite-based survey using ABC methods. *Mol Ecol* **16**: 3069-3083.

618 Pascual M, Schug MD, Aquadro CF (2000). High density of long dinucleotide
619 microsatellites in *Drosophila subobscura*. *Mol Biol Evol* **17**: 1259-1267.

620 Pegueroles C, Ordóñez V, Mestres F, Pascual M (2010a). Recombination and selection
621 in the maintenance of the adaptive value of inversions. *J Evol Biol* **23**: 2709-
622 2717.

623 Pegueroles C, Araúz PA, Pascual M, Mestres F. (2010b). A recombination survey using
624 microsatellites: the O chromosome of *Drosophila subobscura*. *Genetica* **138**:
625 795-804.

626 Pina-Martins F, Paulo OS (2008). CONCATENATOR: Sequence data matrices
627 handling made easy. *Mol Ecol Resour* **8**: 1254-1255.

628 Powell JR (1997). *Progress and prospects in evolutionary biology: The Drosophila*
629 *model*. Oxford University Press, New York.

630 Prevosti A, Ribó G, Serra L, Aguadé M, Balanyà J, Monclus M *et al.* (1988).
631 Colonization of America by *Drosophila subobscura*: Experiment in natural
632 populations that supports the adaptive role of chromosomal inversion
633 polymorphism. *Proc Natl Acad Sci USA* **85**: 5597-5600.

634 Ramos-Onsins SE, Rozas J (2002). Statistical properties of new neutrality tests against
635 population growth. *Mol Biol Evol* **19**: 2092-2100.

636 Ramos-Onsins SE, Segarra C, Rozas J, Aguade M (1998). Molecular and chromosomal
637 phylogeny in the *obscura* group of *Drosophila* inferred from sequences of the
638 *rp49* gene region. *Mol Phyl Evol* **9**: 33-41.

639 Rodríguez-Trelles F, Alvarez G, Zapata C (1996). Time-series analysis of seasonal
640 changes of the O inversion polymorphism of *Drosophila subobscura*. *Genetics*
641 **142**: 179-87.

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

645 Rozas J, Segarra C, Ribó G, Aguadé M (1999). Molecular population genetics of the
646 *rp49* gene region in different chromosomal inversions of *Drosophila*
647 *subobscura*. *Genetics* **151**: 189-202.

648 Rozas J, Segarra C, Zapata C, Alvarez G, Aguade M (1995). Nucleotide polymorphism
649 at the *rp49* region of *Drosophila subobscura*: Lack of geographic subdivision
650 within chromosomal arrangements in Europe. *J Evol Biol* **8**: 355-367.

651 Sánchez-Gracia A, Rozas J (2011). Molecular population genetics of the *OBP83*
652 genomic region in *Drosophila subobscura* and *D. guanche*: contrasting the
653 effects of natural selection and gene arrangement expansion in the patterns of
654 nucleotide variation. *Heredity* **106**:191-201.

655 Schaeffer SW, Anderson WW (2005). Mechanisms of genetic exchange within the
656 chromosomal inversions of *Drosophila pseudoobscura*. *Genetics* **171**: 1729-
657 1739.

658 Schaeffer SW, Goetting-Minesky MP, Kovacevic M, Peoples JR, Graybill JL, Miller
659 JM *et al.* (2003). Evolutionary genomics of inversions in *Drosophila*

660 *pseudoobscura*: Evidence for epistasis. *Proc Natl Acad Sci USA* **100**: 8319-
661 8324.

662 Serra L, Pegueroles G, Mestres F (1987). Capacity of dispersal of a colonizing species:
663 *Drosophila subobscura*. *Genetica* **73**: 223-235.

664 Solé E, Balanyà J, Sperlich D, Serra L (2002). Long-term changes in the chromosomal
665 inversion polymorphism of *Drosophila subobscura*. I. Mediterranean
666 populations from southwestern Europe. *Evolution* **56**: 830-835.

667 Stevison LS, Hoehn KB, Noor MAF (2011). Effects of inversions on within- and
668 between-species recombination and divergence. *Genome Biol Evol* **3**: 830–841.

669 Stump AD, Pombi M, Goeddel L, Ribeiro JMC, Wilder JA, della Torre A *et al.* (2007).
670 Genetic exchange in 2La inversion heterokaryotypes of *Anopheles gambiae*.
671 *Insect Mol Biol* **16**: 703-709.

672 Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998). Comparative
673 phylogeography and postglacial colonization routes in Europe. *Mol Ecol* **7**: 453-
674 464.

675 Tajima F (1989). Statistical methods to test for nucleotide mutation hypothesis by DNA
676 polymorphism. *Genetics* **123**: 585-595.

677 Tamura K, Subramanian S, Kumar S (2004). Temporal patterns of fruit fly (*Drosophila*)
678 evolution revealed by mutation clocks. *Mol Biol Evol* **21**: 36-44.

679 Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA4: Molecular evolutionary
680 genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596-1599.

681 Thompson JD, Higgins DG, Gibson TJ (1994). CLUSTAL W: Improving the sensitivity
682 of progressive multiple sequence alignment through sequence weighting,
683 position-specific gap penalties and weight matrix choice. *Nucl Acid Res* **22**:
684 4673-4680.

685 Torgerson DG, Boyko AR, Hernandez RD, Indap A, Hu X, White TJ *et al.* (2009).
686 Evolutionary processes acting on candidate *cis*-regulatory regions in humans
687 inferred from patterns of polymorphism and divergence. *PLoS Genet* **5**:
688 e1000592.

689 Watterson GA (1975). Number of segregating sites in genetic models without
690 recombination. *Theor Popul Biol* **7**: 256-276.

691 Zivanovic G, Arenas C, Mestres F (2007). The genetic structure of Balkan populations
692 of *Drosophila subobscura*. *Hereditas* **144**: 120-128.

693

694 **Table 1:** Chromosomal location of the 6 genes (newly determined by *in situ*
695 hybridization), the sequenced fragment length, and primers used in the amplification
696 and sequencing

Gene Symbol ⁴	Chrom. location	Length (bp)	Primers
<i>Pif1A</i> (CG42599)	79D ⁵	1820	5'-AATGTATCACAAGGAGAACG-3' ^{1,2,3} 5'-CTCCTGGTAGTACTGCAGAT-3' ^{1,2,3} 5'-GTCCAACACTATGGTTGATGCC-3' ² 5'-ACGACTTTACAGCTAGCTGG-3' ²
<i>Abi</i> (CG9749)	81A ⁵	1566	5'-CCTTGTCCGCATAGTAGTCA-3' ^{1,2} 5'-GTGAGTAGTAATTAGGTTTCG-3' ^{1,2} 5'-ACTCACGTGGCATACTGCTC-3' ²
<i>Sqd</i> (CG16901)	85D ⁵	1436	5'-AACTAACCTTGTCCTCTCC-3' ^{1,3} 5'-TTACACACGCTTCGTCAGTT-3' ^{1,2,3} 5'-CATTATAGTAGCCACCAGGA-3' ² 5'-TCTTGGTGGTTGTTGCTGTG-3' ²
<i>Yrt</i> (CG9764)	86E	911	5'-CTGGACATCATCGAGAAGGA-3' ^{1,2,3} 5'-ACATTGGCCAGCTTCACTTG-3' ^{1,3} 5'-AACTGGAGTCTGGCTGGCGT-3' ² 5'-TACGCCGGATCATCGTTGCT-3' ²
<i>Atpa</i> (CG5670)	87C	1490	5'-TCATAAGATCTCTCCTGAGG-3' ^{1,2,3} 5'-GCAATATCCTCAACGGTCTC-3' ^{1,2,3} 5'-CATCATTGAGGCGCGCAGCT-3' ² 5'-CCAGGGCTAGTTCTATCGTA-3' ²
<i>Fmr1</i> (CG6203)	98A ⁵	1984	5'-ACAGCCAAGTCGTTCTACCA-3' ^{1,2,3} 5'-CCATTCACCAGACCTTCCTT-3' ^{1,2,3} 5'-GGCAAGGTAATCGGCAAGAA-3' ²

697 ¹ primers used in the amplifications; ² primers used for sequencing; ³ primers previously designed

698 (Laayouni *et al.*, 2007); 4 Gene symbol of the homologous gene in *D. melanogaster* and annotation

699 symbol in brackets; ⁵ mislocated genes in Laayouni *et al.* (2007).

700 **Table 2:** Nucleotide variation and divergence for each gene and arrangement (O_{3+4} and O_{ST}) in Mt.
 701 Parnes (MP) and Barcelona (BC) populations

	<i>Pif1A</i>		<i>Abi</i>		<i>Sqd</i>		<i>Yrt</i>		<i>Atpa</i>		<i>Fmr1</i>	
O_{3+4}	MP	BC	MP	BC	MP	BC	MP	BC	MP	BC	MP	BC
<i>n</i>	12	12	12	7	12	10	12	12	12	12	11	10
<i>h</i>	12	12	12	7	12	10	12	12	12	11	11	10
<i>S</i>	120	114	32	24	25	21	33	35	22	27	38	28
singletons	55	42	20	15	16	14	18	21	9	16	25	18
π	0.021	0.022	0.005	0.006	0.005	0.005	0.010	0.010	0.005	0.005	0.005	0.004
π_{sil}	0.023	0.023	0.012	0.014	0.006	0.006	0.034	0.035	0.015	0.016	0.009	0.008
θ_{sil}	0.026	0.024	0.017	0.015	0.008	0.007	0.042	0.041	0.013	0.019	0.012	0.010
$K_{sil-madeir}$	0.034	0.032	0.013	0.013	0.030	0.028	0.078	0.078	0.018	0.018	0.025	0.025
$K_{sil-pseudo}$	0.191	0.190	0.173	0.173	0.116	0.115	0.327	0.326	0.208	0.209	0.151	0.151
<hr/>												
O_{ST}												
<i>n</i>	5	7	2	3	5	6	5	6	5	6	4	6
<i>h</i>	5	7	2	3	5	5	5	6	4	6	4	6
<i>S</i>	82	95	11	16	11	9	23	27	17	12	26	34
singletons	57	49	11	16	8	5	20	22	7	7	16	18
π	0.022	0.023	0.007	0.007	0.004	0.003	0.011	0.011	0.006	0.003	0.007	0.007
π_{sil}	0.023	0.024	0.017	0.017	0.005	0.004	0.037	0.037	0.014	0.011	0.014	0.014
θ_{sil}	0.024	0.025	0.017	0.017	0.005	0.004	0.040	0.043	0.013	0.011	0.014	0.014
$K_{sil-madeir}$	0.032	0.038	0.012	0.013	0.031	0.030	0.077	0.078	0.018	0.015	0.020	0.023
$K_{sil-pseudo}$	0.190	0.194	0.173	0.173	0.116	0.115	0.322	0.326	0.208	0.207	0.149	0.149

702 *n*, sample size; *h*, number of haplotypes; *S*, number of polymorphic sites; π , nucleotide diversity including
 703 all sites; π_{sil} , nucleotide diversity in synonymous sites and non-coding positions; θ_{sil} , heterozygosity in
 704 silent sites; $K_{sil-madeir}$, divergence per silent site between *D. subobscura* and *D. madeirensis*; $K_{sil-pseudo}$,
 705 divergence per silent site between *D. subobscura* and *D. pseudoobscura*.

706 **Table 3:** Genetic differentiation between O_{3+4} and O_{ST} arrangements from Mt. Parnes (MP) and
 707 Barcelona (BC)

		<i>Pif1A</i>	<i>Sqd</i>	<i>Yrt</i>	<i>Atpa</i>	<i>Fmr1</i>	4Conc ¹	5Conc ²
O_{3+4} MP vs O_{3+4} BC	<i>Dxy</i>	0.021	0.004	0.010	0.005	0.004	0.010	0.009
	<i>Fst</i>	-0.017	-0.050	-0.008	-0.044	0.028	-0.0214	-0.015
	<i>P(Snn)</i>	-	-	-	-	-	-	-
O_{ST} MP vs O_{ST} BC	<i>Dxy</i>	0.022	0.003	0.01	0.005	0.007	0.011	0.010
	<i>Fst</i>	-0.014	-0.048	-0.117	0.079	-0.100	0.038	0.013
	<i>P(Snn)</i>	-	-	-	-	-	-	-
O_{3+4} MP vs O_{ST} MP	<i>Dxy</i>	0.020	0.004	0.010	0.005	0.011	0.010	0.010
	<i>Fst</i>	-0.053	-0.031	-0.028	0.035	0.466	-0.054	0.094
	<i>P(Snn)</i>	-	-	-	-	***	-	**
O_{3+4} BC vs O_{ST} BC	<i>Dxy</i>	0.022	0.004	0.010	0.004	0.011	0.010	0.011
	<i>Fst</i>	0.018	0.011	-0.070	-0.048	0.497	0.022	0.155
	<i>P(Snn)</i>	-	-	-	-	***	-	**
O_{3+4} MP vs O_{ST} BC	<i>Dxy</i>	0.022	0.004	0.010	0.004	0.012	0.010	0.011
	<i>Fst</i>	0.006	-0.021	-0.06	-0.037	0.487	0.021	0.153
	<i>P(Snn)</i>	-	-	-	-	***	-	*
O_{3+4} BC vs O_{ST} MP	<i>Dxy</i>	0.021	0.004	0.010	0.006	0.011	0.010	0.010
	<i>Fst</i>	-0.031	0.025	-0.047	0.089	0.487	-0.026	0.116
	<i>P(Snn)</i>	-	-	-	-	**	-	*

708 Statistical significance was assessed using the p-value of the Snn statistic (-, not significant; *,
 709 $0.01 < P < 0.05$; **, $0.001 < P < 0.01$; ***, $P < 0.001$); ¹ concatenated data set of genes *Pif1A*, *Sqd*, *Yrt* and
 710 *Atpa*; ² concatenated data set with the addition of *Fmr1* gene.

711 **Table 4:** Neutrality tests and test of population expansion for each gene separately and
712 all combined genes using Ramos-Onsins and Rozas' R_2 . For *Fmr1* gene, neutrality and
713 demographic tests were calculated within arrangement after removing recombinant
714 individuals. P-values were obtained by coalescent simulations without recombination¹,
715 with estimated levels (ρ) of recombination² and with free recombination³

Mt. Parnes	Tajima's D	Fu and Li's D	R_2	ρ	p-value ¹	p-value ²	p-value ³
<i>Pif1A</i>	-0.723	-1.142	0.101	100	0.129	0.000	0.000
<i>Abi</i>	-1.099	-0.603	0.091	69	0.034	0.010	0.008
<i>Sqd</i>	-1.017	-1.299	0.092	81	0.062	0.035	0.028
<i>Yrt</i>	-0.871	-0.109	0.096	100	0.079	0.019	0.019
<i>Atpα</i>	0.942	1.104	0.176	13	0.881	0.945	0.990
<i>Fmr1</i>	-1.100	-0.753	0.088	27	0.033	0.008	0.000
Concatenate ⁴	-0.799	-0.670	0.100	100	0.070	0.000	0.000
Barcelona							
<i>Pif1A</i>	-0.326	-0.475	0.116	100	0.340	0.101	0.101
<i>Abi</i>	-0.908	-1.569	0.087	68	0.003	0.003	0.002
<i>Sqd</i>	-0.776	-1.335	0.104	62	0.113	0.093	0.079
<i>Yrt</i>	-0.833	-0.418	0.099	100	0.137	0.033	0.033
<i>Atpα</i>	-0.648	-1.033	0.102	79	0.151	0.091	0.069
<i>Fmr1</i>	-0.619	-0.896	0.109	17	0.179	0.133	0.051
Concatenate ⁴	-0.518	-0.810	0.110	100	0.197	0.000	0.000
<i>Fmr1</i> O ₃₊₄ ⁵	-1.696	-1.947	0,062	62	0.001	0.000	0.000
<i>Fmr1</i> O _{ST} ⁶	-0.585	-0.831	0.135	67	0.072	0.227	0.354

716 ⁴ genes *Pif1A*, *Sqd*, *Yrt*, *Atp α* and *Fmr1* included in the concatenate data set. $P < 0.05$ are in bold; ⁵
717 Barcelona and Mount Parnes individuals are included in the analysis; ⁵ Barcelona and Mount Parnes after
718 excluding recombinant individuals.

719

720 **Table 5:** Silent nucleotide diversity and age of the different gene arrangements using
 721 *Fmr1* gene

	Estimates from the two most divergent individuals			Estimates from all individuals but recombinants		
	π_{sil}	Age (Myr) ¹	Age (Myr) ²	π_{sil}	Age (Myr) ¹	Age (Myr) ²
O_{3+4} ³	0.019	2.24	1.01	0.008	0.98	0.44
O_{ST} ³	0.015	1.71	0.77	0.009	1.09	0.49
O_{3+4} MP	0.014	1.70	0.77	0.009	1.00	0.45
O_{3+4} BC	0.016	1.93	0.87	0.008	0.90	0.41
O_{ST} BC	0.014	1.59	0.72	0.009	1.06	0.48

722

723 ¹ age estimation assuming that the species diverged 17.7 Myr ago (Tamura et al., 2004); ² age estimation
 724 assuming that *D. subobscura* and *D. pseudoobscura* diverged 8 Myr ago (Ramos-Onsins et al., 1998); ³
 725 individuals from both populations combined. The least biased estimate is indicated in bold (see
 726 discussion).

727

728

729 **Titles and legends to figures**

730

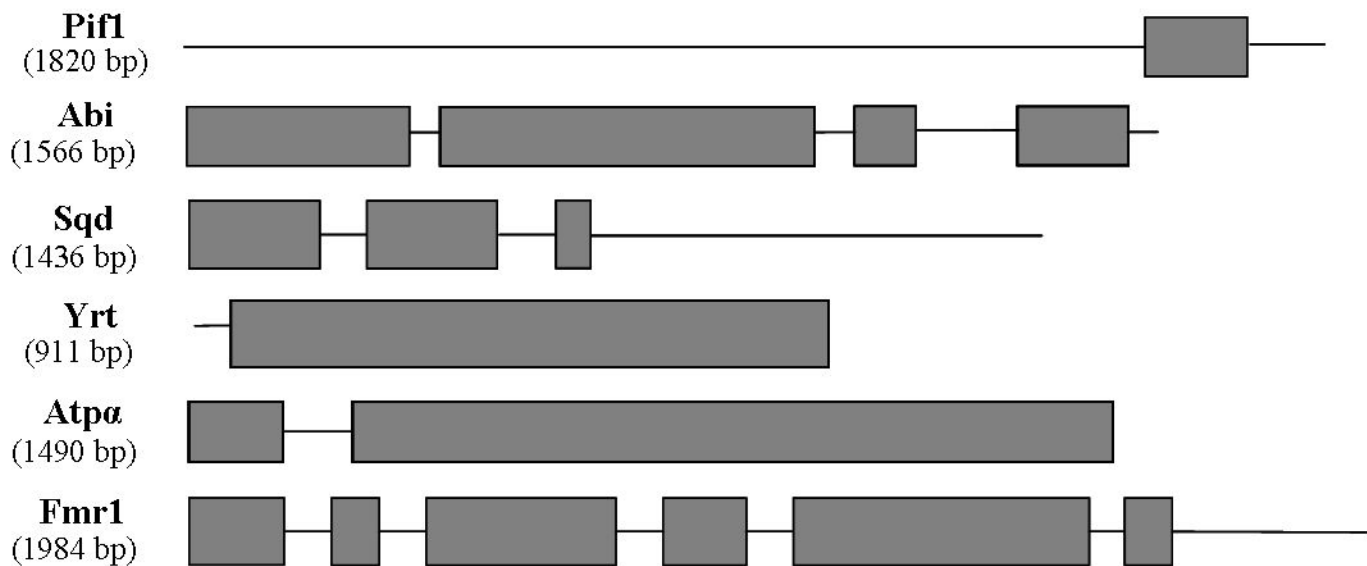
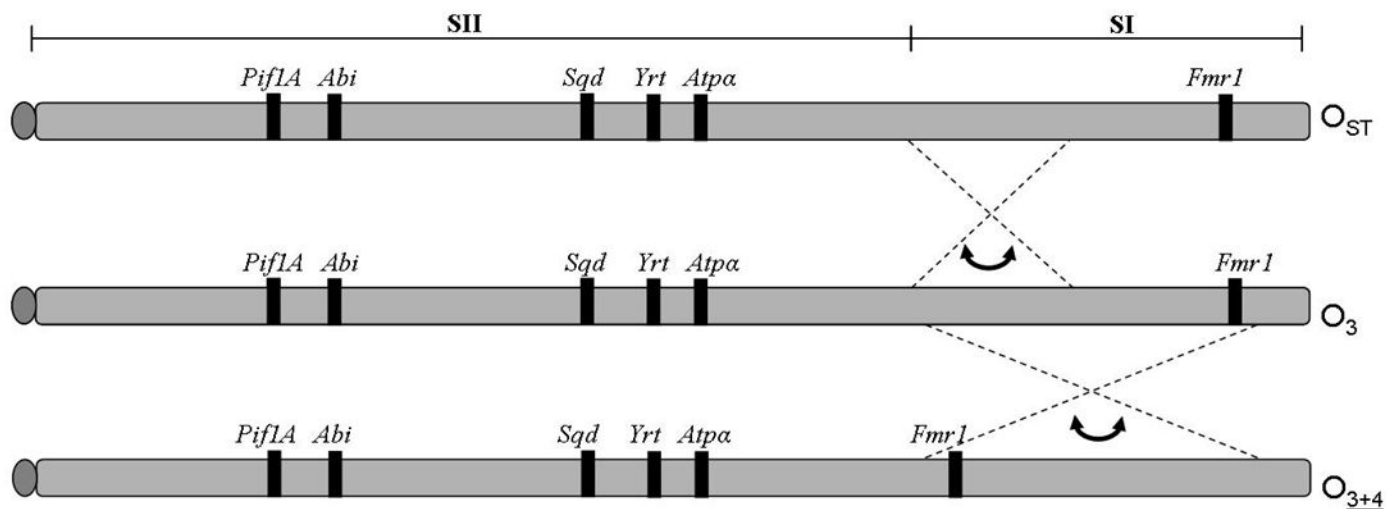
731 **Figure 1:** Top: location of the 6 gene regions studied in the O_{3+4} and O_{ST} arrangements
732 in comparison to the ancestral O_3 arrangement. The O chromosome is divided in
733 sections SI and SII, where SI is the fragment covered by the overlapped inversions O_{3+4} .
734 Bottom: Sequenced fragment for each locus. Exonic regions are identified using grey
735 boxes, and intronic regions are represented by lines. For each gene, the size of the
736 sequenced fragment is given in parentheses.

737

738 **Figure 2:** Polytene chromosome of third instar larvae preparation of O_{3+4}/O_{ST}
739 heterokaryotype. The two overlapped inversions (right of the image) are located
740 opposite to the centromere, in region SI.

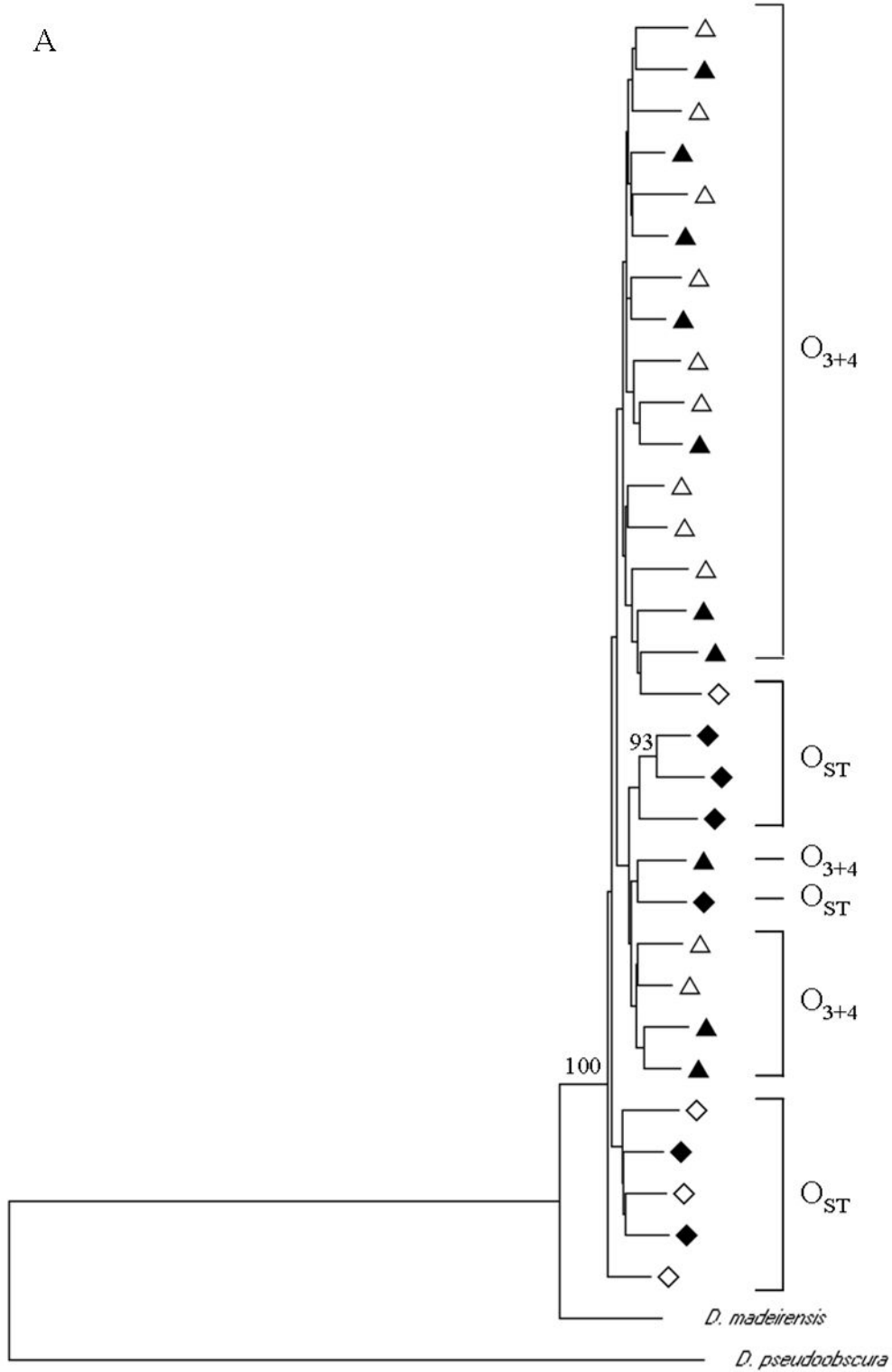
741

742 **Figure 3:** Neighbor-joining gene genealogy based on total nucleotide variation in (A)
743 the concatenated data set of the five genes (*Pif1A*, *Sqd*, *Yrt*, *Atpa* and *Fmr1*) and in (B)
744 the *Fmr1* gene alone. *D. madeirensis* and *D. pseudoobscura* were used as outgroups.
745 Only bootstrap values above 70% are shown. (Δ) O_{3+4} from Mt. Parnes, (\blacktriangle) O_{3+4} from
746 Barcelona, (\diamond) O_{ST} from Mt. Parnes and (\blacklozenge) O_{ST} from Barcelona. Asterisks indicate
747 recombinant individuals.

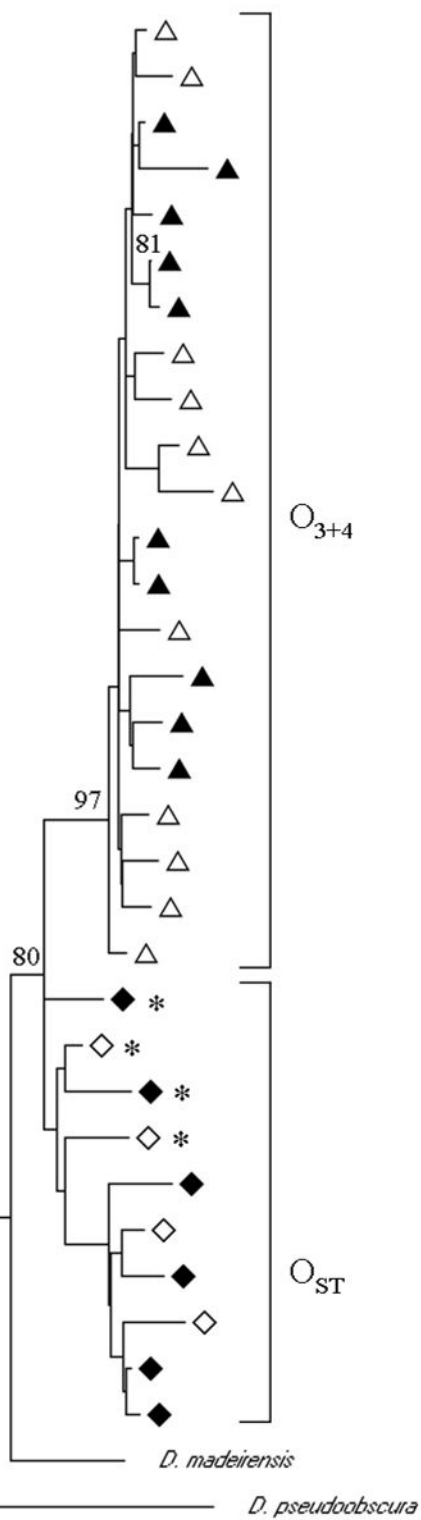




A



B



0.01