

1 Chloroplast microsatellites reveal colonisation and metapopulation dynamics in the  
2 Canary Island pine

3  
4 MIGUEL NAVASCUÉS<sup>1</sup>, ZAFEIRO VAXEVANIDOU<sup>2,3</sup>, SANTIAGO C.  
5 GONZÁLEZ-MARTÍNEZ<sup>2</sup>, JOSÉ CLIMENT<sup>2</sup>, LUIS GIL<sup>3</sup> and BRENT C.  
6 EMERSON<sup>1\*</sup>

7  
8 1 Centre for Ecology, Evolution and Conservation, School of Biological Sciences,  
9 University of East Anglia, Norwich NR4 7TJ, UK

10 2 Departamento de Sistemas y Recursos Forestales, CIFOR-INIA, P.O. Box 8111,  
11 28080 Madrid, Spain

12 3 U.D. Anatomía, Fisiología y Genética Forestal, ETSI de Montes (UPM) Ciudad  
13 Universitaria, 28040 Madrid, Spain

14 \* Author for correspondence (b.emerson@uea.ac.uk)

15

16 **KEYWORDS:** *Pinus canariensis*; Canary Islands; mismatch distribution; chloroplast  
17 microsatellite; population expansion

18

19 **CORRESPONDING AUTHOR:**

20 Brent C. Emerson  
21 School of Biological Sciences  
22 University of East Anglia  
23 Norwich NR4 7TJ, UK  
24 e-mail: b.emerson@uea.ac.uk  
25 fax: (44) 01603 592250

26

27 Running title: *P. canariensis* colonisation and population expansion

28

29 **ABSTRACT**

30 Chloroplast microsatellites are becoming increasingly popular markers for population  
31 genetic studies in plants, but there has been little focus on their potential for  
32 demographic inference. In this work the utility of chloroplast microsatellites for the  
33 study of population expansions was explored. First, we investigated the power of  
34 mismatch distribution analysis and the  $F_S$  test with coalescent simulations of different  
35 demographic scenarios. We then applied those methods to empirical data obtained for  
36 the Canary Island pine (*Pinus canariensis*). The results of the simulations showed that  
37 chloroplast microsatellites are sensitive to sudden population growth. The power of  
38 the  $F_S$  test and accuracy of demographic parameter estimates, such as the time of  
39 expansion, were reduced proportionally to the level of homoplasy within the data. The  
40 analysis of Canary Island pine chloroplast microsatellite data indicated population  
41 expansions for almost all sample localities. Demographic expansions at the island  
42 level can be explained by the colonisation of the archipelago by the pine, while  
43 population expansions of different ages in different localities within an island appear  
44 to be the result of local extinctions and recolonisation dynamics. Comparable  
45 mitochondrial DNA sequence data from a parasite of *P. canariensis*, the weevil  
46 *Brachyderes rugatus*, supports this scenario, suggesting a key role for volcanism in  
47 the evolution of pine forest communities in the Canary Islands.  
48

49 **INTRODUCTION**

50

51 In plants the chloroplast genome is used extensively for evolutionary genetic studies  
52 within species in the same way the mitochondrial genome is used within animal  
53 studies. However, finding enough sequence variation is a challenge due to the low  
54 mutation rates that characterize the chloroplast genome. In contrast, chloroplast  
55 microsatellites, or simple sequence repeats (cpSSRs), present higher levels of  
56 polymorphism and are easily genotyped and this has made them useful and popular  
57 markers for population genetic studies (Provan *et al.*, 2001). Although used  
58 extensively for studying population structure and gene flow, the potential of cpSSRs  
59 to study population demographic history has received little attention. In this study we  
60 investigate in the utility of cpSSR data for the detection of population expansions.

61

62 The study of historical demography by means of genetic information is based on  
63 coalescent theory (see Emerson *et al.*, 2001 for a review). In a stable population  
64 coalescence events are scarcer towards the past giving a genealogy dominated by an  
65 ancient bifurcation with mutations mainly distributed in inter-node branches (King *et al.*,  
66 2000; Reich & Goldstein, 1998). Contrastingly, in the case of sudden population  
67 growth, coalescent events occur mainly during the expansion, leaving a “comblike”  
68 genealogy; and mutations are more abundant along the terminal branches (singleton  
69 mutations) than in inter-node branches (figure 1 shows the main differences of the two  
70 opposing scenarios). As a consequence, population expansions can be detected  
71 because of an excess of singletons (Fu & Li, 1993; Tajima, 1989) or an excess of  
72 haplotypes (as a consequence of the excess of singletons, Fu, 1997). Also, the  
73 divergence between most lineages dates from the time of expansion, producing  
74 unimodal distributions of pairwise genetic distances (Slatkin & Hudson, 1991). The  
75 study of such distributions also allows for the estimation of the time and magnitude of  
76 the population increase (Rogers, 1995; Schneider & Excoffier, 1999).

77

78 The methods for studying population expansions are fairly robust for a genetic marker  
79 evolving under the unrealistic infinite sites model, where singletons and genetic  
80 distances are identified without error. However, in the evolution of sequences under a  
81 finite sites model, parallel and back mutations (i.e. homoplastic mutations) will erase  
82 part of the genetic information producing inaccurate estimates of singletons and  
83 genetic distances. This affects the power of the statistical tests and the estimates of  
84 time and magnitude of the demographic growth (Aris-Brosou & Excoffier, 1996;  
85 Bertorelle & Slatkin, 1995). In nucleotide sequence data, the usual markers for  
86 studying population expansions, the effect of homoplasy is small (Rogers *et al.*, 1996)  
87 and can be accounted for in more sophisticated analyses (Schneider & Excoffier,  
88 1999). In cpSSRs, which evolve in a stepwise fashion, higher levels of homoplasy are  
89 expected in comparison with sequence data and therefore statistical analyses  
90 developed for DNA sequence data may prove unreliable.

91

92 In the present work we have simulated the evolution of cpSSRs under constant  
93 population size and under population expansion to test the usefulness of these markers  
94 for the study of demographic expansions. These theoretical results were then  
95 compared with empirical results from the Canary Island pine (*Pinus canariensis*). The  
96 presence of *P. canariensis* on each of the five volcanic islands on which it occurs  
97 must be through colonisation after the emergence of each island, followed by  
98 population expansion.

99

100 **MATERIALS AND METHODS**

101

102 ***Simulations***

103

104 Demographic histories of population expansions (recent and old) and stable  
 105 population size were modelled with coalescent simulations to obtain theoretical  
 106 expectations of the behaviour of cpSSRs. The coalescent simulation (described in  
 107 Navascués & Emerson, 2005) consists of the generation of a genealogy for a sample  
 108 of individuals under a particular demographic history followed by the distribution of  
 109 mutations randomly onto those lineages. For the population expansions the  
 110 demographic history was modelled with a logistic equation setting the initial  
 111 population size ( $N_0$ ) as one individual (coloniser) at the time of expansion ( $\tau$ , in  
 112 mutational units). Microsatellite evolution was simulated following a symmetrical  
 113 single-step mutation model where mutation rates were either heterogeneous (two-rates  
 114 model) or uniform (one-rate model) across loci. Heterogeneous mutation rates can be  
 115 considered a more realistic scenario taking into account the differences in  
 116 polymorphism among cpSSR loci (see, for example, Gómez *et al.*, 2003). As well as  
 117 being more realistic, heterogeneous mutation rates will also produce higher levels of  
 118 homoplasy by concentrating the mutations onto particular loci, thus providing a more  
 119 rigorous assessment of the demographic utility of cpSSRs. The three different  
 120 demographic histories and the two mutation models gave a combination of six  
 121 different cases considered (table 1). Simulations were performed for a sample size of  
 122 24 individuals and six cpSSR loci. For each case, 1000 replicates were run and their  
 123 output (genetic state of a sample of individuals in the present generation) was  
 124 analysed as described in section (c). For each simulated case the level of homoplasy  
 125 was quantified as the probability that two haplotypes identical in state are not identical  
 126 by descent (homoplasy index, Estoup *et al.*, 2002).

127

128 ***Plant Material and Molecular Markers***

129

130 Empirical data was obtained from two previous studies of *P. canariensis* (Gómez *et al.*  
 131 *et al.*, 2003; Vaxevanidou *et al.*, 2005). Additionally, three populations from Tenerife  
 132 (nine, 12 and 13 in table 2 and figure 2) were also genotyped for the present analysis  
 133 and the compatibility of the data was assured by repeated genotyping of four  
 134 haplotypes from the previous studies. All individuals were genotyped for six cpSSR  
 135 loci: Pt15169, Pt30204, Pt71936, Pt87268, Pt26081 and Pt36480 (Vendramin *et al.*,  
 136 1996).

137

138 ***Data Analysis***

139

140 In order to use Arlequin 2.0 (Schneider *et al.*, 1999) for the analyses, microsatellite  
 141 data was binary coded: the number of repeats were coded with “1” and shorter alleles  
 142 were coded filling the difference in repeats with “0” (Pereira *et al.*, 2002). Analyses  
 143 for the empirical samples were carried out at two levels: (1) sample sites as the unit of  
 144 analysis, (2) islands as the unit of analysis with sample sites within an island pooled  
 145 together.

146

147 A general description of diversity indices and population structure found within *P.*  
 148 *canariensis* using cpSSRs is presented in Gómez *et al.* (2003); thus here we focus on

149 the assessment of demographic history, using two different but complementary  
 150 approaches. Firstly, we performed the  $F_S$  neutrality test for population expansion (Fu,  
 151 1997). This test is based on different expectations for the number of haplotypes when  
 152 comparing a stationary with an expansion demography. The  $F_S$  statistic takes a large  
 153 negative value within a population affected by expansion due to an excess of rare  
 154 haplotypes (recent mutations). Significance of the test was calculated with 10 000 data  
 155 bootstraps (Schneider *et al.*, 1999). An  $F_S$  statistic with  $p(F_S) < 0.02$  ( $\alpha = 0.05$ , due to  
 156 a particular behaviour of this statistic, Fu, 1997) was considered evidence of  
 157 population expansion.

158  
 159 The second analysis consists of the estimation of the demographic model of Rogers &  
 160 Harpending (1992) described with the parameters:  $\tau = 2\mu t$ ,  $\theta_0 = 2\mu N_0$  and  $\theta_1 = 2\mu N_1$   
 161 (where  $\mu$  is the mutation rate,  $t$  is the number of generations since expansion and  $N_0$   
 162 and  $N_1$  are the population sizes before and after expansion). Parameters are estimated  
 163 from the distribution of pairwise differences (difference in number of repeats)  
 164 between individuals within a sample. Although, in our case, the pairwise differences  
 165 calculated cannot be strictly called mismatches, we will refer to their distribution as a  
 166 mismatch distribution as it is the most common term used throughout the literature  
 167 (Harpending *et al.*, 1993). This distribution is affected by the demography of the  
 168 sample; sudden growth produces unimodal distributions while within stationary  
 169 populations distributions are ragged and multimodal (Slatkin & Hudson, 1991). An  
 170 algorithm, which minimizes the sum of squared differences (SSD) between model and  
 171 data, estimates the combination of parameters with the best fit to the empirical data  
 172 (Schneider & Excoffier, 1999). The strength of the estimated model is then evaluated  
 173 from the SSD distribution which is obtained from 10 000 data bootstraps (1000 for the  
 174 simulation output), making  $p(\text{SSD})$  the proportion of bootstraps with the SSD larger  
 175 than the original (Schneider & Excoffier, 1999). A significant SSD value,  $p(\text{SSD}) <$   
 176  $0.05$ , implies the rejection of the estimated demographic model. The confidence  
 177 interval (95% CI) for the estimated parameter  $\hat{\tau}$  is also calculated from the bootstrap  
 178 process (Schneider & Excoffier, 1999). Confidence intervals for parameters related to  
 179 the magnitude of expansion ( $\hat{\theta}_0$  and  $\hat{\theta}_1$ ) will not be discussed as they are usually too  
 180 wide and are of less interest for the interpretation of the results (Excoffier &  
 181 Schneider, 1999). Dating the population expansions was done using the parameter  $\hat{\tau}$   
 182 and its relationship with time and mutation rate:  $\tau = 2l\mu t$  (where  $l$  is the number of  
 183 cpSSR loci and  $\mu$  is the mutation rate per locus).

## 184 185 **RESULTS AND DISCUSSION**

### 186 187 ***Simulations***

188 The results from the simulations are summarised in table 1. In the two analyses  
 189 performed, cpSSR polymorphism was sensitive to population growth; however the  
 190 results were not as precise as would be desirable.

### 191 192 ***F<sub>S</sub> Neutrality Test***

193 In the cases of uniform mutation rate across loci the performance of the  $F_S$  test to  
 194 detect population expansion was acceptable. Type II error for the  $F_S$  test (no evidence  
 195 of population expansion in cases 1 and 2) was very low, and type I error (rejection of  
 196 stationary population size in case 3) was low (11% of the replicates of case 3),  
 197 although greater than expected at the given confidence level (expected 5% for  
 198  $\alpha=0.05$ ).

199

200 In the cases evolving under the two-rate model (cases 4–6), the power of the  $F_S$  test  
201 decreased dramatically, and this was accompanied by an increase in homoplasy.  
202 Detection of recent expansions was especially affected and the reason for this relates  
203 to the estimates of genetic distance and the number of haplotypes used in the test.  
204 First, the test uses the average genetic distance among individuals to calculate the  
205 expected number of haplotypes under a stationary demography scenario. The effect of  
206 homoplasy in this calculation is proportional to the time of expansion, with an average  
207 reduction of 19% in the distance estimates of recent expansions (case 4) and 41% in  
208 the older expansions (case 5). The expected number of haplotypes is then compared to  
209 the observed number of haplotypes. While the effect of homoplasy in genetic distance  
210 estimates was proportional to the time of expansion, homoplasy decreases the  
211 detectable number of haplotypes by approximately 40% both in the recent and older  
212 expansions. It seems that the power of the test varies with the time of expansions  
213 because the error in the estimates of genetic distances and number of haplotypes is  
214 more unbalanced for recent expansions.

215

#### 216 *Demographic Model Estimation*

217 For cpSSRs evolving under the one-rate mutation model, estimates of the time of  
218 expansion were fairly accurate, although older expansion times were slightly  
219 underestimated. The average estimated time of expansion ( $\hat{\tau}$ ) for the recent  
220 expansions (case 1,  $\tau = 1.0$ ) was 1.1 and the true value was always within the 95% CI,  
221 while for older expansions (case 2,  $\tau = 3.0$ ) average  $\hat{\tau}$  was 2.5 and the true value falls  
222 outside the 95% CI in 15% of the replicates. In the simulations using the two-rate  
223 mutation model the estimates for recent expansion (case 5,  $\tau = 1.0$ ) were accurate,  
224 with average  $\hat{\tau} = 1.0$  and the true value fell outside the 95% CI in only 2% of the  
225 replicates. However in older expansions (case 6,  $\tau = 3.0$ ) the expansion time was  
226 largely underestimated for the two-rate mutation model with the average value of  $\hat{\tau}$   
227 being 1.8 and the true value falling outside the 95% CI in 77% of the replicates.  
228 Although these results appear discouraging it is important to note that the relative  
229 times of expansion are still discernable, and that it may be possible to develop new  
230 statistical analyses to improve the estimates as has been done for heterogeneous  
231 mutation rates within sequence data (Schneider & Excoffier, 1999).

232

#### 233 *The Empirical Case: Pinus canariensis*

234 The results for the detection of population expansions in the *P. canariensis* samples  
235 are reported in table 2. For the estimation of the demographic model the algorithm  
236 was unable to find a combination of parameters with a minimum SSD in three  
237 samples (Tamadaba, Chinyero and El Hierro). This inability of the algorithm to  
238 converge sometimes has been observed in previous works (e.g. Stamatis *et al.*, 2004)  
239 and in our simulations. A simple solution is to obtain the estimation from a reduced  
240 sample obtained by randomly removing one individual. This reduction of the sample  
241 size changes the shape of the mismatch distribution slightly enough for the algorithm  
242 to converge while still maintaining a very similar shape to the mismatch distribution  
243 from the original data set. The mismatch distributions from the reduced samples were  
244 used to produce the parameter estimations presented in italics in table 2.

245

246 The demographic expansion model estimated for different sampling sites (including  
247 the grouping of sampling sites at the island level) was, in general, fairly robust  
248 [ $p(\text{SSD}) \gg 0.05$ ] and mismatch distributions were clearly unimodal (figure 2;

249 opposite to the ragged distribution expected with a stable population). The results of  
250 the  $F_S$  test yielded evidence of population expansion for nearly half of the samples. It  
251 is interesting to note that the samples for which the  $F_S$  test could not reject a stable  
252 population scenario [ $p(F_S) > 0.02$ ] were the ones with the lowest  $\hat{\tau}$  values. In the light  
253 of our simulation results it is expected that the  $F_S$  test will have lower power to detect  
254 very recent population expansions, especially under the more realistic scenario of  
255 heterogeneous mutation rates across loci. Thus we could consider that most of the *P.*  
256 *canariensis* populations are likely to have been subject to demographic growth and  
257 the lack of statistical evidence is due to the low power of the  $F_S$  test for the most  
258 recent expansions.

#### 259 *Island Level: Colonisation*

260 Compared to continental areas, oceanic island populations are typically established by  
261 only one or a few individual founders that successfully reproduce, leading to  
262 demographic expansions. Whether the population expansions detected for *P.*  
263 *canariensis* at the island level reflect the initial colonisation of the islands or  
264 subsequent demographic events is difficult to know. However, times of expansion in  
265 relation to the geological history of the archipelago can supply the necessary clues to  
266 discern between both possibilities.

267  
268 Potential maximum times for expansion are bound to the emergence times of the  
269 islands. The maximum subaerial geological age of El Hierro, the youngest island, is  
270 approximately one million years (Carracedo & Day, 2002). If we consider that the  
271 time of the population expansion in El Hierro is  $\hat{\tau} = 1.291$  and the relationship  $\tau =$   
272  $2\mu t$  we obtain a mutation rate estimate of  $1.076 \times 10^{-5}$  per locus per generation  
273 (considering generation time to be 100 years as in Provan *et al.*, 1999). Using this  
274 mutation rate estimate we calculated the maximum age of population expansion for  
275 each sample, reported in table 2.

276  
277  
278 In order to establish a minimum time of expansion we have analysed mtDNA COII  
279 sequence data for *Brachyderes rugatus* from Emerson *et al.* (2000 and unpublished  
280 data). Because the niche of this species is the pine tree, its demographic expansions  
281 must have occurred either during or after the establishment of the pine forest on each  
282 island. Population expansions have been detected (significant  $F_S$  test) for the islands  
283 of La Palma and Tenerife (138 and 182 individuals respectively, sampled throughout  
284 the islands). The times of expansion for *B. rugatus* were estimated from the mismatch  
285 distributions to be approximately 0.72 million years ago (mya) for Tenerife and 1.11  
286 mya for La Palma (considering divergence rates to be between 2% and 2.3% per  
287 million years, Brower, 1994; DeSalle *et al.*, 1987). These dates strengthen the age  
288 estimates for the expansion of the pine forest obtained with the geological age  
289 calibration.

290  
291 These age estimates suggest expansions of the pine tree increasing in age from West  
292 to East, and coinciding broadly with the colonisation ages estimated for *B. rugatus*  
293 (Emerson *et al.*, 2000), as shown in table 2. We interpret the expansions at the island  
294 level as a result of the colonisation process and linked to the volcanic history of the  
295 archipelago. The creation of new emerged landmass by recent (up to 2 mya) volcanic  
296 activity in the younger islands (La Palma and El Hierro) opened new territories for *P.*  
297 *canariensis* to colonise. Note that the age of Tenerife presented in figure 2 refers to its  
298 older massifs which are the remains of two or three smaller precursor islands.

299 However, the majority of the landmass of Tenerife was mainly formed by the activity  
300 of Las Cañadas volcano starting around 2 million years ago (Ancochea *et al.*, 1990)  
301 and it is this event which would appear to be causally related to the pine forest  
302 expansion. On the island of Gran Canaria, an episode of heavy volcanic activity  
303 (Roque Nublo volcano, Pérez-Torrado *et al.*, 1995) is believed to have destroyed  
304 almost all terrestrial ecosystems within the island, with perhaps the exclusion of some  
305 coastal regions, between 5.5 and 3 million years ago (Marrero & Francisco-Ortega,  
306 2001), and this hypothesis has gained recent support from a meta-analysis by Emerson  
307 (2003). The expansion of the pine forest in Gran Canaria after that event can be  
308 explained either by colonisation of *P. canariensis* to the island or by a bottleneck if a  
309 small pocket of pine forest survived through the Roque Nublo eruptive period.  
310

#### 311 *Sample Level: Metapopulation Dynamics*

312 The islands of the Canary archipelago have a geological history marked by recent  
313 dramatic volcanic activity and giant landslides (Carracedo & Day, 2002). These  
314 destructive events would have produced local elimination of pine forest, as has been  
315 recorded for historical volcanic eruptions (del Arco Aguilar *et al.*, 1992; Pérez de Paz  
316 *et al.*, 1994). Also, the Canary Island pine is renowned by its capacity for colonising  
317 lava flows (del Arco Aguilar *et al.*, 1992; Pérez de Paz *et al.*, 1994), which suggest  
318 that a metapopulation dynamic occurs within the pine forest. One of the genetic  
319 signals expected in the local recolonisations after volcanic disturbances are those of  
320 the demographic expansions, as it has been shown in other organisms subject to  
321 similar metapopulation dynamics in other volcanic archipelagos (Beheregaray *et al.*,  
322 2003; Vandergast *et al.*, 2004).  
323

324 It seems very likely that local expansions detected for *P. canariensis* are the product  
325 of metapopulation dynamics. When we consider different samples within the same  
326 island (in Tenerife and Gran Canaria) we observe that the expansion of pine forest at  
327 some areas is younger than the main demographic expansion affecting the island. We  
328 hypothesize that the apparently more recent expansions may be areas recolonised after  
329 geological disturbance. The role of volcanism and giant landslides in the reduction of  
330 genetic diversity has also been proposed to explain the pattern of diversification of  
331 *Brachyderes rugatus* in La Palma, El Hierro, Tenerife and Gran Canaria (Emerson *et*  
332 *al.*, 2000).  
333

## 334 **CONCLUSIONS**

335  
336 This study demonstrates the utility of cpSSRs for the detection of demographic  
337 expansions and the estimation of their relative ages. The application of population  
338 genetic demographic methodology to cpSSR data for *P. canariensis* populations  
339 revealed new insights into the population history of this species. The volcanic activity  
340 of the archipelago appears to be a disturbance agent in the pine forest ecosystem,  
341 conditioning the areas available for the pine tree. Future studies of mitochondrial  
342 DNA data may further complement data from cpSSRs to elucidate the colonisation  
343 and population dynamic history of *P. canariensis* on the Canary Islands. A mtDNA  
344 phylogeographic analysis would reflect the historical seed movements of *P.*  
345 *canariensis*, which are limited relative to pollen and may contain more fine scale  
346 phylogeographic information. Additionally, a sampling design including historical  
347 and isotope-dated lava flows within the pine forest may provide a good test for the  
348 hypothesis of a metapopulation dynamic.

349  
350 Our analyses have revealed homoplasy as a problem for the analyses (mainly in the  
351 detection of younger expansions) because it reduces the power of the  $F_S$  test and  
352 accuracy of absolute expansion time estimates. The development of statistics taking  
353 into account the effects of homoplasy would further improve the usefulness of  
354 cpSSRs as well as other linked microsatellite markers such as Y-chromosome  
355 microsatellites for demographic studies.

#### 356 **Acknowledgements**

357 MN scholarship was funded by the University of East Anglia. We thank the Cabildo  
358 Insular de Tenerife for collecting permits.

#### 360 **REFERENCES**

- 361  
362  
363 Ancochea E, Fúster JM, Ibarrola E, *et al.* (1990) Volcanic evolution of the island of  
364 Tenerife (Canary Islands) in the light of new K-Ar data. *Journal of*  
365 *Volcanology and Geothermal Research* **44**, 231-249.
- 366 del Arco Aguilar MJ, Pérez de Paz PL, Rodríguez Delgado O, Salas M, Wildpret W  
367 (1992) *Atlas Cartográfico de los Pinares Canarios II: Tenerife* Gobierno de  
368 Canarias, Consejería de Política Territorial, Santa Cruz de Tenerife.
- 369 Aris-Brosou S, Excoffier L (1996) The impact of population expansion and mutation  
370 rate heterogeneity on DNA sequence polymorphism. *Molecular Biology and*  
371 *Evolution* **13**, 494-504.
- 372 Beheregaray LB, Ciofi C, Geist D, *et al.* (2003) Genes record a prehistoric volcano  
373 eruption in the Galápagos. *Science* **302**, 75.
- 374 Bertorelle G, Slatkin M (1995) The number of segregating sites in expanding human  
375 populations, with implications for estimates of demographic parameters.  
376 *Molecular Biology and Evolution* **12**, 887-892.
- 377 Brower A (1994) Rapid morphological radiation and convergence among races of the  
378 butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA  
379 evolution. *Proceedings of the National Academy of Sciences of the United*  
380 *States of America* **91**, 6491-6495.
- 381 Carracedo JC, Day S (2002) *Canary Islands* Terra Publishing, Harpenden.
- 382 DeSalle R, Freedman T, Prager EM, Wilson AC (1987) Tempo and mode of sequence  
383 evolution in mitochondrial DNA of Hawaiian *Drosophila*. *Journal of*  
384 *Molecular Evolution* **26**, 157-164.
- 385 Emerson BC (2003) Genes, geology and biodiversity: faunal and floral diversity on  
386 the island of Gran Canaria. *Animal Biodiversity and Conservation* **26**, 9-20.
- 387 Emerson BC, Oromí P, Hewitt GM (2000) Colonization and diversification of the  
388 species *Brachyderes rugatus* (Coleoptera) on the Canary Islands: evidence  
389 from mitochondrial DNA COII gene sequences. *Evolution* **54**, 911-923.
- 390 Emerson BC, Paradis E, Thebaud C (2001) Revealing the demographic histories of  
391 species using DNA sequences. *Trends in Ecology & Evolution* **16**, 707-716.
- 392 Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at  
393 microsatellite loci and their consequences for population genetics analysis.  
394 *Molecular Ecology* **11**, 1591-1604.
- 395 Excoffier L, Schneider S (1999) Why hunter-gatherer populations do not show signs  
396 of Pleistocene demographic expansions. *Proceedings of the National Academy*  
397 *of Sciences of the United States of America* **96**, 10597-10602.

- 398 Fu YX (1997) Statistical tests of neutrality of mutations against population growth,  
399 hitchhiking and background selection. *Genetics* **147**, 915-925.
- 400 Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics* **133**, 693-  
401 709.
- 402 Gómez A, González-Martínez SC, Collada C, Gil L, Climent J (2003) Complex  
403 population genetic structure in an endemic Canary Island pine using  
404 chloroplast microsatellite markers. *Theoretical and Applied Genetics* **107**,  
405 1123-1131.
- 406 Harpending HC, Sherry ST, Rogers AR, Stoneking M (1993) The genetic structure of  
407 ancient human populations. *Current Anthropology* **34**, 483-496.
- 408 King JP, Kimmel M, Chakraborty R (2000) A power analysis of microsatellite-based  
409 statistics for inferring past population growth. *Molecular Biology and*  
410 *Evolution* **17**, 1859-1868.
- 411 Marrero A, Francisco-Ortega J (2001) Evolución en islas: la metáfora espacio-tiempo-  
412 forma. In: *Naturaleza de las Islas Canarias: Ecología y Conservación* (eds.  
413 Fernández-Palacios JM, Martín Esquivel JL), pp. 133-140. Turquesa, Santa  
414 Cruz de Tenerife.
- 415 Navascués M, Emerson BC (2005) Chloroplast microsatellites: measures of genetic  
416 diversity and the effect of homoplasmy. *Molecular Ecology* **14**, 1333-1341.
- 417 Pereira L, Prata MJ, Amorim A (2002) Mismatch distribution analysis of Y-STR  
418 haplotypes as a tool for the evaluation of identity-by-state proportions and  
419 significance of matches - the European picture. *Forensic Science International*  
420 **130**, 147-155.
- 421 Pérez de Paz PL, del Arco MJ, Rodríguez O, *et al.* (1994) *Atlas Cartográfico de los*  
422 *Pinares Canarios III: La Palma* Gobierno de Canarias, Consejería de Política  
423 Territorial, Santa Cruz de Tenerife.
- 424 Pérez-Torrado FJ, Carracedo JC, Mangas J (1995) Geochronology and stratigraphy of  
425 the Roque Nublo Cycle, Gran Canaria, Canary Islands. *Journal of the*  
426 *Geological Society* **152**, 807-818.
- 427 Provan J, Powell W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools  
428 for studies in plant ecology and evolution. *Trends in Ecology & Evolution* **16**,  
429 142-147.
- 430 Provan J, Soranzo N, Wilson NJ, Goldstein DB, Powell W (1999) A low mutation  
431 rate for chloroplast microsatellites. *Genetics* **153**, 943-947.
- 432 Reich DE, Goldstein DB (1998) Genetic evidence for a Paleolithic human population  
433 expansion in Africa. *Proceedings of the National Academy of Sciences of the*  
434 *United States of America* **95**, 8119-8123.
- 435 Rogers AR (1995) Genetic evidence for a Pleistocene population explosion. *Evolution*  
436 **49**, 608-615.
- 437 Rogers AR, Fraley AE, Bamshad MJ, Watkins WS, Jorde LB (1996) Mitochondrial  
438 mismatch analysis is insensitive to the mutational process. *Molecular Biology*  
439 *and Evolution* **13**, 895-902.
- 440 Rogers AR, Harpending H (1992) Population growth makes waves in the distribution  
441 of pairwise genetic differences. *Molecular Biology and Evolution* **9**, 552-569.
- 442 Schneider S, Excoffier L (1999) Estimation of past demographic parameters from the  
443 distribution of pairwise differences when the mutation rates vary among sites:  
444 application to human mitochondrial DNA. *Genetics* **152**, 1079-1089.
- 445 Schneider S, Roessli D, Excoffier L (1999) ARLEQUIN: A software for population  
446 genetics data analysis. University of Geneva, Geneva.

- 447 Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA  
448 sequences in stable and exponentially growing populations. *Genetics* **129**, 555-  
449 562.
- 450 Stamatis C, Triantafyllidis A, Moutou KA, Mamuris Z (2004) Mitochondrial DNA  
451 variation in Northeast Atlantic and Mediterranean populations of Norway  
452 lobster, *Nephrops norvegicus*. *Molecular Ecology* **13**, 1377-1390.
- 453 Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by  
454 DNA polymorphism. *Genetics* **123**, 585-595.
- 455 Vandergast AG, Gillespie RG, Roderick GK (2004) Influence of volcanic activity on  
456 the population genetic structure of Hawaiian *Tetragnatha* spiders:  
457 fragmentation, rapid population growth and the potential for accelerated  
458 evolution. *Molecular Ecology* **13**, 1729-1743.
- 459 Vaxevanidou Z, González-Martínez S, Climent J, Gil L (2005) Tree populations  
460 bordering on extinction: A case study in the endemic Canary Island pine.  
461 *Biological Conservation* **in press**.
- 462 Vendramin GG, Lelli L, Rossi P, Morgante M (1996) A set of primers for the  
463 amplification of 20 chloroplast microsatellites in Pinaceae. *Molecular Ecology*  
464 **5**, 595-598.  
465  
466

467 Figure 1: Coalescent process under two contrasting scenarios: constant population  
468 size and sudden population expansion. For each case, the demographic history and, in  
469 the same timescale measured in mutational units (1 mutational unit =  $1/2\mu$   
470 generations) the simulated genealogy of a random sample of genes is represented,  
471 with stars representing mutational events. Below, chloroplast microsatellite mismatch  
472 distribution and result of the  $F_S$  test and demographic parameters estimates for those  
473 simulated samples are shown.

474  
475 Figure 2: Chloroplast microsatellite mismatch distributions for the islands and map of  
476 the Western Canary Islands. Maximum subaerial age of the islands (Carracedo &  
477 Day, 2002) are shown in parenthesis (in million of years). Sampling localities are  
478 marked with numbers, corresponding to those shown in table 2.

479 Table 1: Population expansion signal on the  $F_S$  test and mismatch distribution analysis and homoplasy level in the six simulated cases.

Case	Expansion time, $\tau$	Mutation rate, $\mu$		Proportion of non-significant $F_S$ test	Proportion of significant SSD	Homoplasy index, $P$
		loci 1-2	loci 3-6			
1	1 (recent)	$5.5 \times 10^{-3}$		0.038	0.052	0.049
2	3 (old)	$5.5 \times 10^{-5}$		0.002	0.051	0.297
3	no expansion	$5.5 \times 10^{-5}$		0.893	0.144	0.065
4	1 (recent)	$1.65 \times 10^{-4}$	$10^{-7}$	0.553	0.095	0.122
5	3 (old)	$1.65 \times 10^{-4}$	$10^{-7}$	0.240	0.060	0.606
6	no expansion	$1.65 \times 10^{-4}$	$10^{-7}$	0.931	0.080	0.263

480

481

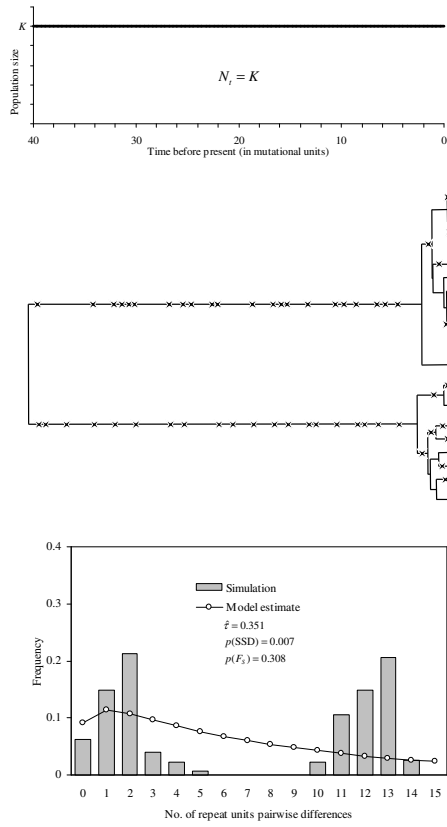
482 Table 2: Results for the  $F_S$  neutrality test for population expansion (Fu, 1997) and population expansion parameters  $\hat{t}$  following Schneider &  
 483 Excoffier (1999). Estimates are presented in italics when the algorithm did not converge (see Results and discussion section for details). The  
 484 time of expansion expressed in million of years before present (mya) is calculated using mutation rates in the range  $1.076 \times 10^{-5}$  per generation  
 485 per locus. The results for the islands (pooling samples from the same island) are presented in bold. Garabato is a monomorphic population and  
 486 tests could not be performed. For comparison, time for the colonisation of *Brachydes rugatus* are also presented from Emerson *et al.* (2000).

Population	$N$	$F_S$	Fu (1997) $p(F_S)$	$\hat{t}$ (95% CI)	Schneider & Excoffier (1999) $t$ (mya)	SSD	$p(SSD)$	Emerson <i>et al.</i> (2000) $B. rugatus$ (MYA)
<b>Gran Canaria</b>	<b>145</b>	<b>-26.260</b>	<b>* &lt; 0.001</b>	<b>2.544 (1.372-5.010)</b>	<b>1.970</b>	<b>0.002</b>	<b>0.569</b>	> 2.56
1 Arguineguín	30	-21.890	* < 0.001	3.703 (2.138-5.636)	2.868	0.003	0.469	
2 Galdar	19	-4.076	* 0.017	2.722 (1.128-3.892)	2.108	0.004	0.554	
3 Mogán	24	-0.849	0.275	0.969 (0.000-1.558)	0.750	0.020	0.077	
4 Tamadaba	24 (23)	-1.304	0.142	0.871 (0.000-1.489)	0.675	0.014	0.138	
5 Tirajana	24	-5.052	* 0.014	1.743 (0.413-5.224)	1.350	0.008	0.399	
6 Tirma	24	-6.845	* 0.001	2.438 (0.779-3.342)	1.888	0.002	0.783	
<b>Tenerife</b>	<b>280</b>	<b>-26.710</b>	<b>* &lt; 0.001</b>	<b>2.374 (1.330-3.144)</b>	<b>1.839</b>	<b>0.000</b>	<b>0.910</b>	1.89-2.56
7 Anaga	24	-0.192	0.185	3.000 (0.523-3.000)	2.323	0.010	0.062	
8 Arico	24	-8.983	* < 0.001	2.314 (0.675-3.185)	1.792	0.003	0.542	
9 Chinyero	50 (49)	-15.290	* < 0.001	2.294 (0.978-2.900)	1.777	0.001	0.636	
10 La Esperanza	24	-3.040	0.022	2.722 (0.483-6.413)	2.108	0.015	0.567	
11 La Guancha	24	-3.545	0.024	2.036 (0.501-2.875)	1.577	0.003	0.623	
12 Güímar	47	-22.280	* < 0.001	3.280 (1.795-4.143)	2.540	0.001	0.702	
13 Ifonche	39	-8.615	* < 0.001	2.482 (1.035-3.172)	1.922	0.007	0.130	
14 Oratava	24	-6.456	* 0.002	2.730 (0.993-6.926)	2.114	0.010	0.285	
15 Vilaflor	24	-0.165	0.437	1.081 (0.000-1.766)	0.837	0.009	0.192	
<b>La Gomera</b>	<b>36</b>	<b>-0.424</b>	<b>0.427</b>	<b>1.558 (0.380-3.379)</b>	<b>1.207</b>	<b>0.008</b>	<b>0.431</b>	
16 Garabato	12	-	-	-	-	-	-	
17 Imada	24	-0.879	0.294	1.307 (0.182-2.064)	1.012	0.002	0.711	
<b>La Palma</b>	<b>48</b>	<b>-2.826</b>	<b>0.072</b>	<b>1.244 (0.311-1.726)</b>	<b>0.963</b>	<b>0.004</b>	<b>0.285</b>	1.58-2.00
18 Fuencaliente	24	-1.063	0.224	1.289 (0.048-2.002)	0.998	0.008	0.305	

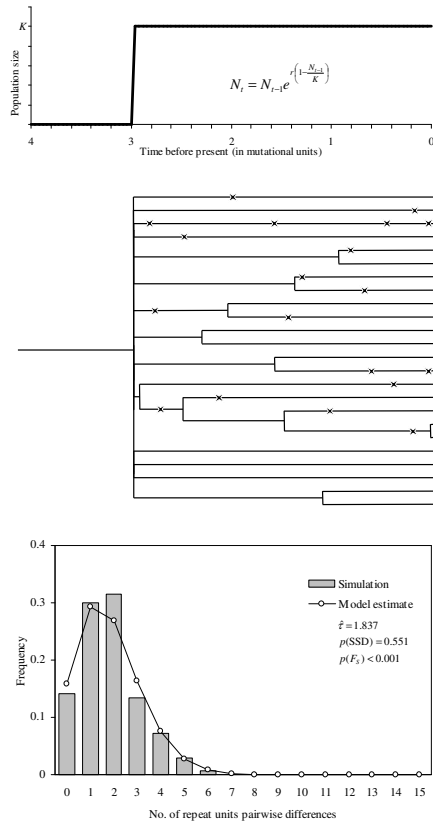
19	Garafía	24	-1.279	0.184	1.206 (0.053-1.904)	0.934	0.002	0.720
<b>20</b>	<b>El Hierro</b>	<b>24 (23)</b>	<b>-3.513</b>	<b>* 0.008</b>	<b>1.291 (0.040-2.035)</b>	<b>1.000</b>	<b>0.008</b>	<b>0.301</b>

487 \* Significant at  $\alpha = 0.05$  ( $p$ -value < 0.02)

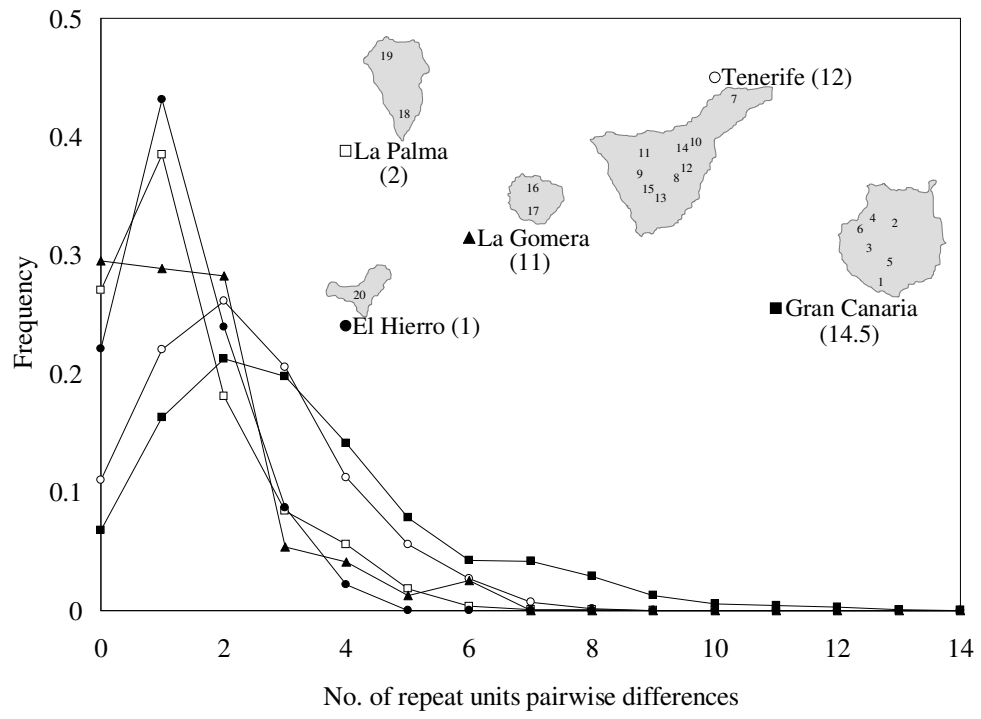
(a) Constant population size (simulation case 3)



(b) Population expansion,  $\tau = 3$  (simulation case 2)



488  
489



490