

Molecular insights into the colonization and chromosomal diversification of Madeiran house mice

D. W. FÖRSTER,* İ. GÜNDÜZ,† A. C. NUNES,‡ S. GABRIEL,*‡ M. G. RAMALHINHO,§ M. L. MATHIAS,‡ J. BRITTON-DAVIDIAN¶ and J. B. SEARLE*

*Department of Biology, University of York, PO Box 373, York YO10 5YW, UK, †Department of Biology, Faculty of Arts and Sciences, University of Ondokuz Mayıs, Samsun, Turkey, ‡Departamento de Biologia Animal, Centro de Biologia Ambiental, Faculdade de Ciências, Universidade de Lisboa, Edifício C2, 3º Piso, Campo Grande, 1746-016 Lisboa, Portugal, §Centro de Biologia Ambiental, Museu Bocage, Universidade de Lisboa, Rua da Escola Politécnica, 1268-102 Lisboa, Portugal, ¶CNRS Institut des Sciences de l'Evolution, Université Montpellier II, CC65, Place E. Bataillon, 34095 Montpellier Cedex 5, France

Abstract

The colonization history of Madeiran house mice was investigated by analysing the complete mitochondrial (mt) D-loop sequences of 156 mice from the island of Madeira and mainland Portugal, extending on previous studies. The numbers of mtDNA haplotypes from Madeira and mainland Portugal were substantially increased (17 and 14 new haplotypes respectively), and phylogenetic analysis confirmed the previously reported link between the Madeiran archipelago and northern Europe. Sequence analysis revealed the presence of four mtDNA lineages in mainland Portugal, of which one was particularly common and widespread (termed the 'Portugal Main Clade'). There was no support for population bottlenecks during the formation of the six Robertsonian chromosome races on the island of Madeira, and D-loop sequence variation was not found to be structured according to karyotype. The colonization time of the Madeiran archipelago by *Mus musculus domesticus* was approached using two molecular dating methods (mismatch distribution and Bayesian skyline plot). Time estimates based on D-loop sequence variation at mainland sites (including previously published data from France and Turkey) were evaluated in the context of the zooarchaeological record of *M. m. domesticus*. A range of values for mutation rate (μ) and number of mouse generations per year was considered in these analyses because of the uncertainty surrounding these two parameters. The colonization of Portugal and Madeira by house mice is discussed in the context of the best-supported parameter values. In keeping with recent studies, our results suggest that mutation rate estimates based on interspecific divergence lead to gross overestimates concerning the timing of recent within-species events.

Keywords: house mouse, island colonization, mitochondrial DNA, *Mus musculus domesticus*, phylogeography

Received 19 December 2008; accepted 27 May 2009

Introduction

The western house mouse *Mus musculus domesticus* is a very widespread mammal, found in Western Europe, Africa, the Americas and Australasia. The subspecies

most probably became commensal (associated with humans) by 8500 BC in the Fertile Crescent of the Middle East (Cucchi *et al.* 2005), and much of its long-range dispersal from that point onwards will have been human mediated (e.g. with agricultural materials) (Pocock *et al.* 2005). In particular, this must have been the case for the oceanic islands that *M. m. domesticus* colonized. This includes our study site, the island of

Correspondence: Daniel W. Förster, Fax: 01904 432 860; E-mail: dwgfoerster@googlemail.com

Madeira (length: 60 km), which is located 600 km off the Atlantic coast of North Africa.

Interest in the house mice of Madeira was intensified after the discovery of several chromosome races on the island characterized by various combinations of Robertsonian (Rb) fusions (Britton-Davidian *et al.* 2000). Such chromosomal variability has been documented elsewhere in *M. m. domesticus* (see Piálek *et al.* 2005). However, the situation on Madeira is exceptional because six highly divergent Rb races inhabit this small geographical area. The fixed karyotypic differences between these races are of sufficient magnitude that they may promote reproductive isolation (Baker & Bickham 1986; King 1993; Searle 1993).

The phylogeographical study of Gündüz *et al.* (2001) yielded a strong link between Madeiran mice and northern Europe based on a comparison of mitochondrial (mt) D-loop sequences from Madeira with those of the literature (Prager *et al.* 1993, 1998; Nachman *et al.* 1994). This was surprising, given that the official human history of Madeira is strongly associated with Portugal, the nation that is historically attributed with its discovery in 1419 AD (Albuquerque & Vieira 1988). However, there is convincing evidence of earlier visitations to the islands by unknown seafarers (mid-14th century or before; reviewed by Albuquerque & Vieira 1988), suggesting that mice could have been introduced into the archipelago from a different source area than Portugal.

Although the study by Gündüz *et al.* (2001) demonstrated close mtDNA sequence similarity between house mice of Madeira and northern Europe, little information was available on mice from Portugal: two mice caught in Lisbon with the same mtDNA haplotype (Prager *et al.* 1993; Gündüz *et al.* 2001). Moreover, a recent allozyme analysis of the Madeiran Rb mice (Britton-Davidian *et al.* 2007) indicates a strong link to Portugal and less so to northern Europe. Thus, the first aim of this study was to reconsider Portugal as a possible source area; not only because of the obvious human association, but also because it is possible that mtDNA haplotypes observed in northern Europe were introduced first to Portugal and secondarily to Madeira. For this purpose, a detailed analysis of D-loop variation in house mice along the Atlantic coast of Portugal was carried out. The second aim was to establish a time frame for the colonization of Portugal and Madeira by *M. m. domesticus* using two molecular dating methods (Rogers & Harpending 1992; Drummond *et al.* 2005). Specifically, by comparing molecular date estimates for several house mouse populations with the recently reviewed zooarchaeological record of *M. m. domesticus* (Cucchi *et al.* 2005), the likely colonization time of Madeira by *M. m. domesticus* could be approached. Additionally, it was investigated if the chromosomal

diversification in Madeira has resulted in a race-specific distribution of mtDNA variation.

Methods

Samples

The collection localities of all specimens used in this study are given in Fig. 1 and Table 1. Localities in mainland Portugal were chosen based on their presence in 15th and 16th century maritime maps for the region, and represent communities with active ports during the Portuguese settlement of Madeira; the only exception is locality 10 (Grândola) that is 30 km inland. The Madeiran collection sites follow the scheme of Gündüz *et al.* (2001); site T has been added (race PSVI, after Piálek *et al.* 2005) for which no data were previously available. Portuguese specimens were collected in spring 2005. The new Madeiran specimens were collected during spring 2004.

Chromosome analysis of Madeiran and Portuguese specimens was carried out as in Nunes *et al.* (2005). Madeiran Rb mice had karyotypes as expected based on previous records for each trapping locality (Britton-Davidian *et al.* 2000). Two mice were typed for most Portuguese sites; the exceptions were sites 1, 4 and 10 (Fig. 1). For Figueira da Foz, there was only one mouse for typing, and in two sites, Caminha and Grândola, no live mice were available for chromosome preparations. Of 23 Portuguese mice karyotyped, one specimen from Porto had a reduced chromosome number ($2n = 39$); the remaining Portuguese mice all had the standard $2n = 40$ chromosome complement. The individual with $2n = 39$ was heterozygous for metacentric 3.8, a Rb fusion found in Madeira (Britton-Davidian *et al.* 2000) as well as Rb populations in Denmark, Spain and Italy (populations DKEA, EGAR and ICDE respectively, following the nomenclature described by Piálek *et al.* 2005).

DNA methods

Tail tips, spleens and/or toe clippings were stored in 100% ethanol and maintained at 4 °C. DNA was extracted using either phenol/chloroform (Sambrook *et al.* 1989) or Qiagen DNeasy Tissue Kit, following the manufacturer's instructions.

The specimens from Madeira were amplified and sequenced in two fragments and in both directions as previously described elsewhere (Gündüz *et al.* 2000). The whole D-loop (879 bp) plus the Thr-tRNA and Pro-tRNA genes was sequenced giving, in total, 1013 bp between positions 15283 and 16295 after Bibb *et al.* (1981). For the specimens from Portugal, another



Fig. 1 Sampling localities for 35 house mouse populations; (a) map overview, (b) localities in mainland Portugal (numbers), (c) localities in Madeira (letters) following the scheme of Gündüz *et al.* (2001). Locality names, numbers of samples and observed mtDNA (D-loop) haplotypes are presented in Table 1. Black boxes (localities) in (b) had the most common Portuguese mtDNA haplotype PORTUGAL.1 present at least once. Black boxes in (c) indicate Madeiran localities that had at least one of six mtDNA haplotypes present that are also found in northern Europe (see text).

Table 1 Collection details and mtDNA haplotypes of *Mus musculus domesticus* specimens from Portugal, Madeira and Porto Santo

Locality code	Locality	N	Race*	2n	Haplotype (frequency)†
Portugal					
1	Caminha	1	?	?	PORTUGAL.1 (1)
2	Viana de Castelo	5	40ST	40	PORTUGAL.1 (5)
3	Porto	4	40ST	40†	PORTUGAL.1 (1), PORTUGAL.10 (3)
4	Figueira da Foz	1	40ST	40	PORTUGAL.1 (1)
5	Peniche	5	40ST	40	PORTUGAL.5 (5)
6	Lisbon	9	40ST	40	PORTUGAL.1 (1), PORTUGAL.8 (2), PORTUGAL.15 (6)
7	Vila Franca de Xira	6	40ST	40	PORTUGAL.1 (1), PORTUGAL.6 (3), PORTUGAL.7 (2)
8	Setubal	8	40ST	40	PORTUGAL.1 (2), PORTUGAL.4 (6)
9	Sines	7	40ST	40	PORTUGAL.1 (1), PORTUGAL.2 (3), PORTUGAL.14 (1), PORTUGAL.13 (2)
10	Grândola	10	?	?	PORTUGAL.2 (8), PORTUGAL.3 (2)
11	Vila Nova Mil Fontes	4	40ST	40	PORTUGAL.12 (4)
12	Lagos	5	40ST	40	PORTUGAL.11 (5)
13	Faro	3	40ST	40	PORTUGAL.9 (3)
14	Tavira	8	40ST	40	PORTUGAL.1 (8)
	Total	76			
Madeira and Porto Santo‡					
A	Funchal	4	PSAN	22§	MADE.1 (4)
B	Funchal	6	PSAN	22	MADE.1 (1), MADE.14 (2), MADE.15 (2), MADE.27 (1)
C	Funchal	5	PSAN	22	MADE.1 (3), MADE.13 (2)
D	Camacha	5	PSAN	22	MADE.1 (1), MADE.6 (4)
E	Santo da Serra	8	PSAN	22	MADE.14 (2), MADE.17 (1), MADE.29 (3), MADE.30 (1), MADE.31 (1)
F	Porta da Cruz	8	PSAN	22	MADE.1 (1), MADE.2 (1), MADE.14 (4), MADE.15 (1), MADE.26 (1)
G	Santana	8	PSAN	22	MADE.1 (1), MADE.2 (5), MADE.15 (2)
H	Ribeira Sao Jorge	7	PSAN	22	MADE.1 (2), MADE.2 (1), MADE.10 (2), MADE.28 (2)
I	Arco de Sao Jorge	8	PSAN	24	MADE.1 (4), MADE.2 (3), MADE.10 (1)
J	Ponta Delgada	10	PPOD	28	MADE.9 (3), MADE.11 (3), MADE.12 (1), MADE.19 (2), MADE.32 (1)
K	Seixal	4	PEDC	24	MADE.4 (1), MADE.8 (1), MADE.18 (1), MADE.20 (1)
L	Achadas da Cruz	1	PADC	24	MADE.5 (1)
M	Sra. do Amparo	4	PADC	24	MADE.8 (1), MADE.12 (1), MADE.23 (2)
N	Paúl do Mar	2	PADC	24	MADE.1 (1), MADE.7 (1)
O	Prazeres	4	PADC	24	MADE.2 (2), MADE.24 (1), MADE.25 (1)
P	Estreito da Calheta	7	PEDC	24	MADE.2 (2), MADE.8 (2), MADE.21 (1), MADE.22 (2)
Q	Lugar de Baixo	6	PLDB	24	MADE.1 (4), MADE.2 (2)
R	Campanario	9	PSAN	22	MADE.1 (2), MADE.2 (1), MADE.3 (1), MADE.25 (5)
S	Funchal Port	4	PSAN	22§	MADE.1 (1), MADE.10 (1), PSANTO.2 (2)
T	São Vicente	4	PSVI	24	MADE.16 (2), MADE.17 (1), MADE.18 (1)
	Porto Santo¶	8	40ST	40	PSANTO.1 (2), PSANTO.2 (4), PSANTO.3 (1), PSANTO.4 (1)
	Total	122			

The locality codes are as in Fig. 1.

*Nomenclature according to Piálek *et al.* (2005).

†One specimen with haplotype PORTUGAL.10 has 2n = 39.

‡Includes mtDNA haplotypes from Madeira and Porto Santo reported in Gündüz *et al.* (2001); MADE = Madeiran haplotypes, PSANTO = haplotypes from Porto Santo.

§Hybridization between PSAN and 40ST mice has been documented at these localities (Britton-Davidian *et al.* 2000), but no karyotypic heterozygotes were present among our new specimens.

¶Note that haplotypes from Porto Santo were given a different name from that given in Gündüz *et al.* (2001), to avoid confusion with Portuguese haplotypes.

approach was adopted using only two primers: L831 (5'-ATACGCCATTCTACGCTCAA-3') positioned in cytochrome *b* and H2228 (5'-TTATAAGGCCAGGACCAAAC-3') positioned in *12SrRNA* (Searle *et al.* 2009a). This approach yielded a total of 1419 bp (between

14969 and 00093 of Bibb *et al.* 1981). Standard concentrations of DNA and reagents (Bilton *et al.* 1998) were used for double-stranded amplification using the following procedure: an initial cycle of denaturation at 94 °C for 4 min, followed by 30 cycles with denaturation at

94 °C for 30 s, annealing at 52 °C for 30 s and extension at 2 °C for 1 min. This was followed by a final extension at 72 °C for 7 min. Amplified products were purified using Qiagen QIAquick Purification Kit by following the manufacturer's instructions and prepared for automated sequencing. Our sequences have been deposited in GenBank (GQ241989–GQ242020).

Phylogenetic and phylogeographical analyses

The region between positions 15363 and 16295 (numbering after Bibb *et al.* 1981), which will hereafter be referred to as the 'D-loop', was used for the phylogenetic analyses to facilitate comparison with other studies (e.g. Prager *et al.* 1993, 1996, 1998; Nachman *et al.* 1994; Gündüz *et al.* 2005). The Portuguese and Madeiran D-loop haplotypes described in this study and previously (Gündüz *et al.* 2001) were combined with 182 published haplotypes from the distribution of the subspecies in these analyses (Prager *et al.* 1993, 1998; Nachman *et al.* 1994; Gündüz *et al.* 2000, 2001, 2005; Searle *et al.* 2009a, b). Three *Mus musculus castaneus* sequences (AF074532, AF074539 and AJ286322) from Prager *et al.* (1998) and Gündüz *et al.* (2000), as well as three *Mus musculus musculus* sequences (U47504, U47530 and U47532) from Prager *et al.* (1996) were used as outgroup sequences.

Neighbour-joining (NJ), maximum-likelihood (ML) and Bayesian methods were used for phylogenetic reconstruction. All phylogenetic analyses were conducted using the HKY + I + Γ substitution model, as selected by MODELTEST version 3.7 (Posada & Crandall 1998) for the data set (without outgroup). The NJ and ML analyses were carried out using PAUP* 4.0b10 (Swofford 2000). The ML analysis involved heuristic searches with stepwise addition (10 random-sequence-addition replicates) and subtree pruning–regrafting (SPR) branch swapping. Bootstrap values were generated with 10 000 (NJ) and 100 (ML) replicates. A 50% majority rule consensus tree was generated by Bayesian analysis with MRBAYES v3.1.2 (Ronquist & Huelsenbeck 2003). The analysis involved two runs for 6 million iterations, using six chains and a burn-in of 30%, and MrBayes was allowed to estimate the parameters within the HKY model. Overall tree topologies were similar in all cases.

Nucleotide (π) and haplotype (h) diversities were calculated as described by Tamura & Nei (1993) in ARLEQUIN v3.11 (Excoffier *et al.* 2005). The partitioning of molecular variance was determined using analysis of molecular variance (AMOVA), also implemented in ARLEQUIN v3.11.

To test for differentiation of the Rb races in Madeira based on their mtDNA sequences, pairwise Φ_{ST} values were calculated between all races and their significance

tested by permutating sequences among chromosome races (Excoffier *et al.* 1992).

The sequence variation of the Portuguese and Madeiran data sets was examined using two statistics, Tajima's (1989) D and Fu's (1997) F_S , which are based on the infinite-site model without recombination and assume that the populations are at equilibrium. Both of these statistics compare different estimates of θ , the population mutation parameter, to test for selective neutrality. Specifically, they compare different estimates of θ to an estimate of θ based on pairwise nucleotide differences, which has been shown to be sensitive to changes in population size. When significantly negative, this suggests the presence of selection or the occurrence of population expansion.

Another approach, following Rogers & Harpending (1992) and Slatkin & Hudson (1991), was used to examine historical changes in population size for the two data sets. The distribution of pairwise genetic differences between sequences was examined as a mismatch distribution, which was compared with the sudden expansion model of Rogers (1995), as implemented in ARLEQUIN v3.11. The sum of square deviations (SSD) between the observed and expected (model) mismatch distributions and Harpending's raggedness index was used as test statistics (denoted P_{SSD} and P_{rag} respectively; Harpending 1994; Schneider & Excoffier 1999).

The mismatch distribution of pairwise genetic differences was also used to calculate expansion times (Rogers & Harpending 1992). The change of effective female population size (N) since population expansion is used to estimate the time since expansion in generations. A population at equilibrium with N_0 changes to N_1 at τ units of mutational time. The modal value τ is determined from the distribution of pairwise genetic differences in the extant population. The parameters of the model are given by: $\theta_0 = 2N_0u$, $\theta_1 = 2N_1u$ and $\tau = 2ut$. Here, t is the time since expansion in generations and u is the mutation rate of the entire DNA fragment (expressed as $u = m_T\mu$, where m_T is the length of the sequence and μ is the mutation rate per nucleotide per generation). A parametric bootstrap approach with 10 000 pseudo-replicates was used to generate confidence intervals for θ_0 , θ_1 and τ (Schneider & Excoffier 1999).

We also used a Bayesian coalescent approach to infer historical demography for each region using a Bayesian skyline plot (BSP) model (Drummond *et al.* 2005) as implemented in BEAST v1.4.8 (Drummond & Rambaut 2007). Markov Chain Monte Carlo simulations were run under the HKY model with four gamma categories using a strict molecular clock (after appraising the clock-like behaviour of the data, as suggested by the authors). Each analysis was run three times for

30 million iterations. The log files from the three runs per sample were then combined using LOGCOMBINER, discarding the first 10 million iterations of each run as burn-in, and results for the combined log (60 million iterations) are presented. Convergence and effective sample sizes of parameters were assessed in TRACER v1.4. Changes in female effective population size through time were analysed in TRACER v1.4 and summarized as BSPs. To infer expansion times, the time to the most recent common ancestor (MRCA) for each sample was estimated in terms of substitutions per site.

We adopted several values for the mutation rate (μ) and the number of generations per year for the molecular dating analyses. Regarding the number of mouse generations per year, several authors suggest a long-term average (over 100 000s of years) of one to two generations per year for feral mouse populations (e.g. Salcedo *et al.* 2007; Geraldès *et al.* 2008), which is consistent with the two generations per year deduced for feral mouse populations in Skokholm (Berry 1970). However, commensal mice in Switzerland appear to have two to three generations per year (B. König, personal communication), and it is known that the number of generations per year can be four in captivity (e.g. Nachman & Searle 1995). Given the commensal status of the house mouse populations concerned in this study, we reject the possibility of only one generation per year, as seasonal effects are unlikely to reduce the reproductive output of these mice so drastically in a commensal habitat. Following this reasoning, we chose to consider two, three and four generations per year for our analyses. Choosing appropriate values of the mutation rate (μ) was more problematic, as the only estimates of this value for house mice have been determined over large timescales with calibration dates several million years in the past. Such substitution rate estimates are likely to be inappropriate for analyses concerning the timescale considered in this study (100s to 1000s of years), because only a fraction of short-term mutations (i.e. polymorphisms) become fixed in populations and contribute to substitution rate estimates (reviewed in Ho & Larson 2006; but see also Emerson 2007; Ho *et al.* 2007). This notwithstanding, we chose to use an approximation of published substitution rate estimates as our conservative low mutation rate. When considering two generations per year as a long-term average in feral mice (above), the mutation rate (μ) for the control region (D-loop) obtained by Goios *et al.* (2007) is 2.28×10^{-8} /site/generation, which is very similar to the value of 2.05×10^{-8} /site/generation estimated by Geraldès *et al.* (2008). For our conservative low mutation rate (μ), we chose a value of 2×10^{-8} /site/generation. When Goios *et al.* used their substitution rate to estimate divergence times for mouse laboratory

strains, their estimates were 10–30 times higher than the documented establishment dates of these strains (Goios *et al.* 2007; and references therein). Similarly, studies in other organisms that employed substitution rates to infer the timing of recent events have produced molecular date estimates that are highly inconsistent with archaeological estimates (by up to two orders of magnitude; see Ho & Larson 2006; and references therein; see also Rajabi-Maham *et al.* 2008). As there are no published estimates of short-term mutation rates available for the mouse mitochondrial genome (which would be based on recent calibration dates), we used the study of Goios *et al.* (2007) to choose two further estimates of μ . These values are 2×10^{-7} and 6×10^{-7} /site/generation, which reflect the 10- and 30-fold difference between the molecular date estimates (based on the substitution rate) and the known age of the mouse strains studied by Goios *et al.* (2007).

Results of the molecular dating were then considered in the context of the recently reviewed zooarchaeological record of the house mouse for Europe and the Middle East (Cucchi *et al.* 2005), to examine which combination of mutation rate (μ) and generations per year produced time estimates in agreement with the subfossil evidence. Additionally, the molecular dating analyses described above were also carried out for *M. m. domesticus* mtDNA (D-loop) sequences from central France (57 individuals, 28 haplotypes; Ihle *et al.* 2006) and Turkey (98 individuals, 54 haplotypes; Gündüz *et al.* 2000, 2005). These published samples were chosen because (i) there is a suitable zooarchaeological record of high confidence for these geographical regions (Cucchi *et al.* 2005) and (ii) they are similar to our study in terms of specimen and haplotype numbers (above).

Lastly, the relationship amongst haplotypes within the two data sets was examined using median-joining (MJ) networks, as produced in NETWORK v4.5.1 (available at <http://www.fluxus-engineering.com>), in which sequences are nodes of a network rather than terminal tips of a tree. As insertions and deletions can occur repeatedly at the same sites in *M. m. domesticus* D-loop sequences (see Gündüz *et al.* 2000, and references therein) and are thus phylogenetically uninformative, the MJ networks were produced with indel weights set to zero.

Results

Portuguese mtDNA (D-loop) haplotypes: sequence variation

Fifteen D-loop haplotypes (PORTUGAL.1–15) were identified among the 76 mice analysed from mainland Portugal (Fig. 2). Four of the haplotypes reported in this study have been previously described, namely:

Table 2 Measures of genetic variation and neutrality tests on *Mus musculus domesticus* from Portugal, Madeira and Porto Santo

Race	N	N _H	Polymorphic sites			Genetic diversity estimates		Neutrality tests	
			Transit	Transv	Indel	$\pi \pm SE$	$h \pm SE$	D (P)	F _S (P)
Portugal (40ST)*	76	15	22	8	4	0.0064 ± 0.0034	0.882 ± 0.024	-0.058 (NS)	0.825 (NS)
Portugal Main Clade (40ST)*	58	10	14	2	2	0.0039 ± 0.0022	0.817 ± 0.036	-0.426 (NS)	0.798 (NS)
Madeira and Porto Santo†	121	35	20	3	6	0.0015 ± 0.0010	0.929 ± 0.014	-1.917 (0.006)	-19.414 (0.001)
PSAN	72	17	10	2	6	0.0012 ± 0.0009	0.878 ± 0.024	-1.479 (0.035)	-6.134 (0.002)
PPOD	10	5	3	0	3	0.0015 ± 0.0011	0.844 ± 0.079	1.001 (NS)	-0.242 (NS)
PADC	11	9	7	1	4	0.0026 ± 0.0017	0.964 ± 0.051	-0.447 (NS)	-1.050 (NS)
PEDC†	10	6	6	0	3	0.0017 ± 0.0012	0.889 ± 0.075	-0.973 (NS)	0.137 (NS)
PSVI	4	2	0	2		0.0012 ± 0.0012	0.833 ± 0.222	0.592 (NS)	-0.658 (NS)
PLDB	6	2	0	0	1	0.0000 ± 0.0000	0.533 ± 0.172	—	—
40ST (Porto Santo)	8	4	1	1	2	0.0005 ± 0.0006	0.750 ± 0.139	-1.310 (NS)	-0.999 (NS)

Number of mice (N) and haplotypes (N_H), as well as nucleotide (π) and haplotype (h) diversity, is given. Number of polymorphic sites is divided into transitions, transversions and insertions/deletions. The neutrality tests are Tajima's (1989) D and Fu's (1997) F_S. NS, non-significant.

*One specimen from Porto (Portugal) had 2n = 39.

†Excludes the single specimen carrying haplotype MADE.20.

(72% of individuals). Prominent in this lineage is PORTUGAL.1, the most widespread haplotype of those detected in Portugal (Table 1; Fig. 1).

A further lineage includes three haplotypes: PORTUGAL.11–13 (Fig. 3). Individuals of this lineage were trapped in the coastal region of southwestern Portugal (locations 9, 11 and 12), yet were not detected at a site 30 km inland (location 10) (see Fig. 1). The remaining two lineages are represented by a single haplotype each. Only one mouse was observed with the PORTUGAL.14 haplotype, which differs from all other Portuguese D-loop sequences at four nucleotide positions (15493, 15530, 15912 and 16268), and has been observed elsewhere (above). PORTUGAL.15 represents the last lineage, and is the only previously reported D-loop haplotype from Portugal (Prager *et al.* 1993; Gündüz *et al.* 2001).

Portuguese mtDNA (D-loop) haplotypes: demographic history

Nucleotide (π) and haplotype (h) diversities were calculated for the total Portuguese data set, and for the

sequences of the Portugal Main Clade separately (Table 2). The nucleotide (π) and haplotype (h) diversities for the whole Portuguese data set (0.0064 and 0.882 respectively) are similar to studies of comparable geographical scale and sample size (e.g. Italy: π = 0.0075, h = 0.960, Castiglia *et al.* 2005; Greece: π = 0.0105, h = 0.891, Tryfonopoulos *et al.* 2005; Turkey, π = 0.0084, h = 0.984, Gündüz *et al.* 2005).

TWO AMOVAS were carried out to examine the partitioning of molecular variance in Portuguese *M. m. domesticus* under different groupings (Table 4). A city-by-city comparison revealed the proportion of variance among cities to be 17.14% (Table 4), with most variance resulting from within-group variation. As most Portuguese haplotypes were unique to their trapping location, this result probably reflects the high frequency of PORTUGAL.1 at most locations (Table 1). When the four clades revealed by the phylogenetic analysis (Fig. 3) were tested, the variance among groups was 72.52%, whereas 27.48% reflected variance within groups. The large variance among groups in this AMOVA was highly significant ($P < 0.001$), and confirms the presence of several

	Portugal Main Clade	PORTUGAL.11,12,13	PORTUGAL.14	PORTUGAL.15
Portugal Main Clade	3.00	10.22	9.12	12.61
PORTUGAL.11,12,13		1.72	10.37	10.44
PORTUGAL.14			—*	13.24
PORTUGAL.15				—*

Table 3 Mean number of nucleotide substitutions within (bold) and between (above diagonal) the Portuguese mtDNA lineages

*Note that two mtDNA lineages are represented by a single haplotype.

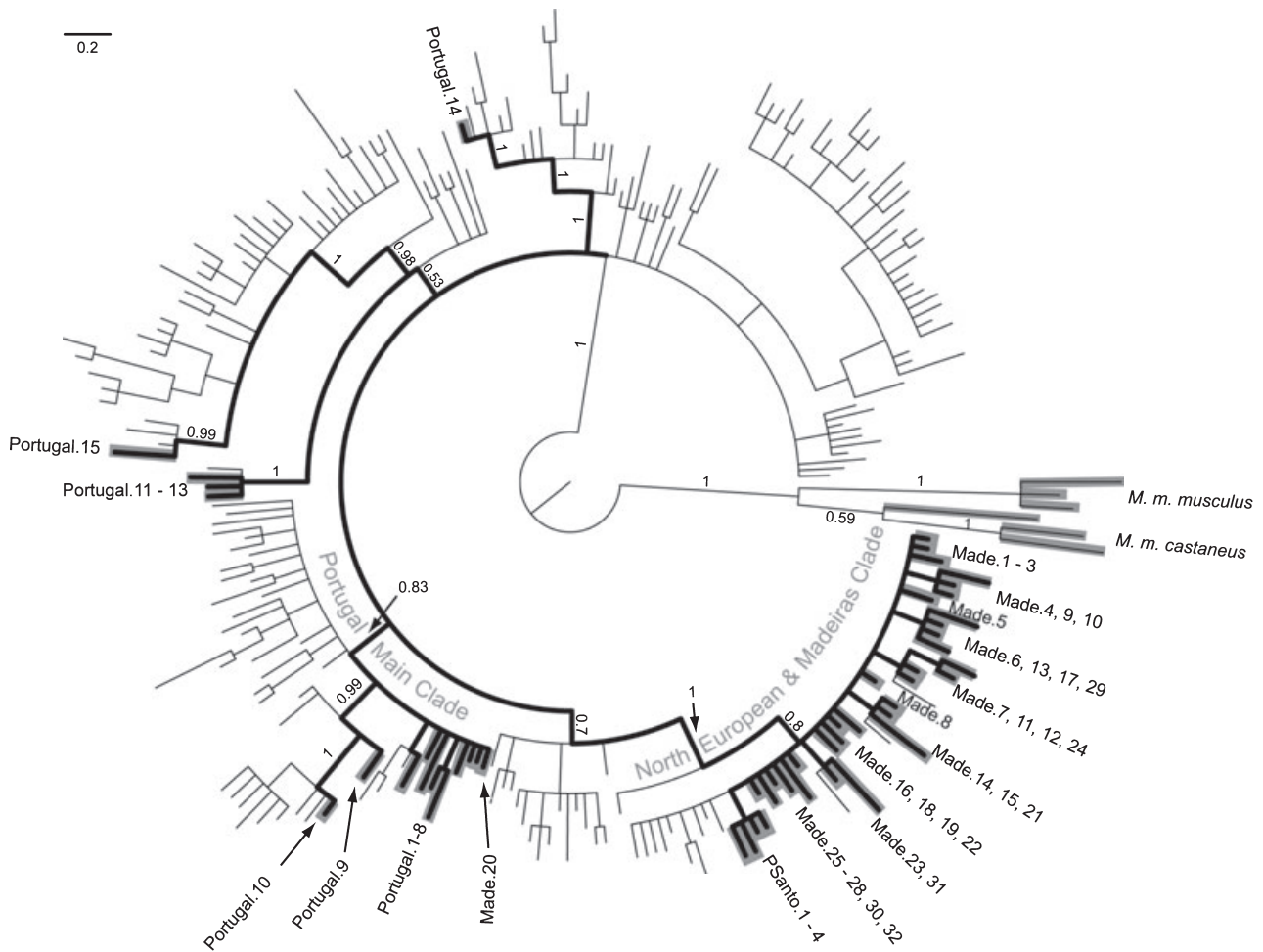


Fig. 3 Fifty per cent majority rule consensus tree after Bayesian analysis of complete D-loop haplotypes obtained in this study together with the haplotypes from Madeira and Porto Santo (Gündüz *et al.* 2001) and 182 equivalent haplotypes from the literature (from Prager *et al.* 1993, 1996, 1998; Nachman *et al.* 1994; Gündüz *et al.* 2000, 2005; Searle *et al.* 2009a, b). Three *Mus musculus castaneus* sequences (AF074532, AF074539 and AJ286322) from Prager *et al.* (1998) and Gündüz *et al.* (2000), as well as three *Mus musculus musculus* sequences (U47504, U47530 and U47532) from Prager *et al.* (1996) were used as outgroup sequences. Posterior probabilities >0.5 for branches leading to haplotypes from Portugal, Madeira and Porto Santo are shown.

distinct mtDNA lineages within the Portuguese data set. This conclusion is corroborated by the analysis of mean pairwise nucleotide substitutions within and between lineages (Table 3).

The mismatch distribution computed for the total Portuguese data set (Fig. 5A) did not differ significantly from Rogers (1995) sudden expansion model ($P_{SSD} = 0.21$, $P_{rag} = 0.07$). Similarly, when considering only sequences belonging to the Portugal Main Clade (Fig. 5B), the mismatch distribution was found to conform to Rogers (1995) model ($P_{SSD} = 0.12$, $P_{rag} = 0.17$). These expansions are also evident from the BSPs generated for these data sets (Fig. 5D,E). As has been previously observed in studies examining the historical demography of *M. m. domesticus* (Gündüz *et al.* 2005; Rajabi-Maham *et al.* 2008), other tests for historical

population expansion, such as Tajima's (1989) D and Fu's (1997) F_s , were not found to be significantly negative (Table 2).

The time since the beginning of these expansions was estimated using two methods (mismatch distribution, MMD: Rogers & Harpending 1992; Bayesian skyline plot, BSP: Drummond *et al.* 2005), and a range of parameter values (see Methods for details). Table 5 shows the results for our samples as well as that of two further *M. m. domesticus* samples (Turkey, Gündüz *et al.* 2000, 2005; central France, Ihle *et al.* 2006). By this approach, we aimed to determine which combination(s) of mutation rate (μ) and mouse generations per year generated molecular date estimates consistent with the zooarchaeological record of *M. m. domesticus* (Cucchi *et al.* 2005). For the three samples with a

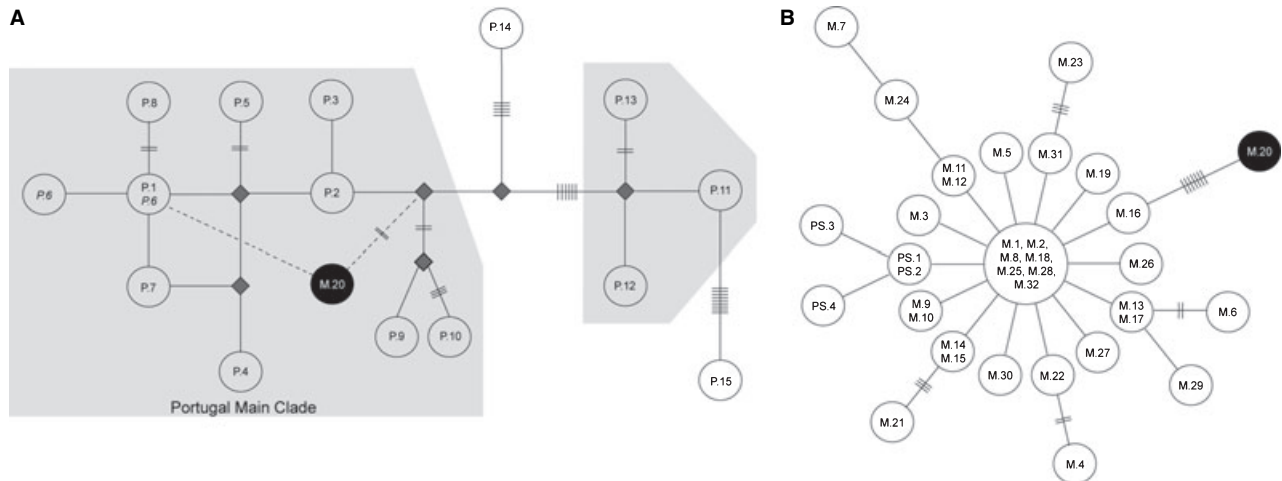


Fig. 4 Median-joining networks of D-loop haplotypes from (A) Portugal and (B) Madeira and Porto Santo (including previously reported sequences from Gündüz *et al.* 2001), with indel weights set to zero. Where there was more than a single nucleotide substitution between haplotypes, this was represented by an equivalent number of bars bisecting a line connecting the two haplotypes (i.e. two bars indicate two nucleotide substitutions, three bars indicate three nucleotide substitutions and so forth). Unsampled or missing nodes are indicated by diamonds. The two Portuguese mtDNA lineages with more than one haplotype are indicated by grey boxes. The Madeiran sequence (MADE.20) that shows greater sequence similarity to Portuguese than Madeiran haplotypes is included in both networks (filled circle). The haplotype PORTUGAL.6 has a substitution at nucleotide position 15550 at which other Portuguese haplotypes are characterized by a deletion (Fig. 2); as this nucleotide position has a weight of zero (above), the haplotype is shown both in its wrongly inferred as well as in its correct position.

suitable subfossil record (Portugal Main Clade, Turkey, central France), two combinations of mutation rate and generations per year produced expansion time estimates in good agreement with subfossil evidence for the first occurrence of house mice in the respective geographical region (Table 5, Fig. 6). Both these combinations include a mutation rate of 2×10^{-7} /site/generation, and involve either three or four house mouse generations per year. Figure 6 summarizes the results for these two combinations, including the 95% confidence

intervals for the MMD analysis, the 95% high posterior density intervals for the BSP analysis and the confidence range of the zooarchaeological record, where available. For the total Portuguese data set, the MMD analyses estimated an expansion time of 9406–7055 years ago and the BSP analyses of 8500–6375 years ago. For the Portugal Main Clade, the expansion dates were estimated to be 2841–2130 years ago in the MMD analyses and 4500–3375 years ago in the BSP analyses.

Source of variation	Variance component	Variance	% Total	Φ -statistic
Portugal				
City-by-city	Among groups	0.55	17.14	0.17
	Within groups	2.67	82.86	
4 Clades	Among groups	4.03	72.52	0.73
	Within groups	1.23	27.48	
Madeira and Porto Santo*				
Location-by-location	Among groups	0.09	12.05	0.12
	Within groups	0.62	87.95	
Chromosome races	Among groups	0.09	8.55	0.09
	Within groups	0.64	91.45	
40ST vs. Rb	Among groups	0.38	35.12	0.35
	Within groups	0.69	64.88	

Table 4 Analysis of molecular variance results for the sequences from Portugal, Madeira and Porto Santo (including sequences from Gündüz *et al.* 2001)

Φ_{ST} value in bold indicates that the among-group variance is significant at the $P < 0.01$ level after compensation for multiple comparisons (Rice 1989). [The significance level (P) is the proportion of 10100 permutations larger than the observed Φ estimate.]

*Excludes specimen with MADE.20.

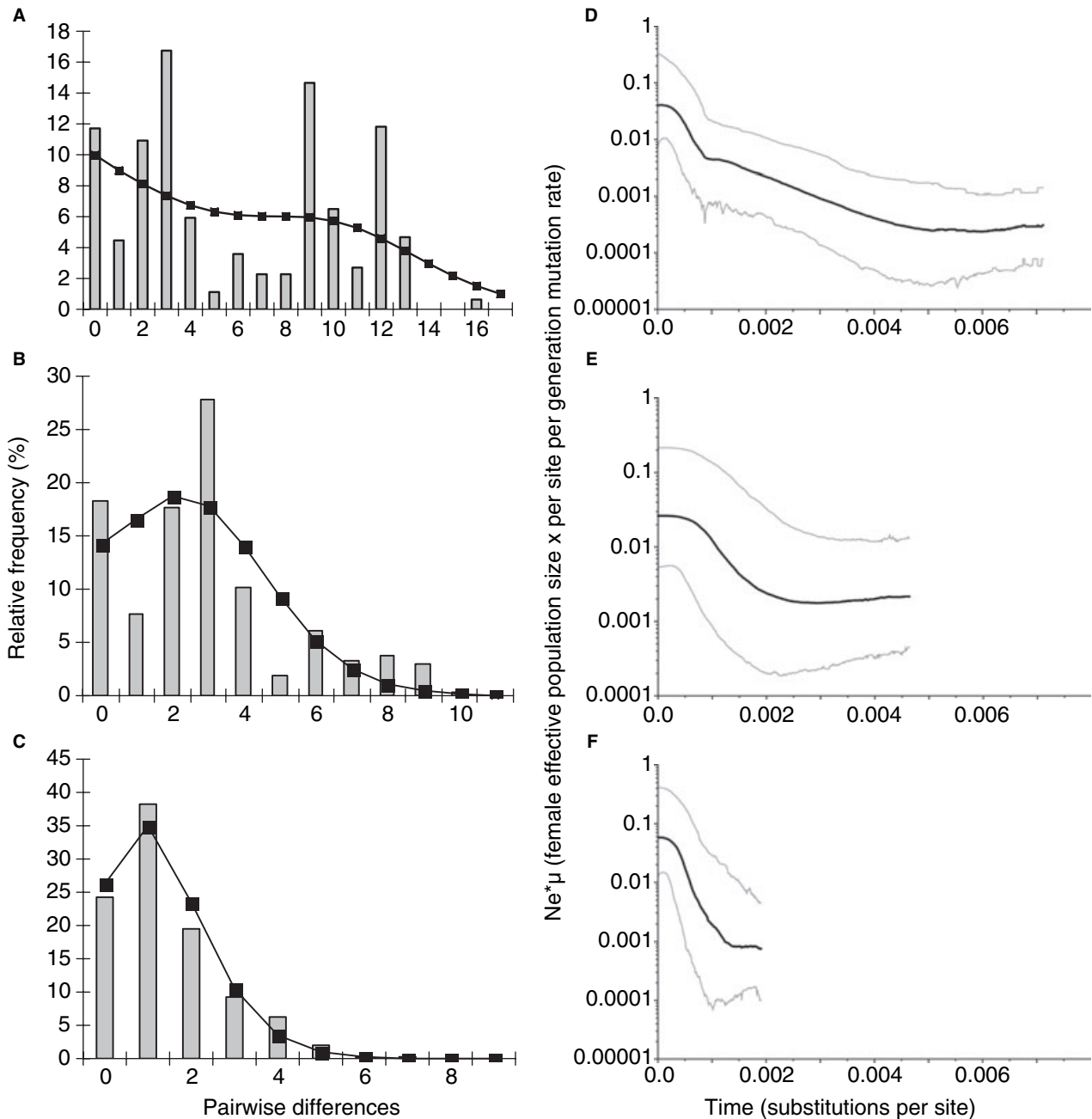


Fig. 5 Demographic history of *Mus musculus domesticus* samples collected in this study inferred from mitochondrial D-loop sequences using two techniques. (A–C) Mismatch distribution of mtDNA sequences (shaded bars) compared with Rogers (1995) sudden expansion model (■). (D–F) Bayesian skyline plots showing the change in female effective population size backwards through time (measured as substitutions per site); the thick solid line is the median estimate and the thin lines show the 95% highest posterior density limits. Sample details in Table 1: (A, D) all Portuguese samples; (B, E) Portugal Main Clade; (C, F) Madeiran archipelago.

Madeiran mtDNA (D-loop) haplotypes: sequence variation

Seventeen new haplotypes (MADE.16–32) were identified among the 80 mice analysed from Madeira (Fig. 2). Only four Madeiran haplotypes reported previously

(Gündüz *et al.* 2001) were not detected again among our new samples (MADE.3,5–7). One previously reported haplotype from Porto Santo (PSANTO.2; PORT.2 in Gündüz *et al.* 2001) was observed in Madeira in this study (Table 1). None of the new sequences from Madeira have been observed elsewhere.

Table 5 Estimates of the time since the beginning of house mouse population expansions (using a range of parameter values) compared with subfossil evidence for the first occurrence of *Mus musculus domesticus* (where available; from Cucchi *et al.* 2005)

		Molecular date estimates (in years)*												
		Mismatch distribution					Bayesian skyline plot					First occurrence based on subfossils (in years)‡		
Location	N	N _H	τ	Mutation rate (μ)					Mutation rate (μ)					Source
				gen/yr	2 × 10 ⁻⁸	2 × 10 ⁻⁷	6 × 10 ⁻⁷	subst/site†	gen/yr	2 × 10 ⁻⁸	2 × 10 ⁻⁷	6 × 10 ⁻⁷		
Portugal (all)	76	15	11.43	2	141 091	14 109	4703	0.0051	2	127 500	12 750	4250	This study	
				3	94 061	9406	3135		3	85 000	8500	2833		
				4	70 545	7055	2352		4	63 750	6375	2125		
Portugal Main Clade	58	10	3.45	2	42 609	4261	1420	0.0027	2	67 500	6750	2250	This study	
				3	28 406	2841	947		3	45 000	4500	1500		
				4	21 304	2130	710		4	33 750	3375	1125		
Madeiran archipelago	121	35	1.34	2	16 486	1649	550	0.0013	2	32 500	3250	1083	This study; Gündüz <i>et al.</i> (2001)	
				3	10 990	1099	366		3	21 667	2167	722		
				4	8243	824	275		4	16 250	1625	542		
Turkey	98	54	11.75	2	144 990	14 499	4833	0.0051	2	127 500	12 750	4250	Gündüz <i>et al.</i> (2000, 2005)	
				3	96 660	9666	3222		3	85 000	8500	2833		
				4	72 495	7250	2417		4	63 750	6375	2125		
Central France¶	57	28	2.75	2	38 537	3854	1285	0.0018	2	45 000	4500	1500	Ihle <i>et al.</i> (2006)	
				3	25 691	2569	856		3	30 000	3000	1000		
				4	19 268	1927	642		4	22 500	2250	750		

Trapping location of mice, number of mice (N), number of haplotypes (N_H) and source of sequence information indicated. †Portugal Main Clade represents a subset of mice from Portugal (see text for details). Mutation rate (μ) is expressed in substitutions per site per generation; gen/yr, generations per year.

*Values in bold were generated using estimates of generations per year and mutation rate (μ) that consistently produced expansion time estimates in good agreement with subfossil evidence (where available). Confidence limits for these estimates are shown in Fig. 6.

†Substitutions per site since the most recent common ancestor.

‡Times for subfossils (considered as ‘certain’, shown as confidence intervals) are taken from Cucchi *et al.* (2005). Possible earlier dates (indicated as ‘probable’ by Cucchi *et al.* 2005) are given below in parentheses.

§These values are for Spanish not Portuguese sites on the Iberian Peninsula (see Cucchi *et al.* 2005).

¶Calculations based on a sequence length of 892 bp.

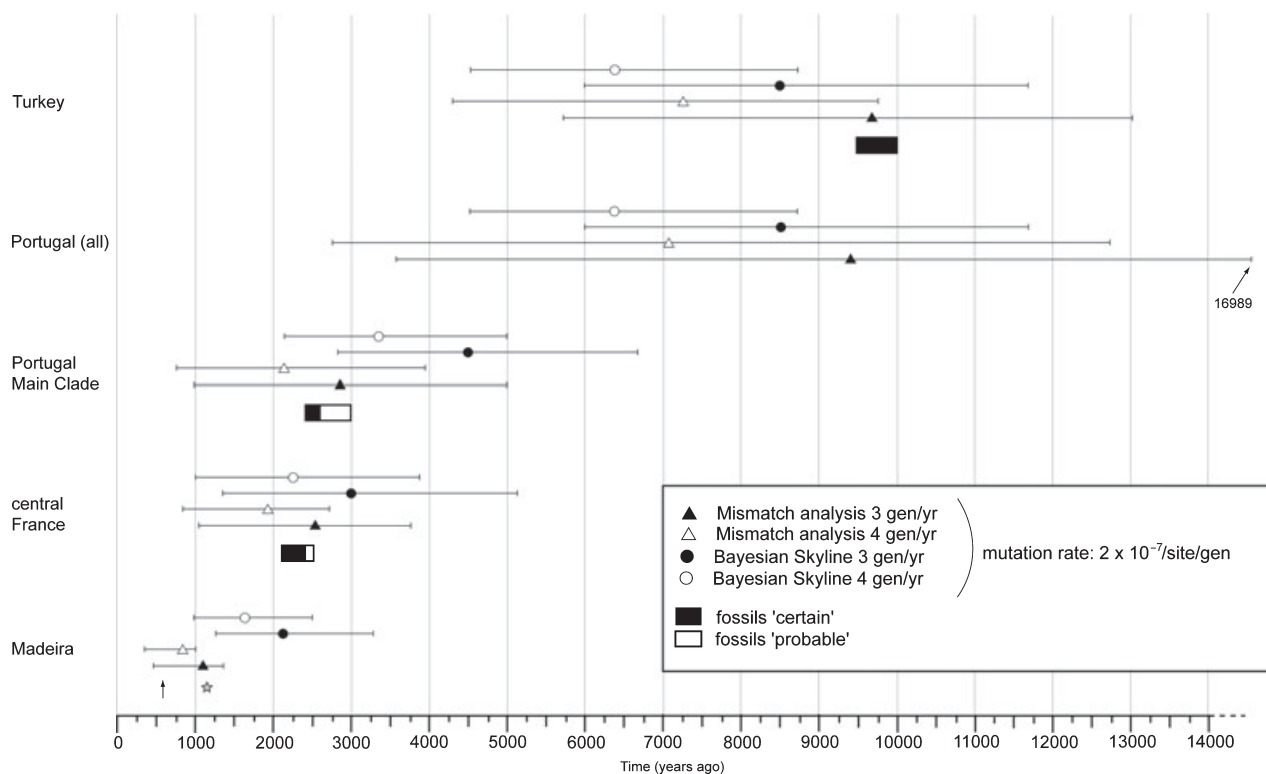


Fig. 6 Schematic summary of molecular date estimates by mismatch distribution (MMD) and Bayesian skyline plot (BSP) methods using the two combinations of parameter values (for mutation rate and mouse generations per year; see legend) that generated expansion time estimates in good agreement with zooarchaeological record (Table 5). Confidence range for MMD (95% CI), BSP (95% HPD) and first occurrence of *Mus musculus domesticus* (where available; from Cucchi *et al.* 2005) is indicated per sample. For the Madeiran sample, the arrow indicates the time of the human settlement in the 15th century and the star indicates the time corresponding to the proposed introduction by Danish Vikings in the 9th century (Gündüz *et al.* 2001).

The number of polymorphic insertions/deletions within the Madeiran data set has increased to six (Fig. 2). Notably, only two further D-loop haplotypes (MADE.22,25) have been detected without the 11-bp insert after 16072, an insert known mostly from north European specimens (e.g. Prager *et al.* 1993; Nachman *et al.* 1994).

MADE.20 is an exception among Madeiran sequences and overall this haplotype is more similar to Portuguese sequences found in this study (Fig. 2), appearing within the monophyletic clade of Portuguese sequences (Portugal Main Clade) reported in this study (Figs 3 and 4). With the exception of MADE.20, all Madeiran haplotypes can be derived from MADE.1 by 0–4 nucleotide substitutions and 0–3 indels. MADE.20 is by far the most divergent sequence, differing from MADE.1 by eight nucleotide substitutions and one insertion (Figs 2 and 4B), whereas it differs from PORTUGAL.1 by only one substitution and two insertions (Figs 2 and 4A). Apart from MADE.20, all Madeiran haplotypes fall into the previously described ‘north European & Madeiras clade’ (Gündüz *et al.* 2001; Fig. 3) that also includes

many north European haplotypes. As in the previous study of Gündüz *et al.* (2001), none of the new D-loop sequences from Madeira have the transversion at position 15540 that characterizes haplotypes from Porto Santo. Unlike the previous study (Gündüz *et al.* 2001), the haplotypes from Porto Santo now form an exclusive clade within the ‘north European & Madeiras clade’ (Fig. 3).

Madeiran mtDNA (D-loop) haplotypes: demographic history

Nucleotide (π) and haplotype (h) diversities were calculated for the Madeiran data set as a whole, as well as for individual chromosome races (Table 2). These values are extremely similar to those previously reported (Gündüz *et al.* 2001). The observed nucleotide diversity of 0.0015 appears low compared with 0.0064 for haplotypes from Portugal (76 individuals, 15 sites; Table 2), 0.0075 from central Italy (73 individuals, 32 sites; Castiglia *et al.* 2005) and 0.0084 from Turkey (98 individuals, 47 sites; Gündüz *et al.* 2005), whereas the haplotype

diversity of 0.929 compares with 0.882, 0.960 and 0.984 for Portugal, central Italy and Turkey respectively. The low nucleotide diversity among the Madeiran haplotypes most probably reflects the presence of many highly related mtDNA sequences (Fig. 5C).

For the whole Madeiran data set, both Fu's (1997) F_S and Tajima's (1989) D are significantly negative (Table 2), consistent with a population expansion. Further, the MJ network constructed with the Madeiran haplotypes has a star-like pattern (Fig. 4B). Such a star-like pattern would be expected as mutations are accumulated in the founding haplotype(s) during a population expansion (Avice 2000). The MMD produced, including specimens from Porto Santo (Gündüz *et al.* 2001), did not differ significantly from Rogers (1995) sudden expansion model ($P_{SSD} = 0.06$, $P_{rag} = 0.15$; Fig. 5C). The expansion is also confirmed by the BSP shown in Fig. 5F.

The time since the beginning of this expansion was estimated using MMD and BSP with a range of parameters (Table 5). Using the combination of parameters that generated molecular dates in good agreement with subfossil evidence with samples from elsewhere in the distribution of *M. m. domesticus* (above), the expansion is estimated to have occurred 1099–824 years ago in the MMD analyses and 2167–1625 years ago in the BSP analyses (Table 5; Fig. 6).

Madeiran mtDNA (D-loop) haplotypes: Robertsonian races

Three AMOVAS were carried out to examine the partitioning of molecular variance among Madeiran mice (Table 4). The first AMOVA grouped samples by their trapping locality in Madeira (letters in Fig. 1 and Table 1), to examine if these populations showed evidence of independent evolution. The proportion of variance among groups (localities) was 12.05%, with most variance attributable to within-group variation (Table 4); this result was not significant ($P = 0.08$). The next AMOVA grouped samples by race, including the standard karyotype race (40ST) from Porto Santo. The proportion of variance among races was low (8.55%)

with most variance resulting from within-race variation (91.45%). This grouping failed to reach significance ($P = 0.16$), indicating that mtDNA variation is not structured according to chromosome race. The last AMOVA comparing the standard karyotype mice from Porto Santo with all Madeiran Rb mice (40ST vs. Rb) revealed the largest among-group variance (35.12%) for the Madeiran data set (Table 4). This result probably reflects the transversion at position 15540 observed only in the Porto Santo haplotypes, and was not significant ($P = 0.10$). The 40ST mice from Porto Santo were also the most distinctive in terms of pairwise Φ_{ST} values (Table 6), and most pairwise comparisons that resulted in statistically significant Φ_{ST} values involved the Porto Santo sample. Only two further pairwise comparisons were found to be significant after correction for multiple comparisons (PSAN & PADC and PSAN & PPOD, Table 6).

Discussion

House mice of Portugal

Our phylogenetic analyses point towards the presence of four D-loop lineages in Portugal (Fig. 3), consistent with multiple colonizations, as suggested for a similar situation in Germany (Rajabi-Maham *et al.* 2008). Following this reasoning, the MRCA of the four lineages would have existed outside Portugal, presumably in the Middle East (Cucchi *et al.* 2005). The expansion times computed using the parameter combinations that generated molecular date estimates in good agreement with the zooarchaeological record (Table 5, Fig. 6) are consistent with this MRCA occurring in the eastern Mediterranean basin, as *Mus musculus domesticus* did not expand into central and western Europe until the Iron Age (1000 BC–300 AD) (Cucchi *et al.* 2005).

The subfossil record confirms the presence of *M. m. domesticus* in Iberia starting 2600–2400 years ago, although an earlier date is considered probable (3000–2900 years ago, Cucchi *et al.* 2005). Our molecular date estimates for the monophyletic group of haplotypes termed the Portugal Main Clade (Table 5, Fig. 6) are

Race	PPOD	PSVI	PEDC	PADC	PLDB	40ST (Porto Santo)
PSAN	0.178	0.118	0.091	0.146	0	0.466
PPOD		0.222	0.145	0.069	0.165	0.541
PSVI			0.126	0.086	0.335	0.634
PEDC				0.102	0.026	0.485
PADC					0.025	0.403
PLDB						0.772

Table 6 Pairwise Φ_{ST} values based on mtDNA (D-loop) sequences among the chromosome races of Madeira and Porto Santo (values in bold are significant after sequential Bonferroni correction)

consistent with the zooarchaeological evidence. The congruence between these expansion time estimates and the subfossil data, and the wide distribution of haplotypes belonging to this lineage in Portugal (particularly PORTUGAL.1) suggest that *M. m. domesticus* with haplotypes of the Portugal Main Clade lineage were important in the early colonization of Portugal.

Origin of Madeiran house mice

The human history of the Madeiran archipelago is strongly linked to Portugal following its documented discovery by Portuguese in 1419 AD (Albuquerque & Vieira 1988), so it may then be expected that Portuguese settlers first introduced house mice to the archipelago in the 15th century. Surprisingly, Gündüz *et al.* (2001) suggested instead that house mice were introduced from northern Europe, based on D-loop sequence similarity. However, one major concern regarding this previous analysis was the scarcity of Portuguese D-loop sequences available at the time, with only two described, constituting a single haplotype from Lisbon (Prager *et al.* 1993; Gündüz *et al.* 2001).

Our new results with 76 mice sampled from 14 localities in Portugal and 80 mice sampled from 20 localities in Madeira support the findings of Gündüz *et al.* (2001). None of the Portuguese mice studied had sequences belonging to the 'north European & Madeiras clade', whereas 99.2% of all mice studied from the Madeiran archipelago had sequences belonging to this clade (which is dominated by haplotypes recorded in Madeira, Denmark, Germany and Sweden, as shown in Fig. S1, Supporting Information). Indeed 54% of *M. m. domesticus* so far sampled from the island of Madeira had six haplotypes (MADE.1–2,10,12,14–15; Table 1) identical to those from northern Europe, which are widely distributed in Madeira (85% of localities, Fig. 1). Only one Madeiran mouse had a haplotype (MADE.20) probably of Portuguese origin (Figs 2 and 4). The mouse sequences in the Madeiran archipelago with a north European affinity most probably derive from a single source area, based on the star-like relationship of haplotypes shown in Fig. 4B and the comparatively low nucleotide diversity (Table 2), particularly taking into account the substantial variation in D-loop sequences when *M. m. domesticus* is considered as a whole (e.g. Prager *et al.* 1993; Nachman *et al.* 1994; Searle *et al.* 2009a). Given the high mutation rate within the D-loop (e.g. Geraldts *et al.* 2008), it is not unreasonable to suppose that new mutations have arisen *in situ* on Madeira following its colonization, as reflected by the MJ network shown in Fig. 4B (e.g. Avise 2000).

Consistent with another of the contentions of Gündüz *et al.* (2001), estimates for the expansion of *M. m. domesticus* in the Madeiran archipelago suggest the arrival of

mice in Madeira prior to the Portuguese settlement in the 15th century (Table 5; Fig. 6; see also Pieper 1981; Mathias & Mira 1992). The expansion of the Portugal Main Clade apparently preceded the Madeiran expansion; therefore, there is good reason to believe that mice with characteristics similar to those seen in Portugal now were available in Portugal to colonize Madeira, but did not do so.

Interestingly, nuclear allozyme markers present a different picture than our results with mitochondrial DNA sequences, showing that the nuclear genome of Madeiran mice appears to be largely Portuguese with only some populations displaying genic similarity to that of northern Europe (Britton-Davidian *et al.* 2007). As these authors have commented, the disparity between the nuclear and mitochondrial genomes may be explained by the observations of Jones *et al.* (1995) regarding another *M. m. domesticus* population. Following an artificial introduction of mice to an island community (Isle of May, Scotland), Jones *et al.* (1995) found that matrilineal markers had not spread as effectively across the island as had nuclear markers, male-specific markers or Rb metacentrics. In that study, asymmetric reproductive success of introduced males over introduced females had resulted in male-biased gene flow into the island community. In terms of introduced alleles, maternal markers were still rare, whereas the nuclear and male-specific markers had spread considerably throughout the study area after only a few years (Jones *et al.* 1995). For example, 3 years after introduction, the majority of males (~62%) already possessed the introduced Y-chromosome markers while still having the pre-introduction mtDNA markers. With regard to the Madeiran mice, a similar colonization-related process might have taken place if, as proposed, the arrival of Portuguese mice represents a secondary colonization event following an earlier arrival of *M. m. domesticus* from northern Europe. Asymmetrical reproduction between early mouse colonists from northern Europe (females) and later arrivals from Portugal (males) would have resulted in the observed pattern of a north European mitochondrial genome and a largely Portuguese nuclear genome (Britton-Davidian *et al.* 2007). If gene flow is biased in favour of introduced (Portuguese) males over females, this could also explain the presence of only one specimen in Madeira with what appears to be a Portuguese mtDNA haplotype (MADE.20).

Whereas the presence of one individual with a Portuguese mtDNA haplotype suggests rare integration of Portuguese female mice into Madeira, the absence of Madeiran mtDNA haplotypes in Portugal suggests that female integration in the other direction is also infrequent. The only uncertain evidence of migration (either male or female) to Portugal by Madeiran mice and

successful interbreeding there is the occurrence of an Rb chromosome fusion 3.8 in Porto, which could have derived from a PSVI, PSAN or PPOD specimen from Madeira. Even comparing the neighbouring islands of Madeira and Porto Santo, mtDNA sequences only reveal one clear case of female migration and integration (a Porto Santo mtDNA sequence found in two individuals in the main harbour of Madeira). Migration and integration of mice (either male or female) in the opposite direction are confirmed by the presence of two mice in Porto Santo with Rb metacentrics (5.14 and 11.12) from PSAN (Britton-Davidian *et al.* 2005).

When the mtDNA results are considered in a human context, the broad range of expansion time estimates (820–2170 years ago: Table 5) using the two molecular dating methods makes for a variety of potential candidates regarding the mariners who (accidentally) introduced *M. m. domesticus* to the archipelago from northern Europe. This includes 9th century Danish Vikings, previously highlighted by Gündüz *et al.* (2001), whose candidature would be in good agreement with one of the time estimates. The basis for suggesting that the mice may have been Danish Viking stowaways is given in detail in Gündüz *et al.* (2001). It is of note that another group of Vikings, the Norwegian Vikings, have been implicated as important human 'vectors' for the transfer of house mice to islands in the north Atlantic (Searle *et al.* 2009a).

Chromosome races of the Madeiran archipelago

The Rb races of house mice in Madeira ($2n = 22-28$) differ in their extent of chromosomal divergence from the standard karyotype race ($2n = 40$), as well as in the combination of chromosomes involved in this karyotypic differentiation (Britton-Davidian *et al.* 2000). There is an expectation that gene flow between Rb races is reduced as a result of fertility reduction experienced by hybrids because of abnormal chromosomal behaviour during gametogenesis (Piálek *et al.* 2005), facilitating the genetic differentiation of such races (e.g. Panithanarak *et al.* 2004). Furthermore, the Madeiran races have probably diverged chromosomally in geographical isolation (Britton-Davidian *et al.* 2000, 2007), and AgNOR activity has been shown to be partly congruent with chromosomal differentiation of the Madeiran Rb races (Ramalhinho *et al.* 2005). In this study, the genetic differentiation of the Madeiran Rb races was assessed by mtDNA (D-loop) sequence analysis. Although we detected significant genetic differences between some races (Table 6), there is no statistical support for mtDNA sequence variation being structured according to karyotype. The most notable mtDNA result is the consistent difference between the standard karyotype ($2n = 40$) mice of Porto Santo and the Madeiran Rb

races ($2n = 22-28$). A single transversion in the D-loop sequences of Porto Santo mice clearly distinguishes them from the Rb populations in Madeira.

There have been other studies examining mtDNA variation among closely related Rb populations that have likewise only revealed consistent differences between Rb populations and neighbouring standard karyotype populations (e.g. *M. m. domesticus*: Hauffe *et al.* 2002; *Graomys griseoflavus*: Catanesi *et al.* 2002). The inability to differentiate such closely related Rb populations using mtDNA sequences may result from the independent evolution of Rb chromosomes and unlinked markers such as mtDNA, as well as the common origin of the Rb populations in such systems. As the strength of a chromosomal barrier to gene flow is expected to be strongest in the late stages of karyotypic divergence, it is possible that there has been insufficient time since the establishment of the chromosomal barrier for mtDNA sequence differences to accumulate between the Rb populations. The latter point in particular is of relevance to the Madeiran chromosomal radiation, which is of very recent origin. This interpretation is also consistent with the allozyme study of Britton-Davidian *et al.* (2007), which was not able to demonstrate a chromosomal barrier using nuclear markers.

Of further significance is the lack of evidence for a reduction in genetic variation among Madeiran mice following karyotypic divergence, both in terms of mtDNA sequence variation (this study) and nuclear genic variation (Britton-Davidian *et al.* 2007). When so many highly divergent races are considered, these results strongly suggest that founder populations during race formation were sufficiently large that the fixation of up to nine Rb metacentric chromosomes was not accompanied by a significant reduction in genetic variation.

Acknowledgements

We are grateful to G. Ganem, J. Catalan, J.-C. Auffray and M. Perriat-Sanguinet for assistance in the collection of mice; to D. Tautz for supplying mtDNA (D-loop) sequence information of 57 French specimens and to F. Jóhannesdóttir for comments on the manuscript. This study benefited from the financial support of the Coopération Scientifique et Technique Ambassade de France/ICCTI/GRICES (Project N°509C1), grants from Fundação para a Ciência e a Tecnologia (PRAXIS/PCNA/C/-BIA/135/96) and SAPIENS (POCTI/BSE/47019/02), and the University of York (Guest Scholarship to Í. Gündüz).

References

- Albuquerque L, Vieira A (1988) *The Arquipélago of Madeira in the XV-century*. Colecção Atlântica. Secretaria Regional do Turismo e Cultura/Centro de Estudos de História do Atlântico, Região Autónoma da Madeira, Funchal, Portugal.

- Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, MA.
- Baker RJ, Bickham JW (1986) Speciation by monobrachial centric fusions. *Proceedings of the National Academy of Science, USA*, **83**, 8245–8248.
- Bayona-Bafaluy MP, Acín-Pérez R, Mullikin JC *et al.* (2003) Revisiting the mouse mitochondrial DNA sequence. *Nucleic Acids Research*, **31**, 5349–5355.
- Berry RJ (1970) The natural history of the house mouse. *Field Studies*, **3**, 219–262.
- Bibb MJ, Van Etten RA, Wright CT, Walberg MW, Clayton DA (1981) Sequence and gene organization of mouse mitochondrial DNA. *Cell*, **26**, 167–180.
- Bilton DT, Mirol PM, Mascheretti S, Fredga K, Zima J, Searle JB (1998) Mediterranean Europe as an area of endemism for small mammals rather than a source for northwards postglacial colonization. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **B265**, 1219–1226.
- Britton-Davidian J, Catalan J, Ramalhinho M *et al.* (2000) Rapid chromosomal evolution in island mice. *Nature*, **403**, 158.
- Britton-Davidian J, Catalan J, Ramalhinho G *et al.* (2005) Chromosomal phylogeny of Robertsonian races of the house mouse on the island of Madeira: testing between alternative mutational processes. *Genetical Research*, **86**, 171–183.
- Britton-Davidian J, Catalan J, Lopez J *et al.* (2007) Patterns of genic diversity and structure in a species undergoing rapid chromosomal radiation: an allozyme analysis of house mice from the Madeira archipelago. *Heredity*, **99**, 432–442.
- Castiglia R, Annesi F, Capanna E (2005) Geographical pattern of genetic variation in the Robertsonian system of *Mus musculus domesticus* in central Italy. *Biological Journal of the Linnean Society*, **84**, 395–405.
- Catanesi CI, Vidal-Rioja L, Crisci JV, Zambelli A (2002) Phylogenetic relationships among Robertsonian karyomorphs of *Graomys griseoflavus* (Rodentia, Muridae) by mitochondrial cytochrome *b* DNA sequencing. *Heredity*, **136**, 130–136.
- Cucchi T, Vigne JD, Auffray J-C (2005) First occurrence of the house mouse (*Mus musculus domesticus* Schwarz & Schwarz, 1943) in the Western Mediterranean: a zooarchaeological revision of subfossil occurrences. *Biological Journal of the Linnean Society*, **84**, 429–445.
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.
- Drummond A, Rambaut A, Shapiro B, Pybus O (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, **22**, 1185–1192.
- Emerson BC (2007) Alarm bells for the molecular clock? No support for Ho *et al.*'s model of time-dependent molecular rate estimates. *Systematic Biology*, **56**, 337–345.
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.
- Excoffier L, Laval G, Schneider S (2005) ARLEQUIN ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.
- Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 913–925.
- Geraldes A, Basset P, Gibson B *et al.* (2008) Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. *Molecular Ecology*, **17**, 5349–5363.
- Goios A, Pereira L, Bogue M, Macaulay V, Amorim A (2007) MtDNA phylogeny and evolution of laboratory mouse strains. *Genome Research*, **17**, 293–298.
- Gündüz İ, Tez C, Malikov V, Vaziri A, Polyakov AV, Searle JB (2000) Mitochondrial DNA and chromosomal studies of wild mice (*Mus*) from Turkey and Iran. *Heredity*, **84**, 458–467.
- Gündüz İ, Auffray J-C, Britton-Davidian J *et al.* (2001) Molecular studies on the colonization of the Madeiran archipelago by house mice. *Molecular Ecology*, **10**, 2023–2029.
- Gündüz İ, Rambaut RV, Tez C, Searle JB (2005) Mitochondrial DNA variation in the western house mouse (*Mus musculus domesticus*) close to its site of origin: studies in Turkey. *Biological Journal of the Linnean Society*, **84**, 473–485.
- Harpending RC (1994) Signature of ancient population growth in a low-resolution mitochondrial mismatch distribution. *Human Biology*, **66**, 591–600.
- Hauffe HC, Fraguadakis-Tsolis S, Mirol PM, Searle JB (2002) Studies of mitochondrial DNA, allozyme and morphometric variation in a house mouse hybrid zone. *Genetical Research*, **80**, 117–129.
- Ho SYW, Larson G (2006) Molecular clocks: when times are a-changin'. *Trends in Genetics*, **22**, 79–83.
- Ho SYW, Shapiro B, Phillips MJ, Cooper A, Drummond AJ (2007) Evidence for time dependency of molecular rate estimates. *Systematic Biology*, **56**, 515–522.
- Ihle S, Ravaoarimanana I, Thomas M, Tautz D (2006) An analysis of signatures of selective sweeps in natural populations of the house mouse. *Molecular Biology and Evolution*, **23**, 790–797.
- Jones CS, Noble LR, Jones JS, Tegelström H, Triggs GS, Berry RJ (1995) Differential male genetic success determines gene flow in an experimentally manipulated mouse population. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **B260**, 251–256.
- King M (1993) *Species Evolution. The Role of Chromosome Change*. Cambridge University Press, Cambridge.
- Mathias ML, Mira A (1992) On the origin and colonization of house mice in the Madeira. *Biological Journal of the Linnean Society*, **46**, 13–24.
- Nachman MW, Searle JB (1995) Why is the mouse karyotype so variable? *Trends in Ecology and Evolution*, **10**, 397–402.
- Nachman MW, Boyer SN, Searle JB, Aquadro CF (1994) Mitochondrial DNA variation and the evolution of Robertsonian chromosomal races of house mice, *Mus domesticus*. *Genetics*, **136**, 1105–1120.
- Nunes AC, Britton-Davidian J, Ramalhinho MG, Capela R, Mathias ML, Ganem G (2005) Influence of physical environmental characteristics and anthropogenic factors on the position and structure of a contact zone between two chromosomal races of the house mouse on the island of Madeira (North Atlantic, Portugal). *Journal of Biogeography*, **12**, 2123–2134.
- Panithanarak T, Hauffe HC, Dallas JF, Glover A, Ward RG, Searle JB (2004) Linkage-dependent gene flow in a house mouse chromosomal hybrid zone. *Evolution*, **58**, 184–192.

- Piálek J, Hauffe HC, Searle JB (2005) Chromosomal variation in the house mouse. *Biological Journal of the Linnean Society*, **84**, 535–563.
- Pieper H (1981) Ein subfossiles Vorkommen der Hausmaus (*Mus musculus* s.l.) auf Madeira. *Bocagiana*, **59**, 1–3.
- Pocock MJO, Hauffe HC, Searle JB (2005) Dispersal in house mice. *Biological Journal of the Linnean Society*, **84**, 565–583.
- Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.
- Prager EM, Sage RD, Gyllensten U *et al.* (1993) Mitochondrial DNA sequence diversity and the colonization of Scandinavia by house mice from East Holstein. *Biological Journal of the Linnean Society*, **50**, 85–122.
- Prager EM, Tichy H, Sage RD (1996) Mitochondrial DNA sequence variation in the eastern house mouse, *Mus musculus*: comparison with other house mice and report of a 75-bp tandem repeat. *Genetics*, **143**, 427–446.
- Prager EM, Orrego C, Sage RD (1998) Genetic variation and phylogeography of central Asian and other house mice, including a major new mitochondrial lineage in Yemen. *Genetics*, **150**, 835–861.
- Rajabi-Maham H, Orth A, Bonhomme F (2008) Phylogeography and postglacial expansion of *Mus musculus domesticus* inferred from mitochondrial DNA coalescent, from Iran to Europe. *Molecular Ecology*, **17**, 627–641.
- Ramalhinho G, Braz C, Catalan J, Mathias ML, Britton-Davidian J (2005) AgNOR variability among Robertsonian races of the house mouse from the island of Madeira: implications for patterns of Rb fusion formation and genetic differentiation. *Biological Journal of the Linnean Society*, **84**, 585–591.
- Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.
- Rogers AR (1995) Genetic evidence for a Pleistocene population explosion. *Evolution*, **49**, 608–615.
- Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.
- Ronquist F, Huelsenbeck JP (2003) MRBAYES3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.
- Salcedo T, Geraldes A, Nachman MW (2007) Nucleotide variation in wild house mice. *Genetics*, **177**, 2277–2291.
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Schneider S, Excoffier L (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics*, **152**, 1079–1089.
- Searle JB (1993) Chromosomal hybrid zones in eutherian mammals. In: *Hybrid Zones and the Evolutionary Process* (ed. Harrison RG), pp. 507–532. Oxford University Press, New York.
- Searle JB, Jones CS, Gündüz İ *et al.* (2009a) Of mice and (Viking?) men: phylogeography of British and Irish house mice. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **B276**, 201–207.
- Searle JB, Jamieson PM, Gündüz İ *et al.* (2009b) The diverse origins of New Zealand house mice. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **B276**, 209–217.
- Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, **129**, 555–562.
- Swofford DL (2000) *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4.0b10. Sinauer Associates, Sunderland, MA.
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.
- Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.
- Tryfonopoulos GA, Chondropoulos BP, Fragedakis-Tsolis SE (2005) Mitochondrial DNA polymorphisms of the house mouse *Mus musculus domesticus* from Greece, focusing on the Robertsonian chromosomal system of north-west Peloponnese. *Biological Journal of the Linnean Society*, **84**, 643–651.

This work formed part of Daniel Förster's PhD studies in Jeremy Searle's laboratory at the University of York. Daniel Förster's research focuses on using molecular techniques to investigate the origin and diversification of lineages. Jeremy Searle is interested in the ecology, population genetics and phylogeography of small mammals, particularly in a European context. This work contributes to a long-term collaborative study on the evolution of Madeiran mice conducted with researchers from Portugal and France.

Supporting information

Additional Supporting Information may be found in the online version of this article.

Fig. S1 The detailed 50% majority rule consensus tree after Bayesian analysis, as summarized in Fig. 3. Posterior probabilities of 0.50 and above are shown. New haplotypes detected in this study are highlighted in blue (Madeiran individuals) and red (Portuguese individuals) (haplotype details in Fig. 2). Codes indicate GenBank number or code used in the publication where the sequence was first reported (Nachman *et al.* 1994; Gündüz *et al.* 2000, 2001, 2005), followed by a list of the countries where the haplotype has been recorded abbreviated as here: CH, Switzerland; DE, Germany; DK, Denmark; EG, Egypt; ES, Spain; FI, Finland; GB, Great Britain; GE, Georgia; GR, Greece; HR, Croatia; IE, Ireland; IL, Israel; IR, Iran; IT, Italy; MA, Morocco; MR, Mauritania; NO, Norway; NZ, New Zealand; PE, Peru; PT, Portugal; PT*, Madeira (Portuguese dependency); SE, Sweden; TR, Turkey; US, United States.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.