Mitochondrial DNA sequence variation and evolution of Old World house mice (Mus musculus)

Authors: Macholán, Miloš, Vyskočilová, Martina Mrkvicová, Bejček, Vladimír, and Šťastný, Karel

Source: Folia Zoologica, 61(3–4) : 284-307

Published By: Institute of Vertebrate Biology, Czech Academy of Sciences

URL: https://doi.org/10.25225/fozo.v61.i3.a12.2012

Folia Zool. – 61 (3–4): 284–307 (2012)

Mitochondrial DNA sequence variation and evolution of Old World house mice (*Mus musculus*)

Miloš MACHOLÁN[1]*, Martina MRKVICOVÁ VYSKOČILOVÁ[1], Vladimír BEJČEK[2] and Karel ŠŤASTNÝ[2]

[1] *Laboratory of Mammalian Evolutionary Genetics, Institute of Animal Physiology and Genetics, Academy of Sciences of the Czech Republic, v.v.i., Veveří 97, 602 00 Brno, Czech Republic; e-mail: macholan.iach.cz*
[2] *Department of Ecology, Faculty of Environmental Sciences, Czech University of Life Sciences, Kamýcká 129, 165 21 Prague, Czech Republic*

Received 30 April 2012; Accepted 25 June 2012

Abstract. We analyzed sequences of two variable segments of the mitochondrial control region (CR) and flanking regions in the house mouse (*Mus musculus*). Most of the material was sampled from the eastern Mediterranean and the Middle East, i.e., a source area for the colonization of Europe. These sequences were supplemented with other samples from the whole range of the species including the Yemeni island of Socotra. This island was shown to harbour mice bearing *M. m. domesticus* and *M. m. castaneus* CR haplotypes. In addition, we found 10 distinct sequences at the same locality that were markedly different from all known CR sequences. Sequencing of the whole mitochondrial genome suggested these sequences to represent nuclear fragments of the mitochondrial origin (numts). We assessed genetic variation and phylogeography within and among the house mouse subspecies and estimated the substitution rate, coalescence times, and times of population expansion. We show the data to be consistent with time dependency of substitution rates and recent expansion of mouse populations. The expansion of European populations of *M. m. musculus* and *M. m. domesticus* estimated from the CR sequences coincide with presumed time of colonization of the continent in the Holocene.

Key words: Bayesian skyline plot, control region, house mouse, numt, phylogeography

Introduction

The house mouse (*Mus musculus* Linnaeus, 1758) is undoubtedly one of the best studied animals. The reason for this is twofold. First, thanks to its close association with humans this species has colonized all continents including some extreme habitats (Berry 1981) and because of consuming and/or contaminating human food and crops, damaging property, and spreading disease, it has become an important "weed animal" (Berry 1995). Second, a great importance of the house mouse stems from its role as a laboratory and model animal, with the whole genome sequence (Mouse Genome Sequencing Consortium 2002) and dense maps of various molecular markers available (Dietrich et al. 1996, Lindblad-Toh et al. 2000, Abe et al. 2004, Petkov et al. 2004, Pletcher et al. 2004, Shifman et al. 2006). However, in the 1980's the most widely used inbred strains were shown to be complex mosaics of genomes of different house mouse subspecies (Ferris et al. 1983, Bishop et al. 1985, Blank et al. 1986, Bonhomme et al. 1987, Nishioka 1987) and this has turned attention to the systematics and biology of wild mouse populations. Another motive triggering studies of wild mice was the increasing demand for new sources of genetic variation and for new inbred lines. These efforts resulted in substantial advances in our knowledge of genetic variation and systematic relationships both within the *Mus musculus* complex (Moriwaki 1994, Boursot et al. 1996, Din et al. 1996, Prager et al. 1996, 1998) and within the whole genus *Mus* (She et al. 1990, Catzeflis & Denys 1992, Lundrigan et al. 2002, Chevret et al. 2003, Tucker et al. 2005, Veyrunes et al. 2005, Cucchi et al. 2006; see Auffray & Britton-Davidian 2012 and Suzuki & Aplin 2012 for recent reviews).

The *Mus musculus* complex consists of at least three subspecies (also referred to as species: for

* *Corresponding Author*

arguments see, e.g., Prager et al. 1993, Sage et al. 1993): *musculus*, occurring from central Europe to the Far East; *domesticus*, which lives in western and southern Europe, North Africa and the Middle East and which has also spread to the Americas, Australia and Africa south of Sahara; and *castaneus* found in south-eastern Asia. Among these taxa, various levels of intergradation exist, from narrow hybrid zone between *musculus* and *domesticus* in Europe (see Baird & Macholán 2012 for a recent review) to large-scale intergradation between *musculus* and *castaneus* in the Far East and Japan giving rise to a hybrid form originally described as *M. m. molossinus* (see Yonekawa et al. 2012 for a recent review). A complex situation exists in the Indian subcontinent, Afghanistan, Pakistan, and Iran, where mice were found to possess predominantly *castaneus*-type mtDNA and an extensive variation in autosomal genes (Boursot et al. 1996, Din et al. 1996). Strikingly, while the *castaneus* Y chromosome occurs across India and most of Pakistan, the *musculus*-type Y was found both to the west (Iran, Afghanistan) and to the north and east (central and south-eastern Asia) from this region (Prager et al. 1996, Boissinot & Boursot 1997, Rajabi Maham 2007).

The central position of south-central populations, together with their high variation, led to proposing the "centrifugal" or "out-of-India" hypothesis of the house mouse evolution, assuming SC Asia a "cradle" of the species where rapid diversification of ancestral populations has happened, followed by radiations to the present-day ranges (Boursot et al. 1996, Din et al. 1996, Guénet & Bonhomme 2003, Duvaux et al. 2011). This hypothesis was subsequently challenged by Prager et al. (1998), who suggested an alternative scenario called a "linear" or "sequential" model. The reason for challenging the then widely accepted model was revelation of a new and deeply divergent mitochondrial (mtDNA) lineage from Yemen (Prager et al. 1998). The population from the southern part of the Arabian Peninsula was described as a distinct subspecies *M. m. gentilulus* by Harrison (1972; see also Harrison & Bates 1991). This lineage was later discovered also on Madagascar (Duplantier et al. 2002). Because of the basal position of the Yemeni sample, Prager et al. (1998) assumed occurrence of an ancestral population in the Near East and a series of branching events giving rise, in turn, to *domesticus* and *gentilulus*, with the *musculus-castaneus* split being the last. Regardless of the phylogenetic pattern, there is now a broad consensus that *M. m. domesticus* has colonized Europe from the Near East through Asia Minor whereas *M. m. musculus* followed a route north of the Black Sea (reviewed in Boursot et al. 1993 and Sage et al. 1993). According to paleontological evidence the two subspecies were not present in Europe before the Holocene, a view challenged by Sage et al. (1990), who estimated the origin of major European *M. m. domesticus* clades at 60000-180000 year ago (60-180 kya) and hypothesized the presence of this subspecies in Europe as early as in 30-40 kya. This hypothesis was refuted by Auffray & Britton-Davidian (1992) and a later reassessment of the paleontological record showed the westward expansion of *M. m. domesticus* to occur in two steps: a quick spread into the eastern Mediterranean around the 8th millennium BC with the second wave to western and northern Europe waiting until the 1st millennium BC (Cucchi et al. 2002, 2005; see also Cucchi et al. 2012 for review).

Mus musculus forms a monophyletic group with three free-living ("aboriginal" *sensu* Sage 1981) species, *M. macedonicus*, *M. spicilegus*, and *M. cypriacus*; the fourth aboriginal species, *M. spretus*, is usually considered a sister taxon of this clade (Bonhomme et al. 1984, She et al. 1990, Suzuki & Kurihara 1994) though some other studies have suggested *M. spretus* to be associated with other aboriginal species to the exclusion of *M. musculus* (see Sage et al. 1993 for review) or yielded equivocal results (Lundrigan et al. 2002, Tucker et al. 2005). With respect to the genetic variation and phylogeography within these species, by far the highest attention has been paid to *M. m. domesticus* (Ferris et al. 1983, Britton-Davidian 1990, Sage et al. 1990, Gündüz et al. 2000, 2005, Rajabi Maham et al. 2008, Gabriel et al. 2011, Jones et al. 2011a, b). Britton-Davidian (1990) found substantial genetic differences between populations from northern and southern Europe, the Levant, and North Africa on the allozyme level, with no gradient in the degree of heterogeneity along an assumed colonization route. They attributed the absence of a heterogeneity gradient to human-mediated, long-distance gene flow. Conversely, Sage et al. (1990) found higher mitochondrial (mtDNA) variation in the Mediterranean region than in north-western Europe. Recently, *M. m. domesticus* populations from Turkey and Iran have been studied by Gündüz et al. (2000, 2005), who reported the existence of two distinct lineages one of which was hypothesized to colonize Europe after the last glacial maximum. This result was corroborated by a study of Rajabi Maham et al. (2008), who analyzed samples from Turkey, Iran, France, Germany, Italy, and Bulgaria. They

concluded that their results are in agreement with a scenario assuming the Fertile Crescent as a cradle of commensalism (Auffray et al. 1988, 1990) and the region where a subsequent westward expansion of this subspecies began. This expansion was hypothesized to occur along at least two separate routes, tentatively termed "Mediterranean" and "Bosphorus-Black Sea", respectively (Rajabi Maham et al. 2008).

Compared to *M. m. domesticus*, studies of genetic variation and phylogeography of other commensal taxa have been rather scarce (Prager et al. 1996, 1998). Based on *t* haplotype diversity, nuclear and mitochondrial restriction fragment length polymorphisms (RFLP), and mitochondrial control region sequence data, *M. m. musculus* has been regarded genetically as more homogenous than *M. m. domesticus* (Figueroa et al. 1987, Klein et al. 1987, 1988, Ruvinsky et al. 1991, Boursot et al. 1996) and with a less deep phylogeny (Prager et al. 1996, 1998). However, other mitochondrial RFLP studies (Ferris et al. 1983), protein electrophoresis data (summarized in Sage et al. 1993), and microsatellite data (Dallas et al. 1995) have shown similar level of genetic variation in the two subspecies. On the other hand, *M. m. castaneus* mtDNA haplotypes have been shown to be the most divergent (Boursot et al. 1996, Boissinot & Boursot 1997) with the phylogeny being the deepest of all the house mouse subspecies (Prager et al. 1998). In this paper we extended previous studies of genetic variation and phylogeography of house mice by sequencing samples from other regions across the *M. m. musculus*, *M. m. domesticus*, and *M. m. castaneus* ranges. The main body of the material originated from the eastern Mediterranean and the Middle East, i.e., areas of proposed origins of mouse commensalism and an important gateway for the colonization of Europe. In addition, we focused on the Yemeni island of Socotra in hopes of finding another *M. m. gentilulus* population. Surprisingly, rather than *gentilulus* we found a new clade of mitochondrial sequences on that island, quite divergent from all other house mouse subspecies. These sequences were suspected to represent parts of nuclear DNA translocated from mtDNA known as "numts". This hypothesis was confirmed through sequencing the whole mitochondrial genome of one individual possessing the suspected numt. Together with GenBank sequences, the total material analyzed consisted of more than 580 individuals and thus allowed us to assess genetic variation and phylogeography within and among the house mouse subspecies, and to estimate, within a Bayesian framework and with a reasonable precision, the substitution rate, coalescence times, and times of population expansion. We show that our data are consistent with the notion of time dependency of substitution rates and recent expansion of mouse populations.

Material and Methods

Specimens

A total of 81 new samples of house mice, collected from 31 sites scattered across the northern and southeastern Europe, Middle East, and the Yemeni island of Socotra (Table 1, Fig. 1) were analyzed. To establish the relative systematic identity of these samples, 17 published sequences of the mitochondrial control region (CR) and flanking regions were retrieved from GenBank and added to the material: five *M. m. domesticus* (accession numbers: U47435, U47452, U47455, U47480, U47496), three *M. m. musculus* (U47498, U47532-U47533), three *M. m. castaneus* (U47534, AF074513, AF074518), and six *M. m. gentilulus* sequences (AF074540-AF074545) (Prager et al. 1993, 1996, 1998), together with a single sequence of an individual of the *castaneus*-derived strain CAS (established from mice collected in Thailand and Indonesia; provided by F. Bonhomme and A. Orth). Sequences of *M. spretus* (U47539), *M. macedonicus* (EU106235), *M. spicilegus* (EU106321), and *M. cypriacus* (EU106195) were used as outgroups.

To assess the pattern of genetic variation within the taxa and to infer phylogeographic relationships among haplotypes, the following sequences were retrieved from GenBank and used alongside of new haplotypes identified in the present study: 82 sequences of *M. m. domesticus* from Great Britain, Spain, Portugal, Switzerland, Italy, Finland, Sweden, Norway, Croatia, Greece, Morocco, Egypt, Israel, Turkey, Iran, Georgia, USA, and Peru (accession numbers: AF074490-AF074503, U47430-U47497), 18 sequences of *M. m. castaneus* from Thailand, Taiwan, China, Pakistan, Afghanistan, and Iran (U47534, AF074513-AF074518, AF074520-AF074522, AF074527-AF074529, AF074532, AF074535, AF074537-AF074539), and 37 sequences of *M. m. musculus* from Czech Republic, Slovakia, Austria, Germany, Poland, Croatia, Serbia, Moldova, Ukraine, Russia, Georgia, Turkmenistan, Afghanistan, and Japan (AF074504-AF074507, U47498-U47509, U47512-U47513, U47515-U47533) (Prager et al. 1993, 1996, 1998). A single unpublished *domesticus* sequence from Tenerife, Canary Islands, was also included. The whole data set used for estimations of genetic

variation consisted of 584 individuals including the GenBank sequences (398 *domesticus*, 142 *musculus*, 34 *castaneus*, and 11 undetermined individuals from Socotra) whereas 212 distinct haplotypes (117 *domesticus*, 47 *musculus*, 32 *castaneus*, 6 *gentilulus*, and 10 undetermined haplotypes from Socotra) were used for phylogeographic analyses. The outgroup of the whole *Mus* sample consisted of five GenBank sequences of *Rattus norvegicus* inbred strains GK/Swe (DQ673913), SS/JrHsd/Mcwi (DQ673914), T2DN/Mcwi (DQ673915), Wild/Mcwi (DQ673916), and Wild/Tku (DQ673917).

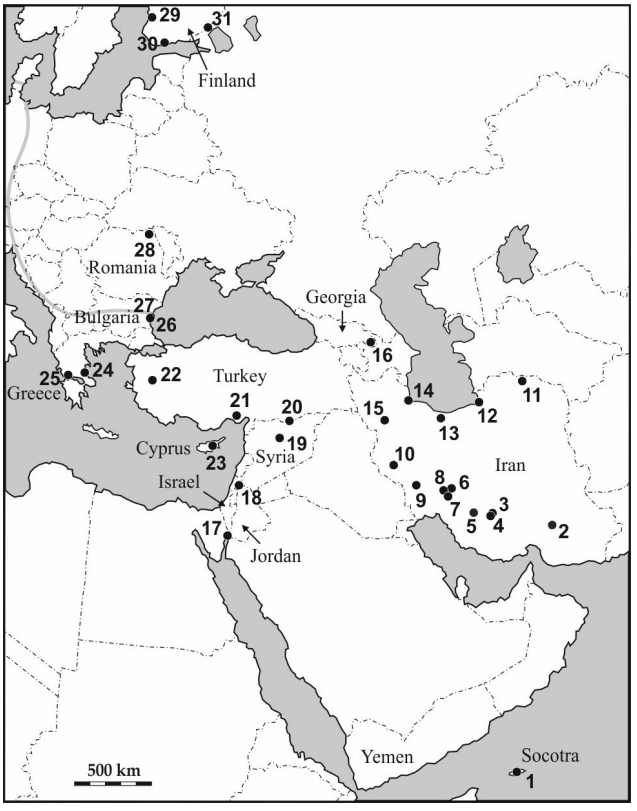

Fig. 1. *Geographical distribution of the house mouse samples analyzed. The thick grey line depicts the hybrid zone between* Mus musculus musculus *and* M. m. domesticus *in Europe.*

Sequencing

DNA was extracted from ethanol-preserved tissues, dried skin or bone using DNeasy Tissue Kit (Qiagen) following manufacturer's instructions. Two segments encompassing variable domains of the mtDNA control region and parts of flanking tRNA genes were amplified using primer pairs L15320-H15782, and L15911-H00072, respectively, described in Prager et al. (1993, 1996; see also Macholán et al. 2007b). Aliquots of 50 ng of DNA were amplified in 30 µl of the PCR reaction buffer with 1.5 mM $MgCl_2$, 200 µM dNTPs, 0.5 units of *Taq* polymerase (Fermentas), and 0.5 µM of each primer. Amplifications were carried out in a gradient RoboCycler thermal cycler (Stratagene) with 35 cycles of 94 °C for 40 s, 53 °C (L15320-H15782) or 55 °C (L15911-H00072) for 40 s, and 72 °C for 2 min. PCR products were checked on 1.5 % agarose gels and then purified using QIAquick MinElute PCR Purification Kit (Qiagen); both light and heavy strands were sequenced in Macrogen Inc. (South Korea). As a result, two segments were obtained, the first one between positions 15313 and 15789 and the second one between 15901 and 00081 of the standard mouse mtDNA sequence (Bibb et al. 1981) so that the length of final concatenated sequences was ≈916 bp (without primer sequences). GenBank accession numbers of new sequences are JX658075-JX658132.

Sequences were aligned with ClustalX v. 1.83 (Thompson et al. 1997) using default values and checked manually. Nucleotide composition and frequencies of substitutions were detected with DAMBE v. 4.2.13 (Xia & Xie 2001). Substitution saturation of the sequences was tested as suggested by Philippe & Douzery (1994) and Hassanin et al. (1998): for the whole data set, values of Tamura-Nei distances with invariant sites and unequal substitution rates (TrN + I + Γ; Tamura & Nei 1993) were plotted against *p*-distances. When the slope of the linear regression $S \approx 1$ there is little or no saturation in the sequences whereas S approaching zero indicates increasing saturation. The distances were computed with the PAUP* 4.0b10 program package (Swofford 2000).

Genetic analyses

Genetic variation was assessed by estimating indices of haplotype diversity, h (Nei 1987), and nucleotide diversity, π (Tajima 1983). Analysis of molecular variance (AMOVA; Excoffier et al. 1992, Excoffier 2007) was applied to partition the variation into hierarchical levels: within populations, among populations within regions, and among regions. The populations were deliberately grouped into four regions: "northern Europe" (Sweden, Norway, Finland, Denmark, northern Germany, and Great Britain), "southern Europe" (Spain, Portugal, southern Germany, Austria, Switzerland, Italy, Serbia, Croatia, and Greece), "Near East" (Israel, Jordan, Syria, Turkey, and Cyprus), and "Iran".

Tajima's D and Fu's F_s statistic was used for estimating potential deviation from selection neutrality and/or

Table 1. *Collection sites, sample sizes (N), mtDNA types, haplotype designations, and site locations of all new samples analyzed.*

No.	Country	Locality	*N*	*mtDNA*	Haplotype	Latitude/Longitude
1	Socotra	Hadiboh	11	SocotraX	*numt*SO1-10	12°39′N 54°01′E
	(Yemen)		9	*domesticus*	*dom*SO1-9[a]	
			7	*castaneus*	*cas*SO1-5	
2	Iran	Deh Bakri	1	*castaneus*	*cas*IR8	29°05′N; 57°55′E
3	Iran	Pasargan	1	*castaneus*	*cas*IR7	30°12′N; 53°10′E
4	Iran	Sivand	2	*castaneus*	*cas*IR5, 6	30°05′N; 52°56′E
5	Iran	Abshar	1	*domesticus*	*dom*IR12	30°23′N; 51°30′E
6	Iran	Ize	2	*domesticus*	*dom*IR8, 9	31°45′N; 49°48′E
7	Iran	Kuli Alireza	2	*domesticus*	*dom*IR10,11	31°15′N; 49°39′E
8	Iran	Simili	5	*domesticus*	*dom*IR5-7	31°41′N; 49°24′E
9	Iran	Choqa Zambíl	1	*domesticus*	*dom*IR4	32°00′N; 48°31′E
10	Iran	Bavineh	2	*domesticus*	*dom*IR2, 3	33°36′N; 47°11′E
11	Iran	Bik	1	*musculus*	*mus*IR1	37°36′N; 57°56′E
12	Iran	Nowkandeh	1	*castaneus*	*cas*IR1	36°42′N; 53°54′E
13	Iran	Valiabad	1	*domesticus*	*dom*IR1	36°15′N; 51°18′E
14	Iran	Asalem	2	*castaneus*	*cas*IR2, 3	37°40′N; 48°50′E
15	Iran	Choplu	1	*castaneus*	*cas*IR4	36°28′N; 47°01′E
16	Georgia	Alazani	1	*domesticus*	*dom*GE1	41°37′N; 45°58′E
17	Jordan	Al-Aqabah	2	*domesticus*	*dom*JOR1, 2	29°31′N; 35°00′E
18	Israel	Kursi Beach	1	*domesticus*	*dom*ISR1	32°49′N; 35°39′E
19	Syria	Lake Asad	1	*domesticus*	*dom*SYR1	35°49′N; 38°28′E
20	Turkey	Harran	1	*domesticus*	*dom*TR2	36°51′N; 39°00′E
21	Turkey	Adana	1	*domesticus*	*dom*TR1	37°00′N; 35°20′E
22	Turkey	Bardakci	1	*musculus*	*mus*TR1	39°10′N; 28°34′E
23	Cyprus	Kornos	2	*domesticus*	*dom*CY1, 2	34°56′N; 33°24′E
24	Greece	Kamena Vourla	2	*domesticus*	*dom*GR2, 3	38°47′N; 22°47′E
25	Greece	Messolongi	1	*domesticus*	*dom*GR1	38°23′N; 21°25′E
26	Bulgaria	Slanchev Briag	1	*musculus*	*mus*BG1	42°42′N 27°43′E
27	Bulgaria	Banya	2	*musculus*	*mus*BG1, *mus*RO1	42°48′N 27°53′E
28	Romania	Botoşani	4	*musculus*	*mus*RO1-4	47°45′N; 26°40′E
29	Finland	Ulvila	2	*musculus*	*mus*FL1-2	61°34′N; 29°34′E
30	Finland	Suitia	9	*domesticus*	*dom*FL1-4	60°01′N; 24°12′E
31	Finland	Parikkala	1	*domesticus*	*dom*FL5	61°25′N; 21°57′E

Note: [a] haplotype domSO9 was revealed by sequencing the whole mitochondrial genome.

recent population expansion or decline. The former method tests the null hypothesis that the estimate of the average number of pairwise nucleotide differences is equal to the estimate of the number of segregating sites (Tajima 1989a). As pointed out by Tajima (1989b), the current population size affects more strongly the number of segregating sites whereas the average number of nucleotide differences is more influenced by the size of the original population. Fu's F_s estimates the probability of obtaining a random neutral sample with a number of alleles equalling or smaller than the observed value given the observed number of pairwise differences (Fu 1997). The *P* values of both of the statistics were tested by generating 10000 random samples under the hypothesis of selective neutrality and/or population equilibrium using a coalescent algorithm adopted from Hudson (1990; see also Schneider et al. 2000). As pointed out by Fu (1997), the F_s statistic is very sensitive to population expansion, leading to highly negative values. The 2 % percentile of its distribution corresponds to the critical value at the 5 % significance level and hence the test can be regarded as significant at 5 % if $P < 0.02$. All the tests were carried out for each taxon using the Arlequin v. 2.000 program (Schneider et al. 2000). The *M. m. gentilulus* samples were excluded from the analyses due to small sample size.

A second test of population demographic equilibrium was based on a mismatch distribution of substitution differences between pairs of haplotypes. The observed

values were compared to the values expected from the population expansion model with parameters estimated using the generalized non-linear least-squares approach of Schneider & Excoffier (1999; see also Li 1977, Harpending 1994, Rogers 1995) using Arlequin, with parameters estimated from the evolutionary models chosen with ModelTest v. 3.7 (Posada & Crandall 1988). The differences between the observed and expected mismatch distributions were expressed as the sum of squared deviations (SSD) and/or as the raggedness index of the observed distribution (Harpending 1994). Sites missing in more than 10 % of the sequences were excluded from all analyses performed with Arlequin.

Finally, a potential population expansion was tested within the likelihood framework. More specifically, we compared two models, the first one assuming stability of population size through time and the second one assuming exponential growth or decline of the population (Kuhner et al. 1995, 1998). The analysis was carried out using Lamarc v. 2.1 (Kuhner 2006). The Hasegawa-Kishino-Yano (HKY) model (Hasegawa et al. 1985) with parameters estimated through ModelTest was used while the exponential growth rate g and population size θ parameters were allowed to vary. The θ parameter is given as $N_e\mu$ where N_e is the effective population size, i.e., twice the number of reproducing females in mtDNA, and μ is the mutation rate per site per generation. Since increasing the number of sequences causes the volume of searched tree-space to increase more than polynomially, getting reliable estimates may be almost impossible for large samples. Therefore, we reduced the sample sizes as follows. First, we excluded sequences from remote areas (Tenerife, USA, Morocco, China, Taiwan, Thailand, and Japan). To minimize the influence of haplotypes present in many copies in the data simply due to intensive sampling of some areas relative to others (e.g., the *M. m. musculus*/*M. m. domesticus* hybrid zone in southern and northern Germany; Tucker et al. 1992, Prager et al. 1993, 1996, 1998), we limited the maximum number of identical sequences to five. From resulting samples, we randomly chose 30 sequences. Two sets of 10 short ("initial") and two long ("final") Metropolis-coupled Monte Carlo Markov chains (MCMCMC) of 6000 and 20000 generations, respectively, were run with the first 1000 and 2000 steps, respectively, discarded as burn-in. Four simultaneous searches, one cold and three heated, were carried out with the temperatures of heated chains adjusted automatically during the run. All these procedures were run 2-3 times for each taxon and for each model.

Phylogenetic analyses and estimation of divergence times

Phylogenetic relationships within and between taxa were analyzed with the neighbor-joining (NJ; Saitou & Nei 1987), maximum parsimony (MP; Fitch 1971), maximum likelihood (ML; Felsenstein 1981), and Bayesian (BA; Rannala & Yang 1996) methods. In MP analyses, an ACCTRAN optimization option was used and gaps were treated as a fifth base. A heuristic search with the tree bisection reconnection (TBR) and nearest neighbor interchanges (NNI) branch swapping methods were applied in the MP and ML analysis, respectively. The robustness of inferences was assessed either by bootstrap resampling (Felsenstein 1985) with 1000 replicates (NJ, MP) or with Bayesian posterior probabilities estimated from 1000000 generations sampled every 1000th generation and excluding the first 250000 steps as burn-in. For NJ and ML analyses the most appropriate models and parameter estimates were chosen with ModelTest (Posada & Crandall 1988) using the likelihood ratio test (LRT) and Akaike criterion (Akaike 1974) for hypothesis testing. When the two criteria preferred different models the simpler one was chosen. In BA, the parameters were estimated directly from the data. All phylogenetic analyses were performed with PAUP* except the Bayesian analysis which was carried out using MrBayes v. 3.1.2 (Ronquist & Huelsenbeck 2003). In *domesticus*, *castaneus*, and *musculus*, both minimum spanning network (MSN) and median-joining network (MJN) were constructed with Arlequin and Network v. 4.2.0.1 (Bandelt et al. 1999), respectively.

Potential differences in the substitution rates between taxa were tested in the whole set of haplotypes using the relative rate test (RRT). Since the tree was highly unbalanced due to overrepresentation of *domesticus* sequences, the weighting scheme of Robinson et al. (1998) using an improved version of Wu & Li's (1985) test which takes into account taxonomic sampling and phylogenetic relationships, was applied. For this purpose, the RRTree v. 1.1 (Robinson-Rechavi & Huchon 2000) program was employed with the NJ tree chosen as the reference topology using *M. macedonicus*, *M. cypriacus*, and *M. spicilegus* sequences as outgroups. The same test was carried out also for each subspecies. In this case, individuals within each taxon were arbitrarily divided into two, more-or-less equally large, groups comprising rapidly and slowly evolving sequences, respectively. Since no significant substitution rate heterogeneity was detected between compared subspecies or among

haplotypes within clades a strict molecular clock model was assumed in all analyses of divergence times (see below).

Approximate times of divergence between the subspecies were estimated in two ways. First, mean Tamura-Nei genetic distances with invariant sites and unequal substitution rates (TrN + I + Γ) between the taxa were computed and corrected for ancestral mtDNA polymorphism according to Edwards (1997): $D = D_{AB} - 1/2\ (D_A + D_B)$, where D_{AB} is the mean genetic distance between lineages A and B, computed from pairwise distances between individuals from different lineages (i.e., A *vs.* B), and D_A and D_B are mean genetic distances within these lineages, respectively. As a reference we used the mean distance between all *Mus* sequences under study and five *Rattus norvegicus* sequences listed above and calibrated with 12 Myr, i.e., the most commonly selected mouse-rat calibration point (Jaeger et al. 1986, Jacobs & Downs 1994, Jacobs & Flynn 2005, Benton & Donoghue 2007).

The second method was based on a Bayesian coalescent analysis under a molecular clock assumption and the HKY + I + Γ model using the Beast v. 1.4.7 (Drummond & Rambaud 2006) program. Before estimation of the time of the most recent ancestor (TMRCA) of all the house mouse subspecies studied we estimated the substitution rate under the exponential model. Because of the absence of any reliable paleontological calibration point for house mice we added 66 sequences of aboriginal species (33 *M. macedonicus*, 18 *M. spicilegus*, 15 *M. cypriacus*) studied by Macholán et al. (2007b) and five *Rattus* GenBank sequences (see above) and used a strong uniform prior for TMRCA bounded between 11 and 12.3 Myr as recommended by Benton & Donoghue (2007) for the *Mus-Rattus* split. The transition/transversion rate ratio κ was assigned the gamma prior G (6, 2) while the shape parameter α had the gamma prior G (1, 1) according to Rannala & Yang (2007). In addition, we also used gamma priors G (2, 1) and uninformative priors drawn from the uniform distribution U (0, 10) for κ and α, and U (0, 1) for the proportion of invariable sites (I). The resulting overall substitution rate estimate, 0.0444×10^{-6}/site/yr (lower and upper bounds of the 95 % highest posterior density [HPD]: 0.0292-0.0617 $\times$ 10^{-6}/site/yr), corresponds to the rate estimated for the three eastern aboriginal species by Macholán et al. (2007b): 0.0451×10^{-6}/site/yr (HPD bounds: 0.0162-0.0783 $\times$ 10^{-6}). The former estimate was then used for all subsequent coalescent analyses. The following priors were used: G (3, 1), U (0, 10) for κ; G (2, 1), U (0, 10) for α, G (1, 1), U (0, 1) for I, U (0, 2) for between-subspecies TMRCA, and (0, 1) for within-subspecies TMRCA. Using the gamma prior for estimation TMRCA is in agreement with conclusions of Kimura (1970), who showed that the overall distribution of time to coalescence is close to this type of distribution. Similarly, the gamma distribution has been used for estimation of other parameters such as recurrent gene flow and range expansion (e.g., Templeton 2005). However, it should be noted that the choice of priors (gamma *vs*. uniform) had no effect on resulting estimates. The parameters were estimated from a range of plausible alternative trees after 2-3 independent MCMC chains of 20 million generations each sampled every 10000th generation and with the first 10 % trees discarded as burn-in. Both constant and exponential coalescent models were assumed for each TMRCA estimate.

Beast was also applied for estimating effective population size against time using the Bayesian skyline plot (BSP) procedure (Drummond et al. 2005). Conditions for MCMC searches were the same as described above and the number of discrete intervals was set to 30 (*domesticus*), and 20 (*castaneus*, *musculus*), respectively.

Results

Genetic variation

Among 82 new sequences, 69 distinct haplotypes were identified: 37 *M. m. domesticus*, 13 *M. m. castaneus*, 9 *M. m. musculus*, and 10 sequences from the island of Socotra which could be ascribed to neither of the house mouse subspecies with known CR sequence. The group of these haplotypes is hereafter referred to as "SocotraX". Interestingly, eight *domesticus* and five *castaneus* haplotypes were found together with these sequences at the same locality (Hadiboh, Socotra; Table 1). Fig. 2 shows variation among consensus sequences of four house mouse subspecies, SocotraX and three aboriginal mouse species. The consensus sequences of the latter species were constructed from the haplotypes published in Macholán et al. (2007b). Among the variable sites, 12 characterize the *M. musculus* complex (we avoid the term synapomorphies here as some of the sites may be polymorphic within the taxa) and 17 are singletons found only in SocotraX. Twenty-five polymorphic sites were identified in 10 SocotraX haplotypes (Fig. 3).

In five *M. m. musculus* sequences from Romania and Bulgaria, a 75-bp direct repeat was found (haplotypes *mus*RO1-4 and *mus*BG1; see Table 1); this repeat was almost identical in sequence and position with

```
                  1111111111111111111111111111111111111111111111111111111111111111
                  5555555555555555555555555555555555555555555555555555555555555555
                  3333333334444444444445555555555555555555555555555555555555555555
                  3366667899014457777999000001111111222333344444444444467777777
                  2336797156281200157579234572345789789034634567891235623456 89
                      *    ↓*       ↓*↓*↓         *                  ↓   ↓  ↓ ↓
SocotraX          TCAAC-AGTGAAATCATTAACCCAAATGTACGGACA--TACAACTGACCCTCACAC-AGC
musculus          .A......CA.C..T.C.T.T.TT..CA..T..C.............TTAA..T..T.-T
castaneus         .........A......C.T.T.T...CA..T..T.............TTA...T..C.AT
domesticus        ..G......A...CT.C.T.T.T...C..T...TT...C....TCA.TTA.T.TCTT.A.
gentilulus        .....T...A......C.T.T.T..G..C....T.............T.AC..T..T.A.
macedonicus       ...-T.G..AG...TG..CCTAT.T..A..TAC.ACTAAC..CA.AT.TAC..T..T.AT
cypriacus         .A..T..A.AG.T.T..ATCTAT..G.AC.T.C..T..ATTT.A.ACT.ACAGT.TT.A.
spicilegus        A...T....AG.......CGTAT...CA..TAC.....AC......T.TTC..T..CTAT


1111111111111111111111111111111111111111111111111111111111111111111111111111110
5555555555555566666666666666666666666666666666666666666666666666666666666666660
5555666667799900000000000000000011111111111111112222222222222222222222222222220
8899003472525700012555666667899900234444455668990123344444555666667788999991
4846032364258057903678134691849908690124814020457823404568012456891579123 59
                      aab
               ↓ * * ↓   ↓↓  *↓ *          ↓        ↓     ↓     ↓   *  *  *
ATTCACCCTATCGTTGAACGATTTGTCATA--TATAGACTTCTCAAACCAAGCCCTAGGAAGCTCATTCTAACAC
C.CT...T....ACC...TTGCCCA...C.....C......T.T.GC...TA....C....ATC....T......
...TG..T....ACC...TTGCCCA.T.C....GC......T.T..C...TA....C...GATC...........
TC.AGT.T....ACC...TTGCCCA...C-....C.....AT.T..C.T.TA..T.C....ATC...........
C..T...T...?.CC.T.TTR.CCAC..C-....C.....CT.T......T.T...C....ATC..........T
T.CT..A..T.TACCT.T.-...CA..GC-A....G---C.T.T...ATTT..T.CCAA..AT...C.TAGGTG-
T.CT..AT.TATACCT.T.-...CA..GC-A...C......T.TT..ATTT..A.CCAA.GAT.T.C..A.GT.-
T.......CT...CCT.T.-...CA.TTC-AAC........TCT......T..A..CAAGGAT..TCC.A..T.-
```

Fig. 2. *Variable sites among consensus mtDNA sequences of the "SocotraX" group of haplotypes, the* Mus musculus *complex (*musculus*,* castaneus*,* domesticus*,* gentilulus*), and aboriginal species* M. macedonicus*,* M. cypriacus*, and* M. spicilegus*. Lower-case letters denote sites added 3′ of the indicated nucleotide relative to the standard mtDNA sequence of Bibb et al. (1981); indels are marked with "-", "?" denote unknown bases. Character states common to* M. musculus *consensus sequences are indicated with asterisks while singletons of the "SocotraX" group consensus are indicated with arrows.*

```
                  1111111111111111111110000
                  5555555555555556666660000
                  3333333445799900000010011
                  1124456139685790225859955
                  4851263370889765370794501
                      aa            a
numtSO1           GAAC--ACAAATGGACCGC-CGCGC
numtSO2           ...........G...........TT
numtSO3           .......................?
numtSO4           ...........G...........?
numtSO5           .....T.....?...T........T
numtSO6           ...........GA..TT..T....T
numtSO7           C....T....TG.C.T...T....T
numtSO8           .CTTT.GA.T.G..TT...TT..??
numtSO9           .??..G-.T..G.C.T.T.T...?T
numtSO10          ..T..G.....G?.TT.TTTTCG??
```

Fig. 3. *Variation at 25 polymorphic sites among 10 haplotypes of the "SocotraX" group shown in the format of Fig. 2.*

the repeat reported by Prager et al. (1996) from Austria, Serbia, Moldova, and Georgia, by Boissinot & Boursot (1997) from Russia, and Iran, and with that found by Liu et al. (2008) in the inbred strain BLG2 derived from *M. m. musculus* from Bulgaria. This segment was excluded from further analyses because of problems with alignment. Another direct repeat, 11-bp long, was found in several *M. m. domesticus* haplotypes: the complete repeat TTTTAACTCTC, identical in both position and sequence motif with that reported from various European localities by Prager et al. (1993, 1996), was found in five haplotypes from Finland (*dom*FL1-5; Table 1) whereas a partial 5-bp repeat CTCTC was found in mice from Iran (*dom*IR1, 2, 4, 5, 9), Georgia (*dom*GE1), Cyprus (*dom*CY1, 2), and Socotra (*dom*SO1, 6, 7; Table 1). Interestingly, two *musculus* mtDNA haplotypes (*mus*FL1-2) were

found at a single locality Ulvila in Finland whereas all other Finnish haplotypes under study were of *M. m. domesticus*.

In order to assess the level and pattern of genetic variation within and among the taxa, the new data were combined with sequences published in GenBank. Significantly unequal base composition was found within each subspecies (Chi-square, $P < 0.05$) with the highest frequency of A (35.43 %) and the lowest of G (11.94 %); no significant differences in nucleotide frequencies were found among the taxa (Chi-square, $P >> 0.05$). There is prevalence of transitions in the studied mtDNA segments with overall transition/transversion rate ratio $\kappa = 3.665$, the only exception being the SocotraX sequences with balanced frequencies of transitions and transversions ($\kappa = 1.054$). No saturation was revealed in the whole data set (including *Rattus*: $S = 0.860$, $R^2 = 0.989$; excluding *Rattus*: $S = 0.953$, $R^2 = 0.991$).

The level of genetic variation in the three subspecies (*domesticus*, *castaneus*, *musculus*) and the SocotraX group is shown in Table 2. Interestingly, *M. m. domesticus* revealed the lowest value of haplotype diversity while *M. m. castaneus* appeared to be the most variable. This subspecies has haplotype diversity significantly higher than in *M. m. domesticus* and *M. m. musculus*, with the highest values of nucleotide diversity and mean genetic distance between populations reaching 140 % and 135 % of that in *domesticus* and *musculus* populations, respectively. Rather low values of haplotype diversity in *M. m. domesticus* can be explained by an unbalanced sampling design where the vast majority of samples have been collected from geographically limited areas in southern Bavaria (Germany) and Scandinavia and East Holstein (Germany) as a part of the *musculus*/*domesticus* hybrid zone studies (Tucker et al. 1992, Prager et al. 1993, 1996). This sampling bias is reflected by low haplotype diversities in the *domesticus* populations from northern and southern Europe (0.7177 and 0.7395, respectively). When only distinct haplotypes are included in the analysis, the genetic distances within the "southern-European" group is highest among the *M. m. domesticus* samples (Table

Table 2. *Genetic variation within* Mus musculus *subspecies and groups of populations. The* gentilulus *sequences were not analyzed due to small sample size and high number of unknown bases. "N Europe": Finland, Denmark, Sweden, Norway, northern Germany, England, and Scotland; "S Europe": Italy, Greece, Serbia, Switzerland, Austria, Spain, and Bavaria (S Germany); "Near East": Israel, Egypt, Jordan, Syria, Cyprus, southern Turkey; "Iran": Iran, and Georgia.*

Taxon and population groupings	No. of individuals	No. of haplotypes	Genetic divergence within subspecies and population groups (% TrN dist.)	Haplotype diversity $h \pm$ SD	Nucleotide diversity $\pi \pm$ SD
domesticus	397	114	1.377	0.9012 ± 0.0114	0.0099 ± 0.0050
castaneus	34	32	1.927	0.9947 ± 0.0096	0.0148 ± 0.0076
musculus	142	46	1.430	0.9532 ± 0.0075	0.0078 ± 0.0041
SocotraX	11	10	0.911	0.9818 ± 0.0463	0.0056 ± 0.0033
N Europe	216	30	0.956	0.7177 ± 0.0314	0.0109 ± 0.0055
S Europe	96	17	1.430	0.7395 ± 0.0321	0.0206 ± 0.0102
Near East	25	23	1.131	0.9933 ± 0.0134	0.0257 ± 0.0131
Iran	20	17	1.078	0.9798 ± 0.0245	0.0297 ± 0.0153

Table 3. *Results of hierarchical AMOVA (Excoffier 2007) of* M. m. domesticus *haplotypes. Populations were grouped into four regions: "northern Europe" (Sweden, Norway, Finland, Denmark, northern Germany, and Great Britain), "southern Europe" (Spain, Portugal, southern Germany, Austria, Switzerland, Italy, Serbia, Croatia, and Greece), "Near East" (Israel, Jordan, Syria, Turkey, and Cyprus), and "Iran";* P *[rand ≥ obs] after 10000 permutations.*

Source of variation		d.f.	% total variation	Φ-statistic	*P*
Among regions	σ_a^2	1	5.41	$\Phi_{CT} = 0.054$	< 0.0001
Among populations within regions	σ_b^2	2	50.53	$\Phi_{SC} = 0.534$	< 0.0001
Within populations	σ_c^2	353	44.06	$\Phi_{ST} = 0.559$	0.6653

2) whereas distances within the "northern-European" group are still very low (~ 67 % of S Europe; $P < 0.001$) and comparable with the genetic divergence among 10 SocotraX haplotypes. Thus genetic variation in the "northern-European" group (Sweden, Norway, Finland, Denmark, northern Germany, and Great Britain) is significantly lower than that in the "south-European" group of populations (Spain, Portugal, southern Germany, Austria, Switzerland, Italy, Serbia, Croatia, and Greece). Hierarchical AMOVA revealed significant differentiation among the four population groups defined in Table 2 although it explains only 5.41 % of the total variation ($P < 0.0001$; Table 3). Variation among populations within these groups explains 50.53 % of the total variation ($P < 0.0001$) while genetic differentiation within populations (44.06 % of total variation) was not significant ($P > 0.05$).

The balanced frequencies of transitions and transversions in the SocotraX sample strongly suggests that these sequences do not represent parts of mitochondrial CR. Rather they are likely to be mtDNA fragments that has inserted into the nucleus or numts (pronounced "new-mights," Bensasson et al. 2001). To test this hypothesis, the whole mitochondrial genome of one of the individuals possessing a suspected numt (haplotype *numt*SO2) was sequenced. Indeed, the CR haplotype ascertained from the sequenced genome appeared to differ from the original *numt*SO2 haplotype being similar to *M. m. domesticus* haplotypes. Since all 10 SocotraX haplotypes form a monophyletic group (see below) also remaining nine original sequences that were not re-sequenced are very likely to be numts.

Both Tajima's and Fu's tests revealed significant deviations from distributions expected under the hypothesis of selective neutrality and/or population equilibrium at the 5 % significance level in all subspecies. The results thus indicate either selection acting on the sequences analyzed, rate heterogeneity, a bottleneck effect, or population growth. Recent population expansion is suggested by the mismatch distribution of differences between pairs of haplotypes (Fig. 4) as the fit of the observed data with theoretical predictions could not be rejected at the 0.05 level for all three subspecies using both SSD (Pr [sim. SSD ≥ obs. SSD] > 0.05) and Harpending's raggedness index (Pr [sim. raggedness ≥ obs. raggedness] > 0.05). Exponential population growth is suggested also by large values of the exponential growth rate parameter *g* for all three subspecies estimated with the Lamarc program (Table 4).

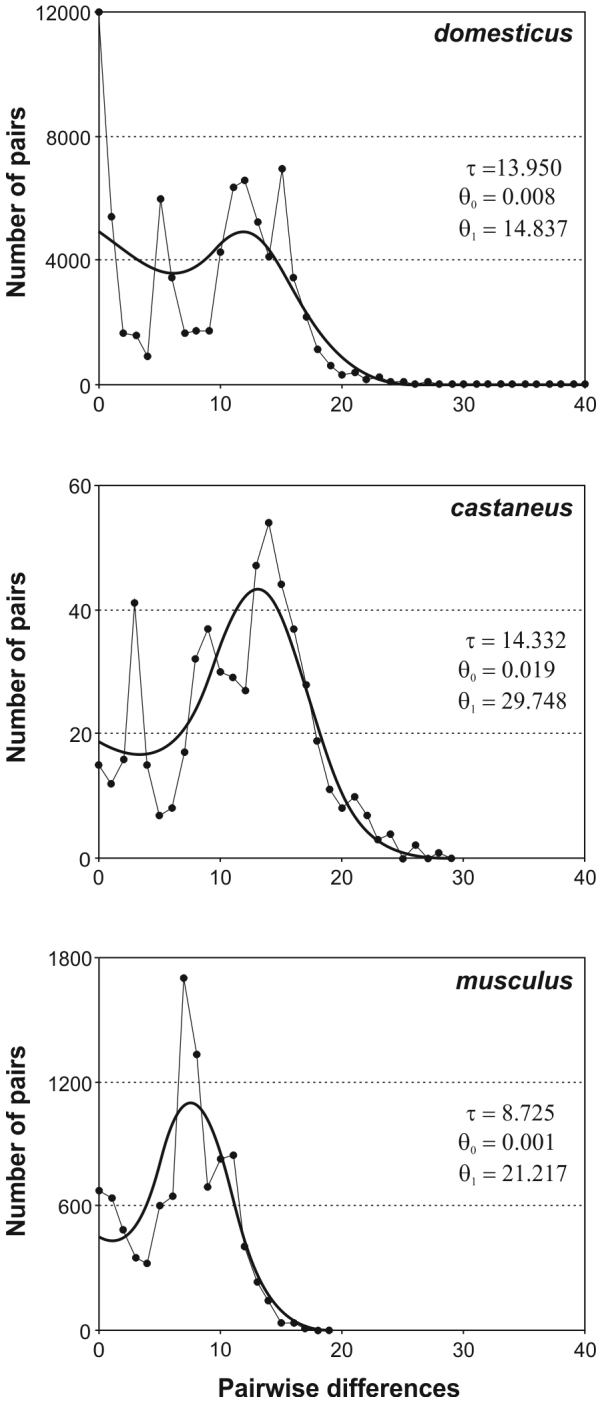

Fig. 4. *Mismatch distribution of mtDNA haplotypes in* M. m. domesticus*,* M. m. musculus*, and* M. m. castaneus*. The observed frequencies (black dots connected with thin lines) are compared to the expected frequencies (thick lines), based on the population expansion function with parameters estimated using a generalized nonlinear least-squares approach. Approximate times of population expansion τ (in 1/2 u units, where u is the mutation rate for the whole sequence) and population sizes before the expansion (θ_0) and at present (θ_1) are given for each group.*

Table 4. *Comparison of likelihoods for the stable population and exponential growth models yielded by the coalescence analysis of* M. m. castaneus*,* M. m. domesticus*, and* M. m. musculus*;* LL *= log-likelihood,* θ *= effective population size,* g *= exponential growth rate. High values of* g*, with confidence intervals not including zero, suggest exponential population growth.*

	castaneus		*domesticus*		*musculus*	
	stable pop.	exp. growth	stable pop.	exp. growth	stable pop.	exp. growth
LL	1.98×10^{-6}	0.0064	0.2855	2.6653	0.4134	0.4690
θ	0.059 (0.039-0.093)	0.206 (0.088-0.755)	0.064 (0.055-0.075)	0.109 (0.084-0.133)	0.021 (0.016-0.029)	0.026 (0.017-0.035)
g	--	464 (234-929)	--	255 (117-406)	--	315 (11-700)

Phylogenetic relationships among haplotypes and coalescence times

Phylogenetic relationships among the haplotypes (including numts) under study are depicted in Fig. 5. The basic branching pattern is the same irrespective of the phylogenetic method used (MP, NJ, ML, BA) even though the support for the relationships within the *musculus-domesticus-castaneus* group is very low (as evidenced also by a high level of homoplasy: homoplasy index *HI* = 0.433, consistency index *CI* = 0.567, retention index *RI* = 0.871, rescaled consistency index *RC* = 0.494). This pattern remains unchanged when all previously published haplotypes (Prager et al. 1993, 1996, 1998, and references therein) are included. The numt clade is basal to and quite divergent from the clades of all other house mouse subspecies: the average TrN distances of the numt sequences to *domesticus*, *castaneus*, *musculus*, and *gentilulus* are 9.61 % (± 0.04 % *SE*), 7.78 % (± 0.10 % *SE*), 7.40 % (± 0.08 % *SE*), and 7.63 % (± 0.16 % *SE*), respectively.

In all trees *M. m. gentilulus* appears as a sister group to the (*domesticus*(*musculus*, *castaneus*)) clade. This is important since the phylogenetic position of this subspecies has been central to two competing hypotheses of the house mouse evolution: the centrifugal model (Boursot et al. 1996, Duplantier et al. 2002) and the sequential model (Prager et al. 1998). The reason for the different position of *gentilulus* between Duplantier et al. (2002) on the one hand and Prager et al. (1998) and this study may be a high number of unknown character states in

Table 5. *Times of the most recent common ancestor (TMRCA) between species and subspecies of the mice studied. Nodes are labeled according to Fig. 6. The dates are in thousands of years; in parentheses, the lower and upper bounds of the 95 % HPD interval. TMRCA 1 = times derived from average TrN distances assuming the* Mus-Rattus *split ≈ 12 MYA; TMRCA 2–3 = times estimated from coalescence trees assuming exponential population growth and stable population, respectively; in both the latter cases, mutation rate was estimated for the whole data set at* 0.044×10^{-6} *substitutions/site/year; TMRCA 4 is based on coalescence assuming exponential growth and mutation rate* 0.500×10^{-6} *substitutions/site/year. In all the cases, a strict molecular clock was assumed.*

Node	TMRCA 1	TMRCA 2	TMRCA 3	TMRCA 4
1		124 (75-177)	193 (117-275)	12 (8-19)
2		208 (144-279)	294 (199-391)	22 (15-30)
3		219 (167-278)	309 (232-397)	20 (15-26)
4	371	340 (250-439)	486 (356-631)	
5	729[a]	493 (385-603)	774 (599-980)	
6	813[a]	548 (418-684)	808 (608-1006)	
7	1878	1318 (1057-1582)	1661 (1349-1993)	
A	382	301 (200-404)[b]		
B	864	491 (344-649)[b]		
C	1147	773 (564-1006)[c]		

[a] *the distance computed as unweighed arithmetic mean of distances between pairs of taxa;* [b] *Macholán et al. (2007b);* [c] *recalculated from the data of Macholán et al. (2007b) assuming* M. macedonicus *and* M. cypriacus *as a monophyletic group.*

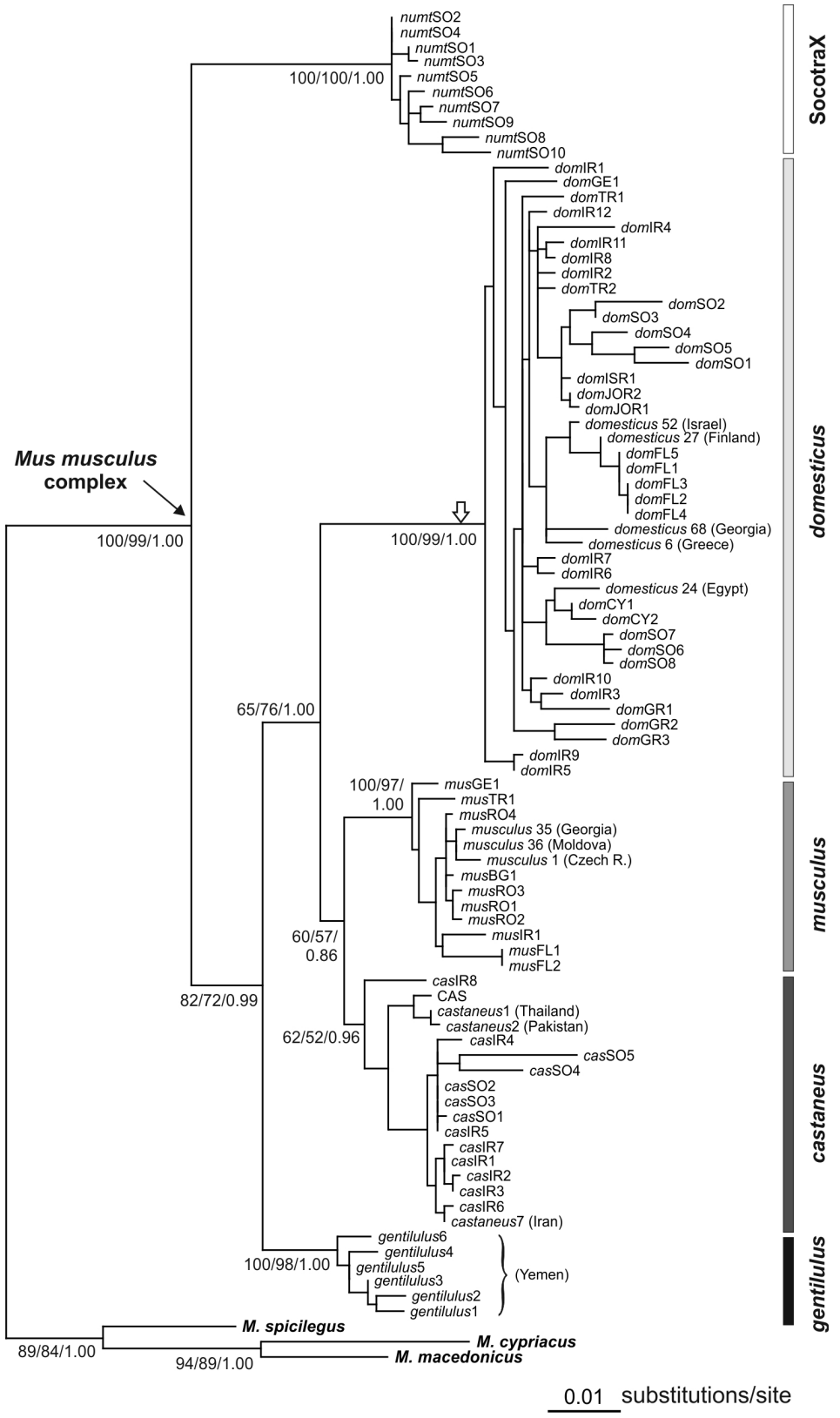

Fig. 5. *Maximum likelihood tree of four subspecies of the* Mus musculus *complex and a group of numts denoted as "SocotraX". The tree is rooted with the group of three aboriginal species,* M. macedonicus*,* M. cypriacus*, and* M. spicilegus*. NJ and MP bootstrap values and Bayesian posterior probabilities are given for each major clade. The position of one Socotra sample for which the whole mtDNA was sequenced is indicated with arrow.*

gentilulus sequences (≈17.5 % of all bases). Although this is unlikely given the high support for the topology (Fig. 5) we tested the potential influence of unknown bases on the phylogenetic position of *M. m. gentilulus* as follows: first, we retrieved from GenBank seven *gentilulus* D-loop sequences from Madagascar (MDG1-3, MDG5-8) published by Duplantier et al. (2002) and then from these seven sequences and six sequences from Yemen analyzed in this study we made a single consensus sequence. In this way the number of unknown character states was decreased to 3.8 %. Again, the basic phylogenetic relationships within the house mouse complex remained the same as in Fig. 5.

Presence of *domesticus* and *castaneus* haplotypes on Socotra raises a question of their geographic origins. The tree in Fig. 5 suggests the origin of the Socotran *domesticus* haplotypes in the Near East: three haplotypes revealed close similarity to those from Egypt (*domesticus* 24) and Cyprus (*dom*CY1, 2) whereas other four appeared as a sister group of *dom*ISR1 from Kursi Beach in Israel and *dom*JOR1, 2 from Al-Aqabah in Jordan. The minimum spanning network (MSN) confirmed the affinity of the former three haplotypes, however, the latter four were connected with a group of haplotypes from Germany (not shown). This MSN had 31 alternative connections but none of those connections changed the position of the Socotran *domesticus* haplotypes nor the position of outgroups. Similar results were revealed in the median-joining network (MJN), however, the number of alternative connections was too high in this case (not shown). Thus we should consider at least two colonization events on Socotra for *M. m. domesticus*. Nevertheless, the real number of independent introductions is likely to be higher since the haplotype *dom*SO9 detected in the whole mtDNA sequence appeared as a sister group to the whole *domesticus* clade, with no close relationship to other Socotran *domesticus* haplotypes, irrespective of phylogenetic method used (marked with arrow in Fig. 5).

Phylogenetic relationships within *M. m. domesticus* are unclear. Even though geographically close haplotypes tend to group together (Fig. 5) the trees inferred by individual phylogenetic methods differ markedly, which is also reflected by the extremely low bootstrap values. Moreover, genetic distances between haplotype groups from different parts of the subspecies' ranges (e.g., southern Bavaria *vs*. Israel and Egypt) are much lower than distances among haplotypes from Iran. The only consistent pattern emerging from the various types of analyses seems to be the position of the Middle Eastern and Balkan haplotypes (Israel, Turkey, Iran, and Greece) close to the presumable root of the *M. m. domesticus* tree.

Both the ML phylogenetic tree (Fig. 5) and MSN (not shown) revealed the Socotran *castaneus* haplotypes to be embedded within the clade of haplotypes from Iran. There are five alternative connections in the network yet these connections change neither the relationships of *M. m. castaneus* from Socotra to other haplotypes nor the position of outgroups. This pattern is corroborated by both the MJN and NJ and MP trees (not shown) comprising all GenBank sequences of Prager et al. (1998). All the trees and networks are consistent in the presence of three clades: "eastern", comprising haplotypes from China, Taiwan, and Thailand, together with one haplotype from Pakistan (*castaneus* 2; Prager et al. 1998) and two "western" clades, one Iranian (+ Socotra and *castaneus* 11 from Pakistan) and the other comprising haplotypes from Iran, Afghanistan, and Pakistan. These clades are all very poorly supported (MP bootstrap values ranging from 6 % to 22 %, NJ bootstrap values: 22-67 %) as is also the support for monophyly of the whole *M. m. castaneus* clade (bootstrap 25 % and 60 % for MP and NJ, respectively; trees not shown). What seems to be clear, however, is the position of haplotypes from Iran, Afghanistan and Pakistan close to the root.

The genealogical relationships among *M. m. musculus* haplotypes are similar to those in *M. m. domesticus*: there are several distinct and rather well supported clades in MP and NJ trees as well as in MSN and MJN networks (e.g., Romania, Afghanistan, Czech Republic-Poland, and Germany-Austria-Croatia) yet the relationships among them and to other haplotypes are unresolved and the branching pattern within the whole clade is rather random. Monophyly of *M. m. musculus* is supported with 73 % and 84 % based on 1000 MP and NJ pseudoreplicates, respectively (trees and networks not shown).

Divergence times based on TrN distances and calibrated with the *Mus-Rattus* split 12 Mya and coalescent TMRCA estimates are given in Table 5 for all nodes numbered in Fig. 6. As expected, estimates based on the exponential growth model (TMRCA 2) are lower than those assuming constant population size (TMRCA 3) since as we go back in time, the effective population size decreases and hence the rate of coalescence increases (Felsenstein 2004, Nordborg 2007). Moreover, we should expect genetic distance-based (TMRCA 1) and constant-size coalescent estimates (TMRCA 3) to be similar as these are based on the same assumption. However,

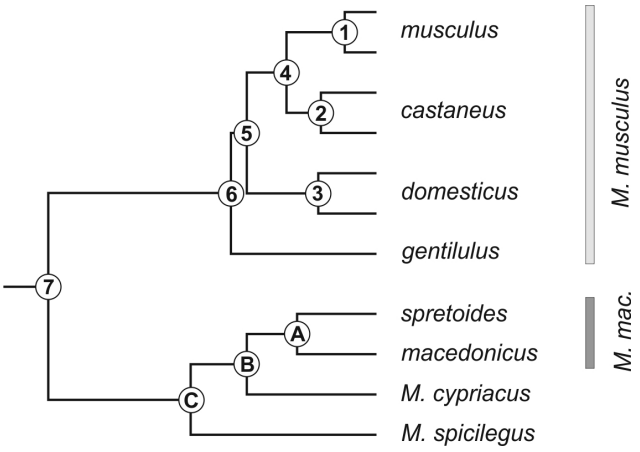

Fig. 6. *Schematic representation of phylogenetic relationships among all the mouse species and subspecies under study;* macedonicus *and* spretoides *are subspecies of* Mus macedonicus *(*M. mac.*). The numbers and letters at each node are referred to in Table 5 and Fig. 7.*

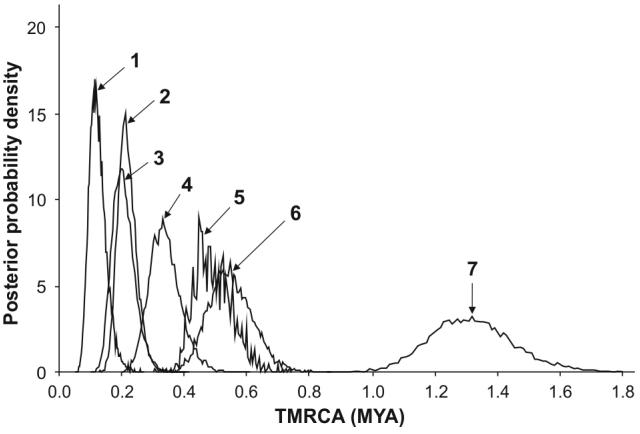

Fig. 7. *Posterior probability distributions (scaled by the mutation rate) estimated for the time of the most recent common ancestor (TMRCA). The numbers correspond to nodes on the tree in Fig. 6.*

the former estimates are considerably higher for the three youngest and two oldest splits (Table 5, Fig. 6). This phenomenon cannot be explained by non-linear relationship between TrN distances and time since Tamura-Nei distance is corrected for saturation of sequences (i.e., is "linearized") nor can this be explained by non-linearity of the coalescence process as this should affect both TMRCA 1 and TMRCA 3 estimates.

From Table 5 (see also Fig. 7) we can conclude that for the mtDNA control region, the most recent common ancestor of the *Mus musculus* complex occurred at ~660 kya, with an upper estimate about 1 Mya under the constant population size assumption. Similarly,

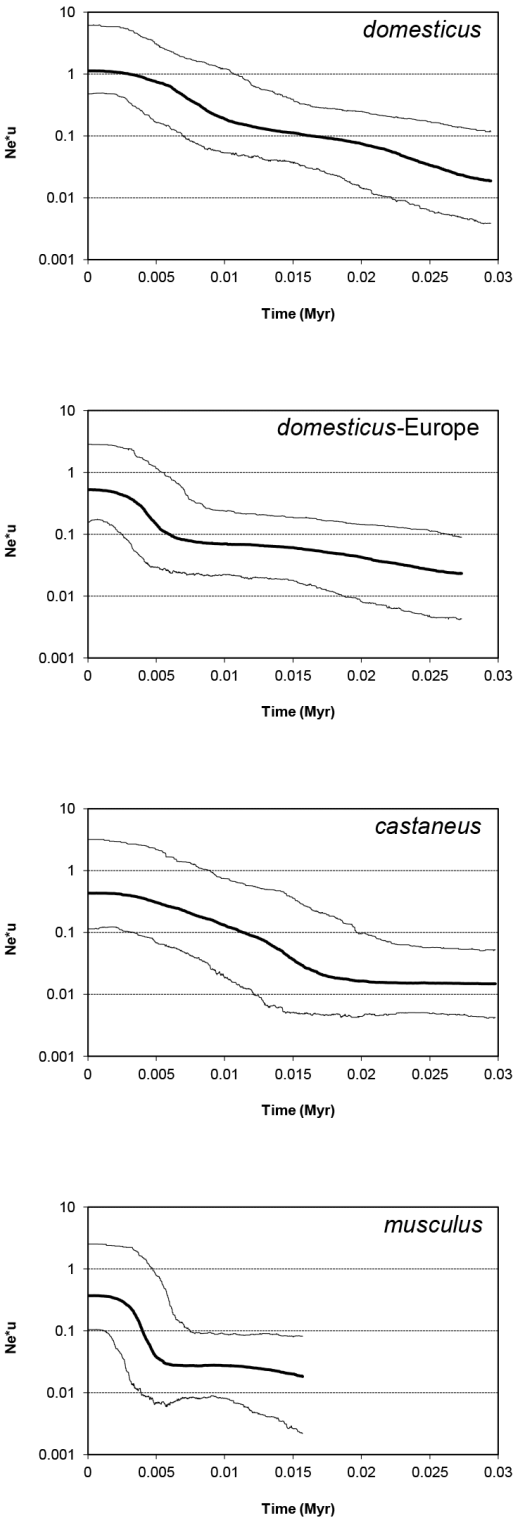

Fig. 8. *Demographic history of* M. m. domesticus*,* M. m. castaneus*, and* M. m. musculus *population depicted as* $N_e * \mu$ *(effective population size*mutation rate) plotted against time (in million years) as estimated with the Bayesian skyline plot model. The thick middle line is the median estimate whereas the thin line indicates upper and lower limits of 95 % highest posterior density (HPD).*

M. m. domesticus and *M. m. musculus* mitochondrial lineages coalesced at ~500 kya, bounded between ~390 kya (lower bound of the 95 % HDP under the exponential growth assumption) and 980 kya (upper bound of the 95 % HDP under the constant population size assumption).

A comparison of TMRCA's for individual house mouse subspecies shows that *M. m. domesticus* and *M. m. castaneus* started to diverge at approximately the same time while the coalescence of *M. m. musculus* mtDNA sequences is much younger, the tree height being only ~60 % of those for the former two taxa.

Demographic analysis

If we assume that the populations grow exponentially we can estimate expansion time, τ, from the mismatch distribution analysis, expressed as $1/2u$ generations (Li 1977, Rogers & Harpending 1992). In addition, we can derive mutation parameters θ_0 and θ_1, given as $\theta_0 = 2uN_0$ and $\theta_1 = 2uN_1$, respectively (see Fig. 4), where u is the mutation rate per *the whole* sequence and N_0 and N_1 is the size of stationary haploid population at the time before expansion and at present, respectively (recall that in the case of mtDNA these parameters equal the effective numbers of females). If we substitute $m_T\mu$ for u, where μ is the substitution rate per site per generation and $m_T \approx 1000$ bp is the number of sites, and assuming two generations per year (Pelikán 1981, Macholán et al. 2007a, b), we can approximate times when individual groups started to grow exponentially according to the formula $t_e = \tau/2\, m_T\mu$ (Li 1977, Rogers & Harpending 1992). This yields the estimates 314, 333, and 197 kya for *domesticus*, *castaneus*, and *musculus*, respectively. These estimates are rather high, roughly corresponding to the coalescence times under the assumption of constant population size (Table 5) and considerably predating a presumed time of colonization of Europe (i.e., not earlier than 12 kya). Therefore, we performed several MCMC analyses using Beast as described above but without using a strong uniform prior for a coalescent tree root height. This resulted in a wide range of substitution rates without any sign of convergence except of a small peak of posterior probabilities between 40 % and 60 % per site per million years, indicating about 10-fold short-term mutation rate compared to the long-term substitution rate as suggested by Ho et al. (2005, 2007). Thus we used 50 % as an approximate mutation rate for subspecific estimates of coalescence times as well as times of population expansions. The former estimates are listed as TMRCA 4 in Table 5. The mismatch distribution yields the expansion times 27.9, 28.6, and 17.5 kya for *domesticus*, *castaneus*, and *musculus*, respectively. If we assume only one generation per year (Raufaste et al. 2005, Rajabi Maham et al. 2008) we get 14.0, 14.3, and 8.7 kya, respectively.

Demographic histories of the four subspecies are depicted also as the Bayesian skyline plots (BSP) in Fig. 8. We can see in the plots that while *M. m. castaneus* has undergone a single population expansion starting around 17 kya, there were at least two expansion events in *M. m. domesticus*: one between 30 and 20 kya and the second one between 10 and 5 kya. The latter event may have been, at least partly, connected with the expansion of the subspecies across Europe. Indeed, BSP of European *domesticus* haplotypes (second panel in Fig. 8) shows a sudden population increase around 5000 years ago. A similar rapid and short population burst appeared during the same time in *M. m. musculus* population (fourth panel in Fig. 8).

Discussion

Patterns of genetic variation and genealogical relationships

Presence of *M. m. domesticus* haplotypes at Finnish sites Suitia and Parikkala (*dom*FL1-5; see Table 1), located within the *M. m. musculus* range (Fig. 1), is consistent with previous studies which have evidenced mice with the *domesticus* mitochondrial and *musculus* nuclear genome from Denmark, Sweden, and Finland (Ferris et al. 1983, Gyllensten & Wilson 1987, Vanlerberghe et al. 1988, Prager et al. 1993). Indeed, the individuals from the two Finnish sites analyzed here were shown to possess the *musculus* sex chromosomes (Laakonen et al. 2007). Moreover, all these haplotypes contained the same 11-bp direct repeat as that reported by Prager et al. (1993, 1996, 1998) from Denmark, Sweden, northern Germany, Finland, Switzerland, Spain, Italy, and Egypt. According to Gyllensten & Wilson (1987) and Prager et al. (1993), Scandinavia has been colonized by mice originating from the *M. m. musculus*/*M. m. domesticus* hybrid zone near the Kiel Bay in East Holstein, northern Germany. According to those authors, *musculus* mice bearing *domesticus* mtDNA arrived to Scandinavia across a chain of islands between northern Germany, Denmark and Sweden.

While the above scenario was supported by a comprehensive survey of Prager et al. (1993) the colonization of Finland is less clear. Theoretically, Finland may have been colonized by (i) *musculus*

mice from the southeast; (ii) *musculus*/*domesticus* hybrids from East Holstein through northern Poland, Lithuania, Latvia, Estonia, and Russia to Finland and then following to Sweden (Model IV of Prager et al. 1993); (iii) mice passing from southeastern Sweden along the Gulf of Bothnia first to northern Finland and then further southwards; or (iv) mice travelling through a chain of islands between eastern Sweden near Uppsala and the Finnish town of Turku. The first scenario is unlikely since colonizing mouse populations from Russia, i.e., from an area deep in the *musculus* territory, must have borne *domesticus* mtDNA which is implausible. Scenario (ii) hypothesizes thousands-kilometres-long migration from the northern-German hybrid zone area across pure-*musculus* populations to Finland, again an implausible assumption. Of the remaining two possibilities, scenario (iv) seems more parsimonious, however, the fact that Prager et al. (1993) found mtDNA with the 11-bp repeat also in central Finland and in Sweden close to northern Finland makes also (iii) reasonably possible. However, the colonization of Finland may have been more complex given the presence of two *musculus* haplotypes (*mus*FL1, 2; Table 1) in Ulvila from southwestern Finland (site 29; Fig. 1). These animals were shown by Laakonen et al. (2007) to bear *musculus* sex chromosomes.

The genetic analyses reported in this paper revealed a rather high level of mtDNA variation in all subspecies under study. *M. m. castaneus* appeared much more variable than *M. m. domesticus*, the latter being comparable with *M. m. musculus* (Table 2). The increased level of variability in *castaneus* is in agreement with results of Din et al. (1996) and Boursot et al. (1996), who found extensive nuclear and mitochondrial diversity in "central" populations from northern Indian subcontinent relative to peripheral subspecies and populations. Because of a complex genetic makeup and unclear systematic status of the mouse populations inhabiting a vast area from Pakistan to SE Asia these have been often dubbed the "oriental" lineage instead of using any taxonomically valid name (Boursot et al. 1996, Boissinot & Boursot 1997). Moreover, some results based on the sequence of the mitochondrial control region (Boissinot & Boursot 1997, Prager et al. 1998: Fig. 9B in their paper), and concatenated 21 nuclear loci have even suggested paraphyly of this group (Liu et al. 2008). Indeed, although all the *castaneus* haplotypes analyzed in this study were found to be monophyletic the bootstrap support was very low compared to other subspecific groups. Interestingly, both haplotype diversity (h) and nucleotide diversity (π) are of the same magnitude in the house mouse subspecies studied (Table 2) as those in three aboriginal species (Macholán et al. 2007b), contrary to results showing reduction of protein variability in *M. spretus*, *M. macedonicus*, and *M. spicilegus* (see Sage et al. 1993 for review).

Another important question is spatial distribution of genetic variation within the subspecies. In this paper we compared variability between four groups of *M. m. domesticus*: northern Europe, southern Europe, Near East, and Iran. Haplotype diversity was significantly lower in the two European groups of populations with northern populations being the least variable and this gradient was even more apparent in nucleotide diversity which was almost twofold in southern European populations compared to northern Europe (Table 2). Thus although delimitation of the groups is rather arbitrary there is a clear gradient of genetic variability from the Middle East to northern Europe. Although differences among the regions explain only 5 % of the total variation this differentiation was highly significant (Table 3). Similar gradient was reported for mtDNA by Sage et al. (1990) whereas Britton-Davidian (1990) and Rajabi Maham et al. (2008) have not found such a gradient using protein electrophoresis and D-loop sequencing, respectively. The decreasing variation is consistent with the hypothesis of colonization of Europe by *domesticus* mice from the Near East and Asia Minor after the last glacial maximum (Auffray et al. 1990).

The genetic structure within and between *M. m. domesticus* populations in Turkey and Iran was studied using D-loop sequences by Gündüz et al. (2000, 2005), who identified two distinct lineages one of which has colonized Europe. Recently, Rajabi Maham et al. (2008) have hypothesized, based on the D-loop sequence data from Iran, Turkey, and Europe, two groups colonizing Europe, provisionally dubbed "Mediterranean" and "Bosphorus-Black Sea". However, though the present study revealed several local groups of haplotypes, the relationships among these groups were unclear and no distinct lineages were identified. Absence of any internal structuring within *M. m. domesticus* suggests a strong homogenizing effect of long-distance, human-mediated, gene flow. According to population genetic theory, recurrent long-distance dispersal has a large evolutionary impact and measures of genetic differentiation among populations (e.g., F_{ST}) are extremely sensitive to even rare long-distance migration events (Templeton 2006).

Coalescence times and demography

An important question concerns the time of divergence among the house mouse subspecies. Prager et al. (1996, 1998), studying genetic variance and phylogeography of house mice based on mitochondrial control region sequences, did not address this question and rather used previously published estimates of the *musculus-domesticus* divergence times of 350 kya (She et al. 1990), 500 kya (Boursot et al. 1993, Suzuki et al. 2004), and 900 kya (Boursot et al. 1996) to root intrasubspecific divergences. Here we avoided using divergence times from other molecular data and rather used the mouse-rat split (12 Mya) based on paleontological record as a calibration point (Jaeger et al. 1986, Jacobs & Downs 1994, Jacobs & Flynn 2005, Benton & Donoghue 2007). For this purpose, five *R. norvegicus* sequences were added to the analysis. Mean TrN distances, corrected for ancestral polymorphism according to Edwards (1997; see Material and Methods), yielded ~370 and ~730 kya for the *musculus-castaneus* and *domesticus-*(*musculus*, *castaneus*) divergence times, respectively (see the TMRCA1 column in Table 5), i.e. higher than the estimate of She et al. (1990) and lower than that of Boursot et al. (1996).

The second method of estimating divergences between and within the taxa was based on a Bayesian coalescent analysis using the *Mus-Rattus* split, bounded between 11 and 12.3 Mya (Benton & Donoghue 2007), as a prior. First, the substitution rate was estimated for the whole *Mus musculus* sample supplied with 66 sequences of aboriginal species *M. macedonicus*, *M. spicilegus*, and *M. cypriacus* and assuming molecular clock and exponential growth of populations. The final estimate (≈4.5 % per site per million years) did not differ from the estimate for the three aboriginal species only (Macholán et al. 2007b) and was similar to that for the control region in laboratory mice (≈5.6 %; Goios et al. 2007). The 4.5 % value was used for subsequent coalescence estimations. This procedure yielded coalescences ~340 kya between *musculus* and *castaneus*, and ~490 kya between *domesticus* and the (*musculus*, *castaneus*) clade under the exponential model, and ~500 kya and ~770 kya, respectively, assuming the constant population model (see TMRCA2 and TMRCA3 in Table 5).

With respect to intrasubspecific coalescences, Prager et al. (1998), using 350 kya (She et al. 1990) and 900 kya (Boursot et al. 1996) for the *musculus-domesticus* split as described above, estimated 70-180, 100-280, and 170-460 kya for *musculus*, *domesticus*, and *castaneus*, respectively. If we assume 770 kya for the *musculus-domesticus* coalescence (Table 5) we get 150, 240, and 390 kya for *musculus*, *domesticus*, and *castaneus*, respectively, i.e. estimates slightly lower than those presented in this study for *musculus* and *domesticus* and higher for *castaneus* under the constant population assumption, i.e., the same assumption used by Prager et al. (1996, 1998) and others. The lower estimates by Prager et al. (1998) can be explained by analyzing more haplotypes in the present paper, however, the different number of haplotypes analyzed cannot be an explanation of the discrepancy in the *castaneus* coalescence estimates. Even though we did not include all the *castaneus* haplotypes published by Prager et al. (1996, 1998) because of too high number of unknown bases in some GenBank sequences we sequenced 13 new haplotypes so the total number of distinct haplotypes analyzed in this study was increased to 31, these sequences covering the whole range of the subspecies (including Socotra). According to Felsenstein (2004), the probability that adding a new haplotype to a set of k haplotypes would lead to a deeper coalescence equals $2/[k\,(k+1)]$, hence the probability that we would find an older root is 0.2 % for *M. m. castaneus*, 0.1 % for *M. m. musculus*, and 0.01 % for *M. m. domesticus*. Thus a more plausible explanation of the younger coalescence of *M. m. castaneus* haplotypes found by us relative to the estimate of Prager et al. (1998) is that we sequenced two segments encompassing variable domains of the mtDNA control region instead of the whole region.

Regardless of the differences in the coalescence times, the *musculus* tree is about 60 % as deep as the *domesticus* tree, a value close to that found by Prager et al. (1996, 1998) and consistent with lower *musculus* variability suggested by t haplotype diversity and/or mitochondrial and nuclear RFLP data (Figueroa et al. 1987, Klein et al.1987, 1988, Ruvinsky et al. 1991, Boursot et al. 1996).

All the estimates of intraspecific coalescences given above, based both on genetic distances calibrated with the paleontological record and the Bayesian coalescent analysis, are crucially dependent on the assumption that short-term mutation rate (e.g., within populations or subspecies) is the same as the long-term substitution rate (e.g., among species or higher taxa). For example, analyses of the mtDNA control region have yielded estimates of mutation rates as high as 32-290 % site^{-1}Myr^{-1} in humans (Parsons et al. 1997, Sigurdardóttir et al. 2000, Howell et al. 2003, Henn et al. 2009) or 95 % site^{-1}Myr^{-1} in Adélie penguins (Lambert et al. 2002). These estimates are

much higher than the recognized substitution rate of 1 % for protein-coding mtDNA sequences (e.g., Brown et al. 1979) or 10 % site^{-1}Myr^{-1} usually used for D-loop sequences (Prager et al. 1993, Gündüz et al. 2005, Rajabi Maham et al. 2008) or 4.5 % site^{-1}Myr^{-1} estimated for mice in this study. Similarly, Goios et al. (2007) found *ca.* 30-times deeper coalescence of mouse laboratory strains in comparison with their known recent history. This discrepancy was explained by Ho et al. (2005, 2007), who found that the relationship between the mutation rate (in average nucleotide changes per site per million years) and the dates used to calibrate the estimates can be described by a vertically translated exponential decay curve. Thus the mutation rate at a site can be a magnitude higher than the rate at which one base is replaced by another. According to Ho et al. (2005) this pattern is most plausibly explained by the action of purifying selection rather than sequencing and calibration error or mutation saturation. These results thus show that we should not simply extrapolate molecular rates across different evolutionary timescales (Ho et al. 2005, Burridge et al. 2008, Henn et al. 2009).

Recently, Rajabi Maham et al. (2008) have followed these findings for correcting their estimate of the time of *M. m. domesticus* population expansion (≈60 kya) based on the 10 % mutation rate commonly used for the evolution of mouse D-loop sequences (Prager et al. 1993, Gündüz et al. 2005). This estimate seems to be rather high given the commonly accepted view that house mice have colonized Europe in the Holocene, i.e., during the last 12000 years. This figure is even higher than the much-disputed 30-40 kya presence of *M. m. domesticus* in Europe (Sage et al. 1990; see above). In addition, if a more realistic 5 % rate (an average between Goios et al. (2007) and this study) is used we get about twice as old date as the estimate of Rajabi Maham et al. (2008), or ~84-132 kya when calibrated with the *domesticus*-(*musculus*, *castaneus*) splits in Table 5. Therefore, Rajabi Maham et al. (2008) recalibrated the mutation rate to reconcile their results with the paleontological evidence and concluded with about four-fold increase of the intrasubspecific mutation rate (≈40 % site^{-1}Myr^{-1}) relative to long-term, intersubspecific, substitution rate.

However, although the recalibrated data of Rajabi Maham et al. (2008) are intuitively more consistent to what is known about the evolution of house mice in the Middle East and Europe, we find their reasoning a bit circular as they base their recalibration of the mutation rate on the assumption (though supported by the fossil record) that the mouse populations expansion *must* have been younger than the first warm period (Alleröd) starting *ca*. 12 kya and then use the resulting re-estimate for computing the time when the expansion started (revealing, not surprisingly, ~12.5 kya). To avoid such the circularity in reasoning we tried to estimate the mutation rate without fixing of or setting strong prior for TMRCA. Without any known reliable calibration point for events younger than 2 Mya for mice it appears very difficult, however, Beast revealed a weak yet consistent signal between 40 % and 60 % site^{-1}Myr^{-1} suggesting an about 10-fold intrasubspecific mutation rate relative to the overall long-term substitution rate. An average of 50 % yields TMRCA of 20 kya for *M. m. domesticus*, 22 kya for *M. m. castaneus*, and 12 kya for *M. m. musculus* (Table 5, TMRCA 4). These estimates are consistent with the results based on the BSP model (Fig. 8) showing a single expansion of *M. m. castaneus* around 17 kya and much younger expansion of *M. m. musculus* around 5 kya. A more complex pattern was revealed in *M. m. domesticus* which has apparently undergone at least two waves of expansion, the first taking place between 20 and 30 kya and the second, most probably connected with colonization of Europe, starting around 5 kya. Thus our results are in pretty good agreement with those of Klein et al. (1987), Auffray et al. (1990), Britton-Davidian (1990), Auffray & Britton-Davidian (1992) and others. In addition, our results suggest that *castaneus* and *domesticus* population expansions followed shortly after their basal divergences. Both these expansions were much older than *musculus* population growth. Nevertheless, in spite of the older divergence and expansion of *M. m. domesticus* both *domesticus* and *musculus* mice have colonized Europe roughly at the same time.

A rather surprising finding is the presence of presumed nuclear mitochondrial DNA (numt) in 11 individuals from the island of Socotra. Numts have been described in a vast array of organisms including humans (see Bensasson et al. 2001 for review). The length of these sequences varies and can reach up to several kilobases such as in the genus *Panthera* (12536 bp; Kim et al. 2006). In the mouse, mitochondrial-DNA-like inserts were only found in normal and tumor cell lines derived from DBA/2 inbred strain (Hadler et al. 1998). To our knowledge no numts have been described in free-living house mouse populations. Detail description of the whole Socotran numt sequences and their place in the nuclear genome is to be published elsewhere.

Acknowledgements

We are grateful to P.K. Tucker for the idea that the unidentified haplotypes from the island of Socotra can

be numts. E.M. Prager, P.K. Tucker, and R.D. Sage are acknowledged for reading an earlier version of the manuscript. We thank D. Čížková and L. Vlčková for assistance with preparing samples for the whole-mtDNA sequencing and J. Piálek for logistic help. This analysis was carried out in the Centre for GeoGenetics, Denmark, in collaboration with T. Gilbert. F. Bonhomme and A. Orth provided the sample of the CAS strain. This study was supported by NSF grants 206/06/0707 and 206/08/0640.

Literature

Abe K., Noguchi H., Tagawa T., Yuzuriha M., Toyoda A., Kojima T., Ezawa K., Saitou N., Hattori M., Sakaki Y., Moriwaki K. & Shiroishi T. 2004: Contribution of Asian mouse subspecies *Mus musculus molossinus* to genomic constitution of strain C57BL/6J, as defined by BAC-end sequence-SKIP analysis. *Genome Res. 14: 2439–2447.*

Akaike H. 1974: A new look at the statistical model identification. *IEEE Trans. Automat. Contr. AC 19: 716–723.*

Auffray J.-C. & Britton-Davidian J. 1992: When did the house mouse colonize Europe? *Biol. J. Linn. Soc. 45: 187–190.*

Auffray J.-C. & Britton-Davidian J. 2012: The house mouse and its relatives: systematics and taxonomy. In: Macholán M., Baird S.J.E., Munclinger P. & Piálek L. (eds.), Evolution of the house mouse. *Cambridge studies in morphology and molecules: new paradigms in evolutionary biology. Cambridge University Press, Cambridge: 1–34.*

Auffray J.-C., Tchernov E. & Nevo E. 1988: Origine du commensalisme de la souris domestique (*Mus musculus domesticus*) vis-´-vis de l'homme. *CR Acad. Sci. Paris, Serie III. 307: 517–522.*

Auffray J.-C., Vanlerberghe F. & Britton-Davidan J. 1990: The house mouse progression in Eurasia: a paleontological and archaeozoological approach. *Biol. J. Linn. Soc. 41: 13–25.*

Baird S.J.E. & Macholán M. 2012: What can the *Mus musculus musculus/M. m. domesticus* hybrid zone tell us about speciation? In: Macholán M., Baird S.J.E., Munclinger P. & Piálek L. (eds.), Evolution of the house mouse. *Cambridge studies in morphology and molecules: new paradigms in evolutionary biology. Cambridge University Press, Cambridge: 334–372.*

Bandelt H.-J., Forster P. & Röhl A. 1999: Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol. 16: 37–48.*

Bensasson D., Zhang D.-X., Hartl D.L. & Hewitt G.M. 2001: Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol. Evol. 16: 314–321.*

Benton M.J. & Donoghue P.C.J. 2007: Paleontological evidence to date the tree of life. *Mol. Biol. Evol. 24: 26–53.*

Berry R.J. 1981: Town mouse, country mouse: adaptation and adaptability in *Mus domesticus* (*M. musculus domesticus*). *Mamm. Rev. 11: 91–136.*

Berry R.J. 1995: An animal weed – man's relationship with the house mouse. In: Macdonald D. (ed.), The encyclopedia of mammals. *Facts on File, New York: 664–665.*

Bibb M.J., Van Etten R.A., Wright C.T., Walberg M.W. & Clayton D.A. 1981: Sequence and gene organization of mouse mitochondrial DNA. *Cell. 26: 167–180.*

Bishop C.E., Boursot P., Baron B., Bonhomme F. & Hatat D. 1985: Most classical *Mus musculus domesticus* laboratory mouse strains carry a *Mus musculus musculus* Y chromosome. *Nature 315: 70–72.*

Blank R.D., Campbell G.R. & D'Eustachio P. 1986: Possible derivation of the laboratory mouse genome from multiple wild *Mus* species. *Genetics 114: 1257–1269.*

Boissinot S. & Boursot P. 1997: Discordant phylogeographic patterns between the Y chromosome and mitochondrial DNA in the house mouse: selection on the Y chromosome? *Genetics 146: 1019–1034.*

Bonhomme F., Catalan J., Britton-Davidian J., Chapman K., Moriwaki K., Nevo E. & Thaler L. 1984: Biochemical diversity and evolution in the genus *Mus*. *Biochem. Genet. 22: 275–303.*

Bonhomme F., Guénet J.-L., Dod B., Moriwaki K. & Bulfield G. 1987: The polyphyletic origin of laboratory inbred mice and their rate of evolution. *Biol. J. Linn. Soc. 30: 51–58.*

Boursot P., Auffray J.-C., Britton-Davidian J. & Bonhomme F. 1993: The evolution of house mice. *Annu. Rev. Ecol. Syst. 24: 119–152.*

Boursot P., Din W., Anand R., Darviche D., Dod B., von Deimling F., Talwar G.P. & Bonhomme F. 1996: Origin and radiation of the house mouse: mitochondrial DNA phylogeny. *J. Evol. Biol. 9: 391–415.*

Britton-Davidian J. 1990: Genic differentiation in *M. m. domesticus* populations from Europe, the Middle East and North Africa: geographic patterns and colonization events. *Biol. J. Linn. Soc. 41: 27–45.*

Brown W.M., George M., Jr. & Wilson A.C. 1979: Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA 76: 1967–1971.*

Burridge C.P., Craw D., Fletcher D. & Waters J.M. 2008: Geological dates and molecular rates: fish DNA sheds light on time dependency. *Mol. Biol. Evol. 25: 624–633.*

Catzeflis F.M. & Denys C. 1992: The African *Nannomys* (Muridae): an early offshoot from the *Mus* lineage – evidence from scnDNA hybridization experiments and compared morphology. *Isr. J. Zool. 38: 219–231.*

Chevret P., Jenkins P. & Catzeflis F. 2003: Evolutionary systematics of the Indian mouse *Mus famulus* Bonhote, 1898: molecular (DNA/DNA hybridization and 12S rRNA sequences) and morphological evidence. *Biol. J. Linn. Soc. 137: 385–401.*

Cucchi T., Auffray J.-C. & Vigne J.D. 2012: On the origin of the house mouse synanthropy and dispersal in the Near East and Europe: zooarchaeological review and perspectives. In: Macholán M., Baird S.J.E., Munclinger P. & Piálek L. (eds.), Evolution of the house mouse. *Cambridge studies in morphology and molecules: new paradigms in evolutionary biology. Cambridge University Press, Cambridge: 65–93.*

Cucchi T., Orth A., Auffray J.-C., Renaud S., Fabre L., Catalan J., Hadjisterkotis E., Bonhomme F. & Vigne J.-D. 2006: A new endemic species of the subgenus *Mus* (Rodentia, Mammalia) on the Island of Cyprus. *Zootaxa 1241: 1–36.*

Cucchi T., Vigne J.-T. & 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. *Biol. J. Linn. Soc. 84: 429–445.*

Cucchi T., Vigne J.-T., Auffray J.-C., Croft P. & Peltenburg E. 2002: Introduction involontaire de la souris domestique (*Mus musculus domesticus*) à Chypre dès le Néolithique précéramique ancien (fin IXe et VIIIe millénaires av. J.-C.). *CR Palevol. 1: 235–241.*

Dallas J.F., Dod B., Boursot P., Prager E.M. & Bonhomme F. 1995: Population subdivision and gene flow in Danish house mice. *Mol. Ecol. 4: 311–320.*

Dietrich W.F., Miller J., Steen R., Merchant M.A., Damron-Boles D., Husain Z., Dredge R., Daly M.J., Ingalls K.A., O'Connor T.J., Evans C.A., DeAngelis M.M., Levinson D.M., Kruglyak L., Goodman N., Copeland N.G., Jenkins N.A., Hawkins T.L., Stein L., Page D.C. & Lander E.S. 1996: A comprehensive genetic map of the mouse genome. *Nature 380: 149–152.*

Din W., Anand R., Boursot P., Darviche D., Dod B., Jouvin-Marche E., Orth A., Talwar G.P., Cazenawe P.-A. & Bonhomme F. 1996: Origin and radiation of the house mouse: clues from nuclear genes. *J. Evol. Biol. 9: 519–539.*

Drummond A.J. & Rambaut A. 2006: Beast v1.4, Bayesian evolutionary analysis sampling trees. *Available from: http://beast.bio.ed.ac.uk*

Drummond A.J., Rambaut A., Shapiro B. & Pybus O.G. 2005: Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol. 22: 1185–1192.*

Duplantier J.-M., Orth A., Catalan J. & Bonhomme F. 2002: Evidence for a mitochondrial lineage originating from the Arabian peninsula in the Madagascar house mouse (*Mus musculus*). *Heredity 89: 154–158.*

Duvaux L., Belkhir K., Boulesteix M. & Boursot P. 2011: Isolation and gene flow: inferring the speciation history of European house mice. *Mol. Ecol. 20: 5248–5264.*

Edwards S.V. 1997: Relevance of microevolutionary processes to higher level molecular systematics. In: Mindell D.P. (ed.), Avian molecular evolution and systematics. *Academic Press, New York: 251–278.*

Excoffier L. 2007: Analysis of population subdivision. In: Balding D.J., Bishop M. & Cannings C. (eds.), Handbook of statistical genetics. 3rd ed. *Wiley, Chichester: 980–1020.*

Excoffier L., Smouse P.E. & Quattro J.M. 1992: Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics 131: 479–491.*

Felsenstein J. 1981: Evolutionary trees from DNA sequences. A maximum likelihood approach. *J. Mol. Evol. 17: 368–376.*

Felsenstein J. 1985: Confidence limits on phylogenies with a molecular clock. *Syst. Zool. 34: 152–161.*

Felsenstein J. 2004: Inferring phylogenies. *Sinauer Associates, Sunderland, MA.*

Ferris S.D., Sage R.D., Huang C.-M., Nielsen J.T., Ritte U. & Wilson A.C. 1983: Flow of mitochondrial DNA across a species boundary. *Proc. Natl. Acad. Sci. USA 80: 2290–2294.*

Figueroa F., Kasahara M., Tichy H., Neufeld E., Ritte U. & Klein J. 1987: Polymorphism of unique noncoding DNA sequences in wild and laboratory mice. *Genetics 117: 101–108.*

Fitch W.M. 1971: Towards defining the course of evolution: minimum change for a specific tree topology. *Syst. Zool. 20: 406–416.*

Fu Y.-X. 1997: Statistical tests of neutrality of mutations against populations growth, hitchhiking and background selection. *Genetics 147: 915–925.*

Gabriel S.I., Stevens M.I., Mathias M.D. & Searle J.B. 2011: Of mice and 'convicts': origin of the Australian house mouse, *Mus musculus*. *PLoS ONE 6(12): e28622.*

Goios A., Pereira L., Bogue M., Macaulay V. & Amorim A. 2007: mtDNA phylogeny and evolution of laboratory mouse strains. *Genome Res. 17: 293–298.*

Guénet J.-L. & Bonhomme F. 2003: Wild mice: an ever-increasing contribution to a popular mammalian model. *Trends Genet. 19: 24–31.*

Gündüz İ., Rambau R.V., Tez C. & Searle J.B. 2005: Mitochondrial DNA variation in the western house mouse (*Mus musculus domesticus*) close to its site of origin: studies in Turkey. *Biol. J. Linn. Soc. 84: 473–485.*

Gündüz İ., Tez C., Malikov V., Vaziri A., Polyakov A.V. & Searle J.B. 2000: Mitochondrial DNA and chromosomal studies of wild mice (*Mus*) from Turkey and Iran. *Heredity 84: 458–467.*

Gyllensten U. & Wilson A.C. 1987: Interspecific mitochondrial DNA transfer and the colonization of Scandinavia by mice. *Genet. Res. 49: 25–29.*

Hadler H.I., Devadas K. & Mahalingam R. 1998: Selected nuclear LINE elements with mitochondrial-DNA-like inserts are more plentiful and mobile in tumor than in normal tissue of mouse and rat. *J. Cell. Biochem. 68: 100–109.*

Harpending R.C. 1994: Signatures of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biol. 66: 591–600.*

Harrison D.L. 1972: The mammals of Arabia, Vol. III. *Ernest Benn, London.*

Harrison D.L. & Bates P.J.J. 1991: The mammals of Arabia, 2nd ed. *Harrison Zoological Museum, Sevenoaks.*

Hasegawa M., Kishino H. & Yano T. 1985: Dating the human-ape split by a molecular clock of mitochondrial DNA. *J. Mol. Evol. 22: 160–174.*

Hassanin A., Lecointre G. & Tillier S. 1998: The 'evolutionary signal' of homoplasy in protein-coding gene sequences and its consequences for a priori weighting in phylogeny. *CR Acad. Sci., Paris, Sciences Vie. 321: 611–620.*

Henn B.M., Gignoux C.R., Feldman M.W. & Mountain J.L. 2009: Characterizing the time dependency of human mitochondrial DNA mutation rate estimates. *Mol. Biol. Evol. 26: 217–230.*

Ho S.Y.W., Phillips M.J., Cooper A. & Drummond A.J. 2005: Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol. Biol. Evol. 22: 1561–1568.*

Ho S.Y.W., Shapiro B., Phillips M.J., Cooper A. & Drummond A.J. 2007: Evidence for time dependency of molecular rate estimates. *Syst. Biol. 56: 515–522.*

Howell N., Smejkal C.B., Mackey D.A., Chinnery P.F., Turnbull D.M. & Herrnstadt C. 2003: The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. *Am. J. Hum. Genet. 72: 659–670.*

Hudson R.R. 1990: Gene genealogies and the coalescent process. In: Futuyma D. & Antonovics J.D. (eds.), Oxford surveys of evolutionary biology. *Oxford University Press, New York: 1–11.*

Jacobs L.L. & Downs W.R. 1994: The evolution of murine rodents in Asia. In: Tomida Y., Li C.K. & Setoguschi T. (eds.), Rodent and lagomorph families of Asian origins and their diversification. *National Science Museum Monograph, Tokyo: 149–156.*

Jacobs L.L. & Flynn L.J. 2005: Of mice ... again: the Siwalik rodent record, murine distribution, and molecular clocks. In: Liebermann D., Smith R. & Kelley J. (eds.), Interpreting the past: essays on human, primate and mammal evolution. *Bill Academic Publishers, Leiden, The Netherlands: 63–80.*

Jaeger J.J., Tong H. & Denys C. 1986: The age of the *Mus-Rattus* divergence – paleontological data compared with the molecular clock. *CR Acad. Sci. II. 302: 917–922.*

Jones E.P., Jensen J.-K., Magnussen E., Gregersen N., Hansen H.S. & Searle J.B. 2011a: A molecular characterization of the charismatic Faroe house mouse. *Biol. J. Linn. Soc. 102: 471–482.*

Jones E.P., Johannesdottir F., Gündüz İ., Richards M.B. & Searle J.B. 2011b: The expansion of the house mouse into north-western Europe. *J. Zool. 283: 257–268.*

Kim J.H., Antunes A., Luo S.J., Menninger J., Nash W.G., O'Brien S.J. & Johnson W.E. 2006: Evolutionary analysis of a large mtDNA translocation (numt) into the nuclear genome of the *Panthera* genus species. *Gene 366: 292–302.*

Kimura M. 1970: The length of time required for a selectively neutral mutant to reach fixation through random frequency drift in a finite population. *Genet. Res. 15: 131–133.*

Klein J., Tichy H. & Figueroa F. 1987: On the origin of mice. *Ann. Univ. Chile. 5 (14): 91–120.*

Klein J., Vincek V., Kasahara M. & Figueroa F. 1988: Probing mouse origins with random DNA probes. *Curr. Top. Microbiol. Immunol. 137: 55–63.*

Kuhner M.K. 2006: LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics Applications Note 22: 768–770.*

Kuhner M.K., Yamato J. & Felsenstein J. 1995: Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. *Genetics 140: 1421–1430.*

Kuhner M.K., Yamato J. & Felsenstein J. 1998: Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics 149: 429–434.*

Laakonen J., Kallio-Kokko H., Vapalahti O., Vaheri A., Vyskočilová M., Munclinger P., Macholán M. & Henttonen H. 2007: The screening of parasites and viral pathogens of small mammals from a farm in southern Finland, and genetic identification of the Finnish house mouse, *Mus musculus*. *Ann. Zool. Fenn. 44: 202–208.*

Lambert D.M., Ritchie P.A., Millar C.D., Holland B., Drummond A.J. & Baroni C. 2002: Rates of evolution in ancient DNA from Adélie penguins. *Science 295: 2270–2273.*

Li W.-H. 1977: Distribution of nucleotide differences between two randomly chosen cistrons in a finite population. *Genetics 85: 331–337.*

Lindblad-Toh K., Winchester E., Daly M.J., Wang D.G., Hirschhorn J.N., Laviolette J.P., Ardlie K., Reich D.E., Robinson E., Sklar P., Shah N., Thomas D., Fan J.B., Gingeras T., Warrington J., Patil N., Hudson T.J. & Lander E.S. 2000: Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nature Genet. 24: 381–386.*

Liu Y.-H., Takahashi A., Kitano T., Koide T., Shiroishi T., Moriwaki K. & Saitou N. 2008: Mosaic genealogy of the *Mus musculus* genome revealed by 21 nuclear genes from its three subspecies. *Genes Genet. Syst. 83: 77–88.*

Lundrigan B.L., Jansa S.A. & Tucker P.K. 2002: Phylogenetic relationships in the genus *Mus*, based on paternally, maternally and biparentally inherited characters. *Syst. Biol. 51: 410–431.*

Macholán M., Munclinger P., Šugerková M., Dufková P., Bímová B., Božíková E., Zima J. & Piálek J. 2007a: Genetic analysis of autosomal and X-linked markers across a mouse hybrid zone. *Evolution 61: 746–771.*

Macholán M., Vyskočilová M., Bonhomme F., Kryštufek B., Orth A. & Vohralík V. 2007b: Genetic variation and phylogeography of free-living mouse species (genus *Mus*) in the Balkans and the Middle East. *Mol. Ecol. 16: 4774–4788.*

Mouse Genome Sequencing Consortium 2002: Initial sequencing and comparative analysis of the mouse genome. *Nature 420: 520–562.*

Moriwaki K. 1994: Wild mouse from a geneticist's viewpoint. In: Moriwaki K., Shiroishi T. & Yonekawa H. (eds.), Genetics in wild mice. *Japan Scientific Societies Press, Tokyo: 13–25.*

Nei M. 1987: Molecular evolutionary genetics. *Columbia University Press, New York, NY.*

Nishioka Y. 1987: Y-chromosomal DNA polymorphism in mouse inbred strains. *Genet. Res. 50: 69–72.*

Nordborg M. 2007: Coalescent theory. In: Balding D.J., Bishop M. & Cannings C. (eds.), Handbook of statistical genetics. 3rd edn. *Wiley, Chichester: 843–877.*

Parsons T.J., Muniec D.S., Sullivan K., Woodyatt N., Alliston-Greiner R., Wilson M.R., Berry D.L., Holland K.A., Weedn V.W., Gill P. & Holland M.M. 1997: A high observed substitution rate in the human mitochondrial DNA control region. *Nat. Genet. 15: 363–368.*

Pelikán J. 1981: Patterns of reproduction in the house mouse. In: Berry R.J. (ed.), Biology of the house mouse. *Academic Press, London: 205–230.*

Petkov P.M., Ding Y.M., Cassell M.A., Zhang W., Wagner G., Sargent E.E., Asquith S., Crew V., Johnson K.A., Robinson P., Scott V.E. & Wiles M.V. 2004: An efficient SNP system for mouse genome scanning and elucidating strain relationships. *Genome Res. 14: 1806–1811.*

Philippe H. & Douzery E. 1994: The pitfalls of molecular phylogeny based on four species as illustrated by the Cetacea/Artiodactyla and postglacial patterns of subdivision in the meadow grasshopper *Chorthippus parallelus*. *Heredity 80: 633–641.*

Pletcher M.T., McClurg P., Batalov S., Su A.I., Barnes S.W., Lagler E., Korstanje R., Wang X.S., Nusskern D., Bogue M.A., Mural R.J., Paigen B. & Wiltshire T. 2004: Use of a dense single nucleotide polymorphism map for in silico mapping in the mouse. *PLoS Biol. 2: 2159–2169.*

Posada D. & Crandall K.A. 1988: ModelTest: Testing the model of DNA substitution. *Bioinformatics 14: 817–818.*

Prager E.M., Orrego C. & Sage R.D. 1998: Genetic variation and phylogeography of Central Asian and other house mice, including a major new mitochondrial lineage in Yemen. *Genetics 150: 835–861.*

Prager E.M., Sage R.D., Gyllensten U., Thomas W.K., Hübner R., Jones C.S., Noble L., Searle J.B. & Wilson A.C. 1993: Mitochondrial DNA sequence diversity and the colonization of Scandinavia by house mice from East Holstein. *Biol. J. Linn. Soc. 50: 85–122.*

Prager E.M., Tichy H. & Sage R.D. 1996: Mitochondrial DNA sequence variation in the eastern house mouse, *Mus musculus*: comparison with other house mouse mice and report of 75-bp tandem repeat. *Genetics 143: 427–446.*

Rajabi Maham H. 2007: Phylogéographie des souris du complexe *Mus musculus* en Iran: coalescence mitochondriale et diversité du chromosome Y. *Ph.D. Thesis, Université Montpellier II.*

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. *Mol. Ecol. 17: 627–641.*

Rannala B. & Yang Z. 1996: Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *J. Mol. Evol. 43: 304–311.*

Rannala B. & Yang Z. 2007: Inferring speciation times under an episodic molecular clock. *Syst. Biol. 56: 453–466.*

Raufaste N., Orth A., Belkhir K., Senet D., Smadja C., Baird S.J.E., Bonhomme F., Dod B. & Boursot P. 2005: Inferences of selection and migration in the Danish house mouse hybrid zone. *Biol. J. Linn. Soc. 84: 593–616.*

Robinson M., Gouy M., Gautier C. & Mouchiroud D. 1998: Sensitivity of the relative-rate test to taxonomic sampling. *Mol. Biol. Evol. 15: 1091–1098.*

Robinson-Rechavi M. & Huchon D. 2000: RRTree: Relative-rate tests between groups of sequences on a phylogenetic tree. *Bioinformatics 16: 296–297.*

Rogers A.R. 1995: Genetic evidence for a Pleistocene population explosion. *Evolution 49: 608–615.*

Rogers A.R. & Harpending H. 1992: Population growth marks waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol. 9: 552–569.*

Ronquist F. & Huelsenbeck J.P. 2003: MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics 19: 1572–1574.*

Ruvinsky A., Polyakov A., Agulnik A., Tichy H., Figueroa F. & Klein J. 1991: Low diversity of *t* haplotypes in the eastern form of the house mouse, *Mus musculus* L. *Genetics 127: 161–168.*

Sage R.D. 1981: Wild mice. In: Foster H.L., Small J.D. & Fox J.G. (eds.), The mouse in biomedical research. *Academic Press, New York: 39–90.*

Sage R.D., Atchley W.R. & Capanna E. 1993: House mice as models in systematic biology. *Syst. Biol. 42: 523–561.*

Sage R.D., Prager E.M., Tichy H. & Wilson A.C. 1990: Mitochondrial-DNA variation in house mice, *Mus domesticus* (Rutty). *Biol. J. Linn. Soc. 41: 105–123.*

Saitou N. & Nei M. 1987: The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol. 4: 406–425.*

Schneider S. & Excoffier L. 1999: Estimation of demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics 102: 1079–1089.*

Schneider S., Roessli D. & Excoffier L. 2000: Arlequin ver. 2.000: a software for population genetics data analysis. *Genetics and Biometry Laboratory, University of Geneva, Switzerland.*

She J.X., Bonhomme F., Boursot P., Thaler L. & Catzeflis F. 1990: Molecular phylogenies in the genus *Mus*: comparative analysis of electrophoretic, scnDNA hybridisation, and mtDNA RFLP data. *Biol. J. Linn. Soc. 41: 83–103.*

Shifman S., Bell J.T., Copley R.R., Taylor M.S., Williams R.W., Mott R. & Flint J. 2006: A high resolution single nucleotide polymorphism genetic map of the mouse genome. *PLoS Biol. 4: 2227–2237.*

Sigurdardóttir S., Helgason A., Gulcher J.R., Stefánsson K. & Donnelly P. 2000: The mutation rate in the human mtDNA control region. *Am. J. Hum. Genet. 66: 1599–1609.*

Suzuki H. & Aplin K.P. 2012: Phylogeny and biogeography of the genus *Mus* in Eurasia. In: Macholán M., Baird S.J.E., Munclinger P. & Piálek L. (eds.), Evolution of the house mouse. *Cambridge studies in morphology and molecules: new paradigms in evolutionary biology. Cambridge University Press, Cambridge: 35–64.*

Suzuki H. & Kurihara Y. 1994: Genetic variation of ribosomal RNA in the house mouse, *Mus musculus*. In: Moriwaki K., Shiroishi T. & Yonekawa H. (eds.), Genetics in wild mice. *Japan Scientific Societies Press, Tokyo: 107–119.*

Suzuki H., Shimada T., Terashima M., Tsuchiya K. & Aplin K. 2004: Temporal, spatial, and ecological modes of evolution of Eurasian *Mus* based on mitochondrial and nuclear gene sequences. *Mol. Phylogenet. Evol. 33: 626–646.*

Swofford D.L. 2000: PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4.0b10. *Sinauer Associates, Sunderland, Massachusetts.*

Tajima F. 1983: Evolutionary relationship of DNA sequences in finite populations. *Genetics 105: 437–460.*

Tajima F. 1989a: Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics 123: 585–595.*

Tajima F. 1989b: The effect of change in population size on DNA polymorphism. *Genetics 123: 597–601.*

Tamura K. & Nei M. 1993: Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol. 10: 512–526.*

Templeton A.R. 2005: Haplotype trees and modern human origins. *Yearbk. Phys. Anthropol. 48: 33–59.*

Templeton A.R. 2006: Population genetics and microevolutionary theory. *John Wiley, Hoboken, New Jersey.*

Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F. & Higgins D.G. 1997: The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res. 25: 4876–4882.*

Tucker P.K., Sage R.D., Warner J., Wilson A.C. & Eicher E.M. 1992: Abrupt cline for sex chromosomes in a hybrid zone between two species of mice. *Evolution 46: 1146–1163.*

Tucker P.K., Sandstedt S.A. & Lundrigan B.L. 2005: Phylogenetic relationships in the subgenus *Mus* (genus *Mus*, family Muridae, subfamily Murinae): examining gene trees and species trees. *Biol. J. Linn. Soc. 84: 653–662.*

Vanlerberghe F., Boursot P., Nielsen J.T. & Bonhomme F. 1988: A steep cline for mitochondrial DNA in Danish mice. *Genet. Res. 52: 185–193.*

Veyrunes F., Britton-Davidian J., Robinson T.J., Calvet E., Denys C. & Chevret P. 2005: Molecular phylogeny of the African pygmy mice, subgenus *Nannomys* (Rodentia, Murinae, *Mus*): implications for chromosomal evolution. *Mol. Phylogenet. Evol. 36: 358–369.*

Wu C.-I. & Li W.-H. 1985: Evidence for higher rates of nucleotide substitutions in rodents than in man. *Proc. Natl. Acad. Sci. USA 82: 1741–1745.*

Xia X. & Xie Z. 2001: DAMBE: Software package for data analysis in molecular biology and evolution. *J. Hered. 92: 371–373.*

Yonekawa H., Sato J.J., Suzuki H. & Moriwaki K. 2012: Origin and genetic status of *Mus musculus molossinus*: a typical example of reticulate evolution in the genus *Mus*. In: Macholán M., Baird S.J.E., Munclinger P. & Piálek L. (eds.), Evolution of the house mouse. *Cambridge studies in morphology and molecules: new paradigms in evolutionary biology. Cambridge University Press, Cambridge: 94–113.*