Genetic Variation and Phylogeography of Central Asian and Other House Mice, Including a Major New Mitochondrial Lineage in Yemen

Ellen M. Prager,*,[1] Cristián Orrego[†,‡,1] and Richard D. Sage[§,,2]**

**Division of Biochemistry and Molecular Biology, University of California, Berkeley, California 94720-3202, [†]Museum of Vertebrate Zoology, University of California, Berkeley, California 94720-3160, [‡]Conservation Genetics Laboratory, Department of Biology, San Francisco State University, San Francisco, California 94132-1722, [§]Division of Biological Sciences, University of Missouri, Columbia, Missouri 65211 and **Department of Biological Sciences, University of California, Santa Barbara, California 93106*

Manuscript received October 20, 1997
Accepted for publication July 7, 1998

ABSTRACT

The mitochondrial DNA (mtDNA) control region and flanking tRNAs were sequenced from 76 mice collected at 60 localities extending from Egypt through Turkey, Yemen, Iran, Afghanistan, Pakistan, and Nepal to eastern Asia. Segments of the *Y* chromosome and of a processed *p53* pseudogene (Ψ*p53*) were amplified from many of these mice and from others collected elsewhere in Eurasia and North Africa. The 251 mtDNA types, including 54 new ones reported here, now identified from commensal house mice (*Mus musculus* group) by sequencing this segment can be organized into four major lineages—*domesticus*, *musculus*, *castaneus*, and a new lineage found in Yemen. Evolutionary tree analysis suggested the *domesticus* mtDNAs as the sister group to the other three commensal mtDNA lineages and the Yemeni mtDNAs as the next oldest lineage. Using this tree and the phylogeographic approach, we derived a new model for the origin and radiation of commensal house mice whose main features are an origin in west-central Asia (within the present-day range of *M. domesticus*) and the sequential spreading of mice first to the southern Arabian Peninsula, thence eastward and northward into south-central Asia, and later from south-central Asia to north-central Asia (and thence into most of northern Eurasia) and to southeastern Asia. *Y* chromosomes with and without an 18-bp deletion in the *Zfy-2* gene were detected among mice from Iran and Afghanistan, while only undeleted *Y*s were found in Turkey, Yemen, Pakistan, and Nepal. Polymorphism for the presence of a Ψ*p53* was observed in Georgia, Iran, Turkmenistan, Afghanistan, and Pakistan. Sequencing of a 128-bp Ψ*p53* segment from 79 commensal mice revealed 12 variable sites and implicated ≥14 alleles. The allele that appeared to be phylogenetically ancestral was widespread, and the greatest diversity was observed in Turkey, Afghanistan, Pakistan, and Nepal. Two mice provided evidence for a second Ψ*p53* locus in some commensal populations.

WITHIN the past two decades, a number of important issues about the genetic variation and phylogenetic relationships of members of the house mouse species group have been resolved, and data are accumulating steadily with respect to several remaining fundamental questions about the extent and organization of the variation in wild mice and the relationships, origin, and radiation of the commensal taxa (*e.g.*, see Boursot *et al.* 1993, 1996; Sage *et al.* 1993; Moriwaki *et al.* 1994; Din *et al.* 1996; Prager *et al.* 1996; Boissinot and Boursot 1997). Thus, it has been demonstrated that the three aboriginal species—*Mus spicilegus*, *M. macedonicus*, and *M. spretus*, each of which occupies limited ranges in Europe, western (W) Asia, and North Africa—lie phylogenetically outside the commensal clade. The preponderance of evidence indicates that *M. spretus* is an outgroup to all the other house mouse taxa.

The native range of the commensal house mice collectively is all of Eurasia plus North Africa. According to the most commonly used system, they can be divided into three or four taxa that, in a binomial classification, are designated *M. domesticus* of W Europe, North Africa, and the Middle East; *M. musculus* of eastern (E) Europe and northern (N) Asia; *M. castaneus* of southeastern (SE) Asia; and *M. bactrianus* of south-central (SC) Asia from Iran to N India. (In the trinomial classification system, these taxa would be called *M. m. domesticus*, *M. m. musculus*, *M. m. castaneus*, and *M. m. bactrianus*.) *M. bactrianus* is the least well defined and characterized taxon, and it is not known whether it is a cohesive genetic entity. On a broader scale, the genetic constitution of the central populations—from the Indian subcontinent, Afghanistan, and Iran—and their genetic affiliations with the other taxa are just now being elucidated, and it has been suggested (Boursot *et al.* 1993, 1996; Din *et al.* 1996) that assignment of a particular

Corresponding author: Ellen M. Prager, Department of Biology, San Francisco State University, 1600 Holloway Avenue, San Francisco, CA 94132-1722. E-mail: emprager@sfsu.edu

[1] *Present address:* Department of Biology, San Francisco State University, San Francisco, CA 94132-1722.

[2] *Present address:* Department of Biological Sciences, University of California, Santa Barbara, CA 93106.

taxonomic name to members of the central populations (including those previously called *M. bactrianus*) be held in abeyance. [Mice from many central populations have been categorized as *M. domesticus* on the basis of morphological criteria (Marshall and Sage 1981).]

The corollary issues being addressed concern the geographic origin of the commensal clade and the modes and routes of radiation giving rise to the diverse species and populations over their present-day ranges. The geological feature of primary importance in understanding the past and present ranges of house mice is the east-west wall of high mountains that runs through Europe and Asia. This backbone of Eurasia, which in Central Asia encompasses the ranges from the Caucasus to the Himalayas, is the major geographic barrier that keeps *M. musculus* in northern Eurasia, away from the commensal taxa that inhabit southern (S) Eurasia. The Zagros Mountains, which run N–S through W and S Iran, may well act in the same way and form the major geographic barrier that keeps *M. domesticus* in the west, away from other SC Asian mice. These mountain massifs act as barriers to mice during both glacial periods (when the higher elevations are colder and even glaciated) and interglacials [when these mountains become forested and, thus, also inhospitable to house mice (Sage 1981)]. Explaining where and how ancestral house mice got from one side of these barriers to the other is a significant challenge for any hypothesis of commensal mouse origin and radiation.

A consensus is lacking as to whether the commensal house mouse taxa should be regarded as full species or as subspecies or perhaps as semispecies (*e.g.*, see Sage *et al.* 1986, 1993; Auffray *et al.* 1990a; Boursot *et al.* 1993; Bonhomme *et al.* 1994; Prager *et al.* 1996; references therein). Thus, on the basis of evidence of separate gene pools, notably of *M. domesticus* and *M. musculus* in Europe, R. D. Sage and E. M. Prager have denoted them as full species, while other investigators, including P. Boursot, F. Bonhomme, and co-workers (*e.g.*, Boursot *et al.* 1993, 1996; Moriwaki *et al.* 1994; Din *et al.* 1996), designate them as subspecies in light of appreciable evidence for a continuum of interbreeding populations over much of Eurasia. These contrasting views become more understandable if *M. musculus* is a ring species (Bonhomme *et al.* 1994; Din *et al.* 1996), with the secondary contact in Europe occurring between the most divergent, longest separated forms. Here we designate the taxa as full species, but recognize that it may ultimately prove appropriate to denote at least some commensal populations as members of subspecies.

Recent investigations have addressed the questions of the genetic make-up of the SC Asian populations and the origin and radiation of house mice by restriction analysis (Boursot *et al.* 1996) and sequencing (Boissinot and Boursot 1997) of mitochondrial DNA (mtDNA); by electrophoresis of proteins encoded by autosomal loci and restriction analysis of three genes on chromosome *6* (Din *et al.* 1996); and by Southern blotting, PCR amplification of a variable length marker and of microsatellites, and sequencing of the *Y* chromosome (Nagamine *et al.* 1992; Boissinot and Boursot 1997). The mice studied came from N and S India, several localities in Pakistan, and N and E Iran. The central populations were found to be highly polymorphic for nuclearly encoded proteins and mtDNA in comparison to the populations recognized as *M. domesticus*, *M. musculus*, and *M. castaneus* from around the periphery of the Eurasian land mass. Most of the mtDNAs fell into a diverse group of types that Boursot *et al.* (1996) and Boissinot and Boursot (1997) call "oriental" (and we call *castaneus*), while some from Iran were *musculus* types. Of two categories of *Y* chromosome, the type found in *M. domesticus* was detected in the Indian and Pakistani mice, while the *Y*s in Iran were of the type found in *M. musculus* and peripheral populations of *M. castaneus*. These molecular and biochemical data provided the foundation for the hypothesis of the northern part of the Indian subcontinent as the cradle of the commensal species, with centrifugal radiations to the west, north, and east giving rise to the peripheral taxa (Boursot *et al.* 1993, 1996; Bonhomme *et al.* 1994; Din *et al.* 1996). Tanooka *et al.* (1995) and Ohtsuka *et al.* (1996), in turn, carried out limited surveys for the presence or absence of a processed *p53* pseudogene (Ψ*p53*) on chromosome *17*. They observed polymorphism in the Central Asian region, in contrast to the invariable presence (in the homozygous state) of this Ψ*p53* in a broad survey of mice recognized phenotypically and genetically as *M. domesticus* and its complete absence in a similar survey of those recognized as *M. musculus* (Prager *et al.* 1997).

In this article, we extend and augment the previously published work in several ways. First, we have filled in genetic "blank spots" on the house mouse map by sampling additional areas—particularly Yemen, Turkey, W and SC Iran, localities throughout Afghanistan, SW as well as N Pakistan, and Nepal. Included are regions, notably Yemen and Nepal, from where anatomical and ecological information is available (*e.g.*, Gruber 1969; Harrison 1972; Marshall 1981; Harrison and Bates 1991), but no molecular work has been done.

Second, our mtDNA study is done by sequencing all or much of the control region and flanking tRNAs, which, relative to restriction analysis (Boursot *et al.* 1996) and sequencing the most variable part of the control region (Boissinot and Boursot 1997), facilitates data analyses involving more distantly related lineages (including those of aboriginal mice), increases resolution, enhances delineation of evolutionary tree structure, and does not require intact high-molecular-weight DNA. In addition to focusing on phylogenetic analyses and biogeographic models, we quantitatively compare the independent duplications of the same tandem repeat.

Third, besides assessing for presence or absence, we carried out a broad survey of sequence variation in a short segment of Ψ*p53*. Fourth, to relate the molecular results to morphologically based categories (*e.g.*, see Marshall 1981; Marshall and Sage 1981), we provide phenotypic and anatomical information for many of the animals we studied.

Finally, our survey of the geographically most interesting areas was carried out largely using museum skins as the DNA source because of the ready availability of specimens from these remote areas. A special value of using museum study skins is that molecular genotypes can be linked to specimens that have been previously classified by taxonomists on the basis of morphological traits conventionally used to define rodent taxa. In addition, these study skins are in public institutions and, thus, available for future analyses by other investigators. Because the DNA in such skins is present in reduced amounts and is generally broken down into small pieces, we used sets of primers that amplify short segments to screen the genetic variability of house mouse specimens. As one must amplify several fragments to sequence the same mtDNA region normally obtained in one or two fragments from total genomic DNA prepared from frozen tissues, our strategy was to sample one or two individuals per locality over a broad range and to survey dozens rather than hundreds of individuals. The markers, *i.e.*, variable sites, we identified among new mtDNA lineages and at a Ψ*p53* locus should facilitate future surveys of variation in house mice from additional localities.

MATERIALS AND METHODS

Specimens: Skin snippets, typically 6 mm^2 per mouse, from 50 of the animals (Table 1, Figure 1) were sent to us in 1991 and 1992 from the Field Museum of Natural History in Chicago. Using ethanol- and flame-sterilized instruments, we cut similarly sized skin snippets from 18 mice in the collection of the Museum of Vertebrate Zoology (MVZ) at the University of California in Berkeley; the 12 samples from mainland China came to the MVZ from the Academia Sinica in Beijing. The Museum of Zoology at the University of Michigan in Ann Arbor sent us frozen tissues of eight Pakistani mice listed by the Field Museum (Table 1); we snipped and extracted them in the same ways as the skin specimens. The mice had been collected during 1951–1954 in Yemen and Turkey, 1961–1975 in Egypt, Iran, Afghanistan, and Nepal, 1990 in Pakistan, and 1945–1978 in eastern Asia. Genomic DNAs, many of them available from previous projects (Prager *et al.* 1993, 1996, 1997), were used along with the skin and tissue extracts to survey the following: (1) types of *Y* chromosomes and (2) presence/absence polymorphism and sequence differences at a ψ*p53* locus. Table 2 provides phenotypic descriptions and measurements of 74 of the commensal mice studied.

Extractions: With sterilized forceps, we rinsed each snippet of skin or tissue through a series of eight 40-μl drops of water before putting it into 250 or 500 μl of extraction solution in a 2-ml screw-cap (for autoclaving) or 1.5-ml locking microcentrifuge tube. Negative controls consisted of (1) sterilized forceps put through the water droplets and then dipped into the extraction tube and (2) untouched extraction solution. Specimens from all 76 individuals were extracted by adding them to a 5% Chelex (Bio-Rad, Richmond, CA) suspension in water, autoclaving for 5 min, and vortexing vigorously for 15 sec. Working stocks containing some Chelex beads were stored at −20°; these sample tubes were vortexed, and the beads were spun down before each PCR. For each 12.5-μl double-stranded amplification of mtDNA and nuclear loci, 1–2 μl of extract was generally used. Fresh snippets of 13 MVZ skins and of the frozen tissues were extracted by a second procedure that, for several skins, markedly improved our ability to amplify at least mtDNA segment 1 (Figure 2) or additional, longer pieces (*e.g.*, segment 4), and for the Pakistani tissues, facilitated amplification of 0.5–0.7-kb fragments. The samples were first heated at 56° for 2 hr in 250 μl of hair lysis buffer, which contains 10 mm Tris-HCl (pH 8.0), 35 mm dithiothreitol, 0.9% Laureth 10 (Macol LA-12; PPG Industries), and 50 μg/ml proteinase K. The tubes were then spun down, 2.5 μl of 10 mg/ml RNase A was added, and the 56° incubation was continued for 1 hr. After the tubes were vortexed, 225 μl of a 5% Chelex suspension in water was added and incubation at 95° was done for 20 min. After centrifugation, 350 μl of the supernatant (without Chelex beads) was removed, stored, and used as the DNA source for PCR as described above.

PCR amplification and sequencing: Figure 2 outlines the strategy for obtaining the sequence of the variable parts of the mitochondrial control region plus flanking tRNAs from extracts of museum skins by amplifying with four pairs of primers. Double-stranded products of segment 2 (the most variable region) from most of the skin specimens from the Field Museum were generated in 25-μl volumes using reactant solution 1 (Prager *et al.* 1993), which has 1 mm of each dNTP and 6.7 mm $MgCl_2$, and adding 1.6 μg of T4 gene 32 protein (United States Biochemical Co., Cleveland, OH; Lessa *et al.* 1992). Amplification was done in a PCR-1000 thermal cycler (Perkin Elmer, Norwalk, CT) for 35–38 cycles of denaturation at 92° for 40 sec, annealing at 60° for 1 min, and extension at 72° for 30 sec. The rest of the double-stranded PCRs were done in 12.5-μl volumes using reactant solution 2 (Prager *et al.* 1997), which has 0.2 mm of each dNTP and $MgCl_2$ at 2.5 mm (primer pairs 1 + 2 and 3 + 4) or 3.5 mm (primers 7 + 9 and 10 + 12); 0–0.8 μg of T4 gene 32 protein or 0.13 μg of *Escherichia coli* SSB (Pharmacia, Piscataway, NJ) was added for segment 1, and 0.2 μg of the T4 protein was added for segments 2 and 4. PCR in a Perkin Elmer 480 cycler was generally carried out for 36–37 cycles; each cycle consisted of 92° for 50 sec (but 3 min for the first cycle), 60° for 45 sec, and 72° for 20 sec (but 3 min for the last cycle). For segment 3, a hot start [as described by Prager *et al.* (1997)] was followed by a "touchdown" procedure: seven precycles, during which an initial annealing temperature of 67° was lowered by 1° for each successive cycle, preceded 36 cycles with annealing at 60°. Amplifications with primer pairs 13 + 14, 13 + 16, 15 + 16, 3 + 16, and 13 + 4 were done for 43–45 cycles, often with a hot start, using reactant solution 2 (2.5 mm $MgCl_2$ for all) and the second cycling protocol given above, except that the annealing temperature was 56° for pairs 15 + 16 and 13 + 4.

For the eight mice from Pakistan, we not only amplified and sequenced segments 1–4, but also amplified the entire region in Figure 2 in two portions, with primer pairs 1 + 6 and 5 + 12, as done previously for genomic DNAs and purified mtDNAs (Prager *et al.* 1993, 1996), and sequenced unidirectionally using primers 1, 3, 8, 9, and 11. Amplification of these two longer fragments from our extracts was appreciably harder than from isolated genomic DNAs. The 5′ portion was amplified from seven individuals with reactant solution 2 (with 2.5 mm $MgCl_2$) and, after a hot start, 45 cycles of 93° for 50

TABLE 1

Collecting localities, mice, mtDNA, and nuclear genotypes

Locality	Mice	mtDNA	*Y* chr	Ψ*p53*
	Egypt			
1 Siwa Oasis, Siwa, Matruh	F121967	*d* 98		P
2 Siwa Oasis, E of Siwa, Matruh	F121966	*d* 28		P
3 Qara Oasis, Qara, Matruh	F101325	*d* 100	A	P
	F101326	*d* 100	A	p[a]
4 Aswan, W bank of Nile R, Aswan	F101654	*d* 101[b]	A	P
5 Aswan, W bank of Nile R, Aswan	F101652	*d* 97		p[a]
6 Kom Ombo, Muneiha, Aswan	F98818	*d* 99		p[a]
	Yemen			
7 Hodeida	F78073	y 6		P
	F78074	y 2	A	P
8 Wadi Maleh, 7 mi E of Ta'izz	F78072	y 3	A	P
9 San'a	F78076	y 1	A	P
	F78077	y 1[c]	A	P
	F78081	y 1		P
	F78082	y 4		P
10 Ma'bar, 2 mi E of	F78075	y 5	ND	P
	Turkey			
11 Lake Emir, Ankara	F74392	*mac* 2		P
	F74393	*mac* 3	NT	P
12 Iskenderun, Hatay[d]	F82208	*d* 107		P
	F82211	*d* 108[e]	A	P
13 Kara Kopru, Urfa	F82241	*d* 105		P
	F82242	*d* 105	A	P
14 Tatvan, Kurtikan, Bitlis	F82204	*d* 106		P
	F82205	*d* 102		P
15 Van, Van	F82214	*d* 102[c]		P
	F82215	*d* 103[e]		P
	Georgia			
16 Batumi, Adzharia[f]	T4567	*d* 68	A	P/N
	T4569	*d* 69	A	P
	T4572	*m* 35		N
	T4575	*d* 68		P/N
	Russia			
17 Sulak, Daghestan	T4558	*m* 23	B	N
	T4562-3	*m* 23		N
	Iran			
18 Orumiyeh, 6 mi SW of, W Azerbaijan	F97541	*d* 104	B	P
19 Kermanshah, 4 mi NW of, Kermanshahan	F112293	*d* 109	A	P
20 Khorramabad, 1 mi S of, Lorestan	F112259[g]	*mac* 4	NT	NT
21 Bandar-e Khomenyi, Khuzestan	F112305	*d* 110	A	P
22 Kuh Rang, 5 mi NW of, Chahar Mahal and Baktiari	F112263	*c* 6	B	P
23 Yasuj, Fars	F112282	*c* 8[h]		P
24 Shiraz, Fars	F112277	*c* 7		P
25 Chalus, Mazandaran	F97537	*m* 37		P/N
26 Gorgan, 10 mi ESE of, Mazandaran	F97559	*c* 16[i]		N
27 Shahrabad-e Kord, Khorasan	F97572	*c* 17		N
	Turkmenistan			
28 Sary-Kamish	T4591	*m* 23	B	P
	T4592	*m* 29		N
	T4593	*m* 25		P
29 Tedzhen	T4600	*m* 21	B	N
	T4601	*m* 21	B	P
	Afghanistan			
30 Herat, Herat	F103698	*c* 20[c]		N

(*continued*)

TABLE 1

(Continued)

Locality	Mice	mtDNA	*Y* chr	Ψ*p53*
31 Maimana, Fariab	F103697	*m* 40	B	P[j]
32 Mazar-e-Sharif, Balkh	F103691	*m* 39		N
33 Faizabad, 16 mi W of, Badakhshan	F103684	*m* 38	B	N
34 Kandahar, 4 mi N of, Kandahar	F103726	*c* 24		P
	F103727	*c* 25[e]	B	ND
35 Kabul, 20 mi W of, Kabul	F103670	*c* 18		P
	F103671	*c* 19	ND	P
36 Jalalabad, Nangarhar	F103712	*c* 9	A	P/N
37 Kamdesh, 5 mi S of, Konar	F103703	*c* 22[c]		ND
	F103704	*c* 23		P
Pakistan				
38 Gwadar, Makran, Baluchistan	F140565	*c* 3		N
39 Nordiz, Makran, Baluchistan	F140564	*c* 21		P
40 Pasni, Makran, Baluchistan	F140567	*c* 2	A	P
41 Dodar, Khuzdar, Baluchistan	F140561	*c* 26[k]	A	P
42 Awaran, Khuzdar, Baluchistan	F140559	*c* 28	A	P
43 Karakar Pass, Swat, Malakand	F140455	*c* 10		P
44 Karakar Pass (Jaba), Swat, Malakand	F140461	*c* 11	A	P
45 Ushu, 6 mi S of, Swat, Malakand	F140552	*c* 27	A	P
Nepal				
46 Jomosom, Mustang	F101734	*c* 12	A	P
47 Phulung Ghyang, Nuwakot	F104287	*c* 14[e]	A	P
48 Thamel, Kathmandu	F104273	*c* 13		p[l]
49 Jamnagaon, Ilam	F94182	*c* 15[c]	A	P
Siberia, Russia				
50 Altai Mountains	T4604-7, 4610	*m* 30	B	N
Northern China				
51 Dengkou, Inner Mongolia	M167362	*m*[m]		NT
52 Chaogegi, Inner Mongolia	M167361	*m*[n]	ND	NT
53 Siziwangqi, Inner Mongolia	M167346	*m*[o]	ND	ND
54 Erlian, Inner Mongolia	M167345	*m* 41[c]	ND	ND
55 Suniteyouqi, Inner Mongolia	M167347	*m* 42		N
56 Abahanaerqi, Inner Mongolia	M167348	*m*[p]		ND
57 Haidian, Beijing	M167343	*m* 43[q]		ND
	M167358	*m*[n]	ND	NT
58 Tongxian, Beijing	M167344	*m* 44[r]		N
59 Zunhua, Beijing	M167359	*m*[n]		NT
	M167360	*m* 45[s]	ND	NT
Other Eastern Asia plus Western Pacific				
60 Seoul, Kyonggi-Do, S. Korea	M123825	*m* 24		N
	M123827	*m* 24	B	N
61 Thon Buri, Thon Buri, Thailand	—[t]	*c* 1	—	—
62 Fuzhou, Fujian, China	M167356	*c* 3[u]		ND
63 Ta-wu-shan, Peng-hu, Taiwan	M152831	*c* 5		P
64 Chian-shan, Peng-hu, Taiwan	M152827	*c* 4	B	P
65 Manila, Rizal, Luzon, Philippines	M109785	*c*[v]		NT
66 Naha, Okinawa, Japan	—[t]	*m* 24	—	—
67 Tinian, Mariana Islands	M104570	*c*[w]		NT

Figure 1 maps the 67 localities listed in this table; mtDNA sequence variation in the mice from localities 16, 17, 28, 29, 50, 61, and 66 has been reported (Prager *et al.* 1996). Letters preceding the identification numbers signify the Field Museum of Natural History (F), the Museum of Vertebrate Zoology (M), and H. Tichy (T). More information about collecting localities is available from the source museums, earlier publications (Prager *et al.* 1996 and references therein), and the authors. Abbreviations for kinds of mtDNAs: *c*, *castaneus*; *d*, *domesticus*; *mac*, *macedonicus*; *m*, *musculus*; y, Yemen. *Y* chromosomes are designated as having an undeleted *Zfy-2* (A) or the 18-bp deletion in *Zfy-2* (B); the Ψ*p53* results (based on PCR with primers Int5S + Int5R) are tabulated as positive (P), negative (N), or heterozygous (P/N) for presence of a *p53* pseudogene. A blank under *Y* chromosome denotes a female (and sex unknown for M123825). ND, not determined; NT, not tested. Except where indicated otherwise

TABLE 1

(Continued)

with footnotes, the mtDNA region between primers 1 and 12 in Figure 2 was sequenced from the mice from localities 38–45 and segments 1, 2, and 4 plus much of segment 3 from the 68 mice where museum skins served as the starting material. In eight cases where limited sequence data were available, diagnostic sites at positions 15333, 15394, 16179, and 16287 facilitated designation as *musculus* mtDNAs, and a diagnostic site at position 15431 facilitated assignment to a *castaneus* mtDNA subcategory.

[a] Possibly heterozygous, perhaps as a result of technical problems. Beginning with total genomic DNA, we scored six more Egyptian mice (three each from Bashtil and Faiyum) as homozygous. There is widespread agreement, based on anatomical and molecular data, that the commensal house mice of Egypt are *M. domesticus*; however, *M. musculus* may have been present in Egypt at least transiently, apparently as a result of human transport (J. T. Marshall, personal communication). For computational purposes (Figure 10, Table 5), we have counted also mice 101326, 101652, and 98818 as having two copies of the Ψ*p53*, though we cannot rule out either actual intra-*M. domesticus* polymorphism for Ψ*p53* P and N or a genetic contribution from non-*M. domesticus* mice as explanations for our observations.

[b] 47 bp within segments 1 and 2 not sequenced.

[c] Segment 3 not sequenced.

[d] A second locality 3 mi northward is included (for individual 82211).

[e] Between 102 and 166 bp of segment 3 not sequenced.

[f] Incorrectly designated as Abkhazia in Prager *et al.* (1996).

[g] The mouse was classified by J. T. Marshall (personal communication) as *M. domesticus* on anatomical grounds (including a long tail of 101 mm and a tail-to-body ratio of 1.16, both uncharacteristic of *M. macedonicus*). mtDNA segment 2 was sequenced but could not be amplified anew 3 yr later; the sequence of segment 1 was viewed as the result of contamination, as it seemed to be a mixture of commensal sequences. We did not attempt to resolve these discrepancies, but we do report the sequence for segment 2, as it is a *mac* sequence distinct from any other we obtained, though it appears not to come from animal 112259.

[h] Segments 3 and 4 not sequenced.

[i] 71 bp of segment 2 not sequenced.

[j] A distinct locus is implicated.

[k] Sequence of 155 bp within the region from positions 15782–16171 not determined.

[l] Two different loci are implicated.

[m] Sequence determined (segment 1 and 101 bp of segment 3) matches *musculus* types 23–25, 32–36, and 41–44.

[n] Sequence determined (segment 1) matches *musculus* 23–28, 32–36, and 41–44.

[o] Sequence determined (segments 1 and 4) matches *musculus* 25, 32–36, 42, and 43.

[p] Sequence determined (segments 1 and 4) matches *musculus* 24 and 41.

[q] Segment 3 and 92 bp of segment 2 not sequenced.

[r] Segment 4 not sequenced.

[s] Only segment 1 sequenced.

[t] Sequence previously determined from purified mtDNA.

[u] Segment 3 and 48 bp of segment 2 not sequenced.

[v] Sequence determined (segment 1) matches *castaneus* types 1 and 3–5.

[w] Sequence determined (segments 1 and 4) matches *castaneus* 1 and 3.

sec (3 min during cycle 1), 60° for 45 sec, and 72° for 20 sec (3 min during cycle 45). The 3′ portion was amplified from four mice using reactant solution 1 and the previous protocol (Prager *et al.* 1993), but with 32 cycles rather than 25, and from three other mice using solution 2, but with the high dNTP and $MgCl_2$ concentrations characteristic of solution 1 and, after a hot start, 36 or 43 cycles of 93° for 50 sec (3 min for cycle 1), 64° for 45 sec, and 72° for 1 min (3 min for the last cycle).

Gel purification of the double-stranded products in 5 μl of the reaction was done in 2% (occasionally 3%) NuSieve agarose as described previously (Prager *et al.* 1993); some or all of the band with the amplified fragment was diluted 2- to 40-fold in water. PCR to yield single-stranded templates for sequencing in both directions was done in 25-μl volumes under a variety of conditions (details available from the authors). Segment 3 in Figure 2 proved to be the hardest from which to obtain templates amenable to sequencing, particularly in the direction of excess primer 7, and we did not sequence the segment fully from any individual. Nearly all 50 skin samples from the Field Museum worked well for PCR and sequencing. In contrast, 14 of the 18 from the MVZ (all but those from Korea and Taiwan) were harder to amplify; from eight, we could sequence only segment 1 or segments 1 and 4 (see Table 1).

Double-stranded amplifications of a short segment of the duplicated *Zfy-1* and *Zfy-2* genes on the *Y* chromosome (with a hot start and 45 cycles for the museum skin and tissue extracts) and of two short segments of a Ψ*p53* plus one of the functional *p53* gene (with 37 cycles or a hot start followed by 42 cycles for the skin and tissue extracts) were done as described by Prager *et al.* (1997). The *Y* primers, Zfy2DF and Zfy2DR, yield PCR products of 184 and 202 bp and bracket the 139- or 157-bp region extending from the second position of codon 467 through the second position of codon 519, with codons 480–485 deleted in *Zfy-2* in one type of *Y*. Ψ*p53* and *p53* primer pair Int5S + Int5R brackets the 89- or l67-bp region extending from the third position of codon 182 to the first position of codon 212, with codons numbered according to the cDNA sequence of the functional gene; the size differ-

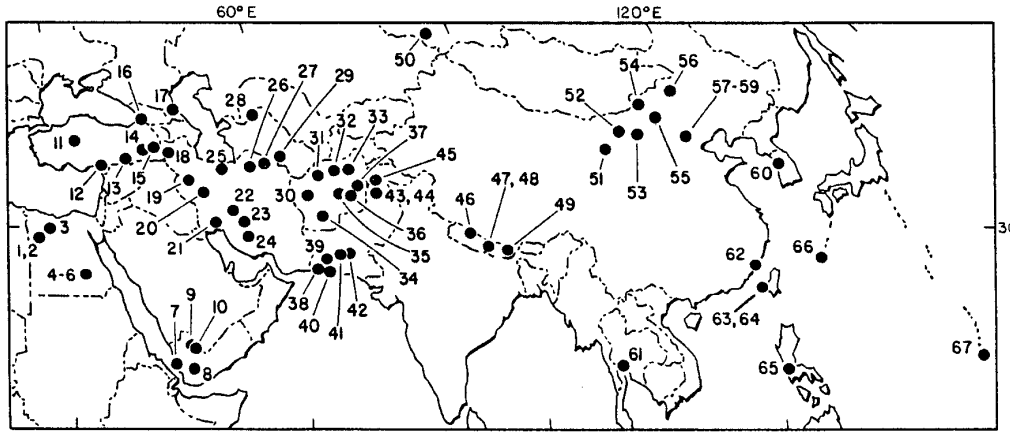


Figure 1.—Map showing 67 collecting localities for mice, as numbered in Table 1. Locality 52 is placed roughly because its latitude and longitude could not be obtained.

ence is caused by the 78-bp intron 5 in *p53*. As the Ψ*p53* and *p53* PCR products of 137 and 215 bp are close in size, one can score presence or absence of Ψ*p53* while confirming successful PCR by appearance of the *p53* product, and can usually also distinguish between individuals homozygous and heterozygous for Ψ*p53* (Prager *et al.* 1997). Primers Exon 4 and Exon 5 bracket a 128-bp piece of the Ψ*p53* in commensal mice and a 133-bp piece in *M. macedonicus* and *M. spicilegus*; these extend from the third position of codon 109 to the third position of codon 153, and the PCR products are 176- or 181-bp long. We tested one or both Ψ*p53* primer pairs on genomic DNA of nine *M. spretus* (four from Catalunya and two from Puerto Real in Spain plus three from Azrou, Morocco) to confirm the previous inference, based on one Spanish mouse (Tanooka *et al.* 1995; Ohtsuka *et al.* 1996), that this species lacks a Ψ*p53*. Gel analysis and purification of PCR products in 3% NuSieve agarose were done as described before (Prager *et al.* 1993, 1997). Single-stranded templates for sequencing were made (details available from the authors) in one direction from the shorter *Y* chromosome fragment (for sequencing with primer Zfy2DR), in both directions with primers Exon 4 and Exon 5, and in one or both directions with primers Int5S and Int5R.

Desalting of templates, which were generally resuspended in 15 μl of water, and dideoxy sequencing were done as described before (Prager *et al.* 1993), except that half volumes were used for the sequencing reactions and usually only wedge gels were required. Segments 1–4 in Figure 2 total 820–835 bp in most of the mtDNAs and 898 bp in those bearing a tandem 76-bp repeat. For mice where museum skins were the starting material, we read for $n = 58$ an average of 744 bp (range, 385–890 bp), and for the $n = 10$ worst results, an average of 233 bp (range, 160–354 bp). Starting with the frozen tissues, we read (of totals of 1043 or 1119 bp) an average of 1059 bp (range, 964–1119 bp; $n = 6$ had no unread sequence). GenBank accession numbers for the 59 new mtDNA sequences we determined are AF074490–AF074548.

Y chromosome sequences of the 139-bp segment (average of 126 bp read) were determined to see whether the same 18 bp had been deleted in *Y*s from diverse areas. The mice assessed were 13 of the 16 with the B type of *Y* in Table 1 (all but that from locality 34 and two from locality 50) plus one each from Croatia, Moldova, and Ukraine, and two from Germany. The GenBank accession no. for the variant sequence found is AF074549.

An average of 126 bp was read for a 128-bp Ψ*p53* fragment flanked by primers Exon 4 and Exon 5 ($n = 79$ commensal mice; localities and individuals detailed in Figure 10). Complete 133-bp sequences (which match the functional gene) inferred to come from a separate Ψ*p53* locus were determined from two commensal mice; to obtain this slightly longer sequence from a mouse yielding both bands, with heteroduplex formation and/or trailing of the shorter fragment in the area of the longer one, we subtracted out the bases found in the shorter piece. The mice and localities that yielded each of the five distinct sequence phenotypes (see results) obtained by sequencing 133 bp (average of 129 bp read; $n = 9$) from aboriginal mice at the locus, designated Ψ*p53-1*, that is shared with most commensal mice are as follows: (1) two *M. macedonicus* from Gradsko, Macedonia, and one from Turkey (no. 74392), plus a *M. spicilegus* from Halbturn, Austria; (2) one *M. spicilegus* from Debeljaca, Serbia, and one from Kishinev, Moldova; (3 and 4) each in one *M. spicilegus* from Srpska Mitrovica, Serbia; (5) one *M. spicilegus* from Debeljaca. By sequencing between primers Int5S and Int5R, we defined one 89-bp sequence for this second segment of Ψ*p53-1* (in Georgian mouse 4569 plus one from Bokhorst, Germany) and two 167-bp sequences for the equivalent part of the functional *p53* (from the data for two German mice from Burg and Dannau). GenBank accession numbers for the 24 Ψ*p53* and two *p53* sequence phenotypes we obtained are AF074551–AF074576.

Calculations: We made use of the 139 published Mus mtDNA sequences for this 1-kb region included by Prager *et al.* (1996): *domesticus* types 1–96, *musculus* types 1–36, *castaneus* and *macedonicus* types 1, *spicilegus* types 1–3, and *spretus* types 1 and 2. Because segments 1–4 encompass almost all the known intracommensal mtDNA sequence variation in the whole control region and flanking tRNAs (Figure 2), we assumed for all sequences considered here a length of 1000 bp for computations of nucleotide variability, which was estimated with the parameters θ and π as before (Nachman *et al.* 1994; Prager *et al.* 1996). This assumed length is very close to the averages read by Prager *et al.* (1996 and references therein), starting with total genomic DNA or purified mtDNA.

Character-state parsimony trees for mtDNAs were constructed with the PAUP (Phylogenetic Analysis Using Parsimony) version 3.0s program with a heuristic search procedure and equal weighting of all character changes, as described in detail previously (Prager *et al.* 1993, 1996). As before, smaller subsets of a given dataset (notably that of 110 *domesticus* mtDNA sequences) were analyzed with PAUP to examine all most-parsimonious arrangements in various sections of the tree and to root trees and relate major commensal mtDNA lineages to one another (see also legends to Figures 5–8). Neighbor-joining mtDNA trees were constructed with the PHYLIP 3.572c program from matrices of pairwise differences computed after weighting transversions five times as heavily as other changes, as well as from matrices of unweighted differences. *M. spretus* mtDNAs served as the outgroup to those of all the other taxa (*cf.* Prager *et al.* 1996).

TABLE 2

Phenotypes and anatomical measurements of commensal house mice

Category, description, and mice	No. of mice and measurements
Long-tailed mice	
1 Dark brownish-gray dorsal and ventral	
Turkey (12, 14, 15)	$n = 6$[a,b], 191 (175–208), 97 (86–110), 1.03
Iran (18, 26)	$n = 2$, 178 (177–178), 84 (83–84), 0.89
2 Brown with reddish tint, whitish belly	
Yemen (7–10)	$n = 8$[c], 140 (119–156), 69 (58–76), 0.97
Iran (25, 27)	$n = 2$, 162 (161–162), 76 (74–77), 0.88
Afghanistan (37)	$n = 1$[d], 177, 91, 1.06.
3 Sand-colored, pure white belly	
Egypt (1–6)	$n = 7$[b,c], 163 (148–179), 81 (73–89), 0.99
Turkey (13)	$n = 2$, 169 (158–180), 81 (76–86), 0.92
Iran (19, 21–24)	$n = 5$[b], 164 (155–172), 81 (73–90), 0.98
Afghanistan (30, 34–37)	$n = 7$[d], 166 (142–186), 78 (63–87), 0.89
Pakistan (38–42)	$n = 5$[e], 160 (142–171), 77 (70–83), 0.93
4 Dark gray, white belly (with slate gray bases to fur)	
Pakistan (43–45)	$n = 3$, 161 (155–167), 82 (78–86), 1.04
Nepal (46, 47, 49)	$n = 3$, 149 (137–158), 77 (74–79), 1.07
5 Chestnut brown dorsal and ventral	
Nepal (48)	$n = 1$[f]
Eastern Asia + Western Pacific (61[g]–65, 67)	$n = 6$, 152 (143–170), 74 (71–80), 0.95
Short-tailed mice	
6 Sand-colored, pure white belly	
Afghanistan (31–33)	$n = 3$, 152 (141–163), 68 (67–71), 0.81
Northern China (51–54)	$n = 4$, 114 (96–133), 50 (45–55), 0.78
7 Dark gray, whitish belly, unusually short tail	
Northern China (55–56)	$n = 2$, 128 (120–135), 45 (40–50), 0.54
8 Brown with reddish tint, whitish belly	
Northern China (57–59)	$n = 5$, 125 (124–127), 57 (55–61), 0.84
South Korea (60)	$n = 2$, 129 (111–146), 51 (45–57), 0.65

For each of the eight phenotypic categories, which are divided between long-tailed (categories 1–5) and short-tailed (categories 6–8) mice, the table provides a brief description of the coat coloration and indicates the mice from Table 1 assigned to each category. After each region in a category, the following are listed in order: locality numbers, number of mice (n), average total length in millimeters (and range), average tail length in millimeters (and range), and average tail-to-body ratio (where body includes the head). Individual identification numbers, as well as information on mice assessed as intermediate between two categories, are given in footnotes where necessary. This table is based in part on personal communications from J. T. Marshall, who included skull traits among the anatomical criteria. A short tail and a low tail-to-body length ratio are characteristic of *M. musculus*; a long tail and high ratio are characteristic of the other commensal species. The name *M. bactrianus* has frequently been applied to the mice in phenotypic category 3 from Iran, Afghanistan, and Pakistan, and probably to others from the same geographic area. The mice in category 4 have generally been designated as the *homourus* subspecies (of *M. domesticus*), while members of category 5 correspond to *M. castaneus*. The remainder of the long-tailed mice are generally designated as *M. domesticus*. Coat color differences are viewed as adaptations to the environment in several instances, notably light colors in deserts (see categories 3 and 6). The shape of the anterior border of the skull's zygomatic plate (a craniofacial feature that is relatively easily compared) is straight and vertical in *M. domesticus*, forward leaning in *M. castaneus*, and convexly curved in *M. musculus* (see Marshall 1981, 1986; Marshall and Sage 1981), but the neotenic state of this trait is rounded (*i.e.*, as in *M. musculus* adults) in all the commensal taxa (J. T. Marshall, personal communication).

[a] Individuals 82204 and 82208 had broken tails; $n = 4$ for the length calculations. Individual 82205 was regarded as intermediate between categories 1 and 2.

[b] Individuals 82211 (Turkey), 101326 (Egypt), and 112293 (Iran) were regarded as intermediates between categories 1 and 3.

[c] Individuals 78073 and 78076 from Yemen and 101325, 101652, and 101654 from Egypt were regarded as intermediates between categories 2 and 3.

[d] Individuals 103704 and 103703 from locality 37 were assigned to categories 2 and 3, respectively.

[e] Individual 140565 (locality 38) was regarded as an intermediate between categories 3 and 5 because its skull was characteristic of *M. castaneus*.

[f] Broken tail; length of head plus body totals 70 mm.

[g] The mouse from Thailand was MVZ 154449, caught at the same locality as the progenitors of the lab animal from which we prepared the mtDNA (Ferris *et al.* 1983).

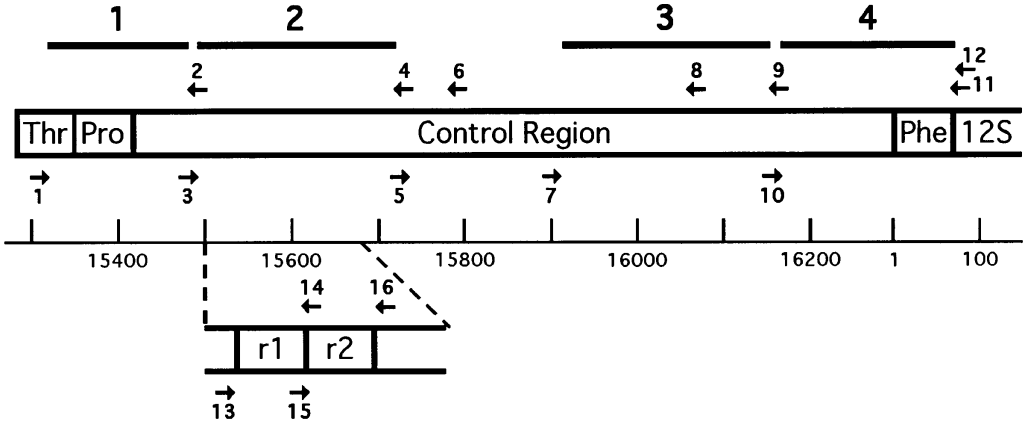


Figure 2.—Strategy for amplification and sequencing of 0.8–0.9 kb of the control region and flanking tRNA genes of mouse mtDNAs retrieved from museum skins. Arrows denote primers, bars 1–4 indicate the individual segments amplified, and r1 and r2 represent, respectively, the 5′ and 3′ tandem repeats. Three-letter abbreviations stand for the tRNA genes; 12S is the small ribosomal RNA gene. Nucleotide positions are numbered throughout this report according to the *domesticus* type 1 sequence as described previously (Prager *et al.* 1993, 1996). The basic strategy was to amplify segments 1–4 with primer pairs 1 + 2, 3 + 4, 7 + 9, and 10 + 12, and to sequence single-stranded templates generated from both strands with the PCR primers used as sequencing primers, except that primer 11 was substituted for primer 12. Additional primer pairs (*e.g.*, 13 + 14, 15 + 16) and internal sequencing primers (*e.g.*, 13–16) were sometimes used. Apart from occasional length variants, the sizes (between primers) of amplified segments 1–4 are in order 160, 224–301, 242–243, and 194 bp. Primers 5, 6, and 8 were used during amplification in two portions and sequencing of the entire 1.0–1.1-kb region from extracts of frozen tissue according to strategies described previously (Prager *et al.* 1993, 1996). The region between segments 2 and 3 is totally invariant among all reported commensal mouse mtDNA sequences; the 12–18 bp where primers 2 and 3 and primers 9 and 10 overlap are conserved, except for three positions, each of which is variant in one *domesticus* or *musculus* mtDNA, and a fourth position that is variant in two *castaneus* mtDNAs (Prager *et al.* 1996 and references therein; Boissinot and Boursot 1997; this report). Primers 3, 5, 6, 8, 9, 11, and 12 correspond, respectively, to primers 2–5, 7, 8, and 9B of Prager *et al.* (1993), and primer 4 corresponds to H15720 of Prager *et al.* (1996). Locations (L, light strand; H, heavy strand; numbers representing positions of the 3′ base) and 5′-to-3′ sequences of the other primers are as follows: 1, L15320, ATTACTCTGGTCTTGTAAACC; 2, H15481, ATGTACTTGCTTATATGCTT; 7, L15911, GTGGTGTCATGCATTTGGTAT; 10, L16171, TTAACTATCAAACCCTATGT; 13, L15537, GGTCATAAAAYAACYATCAACA; 14, H15612, TCATGRTGTATATCAGTT TAGTYA; 15, L15538, AAGACATACCTRTRTTATCTRACT; 16, H15616, AGAGTTTATGACTGTATGGTGTAT.

For the reasons discussed by Prager *et al.* (1996), we assumed the likeliest base at missing variable sites for tree construction and computation of pairwise differences; as argued before, the likelihood of an incorrect assignment is often low and the consequences are in most cases expected to be minor. For the sequence types newly defined here, we do not expect the assumptions made for unsequenced sites to have an effect on any substantial inferences, except perhaps with respect to *castaneus* types 14 and 15. The specific assumptions made beyond those in Prager *et al.* (1996) are as follows: For *musculus* mtDNA types 37–44: as in type 1 at all missing sites. For *domesticus* types 97–110: T at position 15912 in types 97–99 and 102–110, C at position 16012 in type 99, as in type 1 at all other missing sites. Among *castaneus* types 2–28: as in type 1 at all sites missing in types 6, 7, 12, 14, 16, 17, and 19; G at position 15958 in type 23; types 15, 20, 22, and 25 taken as matching, respectively, types 14, 21, 23, and 24. For the mtDNAs in Figure 4, except *castaneus* 8 (for which we made no assumptions), any additional missing sites beyond the 94 sites in that figure were assumed to match *castaneus* type 1; *macedonicus* types 1–3 were assumed to match at all sites missing in any of the three sequences.

Analogous to the procedures described for *musculus* mtDNA types 32–36 (Prager *et al.* 1996), variable positions within the tandem repeats in *castaneus* types 16–28 were entered into the computer only once for PAUP analyses, and seven events were added by hand after tree construction: at position 15548, T to A in the 5′ copy of type 23; at 15550, T to C in the 5′ copy of type 24 or in the 3′ copy of type 25; at 15554, T to C in the 5′ copy along the lineage leading to the clade of types 16–28 and C to T in the 5′ copy of type 22; at 15569, T to C in the 5′ copy of type 28; at 15581, A to G in the 5′ copy of type 27; at 15601, T to C in the 3′ copy of type 18.

Boissinot and Boursot (1997) reported the sequences of mtDNA positions 15443–15742 from 131 commensal mice. Among the 71 mtDNA types they defined, 62 are distinct from the 189 collectively defined by us and Nachman *et al.* (1994). Their segment of 297–374 bp (allowing for length variants and tandem repeats) encompasses segment 2 and part of segment 1 in Figure 2, and includes the most variable part of the control region (Prager *et al.* 1993, 1996). Sequencing this 0.3-kb fragment is likely to detect much of the diversity among the mtDNAs examined, but it lacks many of the variable sites that provide structure and define clades in our parsimony trees based on the 1-kb region in Figure 2. We, therefore, added the Boissinot and Boursot (1997) mtDNAs to the trees in Figures 5A, 6, and 7 by hand after tree construction. Their 29 *castaneus* mtDNAs could be placed with appreciable confidence, so we show them explicitly in Figure 6; placement of their 17 *musculus* (Figure 5A) and 25 *domesticus* (Figure 7) mtDNAs is described in the figure legends. We preface the Boissinot and Boursot (1997) mtDNA type numbers with the letter B, except within Figure 6.

RESULTS

Mitochondrial DNA sequences: Among the 76 newly studied mice from 60 localities, we resolved 61 distinct sequences (Table 1, Figures 3 and 4); 57 of them correspond to types of mtDNA not seen in earlier surveys (Prager *et al.* 1993, 1996; Nachman *et al.* 1994; Boissinot and Boursot 1997). The new types are assignable to four previously recognized clades (*i.e.*, *domesticus*, *musculus*, *castaneus*, and *macedonicus*) and one distinctive new clade (see below). Two types we saw before were

	11111111111111111111111111111111100	
	55555555555555555555566666666666600	
dom	33444555555555555557900000002222200	
mtDNA	66379011223444477891101556995566 7815	
type	38143438560156729278292689330248 1985	
	ab	
1	GTCTTATTATCAATAAACCACAATCC--GAATTTAG	
97	....A....C......GTT.?.......A..C...A	
98	....A....C..G.....T.?.........GC...A	
99	....A..........T..T.?.?........C...A	
100	A...A.C....G...T....T.C........C...A	
101	....A.....A....T..T?T.C........C...A	
102	....A...G....--...TT?T....CCAG.C...C	
103	....A........--...TT?T....C.AG.C...C	
104	....A...G....--...TT?T......AG.C...C	
105	....A....C...--...T.?........G.CCATC	
106	....A..A....G--.....?.....CC.G.C...C	
107	....A.............T.?..CT......C...C	
108	....A.............T.???.T?.....C...C	
109	.-TCAG.......--..TT.?..........C...C	
110	....A...............?....TCC...C...C	

	11111111111111111			111111111111111
	55555555555556666			555555555556666
mus	34445555555561222		*mac*	345555555560011
mtDNA	51471223455687459		mtDNA	671124477840900
type	47908262802819523		type	855754503837311
				a c b a
1	GTTCCACTTCACAGTAC		1	-CAT-ATATCCCCCT
37	....T...C...G....		2	.T.C.G..CT.T-TC
38	.AC....CCT...A.G.		3	TTCC...T...T-TC
39	.AC..T.CCT...A..A		4	??CCC.A.C.T????
40	AAC....CCT...A...			
41	.C....T...G...C..			
42	.C......CT.......			
43	.C......C...?....			
44	.C.?T....T.T.????			
45	.A.T?????????????			

Figure 3.—Variable sites in newly described types of *domesticus*, *musculus*, and *macedonicus* mtDNAs. For each category, the polymorphic sites are listed vertically across the top; lowercase letters denote sites added 3′ of the indicated nucleotide relative to the baseline (*domesticus* 1) sequence. The sequence for each type 1 (previously determined) is shown at all sites; the bases in the other types are shown only where different. —, deletions relative to other sequences; ?, unsequenced sites. Length changes involving >1 bp have been counted as single sites. Among the 110 distinct *domesticus* mtDNAs in Figure 7, 107 sites are variable; 7 more vary among the 20 additional distinct *domesticus* mtDNAs identified from the partial sequences reported by Boissinot and Boursot (1997). Among *musculus* mtDNA types 1–45 (see Figure 5A), 55 sites are variable (with position 15470 uniquely variant in type 45, which is omitted from all analyses), and 10 more vary among the 15 additional distinct *musculus* mtDNAs identified from the partial sequences in Boissinot and Boursot (1997).

dom 28 and *mus* 24; in addition, each of the partial sequences B127 and B136 matches one or two of our *castaneus* types (see Figure 6). For eight animals, our fragmentary sequences allowed classification to the *musculus* and *castaneus* mtDNA categories, but not designation of specific mtDNA types (Table 1).

Our survey revealed *domesticus* mtDNAs in 18 mice—all the commensals from Egypt and Turkey plus three from Iran. The results for the Egyptian mice concur with previous mtDNA (*e.g.*, Ferris *et al.* 1983; Prager *et al.* 1993) and protein electrophoretic (Sage 1981) evidence as well as phenotypic classification. They supplement the earlier mtDNA work on specimens from NE Egypt by documenting *domesticus* mtDNAs in the NW and SE parts of the country. In what appears to be the first mtDNA characterization of Turkish mice, we detected six *domesticus* mtDNAs from four localities in the country's SE quarter, from sea level on the eastern Mediterranean to the mountains bordering Lake Van. This study also marks the first report of *domesticus* mtDNA in Iran, which we found at localities 18, 19, and 21 along the western border, the first two in the Zagros Mountains and the third near the Persian Gulf.

Seventeen of the newly surveyed mice had *musculus* mtDNAs, 13 of them from areas in East Asia known to harbor *musculus* mtDNAs (Yonekawa *et al.* 1988; Nagamine *et al.* 1994b). We found *musculus* mtDNA in NC Iran, at locality 25 on the Caspian Sea, which is consistent with recent detection of *musculus* mtDNAs in E Iran (Boissinot and Boursot 1997). Our study is the first report of *musculus* mtDNA in Afghanistan, which we found at localities 31–33, extending some 500 km across the northern edge of the country, just north of the great central mountain range.

Seven of the newly surveyed mice had sequences (*cas* types 1–5; Figure 4, Table 1) very similar to *castaneus* type 1 known from Thailand. Four of these mice came from Taiwan, SE mainland China, and the Philippines, areas where such *castaneus* mtDNAs are well known (Yonekawa *et al.* 1988; Nagamine *et al.* 1994b; Boursot *et al.* 1996; Boissinot and Boursot 1997); the *M. castaneus* animal from the Mariana Islands also had such a *cas* mtDNA. Types 2 and 3 at localities 38 and 40 on the SW Pakistan coast are the first report of this kind of mtDNA in that country. Transport beyond a natural range by humans via shipping merits consideration. However, as Boursot *et al.* (1996) and Boissinot and Boursot (1997) have documented such *castaneus* mtDNAs in SW and NC India (see Figure 6) and the subspecies (*castaneus* and *tytleri*) of *M. castaneus* are known from much of India, finding this kind of mtDNA in SW Pakistan may not be surprising. The pelage of these two Pakistani mice is not characteristic of *M. castaneus*, but the skull of one of them is (Table 2).

We found a diverse collection of mtDNAs denoted *castaneus* types 6–28 among 23 mice from Central Asia: Iran, Afghanistan, Pakistan, and Nepal. Among the mice with such mtDNAs are those from localities 35–37, which are in the general area of Kabul, and Pakistani localities 43–45, which are in the general area of some of those in the Boursot *et al.* (1996) and Boissinot and Boursot (1997) surveys; the remainder represent previously unsampled areas. As reported in a preliminary account (Prager *et al.* 1996), types *cas* 16–28 have a second,

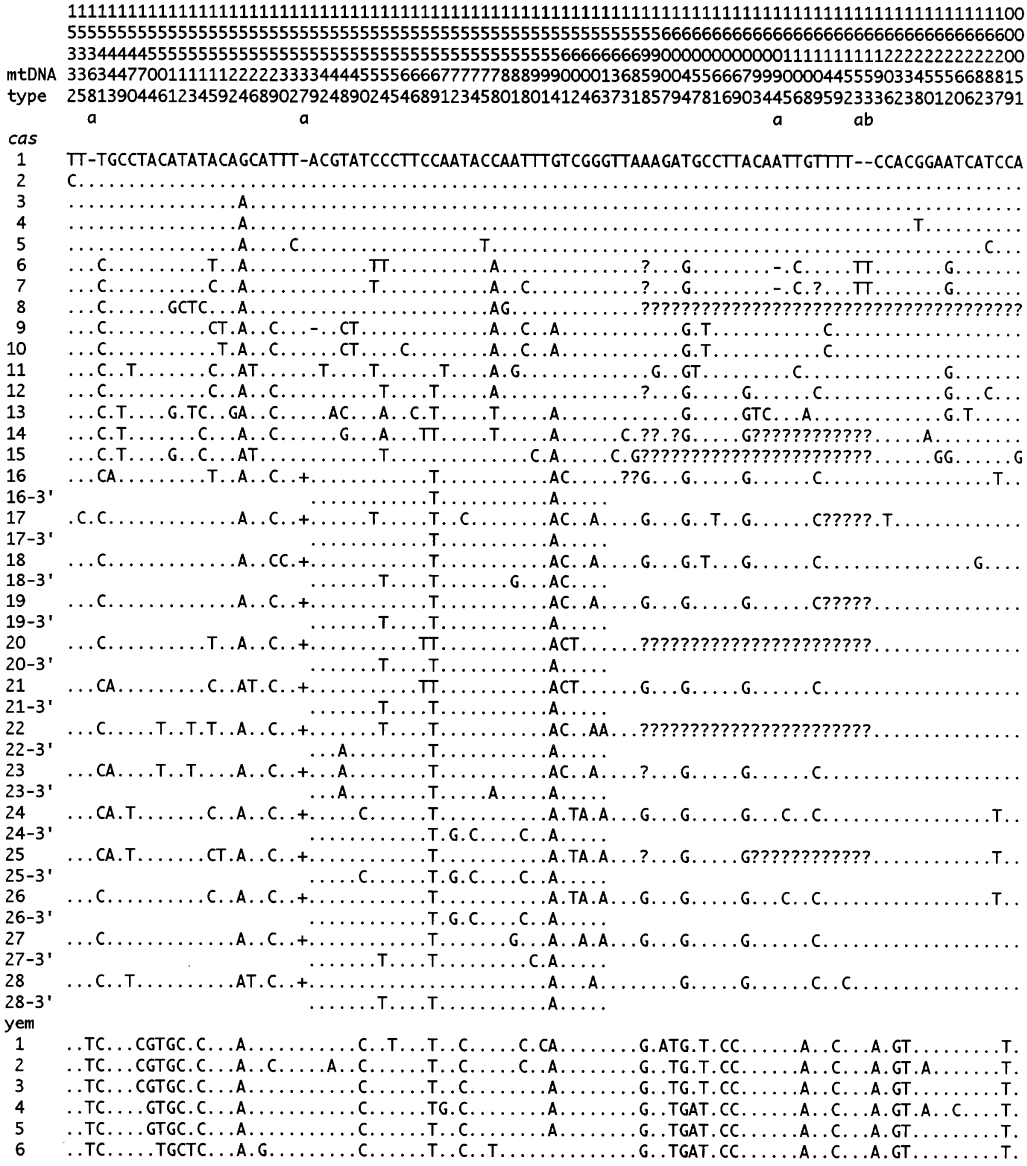


Figure 4.—Variation at 94 polymorphic sites among 34 types of *castaneus* and Yemeni mtDNA sequences shown in the format of Figure 3. The + at 15537a indicates a tandem 76-bp repeat of the sequence from 15538 to 15615; the variation in the 3′ repeat appears on separate lines designated 3′. Among the 28 *castaneus* types, 78 sites are variable, and 16 more vary among the 27 additional distinct *castaneus* mtDNAs identified from the partial sequences in Boissinot and Boursot (1997) (see Figure 6). The tandem repeats in *castaneus* types 16–28 are 76-bp long, rather than the 75 bp in *musculus* types 32–36 (Figure 5A), because all *musculus* mtDNAs have a 1-bp gap at positions 15570–15572. Among the six Yemeni mtDNAs, 16 sites vary.

tandem 76-bp copy of a control region segment that is independently duplicated in the *musculus* 32–36 clade of mtDNAs (*cf.* Figure 5A). Within a given mtDNA type, the repeats differ by one to seven base substitutions and, by several criteria (see below), are considerably more diverse than those of the *musculus* mtDNAs. Boissinot and Boursot (1997) have also documented the independence of the *musculus* and *castaneus* duplications.

We use the name *castaneus* for mtDNA types *cas* 1–28 [and the 29 phylogenetically related types from Boissinot and Boursot (1997)], even though few of the mice bearing these mtDNAs, especially outside the clade of types 1–5, have been called *M. castaneus* on phenotypic and morphological grounds (Table 2). There are two reasons to apply one name to all these mtDNAs: first, mice bearing mtDNAs in the shallow clade with types 1–5 are intermixed throughout the Indo-Pakistan area with mice bearing types outside this clade (see Figure 6), which suggests they belong to the same interbreeding population and are connected by considerable amounts of gene flow. Second, these mtDNAs constitute a phylogeographic unit. Boursot *et al.* (1996) and Boissinot and Boursot (1997) have also recognized the apparent unity of this group, but with the name "oriental." We prefer *castaneus* because it follows the heretofore used protocol of describing gene lineages with names derived from species names of the mice and was already applied to type 1.

Prager *et al.* (1996) reported both *domesticus* and *musculus* mtDNAs in SW Georgia (locality 16), which is consistent with the Orth *et al.* (1996) inference of a broad area of secondary contact and remixing of genomes in Transcaucasia. To the countries with different major lineages of commensal mtDNAs we can now add Iran [also from the results of Boissinot and Boursot (1997)] and Afghanistan.

Most remarkable in our present survey are the six mtDNAs of Yemeni mice (Figure 4). They are similar

and clearly related to one another (pairwise differences of 2–11 bp) but rather different from all the other kinds of mtDNAs of commensal mice (pairwise differences of 24–47 in Table 3 below). Thus, the Yemeni mtDNAs represent a major new lineage from part of the house mouse range previously unexplored at the molecular level. Relevant to our findings, the mice in the southern portion of the Arabian Peninsula were given a distinct subspecific or racial name, *M. m. gentilulus* [Harrison 1972; Harrison and Bates 1991; *M. d. gentilulus* in Marshall and Sage (1981)], in light of their being so conspicuously smaller that Harrison (1972) called them pygmy mice. The Yemeni animals are clearly the smallest long-tailed mice we studied (Table 2). Nine mice from eight nearby localities, to the south and east of ours, had similar traits—with averages (and ranges) for total length, tail length, and tail-to-body ratio, respectively, being 134 mm (111–161), 69 mm (63–83), and 1.07 (0.80–1.31)—as was true also for mice assigned to this taxon from Oman on the SE tip of the Arabian Peninsula and from Bahrain on the Persian Gulf (Harrison 1972). The cranial measurements of the *M. (m.) gentilulus* mice seem even more distinctively small, relative to the mice from the northern Arabian Peninsula and Mesopotamian areas assigned to *M.* (*m.* or *d.*) *praetextus*, than do their external dimensions (Harrison 1972).

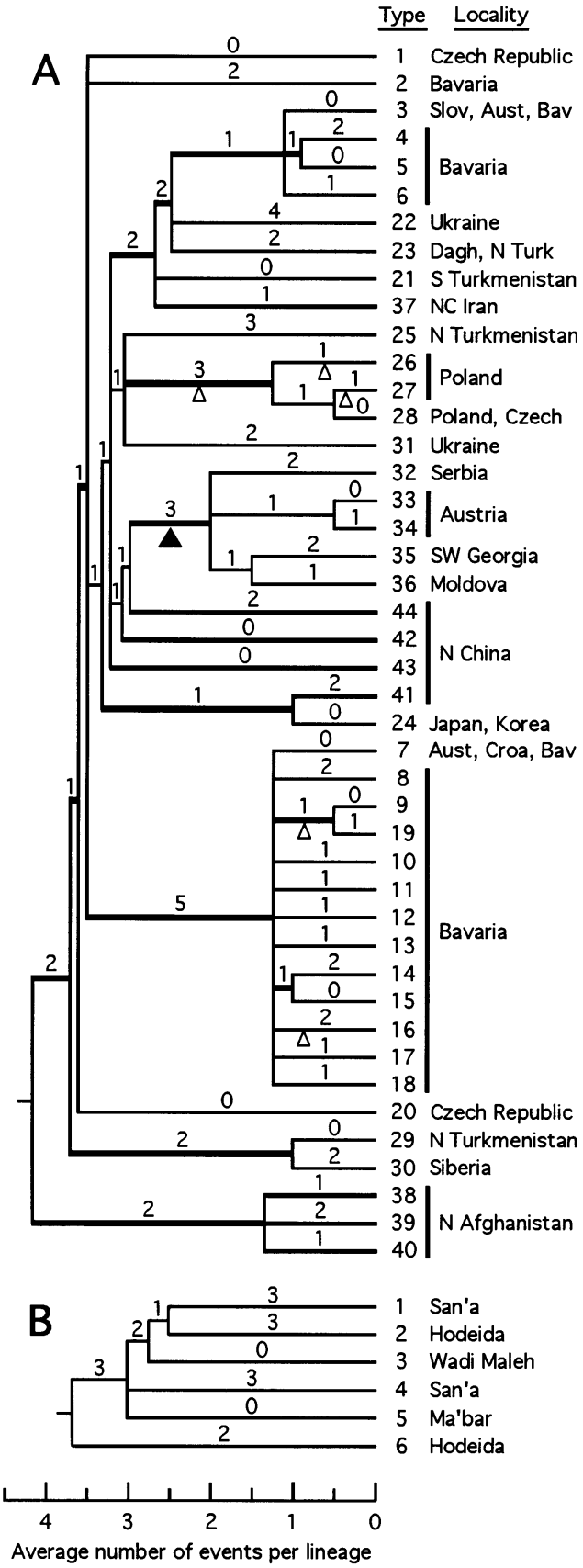
Evolutionary trees and diversity of mtDNAs: Figure 5A presents a rooted parsimony tree relating 44 *musculus* mtDNAs. The present tree differs from the one for *musculus* types 1–36 (Prager *et al.* 1996) in two conspicuous ways: first, it has a new basal clade that is made up of Afghan types 38–40. That the deepest lineage stems from Afghanistan and the next-deepest clade is also from Central Asia accords well with a model [*e.g.*, see Figure 4 of Boursot *et al.* (1996)] postulating the original homeland of *M. musculus* and the start of intraspecific divergence in or near this northern fringe of Afghanistan. Our results for nuclear loci (see below) along with their short tails (Table 2) suggest that these mice are authentic *M. musculus* rather than the products of mtDNA introgression into another species. Second, the average depth of the tree in Figure 5A is ~4.2 events per lineage, 20% deeper than the tree in Prager *et al.* (1996) and close to two-thirds that shown for 110 types of *domesticus* mtDNAs in Figure 7, contrasted to the earlier relative value of about half inferred for the tree for 36 *musculus* mtDNAs *vs.* that for 96 *domesticus* mtDNAs (Prager *et al.* 1996). If we assume that the deepest split among commensal species occurred 350,000–900,000 years ago (She *et al.* 1990; Boursot *et al.* 1993, 1996) and that this split corresponds to the deepest node among commensal mtDNA lineages (at the base of the tree in Figure 8), the implication is that the *musculus* mtDNA lineages examined could have shared a common ancestor some 70,000–180,000 years ago.

Figure 5B shows the most parsimonious rooted tree for the six types of mtDNA from Yemen. The eye-catching feature of the Yemeni tree is that, with a depth of ~3.7 events per lineage, it is nearly as deep as the *musculus* tree in Figure 5A even though it is derived from ~5% of the number of specimens represented in the *musculus* tree. One implication is that the mitochondrial lineages in a limited part of the Arabian Peninsula might have begun diverging nearly as long ago (perhaps 60,000–160,000 years) as did the lineages for extant *musculus* mtDNAs over their entire range of northern Eurasia. The θ and π values in Table 3 suggest that the mice in Yemen are mitochondrially ~60% as variable as is *M. musculus*, an inference supported by the relative ranges of pairwise differences (notably 0–1 *vs.* 0–5 transversions and 0 *vs.* 0–3 length changes). An expectation, also in light of our evolutionary model (see discussion), is that sampling from additional localities on the southern Arabian Peninsula (Harrison and Bates 1991) would reveal more lineages, including deeper ones, in this newly described major branch of the commensal mtDNA tree.

Figure 6 presents a parsimony tree constructed for the 28 *castaneus* mtDNA sequences in Figure 4 and also shows placement of the Boissinot and Boursot (1997) *castaneus* sequences. The tree for 28 sequences has a transition-to-transversion ratio of 4.2, a value lower than those of 5.5 and 6.6, respectively, for the trees in Figures 5A and 7 and indicative of greater sequence divergence. The average depth of the tree in Figure 6 of ~10.6 events per lineage is, respectively, ~2.5 and 1.7 times as deep as those for *musculus* and *domesticus* mtDNAs. The implication is that the mtDNA lineages in Figure 6 began diverging from one another some 170,000–460,000 years ago. The values in Table 3 suggest that these mtDNAs exhibit at least as much genetic diversity as do the *domesticus* mtDNAs.

Members of the shallow clade of *cas* 1–5 and related types (Figure 6) are found across the range of mice designated *M. bactrianus* and *M. castaneus*, from SW Pakistan through NC India to Taiwan, but the southeastern mice have only this category of mtDNA. One possibility is that ships moved mice with this mtDNA lineage around the area and that this lineage is the dominant one in SW Pakistan and SW India. Another interpretation is that *M. castaneus* only recently spread into extreme SE Asia. This latter hypothesis invokes filtering out of the mtDNA diversity from the core Indo-Pakistan area as the mice moved through patchy habitats into E India and SE Asia. Sage and Wolff (1986) have shown how such repeated colonization events lead to erosion of genetic diversity in peripheral populations. Under the filter hypothesis, we would expect to find only this mtDNA clade in future surveys of mice from the extreme southeastern part of the *M. castaneus* range. The out-of-India filter model appears favored over the out-of-Pakistan shipping model because members of this shallow clade also occur in NC India.

Figure 7 shows a rooted parsimony tree for 110 *domes-*



ticus mtDNAs. An important feature is the placement of the easternmost *domesticus* mtDNAs, *i.e.*, those from Iran, Turkey, and Georgia. Under the earlier hypothesis that the commensal clade arose in the east and *M. domesticus* originated via westward migration (see Introduction and discussion), one would predict that the eastern *M. domesticus* mice would have representatives of all the major mtDNA clades and perhaps some clades not detected in the extensive surveys of the Mediterranean (including North African) and western European animals. Instead, all our Iranian, Turkish, and Georgian mtDNAs [and possibly also the Georgian sequences of Boissinot and Boursot (1997)] are limited to the clade comprising the top left quarter of the tree. In contrast, the deepest lineage in our *domesticus* tree (type 96) comes from two Greek mice, and mtDNAs from

Figure 5.—Parsimony trees for 44 *musculus* mtDNAs (A) and six Yemeni mtDNAs (B). The number of mutations inferred to have occurred along each lineage is indicated. The large solid triangle in A marks the lineage where the 75-bp tandem repeat of the sequence from 15538–15615 arose; small open triangles mark the five lineages with inferred additions of 1–2 bp. Aust, Austria; Bavaria and Bav, Bavarian transect (see Prager *et al.* 1996); Croa, Croatia; Czech, Czech Republic; Dagh, Daghestan; N, northern; NC, north central; S, southern; Slov, Slovakia; SW, southwestern; Turk, Turkmenistan. Heavy horizontal lines in A highlight the terminal lineages leading to the eight new *musculus* mtDNA types and also the 15 of 23 internal branches present in 100% of the 4128 minimal-length trees that PAUP found for these 44 mtDNAs plus *castaneus* type 1 used as an outgroup. The *musculus* tree requires 84 mutations: 66 transitions, 12 transversions, and six length changes (consistency index = 0.73). The eight internal branches not highlighted in A occur in 44–86% of all the minimal-length trees. The *musculus* tree was rooted as shown in all PAUP analyses done. The variation in the single most parsimonious network derived for the six types of Yemeni mtDNA can be explained by 16 transitions and one transversion (consistency index = 0.94). The root was placed as shown in B on the basis of diverse analyses that included a variety of commensal or commensal plus aboriginal mtDNAs. Among the 17 distinct *musculus* mtDNAs identified by Boissinot and Boursot (1997) by sequencing positions 15443–15742, type B92 from Latvia matches our types 7, 9, 10, and 16–19 for this 0.3-kb region, and B94 from Georgia matches type 31. Their 15 other *musculus* mtDNAs can be added to the tree in A as follows (see materials and methods for details), with several of the placements being tentative: types B93 from Latvia and Armenia, B95 from Armenia, and B96, B97, B99, and B101–B103 from Georgia emanating from the same basal node as types 25–28 and 31, with B97 + B99 + B103 and B101 + B102 associated in clades; B130 from Moscow in a clade with type 35; B91 from Georgia and B100 from Daghestan emanating from the same basal node as types 32–36 and 44; B98 from Georgia breaking up the deepest internal branch into two branches, such that among the types depicted, only the clade of 38–40 lies deeper within the *musculus* tree; the phylogenetically equivalent Iranian types B118 and B129 from Mashhad and B119 from Kakhk in a clade that shares a common lineage with the clade of types 38–40 or (among additional equally parsimonious options) emanation from the same node as suggested for B98.

TABLE 3

Quantitative comparisons of sequence differences among the mtDNAs of commensal house mice

mtDNAs compared	Range of pairwise differences			Nucleotide variability (%)	
	Total	Transversions	Length	θ	π
Intra-*musculus*	1–15	0–5	0–3	1.24	0.42
Intra-Yemen	2–11	0–1	0	0.70	0.26
Intra-*castaneus*	1–31	0–7	0–3	2.00	0.77
Intra-*domesticus*	1–24	0–8	0–5	2.03	0.56
musculus vs. castaneus	17–34	3–10	1–5	NA	2.62
musculus vs. Yemen	27–38	5–9	2–4	NA	3.32
castaneus vs. Yemen	24–38	2–7	1–3	NA	3.05
musculus vs. domesticus	33–46	7–15	3–6	NA	3.99
castaneus vs. domesticus	27–49	4–13	2–4	NA	3.71
Yemen *vs. domesticus*	35–47	5–11	3–4	NA	4.02

$\theta = p/A$, with $A = \sum_{i=1}^{n-1} 1/i$; p is the number of polymorphic sites among the sequences divided by the sequence length, and n is the number of mtDNA types. $\pi = \sum_{ij} x_i x_j \pi_{ij}$; x_i and x_j are the frequencies of the ith and jth mtDNA types; π_{ij} is the number of sequence differences between types i and j divided by the sequence length. *castaneus* types 1–7 and 9–28, *musculus* types 1–44, and Yemeni types 1–6 were included in all comparisons involving these categories of mtDNAs; *castaneus* type 8 was included only for computing the intra-*castaneus* θ. The intra-*domesticus* values are based on the 110 mtDNAs in Figure 7; for the other comparisons involving *domesticus* mtDNAs, types 1, 6, 7, 10, 16, 25, 27, and 46 were used as described previously (Prager *et al*. 1996). Length changes involving ≥2 adjacent bp were counted as single differences. The π values tabulated resulted from assigning equal frequencies to each type of mtDNA within a category (*cf*. Prager *et al*. 1996). NA, not applicable.

Greek mice are also found in all but one of the other deep clades in this tree. Sampling of the eastern *domesticus* mtDNAs was limited ($n = 11$ mice and $l = 7$ localities from Turkey plus Iran; $n = 8$ and $l = 6$ for Georgia), but the Greek sample size was similar ($n = 11$, $l = 6$). mtDNAs from Spain ($n = 11$, $l = 7$) and Italy ($n = 34$, $l = 18$) are also found as members of diverse deep clades. This tree does further support the view (Prager *et al.* 1996 and references therein) that southern Mediterranean *domesticus* mtDNA lineages are older than northern European ones.

Among the new *domesticus* mtDNAs from Egypt, types 99–101 fall into the same large clade as do the previously characterized Egyptian types 18 and 22–25, type 97 is a deeper lineage in a clade previously containing mtDNAs from NW Europe and Croatia, type 98 constitutes a relatively deep monotypic branch, and type 28 extends the range of mtDNAs with an 11-bp direct repeat to North Africa. Ten Tunisian mtDNAs belong to the clade containing most of our Egyptian mtDNAs (see legend to Figure 7). These results provide increasing evidence for considerable molecular evolution within NE Africa (see also Tucker *et al.* 1989).

The tree in Figure 7 differs structurally from that presented for 96 *domesticus* mtDNAs (Prager *et al.* 1996) in two notable respects: first, it is shallower, with an average depth of ~6.4 events per lineage rather than 7.3. The start of divergence among all 110 lineages is suggested as some 100,000–280,000 years ago. Second, there has been some rearrangement of the deeper lineages. Specifically, the mtDNAs with C at position 00055 (types 53–56, 68, 69, 91–95, and 102–110) no longer form a monophyletic clade, and they have moved from the lower right of the tree to the upper left. Consequently, the G at position 00055 in types 1–6 and 70 arises via a C-to-G transversion rather than an A-to-G transition. In addition, all the mtDNAs with A at position 00055 are united in a clade (from type 7 down to type 98 in the figure). We previously chose from among equally parsimonious alternatives a tree structure that accounted for the four different bases at position 00055 with two transitions and one transversion (see also Prager *et al.* 1993), an option that now does not yield minimal-length trees.

Figure 8 provides an overview of the character-state phylogenetic analyses in Figures 5–7 and relates the four major commensal mtDNA lineages to one another. The neighbor-joining trees in Figure 9 exhibit the same branching order of the major lineages and the cohesiveness of the *musculus*, Yemeni, and *domesticus* clades (each of which is united by 9–14.5 events on the common lineages in Figure 8). The trees reinforce the view that the Yemeni mtDNAs constitute a distinct branch. In both figures, the *domesticus* lineage occupies the ancestral position among the commensal mtDNAs, the Yemeni lineage appears as the next oldest, and the *castaneus* and *musculus* lineages appear to be the two shallowest. This arrangement and rooting of the four commensal lineages are consistent with the π values in Table 3. Leaving out the newly discovered Yemeni

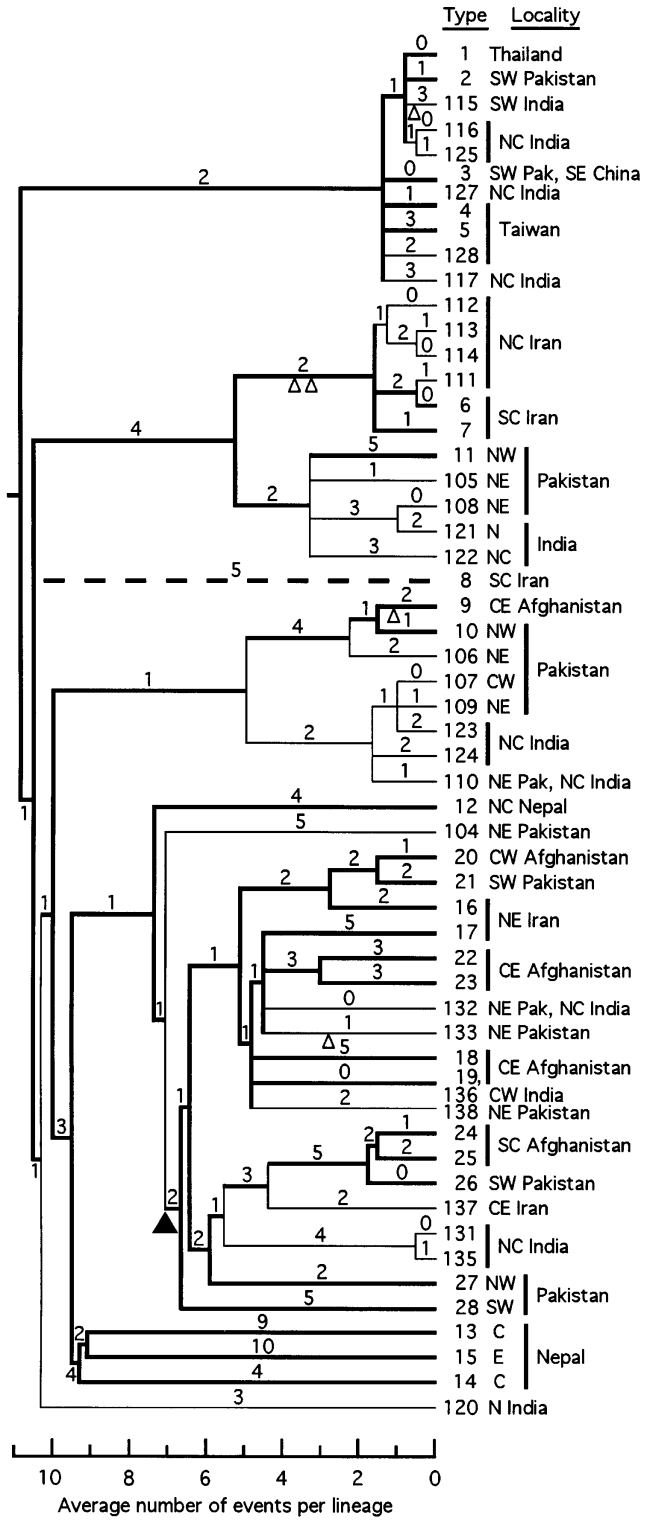
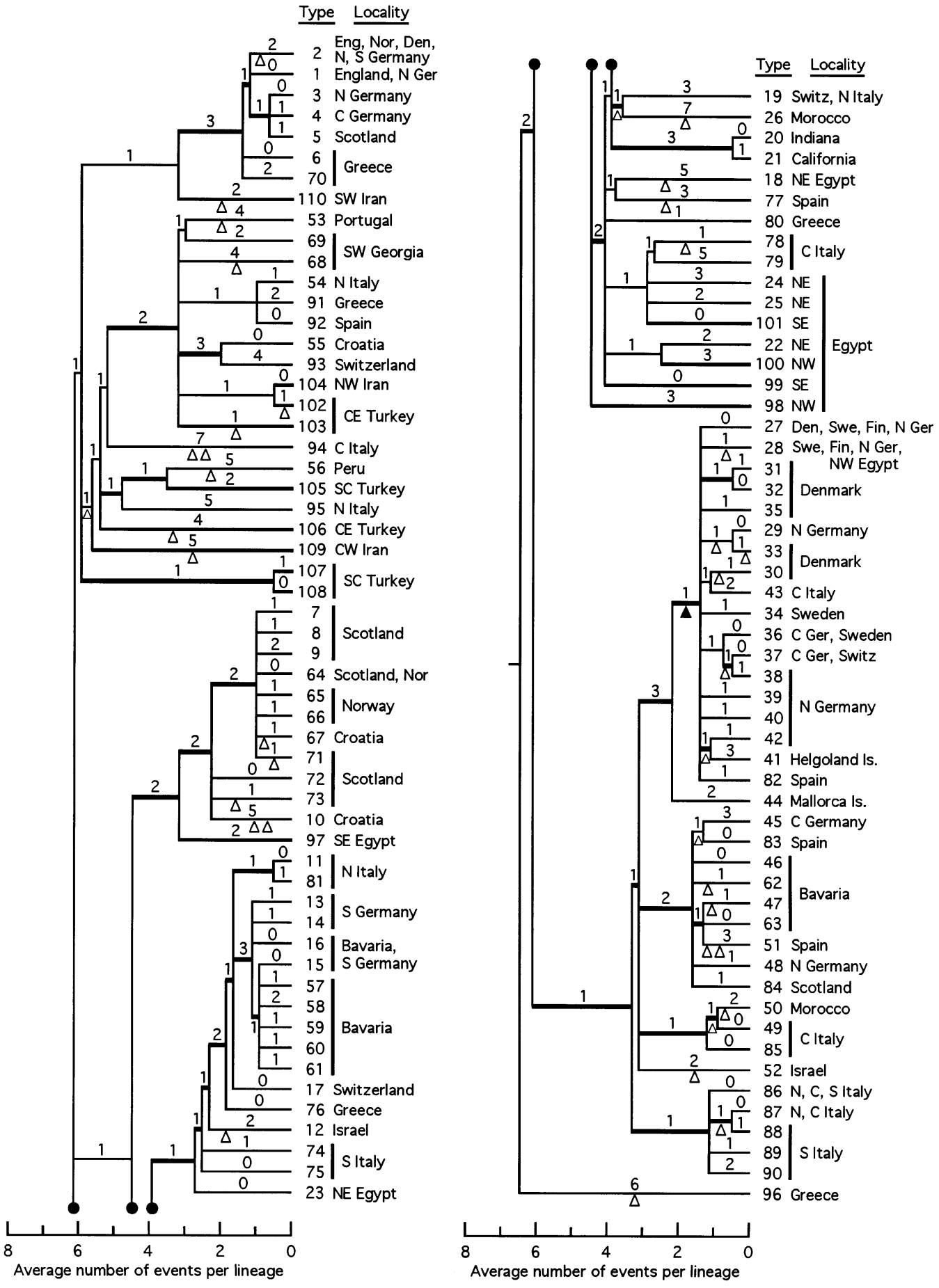


Figure 6.—Parsimony tree constructed for 28 types of *castaneus* mtDNAs shown with heavy lines in the format described for Figure 5. Thin lines indicate the placement (see materials and methods) of and additional branchings generated by the 29 *castaneus* mtDNAs identified by Boissinot and Boursot (1997) from the sequences of positions 15443–15742; type B127 matches our types 3 and 4 in that portion of the control region and B136 matches our type 19. Pak, Pakistan; single letters and two-letter combinations of C, central; E, eastern; N, northern; S, southern; W, western. The source localities for the Boissinot and Boursot (1997) mtDNAs are in order from west to east: Tehran, B111–B114; Birjand, B137; Anga, B107; Rawalpindi, B104–B106, B132, and B133; Islamabad, B108–B110; Gujar Khan, B138; Tahmasapabad, B105; Koorg, B136; Masinigudi, B115; Delhi, B110, B116, B117, B125, B127, B131, B132, and B135; Chaboraha, B122–B124; Leh, B120 and B121; He-Mei, B128. The large solid triangle indicates the origin of the 76-bp tandem repeat; small open triangles mark lineages with inferred additions or deletions of 1–2 bp. Type 8, shown with a dashed line because only segments 1 and 2 were sequenced, was omitted from calculations of node depth, and analyses were done to root the tree. The tree shown with heavy lines is the single most parsimonious network for types 1–7 plus 9–28 and the strict consensus of the three minimal-length networks for types 1–28. This tree requires 135 mutations to explain the variation observed among types 1–28: 106 transitions, 25 transversions, and four length changes (consistency index = 0.68). The root was placed on the lineage indicated from the strict consensus tree of an analysis that included *musculus* mtDNA types 1–9, 11–22, 24, 26, 28–33, and 36–41 plus Yemeni types 1–4 and 6 and *domesticus* types 1, 6, 7, 16, 25, 27, and 46. Analyses that included as outgroups only the six Yemeni sequences or those plus *musculus* and *domesticus* types 1 yielded four minimal-length trees in each case: two with the root placed as shown here, and two with an alternative placement along the lineage separating the clade of types 13–15 from the rest of the *castaneus* sequences.

lineage, the trees in Figures 8 and 9 have the same branching order and root placement as the trees of Boursot *et al.* (1996) and Boissinot and Boursot (1997) (see discussion for details). However, as emphasized in the discussion, the available data do not permit the arrangement and rooting of the four major lineages in Figure 8 to be inferred with statistical confidence, as is true also of the assessment of the cohesiveness of the *castaneus* mtDNAs. The deepest internal branches in Figure 8 have only three to four events, and they are similarly short in Figure 9. An obvious possibility is that cladogenesis has been rapid.

Tandem repeats of 75 and 76 bp: Table 4 quantitatively compares the results of the independent duplications of the same control region segment in *castaneus* and *musculus* mtDNAs. By all criteria, the duplication occurred earlier among the *castaneus* mtDNAs: assuming roughly equal rates of evolution, the tree-based analyses place the duplication point at least twice as long ago for the *castaneus* lineage, with computed depths of ~6.5 *vs.* 3.0 events per lineage. About the same number of events occurred in the areas flanking and within the repeats among the *castaneus* mtDNAs, but none accumulated outside the repeats after the duplication among the *musculus* mtDNAs. Pairwise, the averages and, more importantly, the tops of the ranges are all roughly 1.5- to 3-fold greater among the *castaneus* repeats. Another contrast is that in the *musculus* mtDNAs, the 3′ copy has accumulated more base substitutions, while among the *castaneus* mtDNAs, the 5′ copy seems to have changed more. Finally, the average of 4.2 substitutions between repeats within a given type of *castaneus* mtDNA scarcely


Type Locality
2 Eng, Nor, Den, N, S Germany
1 England, N Ger
3 N Germany
4 C Germany
5 Scotland
6 Greece
70 Greece
110 SW Iran
53 Portugal
69 SW Georgia
68 SW Georgia
54 N Italy
91 Greece
92 Spain
55 Croatia
93 Switzerland
104 NW Iran
102 CE Turkey
103 CE Turkey
94 C Italy
56 Peru
105 SC Turkey
95 N Italy
106 CE Turkey
109 CW Iran
107 SC Turkey
108 SC Turkey
7 Scotland
8 Scotland
9 Scotland
64 Scotland, Nor
65 Norway
66 Norway
67 Croatia
71 Scotland
72 Scotland
73 Scotland
10 Croatia
97 SE Egypt
11 N Italy
81 N Italy
13 S Germany
14 S Germany
16 Bavaria, S Germany
15 Bavaria, S Germany
57 Bavaria
58 Bavaria
59 Bavaria
60 Bavaria
61 Bavaria
17 Switzerland
76 Greece
12 Israel
74 S Italy
75 S Italy
23 NE Egypt
8 6 4 2 0
Average number of events per lineage
Type Locality
19 Switz, N Italy
26 Morocco
20 Indiana
21 California
18 NE Egypt
77 Spain
80 Greece
78 C Italy
79 C Italy
24 NE Egypt
25 NE Egypt
101 SE Egypt
22 NE Egypt
100 NW Egypt
99 SE Egypt
98 NW Egypt
27 Den, Swe, Fin, N Ger
28 Swe, Fin, N Ger, NW Egypt
31 Denmark
32 Denmark
35 Denmark
29 N Germany
33 Denmark
30 Denmark
43 C Italy
34 Sweden
36 C Ger, Sweden
37 C Ger, Switz
38 N Germany
39 N Germany
40 N Germany
42 N Germany
41 Helgoland Is.
82 Spain
44 Mallorca Is.
45 C Germany
83 Spain
46 Bavaria
62 Bavaria
47 Bavaria
63 Bavaria
51 Spain
48 N Germany
84 Scotland
50 Morocco
49 C Italy
85 C Italy
52 Israel
86 N, C, S Italy
87 N, C Italy
88 S Italy
89 S Italy
90 S Italy
96 Greece
8 6 4 2 0
Average number of events per lineage

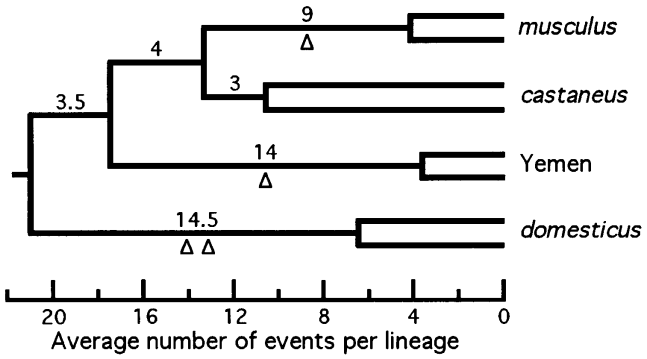


Figure 8.—Parsimony tree for mtDNAs of commensal house mice. First, this tree schematically summarizes the information in Figures 5–7. Thus, for example, the *musculus* portion represents the deepest intra-*musculus* node, plotted at an average of 4.2 events per lineage as in Figure 5A. Second, it adds six lineages that connect the four trees in Figures 5–7 to one another; the 48 events assigned to these lineages (at 47 polymorphic sites) consist of 35 transitions, nine transversions, and four 1–2-bp length changes (open triangles). Assignment of mutations to these six lineages, selection of branching order, and root placement were done by considering 188 commensal mtDNA sequences (*i.e.*, all those in Figures 5 and 7 plus types 1–28 in Figure 6) and those from aboriginal house mice. Parsimony and neighbor-joining (Figure 9) trees, pairwise distances, and estimates of nucleotide variability (Table 3) were taken into account. In analyses of commensal plus aboriginal sequences, arrangement of the deeper commensal lineages and placement of the deepest root were not completely stable to methods of tree construction and choice of representative sequences (*e.g.*, see Figure 9 and its legend). While the *musculus*, Yemeni, and *domesticus* mtDNAs each invariably formed a monophyletic clade (bootstrap values of 97–100% in 1000 replications in two parsimony analyses of 13 commensal plus 6 aboriginal sequences, including the set of types in Figure 9), as did all commensal mtDNAs collectively (100% bootstrap values), this was not true of the *castaneus* mtDNAs. Indeed, in some parsimony analyses, the *domesticus* mtDNAs were implied to emanate from the same node as *castaneus* type 13. The branching order and rooting shown here were, therefore, chosen based on intracommensal parsimony trees and distance values.

exceeds that of 3.9 among 5′ copies, compared to *musculus* mtDNAs with noticeably more differences between repeats within a type than in the 5′ copy among types (averages of 2.9 *vs.* 1.2).

***Y* chromosomes:** In mice from areas where it is clear, based on phenotypic and genotypic criteria, that the nuclear genomes are *M. domesticus* (*e.g.*, Europe and North Africa), the *Zfy-1* and *Zfy-2* genes are the same length; equal-sized genes have been reported also for mice [*M. (m.) bactrianus* or *M. (m.) sp.*] from India and Pakistan (Nagamine *et al.* 1992, 1994b; Boissinot and Boursot 1997; Prager *et al.* 1997). Where the nuclear genomes are *M. musculus* (E Europe and N Asia) by the same criteria, the *Zfy-2* gene is 18 bp shorter; such shorter genes have been found also in *M. castaneus* mice of extreme SE Asia (Nagamine *et al.* 1992, 1994b; Boissinot and Boursot 1997; Prager *et al.* 1997). The shorter *Zfy-2* can be inferred to be the derived state, given that the cladistic study by Tucker *et al.* (1989) of restriction fragment length polymorphisms (RFLPs) of retroviral-related elements in commensal and aboriginal mice places the *Y* of *M. domesticus* as ancestral to those of *M. musculus* and *M. castaneus*.[3] Furthermore, it appears reasonable to assume that this 18-bp deletion occurred only once (see also below). A few other *Y* chromosomal markers have been identified as varying in concordance with the length state of *Zfy-2* (Boissinot and Boursot 1997 and references therein). Nevertheless, it must be recognized that assessment of the *Zfy-2* size class (and of concordant markers) affords little resolving power among *Y* chromosomes of commensal

[3]The 18-bp deletion has been reported as absent in *M. spretus* and in the more distantly related non-house mouse *M. caroli* (Nagamine *et al.* 1994a), but no survey of *M. spretus* and the other two aboriginal house mouse species appears to have been done. The apparent variation in the number of *Zfy* genes among mouse species outside the commensal group (Nagamine *et al.* 1994a) would complicate such a survey and its interpretation.

Figure 7.—Parsimony tree for 110 *domesticus* mtDNAs shown in the format described for Figure 5. Solid circles indicate the connection of the left and right halves of the tree. Den, Denmark; Eng, England; Fin, Finland; Ger, Germany; Nor, Norway; Swe, Sweden; Switz, Switzerland; Scotland includes also localities in the Orkney and Shetland Islands. The solid triangle marks the lineage where the 11-bp direct repeat of the sequence at positions 16073–16083 has arisen; open triangles mark 36 lineages with inferred additions or deletions of 1–5 bp. Heavy horizontal lines highlight the 14 new *domesticus* mtDNA types and the 36 of 54 internal branches that are present in 100% of all minimal-length trees. This tree requires 237 mutations: 171 transitions, 26 transversions, and 40 length changes (consistency index = 0.50). The root was placed as shown from the strict consensus tree of an analysis that included *musculus* mtDNA types 1, 20, 29, 30, and 38–40; Yemeni types 1, 2, and 6; and *castaneus* types 1, 9, 12, and 28. Six of the internal branches not highlighted occurred in 65–90% of all minimal-length trees, nine occurred in 21–50%, and three were not evaluated [see materials and methods and Prager *et al.* (1993, 1996) for further details]. Among the 25 distinct *domesticus* mtDNAs identified by Boissinot and Boursot (1997) by sequencing positions 15443–15742, type B66 from Tunisia matches types 86, 87, 89, and 90 for this 0.3-kb region; Tunisian B67 matches type 94; Tunisian B75 matches types 80 and 99; French B82 matches type 76; and French B84 matches types 15, 16, and 59–61. Twelve of their 20 *domesticus* sequences distinct from types 1–110 could be assigned (see materials and methods) to specific sections of our tree with reasonable confidence, as follows: B83 from Italy emanating from the same node as type 17 and the clades of 11 + 81 and 13–16 + 57–61; clades of Tunisian B64 + B65 and B78 + B79 emanating from the same basal node as types 80, 99, and several other lineages; Spanish B80 + B81 in a clade emanating from the same node as types 20 and 21; Tunisian B76 and B77 emanating from the same node as types 18 and 77; Tunisian B72–B74 in a clade with type 100, with B72 + B73 grouped therein. Possible placements for the remaining eight sequences are as follows: clades of Georgian types B85 + B88 and B86 + B87 emanating from the same node as type 110 and the clade of 1–6 + 70; Tunisian B68, B69, a clade of B70 + B71, and perhaps also B67 (see above) emanating from the same node as type 97 and the clade extending from type 7 to 10.

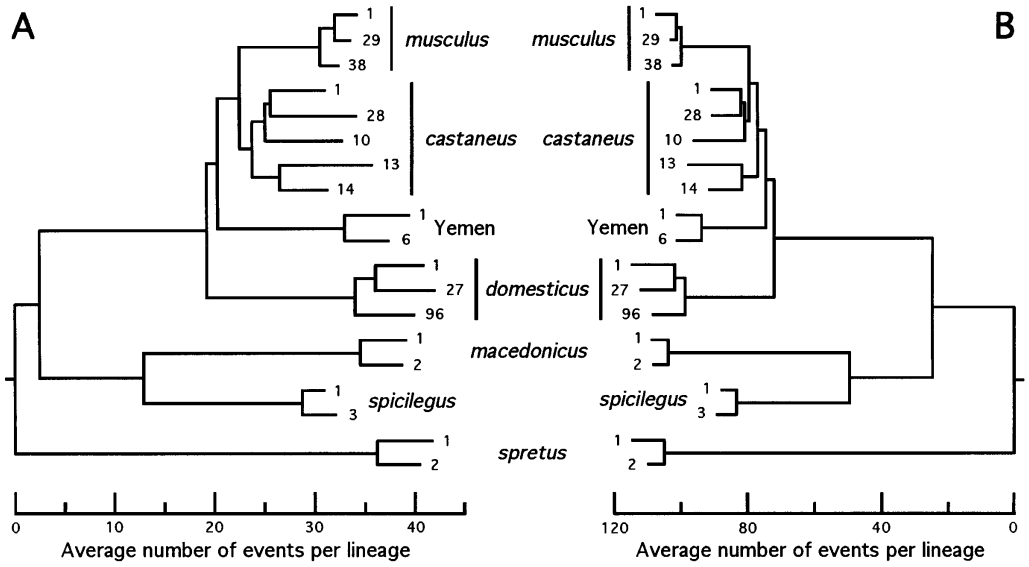


Figure 9.—Neighbor-joining trees for 19 mtDNAs from commensal and aboriginal house mice. In A, all changes were weighted equally; in B, transversions were weighted fivefold relative to transitions and length changes. The deepest and second-deepest commensal clades are, respectively, *domesticus* and Yemeni mtDNAs in both trees. However, the *castaneus* mtDNAs are monophyletic in A but paraphyletic in B. In the analogous analyses for only 14 sequences (with *dom* 96, *mus* 38, and *cas* types 10, 13, and 14 omitted), the branching order of both trees matched A here.

mice compared to the existing variation demonstrated by Tucker *et al.* (1989). That study of sequences presumably spread over a large part of the *Y* chromosome revealed extensive interpopulational variation and regional differentiation within *M. domesticus* and, to a lesser extent, the same phenomena within *M. musculus*. We do not know whether or not *Y*s with equally sized *Zfy-1* and *Zfy-2* genes from outside the well-recognized territory of *M. domesticus* are about as different from *M. domesticus Y*s as are *Y*s bearing the deletion in *Zfy-2*. Until such mice are included in a study having the multistate discriminating power and phylogenetic potential of the Tucker *et al.* (1989) analysis, only limited inferences about *Y* chromosomal variation can come from two-state assays such as that used here.

TABLE 4

Comparison of tandem repeats in two mtDNA lineages of commensal house mice

Comparison	*castaneus*	*musculus*
Pairwise differences		
Tandem repeats		
Within a mtDNA type	4.2, 1–7	2.9, 2–4
Among 5′ copies	3.9, 0–8	1.2, 0–2
Among 3′ copies	2.7, 0–7	2.0, 0–3.5[a]
Among 5′ + 3′ copies	6.6, 0–12	3.2, 1–5
1.1-kb region	11.3, 3–21	3.2, 1–5
Tree analysis		
Duplication point	6.5	3.0
Shallowest node	1.5	0.5
Events within repeats	29	10
Events outside repeats	26	0

The mtDNAs are *castaneus* types 16–28 and *musculus* types 32–36. Under pairwise differences, averages and ranges of base substitutions are given; all possible pairs are included. Duplication point, node depth, and numbers of events are derived from the parsimony trees in Figures 5A and 6. The entries for the shallowest node refer to the most recent divergences between any two types carrying tandem repeats. The events within and outside repeats count the inferred mutations during divergence of all types that have two repeats plus those on their common lineage after the duplication point.

[a] Nonintegral value resulting from apparent heteroplasmy at one site in type 32 (Prager *et al.* 1996).

Consistent with previous reports (Nagamine *et al.* 1992, 1994b; Orth *et al.* 1996; Boissinot and Boursot 1997; Prager *et al.* 1997), we found (Table 1) the A allele (*Zfy-2* same length as *Zfy-1*) in Egypt, SW Georgia, and Pakistan and the B allele (*Zfy-2* shorter) in Daghestan, Korea, and Taiwan. In agreement with well-recognized species distributions and the mtDNA data, the five Siberian males had the B allele. We found only the A allele in Turkey, as expected from *domesticus* mtDNA and anatomical evidence, and in Yemen and Nepal. In light of the mtDNA trees in Figures 8 and 9, finding only undeleted *Zfy-2*s in Yemen strengthens the view that equal lengths are the ancestral condition.

In Turkmenistan, we detected only the B allele, along with only *musculus* mtDNA. Two mice from Iran carried the B allele—at NW locality 18 in an animal with *domesticus* mtDNA and at SC locality 22 in an animal with *castaneus* mtDNA (Table 1). These findings extend to other areas of Iran the published reports (Nagamine *et al.* 1992; Boissinot and Boursot 1997) of *Y*s with the B allele along the NE edge of the country (from Mashhad to 400 km south in Birjand) and in Tehran in the NC region. Two other Iranian mice carried the A allele, which appears to be the first detection of this type of *Y* in Iran. These two animals, collected along the country's western edge (at localities 19 and 21), also had *domesticus* mtDNA and the anatomical features of this species (Table 2). In Afghanistan, we also found both kinds of *Y*. Notably, both males among the three Afghan animals with *musculus* mtDNA had the *Y* B allele,

which, coupled with their appearance (Table 2), supports the idea of *M. musculus* populations across the country's northern edge.

Sequencing the shorter kind of *Zfy-2* from 18 mice from localities extending from N Germany to Korea and Taiwan (see list in materials and methods) confirmed deletion of the identical 18 bp in all cases, which bolsters the view that this deletion was a singular event. A base substitution was noted in the Iranian mouse from locality 18: a G-to-A change in the first position of codon 507 encodes a threonine in place of alanine.

***p53* pseudogenes:** The species and geographic distribution of variation at a locus we designate Ψ*p53-1* is somewhat like that of the *Zfy-2* length states: the Ψ*p53* is present (P in Table 1) in pure *M. domesticus* populations and absent (N in Table 1) in pure *M. musculus* populations, with a more complex pattern of variation and polymorphism in Central and SE Asia (Tanooka *et al.* 1995; Ohtsuka *et al.* 1996; Prager *et al.* 1997; this report). To account for the inter- and intraspecific variation they observed in the genus Mus, Ohtsuka *et al.* (1996) suggest a single reverse transcription of *p53* cDNA, incorporation of this processed gene into the genome of an ancestral mouse, Ψ*p53* fixation in the ancestral mouse population, and Ψ*p53* loss along several lineages. The alternative model invokes nonfixation in the ancestral population and maintenance of an old polymorphism through several speciation events. Whichever model is correct for older intrageneric divergences, the presence of Ψ*p53-1* can be reasonably inferred as the ancestral condition for the commensal clade. This conclusion derives from the branching structure of the mtDNA trees (Figures 8 and 9) and the homozygous Ψ*p53* presence in a broad survey of European and North African *M. domesticus* (Prager *et al.* 1997; this report) and in the Yemeni mice (Table 1); it receives further support from sequence data (see below) implying that the Ψ*p53*-positive aboriginal species, *M. macedonicus* and *M. spicilegus*, which are the sister group to the commensals (Figure 9), also have the Ψ*p53-1* locus. (We confirmed absence of a Ψ*p53* in the phylogenetically more remote *M. spretus* by testing nine mice from Spain and Morocco—see materials and methods.)

All our mice from Egypt, Turkey, and Yemen had the Ψ*p53*. For the Egyptian and Turkish mice, the results for this autosomal locus add to the evidence from mtDNA, the *Y* chromosome, and anatomical traits that they are *M. domesticus* mice. The implication for Yemen is that this area was colonized by founders carrying ancestral traits. The three mice from W Iran were homozygous positive for Ψ*p53*. Two of them, from localities 19 and 21 on the western side of the Zagros Mountains, are *M. domesticus* by mtDNA, the *Y* chromosome, and appearance, and they could be representatives from the eastern edge of pure *M. domesticus* populations. The third one, from locality 18, also has *domesticus* mtDNA and a *M. domesticus* phenotype, but carries the *Y* B allele.

Ψ*p53* absence in all the mice from Daghestan and Siberia fits with other evidence (Frisman *et al.* 1990; Orth *et al.* 1996; Boissinot and Boursot 1997; Table 1) that these are *M. musculus* populations. Like Tanooka *et al.* (1995), we did not detect the Ψ*p53* in animals from Korea and N China, in agreement with mtDNA, *Y* chromosomal, and phenotypic evidence (Yonekawa *et al.* 1988; Nagamine *et al.* 1994b; Tables 1 and 2) that these are *M. musculus* mice. The northern Afghan mice (localities 31–33) are *M. musculus* by mtDNA, the *Y* chromosome, and anatomical traits, and they uniformly lack the pseudogene at Ψ*p53-1*. They could be representatives at the southern edge of pure *M. musculus* populations in Central Asia.

The Ψ*p53* polymorphism we noted in SW Georgia supplements other evidence (*e.g.*, Frisman *et al.* 1990; Milishnikov *et al.* 1990; Orth *et al.* 1996; Table 1) of a contact zone between *M. domesticus* and *M. musculus* in Transcaucasia. Ψ*p53* polymorphism in both S and N Turkmenistan provided our survey's first suggestion of non-*M. musculus* genes in populations in that country. These results are consistent with Turkmenistan's proximity to the highly polymorphic central populations and with evidence from other studies (*e.g.*, Milishnikov *et al.* 1994) that indicate high diversity in the Central Asian republics of the former Soviet Union. They also raise the possibility of residual polymorphism for P and N at Ψ*p53-1* in *M. musculus*. Similar considerations may apply to the mouse with *musculus* mtDNA and Ψ*p53* P and N from NC Iranian locality 25 on the Caspian Sea.

Though the majority of animals with *castaneus* mtDNA carry the Ψ*p53* (Table 1), we found exceptions at localities 26, 27, 30, and 38 in NE Iran, WC Afghanistan, and SW Pakistan and heterozygosity for P and N at locality 36 in EC Afghanistan. Both our *M. castaneus* from Taiwan had Ψ*p53* in the homozygous state, but Ohtsuka *et al.* (1996) reported polymorphism there. These observations for Central and SE Asia suggest that *M. castaneus* is polymorphic for P and N at Ψ*p53-1*. Furthermore, while most of the Ψ*p53*-positive males with *castaneus* mtDNA carry the *Y* A allele, we observed *castaneus* mtDNA, Ψ*p53* P, and the *Y* B allele at SC Iranian locality 22 and SC Afghan locality 34, as well as in Taiwan.

Figure 10 summarizes the results of sequencing from 79 commensal mice a 128-bp piece of Ψ*p53-1* that includes the 3′ end of exon 4 and the 5′ end of exon 5. What makes this region ideal for providing assurance that one is looking at the same locus and the products of the same incorporation event are the deletion of 6 bp plus the insertion of 1 bp relative to the functional gene within a span of 9 bp. We found 12 variable sites and 18 sequence phenotypes, 3 of them widespread and 13 of them each exhibited by only one individual (Figure 10A). To the extent that the positions sequenced overlap, our sequences 1–3 match those reported by Ohtsuka *et al.* (1996) for two *M. domesticus*, a *M. castaneus*, and two *M. bactrianus*, and they correspond to the com-

A Sequence phenotypes

Type	111111111111 111233334444 014856770558 |||||||||||| 322233231131	Mice	Localities
1	CTGAGGCGGACC	27	Widespread
2	G..........	21	Widespread
3	S..........	13	Widespread
4	S..R.....	1	Oxfordshire, England
5	.........Y.	1	Turkey (14-82204)
6	.?.....G..	1	Turkey (12-82211)
7	.....R..	1	Turkey (13-82241)
8	.?..A.....T	1	Iran (25-*h*)
9	.....S.....Y	3	A (35-103670), P (42), N (46)
10	.....C.....T	1	Afghanistan (36-*h*)
11	???...T.....	1	Afghanistan (34)
12	.?...S.....T	1	Pakistan (44)
13	.?..R.....T	1	Pakistan (45)
14	.?R.RC.....Y	1	Pakistan (40)
15	.....A...	1	Nepal (49)
16	.....A.T.	1	Nepal (47)
17	TA.....A....	1	Nepal (48)
18	T.....A....	2	Taiwan

B Alleles inferred

Type	111111111111 111233334444 014856770558 |||||||||||| 322233231131	Present in phenotypes	No. of Mice	No. of Chr	Localities
1	CTGAGGCGGACC	1,3,4,5,7,9	46	73	Widespread
2	G..........	2,3	34	50	Widespread
3	G..G.....	4	1	1	England
4	.........T.	5	1	1	Turkey
5	.....A.T.	16	1	2	Nepal
6	.....A...	15	1	2	Nepal
7	.....G..	6,7	2	3	Turkey
8	.?..A.....T	8,13	2	2	Iran, Pakistan
9	.?.....T	12,13	2	2	Pakistan
10	.....C.....T	9,10,12,14	6	6	A, P, N
11	???...T.....	11	1	2	Afghanistan
12	T.....A....	18	2	4	Taiwan
13	TA.....A....	17	1	2	Nepal
14	.?A.AC.....	14	1	1	Pakistan

Figure 10.—Variable sites, observed sequence patterns (A), and inferred alleles (B) in a 128-bp segment of a *p53* pseudogene among 79 commensal mice from 68 localities, presented in the format of Figures 3 and 4. The 12 variable sites are listed vertically according to codon number and position within the codon; S, R, and Y indicate, respectively, C + G, A + G, and C + T; ?, unsequenced sites. Phenotype 1 and allele 1 at locus Ψ*p53-1* differ from the functional *p53* in having codons 120 and 121 deleted, a T inserted between positions 2 and 3 of codon 122, and a stop signal at codon 143. Locality numbers and, if necessary, identification numbers are included in A for regions in Table 1 where not all mice had the same sequence phenotype. A, Afghanistan; P, Pakistan; N, Nepal; *h*, heterozygosity for presence / absence of the Ψ*p53*. In B, Chr tabulates the number of chromosomes out of 151. Because phenotypic patterns 4, 9, and 14 are each polymorphic at two or three sites, one cannot infer their allele sequences conclusively in the absence of sequencing multiple clones of PCR products. Thus, 14 should be regarded as the minimum number of distinct alleles, and the sequences of postulated alleles 3 and 14 should be viewed as uncertain. (Assignment of both variant bases in phenotype 4 to allele 3 was arbitrary.) At site 111-2 in phenotypes 17 and 18, the Nepalese mouse and Taiwanese animal 152831 may have A + T rather than the single bases tabulated. For the widespread phenotypes 1–3, the localities, numbers of mice, and individual identities are as follows: (1) One mouse each from Mallorca Island, Spain; Val Poschiavo, Switzerland; Tübingen and Helgoland Island, Germany; Erfoud, Morocco; Bashtil, Egypt; Georgia (4569); and Afghanistan (37). Two mice from Turkey (15), three each from Iran (19, 22, 24), Turkmenistan (28, 29), and Pakistan (39, 41, 43), and eight from Yemen. (2) One mouse each from Estany, Spain; Lisbon, Portugal; Neufahrn and Giessen, Germany; four localities in East Holstein, Germany (Burg-*h*, Oldenburg-*h*, Sieversdorf, Bokhorst); Nöfing, Austria (*h*); Metkovič, Croatia; Turkey (12–82208); and Iran (23). Two mice from Georgia (4567-*h*, 4575-*h*) and all seven Egyptian mice in Table 1 (with 22 bp missing for locality 5). (3) One mouse each from Catalunya, Spain; France; Augsburg and Neumünster, Germany; Canton Vaud and Bern, Switzerland; Pisa, Italy; Thebes, Greece; Faiyum, Egypt. Two mice each from Turkey (13–82242, 14–82205) and Iran (18, 21).

mon commensal type II of Tanooka *et al.* (1995). The substitution of A at position 137-3 (in our patterns 17 and 18) corresponds to type III seen by Tanooka *et al.* (1995) in *M. castaneus* from Taiwan and Indonesia.

From the 18 sequence phenotypes, we inferred a minimum of 14 alleles (Figure 10B), which differ pairwise by one to six base substitutions. The alleles can be related in almost a star phylogeny (not shown), which requires 16 mutations (consistency index = 0.81) to explain the observed variation. Typical trees have a nine-way multifurcation at the basal node, whose sequence matches allele 1, with subsequent sharing of common lineages by alleles 2 + 3, 8–10, 12 + 13, and 5 + 4 or 6. Allele 1 can be inferred to be the ancestral allele for the commensal pseudogene at Ψ*p53-1* because at all 12 variable sites in Figure 10, it matches the sequences from all the aboriginal mice examined (see below and materials and methods for details).

The geography of the commensal Ψ*p53-1* alleles is revealing (Figure 10, Table 5). In addition to being phylogenetically ancestral, allele 1 is widespread, occurring in every region we looked at, except Taiwan. The mice from Yemen and Turkmenistan are monomorphic for allele 1. In contrast, allele 2 was found only in mice with *M. domesticus* genomes, except for the Iranian mouse at locality 23. The observed allelic diversity is greater in Turkey, Afghanistan, Pakistan, and Nepal (h = 0.69–0.89; each area has three to five alleles for only 7–16 chromosomes assessed) than in western Europe plus North Africa (h = 0.48). Rare alleles generated *in situ* in western Europe beyond type 3 in England may not have been uncovered because the sampling was not intense in any one area, but two rare alleles (4 and 7) were detected in *M. domesticus* territory in Turkey. The overall scenario suggested is an ancestral allele 1, eastward migration(s) by founder populations carrying this allele, and *in situ* generation of rarer alleles (5, 6, and 8–14) in Central Asian and emigrant populations.

TABLE 5

Ψ*p53* diversity among commensal house mice

Geographic area	Alleles	*n*	*l*	*h*
Europe + Morocco	1, 2, 3	45	3	0.524
Egypt	1, 2	18	2	0.295
Yemen	1	16	1	0
Turkey	1, 2, 4, 7	16	4	0.692
Georgia	1, 2	4	2	0.667
Iran	1, 2, 8	13	3	0.564
Turkmenistan	1	6	1	0
Afghanistan	1, 10, 11	7	3	0.761
Pakistan	1, 8, 9, 10, 14	14	5	0.725
Nepal	1, 5, 6, 10, 13	8	5	0.893
Taiwan	12	4	1	0
Total sample	1–14	151	14	0.657

Diversity (h) was calculated with the equation $h = n(n-1)^{-1}(1 - \sum_{i=1}^{l} x_i^2)$. The parameters n and l are, respectively, numbers of chromosomes and allele types (Figure 10B); x_i is the frequency of the ith type. The 79 animals assessed in Figure 10 are considered here. For Europe plus North Africa (*i.e.*, Egypt added to the first listing), $h = 0.482$.

The aboriginal mice do not share with the commensals the deletion of 6 bp, insertion of 1 bp, and stop at codon 143 (see Figure 10), but our mice plus other representatives of these two species (Tanooka *et al.* 1995; Ohtsuka *et al.* 1996) share a C-to-T substitution at the first position of codon 139. Among nine mice, we found only two polymorphic sites, both in codon 122, but five sequence phenotypes: at the second and third positions, respectively, of this codon, CG (type 1) in both species, as well as CA, YG, CR, and YR (types 2–5 in the order listed) in *M. spicilegus* [with CA being the *M. spicilegus* sequence in Tanooka *et al.* (1995)]. From these observations, we inferred a minimum of three aboriginal alleles: CG, CA, and TG, with respective frequencies of 0.56, 0.33, and 0.11. A strong indication that the Ψ*p53* common in commensal genomes (*i.e.*, Ψ*p53-1*) is the same locus in the aboriginal mice as a consequence of incorporation before the commensal-aboriginal split comes from the sequences of another piece of the Ψ*p53*, the 89 bp extending from the 3′ end of exon 5 through most of exon 6 and bounded by primers Int5S and Int5R. In this second segment, all the house mouse sequences reported by Ohtsuka *et al.* (1996), *i.e.*, including *M. spicilegus* and *M. macedonicus*, as well as a German and a Georgian mouse (see materials and methods) with the prevalent category of commensal Ψ*p53*, share three base substitutions relative to the functional gene: C to A at 193-1 (codon 193, position 1), C to G at 200-1, and G to A at 201-3. The juxtaposed length changes of −6 bp and +1 bp would then be assigned to the common lineage preceding intracommensal divergence and considered diagnostic of commensal Ψ*p53-1*.

The mice from localities 31 in Afghanistan and 48 in Nepal provided evidence for a second processed Ψ*p53* locus (*cf.* Table 1 and materials and methods). With primers Exon 4 + Exon 5, both mice yielded Ψ*p53* fragments matching the coding portion of the functional gene in length and sequence, and sequence data for the segment amplified by primers Int5S + Int5R supported the hypothesis of a locus distinct from Ψ*p53-1* (details available from the authors). The demonstration of a variant Ψ*p53* locus in the northern Afghan mouse fits nicely with its otherwise *M. musculus*-like genotype and phenotype. We infer that these two unusual mice are each probably exemplars of two new and independent retrotranspositions of the *p53* mRNA because it is not apparent how the same new Ψ*p53* would be shared exclusively (in our survey) by two mice whose genotypes and phenotypes are otherwise quite different and that are from localities some 2200 km apart in an area dominated by inhospitable mountainous terrain. The rat *Rattus norvegicus* has multiple Ψ*p53* loci (Weghorst *et al.* 1995), and our findings provide additional impetus for characterizing the Ψ*p53* insertion points in the genomes of house mice. The Nepalese mouse from locality 48 is intriguing, not only in having two Ψ*p53* loci, but also in having a *M. castaneus* phenotype (Table 2), an mtDNA (*cas* 13) rather distantly linked to all others, and a Ψ*p53-1* allele (type 13) with two base changes uniquely shared with *M. castaneus* from Taiwan.

DISCUSSION

Commensal house mice of Yemen: The implication from the evolutionary trees in Figures 8 and 9 and the pairwise comparisons in Table 3 is that the mtDNAs of the Yemeni mice are phylogenetically distinct from the other categories of commensal mtDNAs heretofore recognized. Furthermore, the mtDNAs extant in Yemen appear to have been diverging from one another for an appreciable amount of time, approaching the time characterizing the mtDNA divergence of *M. musculus* over its entire range (Figure 5). At the level of resolution used in this survey, the Yemeni mice have the ancestral states for all three traits at the two nuclear loci examined—the *Y* chromosome A allele, presence of Ψ*p53-1*, and allele 1 at Ψ*p53-1*. The distinct monophyletic clade of their mtDNAs suggests that these Arabian Peninsular animals may represent another recognizable species in the commensal mouse complex. As they have already been given a separate taxonomic designation because of their small size (see results), we will use the name *M. gentilulus* henceforth in this article to refer to them.

The results reported here suggest that more attention be given to the genetics and morphology of *M. gentilulus* than has been done by earlier systematists. As its nuclear gene traits revealed by the present study plus some of its anatomical features are characteristic of mice from diverse areas, additional nuclear loci should be assessed. To investigate further the origin and dispersal of the

Yemeni mice, it becomes desirable to sample for genetic analyses from diverse parts of the Arabian Peninsula, all along the northern shores of the Persian Gulf and the Gulf of Oman, and also the Horn of Africa and adjacent areas. Indeed, discovery of the *gentilulus* mtDNAs provides a strong stimulus for a molecular genetic analysis of house mice from throughout Africa. It has been presumed that, except for North Africa, the continent became populated by commensal house mice because of spreading by humans during recent millennia. Furthermore, it now seems generally believed that these African mice are all *M. domesticus* (*e.g.*, see Klein *et al.* 1987; Boursot *et al.* 1993; K.S.J. 1995; Din *et al.* 1996). Schwarz and Schwarz (1943), however, placed *M. (m.) castaneus* on the coast of East Africa and throughout southern Africa while designating the mice in Somalia as *M. (m.) bactrianus* and those on the Eritrean coast (on the Red Sea) and in northern Sudan as *M. (m.) praetextus.*

Origin and radiation of commensal house mice: The centrifugal model of evolution proposed by Boursot *et al.* (1993, 1996), Bonhomme *et al.* (1994), and Din *et al.* (1996) hypothesizes the northern Indian subcontinent as the cradle of the commensal clade as a whole, and from there, range expansions westward, northward, and eastward to give rise, respectively, to the peripheral populations that are now called *M. domesticus*, *M. musculus*, and *M. castaneus* (designated by them as subspecies of *M. musculus*). They refer to the central populations as *M. m. subspp.* and identify them geographically as Delhi, Pak, and Iran, as their genetic affinities were not clarified. After an initial westward movement of mice along the Arabian Sea and eastern Persian Gulf, the area west of the Zagros Mountains is suggested as a good candidate for the original homeland of *M. domesticus*, from where mice subsequently spread westward to colonize the present-day range of the taxon around the Mediterranean and in NW Europe. The progenitors of *M. musculus* are hypothesized [see Figure 4 in Boursot *et al.* (1996)] to have moved northward between the Kopet Dagh Mountains and the Paropamisus Range (approximately at the corner of NE Iran and NW Afghanistan), with the original homeland of this taxon then suggested as being in Transcaucasia or east of the Caspian Sea. From there, mice subsequently spread further northward and to the east and west to colonize the enormous expanse of N and E Eurasia currently inhabited by this species. A population isolated for only a short time is suggested as having given rise to *M. castaneus* in SE Asia and S India. The *M. gentilulus* lineage implicated by the present mtDNA data makes the centrifugal model somewhat more complicated in that it would need to include a fourth movement out of the postulated N Indo-Pakistan cradle area. This model is based chiefly on the following: (1) Among the commensals, the central populations (included under the name *M. castaneus* by us) have the highest nuclear gene variability, as assessed by electrophoresis of proteins and RFLP of *V*β genes, and the deepest clades of mtDNA lineages. (2) There is a 2-million-year-old Mus fossil of the house mouse group in N India. This hypothesized centrifugal model is already becoming accepted in the literature (KSJ 1995).

In their description of the Indian fossil, Patnaik *et al.* (1996) state that their specimen has several diagnostic traits that are absent in any of the living species of the subgenus Mus. Thus, this mouse cannot be the immediate ancestor of the commensal mice because there are at least eight living species in this subgenus that are ancestral to the commensals and to which this fossil mouse is also ancestral. The collective range of these eight species stretches from China to North Africa and W Europe, and, thus, other places might well be where the commensal mice began their evolution. Indeed, the fossils of the most immediate ancestors of the commensal mice are in Europe and North Africa (Jaeger 1975; Jánossy 1975; Auffray and Britton-Davidian 1992).

The other support for the centrifugal model comes from the variability and degree of divergence of nuclear autosomal loci (chiefly allozymes) and mtDNA sequences. Our present study also supports the claim that the greatest divergence within a monophyletic clade of mtDNA molecules exists among the *M. castaneus* mice. But other clades of mtDNA molecules appear to be older than those in the *castaneus* lineage, which implies that they evolved before those in the present-day *M. castaneus*. The mtDNA lineages leading to the *domesticus* and *gentilulus* clades are apparently ancestral to the lineage giving rise to the *castaneus* clade (Figures 8 and 9). Allozyme heterozygosity is not, *per se*, a demonstration of the ancestral condition. Under the neutral model of molecular evolution, high heterozygosity is the result of both population size and persistence time (Kimura 1983). Thus, the high levels of variability in the Indo-Pakistan mice may imply only that there have been large numbers of mice in that area for a long time. They *could* have been living there for an absolutely longer time than anywhere else and be ancestral, but that historical inference is not proven from levels of variability *per se.* Indeed, the three aboriginal species of house mice that are the immediate ancestors of the commensal mice have low levels of allozyme variability (Sage *et al.* 1993).

What are the strongest kinds of evidence that can support a biogeographic model? Fossils and molecular data with a phylogenetic signal are good information for reconstructing this type of historical record. A continuous fossil record in one stratigraphic column showing the transitional morphological types from the ancestral to the modern condition would be the strongest possible proof for the place of origin of a living species. Unfortunately, such series do not exist for the house mouse. The best series are Late Pleistocene Mus fossils in the Near East (Tchernov 1984; Auffray *et al.* 1990b,c), but because the commensal mice began to

evolve and differentiate in the Early and Middle Pleistocene, these Palestinian fossils are not very suggestive of their place of origin. Tchernov (1986) observed that house mouse fossils tended to be uncommon in strata where other rodents are found, which suggests that we are unlikely ever to find the complete series of transitional fossils leading to the commensal mice.

The most powerful kind of molecular information has a clear phylogenetic signal in it, which means that the ancestral/descendant polarity of the variation is apparent. Using such molecular data to infer geographical histories is frequently done (reviewed in Felsenstein 1982; Avise 1994). Perhaps the best known example of such phylogeographical analysis is the model of the African origin of modern humans (Cann *et al.* 1987), which was proposed primarily because the most ancestral mtDNA lineages these investigators found existed in living African peoples. The use of gene frequencies and matrices of genetic distances derived from them for making phylogeographic inferences has several weaknesses involving the methods of data analysis, the sample sizes, the nature of the information (which is essentially phenetic), and, most importantly, the underlying population genetic events leading to the gene differences observed and the distances computed (*e.g.*, see Felsenstein 1982; Davis and Nixon 1992; Cornuet and Luikart 1996; Din *et al.* 1996; references therein).

Despite their shortcomings with respect to statistical support [which likewise beset previous studies (Boursot *et al.* 1996; Boissinot and Boursot 1997)], our trees in Figures 8 and 9 along with Table 3 stimulated us to develop another model for the origin and radiation of commensal house mice for consideration as an alternative hypothesis to the centrifugal model. We used the phylogeographic approach and assumptions of Avise (1994) to infer the sequence and direction of spread of the mice themselves from the geographic patterning of the mtDNA genes. These assumptions are the following: (1) mitochondrial-gene trees are likely to represent the species tree; (2) in a broad sense, genes originated in the place where the present-day carriers of particular gene lineages live; and (3) spreading of mice carrying the genes of different lineages, rather than gene flow into already established populations, is responsible for the geographic patterning of variation.

The sequential or linear model that we propose postulates a western origin within the range of present-day *M. domesticus* followed by an easterly, arcing spread of new mouse populations to give rise to the progenitors of the other species. We constructed this scenario for the origin and historical route of spreading of commensal mice in Eurasia from the assumption of the relative ages of the mouse lineages inferred from the relative ages implied by the mtDNA trees in Figures 8 and 9. Though they lack the Yemeni mtDNA lineage, the midpoint-folded trees of Boursot *et al.* (1996) and outgroup-rooted tree of Boissinot and Boursot (1997) agree with our trees in having *domesticus* mtDNAs as the sister group to the other commensal mtDNAs. [Boursot *et al.* (1996) suggested placing the root within the *castaneus* lineages, based on She *et al.* (1990), but both of these reports emphasized the uncertainty in root assignment.] Their trees share with ours consistent support for the monophyly of the *musculus* and of the *domesticus* mtDNAs, lack of such support for monophyly of the *castaneus* mtDNAs, and uncertainty in the branching order of the major commensal mtDNA lineages and placement of the root. We do not claim that our trees resolve these questions with more significant support than do those published earlier. Rather, we have used a phylogenetic tree as the foundation for an alternative hypothesis instead of using mainly nontree criteria as was done in the development of the centrifugal model (see above). The Yemeni mtDNA lineage increases the plausibility of considering a tree-based hypothesis.

Our sequential model begins with pre-*M. domesticus* mice arising in WC Asia, within the current range of the mice identified as *M. domesticus* (including subspecies *domesticus*, *brevirostris*, and *praetextus*). Because these mice live so well and are presently most abundant in oases or wet places in arid lands, the ancestral populations may have lived in the Tigris-Euphrates River Valley (*i.e.*, in Mesopotamia). Paleobiological studies suggest that this area has maintained its arid steppe and riverine environments throughout the Pleistocene (Frenzel *et al.* 1992; Vrba *et al.* 1995). The Tigris-Euphrates River Valley could have served as a continuous home to this mouse species. But, given that the deepest lineages in the mtDNA tree in Figure 7 are from around the Mediterranean and we have examined few samples from the Mesopotamian region, we do not rule out, for example, the Nile River Valley as the possible pre-*M. domesticus* homeland. We postulate that the ancestors of *M. gentilulus*, the group that now lives in the southern Arabian Peninsula and has the next-oldest category of mtDNA, moved south from Mesopotamia at a time when desert conditions were not as extreme as they are now. At various times during the Pleistocene, the entire Arabian Peninsula was wetter and more hospitable than it is today (Ripley 1954; Gasperetti 1988).

Our proposed model continues with mice from southern Arabia moving eastward and northward to establish the *M. castaneus-M. musculus* ancestor. The most direct path for the dispersal of *M. gentilulus* mice to the Indian subcontinent would be to have crossed the area where the Strait of Hormuz is presently located (joining the Persian Gulf and the Gulf of Oman). Mice might have rafted across this narrow water barrier (now only 70 km wide) or perhaps had a land route available as a result of sea level lowering, which led to emptying of the Persian Gulf such that the two regions were separated only by the freshwater flow of the Tigris and Euphrates Rivers (Kassler 1973). House mice currently do well in

both salt- and freshwater marshlands in California and the Near East (Sage 1981), so rafting dispersal across a flooding Tigris-Euphrates River may have been a frequent event. An all-land route, with the spread of mouse populations north and then east around the present-day Persian Gulf, seems less likely for two reasons. First, it would have required that *M. domesticus* be displaced from and then return back into the southern Mesopotamian valley region. Also, this route is much longer and requires population expansion through the southern part of the Zagros Mountains, where forests would not be a preferred habitat of feral commensal mice. [The progenitors of the *M. castaneus-M. musculus* stock could also have reached the Indian subcontinental region if two groups at the periphery of the area of origin moved in different directions, one southward and southwestward (to give rise to *M. gentilulus*) and the other eastward (by the all-land route described above) or southward and then eastward across the Strait of Hormuz area. We recognize that this option implies a centrifugal model rather than a linear one.]

In the last aspect of this reconstruction, which is the history of separation of the ancestral *M. castaneus-M. musculus* stock into the two modern species, our proposed scenario is the same as or similar to parts of the centrifugal model. We propose that this ancestral stock spread and occupied the entire Indo-Pakistan area south of three transverse mountain massifs (the Kopet Dagh, the series of ranges from the Paropamisus to the Hindu Kush, and the Himalayas) that separate the SC and NC Asian lowlands. They became the ancestors of this region's present-day *M. castaneus* mice and probably occupied this large area for a comparatively long period of the Pleistocene because there were always large areas of this southland warm enough to support mouse populations. Soon after these ancestral mice occupied the area, a population moved through the mountains into the steppe regions on the north side. This passage, which probably occurred during an interglacial period, may well have been by their dispersing through the Hari River Valley in NW Afghanistan bordering NE Iran. This river system runs between the Kopet Dagh and Paropamisus Mountains. Somewhat to the east, the Amu Darya River system drains the northern slopes of the Hindu Kush mountains of Afghanistan, where present-day mice have *musculus* mtDNAs. A different crossing point from SC to NC Asia could be envisioned somewhat to the west, between the Elburz and Kopet Dagh Mountains and along the SE coast of the Caspian Sea. From this NC location, the mice bearing *musculus* mtDNAs ultimately spread west to central Europe and east to China and Japan.

We propose that the Indo-Pakistan stock then evolved the modern *castaneus* types of mtDNA, as well as a number of distinctive morphological types in this region of much geographic variability. These include the distinctive form called *homourus* in the highlands in and adjacent to Nepal and the form called *castaneus* in the humid lowlands of SE Asia. Most recently, populations spread into SE Asia, carrying a limited diversity of these mtDNA molecules.

The model we propose implies that generation of the deleted states of the *Y* chromosome (*Zfy-2* shorter by 18 bp) and Ψ*p53-1* (absence of the locus) occurred after the ancestral stock arrived in the southern Indo-Pakistan area, so that both loci became polymorphic for the two conditions. Generation of new mutations and persistence of polymorphisms are likelier in the larger populations presumed to have occupied this region. Maintenance of polymorphisms plus sorting and filtering of ancestral lineages (as outlined in results for the *castaneus* mtDNA lineage found in SE Asia) may explain the geographic pattern of variation observed today. Evidence of such ancestral polymorphism is apparent in Iran, Afghanistan, and Pakistan, notably including *castaneus* mtDNA and the *Y* B allele in mice from SC Afghan locality 34 and SC Iranian locality 22 (Table 1) and in several individuals from NC and NE Iran studied by Boissinot and Boursot (1997). This same mtDNA and *Y* combination is found also in extreme SE Asia. Our model does not require secondary sweeps, as proposed by Boissinot and Boursot (1997) to explain the observed distribution of *Y*s with the A and B alleles. (The possible residual polymorphism for P and N at Ψ*p53-1* in mice that otherwise appear to be *M. musculus* at Turkmen locality 29 near the Tedzhen River, which is the northern end of the Hari River, would be consistent with proximity to the initial crossing point from SC to NC Asia.)

That the geographic ranges of the species that are the closest living relatives of the commensal mice are in SW Eurasia provides additional support in favor of a western origin as opposed to an Indo-Pakistan cradle. *M. macedonicus* and *M. spicilegus* occur, respectively, from Macedonia to W Iran and in steppe habitats from SE Austria to the Black Sea. *M. spretus* ranges around the western end of the Mediterranean Sea. The present-day range of *M. domesticus* thus overlaps completely with those of *M. macedonicus* and *M. spretus*, which might suggest that *M. domesticus* also arose in this western area rather than far away from its closest relatives. The centrifugal model requires assuming that the species ancestral to the aboriginal house mouse species lived in the Indo-Pakistan region long enough to have produced another lineage that would become the precommensal lineage and that the whole aboriginal stock then went extinct throughout the entire Indo-Pakistan area, surviving only in the Near Eastern and central European steppelands. Only more distant relatives of the commensal mice (*e.g.*, *M. terricolor* and *M. booduga*) have ranges close to the lands considered ancestral to the commensals in the centrifugal model.

The two models make different and testable predictions about the relative branching order of gene trees

made from commensal mouse DNA sequences and about the geographic location of the oldest fossil remains of these mice. The oldest fossil bones that are morphologically assignable to commensal mice should be found in W Eurasia under the linear model. Under the centrifugal model, these fossils are expected to be found in the Indo-Pakistan area. When DNA sequences with adequate amounts of phylogenetic information are available, the linear model predicts that the *M. domesticus* sequences will be ancestral to those from *M. castaneus* mice, while the centrifugal model predicts the opposite branching order. To date, cladistic analyses of mtDNAs (Boissinot and Boursot 1997; this report), *Y* chromosome DNA (Tucker *et al.* 1989), and *Y* chromosomal *Sry* genes [E. M. Prager, unpublished results based on the GenBank sequences of Albrecht and Eicher (1997)] favor the branching order predicted by the linear model. (However, the cited *Y* chromosome studies examined *M. castaneus* only from one SE Asian locality.)

***p53* pseudogene polymorphisms:** Findings reported here indicate a need to consider three kinds of polymorphism for processed *p53* pseudogenes in the house mouse genome: presence *vs.* absence at a given locus, number of alleles at one locus, and number of Ψ*p53* loci. In a survey of one or two individuals per species, Ohtsuka *et al.* (1996) found that outside the house mouse complex, Ψ*p53* is absent in *M. caroli*, present in *M. booduga* (*M. leggada* in their nomenclature), and absent in *M. platythrix* (which is in another subgenus). Their phylogenetic analyses make it reasonable to assume that the *M. booduga* Ψ*p53* lies at the same place, on chromosome *17*, as mapped for laboratory strains of *M. domesticus* (*i.e.*, at Ψ*p53-1*). As outlined in results, Ohtsuka *et al.* (1996) favored only losses rather than maintenance of an ancient trans-species polymorphism to explain absence of a Ψ*p53* in several mouse lineages. Among the commensals and in light of the linear biogeographic model of origin and radiation, we postulated one loss and then maintenance of the presence/absence polymorphism in *M. castaneus* and lineage sorting or filtering to give only absence (or a low level of polymorphism) in *M. musculus*. However, we cannot rule out multiple independent losses, particularly among the large and collectively diverse *M. castaneus* populations. Furthermore, genomes of different taxa may differ with respect to ease of loss of Ψ*p53-1*.

Our demonstration of at least 14 alleles at one nuclear locus seems indicative of an unusually high level of variability. However, because much of this Ψ*p53-1* variability is geographically partitioned among different taxa and collectively encompasses an enormous territory, this number of alleles inferred at a locus presumably free of functional constraints may not be surprisingly large. The rarer Ψ*p53-1* alleles may serve as useful markers for the timing and routes of spreading of diverse populations, and they may also provide insight into rates of evolution at this locus. Our evidence for a second and likely a third Ψ*p53* locus in house mice suggests that *p53* pseudogene generation and integration may be facile. It invites mapping of the new locus (or loci) and investigation into the presumably viral mediators of the requisite reverse transcription and their geographic and phylogenetic distribution among house mice. Multiple loci and possibly repeated losses at a given locus among commensal mice dictate caution in using scoring for the presence/absence of a Ψ*p53*.

Future directions: A correct understanding of the evolutionary history of commensal house mice is needed because these are the animals that gave rise, via interspecific hybridization by early mouse breeders, to the highly variable inbred strains of laboratory mice that are central to much research on genetic interactions during mammalian development (Sage *et al.* 1993). As the role of gene-gene interactions in development and physiology becomes better understood in mice, researchers need to be alert as to whether the interactions are the result of intra- or interspecific combinations of alleles at the interacting loci.

Our contribution of a new model of commensal mouse origins makes it appropriate to do future comparative molecular surveys in a way that will test the phylogenetic relationships of alleles as predicted by the contrasting models. They should be done using cladistic methods and should use a minimum of four mouse stocks, including at least one aboriginal species as a close outgroup sample and at least one authentic *M. domesticus*, *M. castaneus*, and *M. musculus* (all of which are commercially available, as are their DNAs). The recent availability of some 30 inbred strains from India (K.S.J. 1995) facilitates including members from the center of the highly diverse *M. castaneus* phylogeographic unit. Bringing *M. gentilulus* into laboratory culture for molecular genetic and other studies emerges as a goal from our present investigation.

The work described here provides a stimulus for further work in at least four different arenas. First, additional mouse populations need to be sampled for mtDNA and other genetic analyses, with priority areas being Iraq, the Arabian Peninsula, East Africa, Iran, and along the southern slopes of the Himalayas to Burma. Notably, Iraqi mice need to be surveyed to test the supposition that they have *domesticus* mtDNAs. Second, longer mtDNA sequences, maybe even whole genomes, should be obtained from representatives of all the major commensal mtDNA lineages now identified (preferably including two deep lineages from those in Figure 6) and from the aboriginal species to try to determine definitively the branching order and root position in Figure 8. Third, the generation and maintenance of *p53* pseudogene diversity require elucidation. Cloning and sequencing of PCR products in cases of sequence phenotypes polymorphic at two or more positions are needed to determine allele sequences directly. Sequenc-

ing longer stretches of Ψ*p53-1*, and from more than the 79 commensal mice we surveyed, may yield a better estimate of the actual diversity and permit relating the alleles phylogenetically with greater resolution. Finally, because many future surveys will probably have to depend at least in part on museum skins as the source of genetic information, development of DNA markers for a variety of additional nuclear loci merits attention. Microsatellite loci in general should be accessible via museum skins, as pieces of 100–250 bp are frequently amplified. Diagnostic loci well known from protein electrophoresis may also become assessable at the DNA level.

The late A. C. Wilson provided essential ideas for this survey, notably including Yemen along with the central areas from Afghanistan to Turkey and using museum skins to retrieve mtDNA sequences. We deeply appreciate the efforts of all the mouse collectors, and are especially grateful to L. Heaney, B. D. Patterson, J. D. Phelps, and W. T. Stanley of the Field Museum of Natural History for cutting and sending skin samples, providing information about localities and anatomical measurements, and sending specimens of whole animals on loan and extending hospitality to J. T. Marshall. We are particularly thankful to J. T. Marshall for anatomical assessments of the museum specimens we studied molecularly, communication of unpublished results based on anatomical assessments of many hundreds of other museum specimens, and valuable discussions. We thank P. K. Tucker and J. Arnold for transmitting frozen tissues; C. Cicero, J. L. Patton, and B. R. Stein for providing access to specimens housed at the MVZ at U.C. Berkeley; P. Boursot for communicating an earlier, unpublished version of Boissinot and Boursot (1997), in which mtDNA variation was assessed by restriction analysis; and P. Boursot, A. E. Leviton, S. J. Mack, T. J. Papenfuss, J. L. Patton, T. W. Quinn, A. Sidow, M. Slatkin, W. K. Thomas, G. Thomson, and D. L. Yowe for helpful discussion. E.M.P. thanks K. Le, W. K. Thomas, D. B. Wake, and the MVZ for reagents and supplies, as well as the laboratory of B. N. Ames for use of the Perkin Elmer 480 thermal cycler. This work received support from National Science Foundation grants BSR88-18053 and DEB90-08912 to A.C.W. and (for preparation of genomic DNAs) National Institutes of Health grant R01 AI29800 to R.D.S.

LITERATURE CITED

Albrecht, K. H., and E. M. Eicher, 1997 DNA sequence analysis of *Sry* alleles (subgenus Mus) implicates misregulation as the cause of C57BL/6J-Y^{POS} sex reversal and defines the SRY functional unit. Genetics **147:** 1267–1277.

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

Auffray, J.-C., J. T. Marshall, L. Thaler and F. Bonhomme, 1990a Focus on the nomenclature of European species of *Mus*. Mouse Genome **88:** 7–8.

Auffray, J.-C., E. Tchernov, F. Bonhomme, G. Heth, S. Simson *et al.*, 1990b Presence and ecological distribution of *Mus "spretoides"* and *Mus musculus domesticus* in Israel. Circum-Mediterranean vicariance in the genus *Mus*. Z. Säugetierkd. **55:** 1–10.

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

Avise, J. C., 1994 *Molecular Markers, Natural History and Evolution*. Chapman & Hall, New York.

Boissinot, S., and P. Boursot, 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., R. Anand, D. Darviche, W. Din and P. Boursot, 1994 The house mouse as a ring species?, pp. 13–23 in *Genetics in Wild Mice*, edited by K. Moriwaki, T. Shiroishi and H. Yonekawa. Japan Sci. Soc. Press, Tokyo/S. Karger, Basel, Switzerland.

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

Boursot, P., W. Din, R. Anand, D. Darviche, B. Dod *et al.*, 1996 Origin and radiation of the house mouse: mitochondrial DNA phylogeny. J. Evol. Biol. **9:** 391–415.

Cann, R. L., M. Stoneking and A. C. Wilson, 1987 Mitochondrial DNA and human evolution. Nature **325:** 31–36.

Cornuet, J. M., and G. Luikart, 1996 Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics **144:** 2001–2014.

Davis, J. I., and K. C. Nixon, 1992 Populations, genetic variation, and the delimitation of phylogenetic species. Syst. Biol. **41:** 421–435.

Din, W., R. Anand, P. Boursot, D. Darviche, B. Dod *et al.*, 1996 Origin and radiation of the house mouse: clues from nuclear genes. J. Evol. Biol. **9:** 519–539.

Felsenstein, J., 1982 How can we infer geography and history from gene frequencies? J. Theor. Biol. **96:** 9–20.

Ferris, S. D., R. D. Sage, E. M. Prager, U. Ritte and A. C. Wilson, 1983 Mitochondrial DNA evolution in mice. Genetics **105:** 681–721.

Frenzel, B., M. Pécsi and A. A. Velichko (Editors), 1992 *Atlas of Paleoclimates and Paleoenvironments of the Northern Hemisphere*. Hungarian Academy of Sciences, Bern, Switzerland.

Frisman, L. V., K. V. Korobitsina, L. V. Yakimenko, F. M. Bokshtein and A. I. Muntyanu, 1990 Genetic differentiation of U.S.S.R. house mice: electrophoretic study of proteins. Biol. J. Linn. Soc. **41:** 65–72.

Gasperetti, J., 1988 Snakes of Arabia, pp. 169–392 in *Fauna of Saudi Arabia*, Vol. 9, edited by W. Büttiker and F. Krupp. National Commission for Wildlife Conservation and Development, Riyadh, Saudi Arabia, and Pro Entomologia c/o Natural History Museum, Basel, Switzerland.

Gruber, U. F., 1969 Tiergeographische, ökologische und bionomische Untersuchungen an kleinen Säugetieren in Ost-Nepal. Khumbu Himal, Ergebn. Forsch.-Unternehmen Nepal Himalaya **3:** 197–312.

Harrison, D. L., 1972 *The Mammals of Arabia*, Vol. III. Ernest Benn Ltd., London.

Harrison, D. L., and P. J. J. Bates, 1991 *The Mammals of Arabia*, Ed. 2. Harrison Zoological Museum, Sevenoaks, Kent, UK.

Jaeger, J.-J., 1975 The mammalian faunas and hominid fossils of the Middle Pleistocene of the Maghreb, pp. 399–418 in *After the Australopithecines*, edited by K. W. Butzer and G. Ll. Isaac. Mouton Publishers, The Hague.

Jánossy, D., 1975 Mid-Pleistocene microfaunas of continental Europe and adjoining areas, pp. 375–397 in *After the Australopithecines*, edited by K. W. Butzer and G. Ll. Isaac. Mouton Publishers, The Hague.

Kassler, P., 1973 The structural and geomorphic evolution of the Persian Gulf, pp. 11–32 in *The Persian Gulf*, edited by B. H. Purser. Springer-Verlag, New York.

Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK.

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

K. S. J., 1995 How the mouse's tale began life in India. Nature **377:** 188.

Lessa, E. P., M. F. Smith and C. Orrego, 1992 A cool hot start for PCR amplifications. Ancient DNA Newslett. **1**(1): 40–42.

Marshall, J. T., 1981 Taxonomy, pp. 17–26 in *The Mouse in Biomedical Research*, Vol. 1, edited by H. L. Foster, J. D. Small and J. G. Fox. Academic Press, New York.

Marshall, J. T., 1986 Systematics of the genus *Mus*. Curr. Top. Microbiol. Immunol. **127:** 12–18.

Marshall, J. T., and R. D. Sage, 1981 Taxonomy of the house mouse. Symp. Zool. Soc. Lond. **47:** 15–25.

Milishnikov, A. N., A. N. Rafiev, L. A. Lavrenchenko and V. N. Orlov, 1990 A high level of introgression of the genes of *Mus domesticus* in a *Mus musculus* s. str. population of Transcaucasia. Doklady Akademii Nauk SSSR **311:** 764–768.

Milishnikov, A. N., A. N. Rafiev and V. N. Orlov, 1994 Genetic variation in populations of the house mouse *Mus musculus* L.

1758 s. stricto from western, central, and southeastern parts of the species range. Genetika **30:** 906–912.

Moriwaki, K., T. Shiroishi and H. Yonekawa (Editors), 1994 *Genetics in Wild Mice. Its Application to Biomedical Research.* Japan Sci. Soc. Press, Tokyo / S. Karger, Basel, Switzerland.

Nachman, M. W., S. N. Boyer, J. B. Searle and C. F. Aquadro, 1994 Mitochondrial DNA variation and the evolution of Robertsonian chromosomal races of house mice, *Mus domesticus*. Genetics **136:** 1105–1120.

Nagamine, C. M., Y. Nishioka, K. Moriwaki, P. Boursot, F. Bonhomme *et al.*, 1992 The musculus-type Y chromosome of the laboratory mouse is of Asian origin. Mamm. Genome **3:** 84–91.

Nagamine, C. M., P. Boursot, Y.-F. C. Lau and K. Moriwaki, 1994a Evolution of the Y-chromosome in the wild mouse, pp. 41–55 in *Genetics in Wild Mice*, edited by K. Moriwaki, T. Shiroishi and H. Yonekawa. Japan Sci. Soc. Press, Tokyo / S. Karger, Basel, Switzerland.

Nagamine, C. M., T. Shiroishi, N. Miyashita, K. Tsuchiya, H. Ikeda *et al.*, 1994b Distribution of the molossinus allele of *Sry*, the testis-determining gene, in wild mice. Mol. Biol. Evol. **11:** 864–874.

Ohtsuka, H., M. Oyanagi, Y. Mafune, N. Miyashita, T. Shiroishi *et al.*, 1996 The presence / absence polymorphism and evolution of the p53 pseudogene in the genus *Mus*. Mol. Phylo. Evol. **5:** 548–556.

Orth, A., E. Lyapunova, A. Kandaurov, S. Boissinot, P. Boursot *et al.*, 1996 L'espèce polytypique *Mus musculus* en Transcaucasie. C. R. Acad. Sci. Paris, Sci. Vie / Life Sci. **319:** 435–441.

Patnaik, R., J.-C. Auffray, J.-J. Jaeger and A. Sahni, 1996 House mouse ancestor from late Pliocene Siwalik sediments of India. C. R. Acad. Sci. Paris, Sci. Vie / Life Sci. **319:** 431–434.

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

Prager, E. M., P. Boursot and R. D. Sage, 1997 New assays for Y chromosome and p53 pseudogene clines among East Holstein house mice. Mamm. Genome **8:** 279–281.

Ripley, S. D., 1954 Comments on the biogeography of Arabia with particular reference to birds. J. Bombay Nat. Hist. Soc. **52:** 241–248.

Sage, R. D., 1981 Wild mice, pp. 39–90 in *The Mouse in Biomedical Research*, Vol. 1, edited by H. L. Foster, J. D. Small and J. G. Fox. Academic Press, New York.

Sage, R. D., and J. O. Wolff, 1986 Pleistocene glaciations, fluctuating ranges, and low genetic variability in a large mammal (*Ovis dalli*). Evolution **40:** 1092–1095.

Sage, R. D., J. B. Whitney III and A. C. Wilson, 1986 Genetic analysis of a hybrid zone between domesticus and musculus mice (*Mus musculus* complex): hemoglobin polymorphisms. Curr. Top. Microbiol. Immunol. **127:** 75–85.

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

Schwarz, E., and H. K. Schwarz, 1943 The wild and commensal stocks of the house mouse, *Mus musculus* Linnaeus. J. Mammal. **24:** 59–72.

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

Tanooka, H., A. Ootsuyama, T. Shiroishi and K. Moriwaki, 1995 Distribution of the p53 pseudogene among mouse species and subspecies. Mamm. Genome **6:** 360–362.

Tchernov, E., 1984 Commensal animals and human sedentism in the Middle East, pp. 91–115 in *Animals and Archaeology: 3. Early Herders and Their Flocks*, edited by J. Clutton-Brock and C. Grigson. Brit. Arch. Rep. Int. Ser. 202, Oxford, UK.

Tchernov, E., 1986 The rodents and lagomorphs from 'Ubeidiya Formation: systematics, paleoecology and biogeography, pp. 235–350 in *Les Mammifères du Pléistocène Inférieur de la Vallée du Jourdain à Oubeidiyeh. Mémoires et Travaux du Centre de Recherche Français de Jerusalem*, No. 5, edited by E. Tchernov. Association Paléorient, Paris, France.

Tucker, P. K., B. K. Lee and E. M. Eicher, 1989 *Y* chromosome evolution in the subgenus *Mus* (genus *Mus*). Genetics **122:** 169–179.

Vrba, E. S., G. H. Denton, T. C. Partridge and L. H. Burckle (Editors), 1995 *Paleoclimate and Evolution, with Emphasis on Human Origins.* Yale University Press, New Haven, CT.

Weghorst, C. M., G. S. Buzard, R. J. Calvert, J. E. Hulla and J. M. Rice, 1995 Cloning and sequence of a processed *p53* pseudogene from rat: a potential source of false 'mutations' in PCR fragments of tumor DNA. Gene **166:** 317–322.

Yonekawa, H., K. Moriwaki, O. Gotoh, N. Miyashita, Y. Matsushima *et al.*, 1988 Hybrid origin of Japanese mice "*Mus musculus molossinus*": evidence from restriction analysis of mitochondrial DNA. Mol. Biol. Evol. **5:** 63–78.

Communicating editor: W. F. Eanes