Mitochondrial DNA Sequence Variation in the Eastern House Mouse, *Mus musculus:* Comparison With Other House Mice and Report of a 75-bp Tandem Repeat

Ellen M. Prager,* Herbert Tichy† and Richard D. Sage‡,[1]

**Division of Biochemistry and Molecular Biology, University of California, Berkeley, California 94720-3202, †Abteilung Immungenetik, Max-Planck-Institut für Biologie, D-72076 Tübingen, Germany and ‡Division of Biological Sciences, University of Missouri, Columbia, Missouri 65211*

Manuscript received May 30, 1995
Accepted for publication February 9, 1996

ABSTRACT

The control region and flanking tRNAs were sequenced from 139 *Mus musculus* mitochondrial DNAs (mtDNAs) from mice collected at 44 localities extending from Germany to Japan. Among the 36 types of *M. musculus* mtDNA resolved, five have an added 75-bp direct repeat; the two copies within an individual differ by two to four base substitutions. Among 90 *M. domesticus* mtDNAs sequenced, 12 new types were found; 96 *M. domesticus* types have now been identified by sequencing this segment. Representative mtDNAs from *M. castaneus, M. macedonicus, M. spicilegus* and *M. spretus* were also sequenced. A parsimony tree for the *M. musculus* mtDNAs is about half as deep as the tree for the *M. domesticus* mtDNAs, which is consistent with the idea that *M. musculus* is genetically less diverse and younger than *M. domesticus*. The patterns of variation as a function of position are similar but not identical in *M. musculus* and *M. domesticus* mtDNAs. *M. castaneus* and *M. musculus* mtDNAs are allied, at a tree depth about three times as great as the start of intra-*M. musculus* divergence. The coalescence of the *M. musculus* and *M. castaneus* mtDNAs is about half as deep as their coalescence with the *M. domesticus* mtDNA lineages. The mtDNAs of the aboriginal *M. macedonicus* and *M. spicilegus* are each other's closest relatives, at a tree depth greater than the deepest intracommensal node. The mtDNA results support the view that the aboriginal *M. spretus* is the sister group of the other five species.

MEMBERS of the house mouse species group are the best known vertebrate model organisms for a wide variety of fields, including molecular, chromosomal and organismal evolution, molecular biology, physiology, immunology, genetics, development, population biology, speciation, and ecology (see, *e.g.,* BOURSOT *et al.* 1993; SAGE *et al.* 1993; MORIWAKI *et al.* 1994). The house mouse complex is now viewed as consisting of three or four commensal species, *M. domesticus, M. musculus, M. castaneus,* and *M. bactrianus,* plus three aboriginal species, *M. spicilegus, M. macedonicus,* and *M. spretus* (BOURSOT *et al.* 1993; SAGE *et al.* 1993). The classical strains of laboratory mice carry a single type of *M. domesticus* mtDNA and primarily *M. domesticus* nuclear genes, but some of their nuclear genes, as well as the *Y* chromosome, are derived from other commensal taxa, notably the populations of eastern Asia. Laboratory strains have been established from diverse wild-caught representatives of most of the species in the complex. Indeed, analysis of DNA from crosses and backcrosses involving *M. domesticus* and *M. spretus* is a major tool in high-resolution mapping of the entire mouse genome (COPELAND *et al.* 1993).

Corresponding author: Ellen M. Prager, Division of Biochemistry and Molecular Biology, Barker Hall, University of California, Berkeley, CA 94720-3202.

[1] *Present address:* 131 Sierra Vista, Solvang, CA 93463.

The native range of *M. domesticus,* which is often referred to as the western European house mouse, is western Europe, North Africa, and the Middle East and likely extends into Iran (MARSHALL 1981; BOURSOT *et al.* 1993). It is this taxon that humans during the last 500 years spread worldwide, notably to the Americas (TICHY *et al.* 1994), Australia, and throughout Africa. *M. musculus,* known as the eastern European house mouse, has a native range encompassing eastern Europe and all of northern Asia. *M. castaneus* is found primarily in southeastern Asia. Less genetic information is available for *M. bactrianus,* which is described from the western Indian subcontinent, but it appears likely to be a distinct taxon (BOURSOT *et al.* 1993 and refs. therein). Many workers (*e.g.,* BOURSOT *et al.* 1993; BONHOMME *et al.* 1994; refs. therein) regard these commensal taxa as subspecies; for the reasons recently discussed by PRAGER *et al.* (1993) and SAGE *et al.* (1993), we denote them as full species. The three aboriginal species each occupy relatively limited ranges in Europe, North Africa, and western Asia, essentially allopatric of one another but sympatric with the resident commensal species.

Where *M. domesticus* and *M. musculus* meet in Europe, there is a narrow hybrid zone. SAGE *et al.* (1993) and BOURSOT *et al.* (1993) have reviewed studies at multiple levels of biological organization that have been carried out across four transects, ranging from northern Den-

mark to eastern Bulgaria, and such studies of transect animals are continuing (*e.g.*, ALIBERT *et al.* 1994; MACHOLÁN and ZIMA 1994; DALLAS *et al.* 1995; this report). It appears increasingly likely that there are hybrid zones or areas of introgression wherever any two (or more) of the commensal species meet. While *M. domesticus* has been well studied throughout its range, only recently have genetic data been accumulating for a large part of the range of *M. musculus, i.e.*, central and much of eastern Asia (*e.g.*, FRISMAN *et al.* 1990; RUVINSKY *et al.* 1991; MEZHZHERIN and KOTENKOVA 1992; KOROBITSYNA *et al.* 1993; NAGAMINE *et al.* 1994; YONEKAWA *et al.* 1994). The Indian subcontinent is now also receiving increasing attention (BOURSOT *et al.* 1993, 1996; DIN *et al.* 1996).

Optimal use of house mice as model organisms and optimal interpretation of experimental observations require knowledge of the genetic variation existing in nature and of the phylogenetic relationships of the several species. Despite great progress made during the past two decades, a number of questions exist with respect to the extent and organization of variation in wild house mice. The present report addresses four such questions from the viewpoint of mtDNA with an emphasis on *M. musculus*.

One area of inquiry concerns the relative genetic diversity of *M. musculus* compared to *M. domesticus*. There is some evidence from *t* haplotypes and nuclear DNA restriction fragment length polymorphisms that *M. musculus* may be genetically more uniform (FIGUEROA *et al.* 1987; KLEIN *et al.* 1987, 1988; RUVINSKY *et al.* 1991), while heterozygosity assessed by protein electrophoresis (summarized in SAGE *et al.* 1993) and microsatellite heterozygosity on either side of the Danish hybrid zone (DALLAS *et al.* 1995) implied little or no difference in variability between the two species. An early high-resolution mtDNA restriction analysis study involving only a few *M. musculus* specimens (FERRIS *et al.* 1983) suggested equal depths for the *M. domesticus* and *M. musculus* mtDNA trees and similar intraspecific pairwise differences. A recent, comprehensive restriction study (BOURSOT *et al.* 1996), in contrast, implies that *M. musculus* is mitochondrially less diverse.

While it is now accepted that the commensal taxa form a phylogenetic clade to the exclusion of the aboriginal taxa, questions remain about the relationship of the commensal taxa to one another. Leaving aside the less-studied *M. bactrianus*, there is appreciable evidence favoring a closer association of *M. musculus* and *M. castaneus* to the exclusion of *M. domesticus* from a variety of studies, including electrophoresis of proteins encoded by autosomal genes and analyses of mtDNA, the *Y* chromosome, and a pseudogene locus (*e.g.*, SAGE 1981; MORIWAKI *et al.* 1984; TUCKER *et al.* 1989; BOURSOT *et al.* 1993; HAMMER and SILVER 1993; SAGE *et al.* 1993; NAGAMINE *et al.* 1994; YONEKAWA *et al.* 1994; refs. therein). Yet the relationship is often represented as a trichotomy (*e.g.*, HAMMER and WILSON 1987; DOD *et al.* 1989; BOURSOT *et al.* 1993; HAMMER and SILVER 1993).

With respect to the aboriginal species, there seems to be a consensus that *M. spicilegus* and *M. macedonicus* form a clade (BOURSOT *et al.* 1993; SAGE *et al.* 1993), though divergence of the mitochondrial genomes of these two species does not appear to have been temporally concordant with divergence of their nuclear genomes (FORT *et al.* 1985; BONHOMME 1986; SHE *et al.* 1990). The position of *M. spretus*, however, is a third unresolved issue. On the basis of protein electrophoresis, SAGE (1981; cf. also SAGE *et al.* 1993) associated *M. spretus* with *M. spicilegus* and *M. macedonicus* in an aboriginal clade, an arrangement receiving some support from nuclear gene sequences (MORITA *et al.* 1992). Other protein electrophoretic analyses (BONHOMME *et al.* 1984), some mtDNA data (SHE *et al.* 1990), studies of satellite DNA amplification (BONHOMME 1986; DOD *et al.* 1989), and restriction analysis of nuclear rDNA spacer regions (SUZUKI and KURIHARA 1994) suggested that *M. spretus* could be an outgroup to all the other house mouse taxa, and this evolutionary framework has sometimes been used (*e.g.*, HAMMER and WILSON 1987; HAMMER and SILVER 1993). A trichotomy (commensals, *M. spicilegus* + *M. macedonicus*, *M. spretus*) has often been the consensus or default arrangement depicted (*e.g.*, BONHOMME 1986; DOD *et al.* 1989; SHE *et al.* 1990; AGULNIK *et al.* 1993; BOURSOT *et al.* 1993).

A fourth question concerns the time that house mice colonized Europe. The more widely accepted view is that colonization of the western European mainland was recent, roughly 5000 years ago and in concert with the spread of human activities, including agriculture and seafaring (KLEIN *et al.* 1987; AUFFRAY *et al.* 1990; BRITTON-DAVIDIAN 1990; SAGE *et al.* 1990; AUFFRAY and BRITTON-DAVIDIAN 1992; refs. therein). Though there is agreement that house mice colonized northwestern Europe only recently ($\leq$5000 years ago; cf. PRAGER *et al.* 1993), the fact that in benign climates *M. domesticus* can and does live independently of humans (SAGE 1981; SAGE *et al.* 1993; refs. therein) permits debate about when these animals colonized the western Mediterranean region. SAGE *et al.* (1990) used their results from high-resolution restriction mapping of *M. domesticus* mtDNAs (albeit calibrated with rates derived for creatures other than murine rodents) to suggest that mice might have arrived in Europe much earlier, some 30,000–40,000 years ago in concert with the arrival and rapid expansion of populations of modern humans (see, *e.g.*, DI RIENZO and WILSON 1991). SAGE *et al.* (1990) drew support from genetic and ecological data and pointed out that the absence of a fossil record for house mice could be explained by the rarity of fossilization. AUFFRAY and BRITTON-DAVIDIAN (1992) have advanced cogent arguments based on evidence from paleontological and genetic investigations in urging rejection of the early colonization hypothesis.

To examine these issues, we chose sequencing of a rapidly evolving part of the mitochondrial genome that had proven capable of resolving a large number of dis-

tinct *M. domesticus* mtDNAs (PRAGER *et al.* 1993; NACHMAN *et al.* 1994). Our primary goal was to gather sequence data for *M. musculus* mtDNAs comparable to those available for *M. domesticus* mtDNAs. Another goal was to test the robustness of the phylogenetic and related inferences in PRAGER *et al.* (1993) by increasing the *M. domesticus* dataset by >70% with respect to number of *M. domesticus* mtDNA types and number of localities. Among our 237 study animals are >150 *M. domesticus, M. musculus,* and hybrids from a southern German-Austrian (Bavarian) transect; for these mice, we currently report only the essence of the mtDNA variation (*cf.* legend to Table 1). This investigation provided also the opportunity to compare and contrast the patterns of mtDNA sequence variation as a function of position in two closely related species of mice. Such an analysis may suggest possible selective advantages or disadvantages of a particular mitochondrial genome, especially in hybrid or introgressed populations where the nuclear genome may be predominantly that of another species. Finally, this project led to the discovery of the tandem duplication of a 75-bp segment of the mtDNA control region at relatively high frequency in *M. musculus.*

MATERIALS AND METHODS

Mouse collection and DNA preparation: This sequencing study involved 237 mice collected at 73 localities, 45 of them in or near Bavaria and the remainder extending across Europe and Asia (Table 1, Figure 1). Commensal mice were trapped in or near buildings on farms and in villages, towns, and cities; aboriginal mice were collected in fields. A few laboratory descendants of separate lines of wild-caught females were used, namely mice with *M. musculus* mtDNA from localities 7, 8, 15 (two of 14 individuals), and 28, as well as one or two individuals with *M. castaneus, M. macedonicus, M. spicilegus* (from localities 11 and 21), and *M. spretus* mtDNAs (Table 1). R. D. SAGE trapped the animals from the Bavarian transect in 1984, 1985, and 1992, from localities 9 and 11 in 1979, and from localities 10, 12, and 13 in 1992; he obtained those from localities 3 and 15–17 from other collectors in 1979 and 1982 (cf. FERRIS *et al.* 1983). U. GYLLENSTEN provided purified mtDNA or genomic DNA from mice from locality 2. H. TICHY collected or obtained from other collectors the remaining 37 animals, from localities 4–6, 14, and 19–26, during the 1980s. Voucher specimens are deposited at the Museum of Vertebrate Zoology of the University of California in Berkeley and in the Abteilung Immungenetik of the Max-Planck-Institut für Biologie in Tübingen.

Total genomic DNA was prepared from liver or other tissues of 151 of the 237 mice according to standard procedures similar to that described before (PRAGER *et al.* 1993). Highly purified mtDNA was available from 86 individuals from previous and ongoing projects (FERRIS *et al.* 1983; SAGE *et al.* 1990; PRAGER *et al.* 1993; legend to Table 1): 73 mtDNAs from the Bavarian transect, three from Norway, *M. musculus* mtDNAs from localities 3, 7, 8, 15, and 28, and one mtDNA of each of the last four species listed in Table 1.

PCR amplification and sequencing of mtDNA: The region of interest (Figure 2) was amplified in two portions, single-stranded DNA was generated, and templates were sequenced using primers and procedures described before (PRAGER *et al.* 1993) and slight modifications thereof. The general protocol involved double-stranded amplifications with primers 1B (L15320) plus 4 (H15782) and 3 (L15735) plus 9B (H00072) and sequencing of single-stranded templates generated from one strand with primers 1B, 5 (H16057), and 8 (H00066) and where necessary with primers 2 (L15492), 6 (H16125), and 7 (H16154). L, H, and the numbers refer, respectively, to the light and heavy strands and the positions of the 3′ base. From most of the purified mtDNAs from Bavarian transect mice, single-stranded sequencing templates for both portions were amplified directly from total mtDNA with 37–39 amplification cycles. For six samples, representatives of *M. musculus* mtDNA types 1 and 24 plus types 1 from *M. castaneus, M. macedonicus, M. spicilegus,* and *M. spretus,* the opposite strand of the 5′ portion of this 1-kb region was prepared and sequenced with primer 4 and, for *M. musculus* type 1, also with primer H15720 (5′-CCGGAGCGAGAAGAGGGGCA-3′).

An average of 1005 bp was read per individual for *M. musculus* mtDNAs (n = 139), 1008 bp for *M. domesticus* mtDNAs [n = 342, from PRAGER *et al.* (1993), DALLAS *et al.* (1995), and this report], and 974 bp for mtDNAs from representatives of four other species (n = 8; see Figure 2 and Table 1). One compression, characterized by PRAGER *et al.* (1993), occurred invariably. GenBank accession nos. are U47430-U47497 for the 68 types of *M. domesticus* sequences we determined [1–10 and 12–56 from PRAGER *et al.* (1993) and 57–69 from the present report], U47498-U47533 for the 36 types of *M. musculus* sequences, and U47534-U47539 for the six types of sequences from other taxa.

mtDNA sequences from the literature: For phylogenetic analyses, *M. domesticus* mtDNA sequence types 1–56 (PRAGER *et al.* 1993) and 27 additional distinct *M. domesticus* sequences from NACHMAN *et al.* (1994), types 70–96 in Table 2, were included along with *M. domesticus* types 57–69 reported here. The extreme 3′ ends of representatives of nine types among 1–56 were reanalyzed, so that nearly all the unsequenced variable sites in PRAGER *et al.* (1993) could be filled in as follows: at position 00052, A in types 4, 6, 10, 26, 31, 32, 52, and 56 and G in type 9; at position 00055, G in types 4 and 6, A in 26, T in 31, 32, and 52, and C in 56. In type 26 we found a variant base, C, at position 00056, rather than the T found in 64 of our 69 types (with types 7, 17, 41, and 50 unsequenced at this position). NACHMAN *et al.* (1994) do not report on variation in the far 3′, Phe tRNA, portion of this region; we assumed particular bases at phylogenetically informative sites in this portion as described in the next section. The *M. spretus* mtDNA sequence reported by NACHMAN and AQUADRO (1994) for positions 15352–16142 was denoted type 2; for analytical purposes it was taken as matching type 1 (Figure 2) in regions 5′ and 3′ of the published sequence.

Calculations: Character-state parsimony trees for mtDNAs were constructed with the SWOFFORD (1991) PAUP (Phylogenetic Analysis Using Parsimony) program as described previously (PRAGER *et al.* 1993). All character changes were weighted equally. The bootstrapping option was used for the analysis of 10 sequences representing six species. Because undetermined residues can lead to generation of more equally parsimonious trees and to impractically long computer runs, the likely base at missing variable sites was assumed, as detailed below. The same assumptions were included for computation of pairwise differences. The probability of incorrect assignments at unsequenced sites is low, as the tree placements of the mtDNA types affected were nearly always well defined on the basis of their known sequences at other variable sites. The consequences of an incorrect assignment would most often be failure to detect an extra sequence change; occasionally increased resolution (*i.e.,* an additional branching point) or minor differences in the arrangement of lineages at the tips of trees could be overlooked, but only rarely (for <10% of the assumptions) is there a potential for rearrangements involving deeper lineages.

TABLE 1

mtDNA types and collecting localities

mtDNA type	No. of mice	Locality	Map no.
		M. musculus mtDNA	
1	1[a]	Studenec, near Brno, Czech Republic	7
	3	Petrov, near Prague, Czech Republic[b]	6
	1	Stemplovec, near Opava, Czech Republic	5
2	9	1 in Bavarian transect	B
3	1	Sládečkovce, near Bratislava, Slovakia	8
	2	Halbturn, Austria	10
	8	2 in Bavarian transect	B
4	2	1 in Bavarian transect	B
5	2	1 in Bavarian transect	B
6	1	1 in Bavarian transect	B
7	2	Mönchhof, Austria	9
	1[c]	Čakovec, Croatia	14
	16	9 in Bavarian transect	B
8	4	1 in Bavarian transect	B
9	9	2 in Bavarian transect	B
10	1	1 in Bavarian transect	B
11	3	2 in Bavarian transect	B
12	2	1 in Bavarian transect	B
13	1	1 in Bavarian transect	B
14	5	1 in Bavarian transect	B
15	2	2 in Bavarian transect	B
16	1	1 in Bavarian transect	B
17	4	1 in Bavarian transect	B
18	3	1 in Bavarian transect	B
19	1	1 in Bavarian transect	B
20	3	Stemplovec, Czech Republic	5
21	2	Tedzhen, Turkmenistan	25
22	1	Askanianova, Ukraine	20
	1	Dshankoi, Crimea, Ukraine	21
23	3	Sulak, Daghestan, Russia	23
	1	Sary-Kamish, Turkmenistan	24
24	1	Okinawa, Japan	28
25	1	Sary-Kamish, Turkmenistan	24
26	1	Turew, Poland	3
	2	Dziekanow-Lesny, near Warsaw, Poland	4
27	1	Dziekanow-Lesny, Poland	4
28	1	Dziekanow-Lesny, Poland	4
	1	Stemplovec, Czech Republic	5
29	1	Sary-Kamish, Turkmenistan	24
30	5	Altai Mountains, Siberia, Russia	26
31	1	Dshankoi, Crimea, Ukraine	21
32[d]	14	Belgrade, Serbia	15
	1	Borča, Serbia	16
	1	Besni Fok, Serbia	17
33	7[e]	Kleylehof, Austria	11
	1	Mönchhof, Austria	9
34	1	Mönchhof, Austria	9
35	1	Batumi, Abkhazia, Georgia	22
36	2	Kishinev, Moldova	19
		M. domesticus mtDNA	
2	1	Trondheim, Norway[f]	2
15	30	13 in Bavarian transect	B
16	2	1 in Bavarian transect	B
46	36	11 in Bavarian transect	B
57	2	1 in Bavarian transect	B
58	2	1 in Bavarian transect	B
59	4	1 in Bavarian transect	B
60	2	1 in Bavarian transect	B
61	1	1 in Bavarian transect	B
62	1	1 in Bavarian transect	B

TABLE 1

Continued

mtDNA type	No. of mice	Locality	Map no.
		M. domesticus mtDNA	
63	1	1 in Bavarian transect	B
64	1[g]	Trondheim, Norway	2
65	1	Trondheim, Norway	2
66	2	Trondheim, Norway	2
67	1	Čakovec, Croatia	14
68	2	Batumi, Abkhazia, Georgia	22
69	1	Batumi, Abkhazia, Georgia	22
		M. castaneus mtDNA	
cas 1	1	Thonburi, Thailand	27
		M. macedonicus mtDNA	
mac 1	1	Gradsko, Macedonia	18
		M. spicilegus mtDNA	
spi 1	1	Kleylehof, Austria	11
	1[h]	6 km N of Mönchhof, Austria	12
	1[i]	6 km N of Halbturn, Austria	13
spi 2	1[j]	Kishinev, Moldova	19
spi 3	1[k]	Dshankoi, Crimea, Ukraine	21
		M. spretus mtDNA	
spr 1	1	Puerto Real, near Cádiz, Spain	1

Figure 1 shows the localities designated as map numbers 1–26 and B. The Bavarian transect (B) includes 45 localities encompassing the *M. domesticus-M. musculus* hybrid zone and adjacent areas, beginning in the west at Augsburg in southern Germany (Bavaria) and extending ~180 km eastward to the vicinity of Höhnhart in north-central Austria; 30 of these localities have been listed before (SAGE *et al.* 1986; TUCKER *et al.* 1992). This table indicates the total numbers of transect mice and localities with a given type of mtDNA sequence. Details about specific localities and the mtDNAs found at each will be presented in a report devoted to characterization of several hundred Bavarian transect mice from a variety of molecular and supramolecular viewpoints (R. D. SAGE, E. M. PRAGER and R. KRAFT, unpublished results). Additional earlier publications (FERRIS *et al.* 1983; SAGE *et al.* 1990; RUVINSKY *et al.* 1991; PRAGER *et al.* 1993) and the authors are further sources of more details about collecting sites.

[a] The sequence was determined for two individuals of a lab strain.

[b] NACHMAN *et al.* (1994) obtained the same sequence for two mice trapped in Prague.

[c] This mouse is provisionally assigned to type 7, as the bases at variable positions 00016, 00052, and 00056 in Figure 3 were not determined; the available sequence rules out all but type 10 among the 35 other *M. musculus* mtDNA sequences recognized here.

[d] Subtypes 32a–d are defined based on the relative amounts of T and C at position 15573 in the 3′ copy of the 75-bp tandem repeat (see Figure 3). The mouse from locality 16 had type 32d and that from locality 17 had 32b. The number of Belgrade mice with types 32a–d, respectively, were three, three, six, and two; the sequences amplified from the two purified mtDNAs were of types 32a and 32c.

[e] Assignment of two of these mice, which have the 75-bp insert, to type 33 is provisional as the 5′ portion of the 1-kb segment was only partially sequenced. The assignment is based on the observation that the other five mice from this locality are all of type 33.

[f] The five mice listed as from locality 2 collectively came from three sites in Trondheim.

[g] Assignment of this mtDNA to type 64 is provisional because missing sequence at position 15363 (Figure 4) makes it indistinguishable from type 7. Type 7 has been reported in only one mouse (PRAGER *et al.* 1993), while type 64 (with G at position 15363) is known from four other mice from two localities in northern Scotland and the Orkney Islands (Table 2).

[h] Nine hundred eighty-three basepairs of sequence were read; the undetermined portion is primarily at the far 3′ end.

[i] Nine hundred fifty basepairs of sequence were read; the undetermined portion is primarily within a highly conserved area 3′ of position 15781.

[j] Eight hundred ten basepairs of sequence were read; the undetermined portion is primarily within the region from positions 15782-16039.

[k] Nine hundred forty-one basepairs of sequence were read; the undetermined portion is primarily at the extreme 5′ end and within a highly conserved central region (*cf.* footnote *i*).

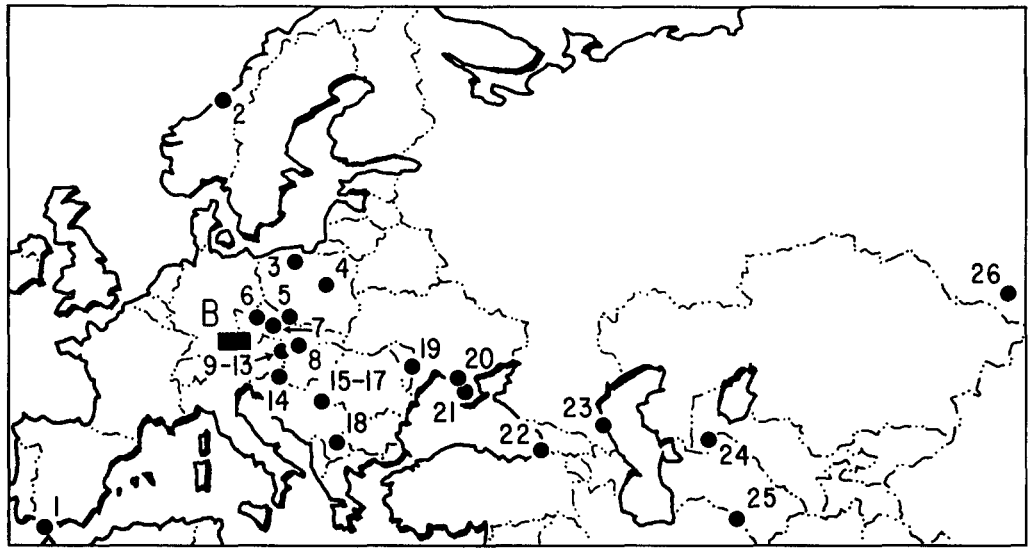

FIGURE 1.—Map showing collecting localities for mice, as numbered in Table 1. B represents the Bavarian transect and encompasses 45 localities in southern Germany and north-central Austria. The far eastern Asian localities 27 and 28 in Table 1 are not shown.

All sequences and variable sites were entered into the computer for analyses of the 36 *M. musculus* plus one *M. castaneus* mtDNA sequences, except the C at position 15550 in the 3′ copy of the 75-bp repeat in *M. musculus* type 35. Variable positions within the tandem repeats in types 32–36 were encoded only once, as at any given position a base substitution occurred in either the 5′ or the 3′ copy but not in both. Type 6 was assumed to have T at positions 00016 and 00056. Type 34 was taken as matching types 1, 2, 21, 32, 33, 35, and 36 in the unsequenced 3′ portion of this segment (Figure 3). For the 36 *M. musculus* mtDNAs, PAUP found 640 most parsimonious (*i.e.*, minimal-length) trees; for the 37 mtDNAs, PAUP found 704 such trees. To root the *M. musculus* + *M. castaneus* mtDNA tree, eight *M. domesticus* mtDNA sequences (types 1, 6, 7, 10, 16, 25, 27, and 46) were used (with a total of 94 variable sites among the 45 sequences); PAUP found 2816 minimal-length trees. The frequencies of occurrence of internal branches among *M. musculus* mtDNAs (cf. Figure 5) were identical for the sets of 37 or 45 sequences. For these interspecific analyses, *M. musculus* mtDNAs were assumed to be invariant at position 15333 (where types 4, 10, 13, 16, 18–21, 27, 28, 30, 34, and 35 were not sequenced) and at position 15363 (where type 35 was undetermined). For the interspecific analyses reported in Table 4 and Figure 8, *M. musculus* 29 was taken as identical to *M. musculus* 1 at variable position 00019, and *M. spicilegus* 3 as identical to *M. spicilegus* 1 at position 15333.

Among the 96 *M. domesticus* mtDNA sequences, 103 sites were variable and 55 of those were phylogenetically informative. Variations only at informative sites were entered into the computer. Sixteen *M. domesticus* sequences (types 4, 16, 21, 31, 32, 34, 38, 39, 48, 57, 60, 63–65, 82, and 84) that were phylogenetically equivalent to others were omitted. Assumptions for undetermined sequences at informative sites were as follows: at position 16056, C in type 35; at position 00052, G in type 71 and A in 50, 70, and 72–96; at position 00055, G in type 70, A in 17 and 71–81, T in 50, 82–90, and 96, and C in 91–95. PAUP done on a Macintosh Quadra 650 did not permit finding and storing all minimal-length trees among 80 *M. domesticus* mtDNA sequences (>6800 trees), and so smaller subsets of 59–65 sequences (and some appreciably smaller) were analyzed. In this way it was possible to examine all most-parsimonious arrangements in various sections of the tree, by focusing in individual analyses on specific sections and eliminating extensive rearrangements elsewhere. To root the *M. domesticus* tree, four *M. musculus* sequences (types 1, 20, 29, and 30) and the *M. castaneus* sequence were added, under which circumstances *M. domesticus* types 21, 31, and 32 became phylogenetically distinct; because it was technically not possible to obtain all shortest trees for the complete dataset of 88 sequences and 80 informative sites, the set of all 1216 minimal-length trees for 67 sequences (with omission of phylogenetically distinct *M. domesticus* types 9, 22, 29, 30, 33, 35–37, 40–43, 54, 62, 66, 67, 71–73, 80, and 83) was used. To assess the stability of root placement on the *M. domesticus* mtDNA tree, the analysis was rerun for 66 sequences, omitting also *M. domesticus* type 96; 2112 minimal-length trees were generated in this case.

A. G. CLARK constructed a neighbor-joining tree for 10 sequences (see Figure 8C) from a matrix of corrected pairwise differences in which transversions were weighted five times as heavily as transitions and length changes. Distance trees were also constructed from this corrected matrix and from matrices of observed differences with the UPGMA (unweighted pair group method with arithmetic mean) technique. For the larger datasets of all *M. musculus* and all *M. domesticus* mtDNAs, we did not calculate distance trees, nor did we weight transversions differentially, in light of the experience of NACHMAN *et al.* (1994), who with a similar set of 42 *M. domesticus* sequences found only minor differences among trees produced with a variety of methods (character-state and distance) and a wide range of transversion weightings.

The *G*-test with one degree of freedom (SOKAL and ROHLF 1981) was used to see if the numbers of mutations accumulated along two lineages of parsimony trees were statistically significantly different. To evaluate the possibility of regional (*in situ*) divergence of several mtDNAs, we computed the average number of differences (with length changes included) in the control region among all possible pairs of individuals from geographically restricted clades in mtDNA parsimony trees, as outlined previously (PRAGER *et al.* 1993). In calculating pairwise differences among *M. musculus* mtDNAs carrying two copies of a 75-bp repeat, we counted the apparently heteroplasmic polymorphism at one position in type 32 as half T and half C.

Two estimates of nucleotide variability, θ and π, were calculated as described in detail by NACHMAN *et al.* (1994). θ (applied only intraspecifically) is a function of number of polymorphic sites, number of mitochondrial types, and sequence

TABLE 2

***M. domesticus* mtDNA types based on control region sequences**

mtDNA type	Locality	Specimen number
2	England	1316-7
6	Greece	1176
19	N Italy	1008-9
44	Mallorca Island, Spain	1272-3
64	Scotland	1345-6, 1356-7
70	Greece	1177
71	Scotland	1334-5
72	Scotland	1328
73	Scotland	1329
74	S Italy	1225, 1248
75	S Italy	1226
76	Greece	1199, 1200
77	Catalunya, Spain	1287
78	C Italy	1067-8
79	C Italy	1081
80	Greece	1150-1
81	N Italy	1368
82	Catalunya, Spain	1311
83	Catalunya, Spain	1286
84	Scotland	1322-3
85	C Italy	1083
86	N, C, S Italy	1036-7, 1128-9, 1214, 1249
87	N, C Italy	1372-3, 1204-5
88	S Italy	1213
89	S Italy	1262
90	S Italy	1263
91	Greece	1173-4
92	Catalunya, Spain	1310
93	Switzerland	1366-7
94	C Italy	1103-4
95	N Italy	1369
96	Greece	1192-3

Of the 42 types of *M. domesticus* mtDNA that NACHMAN *et al.* (1994) distinguished among 56 mice by sequencing the segments containing the ND3 gene plus the control region, the 32 types listed in this table could be resolved based on the sequence variation they reported for the control region and the adjacent Pro tRNA gene. Scotland encompasses northern Scotland and the Orkney Islands; N, C, and S indicate northern, central, and southern, respectively. The actual sequences, tabulated according to specimen number, and details about the collecting localities appear in NACHMAN *et al.* (1994). The first five mtDNA types listed correspond to types identified also by PRAGER *et al.* (1993; this report).

length. π is a function of number of pairwise differences between types, sequence length, and frequencies of each type of mtDNA. Because many of the individual sequences (n = 139 and 399 for *M. musculus* and *M. domesticus* mtDNAs, respectively) resulted, especially for *M. domesticus* mtDNAs, from intense sampling over a limited geographic area, we also computed π with equal frequencies assigned to each type of mtDNA. Such sampling in Scandinavia (which was colonized with mice having *M. musculus* nuclear genomes and *M. domesticus* mtDNAs via founder events during the past 5000 years) plus in the East Holstein source area yielded 110 mice with type 27 *M. domesticus* mtDNA (PRAGER *et al.* 1993; DALLAS *et al.* 1995) and nearly all the other 77 mice in the clade of 18 closely related sequences carrying an added 11-bp direct repeat (see Figure 6). Interspecific π values were computed weighting each individual type within a species equally and, for *M. castaneus vs. M. musculus* mtDNAs, also using observed frequencies.

T. W. QUINN used the Hitachi MacDNASIS Pro RNA Secondary Structure Prediction program, which implements the ZUKER and STIEGLER (1981) method, to find and evaluate potential secondary structures within and adjacent to the 75-bp repeat. The sequences from *M. musculus* types 1 and 32–36 and *M. domesticus* type 1 were examined. The analyses were done separately for four parts of the mitochondrial genome: the first 121 bp of the control region, 5′ of the repeat motif; each repeat of 75 bp (78 bp for the corresponding *M. domesticus* sequence); two tandem repeats together, *i.e.*, 150 bp, for *M. musculus* types 32–36; and the 121 bp 3′ of the repeat motif. The "maximum bulge and interior loop" parameter was set at 2, 15, and 30; the last two settings gave identical results. Unless stated otherwise, free energies are reported as ranges encompassing all sequences examined and all max-bulge settings used.

RESULTS

Mitochondrial DNA sequences: Figure 2 shows the complete sequences of the control region and flanking tRNA genes determined for one type of mtDNA from each of three commensal and three aboriginal house mouse species.

M. musculus mtDNAs were found in 139 mice, 74 of them from 22 localities in the Bavarian region and 65 from 22 localities extending from central Europe to Japan (Table 1). Figure 3 shows the sequence variation among the 36 distinct types of *M. musculus* mtDNA recognized. Of these, 16 types are confined to one or two localities in the Bavarian transect, and 11 others are confined to single localities elsewhere. Types 3 and 7 appear to be most widespread, the former found in 11 mice from four localities and the latter in 19 mice from 11 localities. Type 23 occurs at two rather distant localities separated by the Caspian Sea (23 and 24 in Figure 1). At five of the localities outside the Bavarian transect (4, 5, 9, 21, and 24 in Figure 1), two or three different *M. musculus* mtDNAs were found. Though there is noticeable geographic structuring of mtDNA variability, especially at a microgeographic level (see Figure 5), degree of sequence difference does not consistently correlate with geographic distance. For instance, types 24 from Japan and 21 from Turkmenistan each differ by only two base substitutions from the Czech type 1. In contrast, types 4 and 8 with 15 substitutions coexist in the same Austrian locality, and there are eight to nine substitutions among the three types at locality 24.

At two localities away from the Bavarian hybrid zone and environs, the present survey revealed both *M. musculus* and *M. domesticus* mtDNAs (Table 1). Locality 14 in northeastern Croatia yielded *M. musculus* type 7 and *M. domesticus* type 67, while at locality 22 in Abkhazia *M. musculus* type 35 occurs together with *M. domesticus* types 68 and 69. Finding mtDNAs of both species at locality 14 is not surprising in light of current views about the location of the contact and hybrid zones in south-central

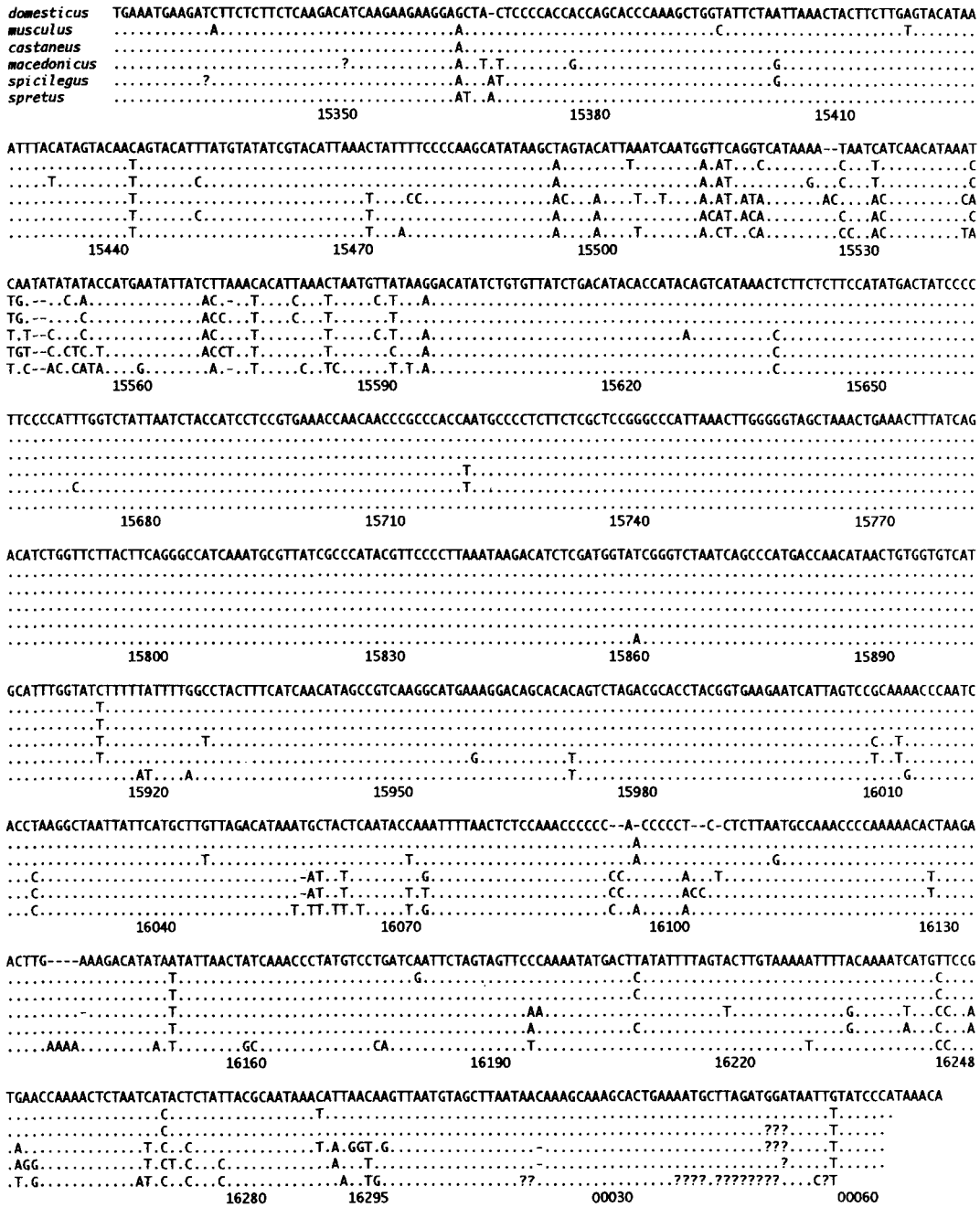

FIGURE 2.—Control region and tRNA gene sequences for mtDNAs of six species of house mice. The 1041-bp sequence of *M. domesticus* type 1 appears in the top row; the sequences of the type 1 mtDNAs of the other species are shown only where different, with identity to *M. domesticus* being indicated by ·. —, deletions relative to other sequences; ?, undetermined residues; blank areas, unsequenced regions. Numbering throughout this report is according to the *M. domesticus* type 1 sequence in BIBB *et al.* (1981), with absence of a bp at positions 15823 and 16240 and another correction as previously (PRAGER *et al.* 1993). The locations shown for length differences occurring in areas of direct repeats are arbitrary (see discussion in PRAGER *et al.* 1993). This segment includes the following parts of the mitochondrial genome: Thr tRNA (partial), 15321–15349; Pro tRNA, 15350–15416; control region, 15417–16295; Phe tRNA, 1–68.

Europe (*e.g.*, VANLERBERGHE *et al.* 1986; SAGE *et al.* 1993). The observation for locality 22 is consistent with growing evidence of the presence of *M. domesticus* as well as *M. musculus* genes in the region between the Black and Caspian Seas that includes the Caucasus and Transcaucasus (*e.g.*, FRISMAN *et al.* 1990; MILISHNIKOV *et al.* 1990; MEZHZHERIN and KOTENKOVA 1992).

Pairwise differences among all 36 *M. musculus* sequences in Figure 3 range from 1–15. The numbers of pairwise transversions and length changes range from 0–4 and 0–3, respectively. Two of the short length changes, at positions 16078ab and 16093a, can be explained by slipped mispairing in areas of direct repeats; addition of a T at 16160a cannot be explained this way, as was true also for one 1-bp length change in a *M. domesticus* mtDNA (PRAGER *et al.* 1993).

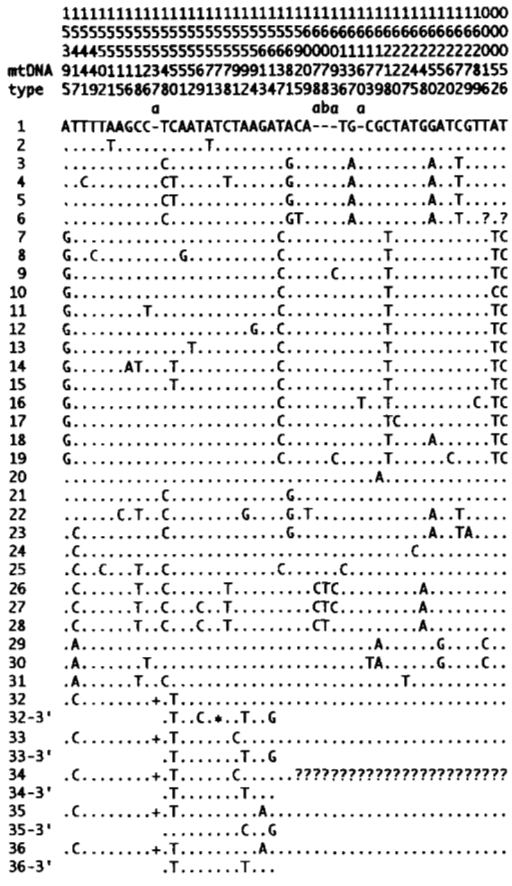

FIGURE 3.—Variable sites in the control region and flanking tRNAs of 36 types of *M. musculus* mtDNA. Listed vertically across the top are 49 polymorphic sites numbered as in Figure 2; lowercase letters denote sites added relative to the baseline (*M. domesticus* type 1) sequence in Figure 2. The sequence for *M. musculus* type 1 is shown at all 49 sites; the bases present in the other types are shown only where different. —, deletions relative to other sequences; + at 15537a indicates presence of a tandem 75-bp repeat of the sequence from 15538–15615; * calls attention to polymorphism at one site in type 32; ?, unsequenced sites. The two locations with length changes involving >1 bp (15537a and 16078ab) have been counted as single sites. The variation in the 3′ copy of the 75-bp repeat is shown on separate lines designated 3′. Subtypes 32a–d are defined based on the approximate relative amounts of T and C at position 15573 in the 3′ tandem repeat as follows: 32a, T only; 32b, 60–70% T; 32c, 50% of each base; 32d, 60–75% C. Footnote *d* to Table 1 gives the numbers and collecting localities of individuals with each of the four subtypes.

In five types of *M. musculus* mtDNA (32–36 in Figure 3), a second, tandem copy of the 75-bp control region segment from positions 15538–15615 was observed. Tandem repeats of about this length and variation in their number have been recognized in the mitochondrial control region in diverse creatures (see, *e.g.*, HOELZEL *et al.* 1994; FUMAGALLI *et al.* 1996). Duplication of these 75 bp seems to be widespread in *M. musculus* mtDNA, having been found in 28 animals from seven localities stretching across some 2000 km from Austria to Abkhazia. Within types, the tandem repeats differ by an average of 2.9 base substitutions, or 3.9%, with a range of two to four substitutions. While the 5′ copy

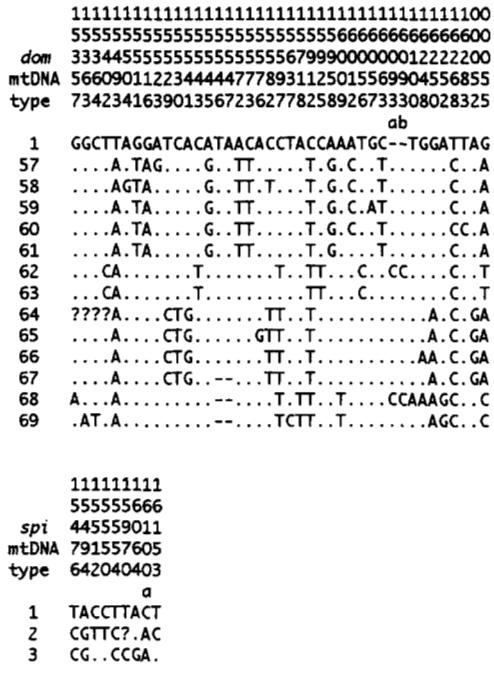

FIGURE 4.—Variable sites in *M. domesticus* and in *M. spicilegus* mtDNA sequences presented in the format described in Figure 3. *M. domesticus* type 64 here was not sequenced 5′ of position 15407 and cannot be distinguished from type 7 but has been equated with type 64 (cf. Table 2) for the reasons given in footnote *g* to Table 1. At 64 sites beyond the 39 at the top of this figure, variation was found among 96 *M. domesticus* mtDNAs; among types 57–69, one of those sites (position 00018) was not sequenced for types 61, 64, and 69. See footnotes *j* and *k* to Table 1 for information about unsequenced portions of *M. spicilegus* types 2 and 3.

differs among types by an average of only 1.2 substitutions (range, 0–2), the 3′ copy differs among types by an average of 2.0 substitutions (range, 0–3.5). Thus, the two copies of the 75-bp segment do not appear to have evolved in concert, though the time since occurrence of the duplication may have been too short to permit homogenization. The 3′ copy seems to have accumulated more substitutions (see also legend to Figure 5). Finally, an instance of polymorphism consistent with heteroplasmy was detected at position 15573 in the 3′ repeat in type 32, with some animals having only T and others having 30–75% C along with T (see legend to Figure 3 and Table 1 for details). Observation of this kind of point-mutational polymorphism in one sequence derived from purified mtDNA makes it unlikely that the apparent heteroplasmy is attributable to the presence of a nuclear copy of this portion of the mitochondrial genome.

The present survey yielded 17 distinct *M. domesticus* mtDNA sequences among 90 mice from 30 localities (Table 1), and Figure 4 shows the variation among the 13 types we sequenced beyond the 56 types described in PRAGER *et al.* (1993). These 17 mtDNAs emanate from five major sections of the *M. domesticus* mtDNA tree (see Figure 6). The four different mtDNAs from western Norway are types typically associated with north-

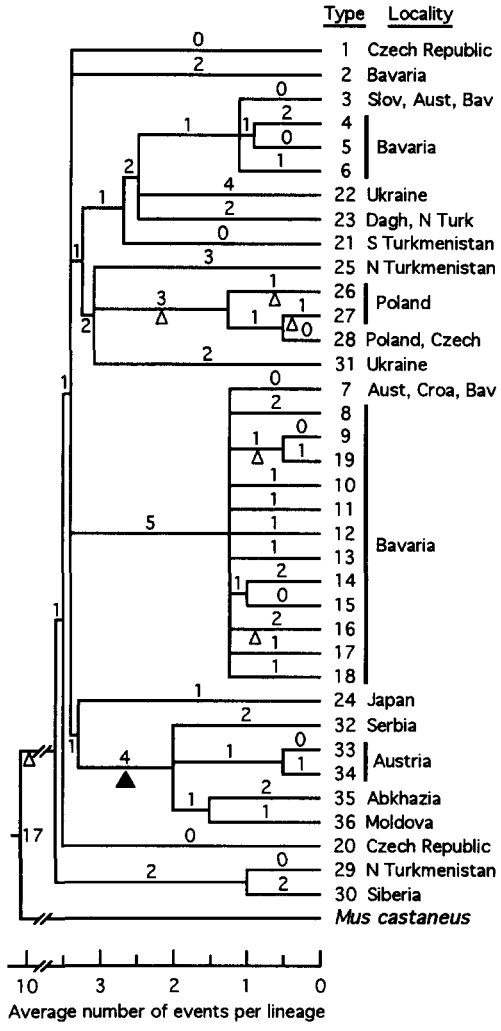

FIGURE 5.—Evolutionary tree constructed by parsimony analysis for 36 *M. musculus* mtDNAs and one *M. castaneus* mtDNA from sequences of the control region and neighboring tRNAs. Abbreviations for localities: Aust, Austria; Bavaria and Bav, Bavarian transect; Croa, Croatia; Czech, Czech Republic; Dagh, Daghestan; N, northern; Slov, Slovakia; S, southern; Turk, Turkmenistan. The number of mutations inferred to have occurred along each lineage is indicated. The large solid triangle marks the lineage where the 75-bp direct repeat of the sequence from 15538–15615 has arisen; small open triangles mark the six lineages where additions or deletions of 1–2 bp are inferred. Length changes were included in computing the depth of each node. The *M. musculus* portion of the tree requires 70 mutations: 54 transitions, 10 transversions, and six length changes at, respectively, 39, nine, and four of the 49 polymorphic sites (consistency index = 0.79). Inclusion of the *M. castaneus* mtDNA yields 59 polymorphic sites and requires 87 events (66 transitions, 14 transversions, and seven length changes; consistency index = 0.77). The root was placed on the lineage indicated in all most-parsimonious trees constructed for these 37 mtDNAs plus eight *M. domesticus* mtDNAs; the 36 *M. musculus* mtDNAs formed a monophyletic clade in 100% of these minimal-length trees. Of the 18 internal branches depicted within the *M. musculus* portion of the tree, 11 occur in 100% of all minimal-length trees; the two deepest intra-*M. musculus* branches occur in 91% of these minimal-length trees. The unions of types 27 + 28, 33 + 34, and 35 + 36 were found 50% of the time, while two other clades—of types 3–6 + 21–23 + 25–28 + 31 and

western Europe. Eight of the *M. domesticus* mtDNAs in Table 1 are known only from the Bavarian transect and two others (types 15 and 16) are known also from other southern German localities (PRAGER *et al.* 1993).

This 1-kb segment was wholly or partially (810–983 bp) sequenced from *M. spicilegus* mice from five localities (Table 1); the lower section of Figure 4 shows the variation among the three kinds of mtDNA identified. The pairwise differences of ≥4 to ≥7 are similar to the intraspecific differences seen among *M. domesticus* and *M. musculus* mtDNAs and the ≥10 differences between the two *M. spretus* mtDNAs considered here.

Evolutionary trees for mtDNAs of commensal mice: Figure 5 presents a rooted parsimony tree relating the 36 *M. musculus* mtDNAs to each other and to a *M. castaneus* mtDNA. The *M. musculus* portion of the tree accounts for the variation at 49 polymorphic sites with 70 mutations. The 5.4 ratio of transitions to transversions matches the value reported for a tree of 56 *M. domesticus* mtDNAs (PRAGER *et al.* 1993). In common with that *M. domesticus* tree and the one in Figure 6 here, the *M. musculus* tree in Figure 5 exhibits noticeable variation in the numbers of events along different lineages emanating from the same node. The average depth of the *M. musculus* tree is ~3.5 events per lineage, *i.e.*, about half the depth of the *M. domesticus* trees. Of interest, a parsimony tree constructed (E. M. PRAGER, unpublished results) for only 19 *M. musculus* mtDNAs, type 1 plus the 18 types detected across the Bavarian transect, is as deep as that shown in Figure 5, which inspires confidence that our 36 *M. musculus* mtDNAs suffice for a realistic assessment of mtDNA tree depth for this species. The depth at which the representative *M. castaneus* mtDNA joins the tree suggests, in turn, that interspecific mtDNA lineage divergence occurred roughly three times as long ago as the start of divergence among the *M. musculus* mtDNA lineages examined here.

As noted earlier for *M. domesticus* mtDNAs (PRAGER *et al.* 1993; NACHMAN *et al.* 1994; refs. therein), the tree in Figure 5 implies some geographic structuring, particularly for the Bavarian transect clades of types 3–6 and 7–19. However, too few localities over the broad range of *M. musculus* have been sampled, particularly outside central Europe, to permit a comprehensive assessment of geographic structuring of mtDNA lineages. Nevertheless, there appears to be a tendency for the deepest lineages in each clade and in the tree as a whole to come from central Asian rather than European localities, which is consistent with the idea of the origin of

of types 24 + 32–36—appeared 73% of the time. This tree postulates four events, in the following order, along the lineage leading to the clade of types 32–36 (cf. Figure 3): a point mutation at position 15550, duplication of the 75-bp segment, and two point mutations in the 3′ copy. During divergence among types 32–36, the tree postulates eight transitions: six in the 3′ copy of the 75-bp repeat and two in the 5′ copy.

the species being in south-central Asia (SCHWARZ and SCHWARZ 1943; MARSHALL 1986; BOURSOT *et al.* 1993, 1996; BONHOMME *et al.* 1994; DIN *et al.* 1996; J. T. MARSHALL, personal communication).

Figure 6 presents a rooted parsimony tree for 96 distinct *M. domesticus* mtDNAs based on control region and adjacent tRNA sequences from 399 mice collected at 125 localities (PRAGER *et al.* 1993; NACHMAN *et al.* 1994; DALLAS *et al.* 1995; this report). These animals represent broad and reasonably intense sampling over the entire western European range of this taxon; included are several localities in northern Africa and the Middle East. The tree is very similar to that for 56 types of *M. domesticus* mtDNA in PRAGER *et al.* (1993) and somewhat different from that in NACHMAN *et al.* (1994) with respect to the relative organization of deeper lineages, likely in part because the latter included sequences for the ND3, Gly tRNA, and Arg tRNA genes but lacked the Phe tRNA segment. The tree in Figure 6 requires 210 mutations; the transition-to-transversion ratio of 6.7 is somewhat higher than previously (see above), perhaps due to the addition of more lineages to shallow clades.

The tree in Figure 6 has a depth of 7.3 events, or 6.5 if only base substitutions are counted; the latter value is the same as for 56 *M. domesticus* mtDNAs (PRAGER *et al.* 1993). That increasing the number of *M. domesticus* mtDNA types by over 70% did not result in a deeper tree provides added confidence for comparing the relative depths of the *M. musculus* and *M. domesticus* trees. The present *M. domesticus* mtDNA tree also strengthens the view (SAGE *et al.* 1990; PRAGER *et al.* 1993) that the southern Mediterranean lineages are older and that the northern lineages originated more recently from southern ones. The tree for 96 mtDNAs further suggests that, at least in the geographic regions already sampled, we are unlikely to find new deep lineages. This tree differs from that in PRAGER *et al.* (1993) in one conspicuous way: it could be rooted by the outgroup method, and in Figure 6 three events separate the clade of types 1–95 from type 96. That the deepest lineage stems from southern Greece agrees well with the model of entry of *M. domesticus* into Europe from the south. In the absence of type 96, the root could not be placed along a specific lineage (see legend to Figure 6).

Figure 7 gives an overview of the phylogenetic analyses of the variation among the mtDNAs of the commensal mice. It shows the accumulation of more changes along the *M. musculus* than the *M. castaneus* lineage, but the difference is not statistically significant by the *G*-test (and it is known that lineages represented by a single taxon or sequence tend to exhibit fewer changes in tree analyses). Similarly, the 23 events along the *M. domesticus* lineage (*i.e.*, 16 + 7.3 in Figure 7) are not significantly more than the 15 (*i.e.*, 5 + the average of 11 + 3.5 and 6) leading to the *M. musculus* plus *M. castaneus* mtDNAs. The tree in Figure 7 implies that the *M. musculus-M. castaneus* mtDNA split is older than the start of intra-*M. domesticus* mtDNA divergence but only about half as old as the divergence of *M. domesticus* mtDNAs from those of these other commensal taxa. If the changes assigned to each lineage are corrected by weighting transversions five times as heavily as transitions, the relative lineage depths are essentially the same (see also the following section).

The values in Table 3 support the idea that *M. domesticus* exhibits greater mitochondrial genetic diversity than *M. musculus*. The ranges of all and of transversional pairwise differences are, respectively, 1.6 times and twice as large among *M. domesticus* as among *M. musculus* mtDNAs. Further, the total numbers of differences for *M. musculus vs. M. castaneus* (18–26) do not overlap with those for *M. domesticus vs.* either of the other species (32–46); the range of transversions extends appreciably higher for the latter two comparisons, and the interspecific π values reinforce the idea of a closer relationship between *M. musculus* and *M. castaneus* mtDNAs. The parameter θ in Table 3 suggests that *M. domesticus* is mitochondrially 1.7 times as variable as *M. musculus*.[2] The parameter π, in turn, suggests that *M. domesticus* is roughly 1.4-fold (if at all; see footnotes *a* and *b* to Table 3) more diverse than *M. musculus*. However, π at the intraspecific level is geared more toward comparing populations where sample sizes are approximately equal than toward assessing the diversity of an entire species when the composite sample results from intense collecting in a few areas (see MATERIALS AND METHODS and Table 1) plus limited sampling over a broad range done in an effort to uncover representatives of as many deeper lineages as possible. Shallow clades, relative to total tree depth, with many members (both individuals and mtDNA types) are more characteristic of the *M. domesticus* mtDNA tree (Figure 6) than of the *M. musculus* tree (Figure 5), and so we could expect π to give a comparative underestimate of the nucleotide diversity of *M. domesticus* mtDNAs. From restriction data, BOURSOT *et al.* (1996) also estimate mtDNA nucleotide diversity to be lower in *M. musculus* (0.2–0.5%) than in *M. domesticus* (0.7–0.9%).

Relationships of commensal and aboriginal mtDNAs: Table 4 gives pairwise total, transversion, and length differences and transition-to-transversion ratios for 10 mtDNAs, one or two from each of three commensal and three aboriginal species. By all criteria the mtDNAs of the commensal mice are appreciably more similar to one another than to those of any of the aboriginal species. *M. macedonicus* mtDNA is more like *M. spicilegus*

[2] This ratio of 1.7 must be interpreted with caution as the θ values may not have plateaued, especially for *M. musculus*. θ computed for *M. domesticus* mtDNA types 1–56 is 1.73%, which is proportionally only slightly less than the 1.99% based on the 1.7-fold larger dataset of 96 types. If for *M. musculus* only mtDNA types 1–19 are considered, θ drops to 0.80%, *i.e.*, about two-thirds the value of 1.18% in Table 3. In contrast, if attention is confined to the 20 *M. musculus* mtDNA types exclusive of those found only in the Bavarian transect, θ is 1.10%, *i.e.*, essentially the same as for the 1.8-fold larger dataset of 36 types.

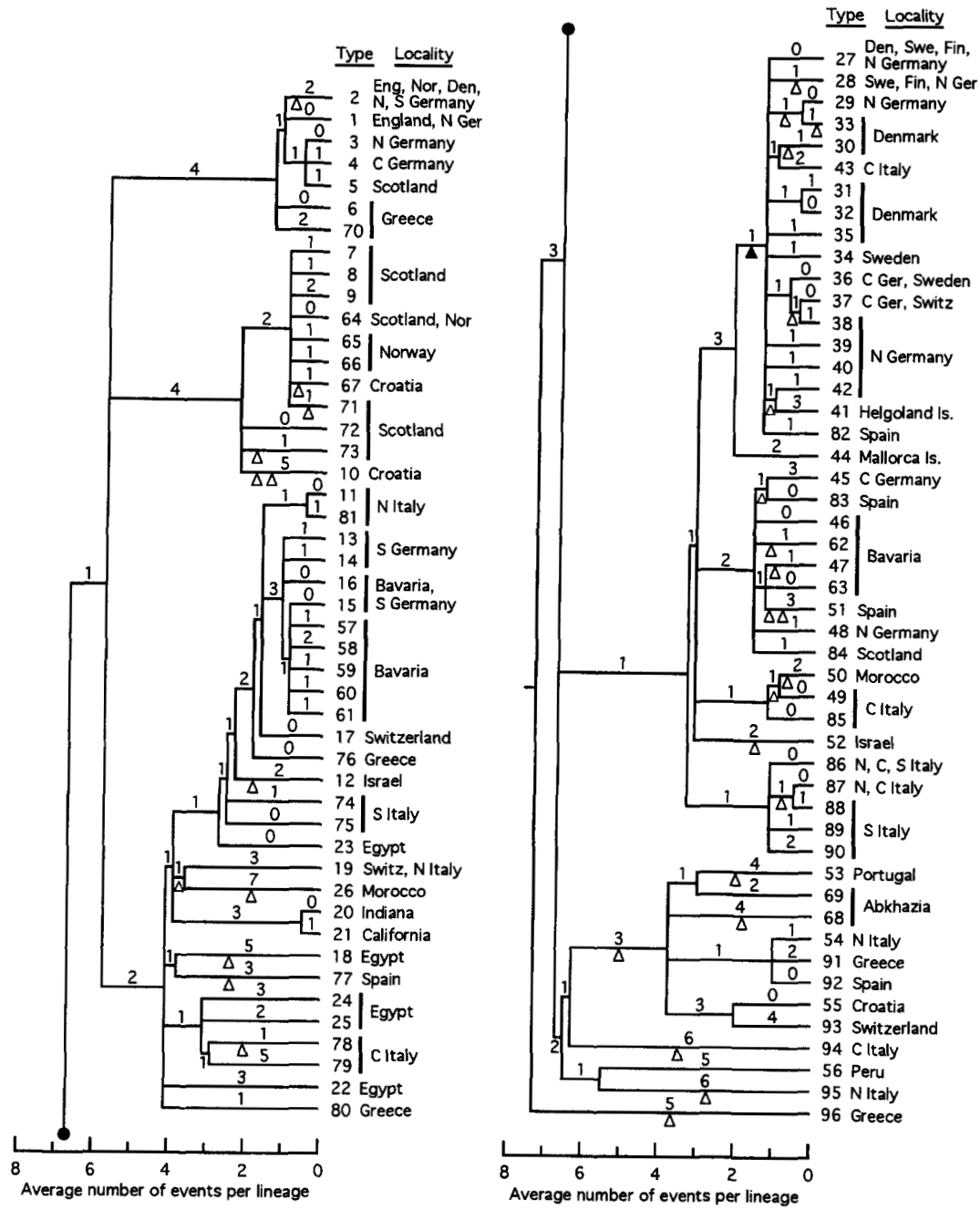

FIGURE 6.—Parsimony tree for 96 *M. domesticus* mtDNAs. Solid circles indicate the connection of the halves of the tree shown at the left and right. Abbreviations for localities: Bavaria, Bavarian transect; C, central; Den, Denmark; Eng, England; Fin, Finland; Ger, Germany; N, northern; Nor, Norway; S, southern; Swe, Sweden; Switz, Switzerland; Scotland includes also localities in the Orkney and Shetland Islands. The number of mutations inferred to have occurred along each lineage is indicated. The solid triangle indicates the lineage where the 11-bp direct repeat of the sequence from 16073–16083 has arisen; the open triangles mark the 31 lineages where additions or deletions of 1–5 bp are inferred. Length changes were included in computing the depth of each node. This tree requires 210 mutations: 153 transitions, 23 transversions, and 34 length changes at, respectively, 83, 19, and 7 of the 103 polymorphic sites (consistency index = 0.53). The root was placed on the lineage indicated from an analysis that included four *M. musculus* and one *M. castaneus* mtDNA sequences; in all most-parsimonious trees in this analysis, *M. domesticus* types 1–95 formed a clade to the exclusion of type 96. The sequences of mtDNAs from aboriginal mice were also considered in assigning three and five events to the lineages leading, respectively, to types 1–95 and type 96. By examining various subsets of the sequences (see MATERIALS AND METHODS), the frequencies of occurrence among all minimal-length trees of each of the 46 internal branches shown were evaluated as follows: 100% for 36 of the internal branches, 71–82% for three (the unions of types 45 + 83 and of types 53 + 69 as well as the branch preceding the trifurcation leading to 20 + 21, 19 + 26, and the clade of 17 sequences extending from types 11 to 23 in the figure), 50% for one (joining types 78 + 79), and 21–30% for three [the branch joining types 24 and 25 to the clade of types 78 + 79, that uniting types 54, 91, and 92, and (for the reason discussed in PRAGER *et al.* 1993) the common lineage uniting types 1–26, 57–61, 64–67, and 70–81]. Three unevalu-

mtDNAs, but the observed differences exceed those among the commensal mtDNAs. By all four criteria in Table 4, the *M. spretus* mtDNAs seem to be as different from those of *M. macedonicus* and *M. spicilegus* as from the mtDNAs of commensal mice. Most striking are the contrasting numbers of transversions and transition-to-transversion ratios for the comparisons of the commensals to *M. macedonicus* and *M. spicilegus vs.* the comparisons of *M. spretus* to the other five species. In the former, there are only 18–26 transversions (average, 22), and the ratio is 1.5–2.0 (average, 1.8). In the latter, there are 33–36 transversions, and the ratio hovers around 1.0 (range, 0.9–1.2). Because a greater proportion of transversions implies more distantly related mtDNA sequences (see PRAGER *et al.* 1993 and refs. therein), the implication is that *M. spretus* mtDNAs and by inference the species *M. spretus* lie outside the other five house mouse lineages.

The tree analyses in Figure 8 support the suggestion that *M. spretus* is the sister taxon to all the other house mouse species, rather than being in a clade with the other two aboriginal taxa, particularly when transversions are weighted more heavily (Figure 8, B and C). A more definitive assessment from mtDNA sequences awaits outgroup rooting (with, for example, *M. cervicolor*) using more slowly evolving parts of the molecule to minimize alignment and multiple-hit problems. The common *M. macedonicus-M. spicilegus* node is ~1.5 times as deep as the node common to all the commensal mtDNAs. The bootstrapping results in Figure 8A provide great confidence in the monophyly of the commensal clade ($P = 100\%$). The clustering of *M. macedonicus* and *M. spicilegus* mtDNAs almost reaches statistical significance ($P = 92\%$), while the association of *M. castaneus* with *M. musculus* appears less certain ($P = 76\%$).

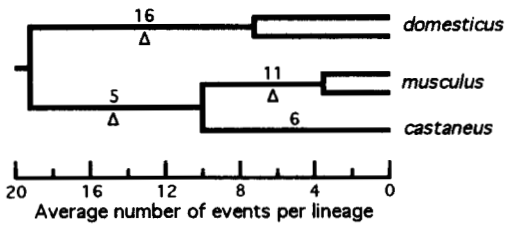

FIGURE 7.—Parsimony tree for mtDNAs of commensal house mice. This tree schematically summarizes the information in Figures 5 and 6, and for four lineages it shows numbers of events and small length changes in the same format used in those figures. Assignment of mutational events to these four lineages and root placement were done by considering the mtDNA sequences of aboriginal house mice as well as all the sequences from the commensal taxa. The *domesticus* portion of the tree shown here represents the deepest intra-*M. domesticus* mtDNA node, plotted at an average of 7.3 events per lineage as in Figure 6. Similarly, the *musculus* portion of the tree here represents the deepest intra-*M. musculus* mtDNA node at an average of ~3.5 events per lineage as in Figure 5. The 11 and six events in this figure show how the 17 events at the far left of Figure 5 were apportioned to two lineages.

DISCUSSION

Diversity and phylogenetic affiliation of *M. musculus* mtDNA: The implication from the trees in Figures 5–8 is that the mtDNAs of *M. musculus* have on average diverged from one another about half as much as have those of *M. domesticus*. Further, if we assume that the *M. musculus-M. domesticus* species split represents the deepest node among commensal lineages and that this split occurred 350,000 (SHE *et al.* 1990) or 500,000 (BOURSOT *et al.* 1993) years ago, these trees imply that the coalescence time of all the *M. musculus* mtDNA lineages we examined could have been as recent as 60,000–80,000 years ago. Such short times accord well with the genetic distance (D) values, based on protein electrophoresis, of only 0.003–0.006 among *M. musculus* mice from Moldova to Kazakhstan (FRISMAN *et al.* 1990). The coalescence time of the *M. domesticus* mtDNA lineages, in contrast, would be roughly 120,000–180,000 years ago, consistent with appreciably larger D values among European *M. domesticus* populations (cf. SAGE *et al.* 1990, 1993).

As discussed in RESULTS, the tree in Figure 6 may well encompass the deepest nodes in the tree for extant *M. domesticus* mtDNAs, given the number of mtDNA types, mice, and localities on which it is based. Moreover, low- and high-resolution restriction analysis has been done on many hundred more individuals from some of these and other localities (*e.g.*, FERRIS *et al.* 1983; VANLERBERGHE *et al.* 1988; RITTE *et al.* 1992; RYAN *et al.* 1993; BOURSOT *et al.* 1996; refs. therein), and no claim of an appreciably deeper lineage has emerged. One caveat is that relatively few animals have been studied from the apparent Indian cradle, where BOURSOT *et al.* (1993, 1996) found quite diverse commensal mtDNA lineages within populations but no *M. domesticus* mtDNAs. However, three novel types of *M. domesticus* mtDNA identified among mice from Iran on the basis of partial control region and flanking sequences (obtained from museum skins; E. M. PRAGER, C. ORREGO, J. T. MARSHALL and R. D. SAGE, unpublished results) are not conspicuously more different from other known *M. domesticus* sequences.

YONEKAWA *et al.* (1994) have presented a UPGMA tree with 32 unpublished *M. musculus* and 30 unpublished *M. castaneus* control region sequences. Their tree depicts the *M. musculus* clade as 17% deeper than the *M. domesticus* clade (represented by sequences corre-

ated branches (joining types 29 + 33 and 30 + 43 and adding type 36 to 37 + 38) were used as before (PRAGER *et al.* 1993). When *M. domesticus* type 96 was omitted from the PAUP analysis used to root this tree, the root did not fall along a single lineage in all minimal-length trees; the strict consensus tree placed it preceding a 10-way multifurcation, *i.e.*, in conceptually the same place as the second-deepest intra-*M. domesticus* node but with the sequences in the clade from type 11 to type 80 emanating from six deep lineages rather than one.

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	0–4	0–3	1.18	0.42[a]
Intra-*domesticus*	1–24	0–8	0–5	1.99	0.57[b]
musculus vs. castaneus	18–26	4–7	1–3	NA[c]	2.26[d]
musculus vs. domesticus	35–46	7–15	3–6	NA	4.02
castaneus vs. domesticus	32–34	5–10	2–3	NA	3.27

All 36 sequences in Figures 3 and 5 were included in the three comparisons involving *M. musculus* mtDNAs and all 96 in Figure 6 for the intra-*M. domesticus* values. For the last two comparisons tabulated, the *M. domesticus* mtDNAs used were types 1, 6, 7, 10, 16, 25, 27, and 46. Length changes involving two or more adjacent bp were counted as single differences. The π values tabulated resulted from assigning equal frequencies to each type of mtDNA within a species (see MATERIALS AND METHODS).

[a] π = 0.41% when observed frequencies of the 36 *M. musculus* mtDNAs are used.

[b] π = 0.44% when observed frequencies of the 96 *M. domesticus* mtDNAs are used.

[c] NA, not applicable.

[d] π = 2.26% also when observed frequencies of the 36 *M. musculus* mtDNAs are used.

sponding to types 1, 20, and probably 11 and 56 in Figure 6) and therefore implies that the tree in Figure 5 may not encompass all major *M. musculus* mtDNA lineages. Possibly their more comprehensive Asian sampling recovered more divergent lineages. Of interest, only four of the 32 *M. musculus* mtDNA types in the YONEKAWA *et al.* (1994) tree have lineages deeper than those of two of their four *M. musculus* mtDNAs from European localities. Another possibility is that the UPGMA computation, which obscures both variation in rates along individual lineages and homoplasious changes, has not produced the likeliest tree with respect to shared derived traits. Indeed, we estimate that a UPGMA tree for our 36 *M. musculus* sequences would be ~85% as deep as a UPGMA tree for the four *M. domesticus* representatives included by YONEKAWA *et al.* (1994), in contrast to the relative depths of ~54% one would infer for the same mtDNAs from the parsimony trees in Figures 5 and 6. [Conversely, the restriction-based UPGMA tree shown by YONEKAWA *et al.* (1994) has a *M. musculus* mtDNA clade only about a quarter as deep as that of *M. domesticus*.]

The possibility of rate differences in the evolution of control region and flanking sequences in *M. domesticus* and *M. musculus* mtDNAs should also be considered. While not statistically significant in Figure 7, such differences, with *M. domesticus* being faster, may in fact exist. On the basis of restriction analyses of the entire mtDNA molecule, BOURSOT *et al.* (1996) have suggested that mtDNA evolution has been significantly slower along the *M. musculus* lineage than along other commensal lineages.

Our mtDNA results encourage further testing of the idea originally based on other evidence (FIGUEROA *et al.* 1987; KLEIN *et al.* 1987, 1988; RUVINSKY *et al.* 1991) that *M. musculus* is genetically more uniform and

TABLE 4

Pairwise differences among the mtDNA sequences of commensal and aboriginal house mice

Sequences compared	*dom* 1	*dom* 27	*mus* 1	*mus* 29	*cas* 1	*mac* 1	*spi* 1	*spi* 3	*spr* 1	*spr* 2
M. domesticus 1	—	11/1	37/3	37/3	33/2	76/7	75/6	76/6	81/6	80/7
M. domesticus 27	3/2.3	—	38/4	38/4	34/3	74/8	73/7	74/7	84/7	83/8
M. musculus 1	10/2.4	9/2.8	—	4/0	19/1	67/8	61/7	62/7	75/3	74/4
M. musculus 29	9/2.8	8/3.3	1/3.0	—	19/1	67/8	59/7	60/7	73/3	72/4
M. castaneus 1	7/3.4	6/4.2	5/2.6	4/3.5	—	69/7	56/6	57/6	75/4	74/5
M. macedonicus 1	26/1.7	25/1.6	24/1.5	23/1.6	21/2.0	—	47/5	45/5	82/10	79/11
M. spicilegus 1	23/2.0	22/2.0	20/1.7	19/1.7	18/1.8	13/2.2	—	6/0	73/7	70/8
M. spicilegus 3	23/2.0	22/2.0	20/1.8	19/1.8	18/1.8	13/2.1	1/5.0	—	77/7	74/8
M. spretus 1	36/1.1	35/1.2	36/1.0	35/1.0	35/1.0	34/1.1	34/0.9	34/1.1	—	10/1
M. spretus 2	35/1.1	34/1.2	35/1.0	34/1.0	34/1.0	33/1.1	33/0.9	33/1.0	1/8.0	—

For each pair of sequences the values above the matrix diagonal give the total number of differences (to the left of the slash) and the number of length differences (to the right of the slash). The values below the diagonal are the number of transversion differences (left of the slash) and the ratio of transitions to transversions (right of the slash). A total of 142 sites were found to be variable among the 10 sequences. Length changes were counted as in Table 3. The unavailable 25% of the *M. spretus* 2 sequence was taken as matching *M. spretus* 1 (see MATERIALS AND METHODS).

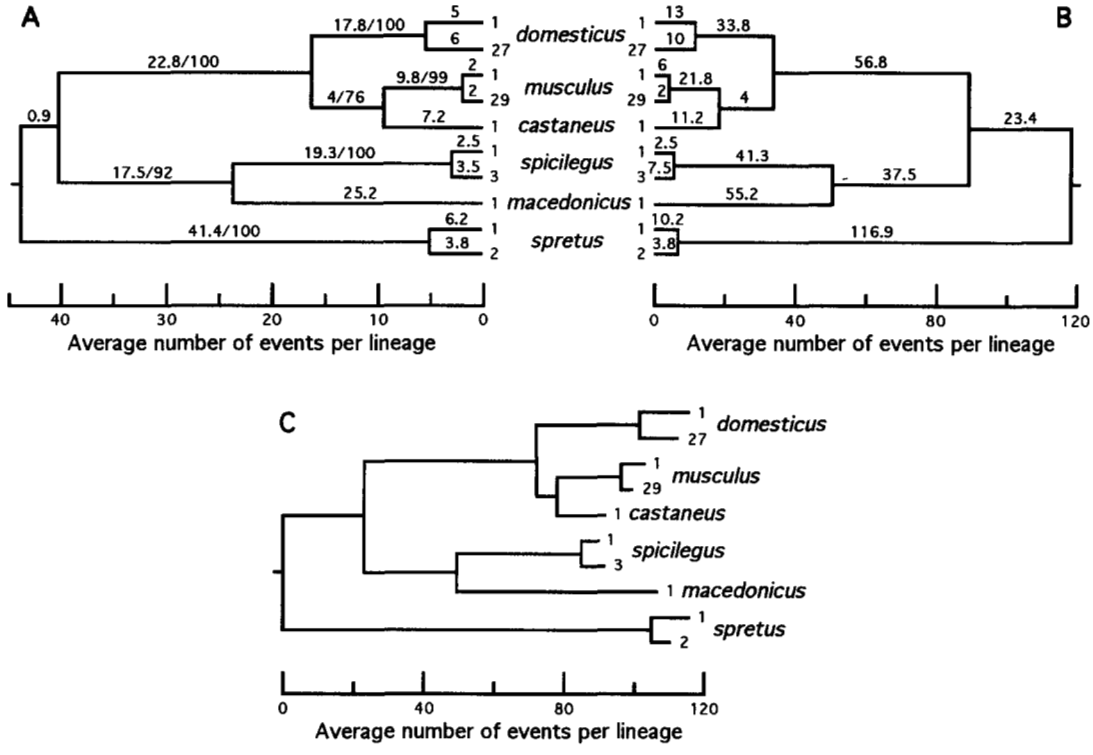

FIGURE 8.—Evolutionary trees constructed by parsimony (A and B) and neighbor-joining (C) analyses of 10 mtDNAs from six species of house mice. For A, analyses were done weighting all changes equally. The topology shown corresponds to the single most parsimonious tree, which requires 197 events (114 transitions, 65 transversions, and 18 length changes of 1–11 bp; consistency index = 0.85). Events were assigned to lineages by hand (which represents a refinement of the output from the ACCTRAN option in the PAUP program) and included averaging equally likely alternative pathways at 20 sites. The root was placed by the midpoint method. Tree A shows the number of events assigned to each lineage and, to the right of the slash, the frequency of occurrence (in percent) of seven internal branches in 5000 bootstrap replications. In B, the tree generated in A has been redrawn and rerooted after weighting transversions five times as heavily as transitions and length mutations; the number of events (of a total of 457) assigned to each branch is indicated. Tree C, constructed from pairwise differences after weighting transversions fivefold, has branch lengths largely in agreement with the numbers assigned to lineages in tree B; the greatest discrepancy is the shorter branch joining *M. spicilegus* and *M. macedonicus* in C. Besides having the same topology shown here, UPGMA trees constructed from the total differences in Table 4 and after weighting transversions fivefold had relative node depths very similar to those in trees A and B, respectively.

younger than *M. domesticus*. It is possible that *M. musculus* and *M. domesticus* are equally old, but that one or more transient or prolonged bottlenecks subsequent to the origin of *M. musculus* reduced the extant genetic variability of this species. There seems to be no *a priori* reason to invoke such bottlenecking. Further, the geographic pattern of variability of *M. musculus* mtDNA, with similar types found at distant localities and highly dissimilar types at the same locality, seems incompatible with bottleneck(s) involving the whole species. KLEIN *et al.* (1987, 1988) have suggested different dispersal patterns as possibly responsible for the apparently greater genetic uniformity of *M. musculus* compared to *M. domesticus*, but the patterns they describe appear unlikely to favor the extinction of deep *M. musculus* mtDNA lineages.

The closer relationship in Figures 7 and 8 of the *M. musculus* and *M. castaneus* control region sequences to the exclusion of those from *M. domesticus* is in agreement with the YONEKAWA *et al.* (1994) tree (which, however, places the start of intracommensal mtDNA divergence less deep relative to the *M. musculus-M. castaneus* split than do our trees). Longer sequences and more *M. castaneus* representatives are needed to support the *M. musculus-M. castaneus* relationship with statistical confidence. Assuming that south-central Asia is indeed the geographic origin of the commensal clade (BOURSOT *et al.* 1993, 1996; DIN *et al.* 1996), the implication is that the populations ancestral to present-day *M. domesticus* moved westward and became separated sooner than the Asian north-south separation between the ancestors of *M. musculus* and *M. castaneus* occurred.

Phylogenetic affiliation of aboriginal mtDNAs: The results in Figure 8 and Table 4 agree entirely with the consensus view that *M. spicilegus* and *M. macedonicus* are sister taxa.[3] However, they appear from our sequences to be mitochondrially as distantly or more distantly related than are the most distant commensal sequences to one another (see also Table 3). SHE *et al.* (1990) reported the mtDNAs of these two species to be simi-

[3] *M. spicilegus* and *M. macedonicus* have also been called, respectively, *M. hortulanus* and *M. abbotti* or *M. spretoides;* these designations are now agreed to be incorrect for the reasons reviewed by BOURSOT *et al.* (1993).

larly distantly grouped on the basis of restriction analysis. FORT *et al.* (1985) found the *M. spicilegus* and *M. macedonicus* sequences for 110 bp of the mitochondrial 16S rRNA gene to be so different that they did not cluster in a phylogenetic tree, a finding consistent with the preliminary restriction studies of FERRIS *et al.* (1983). These observations for mtDNA are somewhat surprising in light of the small amount of divergence measured between *M. spicilegus* and *M. macedonicus* by protein electrophoresis (cf. BONHOMME *et al.* 1984; SHE *et al.* 1990; SAGE *et al.* 1993), hybridization of single-copy nuclear DNA (SHE *et al.* 1990), and assessment of morphological differences (GERASIMOV *et al.* 1990), and they suggest that the mtDNA lineages extant in these taxa could have diverged well before the species divergence (see SHE *et al.* 1990). However, the mtDNA findings are consistent with *Y*-chromosome sequences that did not invariably group these two taxa (TUCKER *et al.* 1989) as well as with the radical difference in their behavioral ecology (reviewed in SAGE *et al.* 1993).

The previously available molecular evidence appears weighted in favor of the view that the lineage leading to *M. spretus* diverged before the lineages leading to the other house mouse species began diverging from one another. The mtDNA sequence evidence presented here clearly supports this scenario. In particular, ~50%, as opposed to 60–67%, transitions in pairwise comparisons (Table 4) can indicate appreciably greater sequence divergence and hence an appreciably earlier divergence (cf. refs. in PRAGER *et al.* 1993). Quantitative immunological comparisons of lysozymes *c* (E. M. PRAGER and M. F. HAMMER, unpublished results) provide another piece of evidence favoring an earlier divergence of *M. spretus*. The proportional depth here for *M. spretus vs.* the intracommensal mtDNA divergence is 2.3–3.5 (Figures 7 and 8), in agreement with estimates of 2.4–3.2 derived from mtDNA restriction analyses (FERRIS *et al.* 1983; SHE *et al.* 1990), the YONEKAWA *et al.* (1994) mtDNA sequencing study, single-copy nuclear DNA hybridization (SHE *et al.* 1990), and 5S rDNA sequences (SUZUKI *et al.* 1994). With the calibration described in the previous section, these ratios fit frequently cited estimates of 1–3 million years ago for the *M. spretus*-commensal divergence.

Time of house mouse colonization of Europe: The geographic progression for the colonization of Europe and Asia by house mice, notably the lineages leading to the present-day commensal taxa, seems to be generally agreed upon (*e.g.*, KLEIN *et al.* 1987; AUFFRAY *et al.* 1990; SAGE *et al.* 1990, 1993; AUFFRAY and BRITTON-DAVIDIAN 1992; BOURSOT *et al.* 1993, 1996; BONHOMME *et al.* 1994; DIN *et al.* 1996). This scenario envisions for the ancestors of *M. domesticus* a westward expansion from a south-central Asian origin first to the Middle East, then to the eastern Mediterranean, subsequently to western Mediterranean Europe, and lastly through Europe north of the Alps. Similarly, the ancestors of *M. musculus* are thought to have expanded northward beyond the Himalayas and then colonized vast areas to the west and east. The European hybrid zone is thus an area of secondary contact. The proposed time scale (which incorporates rapid DNA evolution in murids relative to diverse other vertebrates) and the trees in Figures 7 and 8 make it clear that the initial genetic and geographic divergence occurred in the absence of agriculture and in the absence of modern humans, though it may have occurred in conjunction with *Homo erectus* populations. The times derived for the overall depths of the *M. domesticus* and *M. musculus* mtDNA trees, in turn, imply minimum time spans for the existence of these species. If old mtDNA lineages have been lost or not detected in our surveys, the actual mitochondrial progenitors could have lived appreciably before the dates inferred from Figures 5 and 6. Finally, though the mtDNA and other evidence leave no doubt that the present-day taxa referred to as commensal house mice form a monophyletic clade, the time scale along with archaeozoological and ecological considerations supports the idea of the repeated origin of commensalism (SAGE 1981; BOURSOT *et al.* 1993; SAGE *et al.* 1993; refs. therein).

AUFFRAY and BRITTON-DAVIDIAN (1992) summarize the evidence favoring a European arrival of *M. domesticus* only within about the last 4000–5000 years, though there is fossil evidence placing this taxon in the Middle East 12,000 years ago, and address the arguments advanced by SAGE *et al.* (1990) to urge consideration of a 30,000-year or older time scale. A perhaps surprising corollary of the recent colonization model is that "house mouseless" populations of modern humans, albeit hunter-gatherers with possibly semipermanent settlements, inhabited much of Europe for some 30,000 years or more (*cf.* SHERRY *et al.* 1994). Two quandaries appear mainly responsible for these conflicting hypotheses: (1) How good is the fossil record likely to be for a small mammal, especially if populations were small and perhaps patchily distributed, and moreover, how likely is it that colonizing mouse populations remained small for tens of millennia? (2) Do murids evolve unusually fast at the molecular level?

Computing the average divergence of control region sequences from geographically restricted mtDNA clades, which may have diversified *in situ*, offers one approach to inferring length of habitation. The *M. domesticus* tree in Figure 6 provides two such clades: the 48 mice in southern Germany with types 13–16 and 57–61 and the 13 Italian mice with types 86–90. Using a pairwise rate of control region evolution of ~20% per million years (cf. discussion in PRAGER *et al.* 1993), we derive estimated ages of close to 5000 years for both clades. In Figure 5, the Bavarian transect clades of 13 mice with types 3–6 and 52 mice with types 7–19 afford the opportunity to do a similar calculation for *M. musculus* mtDNAs (assuming that each of these possibly endemic clades was derived from the widespread types 3 and 7). The somewhat older estimated times of 6700–10,000 years are consistent with a somewhat earlier Eu-

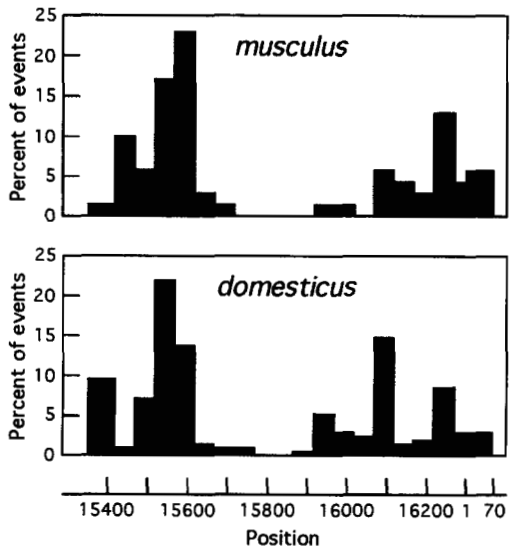

FIGURE 9.—Histograms showing variation as a function of position for the control region and adjacent Pro and Phe tRNA genes for 36 types of *M. musculus* mtDNA and 96 types of *M. domesticus* mtDNA. Horizontally each bar (except those at the extremes) represents a 50-bp segment. Vertically each bar shows in percent the proportion of the 70 events in the tree in Figure 5 and the 210 events in the tree in Figure 6 assigned to positions within each segment.

ropean arrival of *M. musculus* via another route (AUFFRAY *et al.* 1990). Overall, the results of this analysis seem more consistent with the rapid, recent colonization model defended by AUFFRAY and BRITTON-DAVIDIAN (1992).

mtDNA variation as a function of position: KOCHER and WILSON (1991) showed a difference in the patterns of variation for human and chimpanzee mitochondrial control regions, which have observed differences ~13%. The present analysis for the analogous segment of house mouse mtDNAs revealed distinctive patterns in the face of sequence differences below 5%, as shown (Figure 9) by tree-based assessments of the patterns of variation of *M. musculus* and *M. domesticus* mtDNAs in the control region and flanking tRNA genes. While the overall picture is roughly the same in the two species, there are noticeable differences. Perhaps most conspicuous is that the first 50-bp segment at the 5′ end of the control region is 10 times as variable in *M. musculus* as in *M. domesticus*. Conversely, the Pro tRNA gene is about seven times as variable in *M. domesticus*. Two other differences in the profiles in Figure 9 can be explained by a different propensity for length changes at two locations previously recognized as hotspots in *M. domesticus* mtDNAs (PRAGER *et al.* 1993): in *M. domesticus* 2-bp changes are inferred to occur 11 times at 15546–15553ab, while no small length changes occur in the relevant 50-bp segment among *M. musculus* mtDNAs. Positions 16093ab are the most variable ones among *M. domesticus* mtDNAs, with 19 changes of 210 in the tree (9%); in contrast, among *M. musculus* mtDNAs only three of 70 changes (4%) are assigned to this location. Of interest, insertion of a third copy of a 2-bp direct repeat in *M. musculus* types 26–28 (Figure 3) and addition of an 11-bp direct repeat in one *M. domesticus* mtDNA clade (cf. Figure 6) occur in the same spot, in an area where arrays of short repeat motifs are known for many mammalian mtDNAs (HOELZEL *et al.* 1994; FUMAGALLI *et al.* 1996).

Species differences in the pattern of variation are apparent also on the finer scale of individual positions. Of the 12 variable sites with more than one mutation in the tree for 36 *M. musculus* mtDNAs, four (positions 15417, 15578, 15592, and 15614) are invariant among 96 *M. domesticus* mtDNAs. Conversely, of 20 highly variable positions in the *M. domesticus* tree, with three to 19 mutations each, 12 (positions 15363, 15504, 15513, 15530, 15543, 15582, 15588, 15597, 16057, 16100, 16188, and 00055) are invariant among 36 *M. musculus* mtDNAs.

Tandem repeat of 75 bp: The duplication observed in a clade of five types of *M. musculus* mtDNA of the segment numbered 15538–15615 in Figure 2 encompasses much of the most highly variable 5′ section of the control region (Figure 9). The *M. musculus* repeat falls into RS2, one of five locations, designated RS1–RS5, in the control region where tandem repeat motifs have been reported in a wide variety of vertebrate mitochondrial genomes (HOELZEL 1993; HOELZEL *et al.* 1994). [FUMAGALLI *et al.* (1996) use R1 to encompass RS1 and RS2, which are in the 5′ variable domain, and R2 to encompass RS3–RS5, which are in the 3′ variable domain.] In RS2, the evening bat has an array of five to eight copies of an 81-bp motif (WILKINSON and CHAPMAN 1991), the cat has three copies of an 80-bp motif (HOELZEL *et al.* 1994), and shrews have one to eight copies of 78–80-bp repeats plus an imperfect 3′ repeat of 76–79 bp (STEWART and BAKER 1994a,b; FUMAGALLI *et al.* 1996; due to intraindividual heteroplasmy in number of repeats, no shrew has a total of only two repeats in all of its mtDNA molecules). Repeats in RS2 and elsewhere can often form stable secondary structures, and termination-associated sequences (TASs) are often incorporated in repeats at RS1 and RS2.

In common with repeats at RS2 and elsewhere in other creatures, the *M. musculus* tandem array exhibits a variable number of repeats among mtDNAs, point-mutational differences between tandem repeats in the same individual, and point-mutational differences among the same repeats in different individuals. However, we did not observe among the *M. musculus* mtDNAs examined here the common feature of heteroplasmy with respect to the number of repeats within an individual. The competitive displacement model of replication slippage proposed by BUROKER *et al.* (1990) makes length heteroplasmy much more likely if three or more tandem repeats have been generated, and so its absence among the *M. musculus* mtDNAs surveyed may be due to the occurrence of no more than two copies of the repeat motif in any

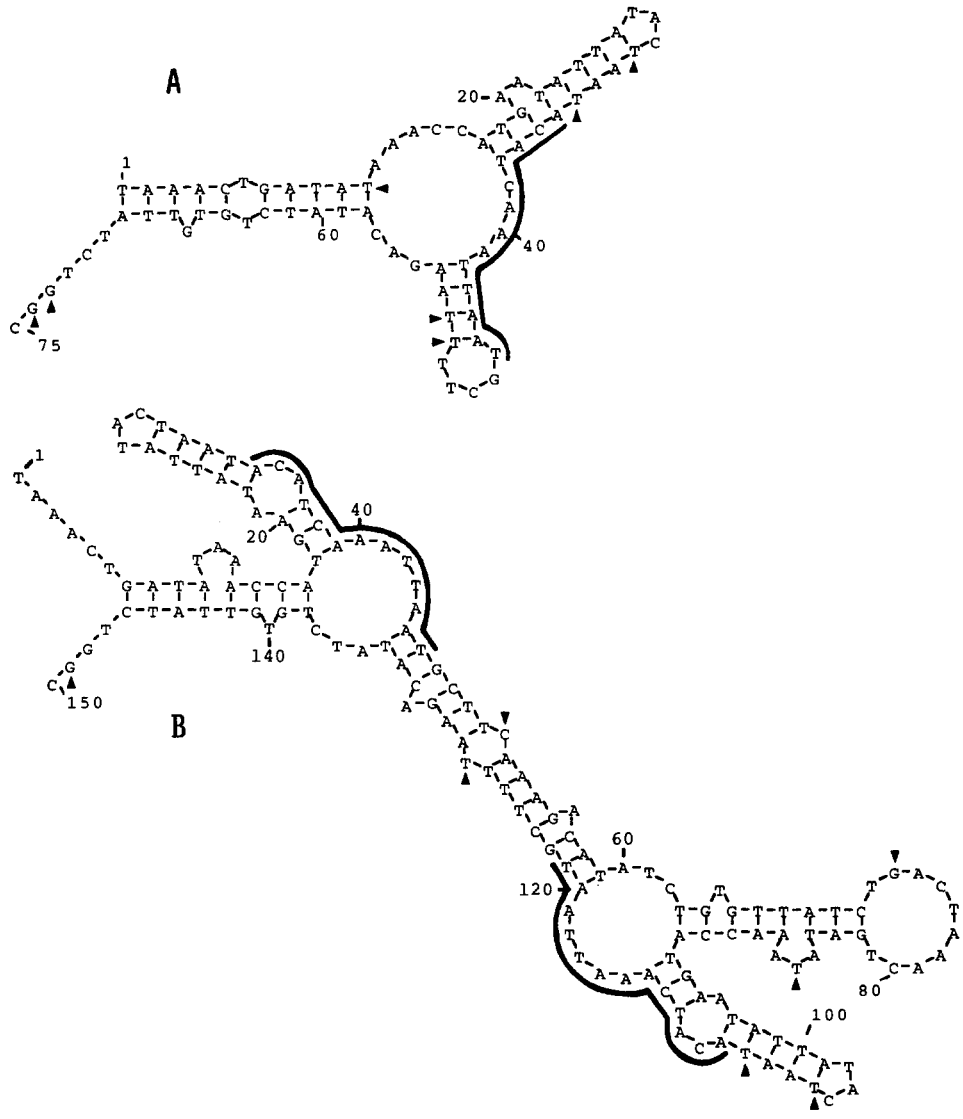

FIGURE 10.—Secondary structures for the 75-bp repeat motif in *M. musculus* mtDNAs. (A) The sequence for the 3′ copy of type 33 (free energy = −8.9 kcal/mol). (B) Both copies from this mtDNA (free energy = −29.5 kcal/mol). Arrowheads in A point to the seven positions variable among all the tandem repeats in types 32–36; those in B point to the two positions variable among the 5′ copies and the five positions variable among the 3′ copies. Heavy lines highlight the TAS that is entirely within the repeated segment. Positions 1 and 76 here correspond to position 15538 in Figures 2 and 3, while 75 and 150 correspond to 15615.

individual. [See FUMAGALLI *et al.* (1996) for support of unidirectional replication slippage as the dominant mechanism in the evolution of repeat arrays.] We also did not find evidence for concerted evolution, another characteristic of many repeat arrays, possibly because the time period has been too short or, alternatively, because the presence of no more than two repeats may make concerted evolution appreciably less likely. Finally, point-mutational heteroplasmy within a repeat seems likely to be rare; our observations for *M. musculus* type 32 (which await definitive assessment via cloning of PCR products) may be analogous to those reported for a coding region in cow mtDNA (HAUSWIRTH and LAIPIS 1982).

The 75-bp repeat unit in *M. musculus* mtDNA encompasses all or most of the two 3′-most of the four TASs identified for *M. domesticus* type 1 mtDNA by DODA *et al.* (1981). It begins 3 bp 3′ of the start of the third TAS (located at 15535–15546 in *M. domesticus* type 1) and extends beyond the fourth TAS (at 15574–15586). Conspicuously, all the *M. musculus* sequences (and, see below, all our other sequences with the tandem repeat) have CTG rather than TCA at positions 15542–15544, while among *M. domesticus* mtDNAs these three base substitutions occur only singly. Perhaps *M. musculus* mtDNA does not use the third TAS during replication.

Figure 10 shows secondary structures that can be produced from one and two tandem repeats. A single *M. musculus* repeat yielded free energies of −6.9 to −8.9 kcal/mol, while two repeats generated proportionally more stable structures, with free energies of −26.6 to −29.5 kcal/mol. Proportionally increased stability from folding two or more repeats of about this size has been reported for other mtDNAs, including those of evening bats [−7.4 to −9.8 kcal/mol for one repeat and −25.4 for two (WILKINSON and CHAPMAN 1991)] and sturgeon [−14.1 kcal/mol for one repeat and −34.4 for two (BUROKER *et al.* 1990)]. The 121 bp 5′ of the repeated segment can likewise form stable secondary structures, with free energies of −19.8 to −26.0 kcal/mol; this observation is consistent with expectations, as that part of the control region contains two TASs and encompasses RS1. The free energies associated with folding the 121 bp 3′ of the repeated segment, in contrast, were

only −12.4 to −13.3 kcal/mol. Different numbers of repeats and more stable structures could have selective consequences, notably with respect to replication.

That there may be a qualitative difference in the *M. domesticus* and *M. musculus* mitochondrial genomes with respect to the ability to generate and/or fix multiple copies of this repeat motif at RS2 is suggested by the fact that none of ~400 *M. domesticus* sequences had a second copy, while in *M. musculus* two tandem repeats were found at a high frequency (14–20%, based on the number of individuals, mtDNA types, or localities). Additionally, in a survey of partial control region and flanking sequences of 37 non-*M. domesticus* commensal house mouse mtDNAs from western and central Asia (using mostly museum skins as the starting material; E. M. PRAGER, C. ORREGO, J. T. MARSHALL and R. D. SAGE, unpublished results), we found duplication of the identical segment in 13 individuals (and 13 of 35 distinct mtDNAs). The tandem repeats in this Asian survey differ from one another within an individual by one to seven point mutations, and the mtDNAs bearing two copies represent one or more lineages phylogenetically not allied with the clade of *M. musculus* types 32–36 in Figure 5. This difference in the mitochondrial genomes of *M. domesticus* and *M. musculus* concerning generation and maintenance of tandem repeats is a counterpart to differences recognized in the nuclear genomes of these closely related house mice, notably with respect to the ability to form and fix Robertsonian chromosomal translocations (REDI *et al.* 1990) and also, for example, as regards the number and organization of satellite DNA sequences (GARAGNA *et al.* 1993).

We thank all the individuals and institutions who facilitated the collecting of mice by R.D.S. and H.T. and who provided us with mice they had caught and/or maintained as laboratory stocks, especially D. KATARANOVSKI (Serbia), R. KRAFT (Germany), F. SEIDL (Austria) and K. UNTERHOLZNER (Austria). We are grateful to U. GYLLENSTEN for DNA, to T. W. QUINN for analyses of DNA secondary structures, to P. BOURSOT, T. W. QUINN and W. K. THOMAS for comments on the manuscript, to S. J. MACK and J. T. MARSHALL for helpful discussion, and to A. G. CLARK for guidance in revising the manuscript. E.M.P. thanks C. A. GAN, C. ORREGO and W. K. THOMAS for reagents and technical advice and S. J. MACK for assistance in generating figures by computer. R.D.S. thanks P. BYERS and M. SAUER, and H.T. thanks L. SANADER for assistance in DNA preparation. This work received support from National Science Foundation grants BSR86-00694, BSR88-18053, and DEB90-08912 to the late A. C. WILSON and National Institutes of Health grant RO1 AI-29800 to R.D.S.

LITERATURE CITED

AGULNIK, S., C. PLASS, W. TRAUT and H. WINKING, 1993 Evolution of a long-range repeat family in chromosome 1 of the genus *Mus*. Mammalian Genome **4:** 704–710.

ALIBERT, P., S. RENAUD, B. DOD, F. BONHOMME and J.-C. AUFFRAY, 1994 Fluctuating asymmetry in the *Mus musculus* hybrid zone: a heterotic effect in disrupted co-adapted genomes. Proc. R. Soc. Lond. B **258:** 53–59.

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., F. VANLERBERGHE and J. BRITTON-DAVIDIAN, 1990 The house mouse progression in Eurasia: a paleontological and archaeozoological approach. Biol. J. Linn. Soc. **41:** 13–25.

BIBB, M. J., R. A. VAN ETTEN, C. T. WRIGHT, M. W. WALBERG and D. A. CLAYTON, 1981 Sequence and gene organization of mouse mitochondrial DNA. Cell **26:** 167–180.

BONHOMME, F., 1986 Evolutionary relationships in the genus *Mus*. Curr. Top. Microbiol. Immunol. **127:** 19–34.

BONHOMME, F., J. CATALAN, J. BRITTON-DAVIDIAN, V. M. CHAPMAN, K. MORIWAKI *et al.*, 1984 Biochemical diversity and evolution in the genus *Mus*. Biochem. Genet. **22:** 275–303.

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.

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:** (in press).

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

BUROKER, N. E., J. R. BROWN, T. A. GILBERT, P. J. O'HARA, A. T. BECKENBACH *et al.*, 1990 Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. Genetics **124:** 157–163.

COPELAND, N. G., N. A. JENKINS, D. J. GILBERT, J. T. EPPIG, L. J. MALTAIS *et al.*, 1993 A genetic linkage map of the mouse: current applications and future prospects. Science **262:** 57–66.

DALLAS, J. F., B. DOD, P. BOURSOT, E. M. PRAGER and F. BONHOMME, 1995 Population subdivision and gene flow in Danish house mice. Mol. Ecol. **4:** 311–320.

DIN, W., R. ANAND, P. BOURSOT, D. DARVICHE, E. JOUVIN-MARCHE *et al.*, 1996 Origin and radiation of the house mouse: clues from nuclear genes. J. Evol. Biol. **9:** (in press).

DI RIENZO, A., and A. C. WILSON, 1991 Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc. Natl. Acad. Sci. USA **88:** 1597–1601.

DOD, B., E. MOTTEZ, E. DESMARAIS, F. BONHOMME and G. ROIZÈS, 1989 Concerted evolution of light satellite DNA in genus *Mus* implies amplification and homogenization of large blocks of repeats. Mol. Biol. Evol. **6:** 478–491.

DODA, J. N., C. T. WRIGHT and D. A. CLAYTON, 1981 Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc. Natl. Acad. Sci. USA **78:** 6116–6120.

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.

FIGUEROA, F., M. KASAHARA, H. TICHY, E. NEUFELD, U. RITTE *et al.*, 1987 Polymorphism of unique noncoding DNA sequences in wild and laboratory mice. Genetics **117:** 101–108.

FORT, P., F. BONHOMME, P. DARLU, M. PIECHACZYK, P. JEANTEUR *et al.*, 1985 Clonal divergence of mitochondrial DNA *vs.* populational evolution of nuclear genome. Evol. Theor. **7:** 81–90.

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.

FUMAGALLI, L., P. TABERLET, L. FAVRE and J. HAUSSER, 1996 Origin and evolution of homologous repeated sequences in the mitochondrial DNA control region of shrews. Mol. Biol. Evol. **13:** 31–46.

GARAGNA, S., C. A. REDI, E. CAPANNA, N. ANDAYANI, R. M. ALFANO *et al.*, 1993 Genome distribution, chromosomal allocation, and organization of the major and minor satellite DNAs in 11 species and subspecies of the genus *Mus*. Cytogenet. Cell Genet. **64:** 247–255.

GERASIMOV, S., H. NIKOLOV, V. MIHAILOVA, J.-C. AUFFRAY and F. BONHOMME, 1990 Morphometric stepwise discriminant analysis of the five genetically determined European taxa of the genus *Mus*. Biol. J. Linn. Soc. **41:** 47–64.

HAMMER, M. F., and L. M. SILVER, 1993 Phylogenetic analysis of the alpha-globin pseudogene-4 (*Hba-ps4*) locus in the house mouse species complex reveals a stepwise evolution of *t* haplotypes. Mol. Biol. Evol. **10:** 971–1001.

HAMMER, M. F., and A. C. WILSON, 1987 Regulatory and structural genes for lysozymes of mice. Genetics **115:** 521–533.

HAUSWIRTH, W. W., and P. J. LAIPIS, 1982 Mitochondrial DNA polymorphism in a maternal lineage of Holstein cows. Proc. Natl. Acad. Sci. USA **79:** 4686–4690.

HOELZEL, A. R., 1993 Evolution by DNA turnover in the control region of vertebrate mitochondrial DNA. Curr. Opin. Genet. Dev. **3:** 891–895.

HOELZEL, A. R., J. V. LOPEZ, G. A. DOVER and S. J. O'BRIEN, 1994 Rapid evolution of a heteroplasmic repetitive sequence in the mitochondrial DNA control region of carnivores. J. Mol. Evol. **39:** 191–199.

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

KLEIN, J., V. VINCEK, M. KASAHARA and F. FIGUEROA, 1988 Probing mouse origins with random DNA probes. Curr. Top. Microbiol. Immun. **137:** 55–63.

KOCHER, T. D., and A. C. WILSON, 1991 Sequence evolution of mitochondrial DNA in humans and chimpanzees: control region and a protein-coding region, pp. 391–413 in *Evolution of Life. Fossils, Molecules, and Culture,* edited by S. OSAWA and T. HONJO. Springer-Verlag, Tokyo.

KOROBITSYNA, K. V., L. V. YAKIMENKO and L. V. FRISMAN, 1993 Genetic differentiation of house mice in the fauna of the former U.S.S.R.: results of cytogenetic studies. Biol. J. Linn. Soc. **48:** 93–112.

MACHOLÁN, M., and J. ZIMA, 1994 Mus domesticus in western Bohemia: a new mammal for the Czech Republic. Folia Zoologica **43:** 39–41.

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. Immun. **127:** 12-18.

MEZHZHERIN, S. V., and E. V. KOTENKOVA, 1992 Biochemical systematics of house mice from the central Palearctic region. Z. Zool. Syst. Evolut.-Forsch. **30:** 180–188.

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.

MORITA, T., H. KUBOTA, K. MURATA, M. NOZAKI, C. DELARBRE *et al.,* 1992 Evolution of the mouse *t* haplotype: recent and worldwide introgression to *Mus musculus.* Proc. Natl. Acad. Sci. USA **89:** 6851–6855.

MORIWAKI, K., H. YONEKAWA, O. GOTOH, M. MINEZAWA, H. WINKING *et al.,* 1984 Implications of the genetic divergence between European wild mice with Robertsonian translocations from the viewpoint of mitochondrial DNA. Genet. Res. **43:** 277–287.

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.

NACHMAN, M. W., and C. F. AQUADRO, 1994 Polymorphism and divergence at the 5′ flanking region of the sex-determining locus, *Sry,* in mice. Mol. Biol. Evol. **11:** 539–547.

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., T. SHIROISHI, N. MIYASHITA, K. TSUCHIYA, H. IKEDA *et al.,* 1994 Distribution of the molossinus allele of *Sry,* the testis-determining gene, in wild mice. Mol. Biol. Evol. **11:** 864–874.

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.

REDI, C. A., S. GARAGNA and M. ZUCCOTTI, 1990 Robertsonian chromosome formation and fixation: the genomic scenario. Biol. J. Linn. Soc. **41:** 235–255.

RITTE, U., E. MARKMAN and E. NEUFELD, 1992 Can the variability of mitochondrial DNA distinguish between commensal and feral populations of the house mouse? Biol. J. Linn. Soc. **46:** 235–245.

RUVINSKY, A., A. POLYAKOV, A. AGULNIK, H. TICHY, F. FIGUEROA *et al.,* 1991 Low diversity of *t* haplotypes in the eastern form of the house mouse, *Mus musculus* L. Genetics **127:** 161–168.

RYAN, A. W., E. J. DUKE and J. S. FAIRLEY, 1993 Polymorphism, localization and geographical transfer of mitochondrial DNA in *Mus musculus domesticus* (Irish house mice). Heredity **70:** 75–81.

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., 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. Immun. **127:** 75–85.

SAGE, R. D., E. M. PRAGER, H. TICHY and A. C. WILSON, 1990 Mitochondrial DNA variation in house mice, *Mus domesticus* (Rutty). Biol. J. Linn. Soc. **41:** 105–123.

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.

SHERRY, S. T., A. R. ROGERS, H. HARPENDING, H. SOODYALL, T. JENKINS *et al.,* 1994 Mismatch distributions of mtDNA reveal recent human population expansions. Human Biol. **66:** 761–775.

SOKAL, R. R., and F. J. ROHLF, 1981 *Biometry.* W. H. Freeman and Co., New York.

STEWART, D. T., and A. J. BAKER, 1994a Patterns of sequence variation in the mitochondrial D-loop region of shrews. Mol. Biol. Evol. **11:** 9–21.

STEWART, D. T., and A. J. BAKER, 1994b Evolution of mtDNA D-loop sequences and their use in phylogenetic studies of shrews in the subgenus *Otisorex* (*Sorex:* Soricidae: Insectivora). Mol. Phylo. Evol. **3:** 38–46.

SUZUKI, H., and Y. KURIHARA, 1994 Genetic variation of ribosomal RNA in the house mouse, *Mus musculus,* pp. 107–119 in *Genetics in Wild Mice,* edited by K. MORIWAKI, T. SHIROISHI and H. YONEKAWA. Japan Sci. Soc. Press, Tokyo/S. Karger, Basel.

SUZUKI, H., K. MORIWAKI and S. SAKURAI, 1994 Sequences and evolutionary analysis of mouse 5S rDNAs. Mol. Biol. Evol. **11:** 704–710.

SWOFFORD, D. L., 1991 *PAUP (Phylogenetic Analysis Using Parsimony)* Version 3.0s. Illinois Natural History Survey, Champaign, IL.

TICHY, H., Z. ZALESKA-RUTCZYNSKA, C. O'HUIGIN, F. FIGUEROA and J. KLEIN, 1994 Origin of the North American house mouse. Folia Biologica (Praha) **40:** 483–496.

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

TUCKER, P. K., R. D. SAGE, J. H. WARNER, A. C. WILSON and E. M. EICHER, 1992 Abrupt cline for sex chromosomes in a hybrid zone between two species of mice. Evolution **46:** 1146–1163.

VANLERBERGHE, F., B. DOD, P. BOURSOT, M. BELLIS and F. BONHOMME, 1986 Absence of *Y*-chromosome introgression across the hybrid zone between *Mus musculus domesticus* and *Mus musculus musculus.* Genet. Res. **48:** 191–197.

VANLERBERGHE, F., P. BOURSOT, J. T. NIELSEN and F. BONHOMME, 1988 A steep cline for mitochondrial DNA in Danish mice. Genet. Res. **52:** 185–193.

WILKINSON, G. S., and A. M. CHAPMAN, 1991 Length and sequence variation in evening bat D-loop mtDNA. Genetics **128:** 607–617.

YONEKAWA, H., S. TAKAHAMA, O. GOTOH, N. MIYASHITA and K. MORIWAKI, 1994 Genetic diversity and geographic distribution of *Mus musculus* subspecies based on the polymorphism of mitochondrial DNA, pp. 25–40 in *Genetics in Wild Mice,* edited by K. MORIWAKI, T. SHIROISHI and H. YONEKAWA. Japan Sci. Soc. Press, Tokyo/S. Karger, Basel.

ZUKER, M., and P. STIEGLER, 1981 Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. **9:** 133–148.

Communicating editor: A. G. CLARK