

## Phylogenetic Relationships in the Genus *Mus*, Based on Paternally, Maternally, and Biparentally Inherited Characters

BARBARA L. LUNDRIGAN,<sup>1</sup> SHARON A. JANSA,<sup>2</sup> AND PRISCILLA K. TUCKER<sup>3</sup>

<sup>1</sup>Michigan State University Museum and Department of Zoology, Michigan State University, East Lansing, Michigan 48824-1045, USA; E-mail: [lundriga@msu.edu](mailto:lundriga@msu.edu)

<sup>2</sup>Division of Mammalogy, Department of Systematic Biology, National Museum of Natural History, Smithsonian Institution, Washington, D.C. 20560-0108, USA

<sup>3</sup>Museum of Zoology and Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109-1079, USA

**Abstract.**—Several species in the rodent genus *Mus* are used as model research organisms, but comparative studies of these mice have been hampered by the lack of a well-supported phylogeny. We used DNA sequences from six genes representing paternally, maternally, and biparentally inherited regions of the genome to infer phylogenetic relationships among 10 species of *Mus* commonly used in laboratory research. Our sample included seven species from the subgenus *Mus*; one species each from the subgenera *Pyromys*, *Coelomys*, and *Nannomys*; and representatives from three additional murine genera, which served as outgroups in the phylogenetic analyses. Although each of the six genes yielded a unique phylogeny, several clades were supported by four or more gene trees. Nodes that conflicted between trees were generally characterized by weak support for one or both of the alternative topologies, thus providing no compelling evidence that any individual gene, or part of the genome, was misleading with respect to the evolutionary history of these mice. Analysis of the combined data resulted in a fully resolved tree that strongly supports monophyly of the genus *Mus*, monophyly of the subgenus *Mus*, division of the subgenus *Mus* into Palearctic (*M. musculus*, *M. macedonicus*, *M. spicilegus*, and *M. spretus*) and Asian (*M. cervicolor*, *M. cookii*, and *M. caroli*) clades, monophyly of the house mice (*M. m. musculus*, "*M. m. molossinus*," *M. m. castaneus*, and *M. m. domesticus*), and a sister-group relationship between *M. macedonicus* and *M. spicilegus*. Other clades that were strongly supported by one or more gene partitions were not strongly supported by the combined data. This appears to reflect a localized homoplasy in one partition obscuring the phylogenetic signal from another, rather than differences in gene or genome histories. [*Coelomys*; molecular phylogeny; Muridae; *Mus*; *Nannomys*; *Pyromys*.]

The genus *Mus* (family Muridae, subfamily Murinae) encompasses some 30–40 species of small, nocturnal, terrestrial rodents (Marshall, 1977, 1981, 1998; Musser and Carleton, 1993). The natural geographic range for *Mus* includes Europe, Africa, and Asia, but one species, the house mouse *Mus musculus*, has been introduced to other parts of the world. Despite their inconspicuous appearance, mice in the genus *Mus* have long held the interest of humans. As early as 12,000 years ago, the house mouse appeared in settlements in the Middle East (Auffray et al., 1988). Since that time, the house mouse has been both defiled, for contaminating food stores, damaging property, and spreading disease; and deified, for its purported powers in prophesy and the art of healing (Morse, 1981; Berry, 1995). During the Edo era (100–400 years ago), Japanese hobbyists, fascinated by the variety of mouse coat color variants, bred house mice to produce new and unusual forms, giving rise to the "fancy mouse" trade, which eventually spread to Europe and England (Morse, 1981;

Moriwaki, 1994). The ease with which these mice could be bred and maintained in captivity also attracted biologists; by the early 1900s, the house mouse had been adopted as a model for biological research (Morse, 1981; Berry, 1995).

Initially, mouse studies focused on a few inbred strains of *M. musculus*. Genetic diversity was relatively high because the founders represented multiple distinct populations of this highly polymorphic taxon. Subsequently, the genetic diversity of laboratory mice was augmented by the introduction of wild mice from additional populations of *M. musculus* and from other species in the genus. This made it possible to investigate inherited variants not found in already established strains, and often for the first time, to take an evolutionary approach to the investigation of character variation, that is, to ask when, where, how many times, and under what circumstances, those characters arose. Today, representatives from at least 12 of the 38 species of *Mus* (listed by Musser and Carleton, 1993) are being raised in captivity

to serve as model research organisms. Comparative studies of these mice have contributed important insights in a wide range of biological disciplines, including genetics, developmental biology, immunology, endocrinology, physiology, functional anatomy, animal behavior, and organismal evolution (e.g., Hammer and Wilson, 1987; Jouvin-Marche et al., 1988; McConnell et al., 1988; Tutter and Riblet, 1989; Bush and Paigen, 1992; Morita et al., 1992; Garagna et al., 1993; Rheume et al., 1994; Sueyoshi et al., 1995; Sharma, 1996; Patris and Baudoin, 1998; Karn and Nachman, 1999). In addition, *Mus* has been the subject of hundreds of field studies, which are illuminating the behavior, ecology, and population biology of wild mice (e.g., Chandrachud, 1974; Muntyanu, 1990; Bhat and Sujatha, 1991; Chou et al., 1998).

#### *Advances and Conflicts in the Systematics of Mus*

Given the importance of *Mus* in comparative studies, it is not surprising that considerable effort has been devoted to establishing a phylogeny for this genus. Initially, the task proved difficult because morphological differences among the forms are often subtle, complicating both the identification of exemplar taxa and the phylogenetic placement of those taxa. The introduction of molecular techniques to *Mus* systematics in the late 1960s provided a large new data set for addressing these problems. Allozyme data were used to reexamine existing taxonomic divisions, identify morphological discriminators, establish species boundaries, and describe hybrid zones (e.g., Selander et al., 1969; Hunt and Selander, 1973; Bonhomme et al., 1978, 1983; Britton and Thaler, 1978; Thaler et al., 1981; Darviche and Orsini, 1982; Orsini et al., 1983; Auffray et al., 1990; Frisman et al., 1990; Gerasimov et al., 1990). Subsequently, historical relationships were explored using a variety of molecular markers, including allozymes (Sage, 1981; Bonhomme et al., 1984; She et al., 1990), mitochondrial DNA restriction fragment length polymorphisms (mtDNA RFLPs; Ferris et al., 1983b; She et al., 1990), RFLPs of nuclear rDNA spacer regions (Suzuki and Kurihara, 1994), single-copy nuclear DNA (scnDNA) hybridization (She et al., 1990; Catzeflis and Denys, 1992), mtDNA sequences (Fort et al., 1984; Sourrouille et al., 1995; Prager et al.,

1996), and nuclear DNA sequences (Jouvin-Marche et al., 1988; Lundrigan and Tucker, 1994).

Despite this wealth of data, presenting a single phylogeny for the genus *Mus* is still problematic. The results of different studies are in agreement with respect to some relationships but conflict with respect to others. There is no objective method for combining the data, because the different data sets differ substantially in the taxa examined, the nature of data collected, and the methods used for generating trees. In 1993, Boursot et al. presented a "synthetic tree" summarizing the results from several of these studies (Fig. 1; see also Bonhomme and Guenet, 1995). In this tree, taxa are grouped whenever that grouping is supported by one or more independent data sets and no compelling evidence contradicts it. Although a useful summary, this tree is difficult to assess because there is no objective way to quantify the strength of these relationships. Clades that are well supported can not be distinguished from those that are poorly supported, nor can character conflict be directly examined; disagreements among data sets are represented as polytomies, without reference to the weight of evidence supporting conflicting topologies.

In this study, we generate a phylogeny for the genus *Mus* by using DNA sequences from six genes, representing paternally, maternally, and biparentally inherited regions of the genome. Our sampling of *Mus* species is too limited to pretend a major resolution of intrageneric relationships. Rather, this phylogeny represents a hypothesis of relationships among laboratory stocks (and the wild populations from which they were derived). We include DNA sequence data from earlier studies, in addition to newly collected sequences, focusing, as Boursot et al. (1993) did, on the species most commonly used in laboratory studies. Our goals are to test their hypothesis of relationships with DNA sequence data from a variety of sources, resolve polytomies, establish relative support for clades, and explore areas of character conflict.

## MATERIALS AND METHODS

### *Taxon Sampling*

The 30–40 species of *Mus* are distributed into four subgenera: *Mus*, *Pyromys*, *Coelomys*, and *Nannomys*, each of which is diagnosed by a combination of morphological and

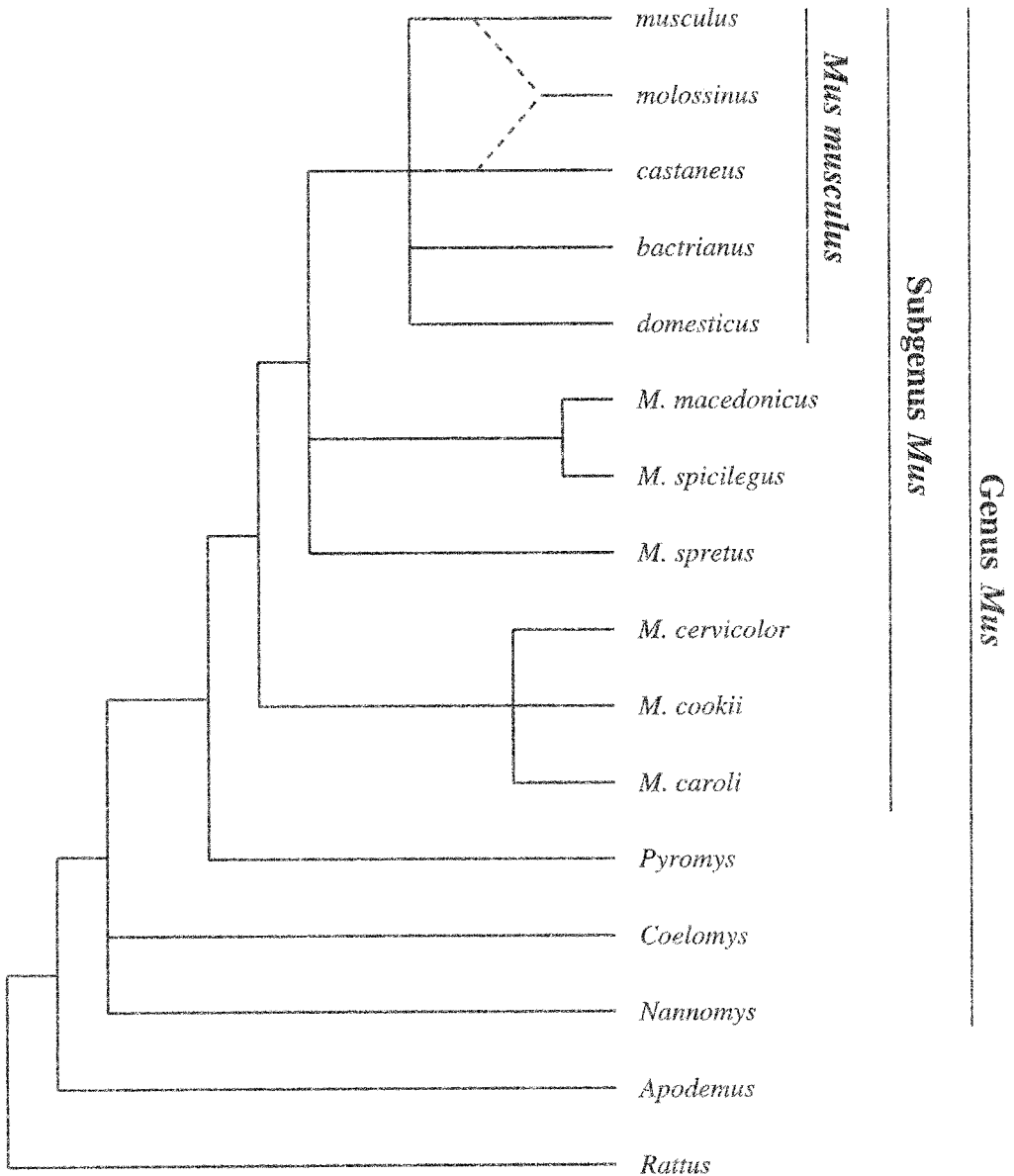


FIGURE 1. "Synthetic tree" redrawn from Boursot et al. (1993) depicting historical relationships within the genus *Mus* based on allozymes (Sage, 1981; Bonhomme et al., 1984; She et al., 1990), *16S* sequences (Fort et al., 1984), *C<sub>1</sub>* sequences (Jouvin-Marche et al., 1988), mtDNA RFLPs, and scnDNA hybridization data (She et al., 1990). No particular algorithm was used to generate this tree; it was rooted with *Apodemus* and *Rattus*. The dashed lines coming from "*M. m. molossinus*" reflect the hybrid origin of this mouse (Yonekawa et al., 1986, 1988; Bonhomme et al., 1989).

molecular characteristics (Marshall, 1977, 1981, 1998; Bonhomme et al., 1984; Catzeflis and Denys, 1992; Musser and Carleton, 1993; Sourrouille et al., 1995). The subgenus *Mus* (which includes the house and rice field mice) is by far the best studied. We sampled seven of the nine species in this subgenus (after Musser and Carleton, 1993; Table 1),

including representatives from each of the three well-defined subspecies of the widespread and highly polymorphic *M. musculus*: *M. m. musculus*, *M. m. castaneus*, and *M. m. domesticus* (a fourth subspecies, *M. m. bactrianus*, was not included because its status is uncertain; Boursot et al., 1996; Din et al., 1996). *Mus m. musculus* and

TABLE 1. Taxonomic sampling and GenBank accession numbers for DNA sequences used in this study. Classification follows Musser and Carleton (1993). Gene abbreviations and regions sequenced are defined in Table 2. Sequences with accession numbers beginning with "AY" were collected specifically for this study. ND = no data.

Taxon			Gene sequences									
Genus	Subgenus	Species	Subspecies	Sry	Cyt b	I2S	B2m	Zp-3 (1)	Zp-3 (2)	Zp-3 (3)	Tcp-1	
<i>Mus</i>	<i>(Mus)</i>	<i>musculus</i>	<i>musculus</i>	a	AY057804	AY057791	AY057800	AY057746	AY057761	AY057776	M61217	
			<i>molossinus</i>	a	AY057806	AY057793	AY057801	AY057748	AY057763	AY057778	X61220	
			<i>castaneus</i>	a	AY057805	AY057792	M84366	AY057747	AY057762	AY057777	X61219	
			<i>domesticus</i>	L29551	AY057807	b	M84363	AY057749	AY057764	AY057779	X61214	
		<i>spretus</i>	L29544	AY057810	AY057796	M84362	AY057752	AY057767	AY057782	AY057780	X61215	
		<i>macedonicus</i>	L29547	AY057808	AY057794	L05899	AY057750	AY057765	AY057780	AY057780	X61213	
		<i>spicilegus</i>	L29550	AY057809	AY057795	L05578	AY057751	AY057766	AY057781	AY057781	X61216	
		<i>caroli</i>	L29552	AY057812	AY057798	L04989	AY057754	AY057769	AY057784	AY057784	AY057819	
		<i>cervicolor</i>	L29548	AY057811	AY057797	L04988	AY057753	AY057768	AY057783	AY057783	X61218	
		<i>cooki</i>	L29549	AY057813	X85946	L05581	AY057755	AY057770	AY057785	AY057785	AY057820	
<i>Mastomys</i>	<i>(Pyromys)</i>	<i>saxicola</i>	ND	AY057815	X85948	AY057802	AY057757	AY057772	AY057787	AY057787	AY057822	
		<i>pahari</i>	L29543	AY057814	X84383	L05579	AY057756	AY057771	AY057786	AY057786	AY057821	
<i>Hylomyscus</i>	<i>(Nannomys)</i>	<i>minutooides</i>	AY057745	AY057816	AY057799	AY057803	AY057758	AY057773	AY057775	AY057780	AY057823	
			L29542	AY057818	X85952	ND	AY057760	AY057775	AY057778	AY057790	AY057824	
			L29541	AY057817	X85953	ND	AY057759	AY057774	AY057789	AY057789	AY057789	
<i>Rattus</i>		L29545	J01436	J01438	Y08531	D78482	D78482	D78482	D78482	X61221		

<sup>a</sup>From Lundrigan and Tucker, 1994.

<sup>b</sup>From Sourrouille et al., 1995.

*M. m. domesticus* exchange genes along a narrow hybrid zone in western Europe (Hunt and Selander, 1973; Ferris et al., 1983a; Sage et al., 1986; Vanlenberghe et al., 1988; Boursot et al., 1989) and in a broad zone of introgression in the Transcaucasus (Mezhzherin et al., 1998); *M. m. musculus* and *M. m. castaneus* hybridize in Japan (Yonekawa et al., 1986, 1988; Bonhomme et al., 1989) and in other parts of eastern Asia (Frisman et al., 1990). Yet, if sampled away from regions of hybridization, these three taxa are genetically and morphologically distinct. As a consequence, they are often accorded species status (e.g., Marshall, 1981, 1998; Sage, 1981; Sage et al., 1993; Prager et al., 1996). To simplify comparisons with Boursot et al. (1993), we refer to them as subspecies throughout this paper. We also included the Japanese mouse, sometimes referred to as "*M. m. molossinus*." This mouse was once considered an independent subspecies of *M. musculus* but is now known to be of hybrid origin, a mixture of *M. m. musculus*, *M. m. castaneus*, and to a lesser extent, *M. m. domesticus* genomes (Yonekawa et al., 1986, 1988; Bonhomme et al., 1989). A single species was sampled from each of the other three subgenera: *Pyromys* (spiny mice), *Coelomys* (shrew mice), and *Nannomys* (African pygmy mice). The Indomalayan *Pyromys* and *Coelomys* each include five well-defined species (Marshall, 1977). In contrast, the subgenus *Nannomys* is poorly defined, with anywhere from 5 to 30 species (Marshall, 1981), several having been described solely on the basis of chromosomal differences.

Three additional murine genera, *Mastomys*, *Hylomyscus*, and *Rattus*, were used as outgroups for rooting our phylogenies. DNA hybridization studies suggest that the *Praomys* group (which includes *Mastomys* and *Hylomyscus*) may be the sister lineage to *Mus* (Catzeflis and Denys, 1992; Chevret et al., 1994). The monophyly of *Mus* with respect to these genera is weakly supported by a cladistic analysis of 12S sequences (Sourrouille et al., 1995). *Rattus* is thought to lie well outside of *Mus*, an inference supported by both scnDNA hybridization data (She et al., 1990) and nuclear DNA sequence data (Jouvin-Marche et al., 1988). Except for the *M. saxicola* 12S sequence collected by Sourrouille et al. (1995), all *Mus* DNA samples used in this study came from common laboratory stock. Samples of other

genera were derived from either laboratory or wild-caught animals. Information on the origin of all DNA samples is provided in the Appendix.

### Genes Sequenced

The complete data set includes nucleotide sequences from six genes (Table 2): the male sex-determining locus (*Sry*), cytochrome *b* (*Cyt b*), 12S ribosomal RNA (12S),  $\beta_2$ -microglobulin (*B2m*), zona pellucida-3 (*Zp-3*), and t-complex polypeptide-1 (*Tcp-1*). These genes represent three different regions of the genome, each with a distinct mode of inheritance: *Sry* is a paternally inherited Y-linked gene; *Cyt b* and 12S are maternally inherited mitochondrial genes; and *B2m*, *Zp-3*, and *Tcp-1* are biparentally inherited autosomal genes. The three autosomal loci have each been mapped to different chromosomes in *M. musculus* (chromosomes 2, 6, and 17, respectively). For the purposes of this study, we assume that none of these autosomal loci are linked in any of the other taxa examined.

Although *Cyt b* and 12S are often used as a source for phylogenetic characters, *Sry*, *B2m*, *Zp-3*, and *Tcp-1* are not as familiar to systematists. The Y chromosome-linked gene, *Sry*, codes for testis determination in mammals (Gubbay et al., 1990; Sinclair et al., 1990). It includes a single exon with an open reading frame consisting of a conserved 237-base-pair (bp) high-mobility group DNA binding domain (HMG box) flanked by N-terminal and C-terminal regions of variable length. We sequenced 76 bp 5' of the start codon, the N-terminal region, the HMG box, and 187 bp of the C-terminal region, using primers developed in an earlier study (Lundrigan and Tucker, 1994). We were unable to obtain *Sry* sequences for *M. saxicola*.

$\beta_2$ -Microglobulin is a small secretory protein that appears to have a number of functions, the most notable involving the immune response (Bjorkman et al., 1987; Simister and Mostov, 1989; Williams et al., 1989; Vitiello et al., 1990). The gene has four exons, the second of which encodes 92 of the 99 amino acids found in the mature protein (Parnes and Seidman, 1982). We amplified exon 2 using various combinations of four primers, two designed by Hermel et al. (1993) and two designed by us for this study (Table 2). Sequences from *B2m* for two of the outgroup taxa, *Mastomys* and *Hylomyscus*, were not

TABLE 2. Regions sequenced, length of aligned sequences before and after removal of alignment ambiguous sites, and primers or source of primers used. For zona pellucida-3, three noncontiguous segments were sequenced (labeled 1-3 here and in Table 1). F = forward primer, R = reverse primer.

Locus	Abbr.	Region sequenced	Length		Primers/Reference
			Total	Ambiguous removed	
Male sex-determining locus Cytchrome <i>b</i> 12S ribosomal RNA	<i>Sry</i>	Beginning 76 bp 5' to start codon	506	506	Lundrigan and Tucker, 1994
	<i>Cyt b</i>	Entire locus	1,145	1,145	Jansa et al., 1999
	<i>12S</i>	Entire locus	987	894	Kocher et al., 1989 (primer L1091) F: 5'-AGGCACCTGAAAATGCCCTAGATG-3' R: 5'-TCGATTATAGAACACAGGCTCCCTCTAG-3' R: 5'-TGATTACGCTACCCTT(C/T)GCACGG-3'
$\beta_2$ -Microglobulin	<i>B2m</i>	Exon 2	278	278	Hermel et al., 1993 (primers P3 and P4) F: 5'-CACGCCACCCACCGAGAAATG-3' R: 5'-ACTGACATGCTTAACTCTGCCA-3'
Zona pellucida-3	<i>Zp-3</i>	(1) Exon 1	240	240	F: 5'-TGTCTCCTGCTGTGTGGAGGC-3' R: 5'-CTGCACCCCTGTGCTGCA-3'
		(2) Exons 3 and 4, intron 3	335	335	F: 5'-GCAGGGCAATGTGACAGCCAC-3' R: 5'-CCDTGGAAAGTCCACDATGA-3'
		(3) Exons 6 and 7, intron 6	325	307	F: 5'-TCTACATCACCTGCCATCTCAAAG-3' R: 5'-CGTGCCTGGGTTTCGAGA-3'
T-complex polypeptide-1 All combined	<i>Tcp-1</i>	Exons 8-10, introns 8 and 9	1,454 5,270	1,408 5,113	Morita et al., 1992

obtainable with any of the four possible combinations of these primers.

The *Zp-3* gene codes for the primary sperm receptor in mammals and appears to be responsible for controlling species-specific gamete interactions (Bleil and Wassarman, 1980a,b; Ringuette et al., 1988; Wassarman, 1990; but see Rankin et al., 1998). In *M. m. domesticus*, the gene includes eight exons, ranging from 92 to 338 bp long. Three regions were targeted for polymerase chain reaction (PCR) amplification: Region one includes most of exon 1 (bp 81–320; numbering after Kinloch and Wassarman, 1989), region two includes most of exons 3 and 4 and all of intron 3 (bp 4,105–4,435), and region three includes most of exons 6 and 7 and all of intron 6 (bp 7,220–7,523). A total of 668 bp of exon sequence and 232 bp of intron sequence were collected.

The *Tcp-1* gene has been mapped to the region of chromosome 17 responsible for transmission-ratio distortion and male sterility produced by the mouse t complex (Silver et al., 1979; Lyon, 1991). It is believed to code for a protein that is part of the testicular cell surface matrix (Silver and White, 1982). We targeted a single region of *Tcp-1* that includes exons 8, 9, and 10 and introns 8 and 9 (bp 1–1,231; numbering after Morita et al., 1992), using primers developed by Morita et al. (1992). A total of 399 bp of exon sequence and 1,055 bp of intron sequence were collected.

Forty-four of the sequences used as a source of phylogenetic characters were obtained from GenBank or the literature; the remaining 81 sequences were collected specifically for this study (Table 1).

#### DNA Amplification and Sequencing

Genomic DNA was isolated from frozen tissue according to the methods of Jenkins et al. (1982). Sequences were enzymatically amplified by PCR (Saiki et al., 1985) using primer sequences obtained from the literature or designed specifically for this project (Table 2). Amplifications were done as standard 100- $\mu$ l reactions, on a Perkin-Elmer 480 Thermal Cycler, using the recommended concentrations of primers, unincorporated nucleotides, buffer, and  $MgCl_2$ , and the following reaction conditions: 25 to 30 cycles, with denaturation at 95°C for 1 min, annealing at 50–65°C for 1 min, and extension at 72°C for 1 min 15 s. All amplifications were

preceded by a 95°C soak for 3 min and followed by a 7-min extension at 72°C. PCR products were prepared for automated sequencing by separation on a 2% agarose gel (NuSieve GTG; FMC Bioproducts) followed by purification using a QIA-quick Gel Extraction Kit (Qiagen Inc.). Purified PCR products were sequenced in both directions using a Perkin-Elmer Dye Termination Sequencing Kit and an ABI 377 automated sequencer. All sequences were proofed and edited using Sequence Navigator ver. 1.0 (Applied Biosystems).

#### Sequence Alignment

Sequences from the five protein-coding genes (*Sry*, *Cyt b*, *B2m*, *Zp-3*, and *Tcp-1*) were aligned using the multiple alignment algorithm in CLUSTAL W (Thompson et al., 1994) and adjusted by eye. For the 12S data, we constructed an alignment based on the secondary structure model of Springer and Douzery (1996). Aligning the *Sry*, *Cyt b*, and *B2m* sequences was nonproblematic: Both *Cyt b* and *B2m* are length-invariant for these taxa and the *Sry* alignment required only one single-base-pair insertion. The *Zp-3*, *Tcp-1*, and 12S matrices, however, each include several regions of variable length and gaps were introduced to align these sequences. Regions that could not be aligned with reasonable certainty were excluded from the phylogenetic analysis, resulting in the elimination of 157 bp: 18 from *Zp-3* intron 6, 46 from *Tcp-1* introns 8 and 9, and 93 from loop regions of 12S (Table 2). Sequence alignments for all six genes, along with a list of alignment ambiguous sites, can be downloaded from <http://treebase.org>.

#### Data Set Partitioning

Considerable controversy exists over whether data from different sources (e.g., mitochondrial vs. nuclear genes) should be analyzed separately or combined into a single "total evidence" analysis (e.g., Kluge, 1989; Bull et al., 1993; de Queiroz, 1993; Eernisse and Kluge, 1993; Kluge and Wolf, 1993; de Queiroz et al., 1995; Miyamoto and Fitch, 1995; Huelsenbeck et al., 1996a; DeSalle and Brower, 1997; Wiens, 1998). Advocates for separate analyses stress that characters within a data set are more likely to be nonindependent estimators of phylogeny than are characters from different data sets.

They warn that combining incongruent data sets can mislead with respect to historical relationships among the organisms of interest or obscure those relationships (e.g., Lanyon, 1993; Miyamoto and Fitch, 1995). In contrast, advocates for "total evidence" prefer the combined approach because it does not require the recognition of partitions and maximizes the descriptive and explanatory power of the evidence (e.g., Kluge, 1989; Eernisse and Kluge, 1993; Kluge and Wolf, 1993; see also Miyamoto, 1985). In this study, we initially partitioned the data to facilitate comparison of various subsets of the data matrix. Subsequently, all data were combined in a single simultaneous analysis.

One difficulty with partitioning is the lack of definitive criteria for identifying appropriate boundaries for data partitions (Kluge and Wolf, 1993; Siddall, 1997; DeSalle and Brower, 1997; but see Miyamoto and Fitch, 1995). We partitioned our data into subsets that might be expected to have a high probability of incongruence because of their differing biological properties. Specifically, we targeted differences in function, linkage group, and transmission properties—first partitioning along the six gene boundaries (*Sry*, *Cyt b*, *12S*, *B2m*, *Zp-3*, and *Tcp-1*), then among the five linkage groups (the Y chromosome, mitochondrion, and chromosomes 2, 6, and 17), and finally among the three modes of inheritance (paternal, maternal, and biparental).

#### *Phylogenetic Analyses*

*Maximum parsimony.*—We performed a maximum parsimony analysis on each partition and on the combined matrix, using the branch-and-bound option in PAUP\* ver. 4.0b4a (Swofford, 1998). Phylogenetically informative characters were unordered and equally weighted; gaps were treated as missing characters. Clade stability was assessed by bootstrap analysis (Felsenstein, 1985) and Bremer support analysis (Bremer, 1988, 1994; also known as decay analysis, Donoghue et al., 1992). All bootstrap analyses included 1,000 replicates; searches were heuristic with 20 random taxon additions and tree-bisection-and-reconnection (TBR) branch swapping. TreeRot ver. 2 (Sorenson, 1999) was used to construct PAUP\* command files for use in Bremer support calculations. Patterns of character change were explored using MacClade ver. 3.06 (Maddison and Maddison, 1992).

*Maximum likelihood.*—We also conducted a maximum likelihood analysis of each data partition, using PAUP\*. The best-fit maximum likelihood model for each partition was determined using the following strategy. Jukes–Cantor-corrected distances (JC; Jukes and Cantor, 1969) were used to calculate a neighbor-joining tree. This tree was used to estimate the log-likelihood scores and parameter values under eight models of nucleotide substitution: Jukes and Cantor (1969; JC69), Felsenstein (1981; F81), Hasegawa et al. (1985; HKY85), Tamura and Nei (1993; TrN), Kimura (1980; K2P), Kimura (1981; K3P), Zharkikh (1994; SYM), and Rodriguez et al. (1990; GTR). We also assessed whether allowing for gamma-distributed heterogeneity of the substitution rate across sites (Yang, 1994;  $\Gamma$ -shape parameter) improved the fit of each model to the data, and whether allowing for a proportion of invariant sites (I) improved the fit of each model to the data. The best-fit model was the one for which additional parameters no longer significantly improved the log-likelihood score, as determined with a likelihood-ratio test (Goldman, 1993; Huelsenbeck and Rannala, 1997).

Finally, we evaluated whether enforcing a molecular clock provided a better fit to the data than did allowing for different rates across the tree. To provide the most conservative test for a clock-like model of evolution, we calculated a UPGMA tree based on JC distances and calculated the likelihood score for the best-fit model with no clock enforced versus the same model with a clock enforced.

Subsequent to model evaluation and selection, the maximum likelihood tree for each data set was determined using a heuristic search in which the parameter values under the best-fit model were fixed and a neighbor-joining tree was used as a starting point for TBR branch swapping. Parameter values were subsequently reestimated on this tree. Bootstrap estimates for maximum likelihood were obtained from 100 replicates using the "fast" stepwise-addition option.

#### *Data Set Heterogeneity*

*Incongruence length difference (ILD).*—To assess character congruence among the data partitions under each partitioning scheme, we used the Mickevich–Farris character incongruence metric (Mickevich and Farris, 1981). Statistical significance of



incongruence was calculated with the ILD test as described in Farris et al. (1994) and implemented in PAUP\* as the partition homogeneity test. Each test included 1,000 replicates; searches were heuristic with simple taxon addition and TBR branch swapping. All analyses were done using informative characters only (see Carpenter, 1996; Cunningham, 1997).

*Parametric bootstrapping.*—As an alternative test for evaluating whether a combined analysis of the mitochondrial and nuclear data sets resulted in a tree that was inconsistent with trees resulting from either data set analyzed alone, we used a parametric bootstrapping approach (Huelsenbeck et al., 1996b; Huelsenbeck and Rannala, 1997; Goldman et al., 2000). We fit the mitochondrial data to the combined data tree using the GTR + I +  $\Gamma$  model with clock enforced and calculated the difference in likelihood scores ( $-\ln\Lambda$ ) between this tree and the mitochondrial likelihood tree. The significance of this difference was assessed by comparing it against a random distribution of  $-\ln\Lambda$  values generated by Monte Carlo simulation. To generate this distribution, the tree and parameter estimates that resulted from fitting the mitochondrial data to the combined data topology were used to generate 100 random data sets in Seq-Gen ver. 1.2.4 (Rambaut and Grassly, 1997). For each of these simulated datasets,  $-\ln\Lambda$  was calculated using the partial optimization method (*posPpud*) as described in Goldman et al. (2000) and implemented in PAUP\*. A similar procedure was followed to assess whether the nuclear data were consistent with the combined data tree.

## RESULTS

### Maximum Parsimony Analyses

Parsimony analysis of the six individual gene matrices yielded from 1 (*Sry*) to 15 (*Zp-3*) minimum length trees (Table 3); the strict consensus trees are shown in Figure 2. These trees differ in topology, degree of resolution, and amount of homoplasy. No two gene trees are exactly alike, but several clades are shared among gene trees. Five of the six support monophyly of the genus *Mus* with respect to the outgroup taxa *Mastomys*, *Hylomyscus*, and *Rattus* (not shown in Fig. 2); the sixth, *B2m*, include only one outgroup taxon. Five of the six support monophyly of the subgenus *Mus*; *12S* is not resolved in that part of the topology. Within the subgenus *Mus*, a clear division is apparent between the Palaearctic species, *M. musculus*, *M. macedonicus*, *M. spicilegus*, and *M. spretus* (all of which are restricted to the Palaearctic Region except *M. musculus*, which extends into the Oriental), and the strictly Asian species, *M. cervicolor*, *M. cookii*, and *M. caroli*. A monophyletic Palaearctic clade is supported by all six gene trees, and a monophyletic Asian clade is supported by all except the *Zp-3* tree, which is unresolved in that part of the topology. Within the Palaearctic clade, four of six gene trees unite the house mice (*musculus*, "*molossinus*," *castaneus*, and *domesticus*), and four of six support a sister-group relationship between the eastern Mediterranean short-tailed mouse, *M. macedonicus*, and the mound-builder, *M. spicilegus*.

Although there is considerable concordance among gene tree topologies, three

TABLE 3. Tree statistics for each data set analyzed separately and mapped on combined data tree. Tree lengths and consistency indices (CI) were calculated after removal of uninformative characters. MP = maximum parsimony; RI = retention index.

	Partition analyzed separately					Partition mapped on combined data tree		
	Number of informative characters	Number of MP trees	Tree length	CI	RI	Tree length	CI	RI
<i>Sry</i>	42	1	63	0.73	0.85	66	0.70	0.82
<i>Cyt b</i>	321	4	1103	0.43	0.44	1112	0.42	0.43
<i>12S</i>	83	2	188	0.57	0.59	193	0.55	0.57
<i>B2m</i>	47	2	87	0.69	0.77	93	0.65	0.72
<i>Zp-3</i>	76	15	131	0.73	0.76	139	0.69	0.70
<i>Tcp-1</i>	89	2	133	0.75	0.86	135	0.74	0.85
Mitochondrial	404	6	1297	0.45	0.46	1305	0.44	0.45
Nuclear	254	2	431	0.70	0.79	433	0.70	0.78
All combined	658	1	1738	0.51	0.56			

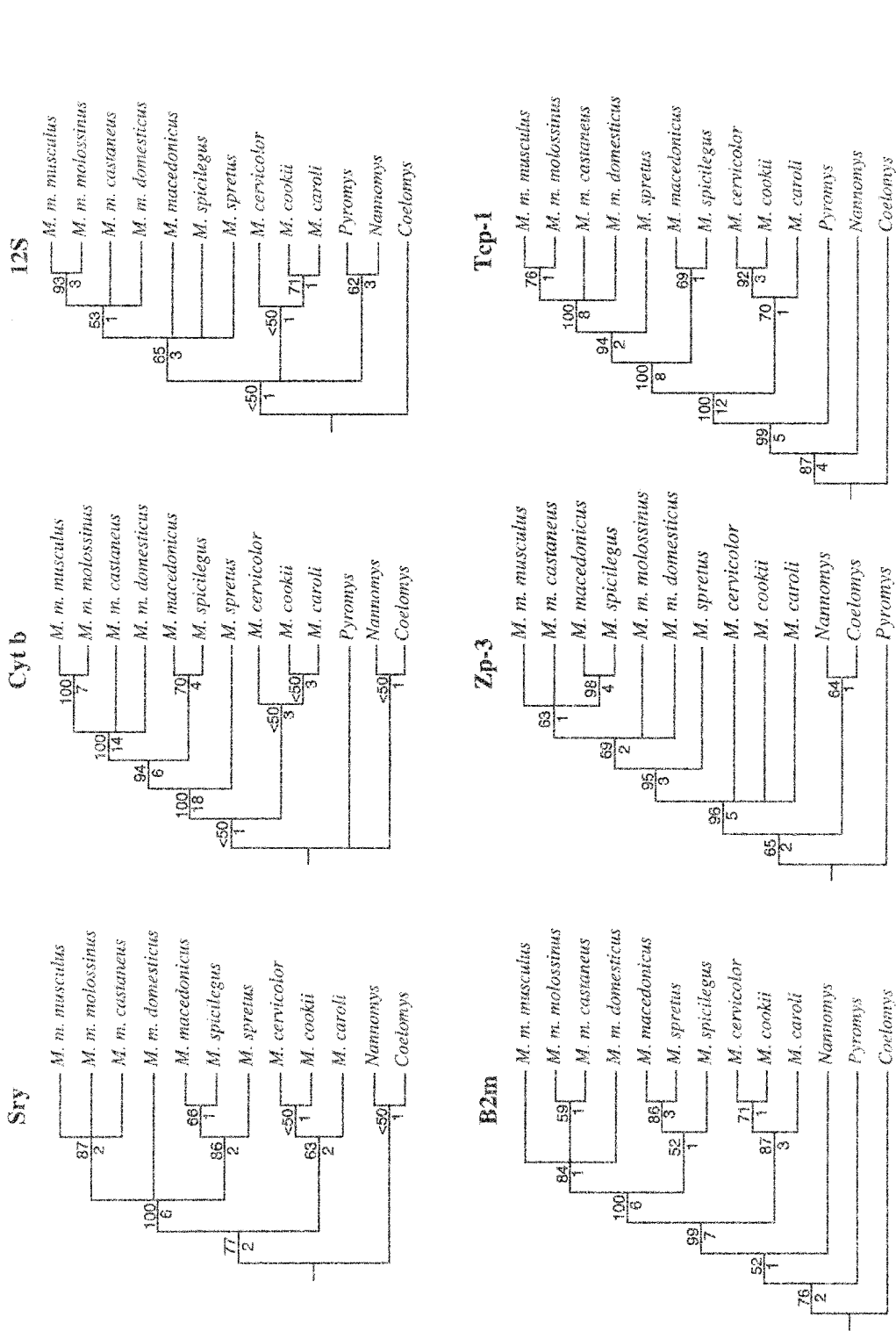


FIGURE 2. Results of individual parsimony analyses of the six gene partitions. Gene abbreviations are defined in Table 2 and associated tree statistics are provided in Table 3. The bootstrap proportion is shown above and Bremer value below each nonterminal branch. Trees were rooted with *Mastomys*, *Hylomyscus*, and *Rattus* (not shown).

regions of conflict are notable: (1) *M. spretus* is sometimes basal to the rest of the Palaearctic clade (*Cyt b* and *Zp-3*) and sometimes embedded within that clade (*Sry*, *B2m*, and *Tcp-1*); (2) some gene trees support a sister-group relationship between *M. cervicolor* and *M. cookii* (*Sry*, *B2m*, and *Tcp-1*) and others between *M. caroli* and *M. cookii* (*Cyt b* and *12S*); and (3) relationships among the four subgenera of *Mus* (*Mus*, *Pyromys*, *Coelomys*, and *Nannomys*) are inconsistent. The nodes defining these relationships are unresolved or poorly supported except in the *Tcp-1* tree, where support is strong for a sister-group relationship between *Mus* and *Pyromys* (bootstrap = 99, Bremer = 5) and moderate for the placement of *Nannomys* immediately basal to that clade (bootstrap = 87, Bremer = 4), leaving *Coelomys* as the most basal subgenus.

An ILD test including all genes, with the data set partitioned along the six gene boundaries, failed to reject the null hypothesis of data set homogeneity ( $P = 0.10$ ), as did tests of each gene individually against

the remainder of the data ( $P$ -values ranged from 0.11 to 0.85). Only 2 of the 15 pairwise comparisons between genes had uncorrected  $P$ -values of  $<0.01$  (the cutoff recommended by Cunningham, 1997). Both involved the *Zp-3* gene, which was significantly incongruent with *B2m* ( $P = 0.006$ ) and with *Tcp-1* ( $P = 0.001$ ). Because *Zp-3* was not significantly incongruent with any other gene in the sample, it is not clear how this apparent conflict could be objectively applied to partitioning the larger data set (see also Baker and DeSalle, 1997). Except for those two pairwise comparisons, none of the ILD tests based on linkage group or mode of inheritance resulted in significant incongruence. The lowest  $P$ -value obtained was for a partition separating mtDNA from the remaining data, represented by the four nuclear genes combined ( $P = 0.10$ ).

A comparison of phylogenies based on the mtDNA and nuclear partitions (Fig. 3) summarizes and highlights the local regions of disagreement that were apparent among

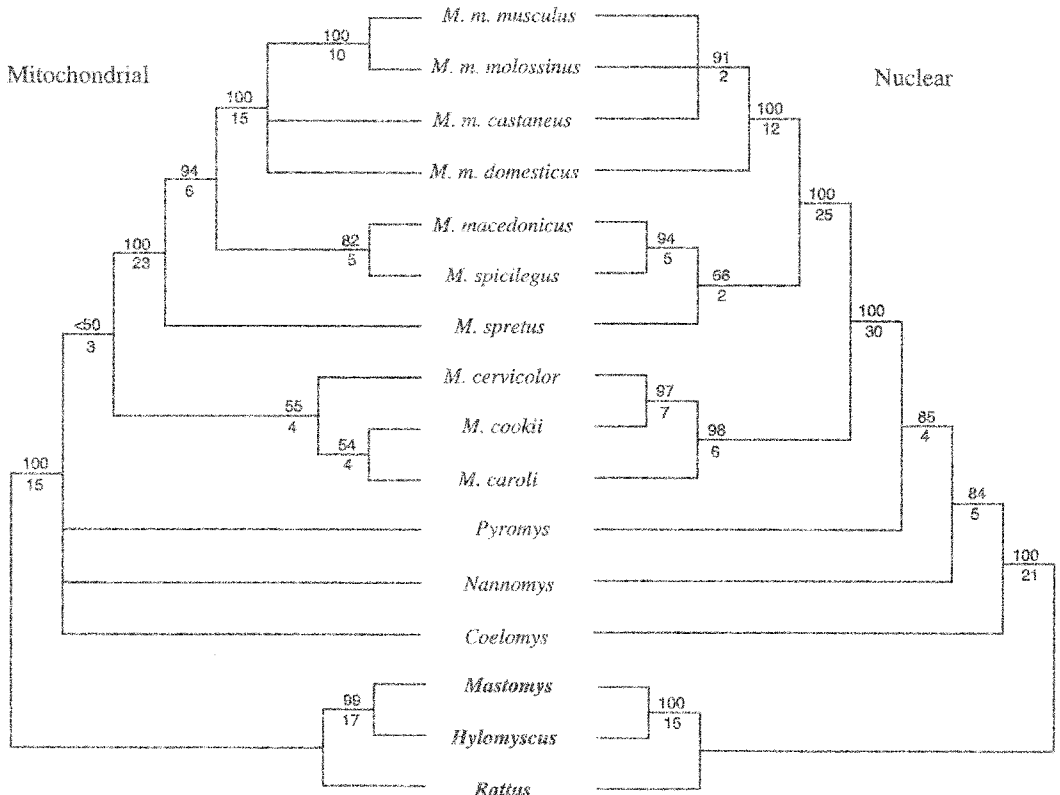


FIGURE 3. Results of parsimony analyses of combined mitochondrial data (*Cyt b* and *12S*) and combined nuclear data (*Sry*, *B2m*, *Zp-3*, and *Tcp-1*). Associated tree statistics are provided in Table 3. The bootstrap proportion is shown above and Bremer value below each nonterminal branch. Outgroup taxa are in bold.

individual gene trees. The mitochondrial tree places *M. spretus* at the base of the Palaearctic clade and supports a sister-group relationship between *M. cookii* and *M. caroli*, whereas the nuclear tree places *M. spretus* in a clade with *M. macedonicus* and *M. spicilegus* and unites *M. cookii* with *M. cervicolor*. Subgeneric relationships are fully resolved in the nuclear tree and completely unresolved in the mitochondrial tree. For each of the conflicting nodes, one of the partitions provides strong evidential support (as measured by bootstrap proportions and Bremer values) and the alternative is poorly supported.

A parsimony analysis of the combined data resulted in a single most-parsimonious tree of 1,738 steps, with a consistency index (excluding uninformative characters) of 0.51 and a retention index of 0.56 (Table 3, Fig. 4). This tree supports all of the clades that are shared by four or more of the individual gene trees, but with stronger support for those clades than was present in any individual gene tree. The three areas of conflict among gene trees are fully resolved in the combined data tree, but support values for those nodes are lower. *Mus spretus* is located at the base of the Palaearctic clade (bootstrap = 88; Bremer = 6); six additional steps are required to unite *M. spretus* with the *M. spicilegus*/*M. macedonicus* clade (the topology observed in the nuclear gene tree). The sister-group relationship between *M. cervicolor* and *M. cookii*, which was contraindicated by the mitochondrial tree, is recovered in only 65% of the bootstrap replicates and has a Bremer value of only 2. Two additional steps are required to unite *M. cookii* with *M. caroli*. Finally, subgeneric relationships are identical to those shown in the nuclear tree (Fig. 3), but with much lower support values (bootstrap recovery percentages drop from 85 to 70 for the clade uniting *Mus* with *Pyromys* and from 84 to 57 for the clade uniting those two subgenera with *Nannomys*).

#### Maximum Likelihood Analyses

Likelihood-ratio tests among competing nested models identified the GTR + I +  $\Gamma$  model with clock enforced as the best model for the *Cyt b* and 12S data sets, the two mitochondrial genes combined, and all six genes combined. The HKY +  $\Gamma$  model with clock enforced provided the best-fit model for each of the nuclear genes and those four genes

combined. A search for the most likely tree for each data set under the best-fit model yielded from 1 to 15 equally likely trees (Table 4).

The strict consensus trees resulting from maximum likelihood analysis (not shown) are either identical to their maximum parsimony counterparts (i.e., *Tcp-1*, nuclear genes combined, and all data combined), or very similar to those trees (i.e., *Sry*, *Cyt b*, 12S, *B2m*, *Zp-3*, and mitochondrial genes combined). All clades with >65% bootstrap support in the parsimony trees (Figs. 2–4) are also recovered when using maximum likelihood. For *Sry*, *Zp-3*, and the mitochondrial genes combined, the likelihood and parsimony topologies differ only in extent of resolution, the likelihood tree being more resolved in each case. However, for *Cyt b*, 12S, and *B2m*, the maximum likelihood trees conflict somewhat with those obtained using maximum parsimony.

In contrast to the *Cyt b* maximum parsimony tree (Fig. 2), the *Cyt b* maximum likelihood tree supports a sister-group relationship between *M. cervicolor* and *M. cookii*, with *M. caroli* as the basal member of the Asian clade; these nodes are recovered in 70% and 79% of maximum likelihood bootstrap replicates, respectively. In addition, the likelihood *Cyt b* tree resolves relationships among the subgenera as a grade, with *Coelomys* immediately basal to a monophyletic *Mus* (bootstrap = 53), *Pyromys* as the next taxon towards the base of the tree (bootstrap = 63), and *Nannomys* as the most basal subgenus.

The 12S and *B2m* maximum likelihood trees differ from their parsimony counterparts (Fig. 2) only with respect to subgeneric relationships. In the 12S maximum likelihood tree, *Coelomys* and *Nannomys* are sister taxa (bootstrap = 65), with *Pyromys* as the most basal taxon (bootstrap < 50); in the *B2m* maximum likelihood tree, *Nannomys* and *Pyromys* are sister taxa (bootstrap < 50), with *Coelomys* the most basal taxon (bootstrap < 50). As was the case among maximum parsimony gene trees, there is little agreement among maximum likelihood gene trees regarding the relative placement of subgenera.

Parametric likelihood-based tests indicate that both the mitochondrial data set and the nuclear data set can reject the combined data topology ( $-\ln\Lambda = 4.71$ ,  $P < 0.03$ ;  $-\ln\Lambda = 10.16$ ,  $P < 0.01$ , respectively). These results

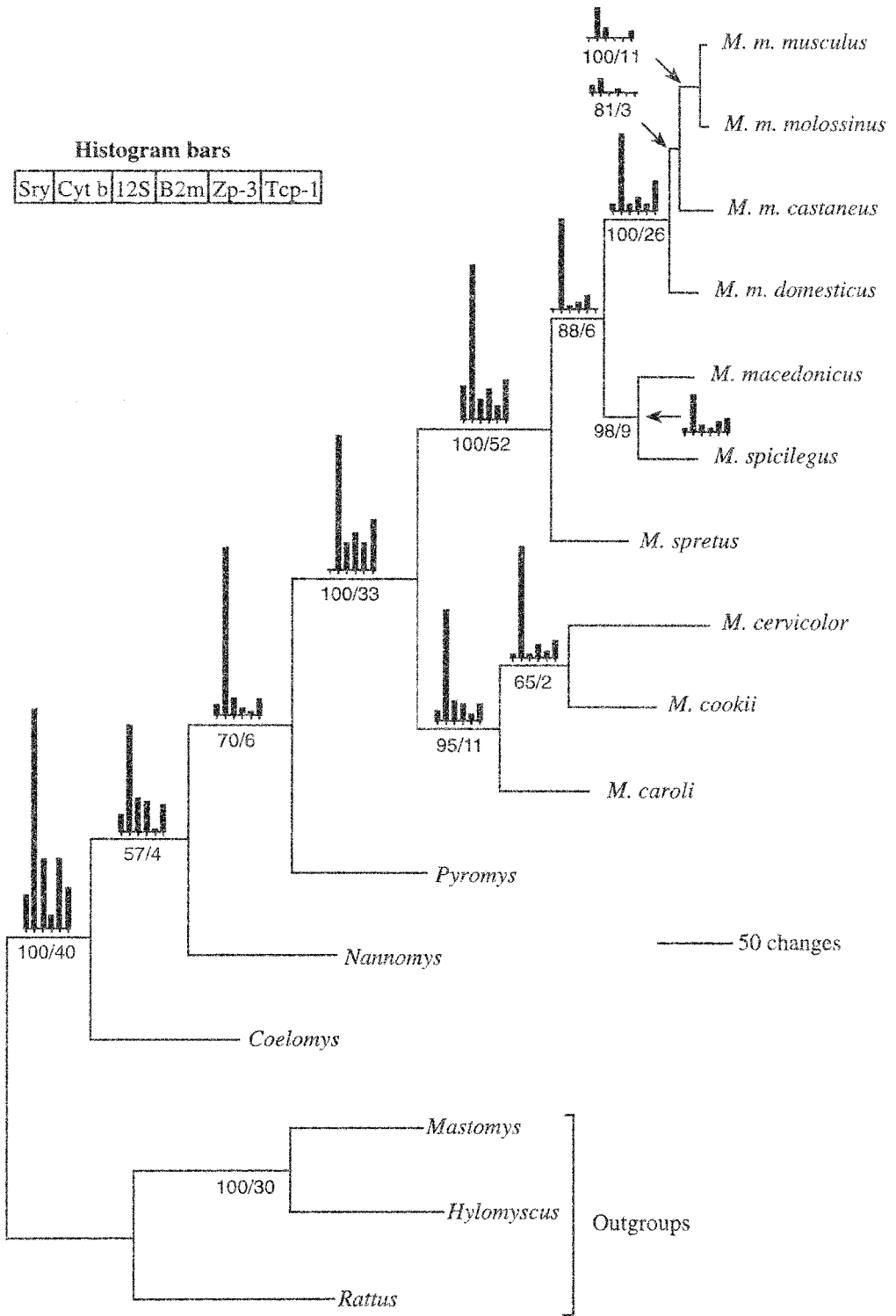


FIGURE 4. The single most-parsimonious tree from an analysis of the combined data. Associated tree statistics are provided in Table 3. The bootstrap proportion is shown to the left of the slash and Bremer value to the right under each nonterminal branch. The histograms above nonterminal branches depict the contribution of each gene to total branch length, based on ACCTRAN optimization in PAUP.

TABLE 4. Log-likelihood scores and parameter estimates under the best-fit model for each dataset. Methods for obtaining the best-fit model and parameter estimates are described in Methods. Where more than one tree was obtained, the average value of each parameter was calculated. ML = maximum likelihood; I = proportion of invariable sites; G =  $\Gamma$ -shape parameter.

	<i>Sry</i>	<i>Cyt b</i>	<i>12S</i>	<i>B2m</i>	<i>Zp-3</i>	<i>Tcp-1</i>	mtDNA	Nuclear	Combined
No. of ML trees	15	1	3	15	3	3	1	1	1
$-\ln\Lambda$	1501.71	6311.22	2615.43	1107.85	2718.16	3430.09	9065.04	8931.57	18422.19
Freq A	0.319	0.322	0.363	0.312	0.215	0.273	0.340	0.267	0.294
Freq C	0.242	0.307	0.206	0.280	0.285	0.168	0.262	0.229	0.239
Freq G	0.236	0.116	0.179	0.198	0.260	0.240	0.146	0.242	0.206
Freq T	0.203	0.255	0.252	0.211	0.240	0.319	0.252	0.263	0.261
Rate A-C	1.0	2.393	1.810	1.0	1.0	1.0	2.461	1.0	2.280
Rate A-G	2.603	9.463	7.917	1.03	2.178	2.491	8.049	2.056	5.212
Rate A-T	1.0	4.032	4.173	1.0	1.0	1.0	4.574	1.0	2.331
Rate C-G	1.0	0.071	0.000	1.0	1.0	1.0	0.053	1.0	1.068
Rate C-T	2.603	39.439	20.750	1.0372	2.178	2.491	40.336	2.056	16.001
Rate G-T	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
I	0.0	0.566	0.691	0.0	0.0	0.0	0.625	0.0	0.449
G	1.421	1.711	0.990	0.455	0.443	0.776	1.168	0.508	0.540
Clocklike?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

contrast with the results of the parsimony-based ILD test, which found no significant character incongruence between mitochondrial and nuclear partitions ( $P = 0.10$ ). How can we reconcile significant topological incongruence, as indicated by the parametric bootstrapping approach, with nonsignificant character incongruence, as indicated by the ILD test? One possibility is that the ILD test is not sensitive enough to detect incongruence between these data sets. Dolphin et al. (2000) recently criticized the utility of the ILD test for examining incongruence among data sets. However, their main criticism is that the test is overly sensitive; that is, it tends to reject the null hypothesis of data set homogeneity when data sets are in fact congruent (see also Barker and Lutzoni, in press). Conversely, the parametric likelihood-based tests, with their implicit reliance on models of sequence evolution, may be overly sensitive in detecting topological incongruence between trees based on different data sets. These tests have only recently been developed and their sensitivity under various conditions will require further evaluation (Goldman et al., 2000).

We maintain that the best estimate of phylogeny results from combining data from multiple sources and present our combined data phylogeny (Fig. 4) as the best working hypothesis of relationships among these taxa. Nonetheless, we present the phylogenies resulting from separate analyses of individual data partitions (Figs. 2, 3) as a convenient means to examine regions of conflict

and agreement among these different sources of data.

## DISCUSSION

Our combined data phylogeny (Fig. 4) is completely concordant with the "synthetic tree" presented in Boursot et al. (1993; see Fig. 1), despite the fact that none of the gene sequences included in our analysis were considered in constructing the Boursot et al. tree. The only difference between that tree and ours is in the extent of resolution. In the Boursot et al. tree, areas of conflict among data sets are depicted as polytomies. Although our combined data tree is fully resolved, the analysis revealed some of the same conflicts among data sets, here evidenced by the individual gene trees (Fig. 2) and the examination of mitochondrial versus nuclear topologies (Fig. 3).

### *Monophyly of the Subgenus Mus and Genus Mus*

The most recent taxonomic treatments of the genus *Mus* (Marshall, 1977, 1981, 1998) recognize four subgenera: *Mus*, *Pyromys*, *Coelomys*, and *Nannomys*. Monophyly of the subgenus *Mus* is supported by a variety of data sets, including allozymes (Bonhomme et al., 1984), scnDNA hybridization data (She et al., 1990), and nuclear DNA sequences (Jouvin-Marche et al., 1988). Our results corroborate this finding. None of the gene trees contradict monophyly of this subgenus, and

in the combined data analysis (Fig. 4), a monophyletic subgenus *Mus* was found in 100% of bootstrap replicates and received a Bremer support value of 33.

Monophyly of the genus *Mus* is not as firmly established. Most of the morphological character states that Marshall (1977) listed as uniting the four subgenera are plesiomorphic and hence do not provide evidence for monophyly (as discussed in Thaler, 1986). In addition, the four subgenera are quite divergent genetically, in some instances as divergent from one another as from other murid genera (Bonhomme et al., 1985). In his 1986 paper, Bonhomme concluded from allozyme data that the four subgenera of *Mus* were sufficiently divergent to merit generic status (an opinion shared by Thaler, 1986); however, the historical relationships among these four taxa relative to other genera could not be resolved without making restrictive assumptions about protein evolution (Bonhomme et al., 1985). In our analysis, three additional murine genera, *Mastomys*, *Hylomyscus*, and *Rattus*, were used as outgroups for rooting phylogenies. We ran all phylogenetic analyses with the ingroup and outgroup taxa unconstrained. Therefore, these three genera provide a partial test of monophyly of the genus *Mus*, in so far as it is represented by this limited taxon sampling. Our analysis strongly supports monophyly of the genus *Mus* with respect to *Mastomys*, *Hylomyscus*, and *Rattus* (Fig. 4; bootstrap = 100, Bremer = 40). Clearly, a broader sampling of *Mus* species and murine genera will be needed to provide a rigorous test of *Mus* monophyly, but for now there is no phylogenetic justification for excluding *Pyromys*, *Coelomys*, or *Nannomys* from the genus *Mus*.

#### *Relationships within the Subgenus Mus*

We uncovered three major clades within the subgenus *Mus* (Fig. 4): a house mouse clade that includes *M. m. musculus*, "*M. m. molossinus*," *M. m. castaneus*, and *M. m. domesticus*; a Palaearctic clade that includes those four taxa plus *M. macedonicus*, *M. spicilegus*, and *M. spretus*; and an Asian clade that includes *M. cervicolor*, *M. cookii*, and *M. caroli*.

*The house mouse clade.*—Our results agree with several earlier studies (e.g., Sage, 1981; Ferris et al., 1983b; Bonhomme et al., 1984; Tucker et al., 1989; She et al., 1990; Suzuki

and Kurihara, 1994; Prager et al., 1996) in supporting a clade of commensal house mice (Fig. 4). Within the house mouse clade, the hybrid Japanese mouse, "*M. m. molossinus*," is united with one of its parent subspecies, *M. m. musculus*. Support for this relationship comes almost entirely from the two mitochondrial genes, *Cyt b* and *12S* (the four nuclear genes exhibit little or no variation in this part of the phylogeny). Our "*M. m. molossinus*" mitochondrial sequences are derived from a common laboratory strain (MOLO) founded by mice from northern Kyushu. The sister-group relationship between "*M. m. molossinus*" and *M. m. musculus* in our phylogeny is consistent with the finding of Yonekawa et al. (1988) that the "*M. m. molossinus*" mtDNA from this region is derived predominantly from *M. m. musculus*.

Our data also support the basal placement of *M. m. domesticus* within the house mouse clade. This conflicts with Moriwaki (1994) who argues (based on C-banding patterns and the distribution of various alleles) for a sister-group relationship between *M. m. castaneus* and *M. m. domesticus*, with *M. m. musculus* basal. The latter topology is also supported by allozyme data (Bonhomme, 1986) and nuclear RFLPs (Santos et al., 1993; Suzuki and Kurihara, 1994). However, various other molecular studies support the topology we found, that is, a sister-group relationship between *M. m. musculus* and *M. m. castaneus*, to the exclusion of *M. m. domesticus* (e.g., allozymes, Sage, 1981; mtDNA RFLPs, Yonekawa et al., 1988; Y-specific RFLPs, Tucker et al., 1989; and mtDNA sequences, Prager et al., 1996). Given the recent origin of *M. musculus* (estimated at 0.9 million years ago; Bonhomme et al., 1994), the geographic overlap of subspecies, their passive transport by humans, and the fact that the subspecies hybridize when they come into contact, it is not surprising that different data sets are incongruent with respect to historical relationships among these taxa (as discussed in Bonhomme, 1986).

*The Palaearctic clade.*—A Palaearctic clade that includes the house mouse plus *M. macedonicus*, *M. spicilegus*, and *M. spretus*, is recovered in all six individual gene trees (Fig. 2) and is strongly supported by the combined data (Fig. 4). The species in this clade have been the subject of numerous studies; their monophyly is supported not only by our gene sequence data, but also by allozyme

data (Bonhomme et al., 1984) and scnDNA hybridization data (She et al., 1990). Within the Palaearctic clade, the sister-group relationship between the eastern Mediterranean short-tailed mouse (*M. macedonicus*) and the mound-building mouse (*M. spicilegus*) is also well established (Fig. 4; and Sage, 1981; Bonhomme et al., 1984; She et al., 1990; Prager et al., 1996). The failure of 16S sequence data (Fort et al., 1984) to recover this grouping may reflect the small number of informative characters in that data set (15, excluding intraspecific synapomorphies) rather than different histories of mitochondrial versus nuclear DNAs, as has been suggested (Fort et al., 1984; Bonhomme, 1992). In our study, the sister-group relationship between *M. macedonicus* and *M. spicilegus* is supported by both mitochondrial and nuclear markers (Fig. 3).

While there is little argument that *M. macedonicus* and *M. spicilegus* are sister taxa, the correct placement of the western Mediterranean short-tailed mouse, *M. spretus*, has been difficult to determine. Different data sets disagree with respect to where *M. spretus* belongs, sometimes placing it in a clade with *M. macedonicus* and *M. spicilegus* (e.g., allozymes, Sage, 1981; nuclear DNA sequences, Morita et al., 1992) and sometimes as sister to the rest of the Palaearctic clade (e.g., allozymes, Bonhomme et al., 1984; microsatellite DNA, Dod et al., 1989; RFLPs of nuclear rDNA spacer regions, Suzuki and Kurihara, 1994; and mtDNA sequences, Prager et al., 1996). Our combined data analysis (Fig. 4) is in accord with the latter hypothesis; that is, it supports a clade that includes *M. musculus*, *M. macedonicus*, and *M. spicilegus* to the exclusion of *M. spretus*. However, the alternative (*M. spretus* in a clade with *M. spicilegus* and *M. macedonicus*) is favored by the combined nuclear DNA data (Fig. 3). Indeed, almost all of the support for the basal position of *M. spretus* in the combined data tree comes from a single mitochondrial gene, *Cyt b*. Removal of *Cyt b* from the combined data set causes the Palaearctic clade to collapse into an unresolved trichotomy (as in the Boursot et al. [1993] see Fig. 1). The inability of this large data set to place *M. spretus* with confidence may reflect rapid speciation of the Palaearctic lineages (as suggested in She et al., 1990; and Bonhomme, 1992). The branch connecting *M. spretus* to the rest of the Palaearctic clade is relatively short (Fig. 4),

and of the six genes we considered, apparently only *Cyt b* is evolving rapidly enough to contribute more than a few informative characters toward placement of this taxon.

*The Asian clade.*—The Asian species, *M. cervicolor*, *M. cookii*, and *M. caroli* are not as well studied as the Palaearctic species of *Mus*. Geographic ranges are incompletely known, and geographic variation within species has not been assessed. A few individuals collected in the 1980s provided the source for most laboratory stocks used today. Our data strongly support monophyly of this group; these species are united in five of the six gene trees (Fig. 2), and the support indices in the combined data tree are high (Fig. 4; bootstrap = 95, Bremer = 11). Although a close relationship among these three species was suggested by allozyme and mtDNA data (She et al., 1990), until our study, the hypothesis that they make up a monophyletic group (as portrayed in Fig. 1), was based entirely on scnDNA hybridization data (She et al., 1990).

Phylogenetic relationships among *M. cervicolor*, *M. cookii*, and *M. caroli* are not as clear. The scnDNA hybridization data (She et al., 1990) were unable to resolve the branching pattern among these species; hence, they form a polytomy in the Boursot et al. (1993) tree (Fig. 1). In our study, the combined nuclear data strongly support an *M. cervicolor/M. cookii* clade, whereas the combined mitochondrial data provide weak support for a sister-group relationship between *M. cookii* and *M. caroli*. The nuclear topology is retained in the combined data tree (Fig. 4), but support indices are lower than in the nuclear tree, reflecting the conflict introduced by the mitochondrial genes.

It is tempting to hypothesize that the disagreement between mitochondrial and nuclear phylogenies with respect to these taxa reflects different histories of the mitochondrial and nuclear genomes, resulting, for example, from mitochondrial introgression, no compelling evidence supports this hypothesis. The sister-group relationship between *M. cookii* and *M. caroli* is not well supported by the mtDNA data; only two additional steps are required to force a sister-group relationship between *M. cervicolor* and *M. cookii*. Moreover, although the clade joining *M. cervicolor* and *M. cookii* is not present in either the *Cyt b* or 12S parsimony gene trees, it is recovered in the *Cyt b* maximum likelihood tree. Although we accept



the combined data topology as the best working hypothesis for relationships among these species, we note that this part of the topology may be particularly subject to being overthrown by additional data. Of the three possible relationships among these species, only the *M. cervicolor*/*M. caroli* sister group is resoundingly contradicted by our data; an additional 14 steps are required to constrain a relationship between those two taxa.

#### *Relationships Among the Subgenera*

Attempts to unravel relationships among the four subgenera of *Mus* (*Mus*, *Pyromys*, *Coelomys*, and *Nannomys*) have met with limited success. Either the branches joining these taxa are short and unstable (Jouvin-Marche et al., 1988) or the relationships are unresolved (She et al., 1990; Catzefflis and Denys, 1992; Sourrouille et al., 1995). In the Boursot et al. (1993) summary tree (Fig. 1), a *Mus*/*Pyromys* clade forms a trichotomy with *Coelomys* and *Nannomys* (based on scnDNA hybridization data from She et al., 1990). The uncertainty surrounding subgeneric relationships is particularly troublesome because many characters of interest (e.g., many nuclear genes) evolve too slowly to be effectively examined from an evolutionary perspective within the subgenus *Mus*. A robust phylogeny including all four subgenera would provide a framework to examine the evolution of those characters.

Our individual gene trees are not in agreement with respect to relationships among the subgenera of *Mus* (Fig. 2). Except in the *Tcp-1* tree, either subgeneric relationships are unresolved or the topology in this region of the tree is poorly supported. *Tcp-1* strongly favors a sister-group relationship between *Pyromys* and *Mus* (bootstrap = 99, Bremer = 5), with *Nannomys* sister to that clade (bootstrap = 87, Bremer = 4) and *Coelomys* being the most basal subgenus. This topology is retained in the combined data tree (Fig. 4), but the support indices for those clades are much lower than in the *Tcp-1* tree.

What is the explanation for the absence of well-supported branches in this part of the combined data tree? The conflict does not seem to fall along gene, linkage group, or mitochondrial/nuclear divisions. Rather, phylogenetic signal in this region of the tree apparently is diminished by a high amount of homoplasy, introduced especially by the

*Cyt b* sequences (which contribute the largest number of informative characters to resolving these relationships). Of the six genes, only *Tcp-1* produces a clear signal with respect to phylogenetic relationships among subgenera. If *Tcp-1* is removed from the combined data analysis, these relationships are completely unresolved. Additional sampling within *Pyromys*, *Coelomys*, and *Nannomys* might improve the signal from *Cyt b* (and the other genes) by subdividing the relatively long terminal branches representing these three lineages. Although the subgenus *Mus* is well sampled in our study (i.e., seven of nine species are included), we have included only one species from each of the other subgenera. The relatively large subgenus *Nannomys* provides some of the thorniest taxonomic problems and is perhaps the richest source of material for future studies. Although chromosomal and biochemical surveys support the monophyly of *Nannomys* (Van Rompaey et al., 1984; Catzefflis and Denys, 1992; Bonhomme et al., 1994; Sourrouille et al., 1995), many of the species in this subgenus are not clearly defined and only a few of the 19 species recognized by Musser and Carleton (1993) have been included in any phylogenetic study.

#### *Conclusions*

We present our combined data phylogeny as the best hypothesis of *Mus* relationships to date. It is based on a large sample of nucleotide sites, representing various parts of the genome, and is concordant with earlier studies based on other kinds of data (e.g., allozymes and scnDNA hybridization). Although there is some disagreement among genes with respect to the placement of *M. spretus*, relationships among Asian taxa in the subgenus *Mus*, and relationships among the subgenera of *Mus*, we find no convincing evidence that this conflict is the result of different gene or genome histories. Rather, these differences appear to result from a localized homoplasy in one partition that obscures phylogenetic signal from another. Ultimately, the placement of *M. spretus* and relationships among species in the Asian clade may prove difficult to determine. We have sampled taxa densely in these regions of the phylogeny without obtaining a stable topology, which may reflect a rapid radiation of these clades (as suggested by She et al.,

1990; and Bonhomme, 1992). In contrast, a well-supported hypothesis of relationships among the subgenera of *Mus* might be obtained by including additional taxa from the subgenera *Pyromys*, *Coelomys*, and *Nannomys*.

#### ACKNOWLEDGMENTS

We thank M. Potter (National Cancer Institute Contract N01-CB-71085), E. Eicher (The Jackson Laboratories), and R. D. Sage (University of Missouri), for providing frozen tissue or DNA from various stocks and strains of wild mice, and B. Patterson and L. Heaney (The Field Museum of Natural History), for providing frozen organs from wild mice collected in the field. Michael Carleton, Keith Barker, Dick Olmstead, Jack Sullivan, and two anonymous reviewers provided numerous suggestions that substantially improved the final paper. S. A. J. was supported by a Kalbfleisch Postdoctoral Research Fellowship from the American Museum of Natural History and a Scholarly Studies Fellowship from the Smithsonian Institution during the analysis and writing stages of this project. This research was supported by National Science Foundation Grant DEB-9209950 (to P. K. T).

#### REFERENCES

- AUFFRAY, J.-C., E. TCHERNOV, F. BONHOMME, G. HETH, S. SIMSON, AND E. NEVO. 1990. Presence and ecological distribution of *Mus "spretoides"* and *Mus musculus domesticus* in Israel. Circum-Mediterranean vicariance in the genus *Mus*. *Z. Saug.* 55:1–10.
- AUFFRAY, J.-C., E. TCHERNOV, AND E. NEVO. 1988. Origine du commensalisme de la souris domestique (*Mus musculus domesticus*) vis-à-vis de l'homme. *C. R. Acad. Sci. Paris* 307:517–522.
- BAKER, R. H., AND R. DE SALLE. 1997. Multiple sources of character information and the phylogeny of Hawaiian Drosophilids. *Syst. Biol.* 46:654–673.
- BARKER, F. K., AND F. M. LUTZONI. 2002. The utility of the incongruence length difference test. *Syst. Biol.* (in press).
- BERRY, R. J. 1995. An animal weed—Man's relationship with the house mouse. Pages 664–665 in *The encyclopedia of mammals* (D. Macdonald, ed.). Facts on File, New York.
- BHAT, S. K., AND A. SUJATHA. 1991. Population structure and breeding season of the field mouse, *Mus booduga booduga* Gray. *Mammalia* 55:389–396.
- BJORKMAN, P. J., M. A. SAPER, B. SAMRAOUL, W. S. BENNETT, J. L. STROMINGER, AND D. C. WILEY. 1987. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* 329:506–512.
- BLEIL, J. D., AND P. M. WASSARMAN. 1980a. Structure and function of the zona pellucida: Identification and characterization of the proteins of the mouse oocyte's zona pellucida. *Dev. Biol.* 76:185–202.
- BLEIL, J. D., AND P. M. WASSARMAN. 1980b. Mammalian sperm-egg interaction: Identification of a glycoprotein in mouse egg zona pellucida possessing receptor activity for sperm. *Cell* 20:873–882.
- BONHOMME, F. 1986. Evolutionary relationships in the genus *Mus*. *Curr. Top. Microb. Immunol.* 127:19–34.
- BONHOMME, F. 1992. Genetic diversity and evolution in the genus *Mus*. Pages 41–56 in *Techniques for genetic analysis of brain and behavior* (D. Goldwits, D. Wahlsten, and R. E. Wimer, eds.). Elsevier Science Publishers, Amsterdam, The Netherlands.
- BONHOMME, F., R. ANAND, D. DARVICHE, W. DIN, AND P. BOURSOT. 1994. The house mouse as a ring species? Pages 13–23 in *Genetics in wild mice* (K. Moriwiki, T. Shiroishi, and H. Yonekawa, eds.). Japan Scientific Societies Press, Tokyo.
- BONHOMME, F., J. BRITTON-DAVIDIAN, L. THALER, AND C. TRIANTAPHYLIDIS. 1978. Sur l'existence en Europe de quatre groupes de Souris (genre *Mus* L.) du rang espèce et semi-espèce, démontrée par la génétique biochimique. *C. R. Acad. Sci. Paris* 287:631–633.
- BONHOMME, F., J. CATALAN, J. BRITTON-DAVIDIAN, V. M. CHAPMAN, K. MORIWAKI, E. NEVO, AND L. THALER. 1984. Biochemical diversity and evolution in the genus *Mus*. *Biochem. Genet.* 22:275–303.
- BONHOMME, F., J. CATALAN, S. GERASIMOV, P. ORSINI, AND L. THALER. 1983. Le complexe d'espèces du genre *Mus* en Europe Centrale et Orientale. *Z. Saug.* 48: 78–85.
- BONHOMME, F., AND J. GUENET. 1995. The laboratory mouse and its wild relatives. Pages 1577–1596 in *Inbred and genetically defined strains of laboratory animals* (P. L. Altman and D. D. Katz, eds.). FASEB, Bethesda, Maryland.
- BONHOMME, F., D. ISKANDAR, L. THALER, AND F. PETTER. 1985. Electromorphs and phylogeny in muroid rodents. Pages 671–683 in *Evolutionary relationships among rodents. A multidisciplinary analysis* (W. P. Luckett and J. Hartenberger, eds.). Plenum Press, New York.
- BONHOMME, F., N. MIYASHITA, P. BOURSOT, J. CATALAN, AND K. MORIWAKI. 1989. Genetical variation and polyphyletic origin in Japanese *Mus musculus*. *Heredity* 63:299–308.
- 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., F. BONHOMME, J. CATALAN, AND K. MORIWAKI. 1989. Variation of a Y chromosome repeated sequence across subspecies of *Mus musculus*. *Heredity* 63:289–297.
- BOURSOT, P., W. DIN, R. ANAND, D. DARVICHE, B. DOD, F. VON DEIMLING, G. P. TALWAR, AND F. BONHOMME. 1996. Origin and radiation of the house mouse: Mitochondrial DNA phylogeny. *J. Evol. Biol.* 9:391–415.
- BREMER, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* 42:795–803.
- BREMER, K. 1994. Branch support and tree stability. *Cladistics* 10:295–304.
- BRITTON, J., AND L. THALER. 1978. Evidence for the presence of two sympatric species of mice (genus *Mus* L.) in southern France based on biochemical genetics. *Biochem. Genet.* 16:213–225.
- BULL, J. J., J. P. HUELSENBECK, D. W. CUNNINGHAM, D. L. SWOFFORD, AND P. J. WADDELL. 1993. Partitioning and combining data in phylogenetic analysis. *Syst. Biol.* 42:384–397.
- BUSH, R. M., AND K. PAIGEN. 1992. Evolution of  $\beta$ -glucuronidase regulation in the genus *Mus*. *Evolution* 46:1–15.
- CARPENTER, J. M. 1996. Uninformative bootstrapping. *Cladistics* 12:177–181.
- CATZEFELIS, F. M., AND C. DENYS. 1992. The African *Nannomys* (Muridae): An early offshoot from the *Mus*

- lineage—Evidence from scnDNA hybridization experiments and compared morphology. *Isr. J. Zool.* 38:219–231.
- CHANDRAHAS, R. K. 1974. Ecology of the brown spiny mouse *Mus P. Platythrix* (Bennett) and the Indian field mouse *Mus b. Booduga* (Gray). *Indian J. Med. Res.* 62:264–280.
- CHEVRET, P., L. GRANJON, J. DUPLANTIER, C. DENYS, AND F. CATZEFELIS. 1994. Molecular phylogeny of the *Praomys* complex (Rodentia: Murinae): A study based on DNA/DNA hybridization experiments. *Zool. J. Linn. Soc.* 112:425–442.
- CHOU, C.-W., P.-F. LEE, K.-H. LU, AND H.-T. YU. 1998. A population study of house mice (*Mus musculus castaneus*) inhabiting rice graineries in Taiwan. *Zool. Stud.* 37:201–212.
- CUNNINGHAM, C. W. 1997. Can three incongruence tests predict when data should be combined? *Mol. Biol. Evol.* 14:733–740.
- DARVICHE, D., AND P. ORSINI. 1982. Critères de différenciation morphologique et biométrique de deux espèces de souris sympatriques: *Mus spretus* et *Mus musculus domesticus*. *Mammalia* 46:205–217.
- DESALLE, R., AND A. BROWER. 1997. Process partitions, congruence, and the independence of characters: Inferring relationships among closely related Hawaiian *Drosophila* from multiple gene regions. *Syst. Biol.* 46:751–764.
- DE QUEIROZ, A. 1993. For consensus (sometimes). *Syst. Biol.* 42:368–372.
- DE QUEIROZ, A., M. J. DONOGHUE, AND J. KIM. 1995. Separate versus combined analysis of phylogenetic evidence. *Annu. Rev. Syst.* 26:657–681.
- DIN, W., R. ANAND, P. BOURSOT, D. DARVICHE, B. DOD, E. JOUVIN-MARCHE, A. ORTH, G. P. TALWAR, P.-A. CAZENAVE, AND F. BONHOMME. 1996. Origin and radiation of the house mouse: Clues from nuclear genes. *J. Evol. Biol.* 9:519–539.
- DOD, B., E. MOTTEZ, E. DESMARAIS, F. BONHOMME, AND G. ROIZES. 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.
- DOLPHIN, K., R. BELSHAW, C. D. L. ORME, AND D. L. J. QUICKE. 2000. Noise and incongruence: Interpreting results of the incongruence length difference test. *Mol. Phylogenet. Evol.* 17:401–406.
- DONOGHUE, M. J., R. G. OLMSTEAD, J. F. SMITH, AND J. D. PALMER. 1992. Phylogenetic relationships of Dipsacales based on *rbcL* sequences. *Ann. Mo. Bot. Gard.* 79:333–345.
- EERNISSE, D. J., AND A. G. KLUGE. 1993. Taxonomic congruence versus total evidence, and the phylogeny of amniotes inferred from fossils, molecules, and morphology. *Mol. Biol. Evol.* 10:1170–1195.
- FARRIS, J. S., M. KÄLLERSJÖ, A. G. KLUGE, AND C. BULT. 1994. Testing significance of incongruence. *Cladistics* 10:315–319.
- FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* 17:368–372.
- FELSENSTEIN, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783–791.
- FERRIS, S. D., R. D. SAGE, C.-M. HUANG, J. T. NIELSEN, U. RITTE, AND A. C. WILSON. 1983a. Flow of mitochondrial DNA across a species boundary. *Proc. Natl. Acad. Sci. USA* 80:2290–2294.
- FERRIS, S. D., R. D. SAGE, E. M. PRAGER, U. RITTE, AND A. WILSON. 1983b. Mitochondrial DNA evolution in mice. *Genetics* 105:681–721.
- FORT, P., F. BONHOMME, P. DARLU, M. PIECHACZYK, P. JEANTEUR, AND L. THALER. 1984. Clonal divergence of mitochondrial DNA versus population evolution of nuclear genome. *Evol. Theory* 7:81–90.
- FRISMAN, L. V., K. V. KOROBITSINA, L. V. YAKIMENKO, F. M. BOKSHEIN, 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.
- GARAGNA, S., C. A. REDI, E. CAPANNA, N. ANDAYANI, R. M. ALFANO, P. DOI, AND G. VIALE. 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, AND V. MIHAILOVA. 1990. Morphometric stepwise discriminant analysis of the five genetically determined European taxa of the genus *Mus*. *Biol. J. Linn. Soc.* 41:47–64.
- GOLDMAN, N. 1993. Simple diagnostic statistical tests of models for DNA substitution. *J. Mol. Evol.* 37:650–661.
- GOLDMAN, N., J. P. ANDERSON, AND A. G. RODRIGO. 2000. Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.* 49:652–670.
- GUBBAY, J., J. COLLIGNON, P. KOOPMAN, B. CAPEL, A. ECONOMOU, A. MUNSTERBERG, N. VIVIAN, P. GOODFELLOW, AND R. LOVELL-BADGE. 1990. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* 346:245–250.
- HAMMER, M. F., AND A. C. WILSON. 1987. Regulatory and structural genes for lysozymes of mice. *Genetics* 115:521–533.
- HASEGAWA, M., H. KISHINO, AND T. A. YANO. 1985. Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22:160–174.
- HERMEL, E., P. J. ROBINSON, J. X. SHE, AND K. F. LINDAHL. 1993. Sequence divergence of *B2m* alleles of wild *Mus musculus* and *Mus spretus* implies positive selection. *Immunogenetics* 38:106–116.
- HUELSENBECK, J. P., J. J. BULL, AND C. C. CUNNINGHAM. 1996a. Combining data in phylogenetic analysis. *Trends Ecol. Evol.* 11:152–158.
- HUELSENBECK, J. P., D. M. HILLIS, AND R. JONES. 1996b. Parametric bootstrapping in molecular phylogenetics: Applications and performance. Pages 519–545 in *Molecular zoology: Advances, strategies, and protocols* (J. D. Ferraris and S. R. Palumbi, eds.). Wiley-Liss, New York.
- HUELSENBECK, J. P., AND B. RANNALA. 1997. Phylogenetic methods come of age: Testing hypotheses in an evolutionary context. *Science* 276:227–232.
- HUNT, W. G., AND R. K. SELANDER. 1973. Biochemical genetics of hybridisation in European house mice. *Heredity* 31:11–33.
- JANSA, S. A., S. M. GOODMAN, AND P. K. TUCKER. 1999. Molecular phylogeny and biogeography of the native rodents of Madagascar (Muridae: Nesomyinae): A test of the single-origin hypothesis. *Cladistics* 15:253–270.
- JENKINS, N. A., N. G. COPELAND, B. A. TAYLOR, AND B. K. LEE. 1982. Organization, distribution and stability of endogenous ecotropic murine leukemia virus DNA sequence in chromosomes of *Mus musculus*. *J. Virol.* 43:26–36.
- JOUVIN-MARCHE, E., A. CUDDIHY, S. BUTLER, J. N. HANSEN, W. M. FITCH, AND S. RUDIKOFF. 1988.

- Modern evolution of a single-copy gene: The immunoglobulin C<sub>k</sub> locus in wild mice. *Mol. Biol. Evol.* 5:500–511.
- JUKES, T. H., AND C. R. CANTOR. 1969. Evolution of protein molecules. Pages 21–132 in *Mammalian protein metabolism* (H. N. Munro, ed.). Academic Press, New York.
- KARN, R. C., AND M. W. NACHMAN. 1999. Reduced nucleotide variability at an androgen-binding protein locus (Abpa) in house mouse: Evidence for positive natural selection. *Mol. Biol. Evol.* 16:1192–1197.
- KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16:111–120.
- KIMURA, M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. *Proc. Natl. Acad. Sci. USA* 78:454–458.
- KINLOCH, R. A., AND P. M. WASSARMAN. 1989. Nucleotide sequence of the gene encoding zona pellucida glycoprotein ZP3—The mouse sperm receptor. *Nucleic Acids Res.* 17:2861–2863.
- KLUGE, A. G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Syst. Zool.* 38:7–25.
- KLUGE, A. G., AND A. J. WOLF. 1993. Cladistics: What's in a word? *Cladistics* 9:183–199.
- KOCHER, T. D., W. K. THOMAS, A. MEYER, S. V. EDWARDS, S. PAABO, F. X. VILLABLANCA, AND A. C. WILSON. 1989. Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci. USA* 86:6196–6200.
- LANYON, S. M. 1993. Phylogenetic frameworks: Towards a firmer foundation for the comparative approach. *Biol. J. Linn. Soc.* 49:45–61.
- LUNDRIGAN, B., AND P. K. TUCKER. 1994. Tracing paternal ancestry in mice, using the Y-linked, sex-determining locus, *Sry*. *Mol. Biol. Evol.* 11:483–492.
- LYON, M. F. 1991. The genetic basis of transmission ratio distortion and male sterility due to the t complex. *Am. Nat.* 147:349–358.
- MADDISON, W. P., AND D. R. MADDISON. 1992. MacClade, version 3.01. Sinauer Associates, Sunderland, Massachusetts.
- MARSHALL, J. T. 1977. A synopsis of Asian species of *Mus* (Rodentia, Muridae). *Bull. Am. Mus. Nat. Hist.* 158:174–220.
- MARSHALL, J. T. 1981. Taxonomy. Pages 17–26 in *The mouse in biomedical research*, volume 1 (H. L. Foster, J. D. Small, and J. G. Fox, eds.). Academic Press, New York.
- MARSHALL, J. T. 1998. Identification and scientific names of Eurasian house mice and their European allies, subgenus *Mus* (Rodentia: Muridae). Privately published. Available from author at National Museum of Natural History, Room 378, Washington, DC 20560-0111.
- MCCONNELL, T. J., W. S. TALBOT, R. A. MCINDOE, AND E. K. WAKELAND. 1988. The origin of MHC class II gene polymorphism within the genus *Mus*. *Nature* 322:651–654.
- MEZHHERIN, S. V., E. V. KOTENKOVA, AND A. G. MIKHAILENKO. 1998. The house mice, *Mus musculus* s. l., hybrid zone of Transcaucasus. *Z. Saug.* 63:154–168.
- MICKEVICH, M. F., AND J. S. FARRIS. 1981. The implications of congruence in *Menidia*. *Syst. Zool.* 30:351–370.
- MIYAMOTO, M. M. 1985. Consensus cladograms and general classifications. *Cladistics* 1:186–189.
- MIYAMOTO, M. M., AND W. M. FITCH. 1995. Testing species phylogenies and phylogenetic methods with congruence. *Syst. Biol.* 44:64–76.
- MORITA, T., H. KUBOTA, K. MURATA, M. NOZAKI, C. DELARBRE, K. WILLISON, Y. SATTA, M. SAKAIZUMI, N. TAKAHATA, G. GACHELIN, AND A. MATSUSHIRO. 1992. Evolution of the mouse *t* haplotype: Recent and worldwide introgression to *Mus musculus*. *Proc. Natl. Acad. Sci. USA* 89:6851–6855.
- MORIWAKI, K. 1994. Wild mouse from a geneticist's viewpoint. Pages xiii–xxv in *Genetics in wild mice* (K. Moriwaki, T. Shiroishi, and H. Yonekawa, eds.). Japan Scientific Societies Press, Tokyo.
- MORSE, H. C. 1981. The laboratory mouse—A historical perspective. Pages 1–16 in *The house mouse in biomedical research*, Volume 1 (H. L. Foster, J. D. Small, and J. G. Fox, eds.). Academic Press, New York.
- MUNTYANU, A. I. 1990. Ecological features of an overwintering population of the hillock mouse (*Mus hortulanus* Nordm.) in the southwest of the U.S.S.R. *Biol. J. Linn. Soc.* 41:73–82.
- MUSSER, G. G., AND M. D. CARLETON. 1993. Family Muridae. Pages 501–755 in *Mammal species of the world*. (D. E. Wilson and D. M. Reeder, eds.). Smithsonian Institution Press, Washington, D.C.
- ORSINI, P., F. BONHOMME, J. BRITTON-DAVIDIAN, H. CROSET, S. GERASIMOV, AND L. THALER. 1983. Le complexe d'espèces du genre *Mus* en Europe Centrale et Orientale. *Z. Saug.* 48:86–95.
- PARNES, J. R., AND J. G. SEIDMAN. 1982. Structure of wild-type and mutant mouse  $\beta_2$ -microglobulin genes. *Cell* 29:661–669.
- PATRIS, B., AND C. BAUDOIN. 1998. Female sexual preferences differ in *Mus spicilegus* and *Mus musculus domesticus*: The role of familiarization and sexual experience. *Anim. Behav.* 56:1465–1470.
- PRAGER, E. M., H. TICHY, AND R. D. SAGE. 1996. Mitochondrial DNA sequence variation in the eastern house mouse, *Mus musculus*: Comparison with other house mice and report of a 75-bp tandem repeat. *Genetics* 143:427–446.
- RAMBAUT, A., AND N. C. GRASSLY. 1997. Seq-Gen: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput. Appl. Biosci.* 13:235–238.
- RANKIN, T. L., Z.-B. TONG, P. E. CASTLE, E. LEE, R. GORE-LANGTON, L. M. NELSON, AND J. DEAN. 1998. Human ZP3 restores fertility in *Zp3* null mice without affecting order-specific sperm binding. *Development* 125:2415–2424.
- RHEAUME, C., R. L. GOODWIN, J. J. LATIMER, H. BAUMANN, AND F. G. BERGER. 1994. Evolution of murine  $\alpha_1$ -proteinase inhibitors: Gene amplification and reactive center divergence. *J. Mol. Evol.* 38:121–131.
- RINGUETTE, M. J., M. E. CHAMBERLIN, A. W. BAUR, D. A. SOBIESKI, AND J. DEAN. 1988. Molecular analysis of cDNA coding for ZP3, a sperm binding protein of the mouse zona pellucida. *Dev. Biol.* 127:287–295.
- RODRIGUEZ, F., J. L. OLIVER, A. MARIN, AND J. R. MEDINA. 1990. The general stochastic model of nucleotide substitution. *J. Theor. Biol.* 142:485–501.
- SAGE, R. D. 1981. Wild mice. Pages 39–90 in *The mouse in biomedical research*, volume 1 (H. L. Foster, J. D. Small, and J. G. Fox, eds.). Academic Press, New York.

- SAGE, R. D., W. R. ATCHLEY, AND E. CAPANNA. 1993. House mice as models in systematic biology. *Syst. Biol.* 42:523–561.
- SAGE, R. D., D. HEYNEMAN, K.-C. LIM, AND A. C. WILSON. 1986. Wormy mice in a hybrid zone. *Nature* 324:60–63.
- SAIKI, R. K., S. SCHARF, F. FALOONA, K. B. MULLIS, G. T. HORN, A. ERLICH, AND N. ARNHEIM. 1985. Enzymatic amplification of B-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230:1350–1354.
- SANTOS, J., Y. COLE, AND A. PELLICER. 1993. Phylogenetic relationships among laboratory and wild-origin *Mus musculus* strains on the basis of genomic DNA RFLPs. *Mamm. Genome* 4:485–492.
- SELANDER, R. K., W. G. HUNT, AND S. Y. YANG. 1969. Protein polymorphism and genetic heterozygosity in two European subspecies of the house mouse. *Evolution* 23:379–390.
- SHARMA, T. 1996. Chromosomal and molecular divergence in the Indian pygmy field mice *Mus boodugatericolor* lineage of the subgenus *Mus*. *Genetica* 97:331–338.
- 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.
- SIDDALL, M. E. 1997. Prior agreement: Arbitration or arbitrary? *Syst. Biol.* 46:765–769.
- SILVER, L. M., K. ARZT, AND D. BENNETT. 1979. A major testicular cell protein specified by a mouse T/t complex gene. *Cell* 17:275–284.
- SILVER, L. M., AND M. WHITE. 1982. A gene product of the mouse t complex with chemical properties of a cell surface-associated component of the extracellular matrix. *Dev. Biol.* 91:423–430.
- SIMISTER, N. E., AND K. E. MOSTOV. 1989. An Fc receptor structurally related to MHC class I antigens. *Nature* 337:184–187.
- SINCLAIR, A. H., P. BERTA, M. S. PALMER, J. R. HAWKINS, B. L. GRIFFITHS, M. J. SMITH, J. W. FOSTER, A. FRISCHAUF, R. LOVELL-BADGE, AND P. N. GOODFELLOW. 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* 346:240–244.
- SORENSEN, M. D. 1999. TreeRot, version 2. Boston Univ., Boston, Massachusetts.
- SOURROUILLE, P., C. HANNI, M. RUEDI, AND F. M. CATZEFLIS. 1995. Molecular systematics of *Mus crociduroides*, an endemic mouse of Sumatra (Muridae: Rodentia). *Mammalia* 59:91–102.
- SPRINGER, M. S., AND E. DOUZERY. 1996. Secondary structure and patterns of evolution among mammalian mitochondrial 12S rRNA molecules. *J. Mol. Evol.* 43:357–373.
- SUEYOSHI, T., R. KOBAYASHI, K. NISHIO, K. AIDA, R. MOORE, T. WADA, H. HANDA, AND M. NEGISHI. 1995. A nuclear factor (NF2d9) that binds to the male-specific P450 (*Cyp 2d-9*) gene in mouse liver. *Mol. Cell. Biol.* 15:4158–4166.
- SUZUKI, H., AND Y. KURIHARA. 1994. Genetic variation of ribosomal RNA in the house mouse, *Mus musculus*. Pages 107–119 in *Genetics in wild mice: Its application to biomedical research* (K. Moriwaki, T. Shiroishi, and H. Yonekawa, eds.). Japan Science Society Press, Tokyo.
- SWOFFORD, D. 1998. PAUP\*: Phylogenetic analysis using parsimony, beta version 4.0b4a. Sinauer Associates, Sunderland, Massachusetts.
- TAMURA, K., AND M. NEI. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10:512–526.
- THALER, L. 1986. Origin and evolution of mice: An appraisal of fossil evidence and morphological traits. *Curr. Top. Microb. Immunol.* 127:3–11.
- THALER, L., F. BONHOMME, J. BRITTON-DAVIDIAN, AND M. HAMAR. 1981. The house mouse complex of species: Sympatric occurrence of biochemical groups *Mus* 2 and *Mus* 4 in Rumania. *Z. Saug.* 46:169–173.
- THOMPSON, J. D., G. D. HIGGINS, AND T. J. GIBSON. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673–4680.
- TUCKER, P. K., B. K. LEE, AND E. M. EICHER. 1989. Y chromosome evolution in the subgenus *Mus* (Genus *Mus*). *Genetics* 122:169–179.
- TUTTER, A., AND R. RIBLET. 1989. Evolution of the immunoglobulin heavy chain variable region (*Igh-V*) locus in the genus *Mus*. *Immunogenetics* 30:315–329.
- VANLENBERGHE, F., P. BOURSOT, J. T. NIELSEN, AND F. BONHOMME. 1988. A steep cline for mitochondrial DNA in Danish mice. *Genet. Res.* 52:185–193.
- VAN ROMPAEY, J., W. VERHEYEN, AND M. SELENS. 1984. Genetic differences between three species of pigmy-mice in Rwanda (Africa). *Rev. Zool. Afr.* 98:886–890.
- VITIELLO, A., T. A. POTTER, AND L. A. SHERMAN. 1990. The role of  $\beta_2$ -microglobulin in peptide binding by class I molecules. *Science* 250:1423–1426.
- WASSARMAN, P. M. 1990. Profile of a mammalian sperm receptor. *Development* 108:1–17.
- WIENS, J. 1998. Combining data sets with different phylogenetic histories. *Syst. Biol.* 47:568–581.
- WILLIAMS, D. B., B. H. BARBER, R. A. FLAVELL, AND H. ALLEN. 1989. Role of  $\beta_2$ -microglobulin in the intracellular transport and surface expression of murine class I histocompatibility molecules. *J. Immunol.* 142:2796–2806.
- YANG, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J. Mol. Evol.* 39:306–314.
- YONEKAWA, H., O. GOTOH, Y. TAGASHIRA, Y. MATSUSHIMA, L.-I. SHI, W. S. CHO, N. MIYASHITA, AND K. MORIWAKI. 1986. A hybrid origin of Japanese mice "*Mus musculus molossinus*." *Curr. Top. Microb. Immunol.* 127:62–67.
- YONEKAWA, H., K. MORIWAKI, O. GOTOH, N. MIYASHITA, Y. MATSUSHIMA, L. SHI, W. S. CHO, X. ZHEN, AND Y. TAGASHIRA. 1988. Hybrid origin of Japanese mice "*Mus Musculus molossinus*": Evidence from restriction analysis of mitochondrial DNA. *Mol. Biol. Evol.* 5:63–78.
- ZHARKIKH, A. 1994. Estimation of evolutionary distances between nucleotide sequences. *J. Mol. Evol.* 39:315–329.

First submitted 30 October 2000; revision submitted

19 December 2000; final accepted 31 December 2001

Associate Editor: Jack Sullivan

APPENDIX. ORIGIN OF MICE USED  
IN THIS STUDY

*Mus musculus musculus*: Sry, Cyt b, 12S, B2m, Zp-3 (CZECH-II, Czechoslovakia); Tcp-1 (MBT, Bulgaria).

*Mus musculus molossinus*: Sry, Cyt b, 12S, B2m, Zp-3 (MOLO, Japan); Tcp-1 (MOL, Japan).

*Mus musculus castaneus*: Sry, Cyt b, 12S, Zp-3 (CAST, Thailand); B2m (CAST/Ei, Thailand); Tcp-1 (CAS, Indonesia).

*Mus musculus domesticus*: Sry, Cyt b, Zp-3 (POSCHIL, Switzerland); 12S (lab-bred *brevirostris*); B2m (BREV, Morocco); Tcp-1 (129/sv).

*Mus macedonicus*: Sry, Cyt b, 12S, Zp-3 (lab strain originating in Yugoslavia); B2m (XBJ, Bulgaria); Tcp-1 (XBS, Bulgaria).

*Mus spicilegus*: Sry, Cyt b, 12S, Zp-3 (HORT, Austria); B2m (ZYD, Yugoslavia); Tcp-1 (ZBN, Bulgaria).

*Mus spretus*: Sry, Cyt b, 12S, Zp-3 (SPRET-2, Morocco); B2m (Spain.Pt); Tcp-1 (SEL, Spain).

*Mus cervicolor*: Sry (lab strain originating in Thailand); Cyt b, 12S, Zp-3 (CERV, Thailand); B2m, Tcp-1 (CRP, Thailand).

*Mus cookii*: Sry, Cyt b, Zp-3, Tcp-1 (COOK, Thailand); 12S (lab strain originating in Thailand); B2m (COK, Thailand).

*Mus caroli*: Sry, Cyt b, 12S, Zp-3, Tcp-1 (Thailand); B2m (KAR, Thailand).

*Mus (Pyromys) saxicola*: Cyt b, B2m, Zp-3, Tcp-1 (SAXI, India); 12S (wild-caught in India).

*Mus (Coelomys) pahari*: Sry, Cyt b, Zp-3, Tcp-1 (PAHA, Thailand); 12S (lab strain originating in Thailand); B2m (PAH, Thailand).

*Mus (Nannomys) minutoides*: Sry, Cyt b, 12S, B2m, Zp-3, Tcp-1 (MINU, Kenya).

*Mastomys hildebrandtii*: Sry, Cyt b, Zp-3, Tcp-1 (wild-caught in Kenya).

*Mastomys erythroleucus*: 12S (wild-caught in Senegal).

*Hylomyscus alleni*: Sry, Cyt b, Zp-3, Tcp-1 (wild-caught in Gabon).

*Hylomyscus stella*: 12S (wild-caught in Burundi).

*Rattus exulans*: Sry (wild-caught in the Philippines).

*Rattus norvegicus*: Cyt b, 12S (stock/strain not provided); B2m (RT1o); Zp-3 (Sprague-Dawley); Tcp-1 (Fischer).