

---

## ***The genus Mus as a model for evolutionary studies***

*Edited by J. Britton-Davidian and J. B. Searle*

# **Phylogenetic relationships in the subgenus *Mus* (genus *Mus*, family Muridae, subfamily Murinae): examining gene trees and species trees**

PRISCILLA K. TUCKER<sup>1\*</sup>, SARA A. SANDSTEDT<sup>1</sup> and BARBARA L. LUNDRIGAN<sup>2</sup>

<sup>1</sup>*Museum of Zoology and Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109–1079, USA*

<sup>2</sup>*Michigan State University Museum and Department of Zoology, Michigan State University, East Lansing, Michigan 48824–1045, USA*

*Received 31 October 2003; accepted for publication 7 October 2004*

---

Over eight kilobases (kb) of sequence from eight genes including two mitochondrial loci, *Cytb* and 12S, and six nuclear loci, *B2m*, *Zp3*, *Tcp1*, *Sry*, *Smcx* and *Smcy*, were used to investigate phylogenetic relationships among 11 taxa representing eight species within the rodent genus *Mus*. Particular attention was given to discerning relationships among species within the subgenus *Mus* including members of a Palearctic clade (*M. musculus*, *M. spicilegus*, *M. macedonicus* and *M. spretus*) and members of an Asian clade (*M. caroli*, *M. cookii* and *M. cervicolor*), as previous studies using different datasets have produced different topologies for taxa within these two groups. While parsimony analyses of the combined eight-gene dataset yielded a single, fully resolved tree, support values were lower for nodes resolving relationships within the Palearctic and Asian clades than they were elsewhere in the tree. In addition, a maximum likelihood analysis of the same eight-gene dataset yielded different topologies for both the Palearctic and the Asian clades. Both observations are indicative of clade instability. The nature of this instability was explored through a comparison with our previous study in which we included the two mitochondrial loci and only four of the six nuclear genes, and through an analysis of partitioned data, specifically mitochondrial vs. nuclear genes. This study underscores the importance of considering among-site rate variation in phylogeny reconstruction. © 2005 The Linnean Society of London, *Biological Journal of the Linnean Society*, 2005, 84, 653–662.

**ADDITIONAL KEYWORDS:** likelihood – mtDNA – nuclear genes – parsimony – phylogeny.

---

### INTRODUCTION

The rodent genus *Mus* (family Muridae, subfamily Murinae), comprising approximately 30–40 species of mice (Musser & Carleton, 1993), has been the subject of numerous phylogenetic investigations over the last two decades (see Berry & Scriven, 2005, this issue). These studies focused primarily on taxa in the subgenus *Mus* and incorporated comparative data from allozymes (Sage, 1981; Bonhomme *et al.*, 1984; She

*et al.*, 1990), mitochondrial DNA (mtDNA) restriction fragment length polymorphisms (RFLPs) (Ferris *et al.*, 1983; She *et al.*, 1990), RFLPs of nuclear rDNA spacer regions (Suzuki & Kurihara, 1994), single-copy nuclear DNA (scnDNA) hybridization (She *et al.*, 1990; Catzeflis & Denys, 1992; Chevret, Jenkins & Catzeflis, 2003), mtDNA sequences (Fort *et al.*, 1984; Sourrouille *et al.*, 1995; Prager, Tichy & Sage, 1996; Prager, Orrego & Sage, 1998; Lundrigan, Jansa & Tucker, 2002; Chevret *et al.*, 2003) and nuclear sequences (Jouvin-Marche *et al.*, 1988; Lundrigan & Tucker, 1994; Lundrigan *et al.*, 2002; Auffray *et al.*, 2003).

---

\*Corresponding author. E-mail: ptuck@umich.edu

Results from many of these studies have been summarized in a 'synthetic' evolutionary tree (Guénet & Bonhomme, 2003) in which agreements among datasets are depicted as monophyletic groupings. These include the subgenus *Mus* comprising a Palearctic clade and two Asian clades. The Palearctic clade includes *M. musculus* and *M. spretus* as well as the sister taxa, *M. macedonicus* and *M. spicilegus*. One Asian clade includes the newly recognized species *M. fragilicauda* and *M. famulus* (Auffray *et al.*, 2003; Chevret *et al.*, 2003) and is positioned as sister to the Palearctic clade, while the other Asian clade includes *M. cervicolor*, *M. cookii* and *M. caroli*. Disagreements among datasets are depicted as polytomies. Within the subgenus *Mus*, these include relationships among the Asian species, *M. cervicolor*, *M. cookii* and *M. caroli*, relationships among the Palearctic species, and relationships among subspecies of the house mouse *M. musculus*.

Previously we (Lundrigan *et al.*, 2002) published a phylogeny for the genus *Mus* using DNA sequences from six genes representing both the nuclear (*B2m*, *Zp3*, *Tcp1*, *Sry*), and mitochondrial (*Cytb*, 12S) genomes (see also Chevret, Veyrunes & Britton-Davidian, 2005, this issue). Phylogenetic analyses were conducted not only to test hypotheses of relationships, but also to assess clade support and to examine congruence among characters. Maximum parsimony analyses of the combined six-gene dataset resulted in a fully resolved tree with strong support, based on bootstrap and Bremer support values, for the monophyly of: the subgenus *Mus*; the Palearctic clade; the Asian clade comprising *M. cervicolor*, *M. cookii* and *M. caroli*; *M. musculus*; and for a sister-group relationship between *M. macedonicus* and *M. spicilegus*. However, support values were lower for nodes resolving relationships among some species within the Palearctic clade and for relationships among *M. cervicolor*, *M. cookii* and *M. caroli*. In particular, the basal positions of *M. spretus* within the Palearctic clade and *M. caroli* within the Asian clade were not as well supported. Through analyses of partitioned data the lower support was attributed to differences between nuclear and mitochondrial genes. Specifically, in parsimony analyses of the combined mitochondrial data, *M. spretus* was basal within the Palearctic clade and *M. cervicolor* was basal within the Asian clade. In parsimony analyses of the combined nuclear data, *M. caroli* was basal within the Asian clade and *M. spretus* was sister to the clade comprising *M. macedonicus* and *M. spicilegus* within the Palearctic clade. Similar results were found using likelihood, with one exception. In analyses of the mitochondrial dataset, *M. caroli* was basal within the Asian clade.

We suggested that differences between the nuclear and mitochondrial gene trees for Palearctic and Asian

species are due to rate differences (Lundrigan *et al.*, 2002). Unlike each one of the nuclear genes, *Cytb* is evolving rapidly enough to contribute a considerable number of informative characters to resolve relationships among these closely related taxa. However, the *Cytb* data were homoplasious; for example, the sister-group relationship between *M. caroli* and *M. cookii* in the mitochondrial parsimony tree was weakly supported. We concluded that homoplasy in one data partition, specifically *Cytb*, may be obscuring phylogenetic signal in the combined data.

This paper, an extension of our previous study, utilized a larger nuclear dataset to reassess phylogenetic relationships among species within the Palearctic clade, in particular the position of *M. spretus*, and the relationships among species in the Asian clade. Specifically, we increased the number of informative nuclear characters by sequencing over 3 kb from both introns and exons of two nuclear genes, *Smcx* and *Smcy*, producing a combined eight-gene dataset. We also re-examined, in greater detail than in our previous paper (Lundrigan *et al.*, 2002), the contribution of nucleotide characters to the phylogenetic signal, including noncoding sequences and coding sequences by codon position.

## MATERIAL AND METHODS

### SPECIES

A total of 13 taxa representing ten species from three genera, *Mus*, *Mastomys* and *Hylomyscus*, were used in this study. The genus *Mus* was represented by two of the four subgenera, *Mus* and *Coelomys*; the latter subgenus was represented by a single taxon, *M. pahari*. The origins for all samples were provided in the appendix of our earlier study (Lundrigan *et al.*, 2002). Three of the species used in the earlier study were excluded from this study because additional sequence data from *Smcx* and *Smcy* were not available. All of the species are members of the rodent family Muridae and the subfamily Murinae, the Old-World mice and rats. Muridae is the largest mammalian family, containing 281 genera and 1326 species (Musser & Carleton, 1993). The sister group to *Mus* is unknown, although there is evidence from DNA hybridization (Catzeflis & Denys, 1992; Chevret *et al.*, 1994) and nuclear gene data (Jansa & Weksler, 2004) for a sister-group relationship with the *Praomys* group. Thus, members of the *Praomys* group (*Hylomyscus alleni* and *Mastomys hildebrandtii*) were used as outgroups to root trees in this study.

The house mice *M. musculus musculus* and *M. m. domesticus* were represented as subspecies due to their ability to hybridize along contact zones (Boursot *et al.*, 1993). However, these hybrids suffer from a loss of fertility and the species are genetically and

morphologically distinct away from the hybrid zones. Thus, *M. m. domesticus* is sometimes referred to as a distinct species, *M. domesticus* (Marshall, 1981; Prager *et al.*, 1998). *M. m. molossinus* is the result of hybridizations of *M. m. musculus*, *M. m. castaneus*, and occasionally *M. domesticus*, in Japan (Yonekawa *et al.*, 1988).

#### GENES SEQUENCED

The complete dataset included nucleotide sequences from eight genes. Six of these genes, the male sex-determining locus (*Sry*), cytochrome b (*Cytb*), 12S ribosomal RNA (12S), B2-microglobulin (*B2m*), zona pellucida-3 (*Zp3*), and t-complex polypeptide-1 (*Tcp1*), were used in our earlier study and are described in Lundrigan *et al.* (2002). *Smcy* (selected mouse cDNA Y) and *Smcx* are members of the jumonji family of transcription factors (Takeuchi *et al.*, 1995) and are thought to be involved in chromatin-mediated transcriptional regulation (Aasland, Gibson & Stewart, 1995). *Smcy* encodes several of the H-Y antigens that can cause the rejection of male-donor tissue transplants in female recipients (Scott *et al.*, 1995; Wang *et al.*, 1995). *Cytb* and 12S are mitochondrial genes. *Sry* and *Smcy* are located on the Y chromosome. *Smcx* is located on the X chromosome. *B2m*, *Zp3* and *Tcp1* are located on autosomes.

#### DNA AMPLIFICATION AND SEQUENCING

Genomic DNA was extracted from frozen tissues (Jenkins *et al.*, 1982). All sequences were enzymatically amplified by PCR (Saiki *et al.*, 1985). The primers used, sequencing methods employed for six of the genes (*Sry*, *Cytb*, 12S, *B2m*, *Zp3*, *Tcp1*) and GenBank accession numbers are given elsewhere (Lundrigan *et al.*, 2002). A fragment of *Smcy* including two complete exons (600 bp and 82 bp in length), a 9-bp portion of one other exon, and three introns (ranging from 176 to 204, 401 to 780, and 128 to 129 bp in length) was isolated. The beginning of the fragment corresponded to base 3443 and the end to base 4129 of an *Smcy* sequence from *M. musculus*, accession number AF127244, obtained from GenBank. The corresponding fragment of *Smcx* including two complete exons of 582 and 70 bp in length, a 9-bp portion of another exon, and introns ranging from 171 to 176, 266 to 282, and 145 to 146 bp in length was also isolated. The beginning of this fragment corresponded to base 3728 and the end to base 4387 of an *Smcx* sequence from *M. musculus*, accession number AF127245, obtained from GenBank. The GenBank sequences of *Smcy* and *Smcx* used as references did not contain introns. Amplifications were performed using primers under various cycling conditions determined empirically

using several primer sets in different taxa (Sandstedt & Tucker, 2004). Sequencing was done as described in Sandstedt & Tucker (2004). The *Smcx* and *Smcy* sequences have been deposited in GenBank (accession numbers AY260478—AY260503).

#### ALIGNMENT

*Smcy* and *Smcx* nucleotide sequences were aligned using Clustal X (Thompson *et al.*, 1997) and adjusted by eye using amino acid alignments as a guide. These sequences were then combined with aligned sequences from the six genes from Lundrigan *et al.* (2002) to produce an eight-gene dataset.

#### PHYLOGENETIC ANALYSES

Maximum parsimony analyses were performed on *Smcx*, *Smcy*, the combined nuclear six-gene dataset, the combined mitochondrial two-gene dataset, and the combined eight-gene dataset, using the branch and bound option as implemented in PAUP\* version 4.0b10 (Swofford, 2002). In all but one analysis, phylogenetically informative characters were equally weighted and unordered. Gaps were treated as missing characters. To assess the influence of *Cytb* transitional substitutions on the tree topology, transitional substitutions were down-weighted in one analysis of the combined eight-gene dataset. Trees were rooted using the outgroup taxa, *H. alleni* and *Ma. hildebrandtii*. Bootstrap analyses were performed using the full heuristic option with 1000 replicates and tree-bisection-and-reconnection (TBR) branch swapping. TreeRot version 2 (Sorenson, 1999) was used to construct PAUP\* command files for use in the Bremer support (Bremer, 1988, 1994) calculations.

Maximum likelihood analyses were performed on the same datasets as described above using PAUP\* ver. 4.0b10 (Swofford, 2002). The program Modeltest (Posada & Crandall, 1998) was used to identify best-fit models for *Smcx* (HKY + G), *Smcy* (HKY + G), the combined six-gene nuclear dataset (HKY + G), the combined two-gene mitochondrial dataset (TrN + G + I) and the combined eight-gene dataset (TrN + G + I). The maximum likelihood trees were determined using heuristic searches and TBR branch swapping. Starting trees were obtained via random stepwise addition with 1000 replicates. Bootstrap analyses were performed using the full heuristic option with 500 pseudoreplicates.

Bayesian phylogenetic analyses were conducted with Mr Bayes 3.0b4 (Huelsenbeck & Ronquist, 2001) with parameters estimated separately for each of the genes. Three separate analyses using the combined eight-gene, the combined-six gene nuclear and the combined two-gene mitochondrial datasets were con-

ducted. Each analysis was initiated with random starting trees and was run for  $2 \times 10^6$  generations, sampling every 100th generation. Four continuous chains were run. To check that stationarity had been reached, the fluctuating value of the likelihood was checked graphically. The initial 1000 trees were discarded as burn-in. The effective sample size of all parameters as determined in Tracer (Rambaut & Drummond, 2003) was greater than 100. This indicated that parameters were sampled appropriately. The simulation was conducted twice.

#### ILD tests

To assess character congruence among genes, the Mickevich–Farris character incongruence metric (Mickevich & Farris, 1981) was used. Statistical significance was calculated with the ILD test as described in Farris *et al.* (1994) and implemented in PAUP\* version 4.0b10. Each test included 1000 replicates. Searches were heuristic with simple taxon addition and TBR branch swapping. Only informative characters were used.

#### Shimodaira–Hasegawa test

To determine whether likelihood scores for alternate topologies were significantly different, Shimodaira–Hasegawa tests (Shimodaira & Hasegawa, 1999) were performed as implemented in PAUP\* version 4.0b10.

## RESULTS

### PHYLOGENETIC ANALYSES

#### Basic statistics

Basic statistics for all eight genes, including non-coding and coding sequences (partitioned by codon position), are given in Table 1. These statistics are base composition, total number of nucleotides analysed, numbers of variable and parsimony-informative characters and consistency indices (CIs) excluding uninformative characters. CIs are given for both the eight-gene parsimony tree and the eight-gene likelihood tree to assess homoplasy for alternative topologies. Separate CIs for transition and transversion substitutions are given for each. *Cytb* contributed the largest number of variable characters (28% of the total number) and the largest number of parsimony-informative characters (37% of the total number) for any single gene. *Cytb* third-position sites contributed 21% of the total number of variable characters and 30% of the total number of parsimony-informative characters. *Smcy* contributed the second largest number of variable characters (18% of the total number) and the second largest number of parsimony-informative characters (18%). *Smcy*'s non-coding sequence contributed 13% of the total number of

variable characters and 12% of the total number of parsimony-informative characters.

*Cytb* had the lowest CIs for transition substitutions at each codon position (range, 0.3848–0.5556) of all genes (range, 0.5556–1) with the exception of *B2m*; the second codon position of *Cytb* and the first codon position of *B2m* had identical CIs. The CIs for transition substitutions were higher when the mitochondrial gene characters were mapped on to the combined eight-gene maximum parsimony tree vs. the combined eight-gene maximum likelihood tree. The exception was second-position sites for *Cytb*, where the CIs were identical. The CIs for transition substitutions on the eight-gene maximum likelihood tree were higher than or identical to the CIs for transitional substitutions on the eight-gene parsimony tree when the nuclear characters were mapped. The CIs for transversion substitutions ranged from 0.5 to 1. No general pattern for CIs was observed between mitochondrial and nuclear genes, or among codon positions, when transversions were mapped onto the combined eight-gene maximum likelihood tree or the combined eight-gene maximum parsimony tree.

#### Individual-gene analyses

Parsimony and likelihood analyses for *Smcx* produced identical trees (Fig. 1). *M. spretus* was sister to the *M. spicilegus*/*M. macedonicus* clade and the relationship among the Asian taxa, *M. cervicolor*, *M. caroli* and *M. cookii*, was unresolved. Parsimony and likelihood analyses for *Smcy* yielded trees that differed in the placement of *M. spretus* (Fig. 1). In the parsimony analysis two most parsimonious trees were recovered. In one tree, *M. spretus* was sister to the *M. spicilegus*/*M. macedonicus* clade. In the other tree, *M. spretus* was basal to a clade comprising the *M. spicilegus*/*M. macedonicus* clade and *M. musculus*. In the likelihood analysis *M. spretus* was sister to the *M. spicilegus*/*M. macedonicus* clade, although bootstrap support for this arrangement was less than 50%. In all trees, the Palearctic species of *Mus* and the Asian species of *Mus* each formed monophyletic groups and the subgenus *Mus* constituted a monophyletic group. Parsimony and likelihood bootstrap values for both genes were similar (Fig. 1).

#### ILD tests

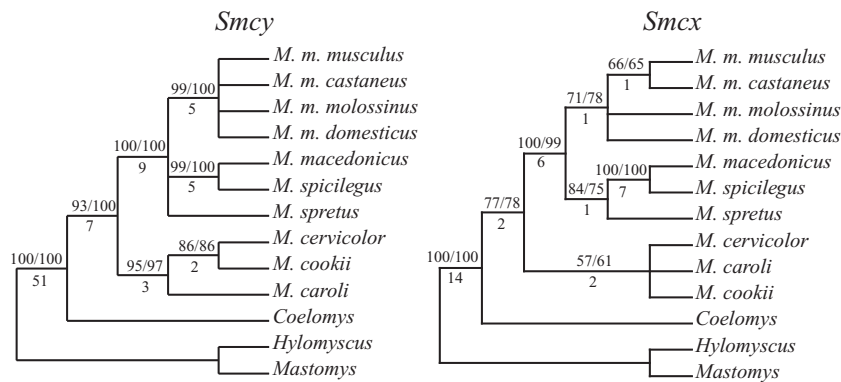
Pairwise ILD tests for *Smcx* with each of the nuclear and mitochondrial genes and for *Smcy* with each of the nuclear and mitochondrial genes failed to reject the null hypothesis of dataset homogeneity ( $P < 0.01$  as recommended by Cunningham, 1997).  $P$ -values ranged from 0.024 to 1.0. This result is consistent with previous tests for the other genes in the dataset (Lundrigan *et al.*, 2002). ILD tests on mitochondrial (*Cytb* and 12S) and nuclear partitions (*Smcx*, *Smcy*, *Sry*,



**Table 1.** Basic statistics

Gene		Base composition				total	Number of characters		CI: transitions		CI: transversions	
		A	C	G	T		var.	pars. info.	ML	MP	ML	MP
12S		0.38	0.20	0.16	0.27	987	175	95	0.5873	0.5902	0.6849	0.6849
<i>Cytb</i>	Position 1	0.30	0.23	0.21	0.25	382	71	45	0.4203	0.4762	0.70	0.7368
	Position 2	0.20	0.24	0.13	0.43	382	17	6	0.5556	0.5556	1	1
	Position 3	0.45	0.31	0.27	0.21	381	291	222	0.3848	0.3906	0.665	0.6585
<i>B2m</i>	Position 1	0.35	0.25	0.22	0.19	92	19	14	0.5556	0.5556	0.75	0.8182
	Position 2	0.38	0.25	0.11	0.25	93	13	6	na	na	0.6	0.5
	Position 3	0.21	0.34	0.25	0.20	93	28	6	1	1	0.5556	0.5
<i>Zp3</i>	Non-coding	0.20	0.23	0.26	0.29	230	56	24	0.6667	0.6667	0.8	0.8
	Position 1	0.20	0.26	0.29	0.25	225	20	11	0.8	0.7273	0.6	0.6
	Position 2	0.29	0.29	0.17	0.24	223	16	6	1	1	0.6667	0.6667
<i>Tcp1</i>	Position 3	0.13	0.34	0.33	0.21	222	61	25	0.7586	0.7241	0.6667	0.6667
	Non-coding	0.25	0.17	0.22	0.36	1053	140	56	0.8049	0.75	0.9167	0.9167
	Position 1	0.31	0.13	0.40	0.16	134	9	3	1	1	0.5	0.5
<i>Smcx</i>	Position 2	0.28	0.21	0.18	0.32	134	5	4	0.6667	0.6667	na	na
	Position 3	0.25	0.13	0.25	0.37	133	26	11	1	0.9091	1	1
	Non-coding	0.20	0.25	0.24	0.32	612	66	27	0.7273	0.6957	0.9167	0.8462
<i>Smcy</i>	Position 1	0.18	0.30	0.34	0.17	220	6	4	1	1	1	1
	Position 2	0.24	0.27	0.24	0.25	220	3	3	0.6667	0.6667	1	1
	Position 3	0.21	0.24	0.37	0.19	220	17	7	0.6000	0.6000	1	1
<i>Sry</i>	Non-coding	0.29	0.19	0.19	0.33	1144	179	88	0.7907	0.7907	0.9545	0.913
	Position 1	0.25	0.24	0.30	0.21	228	21	11	0.9	0.8	1	1
	Position 2	0.31	0.22	0.20	0.26	228	13	7	0.8571	0.75	1	1
<i>Sry</i>	Position 3	0.28	0.20	0.28	0.24	228	39	26	0.8696	0.8696	1	0.8
	Non-coding	0.32	0.20	0.25	0.23	76	14	4	1	1	1	1
	Position 1	0.33	0.26	0.22	0.19	144	17	7	0.67	0.60	1	1
<i>Sry</i>	Position 2	0.37	0.21	0.22	0.20	143	20	9	1	1	1	0.7143
	Position 3	0.26	0.25	0.29	0.20	143	33	14	0.8462	0.8462	1	1

Base composition, length of sequence examined (total), number of variable characters (var.) and parsimony informative characters (pars. info.) for non-coding sequence and coding sequence by codon position for the eight genes used in the phylogenetic analyses. Consistency indices (CIs) excluding uninformative characters for characters mapped on the combined eight-gene maximum parsimony (MP) and maximum likelihood (ML) trees are given for transitions and transversions.



**Figure 1.** Maximum parsimony (MP) gene trees for the subgenus *Mus* based on comparative DNA sequence from *Smcy* and *Smcx* with MP/maximum likelihood (ML) bootstrap and Bremer support values given above and below the branches, respectively. The *Smcy* tree is a strict consensus of two trees with length 282 and a consistency index (CI) excluding uninformative characters of 0.8730. The MP *Smcx* tree has a length of 104 and a CI excluding uninformative characters of 0.8269. Corresponding ML trees in both cases were similar and described in the text.

*B2m*, *Zp3*, *Tcp1*) and on maternally inherited (*Cytb* and 12S) and biparentally inherited (*Smcx*, *B2m*, *Zp3*, *Tcp1*) partitions were significantly incongruent; *P*-values were 0.001 and 0.008, respectively. ILD tests on maternally inherited (*Cytb* and 12S) and paternally inherited (*Smcy*, *Sry*) partitions and biparentally inherited (*Smcx*, *B2m*, *Zp3*, *Tcp1*) and paternally inherited (*Smcy*, *Sry*) partitions failed to reject the null hypothesis of dataset homogeneity; *P*-values were 0.06 and 1, respectively.

#### Combined eight-gene analysis

Parsimony analyses of the combined eight-gene dataset yielded a single, fully resolved tree (Fig. 2A). Within the Palearctic clade, *M. spretus* was basal to a group comprising the *M. spicilegus*/*M. macedonicus* clade and the *M. musculus* clade. Bootstrap and Bremer support values for this arrangement were 71% and 3, respectively. Within the Asian clade, *M. caroli* and *M. cookii* were sister taxa, and *M. cervicolor* was sister to that clade, although this relationship was not well supported, with bootstrap and Bremer support values of 54% and 1, respectively.

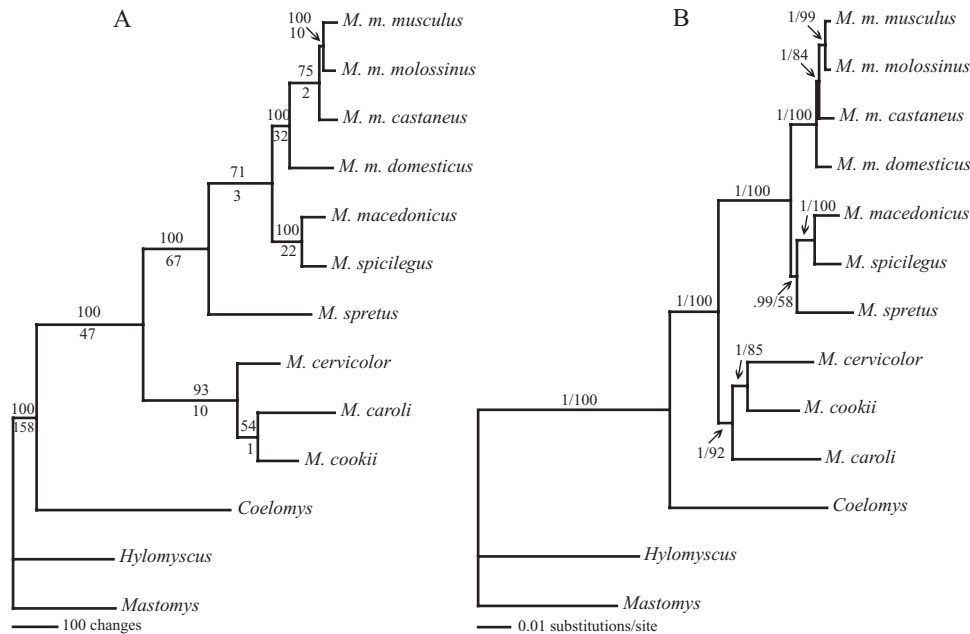
In parsimony analyses of the combined eight-gene dataset where transitional substitutions were down-weighted to 0.9, *M. caroli* switched from being sister to *M. cookii* to a basal position in the Asian clade. This tree topology remained stable until *Cytb* transitional substitutions were successively down-weighted to 0.3,

at which point *M. spretus* switched from a basal position in the Palearctic clade to being sister to *M. spicilegus*/*M. macedonicus*.

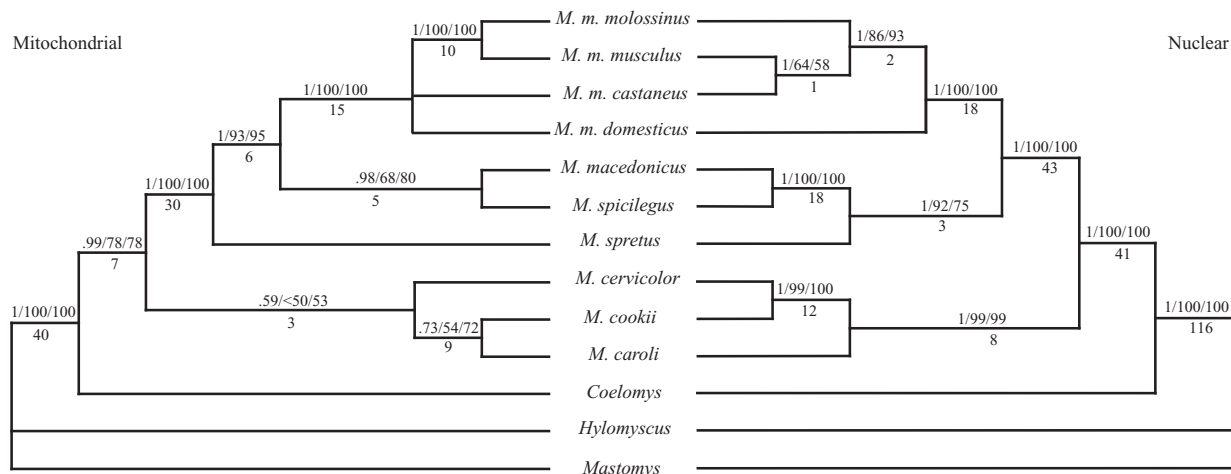
A maximum likelihood analysis and a Bayesian analysis of the combined eight-gene dataset produced fully resolved trees (Fig. 2B). In contrast to the parsimony tree, *M. spretus* was sister to the *M. spicilegus*/*M. macedonicus* clade within the Palearctic clade with bootstrap support of 58% and posterior probability of 0.99, and *M. caroli* was basal to an *M. cervicolor* and *M. cookii* clade with bootstrap support of 85% and posterior probability of 1.

#### Combined mitochondrial analysis

Parsimony analyses of the combined two-gene mitochondrial dataset yielded two most parsimonious trees differing only with respect to the relationship of *M. m. castaneus* and *M. m. domesticus* to an *M. m. musculus*/*M. m. molossinus* clade (Fig. 3). Relationships among species within the Palearctic clade were identical to the combined eight-gene parsimony tree. Bootstrap and Bremer support values for the placement of *M. spretus* as basal to the other species in the Palearctic clade were 95% and 6, respectively. Likewise, relationships among species within the Asian clade were identical to the combined eight-gene parsimony tree. Bootstrap and Bremer support values for the placement of *M. cervicolor* as basal to an *M. caroli*/*M. cookii* clade were 72% and 9, respectively.



**Figure 2.** A, maximum parsimony (MP) and B, maximum likelihood (ML) trees for the subgenus *Mus* for the combined eight-gene dataset. The MP tree has a length of 2114 and a consistency index excluding uninformative characters of 0.6242. MP bootstrap and Bremer support values are given above and below the branches, respectively, in A. Bayesian posterior probabilities (PP) and ML bootstrap values (PP/ML) are given above the branches in B.



**Figure 3.** Two maximum parsimony (MP) trees, one based on the mitochondrial sequences only and the other on the nuclear sequences only, for the subgenus *Mus*. The MP mitochondrial tree has a length of 1141 and a consistency index (CI) excluding uninformative characters of 0.5431. The MP nuclear tree has a length of 957 and a CI excluding uninformative characters of 0.8124. Maximum likelihood (ML) and Bayesian trees for both analyses were identical. Posterior probabilities, ML bootstrap values, and MP bootstrap values (PP/ML/MP) and Bremer support values are given above and below the branches, respectively.

Maximum likelihood and Bayesian analyses of the combined two-gene mitochondrial dataset produced trees identical to the most parsimonious tree in which *M. m. castaneus* was sister to the *M. m. musculus*/*M. m. molossinus* clade. Bootstrap support for this arrangement was less than 50% in the likelihood analysis and was 0.54 in the Bayesian analysis. Bootstrap values for the likelihood analysis and Bayesian posterior probabilities were generally similar to bootstrap values for the parsimony tree with the exception of the *M. caroli*/*M. cervicolor* clade where the maximum likelihood bootstrap value of 54% was considerably lower than a parsimony bootstrap value of 72% and a Bayesian posterior probability of 0.73 (Fig. 3).

#### Combined nuclear analysis

Parsimony analyses of the combined six-gene nuclear dataset yielded a single most parsimonious tree (Fig. 3) identical to the eight-gene tree. Within the Palearctic clade, *M. spretus* was sister to the *M. spicilegus*/*M. macedonicus* clade with bootstrap and Bremer support values of 75% and 3, respectively. Within the Asian clade, *M. caroli* was basal to an *M. cervicolor*/*M. cookii* clade with bootstrap and Bremer support values of 99% and 8, respectively. A maximum likelihood analysis and a Bayesian analysis of the combined six-gene nuclear dataset produced identical trees. Bayesian posterior probabilities and bootstrap values for the likelihood tree were similar to bootstrap values for the parsimony tree with one exception: support for *M. spretus* as sister to the *M. spicilegus*/*M. macedonicus* clade was considerably

higher, with a bootstrap value of 92% and a posterior probability of 1 (Fig. 3).

#### Shimodaira–Hasegawa test

A Shimodaira–Hasegawa (Shimodaira & Hasegawa, 1999) test was performed on the eight-gene dataset to determine whether the likelihood score ( $-\ln L$  22504) for a tree identical to the eight-gene parsimony tree was significantly different from the tree with the highest likelihood ( $-\ln L$  22490). No significant difference (at the  $P < 0.05$  level) was found ( $P = 0.162$ ). Shimodaira–Hasegawa tests were also performed on the two-gene mitochondrial dataset to determine whether the likelihood score ( $-\ln L$  7990) for a tree identical to the mitochondrial tree was significantly different from the likelihood score ( $-\ln L$  8037) for a tree identical to the nuclear tree as well as on the six-gene nuclear dataset to determine whether the likelihood score ( $-\ln L$  14269) for a tree identical to the mitochondrial tree was significantly different from the likelihood score ( $-\ln L$  14220) for a tree identical to the nuclear tree. For both tests scores were significantly different.

## DISCUSSION

The addition of over 3 kb of sequence, from both introns and exons of two nuclear genes, to our comparative molecular dataset (Lundrigan *et al.*, 2002) generated several interesting results. In contrast to our previous study, ILLD tests indicated incongruence between the mitochondrial and nuclear datasets. Not

surprisingly, Shimodaira–Hasegawa tests comparing alternate (nuclear vs. mitochondrial) topologies for the separate mitochondrial and nuclear datasets were also significantly different. Differences between the mitochondrial and nuclear phylogenies could reflect different histories for these two genomes resulting from differential introgression of mitochondrial vs. nuclear genes or it could reflect lineage sorting or rate heterogeneity.

#### THE ASIAN CLADE

The basal position of *M. caroli* within the Asian clade in the combined nuclear gene tree (Fig. 3) contrasted with the basal position of *M. cervicolor* in the combined mitochondrial tree (Fig. 3; Lundrigan *et al.*, 2002: fig. 3). However, there was no compelling evidence to support the hypothesis of a different history for the mitochondrial and nuclear genomes because the position of *M. cervicolor* as basal was weakly supported in the mitochondrial tree (bootstraps of 54% and 72% for maximum likelihood and maximum parsimony trees, respectively, Fig. 3). In the combined eight-gene parsimony tree (Fig. 2A) *M. cervicolor* was basal but bootstrap/Bremer support values (54%/1) were also low. In contrast, the basal position of *M. caroli* was more strongly supported in the eight-gene likelihood tree (bootstrap of 85%, Fig. 2B) and in the combined nuclear tree (bootstraps of 99% and 100% for maximum likelihood and maximum parsimony trees, respectively, Fig. 3). Because the mitochondrial gene *Cytb* contributed the greatest number of informative characters for any single gene, with transition substitutions, especially at first- and third-position sites, exhibiting considerable homoplasy (Table 1), we explored the contribution of transition substitutions in *Cytb* to the phylogenetic signal through successive down-weighting in a combined eight-gene parsimony analysis. A slight decrease in weighting, from 1.0 to 0.9, resulted in the repositioning of *M. caroli* to a basal position in the Asian clade. This suggests instability for the topology in which *M. caroli* is sister to *M. cookii* in the eight-gene parsimony tree (Fig. 2A) and the two-gene mitochondrial tree (Fig. 3) and is consistent with the observed low bootstrap and Bremer support values for that arrangement in those two trees.

#### THE POSITION OF *M. SPRETUS*

The addition of two nuclear genes to our comparative molecular dataset resulted in lower bootstrap and Bremer support values for the position of *M. spretus* as basal to the rest of the Palearctic taxa in the combined eight-gene parsimony analysis (71%/3) (Fig. 2A) vs. the combined six-gene parsimony analysis (88%/6)

(Lundrigan *et al.*, 2002; Fig. 4). This result is most likely due to the influence of *Smcx* and possibly *Smcy*. The *Smcx* gene tree supported the alternate topology, in which *M. spretus* was sister to *M. spicilegus*/*M. macedonicus*, and the *Smcy* tree was unresolved with respect to the position of *M. spretus* within the Palearctic clade. There was also increased bootstrap and Bremer support (75%/3) for *M. spretus* as sister to *M. spicilegus*/*M. macedonicus* in the combined six-gene nuclear analysis (Fig. 3) in comparison with the combined four-gene nuclear analysis (66%/2, Lundrigan *et al.*, 2002, fig. 3).

We also explored the contribution of transition substitutions in *Cytb* to the phylogenetic signal with respect to the position of *M. spretus*. The repositioning of *M. spretus* from a basal position within the Palearctic clade to a sister relationship with *M. spicilegus*/*M. macedonicus* when transitions were down-weighted from 0.9 to 0.3 is more difficult to interpret. While *Cytb* transition substitutions appeared largely responsible for the basal position of *M. spretus* in the Palearctic clade, this signal was congruent across *Cytb* and 12S as support values were high for this arrangement in the combined two-gene mitochondrial trees (bootstraps of 93 and 95% for maximum likelihood and maximum parsimony trees, respectively, Fig. 3). Two explanations for this observation can be given. First, there are real differences in the mitochondrial and nuclear genome histories for some of the Palearctic taxa. The difference in the position of *M. spretus* in the mitochondrial vs. nuclear trees could possibly result from historical introgression of nuclear genes between *M. spretus* and *M. spicilegus*/*M. macedonicus*, provided the ranges of these taxa overlapped historically; the range of *M. spretus* does not currently overlap with either *M. spicilegus* or *M. macedonicus*.

Second, the basal position of *M. spretus* (Fig. 2A) in the combined eight-gene maximum parsimony tree could be due to a large number of variable and homoplasious sites in *Cytb* that obscure the phylogenetic signal in the other, less variable, data partitions. The maximum likelihood analysis of the eight genes resulted in a maximum likelihood tree (Fig. 2B) that was different from the maximum parsimony combined eight-gene tree (Fig. 2A). While these alternate topologies were not significantly different based on the Shimodaira–Hasegawa test, different optimal topologies from methods that model (maximum likelihood) or do not model (maximum parsimony) rate heterogeneity suggest that rate heterogeneity is a factor.

Other phylogenetic studies of rodents (Huchon *et al.*, 2002; Adkins, Walton & Honeycutt, 2003; DeBry, 2003) using multigene datasets have been concerned with problems of rate heterogeneity. In particular, DeBry (2003) demonstrated lack of congruence between data partitions, with apparent homoplasy in



a single partition (third position sites of interphotoreceptor binding protein (IRBP)) obscuring the phylogenetic signal elsewhere in the data. All of these studies were concerned with higher-level relationships among rodent families where one might expect problems in resolving deep nodes due to homoplasy. This study has served to highlight a similar problem for closely related taxa, i.e. members of a rodent subgenus.

#### ACKNOWLEDGEMENTS

We thank M. Potter (National Cancer Institute Contract N01-CB-71085), E. Eicher (the Jackson Laboratory) and R.D. Sage for providing frozen tissue or DNA from various stocks and strains of wild mice, and R. J. Baker (The Museum, Texas Tech University) for providing frozen tissue from wild mice collected in the field. L. W. Harris and J. Rest provided analytical assistance. The research was supported in part by National Science Foundation Grant DEB-9209950 and National Institutes of Health Grant EY12994 (to P.K.T.), and grants from the Museum of Zoology, University of Michigan (to S.A.S.).

#### REFERENCES

- Aasland R, Gibson TJ, Stewart AF. 1995.** The Phd finger-implications for chromatin-mediated transcriptional regulation. *Trends in Biochemical Sciences* **20**: 56–59.
- Adkins RM, Walton AH, Honeycutt RL. 2003.** Higher-level systematics of rodents and divergence time estimates based on two congruent nuclear genes. *Molecular Phylogenetics and Evolution* **26**: 409–420.
- Auffray J-C, Orth A, Catalan J, Gonzalez J-P, Desmarais E, Bonhomme F. 2003.** Phylogenetic position and description of a new species of subgenus *Mus* (Rodentia, Mammalia) from Thailand. *Zoologica Scripta* **32**: 119–127.
- Berry RJ, Scriven PN. 2005.** The house mouse: a model and motor for evolutionary understanding. *Biological Journal of the Linnean Society* **84**: 335–347.
- Bonhomme F, Catalan J, Britton-Davidian J, Chapman VM, Moriwaki K, Nevo E, Thaler L. 1984.** Biochemical diversity and evolution in the genus *Mus*. *Biochemical Genetics* **22**: 275–303.
- Boursot P, Auffray JC, Britton-Davidian J, Bonhomme F. 1993.** The evolution of house mice. *Annual Review of Ecology and Systematics* **24**: 119–152.
- Bremer K. 1988.** The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* **42**: 795–803.
- Bremer K. 1994.** Branch support and clade stability. *Cladistics* **10**: 295–304.
- Catzeffis FM, Denys C. 1992.** The African *Nannomys* (Muridae): An early offshoot from the *Mus* lineage. Evidence from scDNA hybridization experiments and compared morphology. *Israel Journal of Zoology* **38**: 219–231.
- Chevret P, Granjon J, Duplantier J, Denys C, Catzeffis F. 1994.** Molecular phylogeny of the *Praomys* complex (Rodentia: Murinae): a study based on DNA/DNA hybridization experiments. *Zoological Journal of the Linnean Society* **112**: 425–442.
- Chevret P, Jenkins P, Catzeffis F. 2003.** Evolutionary systematics of the Indian mouse *Mus famulus* Bonhote, 1898: molecular (DNA/DNA hybridization and 12S rRNA sequences) and morphological evidence. *Zoological Journal of the Linnean Society* **137**: 385–401.
- Chevret P, Veyrunes F, Britton-Davidian J. 2005.** Molecular phylogeny of the genus *Mus* (Rodentia: Murinae) based on mitochondrial and nuclear data. *Biological Journal of the Linnean Society* **84**: 417–427.
- Cunningham CW. 1997.** Can three incongruence tests predict when data should be combined? *Molecular Biology and Evolution* **14**: 733–740.
- DeBry RW. 2003.** Identifying conflicting signal in a multigene analysis reveals a highly resolved tree: the phylogeny of Rodentia (Mammalia). *Systematic Biology* **52**: 604–617.
- Farris JS, Kallersjo M, Kluge AG, Bult C. 1994.** Testing significance of incongruence. *Cladistics* **10**: 315–319.
- Ferris SD, Sage RD, Huang C-M, Nielsen JT, Ritte U, Wilson AC. 1983.** Flow of mitochondrial DNA across a species boundary. *Proceedings of the National Academy of Sciences, USA* **80**: 2290–2294.
- Fort P, Bonhomme F, Darlu P, Piechaczyk M, Jeanteur P, Thaler L. 1984.** Clonal divergence of mitochondrial DNA versus population evolution of nuclear genome. *Evolutionary Theory* **7**: 81–90.
- Guénet J-L, Bonhomme F. 2003.** Wild mice: an ever-increasing contribution to a popular mammalian model. *Trends in Genetics* **19**: 24–31.
- Huchon D, Madsen O, Sibbald MJJB, Ament K, Stanhope MJ, Catzeffis F, DeJong WW, Douzery EJP. 2002.** Rodent phylogeny and a timescale for the evolution of Glires: evidence from an extensive taxon sampling using three nuclear genes. *Molecular Biology and Evolution* **19**: 1053–1065.
- Huelsenbeck JP, Ronquist F. 2001.** MR BAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754–755.
- Jansa SA, Weksler M. 2004.** Phylogeny of muroid rodents: relationships within and among major lineages as determined by IRBP gene sequences. *Molecular Phylogenetics and Evolution* **31**: 256–276.
- Jenkins NA, Copeland NG, Taylor BA, Lee BK. 1982.** Organization, distribution and stability of endogenous ectropic murine leukemia virus DNA sequence in chromosomes of *Mus musculus*. *Journal of Virology* **43**: 26–36.
- Jouvin-Marche E, Cuddihy A, Butler S, Hansen JN, Fitch WM, Rudikoff S. 1988.** Modern evolution of a single-copy gene: The immunoglobulin C<sub>k</sub> locus in wild mice. *Molecular Biology and Evolution* **5**: 500–511.
- Lundrigan BL, Jansa SA, Tucker PK. 2002.** Phylogenetic relationships in the genus *Mus*, based on paternally, maternally and biparentally inherited characters. *Systematic Biology* **51**: 410–431.
- Lundrigan B, Tucker PK. 1994.** Tracing paternal ancestry in mice, using the Y-linked, sex-determining locus, *Sry*. *Molecular Biology and Evolution* **11**: 483–492.

- Marshall JT. 1981.** Taxonomy. In: Foster HL, Small JD, Fox JG, eds. *The mouse in biochemical research*, Vol. 1. New York: Academic Press, 17–26.
- Mickevich MF, Farris JS. 1981.** The implications of congruence in Menidia. *Systematic Zoology* **30**: 351–370.
- Musser GG, Carleton MD. 1993.** Family Muridae. In: Wilson DE, Reeder DM, eds. *Mammal species of the world*. Washington D.C.: Smithsonian Institution Press, 510–755.
- Posada D, Crandall KA. 1998.** Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**: 817–818.
- Prager EM, Orrego C, Sage RD. 1998.** Genetic variation and phylogeography of Central Asian and other house mice, including a major new mitochondrial lineage in Yemen. *Genetics* **150**: 835–861.
- Prager EM, Tichy H, Sage RD. 1996.** Mitochondrial DNA sequence variation in the eastern house mouse, *Mus musculus*: Comparison with other house mice and report of a 75-bp tandem repeat. *Genetics* **142**: 427–446.
- Rambaut A, Drummond AJ. 2003.** *Tracer v1.0*. <http://evolve.zoo.ox.ac.uk/>
- Sage RD. 1981.** Wild mice. In: Foster HL, Small JD, Fox JG, eds. *The mouse in biochemical research*, Vol. 1. New York: Academic Press, 39–90.
- Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. 1985.** Enzymatic amplification of b-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**: 1350–1354.
- Sandstedt S, Tucker PK. 2004.** Evolutionary strata on the X chromosome in the mouse genus *Mus*. *Genome Research* **14**: 267–272.
- Scott DM, Ehrmann IE, Ellis PS, Agulnik AI, Bishop CE, Simpson E, Mitchell MJ. 1995.** Identification of a mouse male-specific transplantation antigen, H-Y. *Nature* **376**: 695–698.
- She JX, Bonhomme F, Boursot P, Thaler L, Catzeflis F. 1990.** Molecular phylogenies in the genus *Mus*: Comparative analysis of electrophoretic, scnDNA hybridization, and mtDNA RFLP data. *Biological Journal of the Linnean Society* **41**: 83–103.
- Shimodaira H, Hasegawa M. 1999.** Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* **16**: 1114–1116.
- Sorenson MD. 1999.** *Treerot*, Version 2. Boston: Boston University.
- Sourrouille P, Hanni C, Ruedi M, Catzeflis FM. 1995.** Molecular systematics of *Mus crociduroides*, an endemic mouse of Sumatra (Muridae: Rodentia). *Mammalia* **59**: 91–102.
- Suzuki H, Kurihara Y. 1994.** Genetic variation of ribosomal RNA in the house mouse, *Mus musculus*. In: Moriwaki K, Shiroishi T, Yonekawa H, eds. *Genetics in wild mice: its application to biomedical research*. Tokyo: Japan Science Society Press, 107–119.
- Swofford DL. 2002.** *PAUP\*: phylogenetic analysis using parsimony (\*and other methods)*. Sunderland: Sinauer Associates.
- Takeuchi T, Yamazaki Y, Katohfukui Y, Tsuchiya R, Kondo S, Motoyama J, Higashinagawa T. 1995.** Gene trap capture of a novel mouse gene, *Jumonji*, required for neural-tube formation. *Genes and Development* **9**: 1211–1222.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997.** The CLUSTAL windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**: 4876–4888.
- Wang W, Meadows LR, den Haan JMM, Sherman NE, Chen Y, Blokland E, Shabanowitz J, Agulnik AI, Hendrickson RC, Bishop CE, Hunt DF, Goulmy E, Engelhard VH. 1995.** Human H-Y: a male-specific histocompatibility antigen derived from the SMCY protein. *Science* **269**: 1588–1590.
- Yonekawa H, Moriwaki K, Gotoh O, Miyashita N, Matsushima Y, Shi L, Cho WS, Zhen X, Tagashira Y. 1988.** Hybrid origin of Japanese mice '*Mus musculus molossinus*': Evidence from restriction analysis of mitochondrial DNA. *Molecular Biology and Evolution* **5**: 63–78.