Genes Genet. Syst. (2008) **83**, p. 77–88

Mosaic genealogy of the *Mus musculus* genome revealed by 21 nuclear genes from its three subspecies

Yu-Hua Liu[1,4], Aya Takahashi[1], Takashi Kitano[1,5], Tsuyoshi Koide[2], Toshihiko Shiroishi[2], Kazuo Moriwaki[3] and Naruya Saitou[1]*

[1]*Division of Population Genetics, National Institute of Genetics, Mishima, 411-8540, Japan*
[2]*Mammalian Genetics Laboratory, National Institute of Genetics, Mishima, 411-8540, Japan*
[3]*RIKEN Bioresource Center, RIKEN Tsukuba Institute, Tsukuba, 305-0074, Japan*
[4]*Present address: Department of Pathology, University of Washington, Seattle, WA 98915-6340, USA*
[5]*Present address: Department of Biomolecular Function Engineering, Ibaraki University, Hitachi, 316-8511, Japan*

(Received 26 March 2007, accepted 4 November 2007)

Patterns of genetic variation provide insight into the evolutionary history of a species. Mouse (*Mus musculus*) is a good model for this purpose. Here we present the analysis of genealogies of the 21 nuclear loci and one mitochondrial DNA region in *M. musculus* based on our nucleotide sequences of nine inbred strains from three *M. musculus* subspecies (*musculus*, *domesticus*, and *castaneus*) and one *M. spicilegus* strain as an outgroup. The mitochondrial DNA gene genealogy of those strains confirmed the introgression pattern of one *musculus* strain. When all the nuclear DNA data were concatenated to produce a phylogenetic tree of nine strains, *musculus* and *domesticus* strains formed monophyletic clusters with each other, while the two *castaneus* strains were paraphyletic. When each DNA region was treated independently, the phylogenetic networks revealed an unnegligibly high level of subspecies admixture and the mosaic nature of their genome. Estimation of ancestral and derived population sizes and migration rates suggests the effects of ancestral polymorphism and gene flow on the pattern of genetic variation of the current subspecies. Gene genealogies of *Fut4* and *Dfy* loci also suggested existence of the gene flow between *M. musculus* and *M. spicilegus* or other distant species.

Key words: mouse, phylogenetic network, introgression, inbred strains

INTRODUCTION

Allele frequency data of many loci are often averaged to obtain genetic distances between closely related populations such as different subspecies of the same species (e.g., Nei, 1987). Admixture may occur among subspecies, but the signature of admixture will eventually disappear by backcrossing of hybrid individuals in the later generations. Yet the introgression of a small DNA region from one subspecies to another may remain for a relatively long evolutionary time. Therefore, comparison of many gene genealogies of different loci from different subspecies can tell us a detailed evolutionary history of their genomic structure.

Mouse (*Mus musculus*) is an appropriate organism to delineate such a detailed history of a genome. The variation in the genetic background and morphological characters of wild mice has been well documented (e.g., Bonhomme et al., 1984; Moriwaki et al., 1994; Prager et al., 1993; Sage et al., 1993; She et al., 1990). *M. musculus* is genetically classified largely into four subspecies, *M. m. domesticus*, *M. m. musculus*, *M. m. castaneus*, and *M. m. bactrianus* (Bonhomme and Guenet, 1989; Sage et al., 1993; Yonekawa et al. ,1981) with several other minor groups such as *M. m. molossinus* and *M. m. brevirostris* used in this study. They are found in different geographic areas of the world. The subspecies *M. m. domesticus* is known to be indigenous to west Europe and is also found in America, Australia, and various Atlantic and Pacific islands (Boursot et al., 1993; Marshall, 1981; Tichy et al., 1994). *M. m. brevirostris* is treated as a local form of *M. m. domesticus* in this study. *M. m.*

Edited by Yoko Satta
* Corresponding author. E-mail: saitounr@lab.nig.ac.jp
Note: Supplementary materials in this article are at http://wwwsoc.nii.ac.jp/gsj3/sup/83(1)Liu/

musculus has a native range encompassing east Europe and east Asia, while *M. m. castaneus* is found primarily in southeastern Asia (Bonhomme et al., 1994; Boursot et al., 1993; Prager et al., 1993; Sage et al., 1993). The subspecies that is currently grouped as *M. m. bactrianus* inhabits through Iran to Myanmar harboring heterogeneous forms. *M. m. molossinus*, known to have originated from the hybrids of *M. m. musculus* and *M. m. castaneus* (Bonhomme et al., 1989; Yonekawa et al., 1988), inhabits Japan (Yonekawa et al., 1981, 1988). Since this hybrid is mostly *musculus* origin and a very small *castaneus* counterpart, we treat this group as one of the local representatives of *M. m. musculus* in this study. Abe et al. (2004) determined BAC-end sequences of MSM strain of *M. m. molossinus*, and found that there was as high as ca. 1% nucleotide difference between this strain and a commonly used laboratory strain, C57BL/6J, whose genome was derived from mostly *M. m. domesticus*. Thus, leaving out *M. m. bactrianus*, we consider three large subspecies groups, *M. m. domesticus*, *M. m. musculus,* and *M. m. castaneus*, in this study.

Many genetic markers have been used to investigate the origin and the radiation of these house mice as above. There have been studies on chromosomal C-band patterns (Moriwaki et al., 1990), protein electrophoresis (Bonhomme et al., 1984; Boursot et al., 1989; Miyashita et al. ,1985), and DNA RFLP (Ferris et al., 1983a, 1983b; Redi et al., 1990; Suzuki et al., 1986; Yonekawa et al., 1981). Historically, because of the rather conservative morphology of mice, the definition of subspecies is largely based on the combinations of allele frequencies at many nuclear DNA loci (Bonhomme et al., 1984).

The above classification into four subspecies has received good support from the further studies of mtDNA (Boursot et al., 1996; Prager et al., 1998, 1996). However, the studies of Y-chromosome revealed two major Y-chromosome lineages whose distributions are apparently discordant with that of the mtDNA lineages (Bishop et al., 1985; Boissinot and Boursot, 1997). Boissinot and Boursot (1997) suggested that selection has played a role in the rapid spread of Y-chromosome haplotypes across subspecies. This pattern of shared genetic variants across subspecies is also observed in several other loci such as in the class I MHC (Moriwaki et al., 1990) and in the *p53* pseudogene (Ohtsuka et al., 1996), and is expected to be found in many other nuclear DNA genes. After the genome sequence of a laboratory inbred strain, C57BL/6J, was assembled (Mouse Genome Sequencing Consortium, 2002), there have been number of studies which indicated complex origin of the laboratory inbred strain (Wade et al., 2002; Frazer et al., 2004; Zhang et al., 2005). The main focus of these studies has been on the mosaic origin of the laboratory inbred mouse and its relevance to designing QTL mapping and positional cloning experiments. Recently, Baines and Harr (2007) have sequenced 6 X-linked and 7 autosomal loci from multiple population samples of *M. musculus*. Their intention was to contrast the diversity pattern of X-linked and autosomal loci, and they indeed showed that the X-linked diversity is too large to be explained by a simple demographic model in some of the populations. Different from those studies, our purpose is to construct the phylogenies of multiple loci from wild-derived strains and to investigate the clustering pattern for each locus. Also, we want to obtain a rough idea about the ancestral population size and migration rates to understand the evolutionary history of this species complex, particularly the features of population subdivision with migration model.

We sequenced about 1 kb regions of 21 nuclear genes distributed among 16 chromosomes and the control region of mtDNA from nine mouse strains from three *M. musculus* subspecies and one *Mus spicilegus* strain in the present study. Takahashi et al. (2004) analyzed these sequences and detected a positive correlation between the recombination rate and nucleotide diversity within subspecies, and a negative correlation between that and G_{ST}. The same set of nine strains was also used for the ABO blood group gene studied by Yamamoto et al. (2001). We constructed phylogenetic network for each nuclear DNA locus, as well as a phylogenetic tree of concatenated sequences. We also provide rough estimates of the ancestral and derived population sizes and migration rates between *M. m. domesticus* and *M. m. musculus*. Possibility of intersubspecific and interspecific introgressions based on the incongruent gene genealogies is discussed.

MATERIALS AND METHODS

Mouse strains The mouse strains used in this study are listed in Table 1. One old laboratory inbred strain of *M. m. domesticus* (C57BL/10J, abbreviated as B10), eight strains of wild-derived *M. musculus* (PGN2, BFM/2, BLG2, NJL, MSM, SWN, CAST/Ei, and HMI), and one strain of wild-derived *Mus spicilegus* (ZBN) were used. The wild-derived strains were captured from different localities over the world (Table 1), and have been established as inbred strains in the National Institute of Genetics for more than 20 generations (Koide et al., 2000). All of these strains except C57/10J (B10) are wild-derived and the origin is controlled in contrast to some "old inbred" lines whose multiple origin is suspected (Bonhomme et al., 1987). Two *M. m. musculus* strains (BLG2 and NJL) originated near the well-studied hybrid zone between *M. m. musculus* and *M. m. domesticus* in Europe (reviewed in Sage et al., 1993). The source *M. m. domesticus* strains of the DNA sequences from the DDBJ/EMBL/GenBank International Nucleotide Sequence Database used in the analyses are listed in Supplementary Table 1.

Table 1. Inbred mouse strains used for DNA sequencing in this study

Species and subspecies	Strain[a]	Abbreviation	Origin
Mus musculus			
domesticus	B10 (C57BL/10J)	D1	Unknown
	PGN2	D2	Canada
(*brevirostris*)[b]	BFM/2	D3	France
musculus	BLG2	M1	Bulgaria
	NJL	M2	Denmark
(*molossinus*)[c]	MSM	M3	Japan
	SWN	M4	Korea
castaneus	CAST/Ei	C1	Thailand
	HMI	C2	Taiwan
Mus spicilegus (outgroup)	ZBN	S	Bulgaria

[a]: All the strains except B10 are wild-derived and maintained at the National Institute of Genetics, Japan. Details are described in Koide et al. (2000).
[b]: Treated here as a local form of *M. m. domesticus*.
[c]: Originated from hybridization between *M. m. musculus* and *M. m. castaneus* (Bonhomme et al., 1989; Yonekawa et al., 1988). Treated here as one of the representatives of *M. m. musculus*.

Genes sequenced We chose a total of 22 loci for sequence determination, including 19 autosomal loci, one X-linked locus, one Y-linked locus, and a part of the mitochondrial DNA. We chose autosomal loci from different chromosomes rather randomly, though some loci are in the same chromosome. The list of the genes sequenced is shown in Table 2 with the chromosome location and the accession numbers of the sequences used to design primers. The sequenced region and the intron-exon structure for each gene are shown in Supplementary Fig. 1.

PCR and sequencing DNA was extracted from the liver tissue of mice by SDS-proteinase-K method (Sambrook et al., 1989). An internal fragment, approximately 1 kb in size, of each of the 21 nuclear genes and the control region of the mtDNA were sequenced from nine stains of *M. musculus* and one strain of *M. spicilegus* described above, using the PCR-direct sequencing method. A hot start-PCR was carried out in 25 μl reaction volume, containing 2.5 μl of 10x Universal Buffer™ (Nippon Gene), 1.5 μl of 25mM $MgCl_2$, 2.5 μl of Gene Amp

Table 2. The list of the 22 mouse genes analyzed in this study

Symbol	Name	Chromosome Location	Accession Number[a]	Source Strain[b] (DB)	Length (bp) Sequenced[c]
A. *b3GT1*	β1,3-galatosyltranferase-I	2	AF029790	129/svj	917
B. *b3GT2*	β1,3-galatosyltranferase-II	1	AF029791	129/svj	1228
C. *b3GT3*	β1,3-galatosyltranferase-III	3	AF029792	129/svj	996
D. *b3GT4*	GM1/GD1b/GA1 synthase	17	AF082504	– [d]	1112
E. *BNP*	Brain natriuretic peptide	4	D82049	129/sv	1112
F. *CD14*	Monocyte and granulocyte cell surface glycoprotein homolog	18	M34510	BALB/c	1254
G. *Cramp*	Cathelicidin (Cramp) gene	9	AF035680	129/svj	987
H. *Dfy*	Duffy-like blood group	1	AF016697	B6CBA/F1	1509
I. *Fau*	Cellular homolog of the fox sequence of FBR-MuSV	19	L33715	BALB/c	1241
J. *fisp-12*	Growth factor inducible gene	10	M70641	– [d]	1156
K. *Fut1*	α (1,2)fucosyltransferase	7	Y09883	ICR	1134
L. *Fut2*	α (1,2)fucosyltransferase	7	AF064792	ICR	963
M. *Fut4*	α (1,3)fucosyltransferase	9	D63380	129/sv	1170
N. *GdX*	Housekeeping protein DXS254E	X	J04761	BALB/c	1013
O. *Hox-1.11*	Homeobox protein 1.11	6	M93148	ICR	907
P. *MECL1*	Proteasome subunit MECL1	8	D85562	BALB/c	652
Q. *sec1*	α (1,2)fucosyltransferase	7	Y09882	ICR	1107
R. *Sox15*	HMG-box transcription factor 15	11	AB014474	129	1340
S. *TNF*	Tumor necrosis factor	17	M20155	– [d]	1201
T. *Tspy*	Testis-specific Y-encoded protein	Y	AJ001379	C57BL/6	1284
U. *Wnt-1*	Mammary proto-oncogene	15	K02593	BALB/c	1044
mtDNA	Mitochondrial DNA		J01420	LA9 or 111-OB3 (L-cell lines)	975

[a]: Accession numbers of the sequences used to design primers for PCR.
[b]: Source strains of the sequences used to design primers for PCR. Each of these strains are represented as "(DB)" in Figs. 3, 4, and 5.
[c]: Size of the sequenced fragment.
[d]: Information not provided.

dNTP Mix™ (20 mM), 0.5U of AmpliTaq Gold™ (Applied Biosystems), 12.5 pmol of primer, and 5–20 ng genomic DNA. Annealing temperature was between 50°C and 65°C, depending on the primers. PCR products were purified by MicroSpin columns S-300HR™ (Amersham Pharmacia Biotech). Sequencing reactions were performed using the Big Dye Terminal Cycle Sequencing Ready Reaction Kit™, and DNA sequence data were obtained by using ABI PRISM310™ and ABI PRISM 377™ DNA automatic sequencer (Applied Biosystems). PCR primers used in this study are listed in Supplementary Table 1.

Sequence alignment and phylogenetic analyses CLUSTAL W version 1.6 (Thompson et al., 1994) was used for multiple sequence alignment. The neighbor-joining method (Saitou and Nei. 1987) was used for constructing phylogenetic trees. Phylogenetic networks (see Bandelt, 1994) were constructed manually, as done for the primate ABO blood group gene sequences (Saitou and Yamamoto, 1997). Alignment gaps were not used for tree and network construction. The phylogenetic networks have the advantage of showing all the possible pathways of nucleotide changes between sequences, thus, visualize more sequence information about the discordant partition among nucleotide positions compared to the phylogenic trees (Saitou, 1996; Saitou and Yamamoto, 1997).

Estimation of demographic parameters The population genetic parameters (ancestral and derived population sizes, population split time, and migration rates) included in the "isolation with migration model" was estimated using the method suggested by Hey and Nielsen (2007). The computer program "IMa", provided by these authors were utilized to conduct the Marcov chain Monte Carlo (MCMC) simulations. First, 10,000 genealogies were generated from the multilocus sequence data after the initial burn-in period of 100,000 steps. Then, the joint parameter estimate was conducted from these saved genealogies. Convergence by the Markov chain simulations was assessed by monitoring multiple independent chains at different starting points and by assessing the autocorrelation of the parameter values over the course of the runs. Each locus was assigned an inheritance scalar to adjust for its relative effective population size: 1.0 for autosomal loci, 0.75 for X-linked and 0.25 for Y-linked loci.

RESULTS

Determined Nucleotide Sequences A total of approximately 243 kb sequences were determined from 22 loci of nine *M. musculus* strains and one *M. spicilegus* strain. There were no heterozygous sites observed in any of the loci from each inbred strain. DDBJ/EMBL/GenBank International Nucleotide Sequence Database accession numbers for those sequences are AB039044-AB039263.

Phylogenetic trees based on mitochondrial and nuclear DNA sequences The 975 bp of the control region and its flanking tRNA gene of mouse mtDNA (see Supplementary Fig. 1V) were sequenced. The neighbor-joining tree of this locus from sequences of the nine *M. musculus* strains and a sequence from the nucleotide sequence database (designated as "DB"), rooted by an outgroup sequence from *M. spicilegus*, is shown in Fig. 1. The strains that belong to the same subspecies cluster together by branches with high bootstrap values (> 90%) except for one strain of *M. m. musculus* (NJL), which clustered more closely with *M. m. domesticus* strains than with themselves (Fig. 1). Comparing the above sequences with those in Prager et al. (1993, 1996) as described in detail below, the interpretation of this exception would be that the sequences of BLG2 (M1), MSM (M3), and SWN (M4) represent the authentic *M. m. musculus* lineage and

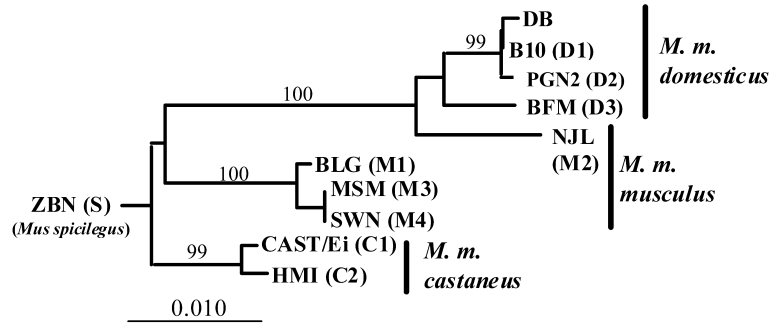


Fig. 1. The neighbor-joining tree constructed from mtDNA sequences of nine inbred strains from three *Mus musculus* subspecies. One sequence from the database (J01420) designated as "DB" is also included as a *M. m. domesticus* sequence. The tree is rooted by *Mus spicilegus* sequence (ZBN). The numbers on the interior branches are bootstrap probabilities (%) based on 1000 bootstrap resampling (only those higher than 90% are shown). The scale bar indicates the number of nucleotide substitutions per site.

that of NJL (M2) represents an *domesticus* mtDNA introgressed at the hybrid zone. This mtDNA status of northern Denmark mice has been described earlier (Ferris et al., 1983a; Gyllensten and Wilson, 1987; Vanlerberghe et al., 1988).

Six of the nine mtDNA sequences determined in this study are identical to some known haplotypes determined by Prager et al. (1993, 1996). The sequences from B10 (D1) and PGN2 (D2) are identical to *M. m. domesticus* mtDNA type 1 (U47430) and type 2 (U47431), respectively. MSM and SWN are both identical to *M. m. musculus* mtDNA type 34 (U47531). CAST/Ei (C1) was identical to *M. m. castaneus* mtDNA type 1 (U47534), and NJL (M2) was identical to *M. m. domesticus* mtDNA type 27 (U47455). The 75-bp tandem repeat reported to be present in some strains of *M. m. musculus* (Prager et al., 1996) was observed in strain BLG2 (M1) but not in other strains.

We constructed the neighbor-joining tree for the concatenated sequence of all the 21 nuclear loci (Fig. 2). Its clustering pattern of the three subspecies is somewhat different from that for mtDNA tree (Fig. 1). All four *M. m. musculus* strains (M1–M4) cluster together with high bootstrap values in contrast to the mtDNA pattern, while two *M. m. castaneus* strains (C1 and C2) did not form a cluster.

Clustering of strains belonging to the same subspecies Figure 3 shows the phylogenetic network of each nuclear locus. Inconsistent clustering patterns are commonly seen among these networks in Fig. 3. The two strains of *M. m. castaneus* (C1 and C2) do not cluster together in the networks of *b3GT3*, *Dfy*, *Fau*, and *TNF* (Figs. 3C, 3H, 3I, and 3S), while their sequences were identical to each other in six other loci, *b3GT1, CD14, fisp-12, Fut4, Hox-1.11, and Tspy* (Figs. 3A, 3F, 3J, 3M, 3O, and 3T). One *M. m. musculus* strain (M2) that revealed a *musculus* to *domesticus* introgression event in the mtDNA clustered together with other *M. m. musculus* strains in many of the nuclear loci. The same clustering pattern as that of mtDNA for the M2 strain is observed only at the *Fut1* locus (Fig. 3K). *Hox-1.11* sequences were identical in all strains except for M2, which is different at one nucleotide site (Fig. 3O). The *BNP* and *sec1* loci show complicated networks suggesting past recombination events within and among subspecies (Figs. 3E and 3Q). Hence these networks indicate that the mouse genome is highly heterogeneous in terms of subspecies genealogy.

Partial sequences of *Tspy*, a Y-chromosome gene that has been determined as a pseudogene in mouse (Mazeyrat and Mitchell, 1998), were obtained from all the 10 strains (Fig. 3T). There were four Y-chromosome haplotypes depicted from this locus in these strains. The two major Y-chromosome lineages in Boissinot and Boursot (1997), one found in *M. m. domesticus* and the other found in *M. m. musculus* and *M. m. castaneus*, are not apparent from the present study. The haplotype of the ZBN, a *M. spicilegus* strain, was identical to one of the four types present in *M. musculus* strains. A similar pattern was seen in other murine species at this locus (Schubert et al., 2000).

Most of the laboratory mice strains are considered to have originated from *domesticus* strains. If we include sequences retrieved from database (designated as DB in Fig. 3; see Table 1 for specific strain names used) to *domesticus* subspecies, most of the loci showed clustering of the strains within this locus, except for two loci. The DB sequence is one nucleotide different from the three identical *domesticus* sequences (D1–D3) for the *Fut2*

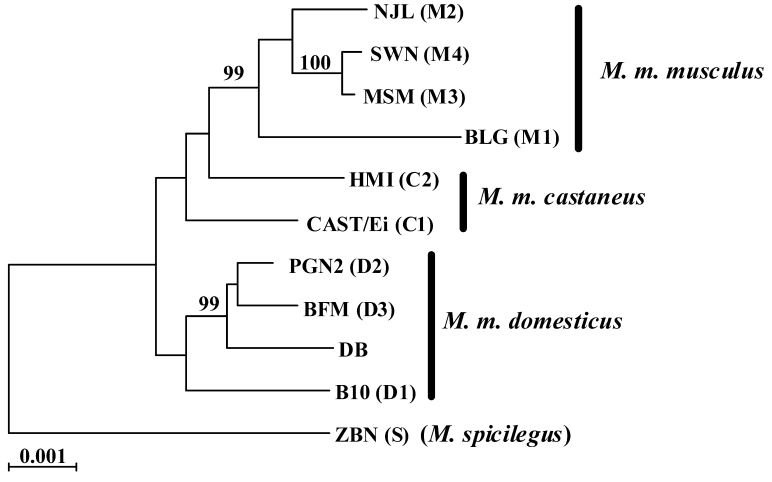
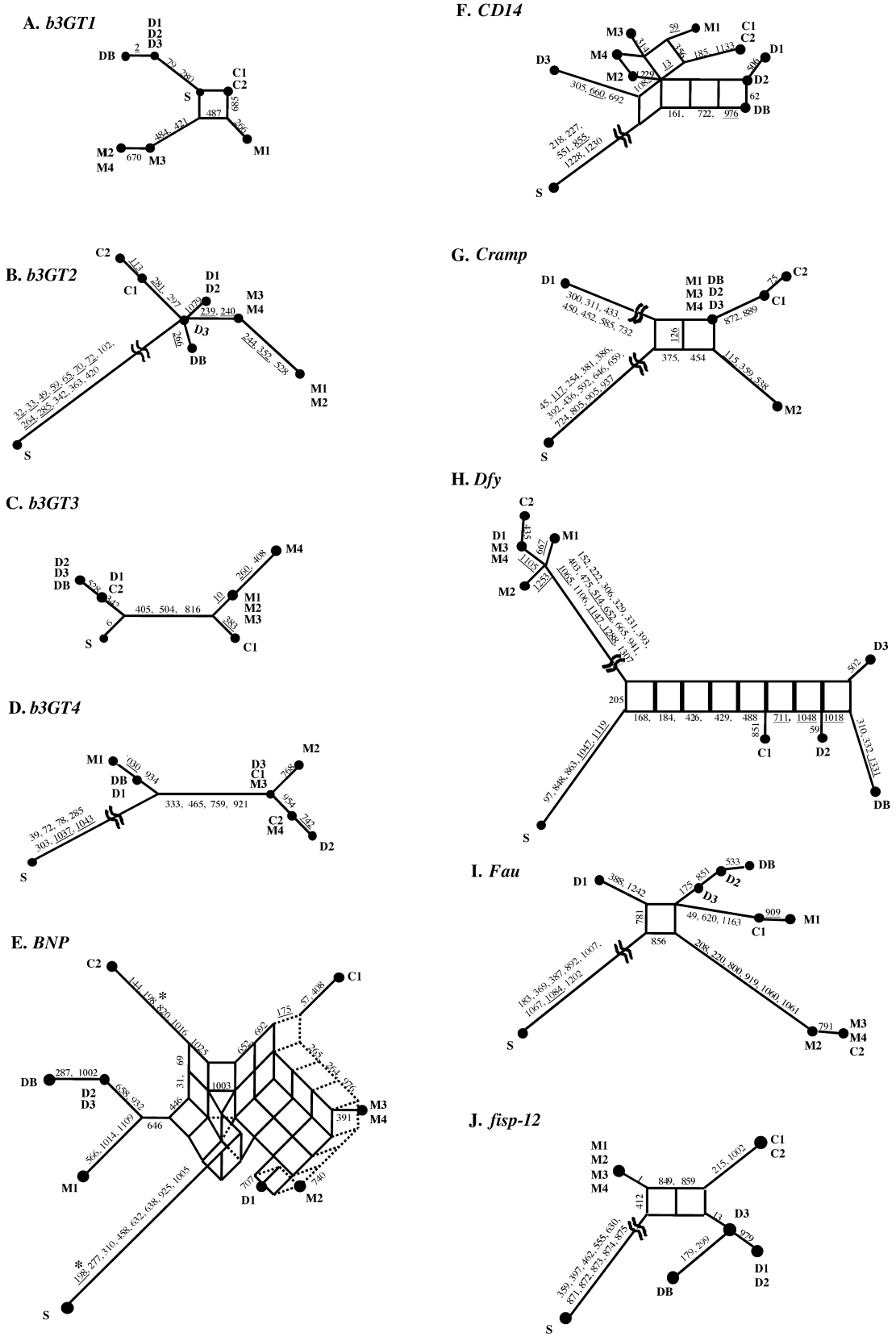


Fig. 2. The neighbor-joining tree constructed from the concatenated sequence of all the 21 nuclear loci of nine inbred strains from three *Mus musculus* subspecies, rooted by *Mus spicilegus* sequence (ZBN). The sequences retrieved from the database described in Table 2 are also concatenated as *M. m. domesticus* sequences. All the notations are the same as in Fig. 1.


A. b3GT1
B. b3GT2
C. b3GT3
D. b3GT4
E. BNP
F. CD14
G. Cramp
H. Dfy
I. Fau
J. fisp-12

(continued)

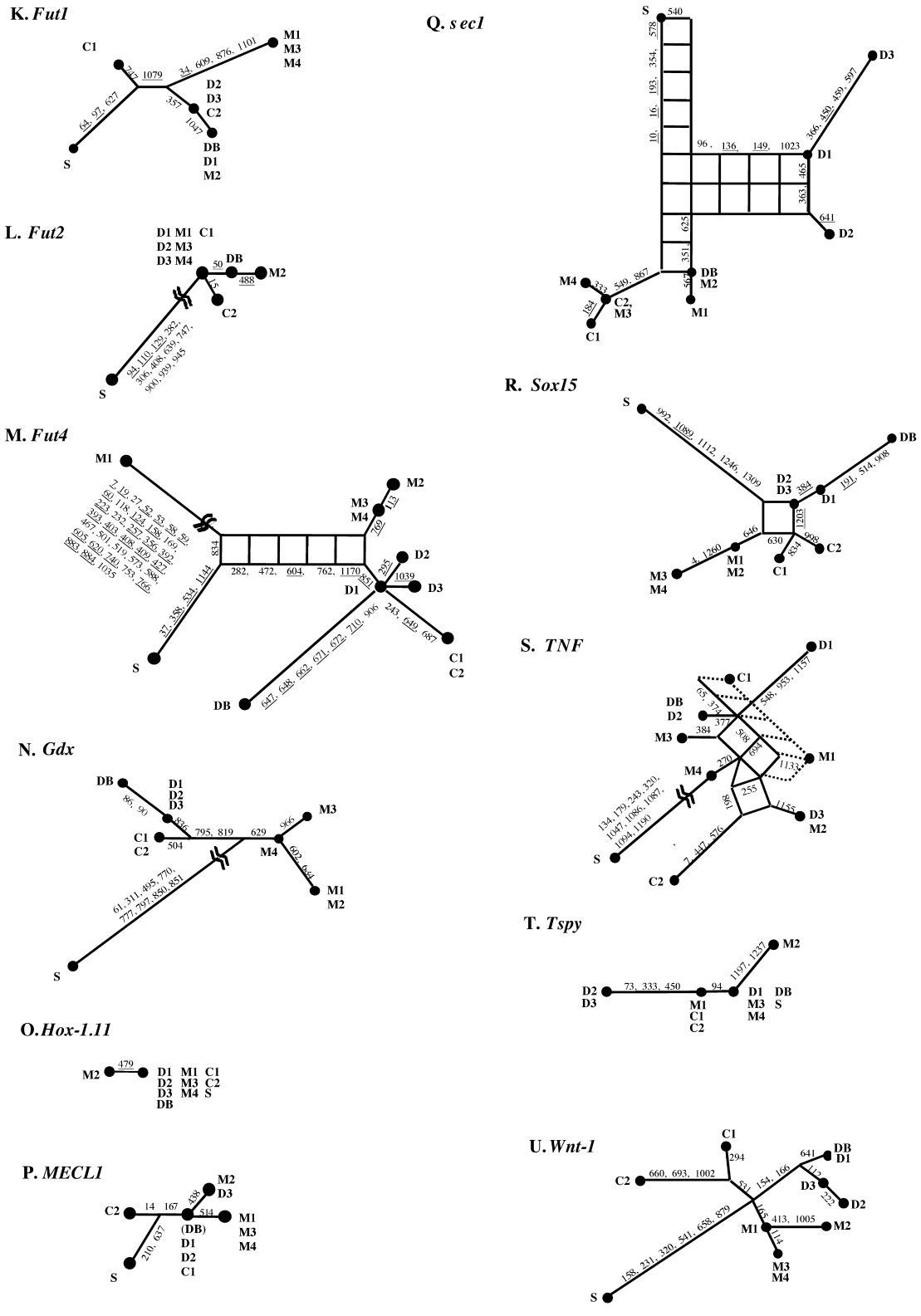


Fig. 3. The phylogenetic networks of the 21 nuclear loci in *Mus musculus*. The full circles denote sequences of each inbred mouse strain from three subspecies. Sequences from one *Mus spicilegus* strain (S) are also included in the networks. The numbers on the edges show variant nucleotide positions responsible for corresponding splits (see multiple alignment data presented in our website for nucleotide positions). The underlined numbers indicate positions of the non-synonymous nucleotide changes. The numbers indicated by * in the BNP gene is interpreted as two independent substitution events. "DB" indicates the sequence of unknown *M. m. domesticus* strain in the DDBJ/EMBL/GenBank database (see Table 2).

locus, and the DB sequence is identical with one of *musculus* strain (M2) for the locus *sec1* (Fig. 3Q). This mosaic nature of the inbred laboratory mouse strains observed in this study was also reported by other authors (Wade et al., 2002; Frazer et al., 2004; Zhang et al., 2005).

One strain (S or ZBN; see Table 1) of *M. spicilegus* was used as outgroup in the present study. We expect that the branch leading to strain S is much longer than those connecting sequences of *M. musculus* strains. Many loci in fact showed this expected pattern (see Fig. 3). However, some loci showed unexpected patterns between *M. musculus* and *M. spicilegus*. The *b3GT3* and *Fut1* had very short branches leading to strain S (Figs. 3C and 3K), while sequences of strain S falled within the variation of *M. musculus* sequences in *b3GT1* and *Tspy* (Figs. 3A and 3T). The sequence of strain S of *Hox-1.11* falls within the variation of *M. musculus* sequences (Fig. 3O), though the pattern may be due to very low evolutionary rate of this locus.

Strain M1 (BLG of *M. m. musculus*) exhibits a clear anomaly at the *Fut4* locus, i.e. considerably larger number (35) of nucleotide changes accumulated than that (4) in the outgroup sequence S (Fig. 3M). The remaining three strains (M2–M4) of the *musculus* subspecies are similar with each other, and 5 other strains of *M. musculus* also cluster with these *musculus* subspecies strains. Various mechanisms that can explain this pattern will be discussed in Discussion.

We observed yet another unusual pattern in the *Dfy* locus (Fig. 3H). The phylogenetic network of this locus shows a large divergence within *M. musculus*. Seventeen nucleotide differences divide the ten strains into two clusters, one of which contains sequences of C2, D1, M1, M2, M3, and M4, while the other contains those of C1, D2, D3 and DB. Possible mechanisms that can explain this pattern will be discussed in Discussion.

Parameter estimation In order to obtain quantitative measures of the level of admixture, we estimated population genetic parameters (ancestral and derived population sizes, population split time, and migration rates) included in the "isolation with migration model" (Hey and Nielsen, 2007). This model captures demographic phenomena that occur when one ancestral population splits into two descendant populations. The three populations could differ in size, but the model does not implement population size change. Gene exchanges are allowed during the time since population split.

We used 4 strains each from two subspecies, *M. m. musculus* and *M. m. domesticus*, for the analyses. A total of concatenated introns from 9 sequenced genes (*Cramp, Fau, fisp12, Gdx, Hox-1.11, MECL1, Sox15, Tspy, Wnt-1*) were subjected to the multilocus analyses of the joint parameter estimates using MCMC simulations. Note that since Hey and Nielsen (2007)'s model assumes no recombination within loci, we had excluded introns that showed reticulations in their phylogenetic networks. Repeated runs of the IMa computer program revealed unambiguous marginal posterior probability distributions of the parameters except for the population split time, t (Supplemental Fig. 2). The peak and the range of the parameters are summarized in Table 3. The output of the program gives estimated parameters scaled by per locus mutation rate. We used the mutation rate $u = 3.94 \times 10^{-7}$ per locus per generation calculated from the average divergence between *M. musculus* and *M. spicilegus* of the nine loci used in the analyses.

The 90% Highest Posterior Density (HPD) intervals of the parameter estimates are large due to complex lineage sorting and limited data (Table 3). However, the locations of peaks in the marginal posterior probability distributions in Table 3 suggest that the ancestral population size (N_A) of *M. m. domesticus* and *M. m. musculus* maybe ~3–8 fold larger than the population sizes of the current subspecies (N_1 and N_2). The range of population migration rates ($4N_1m_1$ and $4N_2m_2$) could not be obtained from the analyses, however, 90% HPDs of per generation migration rates (m_1 and m_2) exclude zero.

Table 3 Estimates of the demographic parameters obtained by the method proposed by Hey and Nielsen (2007)

Parameters[a]	N_1	N_2	N_A	$4N_1m_1$	$4N_2m_2$	t[d]
Peak[b]	7.8×10^4	3.6×10^4	3.1×10^5	0.017	0.119	–
Lower 90% HPD[c]	1.9×10^4	7.6×10^3	5.1×10^3			–
Upper 90% HPD[c]	4.6×10^5	1.7×10^5	3.1×10^6			–

[a]: N_1, N_2, and N_A are effective population sizes of current *M. m. domesticus*, *M. m. musculus*, and their ancestral population, respectively. $4N_1m_1$ and $4N_2m_2$ are the population migration rates of *M. m. domesticus* and *M. m. musculus*, respectively.

[b]: Locations of peaks in the marginal posterior probability distributions (Supplementary Fig. 2).

[c]: Highest Posterior Density (HPD) intervals are the shortest spans that contain 90% of the posterior probability.

[d]: Unambiguous marginal posterior probability distribution could not be obtained for the population split time, t (Supplemental Fig. 2).

DISCUSSION

A High level of mice subspecies admixture The mosaic genealogy of genes among subspecies in the nuclear DNA could have occurred via either or both of the two factors besides the possibility of any selective force involved. First is the large ancestral population size that harbored many variable sites. A major proportion of the genome admixture could have happened in the ancestral population before the divergence of the subspecies. A large ancestral population and a short divergence time could produce such incomplete lineage sorting, i.e. mosaic genealogy.

The second possible factor is the vague reproductive barrier among the current subspecies. In mice, despite capable hybridization in captivity, the reproductive barrier among subspecies is documented well in the hybrid zone studies (reviewed in Sage et al., 1993), which may be attributed to the potential fitness reduction of the hybrids such as the increased loads of intestinal parasites (Moulia et al.. 1993; Sage et al., 1986). However, the information is scarce in other regions of the world. It is of our interest to know to what extent this barrier has affected the genetic structure of the current *M. musculus* species genome.

Phylogenic networks obtained from 21 nuclear DNA regions of the *M. musculus* genome (Fig. 3) suggest that the genetic exchange at the subspecies level is fairly common in this species. Our data confirms the existence of imperfect separation of the peripheral subspecies according to their nuclear genes and the mosaicism in this species complex implicated earlier (Bonhomme et al., 1994; Wade et al., 2002; Frazer et al., 2004; Zhang et al., 2005). Those networks of the loci showed that the strains of the same subspecies, especially those of *M. m. castaneus* and *M. m. domesticus*, do not cluster together closely in some genes (Fig. 3, see also Result).

Estimated parameters of population size and migration In addition to the observation of the clustering pattern in the phylogenetic networks (Fig. 3), we obtained a rough estimates of the demographic parameters by the method proposed by Hey and Nielsen (2007). The method is useful for analyzing a pair of closely related populations or species such as subspecies in our study. The estimated parameters in Table 3 indicate large ancestral population size with frequent migration, although the ranges of estimates are large. Recently, Ideraabdullah et al. (2004) showed that the extent of ancient polymorphism is substantial among the wild-derived inbred strains of *M. musculus*. Thus, the mosaic nature of the *M. musculus* genome is likely to be due to both ancestral polymorphism and migration among subspecies.

Eyre-Walker et al. (2002) has estimated the effective population size (N_e) of *M. m. domesticus* to be 1.6×10^5 or 2.9×10^5 depending on different divergence time estimates between *M. m. domesticus* and *M. caroli*. Baines and Harr (2007) estimated N_e of Iranian population of *M. m. domesticus* to be $4.4 \times 10^5 - 7.9 \times 10^5$. Those N_e estimates are not directly comparable to our estimates of N_1 and N_2 based on the "isolation with migration model" (Hey and Nielsen, 2007), because in this model, implementation of ancestral polymorphism and migration allows N_1 and N_2 to take lower values. Nevertheless, our population size estimates maybe slightly under estimated because we intentionally chose loci that had no reticulation in the phylogenetic network to avoid recombination within loci. This may have caused the sampled genealogies to coalesce faster than other average genomic regions (Hey and Nielsen, 2004). Indeed the loci we chose had lower average nucleotide diversity compared to the average silent π of the 19 autosomal loci used in this study (Takahashi et al., 2004). Nucleotide diversity, divergence between *M. musculus* and *M. spicilegus*, and population divergence (G_{ST}) of those 19 loci are listed in Takahashi et al. (2004). It should also be noted that the current estimate of N_A using only two subspecies maybe smaller than the ancestral population size of the whole *M. musculus* subspecies. In any case, these population size estimates need to be treated with caution because of the small sample size.

Unique position of the BLG2 strain in the *Fut4* locus We observed that the BLG2 (M1) strain of the *M. m. musculus* was very different from other strains in the *Fut4* locus (Fig. 3M). There may be four possible explanations for the unique position of this strain. First is the acceleration of the evolutionary rate due to directional natural selection or loss of function. The second possibility is the gene conversion from other paralogous genes. The third is introgression of a distant species, far apart from *M. speciligus*, used in this study as an outgroup species. The fourth one is long-term coexistence of two distinct lineages through balancing selection.

We predict a high ratio of non-synonymous to synonymous substitutions in the BLG2 lineage under the first possibility. The MK test (McDonald and Kreitman, 1991) was performed to investigate whether there had been particularly strong positive selection or relaxation of selection on the M1 lineage. In this study, the test was applied to the intraspecific data. Instead of comparing a sequence from another species versus multiple sequences from one species in the ordinary MK test, we compared the M1 sequence versus the remaining conspecific sequences of the *Fut4* locus. There are 26 non-synonymous and 14 synonymous substitutions that are not shared between M1 and any of the other *M. musculus* strains, whereas there are 12 non-synonymous and four synonymous sites that are variable within other *M.*

musculus strains. Application of the MK test showed that there is no particular acceleration in amino acid substitutions over synonymous substitutions in the M1 lineage compared to the rate within other lineages (Fisher's exact test; $p > 0.10$).

To examine the second possibility, we conducted a BLAST search (Altschul et al., 1990) of the latest DDBJ/EMBL/GenBank International Nucleotide Sequence Database using the M1 *Fut4* sequence as the query. However, we found only the mouse *Fut4* gene itself, and no other regions of the completely sequenced mouse genome (Mouse Genome Sequencing Consortium, 2002) can be aligned with the M1 *Fut4*. Therefore, there seems to exist no gene paralogous to the *Fut4* gene.

Therefore, we are left with the last two possibilities: introgression or balancing selection. Unfortunately, however, there is no way to test these hypotheses at this moment. The nucleotide sequence of the donor species of introgression should be investigated in the future. When balancing selection operates, two allelic lineages are expected to coexist, as observed for the *Dfy* locus (see Fig. 3H), as we discuss below. Therefore, we are inclined to choose the introgression hypothesis for this locus, because only one strain (M1) showed anomaly in the *Fut4* locus.

Existence of the two divergent lineages in the *Dfy* locus The mouse *Dfy* gene, homolog of human *FY* (gene for Duffy blood group), is a member of the superfamily of chemokine receptors (Luo et al., 1997). Two distinct alleles exist among the 10 mice strains, and both variants are present in at least two of the three subspecies. (Fig. 3H). We can consider three nonexclusive possibilities responsible for this peculiar pattern of *Dfy* sequences; (1) they have been maintained by some kind of a balancing selection through the divergence of the subspecies, (2) introgression of alleles from one subspecies to other, and (3) the ancestral polymorphism is kept to present-day populations. If balancing selection is quite strong, both lineages may coexist in many local populations of all the subspecies. This can be examined in future. Possibilities (2) and (3) are not easy to be distinguish, for they are not exclusive.

Introgression between *M. musculus* and other species Our data suggest that even the reproductive barrier between *M. musculus* and other species is incomplete. A rough estimation of the divergence time between *M. musculus* and *M. spicilegus* without considering the complicated lineage sorting becomes about 2.4 Myr, using the mutation rate $\lambda = 4.8 \times 10^{-9}$ per site per year in introns of rodent lineages (Li et al., 1996). The divergence time (T) is obtained under the assumption of rate constancy, where the evolutionary distance $d = 2\lambda T$. The average number of synonymous substitutions per site (d_s; Ina, 1995) between most distant pairs of strains in the two species was used as d. The divergence time estimate between M1 (BLG2) and M3 (MSM) in *Fut4* is about 5.7 Myr, which is more than two times larger than the estimated *musculus – spicilegus* divergence time.

Difference in genetic distances between *M. musculus* and *M. spicilegus* also indicates the recent introgression from an outgroup species. Two loci show very short branches leading to an outgroup S (Figs. 3C and 3K), and the sequence from S falls within the variation of *M. musculus* sequences in three loci (Figs. 3A, 3O, and 3T). The divergence time between *M. musculus* and *M. spicilegus* is estimated to be about 1.1Myr (She et al., 1990) or larger (Moriwaki et al., 1994; see also our estimate above). From these estimates, although we need again to consider the effect of ancestral polymorphism, the observation that about 25% of the loci (five out of 21) show the above pattern seems substantial, and indicates that the reproductive barrier between even different species of genus *Mus* is lower than one might have imagined.

In conclusion, an unnegligibly high level of subspecies admixture and the mosaic pattern of the genome were found from *M. musculus* by comparing 21 nuclear gene genealogies. This pattern is likely to be formed by ancestral polymorphism and frequent migration. Unique genealogy patterns of some loci could be resulted from introgression from other species and/or balancing selection. The nucleotide sequences of multiple loci sampled from many genomic regions in this study provide informative genealogical data for obtaining complicated picture of the evolutionary history of this species.

Credits. This study was planned by N. Saitou. All the mouse samples were provided by T. Koide, T. Shiroishi, and K. Moriwaki. Sequencing was done by Y.-H. Liu for 18 loci and by T. Kitano for 4 loci. Initial sequence analyses were conducted by Y.-H. Liu and later extended by A. Takahashi. Manuscripts were written by Y.-H. Liu, A. Takahashi, and N. Saitou.

This study was partially supported by grants-in-aid for scientific studies from the Ministry of Education, Science, Sport, and Culture, Japan, to N.S., the COE foreign visiting researcher fellowship to Y.H.L, and JSPS Postdoctoral Fellowship to A.T. We appreciate Ms. H. Kobayakawa for her help in the laboratory, and Drs. R. Noda and K. Kryukov for their help on the data analyses.

REFERENCES

Abe, K., Noguchi, H., Tagawa, K., Yuzuriha, M., Toyoda, A., Kojima, T., Ezawa, K., Saitou, N., Hattori, M., Sakaki, Y., Moriwaki, K., and Shiroishi, T. (2004) Contribution of Asian mouse subspecies *Mus musculus molossinus* to genomic constitution of strain C57BL/6J, as defined by BAC end sequence-SNP analysis. Genome Res. **14**, 2239–2247.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. J. Mol. Biol. **215**, 403–410.

Baines, J. F., and Harr, B. (2007) Reduced X-linked diversity in derived populations of house mice. Genetics **175**, 1911–1921.

Bandelt, H. J. (1994) Phylogenetic networks. Verh. Naturewiss. Ver. Hamburg. (NF). **34**, 51–71.

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

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

Bonhomme, F., Anand, R., Darviche, D., Din, W., and Boursot, P. (1994) The house mouse as a ring species? In: Genetics in Wild Mice (eds.: K. Moriwaki, T. Shiroishi, and H. Yonekawa) pp. 13–23. Japan Sci. Soc. Press, Tokyo/S Karger, Basel.

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

Bonhomme, F., and Guenet, J.-L. (1989) The wild house mouse and its relatives. In: Genetic Variants and Strains of the Laboratory Mouse (eds.: M. F. Lyon, and A. G. Searle), pp. 649–662. Gustav Fischer Verlag, Stuttgart.

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

Bonhomme, F., Miyashita, N., Boursot, P., Catalan, J., and Moriwaki, K. (1989) Genetical variation and polyphyletic origin in Japanese *Mus musculus*. Heredity **63**, 299–308.

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

Boursot, P., Bonhomme, F., Catalan, J., and Moriwaki, K. (1989) Variations of a Y chromosome repeated sequence across subspecies of *Mus musculus*. Heredity **63**, 289–297.

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

Eyre-Walker, A., Keightley, P. D., Smith, N. G. C., and Gaffney, D. (2002) Quantifying the slightly deleterious mutation model of molecular evolution. Mol. Biol. Evol. **19**, 2142–2149.

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

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

Frazer, K. A., Wade, C. M., Hinds, D. A., Patil, N., Cox, D. R. and Daly, M. J. (2004) Segmental phylogenetic relationships of inbred mouse strains revealed by fine-scale analysis of sequence variation across 4.6Mb of mouse genome. Genome Res. **14**, 1493–1500.

Gyllensten, U., and Wilson, A. C. (1987) Interspecific DNA transfer and the colonization of Scandinavia by mice. Gen. Res. **49**, 25–29.

Hey, J., and Nielsen, R. (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics **167**, 747–760.

Hey, J., and Nielsen, R. (2007) Integration within Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proc. Natl. Acad. Sci. USA **104**, 2785–2790.

Ideraabdullah, F. Y., de la Casa-Esperon, E., Bell, T. A., Detwiler, D. A., Magnuson, T., Sapienza, C., and de Villena, F. P.-M. (2004) Genetic and haplotype diversity among wild-derived mouse inbred strains. Genome Res. **14**, 1880–1887.

Ina, Y. (1995) New methods for estimating the numbers of synonymous and nonsynonymous substitutions. J. Mol. Evol. **40**, 190–226.

Koide, T., Moriwaki, K., Ikeda, K., Niki, H., and Shiroishi, T. (2000) Multi-phenotype behavioral characterization of inbred strains derived from wild stocks of *Mus musculus*. Mamm. Genome **11**, 664–670.

Li, W. H., Ellsworth, D. L., Krushkal, J., Chang, B. H., and Hewett-Emmett, D. (1996) Rates of nucleotide substitution in primates and rodents and the generation-time effecthypothesis. Mol. Phylogenet. Evol. **5**, 182–187.

Luo, H., Chaudhuri, A., Johnson, K. R., Neote, K., Zbrzezna, V., He, Y., and Pogo, A. O. (1997) Cloning, characterization, and mapping of a murine promiscuous chemokine receptor gene: homolog of the human Duffy gene. Genome Res. **7**, 932–941.

Marshall, J. T. (1981) Taxonomy. In: The Mouse in Biomedical Research (eds.: H. L. Forter, H. J. D. Small, and J. G. Fox), pp. 17–26. Academic Press, New York.

Mazeyrat, S., and Mitchell, M. J. (1998) Rodent Y chromosome *TSPY* gene is functional in rat and non-functional in mouse. Hum. Mol. Genet. **7**, 557–562.

McDonald, J. H., and Kreitman, M. (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature **354**, 114–116.

Miyashita, N., Moriwaki, K., Minezawa, M., Yonekawa, H., Bonhomme, F., Migita, S., Yu, Z. C., Lu, D. Y., Cho, W. S., and Thohari, M. (1985) Allelic constitution of the hemoglobin beta chain in wild populations of the house mouse *Mus musculus*. Biochem. Genet. **23**, 975–986.

Moriwaki, K., Shiroishi, T., and Yonekawa, H. (1994) Genetics in Wild Mice. Jpn. Sci. Soc. Press, Tokyo/S Karger, Basel.

Moriwaki, K., Sagai, T., Shiroishi, T., Bonhomme, F., Chenghuai, W., Xinquao, H., Meilei, J., and Zhengan, W. (1990) Mouse subspecies differentiation and H-2 polymorphism. Biol. J. Linnean Soc. **41**, 125–139

Moulia, C., le Brun, N., Dallas, J., Orth, A., and Renaud, F. (1993) Experimental evidence of genetic determinism in high susceptibility to intestinal pinworm infection in mice: a hybrid zone model. Parasitol. **106**, 387–393.

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

Nei, M. (1987) Molecular Evolutionary Genetics. Columbia University Press, New York.

Ohtsuka, H., Oyanagi, M., Mafune, Y., Miyashita, N., Shiroishi, T., Moriwaki, K., Kominami, R., and Saitou, N. (1996) The presence/absence polymorphism and evolution of the p53 pseudogene in the genus *Mus*. Mol. Phylogenet. Evol. **5**, 548–556.

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

Prager, E. M., Sage, R. D., Gyllensten, U., Kelley, T., W., Hübner, R., Jones, C. S., Noble, L., Searle, J. B., and Wilson, A. C.

(1993) Mitochondrial DNA sequence diversity and the colonization of Scandinavia by house mice from East Holstein. Biol. J. Linnean Soc. **50**, 85–122.

Prager, E. M., Tichy, H., and Sage, R. D. (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.

Redi, C. A., Garagna, S., Della, Valle, G., Bottiroli, G., Dell'Orto, P., Viale, G., Peverali, F. A., Raimondi, E., and Forejt, J. (1990) Differences in the organization and chromosomal allocation of satellite DNA between the European long tailed house mice *Mus domesticus* and *Mus musculus*. Chromosoma. **99**, 11–17.

Sage, R. D., Atchley, W. R., and Ernesto, C. (1993) House mice as models in systematic biology. Syst. Biol. **42**, 523–561.

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

Saitou, N. (1996) Reconstruction of gene trees from sequence data. Methods Enzymol. **266**, 427–449.

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

Saitou, N., and Yamam oto, F. (1997) Evolution of primate ABO blood group genes and their homologous genes. Mol. Biol. Evol. **14**, 399–411.

Sambrook, J., Fritisch, E. F., and Maniatis, T. (1989) Molecular Cloning: a Laboratory Manual. Cold Spring Habor Laboratory Press, Cold Spring Harbor, NY.

Schubert, S., Dechend, F., Skawran, B., Krawczak, M., and Schmidtke, J. (2000) Molecular evolution of the murine *tspy* genes. Cytogenet. Cell Genet. **91**, 239–242.

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

Suzuki, H., Miyashita, N., Moriwaki, K., Kominami, R., Muramatsu, M., Kanehisa, T., Bonhomme, F., Petras, M. L., Yu, Z. C., and Lu, D. Y. (1986) Evolutionary implication of heterogeneity of the nontranscribed spacer region of ribosomal DNA repeating units in various subspecies of *Mus musculus*. Mol. Biol. Evol. **3**, 126–137.

Takahashi, A., Liu, Y.-H., and Saitou, N. (2004) Genetic variation versus recombination rate in a structured population of mice. Mol. Biol. Evol. **21**, 404–409.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (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.

Tichy, H., Zaleska-Rutczynska, Z., O'Huigin, C., Figueroa, F., and Klein, J. (1994) Origin of the North American house mouse. Foli. Biola. (Praha) **40**, 483–496.

Vanlerberghe, F., Boursot, P., Nielsen, J. T., and Bonhomme, F. (1988) A steep cline for mitochondrial DNA in Danish mice. Gen. Res. **52**, 185–193.

Wade, C. M., Kulbokas, III E. J., Kirby, A. W., Zody, M. C., Mullikin, J. C., Lander, E. S., Lindblad-Toh, K., and Daly, M. J. (2002) The mosaic structure of variation in the laboratory mouse genome. Nature **420**, 574–578.

Yamamoto, M., Lin, X.-H., Kominato, Y., Hata, Y., Noda, R., Saitou, N., and Yamamoto, F. (2001) Murine equivalent of the human histo-blood group ABO gene is a *cis*-AB gene and encodes a glycosyltransferase with both A and B transferase activity. J. Biol. Chem. **276**, 13701–13708.

Yonekawa, H., Moriwaki, K., Gotoh, O., Hayashi, J. I., Watanabe, J., Miyashita, N., Petras, M. L., and Tagashira, Y. (1981) Evolutionary relationships among five subspecies of *Mus musculus* based on restriction enzyme cleavage patterns of mitochondrial DNA. Genetics **98**, 1–16.

Yonekawa, H., Moriwaki, K., Gotoh, O., Miyashita, N., Matsushima, Y., Shi, L. M., Cho, W. S., Zhen, X. L., and Tagashira, Y. (1988) Hybrid origin of Japanese mice "*Mus musculus molossinus*": evidence from restriction analysis of mitochondrial DNA. Mol. Biol. Evol. **5**, 63–78.

Zhang, J., Hunter, K. W., Gandolph, M., Rowe, W. L., Finney, R. P., Kelley, J. M., Edmonson, M., and Buetow, K. H. (2005) A high-resolution multistrain haplotype analysis of laboratory mouse genome reveals three distinctive genetic variation patterns. Genome Res. **15**, 241–249.