MOLECULAR PHYLOGENETICS AND EVOLUTION
Vol. 5, No. 3, June, pp. 548–556, 1996
ARTICLE NO. 0049

The Presence/Absence Polymorphism and Evolution of the p53 Pseudogene in the Genus *Mus*

HIROSHI OHTSUKA,* MITURU OYANAGI,* YOSHIO MAFUNE,* NOBUMOTO MIYASHITA,†,[1]
TOSHIHIKO SHIROISHI,†,[2] KAZUO MORIWAKI,†,[3] RYO KOMINAMI,*,[4]
AND NARUYA SAITOU‡

**First Department of Biochemistry, Niigata University School of Medicine, Niigata 951, Japan; †Laboratory of Cytogenetics, National Institute of Genetics, Mishima 411, Japan; and ‡Laboratory of Evolutionary Genetics, National Institute of Genetics, Mishima 411, Japan*

Received April 6, 1995; revised July 28, 1995

Distribution of the p53 pseudogene within the house mouse species (genus *Mus*) was studied with polymerase chain reaction for 37 individuals that were caught at different localities. Pseudogene-specific fragments were detected in some, but not all, individuals of *Mus musculus* subspecies regardless of locality and type of subspecies. In addition, 3 of 7 individuals belonging to different *Mus* species carried the pseudogene in their genomes. These results show the existence of an interspecific presence/absence polymorphism of the p53 pseudogene in mice. Sequence analysis of 11 amplified 0.3-kb fragments suggested that the pseudogene originated in an ancestral mouse about 7 million years ago. Thus alleles with and without the p53 pseudogene have persisted through the mice speciation. The evolutionary rate for the p53 functional gene was also estimated to be about 3.3×10^{-9} per nucleotide site per year.

INTRODUCTION

p53 is a cell cycle-dependent nuclear phosphoprotein and acts as a negative regulator of cell proliferation. Inactivation of this gene is frequently observed in tumor cells and is hence thought to be required for neoplastic transformation (Vogelstein and Kinzler, 1992). The p53 functional gene is located on mouse chromosome 11, whereas a processed pseudogene found in the mouse genome has been assigned to chromosome 14 using the analysis of a hybrid-cell panel (Zakut-Houri *et al.*, 1983; Czosnek *et al.*, 1984). However, another mapping study using recombinant inbred strains suggests that the pseudogene is located on chromosome 17 (Aitman *et al.*, 1991). We carried out a more detailed linkage analysis using microsatellite probes (Dietrich *et al.*, 1992) and determined that the pseudogene is located about 30 centiMorgans from the centromere on chromosome 17 (Kominami *et al.*, unpublished data).

In the course of the mapping study, we noted that a laboratory strain originating from Japanese wild mice did not carry this pseudogene in the genome. This suggests a unique distribution of the pseudogene within the house mouse. Hence, we have examined wild mice trapped in various localities and here show that the pseudogene is present in some individuals of species and subspecies of the mouse and is absent in other individuals. This suggests that a presence/absence polymorphism of the p53 pseudogene exists in the genus *Mus*. This unique distribution is discussed from the evolutionary viewpoint of polymorphism, as well as the origin and evolution of this pseudogene and its functional counterpart.

MATERIALS AND METHODS

Mice

Wild mice examined in this study were trapped at various localities in the world. Their taxonomy is described in Table 1 according to studies of genetic polymorphisms (Bonhomme and Guenet, 1989). Several *Mus* species and four main groups of *Mus musculus* subspecies are distributed throughout the Old World.

The nucleotide sequence data reported in this paper will appear in the DDBJ/EMBL/GenBank nucleotide sequence databases with the following accession numbers: D63399, p53MLEGF; D63400, p53MLEGP; D63401, p53MMBAP; D63402, p53MMBRP; D63403, p53MMCAP; D63404, p53MPLAF; D63405, p53MSPIP; D63406, p53MSPRP.

[1] Present address: Kagawa Medical School, Miki, Kagawa 761-07, Japan.

[2] Present address: Mammalian Genetics Laboratory, National Institute of Genetics, Mishima 411, Japan.

[3] Present address: Graduate University for Advanced Studies, Hayama 240-01, Japan.

[4] To whom correspondence should be addressed at, First Department of Biochemistry, Niigata University School of Medicine, 1-757 Asahimachi-doori, Niigata 951, Japan. Fax: +81-25-223-0237.

1055-7903/96 $18.00

TABLE 1

List of Wild Mice and the Presence or Absence of p53 Pseudogene

Origin of mouse	Locality (sample size)	Presence/absence of p53 pseudogene[a]
Mus musculus		
M. m. domesticus	Canada (2)	P, P
	Spain (1)	P
	Bulgaria (1)	P
	France (3)	P, P, P
M. m. castaneus	Malaysia (1)	P
	Indonesia (2)	P, P
	Taiwan (2)	P, A
	Philippines (2)	P, P
M. m. bactrianus	Pakistan (2)	P, A
	Iran (2)	A, A
	Afghan (1)	P
M. m. musculus	Poland (1)	A
	Bulgaria (2)	P, A
	Denmark (2)	A, A
M. m. molossinus[b]	Japan (6)	P, A, A, A, A, A
Other *Mus* species		
M. spretus	Spain (1)	A
M. spicilegus	Bulgaria (1)	P
M. spretoides[c]	Bulgaria (1)	P
M. caroli	Thailand (2)	A, A
M. leggada	Sri Lanka (1)	P
M. platythrix	India (1)	A

[a] P indicates the presence of p53 pseudogene and A indicates its absence.

[b] Mice having gene exchange between *M. m. castaneus* and *M. m. musculus* in Japan.

[c] It is also called *M. macedonicus.*

A schematic phylogeny of the wild mice is illustrated in Fig. 1, after Bonhomme and Guenet (1989). Since *Mus musculus molossinus* living in Japan is regarded as hybrid between *Mus musculus musculus* and *Mus musculus castaneus,* that subspecies is omitted from Fig. 1.

PCR Primers and Conditions

The oligonucleotide primers used for amplification of the p53 pseudogene were F1: 5′-ATGAACCGCCGACC TATCCTT and R1: 5′-CTTGAGGGTGAAATACTCT CC, and F2: 5′-TACTCTCCTCCCCTCAATAAGCT and R2: 5′-CAGAAGGTTCCCACTGGAGTCTTC for exons 7–9 and exons 5–7, respectively. Amplification of DNA by the polymerase chain reaction (PCR) method was carried out in 20 μl containing 20 ng of high-molecular-weight genomic DNA, 0.1 μg of each primer, 10 m*M* Tris–HCl, pH 8.3, 50 m*M* potassium chloride, 1.5 m*M* magnesium chloride, 0.01% gelatin, and 0.5 units of Ampli*Taq* polymerase (Takara Inc., Japan) and overlayered with 20 μl of mineral oil (Saiki *et al.,* 1988). Thermal cycling was carried out as follows: denaturing at 95°C for 1 min, annealing at 55°C for 1.5 min, polymerization at 72°C for 1.5 min, for 30–35 cycles.

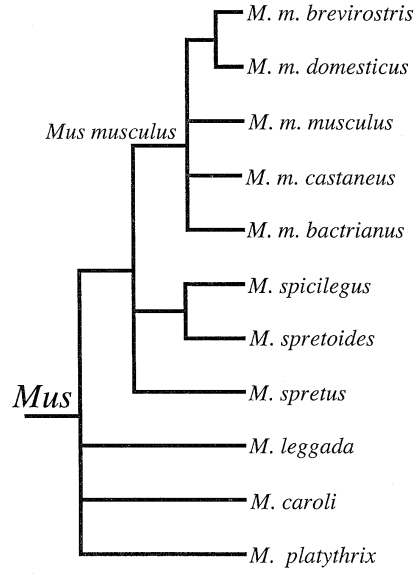

FIG. 1. A schematic phylogeny of the house mouse species and subspecies (after Bohnomme and Guenet, 1989). *M. spretoides* is also called *M. macedonicus.*

Analysis of PCR Products

Five-microliter aliquots of product were separated on 4% NuSieve agarose gel electrophoresis. The gel was stained in 0.5 μg/ml ethidium bromide and photographed.

Sequence Analysis

PCR products were isolated from NuSieve agarose gels by centrifugation with Ultrafree (Millipore). Approximately 100 ng of gel-purified PCR product was used as a template for sequence analysis by the chain termination method (Sanger *et al.,* 1977). For this we used the Sequenase kit (United States Biochemical Corp.). The reaction products were separated on 6% polyacrylamide gels.

Phylogenetic Analysis

Phylogenetic trees were constructed by using the neighbor-joining method (Saitou and Nei, 1987), the maximum parsimony method (Fitch, 1977), and the maximum likelihood method (Felsenstein, 1981). Those three methods have been shown to reconstruct true phylogenetic relationships with high probabilities when closely related sequences are compared (Saitou and Imanishi, 1989; Kuhner and Felsenstein, 1994). Computer programs NJBOOT2 (kindly provided by Dr. K. Tamura), PAUP 3.1.1 (University of Illinois Natural History Survey), and DNAML in PHYLIP 3.5 (Felsenstein, 1993) were used for assisting the neighbor-joining, the maximum parsimony, and the maximum likelihood analyses, respectively. Numbers of nucleotide substitutions between sequences used in the neighbor-joining analysis were estimated by using Kimura's (1980) method. Branch length estimation of a

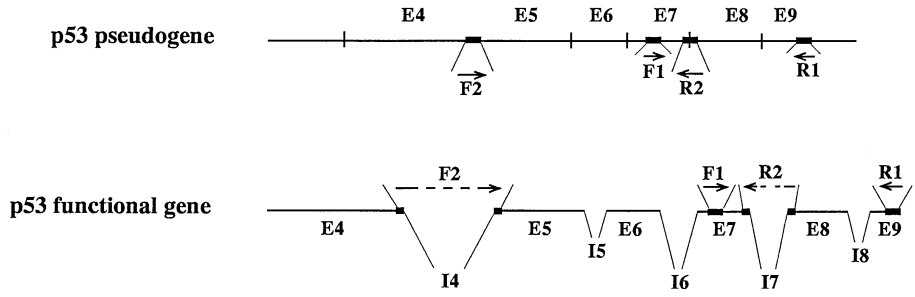

FIG. 2. Schematic illustration of the p53 pseudogene and the functional gene. The p53 pseudogene does not contain introns. Exons and introns of the functional gene are shown by E's and I's, respectively. Arrows indicate the positions of the forward and reverse primers used. F1 and R1 primers are located on exons 7 and 9, respectively, whereas both F2 and R2 primers are composed of sequences of two exons as illustrated.

neighbor-joining tree was obtained by using the method described in Ishida *et al.* (1995).

RESULTS

p53 Pseudogene Presence/Absence Polymorphism

Thirty-seven wild mice caught at different localities were used to examine whether or not they carry the p53 pseudogene in the genomes. PCR was carried out using a primer pair complementary to both of the gene and pseudogene sequences (Zakut-Houri *et al.*, 1983; Bienz *et al.*, 1984). The primers F1 and R1 were located on exons 7 and 9, respectively (Fig. 2). The electrophoresis of PCR products gave one or two DNA fragments (Fig. 3A). A larger 660-bp fragment was present in all the mice except for *Mus caroli,* which had a fragment of longer size. This large fragment was derived from the p53 functional gene because it contained introns 7 and 8 (data not shown). The smaller 260-bp fragment was derived from the processed pseudogene that lacked the intron sequences. This difference between these two sizes of fragments was determined by sequence analysis (data not shown).

Some of the wild mice failed to give the 260-bp fragment, suggesting that these animals do not contain the pseudogene. Table 1 summarizes the results for the 37 wild mice and their localities. All seven individuals of *M. musculus domesticus* carried the pseudogene, but mice belonging to the other subspecies showed variation. To confirm this, PCR amplification using another primer set, F2 and R2, was done. The primers span two exons (Fig. 2). The 5′ half of the forward primer (F2) is on the 3′ part of exon 4 and the 3′ half is on the 5′ part of exon 5; and the 5′ half of the reverse (R2) is on the 3′ part of exon 7 and the 3′ half is on the 5′ part of exon 8. Since these primer sequences were not able to form stable duplexes with the p53 functional gene sequence, they gave only a 418-bp fragment that was derived from the pseudogene sequence. As shown in Fig. 3B, the obtained result was consistent with that using the F1 and R1 primers; all wild mice thought to lack the pseudogene from the F1 and R1 probe failed to give the pseudogene fragment of 418 bp.

Seven individuals of *M. musculus domesticus* contained a whole or part of the pseudogene sequence in the genomes, suggesting that the mice belonging to this subspecies carry the p53 pseudogene. Among Asian subspecies, however, there was variation: 6 *castaneus,* 2 *bactrianus,* 1 *musculus,* and 1 *molossinus* carried the pseudogene, whereas the other 13 individuals of *castaneus, bactrianus, musculus,* and *molossinus* did not (Table 1). Interestingly, individuals of *M. spicilegus, M. spretoides,* and *M. leggada* belonging to other *Mus* species had the pseudogene, but *M. spretus, M. caroli,* and *M. platythrix* did not. These results clearly show the presence/absence polymorphism of the p53 pseudogene within species of the genus *Mus.*

Phylogenetic Analysis

To know if the pseudogenes have the same evolutionary origin, we determined the sequence spanning the exons 5 to 7 of five subspecies and three different species of mice. Figure 4 shows the multiple alignment of the 10 sequences thus determined as well as 3 p53 cDNA sequences retrieved from the DDBJ/EMBL/GenBank international nucleotide sequence database. Table 2 shows the evolutionary distances among those sequences. Two individuals were examined for the case of *M. musculus bactrianus* pseudogene, but those were identical. The two *M. musculus bactrianus* pseudogene sequences turned out to be identical with that of *M. musculus castaneus. M. musculus brevirostris* pseudogene and *M. musculus domestics* pseudogene were identical with each other, so was *Mus spicilegus* and *Mus spretoides* pseudogenes (see also Fig. 5). Those identical sequences were represented by one of those in the phylogenetic analysis. All the gaps in Fig. 4 were not used in the phylogenetic analysis, and a total of 323 nucleotide sites were compared.

Figure 5 is a neighbor-joining tree for all 14 sequences based on the evolutionary distances of Table 2. All the pseudogene sequences (designated by ψ) constitute a clear monophyletic cluster at branch C (the

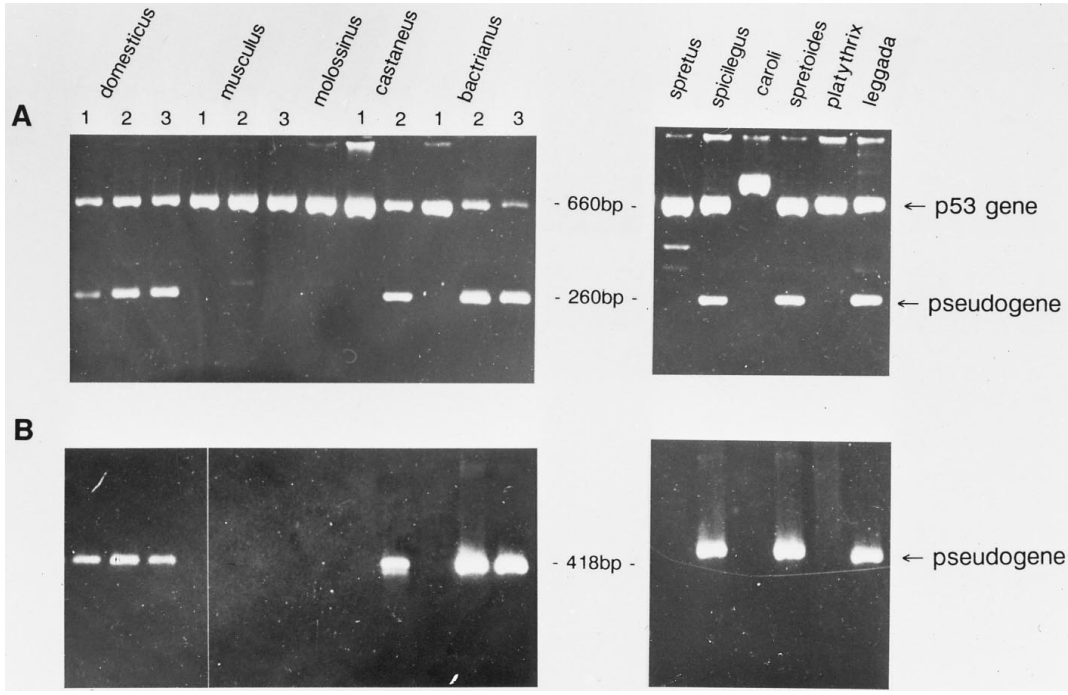

FIG. 3. Detection of the p53 pseudogene in mouse DNAs. (A) PCR was carried out using the F1 and R1 primers for 18 wild mice. The products were electrophoresed in 4% NuSieve agarose gel and stained with ethidium bromide. Wild mice used are listed in Table 1. Numbers indicate different individuals belonging to the same subspecies. Details of these wild mice are described in Bonhomme and Guenet (1989). (B) PCR analysis was carried out as described in A except for using the F2 and R2 primers.

bootstrap probability was 97%), while the phylogenetic relationship among the functional genes is not clear because of low bootstrap probabilities at branches F, G, and H. In any case, it is apparent that the p53 pseudogene has been evolving with a much higher rate than the functional gene. It should also be noted that the phylogenetic pattern in the pseudogene cluster is compatible with that of Fig. 1.

We found eight equally maximum parsimonious trees, and all the trees required 112 nucleotide substitutions (see Table 3). Tree 6 has the identical topology with that of the neighbor-joining tree (Fig. 5). We also examined nine sequences by excluding *Homo sapiens* cDNA, but obtained the same eight trees. Table 4 shows the result of bootstrapping. Partitions A–D and F–H correspond to interior branches of the neighbor-joining tree in Fig. 5. As in the case of Fig. 5, partitions A–D showed high bootstrap probabilities, while partitions E–H showed low probabilities. Partitions A–G are mutually compatible with each other, while partition H is incompatible with partition E.

After the p53 processed pseudogene was created, both the pseudogene and the functional gene lineages must have diverged following the speciation events. Thus *M. leggada* and *M. musculus* functional genes should be monophyletic, for their pseudogene counterparts were clearly monophyletic. It should be noted that *Mus musculus domesticus* functional gene and its cDNA must also cluster. However, none of the eight maximum parsimonious trees showed this topology (Table 3). Therefore, we examined topologies under these constraints, and found the three submaximum parsimonious trees in which 113 nucleotide substitutions were required (Fig. 6).

We also used the maximum likelihood method. User tree option of DNAML was applied to examine the eight maximum parsimonious trees and the three submaximum parsimonious trees. The result is presented in Table 3. Tree 4 showed the highest log-likelihood value among the 11 topologies examined. The three submaximum parsimonious trees of Fig. 6 showed somewhat lower log-likelihood values, though none of the 10 subbest trees were significantly worse than the best tree with the highest likelihood. Branch lengths of the trees in Fig. 6 are based on the maximum likelihood estimates.

All three trees of Fig. 6 suggest that the *M. platythrix* lineage diverged first followed by the *M. leggada* lineage. This branching pattern was also obtained in the neighbor-joining tree (see Fig. 5), though the bootstrap probabilities were not high. If we accept this divergence pattern, there are two possibilities for the emergence of the p53 pseudogene. If the pseudogene was created before the divergence of the *platythrix* lineage, then tree 11 of Fig. 6 is compatible with this hypothesis. If the pseudogene was created after the divergence of the

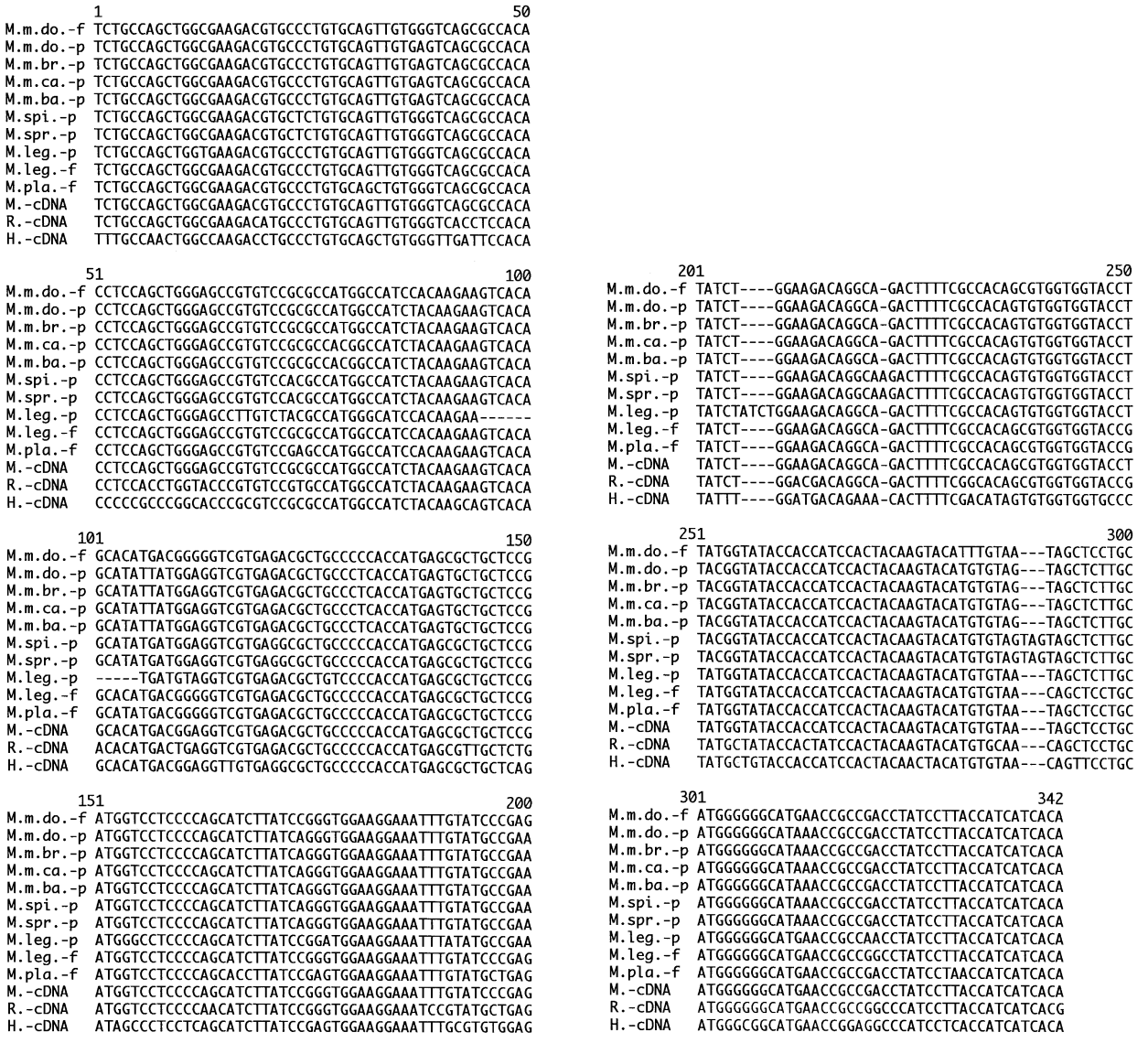

FIG. 4. Alignments of nucleotide sequences of three functional genes and seven pseudogenes from five rodent species as well as three cDNA sequences retrieved from the DDBJ/EMBL/GenBank database. Gaps are denoted by dashes. The nucleotide position 1 corresponds to the 390th from the initiation codon site (Bienz *et al.*, 1984). Full species names of abbreviated forms are as follows: M.m.do., *Mus musculus domesticus;* M.m.br., *Mus musculus brevirostris;* M.m.ca., *Mus musculus castaneus;* M.m.ba, *Mus musculus bactrianus;* M.spi., *Mus spicilegus;* M.spr., *Mus spretoides;* M.leg., *Mus leggada;* M.pla., *Mus platythrix;* M.-cDNA, *Mus musculus* cDNA (accession number K01700), R.-cDNA, *Rattus norvegicus* cDNA (accession number = X13058); and H.-cDNA, *Homo sapiens* cDNA (accession number K03199). Designations "f" and "p" after a dash mean functional genes and pseudogenes, respectively.

M. platythrix lineage, tree 10 of Fig. 6 is compatible. In this case, the emergence of the pseudogene must occurred before the divergence of the *M. leggada* lineage. In either case, tree 9 of Fig. 6 is incompatible with any hypothesis. We will therefore consider only trees 10 and 11 in the following.

Because a rough constancy of evolutionary rate seems to hold within the pseudogenes and within the functional genes, we estimated the time of node P, the divergence time of the pseudogene and the functional gene, and that of node Q, the divergence time of the *M. platythrix* lineage from the remaining *Mus* lineages, in trees 10 and 11 of Fig. 6. The estimation method of Ishida *et al.* (1995) was used in the following analysis.

If we consider only the functional gene lineages, the average number of nucleotide substitutions between node Q and the present time becomes 5.5 and 6.1 for trees 10 and 11, respectively. The corresponding numbers for the divergence point of *Mus* and *Rattus* lineages are 16.3 and 16.6 for trees 10 and 11, respec-

TABLE 2

Estimated Number of Nucleotide Substitutions per Site between Every Pair of 14 Sequences (see Fig. 4); Gaps Were Eliminated from the Comparison, and a Total of 323 Nucleotide Sites Were Compared (Kimura's (1980) Two-Parameter Method Was Used)

	1	2	3	4	5	6	7	8	9	10	11	12	13
2	0.052												
3	0.052	0.0											
4	0.055	0.003	0.003										
5	0.055	0.003	0.003	0.0									
6	0.055	0.003	0.003	0.0	0.0								
7	0.048	0.022	0.022	0.025	0.025	0.025							
8	0.048	0.022	0.022	0.025	0.025	0.025	0.0						
9	0.058	0.065	0.065	0.069	0.069	0.069	0.055	0.055					
10	0.192	0.218	0.218	0.223	0.223	0.223	0.205	0.205	0.231				
11	0.009	0.042	0.042	0.045	0.045	0.045	0.038	0.038	0.055	0.180			
12	0.093	0.122	0.122	0.126	0.126	0.126	0.119	0.119	0.137	0.186	0.082		
13	0.013	0.058	0.058	0.062	0.062	0.062	0.055	0.055	0.065	0.180	0.016	0.079	
14	0.028	0.069	0.069	0.072	0.072	0.072	0.065	0.065	0.075	0.183	0.032	0.093	0.029

Note. Sequence IDs are; 1, *M. m. domesticus* functional gene; 2, *M. m. domesticus* pseudogene; 3, *M. m. brevirostris* pseudogene; 4, *M. m. castaneus* pseudogene; 5, *M. m. bactrianus* pseudogene A; 6, *M. m. bactrianus* pseudogene B; 7, *M. spicilegus* pseudogene; 8, *M. spretoides* pseudogene; 9, *M. leggada* pseudogene; 10, *Homo sapiens* cDNA; 11, *M. m.* cDNA; 12, *Rattus norvegicus* cDNA; 13, *M. leggada* functional gene; 14, *M. platythrix* functional gene.

tively. Thus, the relative divergence time for node Q is 0.34 (=5.5/16.3) and 0.37 (=6.1/16.6) for trees 10 and 11, respectively, when that for *Mus* and *Rattus* is 1.0. Since the divergence time of *Mus* and *Rattus* was estimated to be about 1.5×10^7 years (Li and Tanimura, 1987), the divergence times for node Q become 5.1×10^6 and 5.6×10^6 years for trees 10 and 11, respectively.

Because 323 nucleotides were compared in the phylogenetic analysis, 16.3/323 and 16.6/323 correspond to $d/2$ in the equation $d = 2\lambda t$, where d is the pairwise evolutionary distance, λ is the evolutionary rate, and t is the divergence time. The rate ($\lambda \equiv d/2t$) of nucleotide substitution for the p53 functional gene is thus estimated to be 3.3×10^{-9} per nucleotide site per year, us-

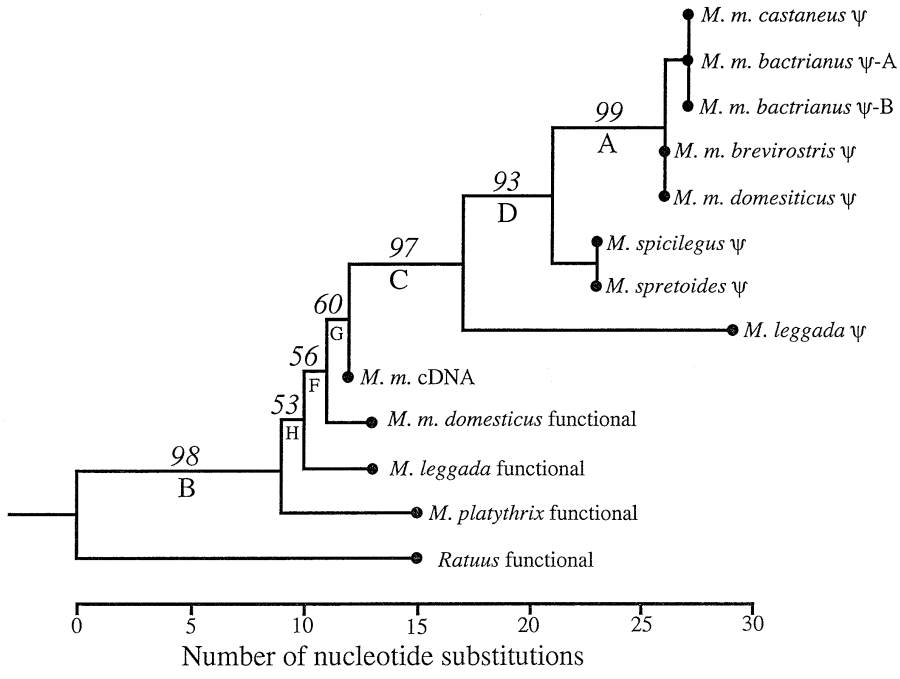

FIG. 5. A neighbor-joining tree of the p53 functional genes, pseudogenes, and cDNA. Numbers on interior branches are bootstrap probabilities (%) based on 1000 bootstrap resampling. The tree was rooted by assuming that *Homo sapiens* cDNA sequence was the outgroup. The bottom scale is the number of nucleotide substitutions that occurred in the region of 323 compared nucleotides.

TABLE 3

Results of the Maximum Parsimony and Maximum Likelihood Analyses

Tree ID[a]	Tree topology[b]	RNM[c]	Diff. Ln L[d]
1	(((((((2,3),4),5),9),6),1),8,(7,10))	112	−4.07
2	((((8,9),1),6),(((2,3),4),5),(7,10))	112	−0.29
3	(1,6,(((((2,3),4),5),9),((7,10),8)))	112	−4.09
4	(((((((2,3),4),5),6),1),9),8,(7,10))	112	Best
5	((((((2,3),4),5),(6,1)),9),8,(7,10))	112	−5.39
6	(((((((2,3),4),5),6),1),8),9,(7,10))	112	−2.55
7	(((((2,3),4),5),(6,(1,8))),9,(7,10))	112	−4.31
8	(((((2,3),4),5),6),((1,8),9),(7,10))	112	−6.10
9	(((1,6),8),((((2,3),4),5),9),(7,10))	113	−9.02
10	((((1,6),8),(((2,3),4),5)),9,(7,10))	113	−8.80
11	((((1,6),8),9),(((2,3),4),5),(7,10))	113	−8.99

[a] Trees 9–11 are submaximum parsimonious trees 9–11 of Fig. 6, respectively.

[b] Sequence IDs are as follows: 1, *M. m. domesticus* functional gene; 2, *M. m. domesticus* pseudogene; 3, *M. m. castaneus* pseudogene; 4, *M. spicilegus* pseudogene; 5, *M. leggada* pseudogene; 6, *M. m.* cDNA; 7, *Rattus norvegicus* cDNA; 8, *M. leggada* functional gene; 9, *M. platythrix* functional gene; and 10, *Homo sapiens* cDNA.

[c] Required number of mutations when the maximum parsimony method was applied.

[d] Differences of log-likelihood values from that of the best tree (tree 4; its log-likelihood was −1016.75).

ing the average distance value (0.10 = [16.3 + 16.6]/323) of trees 10 and 11 and 1.5×10^7 years as *t* between *Mus* and *Rattus*. If we use this rate, the divergence times for node P of trees 10 and 11 become 2.8×10^6 and 6.7×10^6 years, respectively.

When the pseudogene lineages are considered, the average number of nucleotide substitutions between the node P and the present time becomes 15.9 and 15.6 for node P of trees 10 and 11, respectively. We applied these *d* values and the divergence times for P estimated above by using the evolutionary rate for functional genes, and the rate of nucleotide substitution of p53 pseudogenes is estimated to be 17.6×10^{-9} and 7.2×10^{-9} for the case of trees 10 and 11, respectively. The rate of synonymous substitution for rodent genes (comparison between mice and rats) was estimated to be 7.9×10^{-9} per nucleotide site per year (Li and Tanimura, 1987). The estimated rate (7.2×10^{-9}) of pseudogene nucleotide substitution for tree 11 is closer to this value than that for tree 10. Therefore, tree 11 seems to be more plausible than tree 10, and the p53 pseudogene

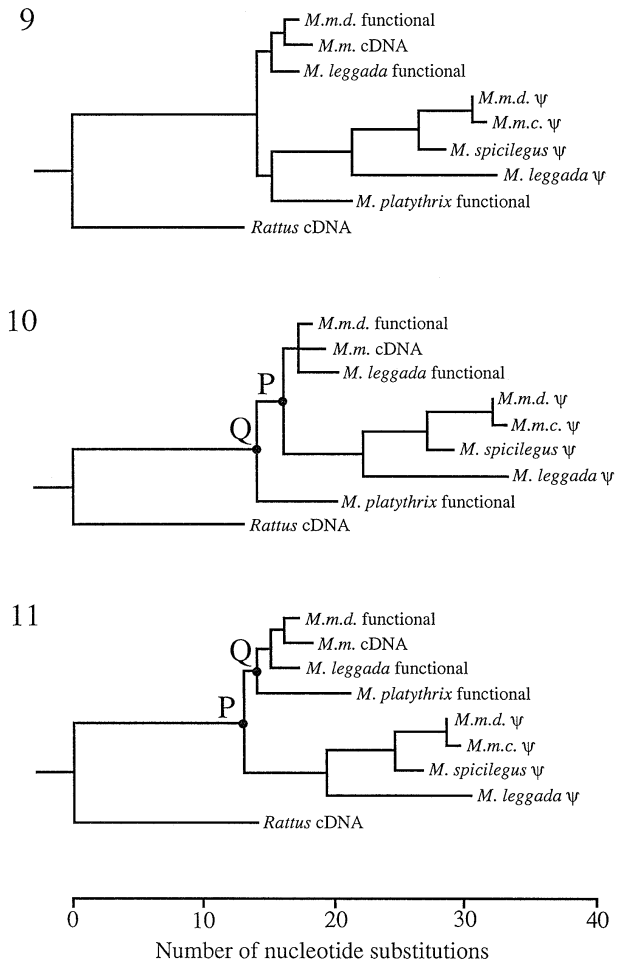

FIG. 6. Three submaximum parsimonious trees found by using the "constraint" option of PAUP 3.1.1. The constraint used was: ((*M.m.d.* pseudogene, *M.m.c.* pseudogene, *M. spicilegus* pseudogene, *M. leggada* pseudogene), ((*M.m.* cDNA, *M.m.d.* functional gene), *M. leggada* functional gene), *M. platythrix* functional gene, *Rattus* cDNA, *Homo* cDNA). All three trees required 113 nucleotide substitutions. *Homo sapiens* p53 cDNA was assumed to be the outgroup. Numbers on branches are branch length estimates (number of nucleotide substitutions). Those estimates were obtained by rounding the maximum likelihood estimates (per nucleotide site) multiplied by the number of compared nucleotides (323). Node P of trees 10 and 11 denotes the divergence point of the pseudogene and functional gene, while node Q denotes the divergence point of the *Mus platythrix* lineage from the remaining *Mus* lineages.

TABLE 4

Bootstrap Probabilities for Eight Partitions When the Maximum Parsimony Method Was Used

Partition[a]	Sequence ID[a] 1	2	3	4	5	6	7	8	9	10	Probability (%)
A	#	@	@	#	#	#	#	#	#	#	98.9
B	#	#	#	#	#	#	@	#	#	@	96.8
C	#	@	@	@	@	#	#	#	#	#	94.8
D	#	@	@	@	#	#	#	#	#	#	92.1
E	#	#	#	#	#	#	@	@	#	@	41.7
F	#	#	#	#	#	#	@	@	@	@	32.6
G	#	@	@	@	@	@	#	#	#	#	27.0
H	#	#	#	#	#	#	@	#	@	@	21.5

This result is based on the 1000 bootstrap resampling using PAUP 3.1.1.

[a] These partitions correspond to interior branches A–H of Fig. 5, except for partition E.

[b] Sequence IDs are the same as those of Table 3.

is estimated to have emerged about 7 million years ago (node P of tree 11). If we compare the relative evolutionary rate of the p53 functional gene to that of the p53 pseudogene after node P of tree 11, it becomes 0.39 (=6.1/15.6). This is not much different from the ratio (0.46 = 3.3/7.2) between the two estimates of absolute evolutionary rates.

DISCUSSION

We analyzed the sequences and the distribution of a p53 pseudogene in various species of the genus *Mus*. It is demonstrated that the pseudogene is present in some individuals of mouse species and subspecies and absent in others, indicating variation within populations of the house mouse species. The findings, together with sequence analysis, suggest the following phylogenetic interpretation of the presence/absence polymorphism of the p53 pseudogene. A reverse transcription of the p53 gene took place only once and the complementary DNA was integrated into a site on chromosome 17 in an ancestral mouse, probably slightly before the speciation of the *M. platythrix* from the remaining *Mus* lineages (see tree 11 of Fig. 6).

The allele having the pseudogene subsequently spread and probably was fixed in the ancestral mouse population. If this is the case, the pseudogene sometimes disappeared during speciation and this disappearance must have occurred several times, since there are many mice that do not carry the pseudogene in their genomes.

If we could compare sequences flanking to the pseudogene and sequences of corresponding regions in the alleles without the pseudogene, it would give clues to verify this interpretation; i.e., such an analysis might give evidence for deletion of the pseudogene in the alleles lacking the pseudogene. We thus synthesized two primers complementary to flanking sequences and tried to amplify the corresponding sequence. However, we failed to isolate fragments corresponding to the flanking regions; this PCR gave only L1 sequence fragments, probably because of the presence of the pseudogene within an L1 repetitive sequence (Zakut-Houri *et al.*, 1983: sequence homology data not shown there). It is well known that repetitive sequences frequently undergo recombination and evolve rapidly (Smith, 1976; Fanning and Singer, 1988). Therefore, the presence of the pseudogene in L1 may be why deletion of the pseudogene frequently occurred in populations of the mouse subspecies and subgenera.

There may be another possible explanation; the pseudogene allele was not fixed in the ancestral mouse population. If this is the case, the presence/absence polymorphism of p53 pseudogene implies that the two alleles with and without the pseudogene have persisted through speciation and both forms remain in modern *Mus* species. That is, the polymorphism has passed from one species to another. This phenomenon, the so-called trans-species polymorphism, has been reported for the major histocompatibility complex (MHC) genes in mice (McConnell *et al.*, 1988; Sagai *et al.*, 1989). However, the evolutionary mechanism responsible for the p53 trans-species polymorphism is probably different from that for the MHC genes. Pseudogenes are considered to undergo neutral evolution (Kimura, 1983), while MHC genes are known to be under some kind of natural selection such as overdominant selection (Hughes and Nei, 1987).

If there is no selective force, why could the p53 pseudogene persist for a long time? We suggested that tree 11 of Fig. 6 is the most likely phylogenetic tree for the *Mus* p53 functional genes and pseudogenes, and the p53 pseudogene probably emerged about 7 million years ago. On the other hand, the coalescence time for a pair of neutrally evolving genes is $2N$ generations for diploid genes, where N is the long-term effective population size (Nei, 1987). Effective population size of dozens of rodent species was estimated to be in the range of 0 to 7×10^5 (Nei and Graur, 1984). If we assume N of an ancestral *Mus* species to be about 5×10^5 individuals and the generation time to be about 1 year, the expected coalescence time for a pair of gene copies in that species group becomes 1 million years. Therefore, if a timespan of two successive speciation events within the genus *Mus* was less than 1 million years, it is possible that a polymorphism on any neutrally evolving gene is shared in different *Mus* species. The p53 pseudogene may be an example of this.

ACKNOWLEDGMENTS

This study was partially supported by grants-in-aids for scientific research of Ministry of Education, Science, Sports, and Culture, Japan, to R.K., T.S., K.M., and N.S.

REFERENCES

Aitman, T. J., Hearne, C. M., McAleer, M. A., and Todd, J. A. (1991). Mononucleotide repeats are an abundant source of length variants in mouse genomic DNA. *Mammal. Genome* **1:** 206–210.

Bienz, B., Zakut-Houri, R., Givol, D., and Oren, M. (1984). Analysis of the gene coding for the murine cellular tumor antigen p53. *EMBO J.* **3:** 2179–2183.

Bonhomme, F., and Guenet, J-L. (1989). The wild house mouse and its relatives. *In* "Genetic Variants and Strains of the Laboratory Mouse" (M. F. Lyon and A. G. Seale, Eds.), pp. 649–662, Oxford Univ. Press, Oxford.

Czosnek, H. H., Bienz, B., Givol, D., Zakut-Houri, R., Pravtcheva, D. D., Raddle, F. K., and Oren, M. (1984). The gene and the pseudogene for mouse p53 cellular tumor antigen are located on different chromosomes. *Mol. Cell. Biol.* **4:** 1638–1640.

Dietrich, W., Katz, H., Lincoln, S. E., Shin, H-S., Friedman, J., Dracopoli, N. C., and Lander, E. S. (1992). A genetic map of the mouse suitable for typing interspecific crosses. *Genetics* **131:** 423–447.

Fanning, T. G., and Singer, M. F. (1988). LINE-1: A mammalian transposable element. *Biochem. Biophys. Acta* **910:** 203–212.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* **17:** 368–376.

Felsenstein, J. (1993). "PHYLIP (Phylogeny Inference Package) Version 3.5." University of Washington.

Fitch, W. M. (1977). On the problem of discovering the most parsimonious tree. *Am. Nat.* **111:** 223–257.

Hughes, A. L., and Nei, M. (1987). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335:** 167–170.

Ishida, N., Oyunsuren, T., Mashima, S., Mukoyama, H., and Saitou, N. (1995). Mitochondrial DNA sequences of various species of the genus *Equus* with a special reference to the phylogenetic relationship between Przewalskii's wild horse and domestic horse. *J. Mol. Evol.* **47:** 180–188.

Kimura, M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16:** 111–120.

Kimura, M. (1983). "The Neutral Theory of Molecular Evolution," Cambridge Univ. Press, Cambridge.

Kuhner, M. K., and Felsenstein, J. (1994). A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol. Biol. Evol.* **11:** 459–468.

Li, W.-H., and Tanimura, M. (1987). The molecular clock runs more slowly in man than in apes and monkeys. *Nature* **326:** 93–96.

McConnell, T. J., Talbot, W. S., McIndoe, R. A., and Wakeland, E. K. (1988). The origin of MHC class II gene polymorphism within the genus *Mus. Nature* **332:** 651–654.

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

Nei, M., and Graur, D. (1984). Extent of protein polymorphism and the neutral mutation theory. *Evol. Biol.* **17:** 73–118.

Sagai, T., Sakaizumi, M., Miyashita, N., Bonhomme, F., Petras, M. N., Nielsen, J. T., Shiroishi, T., and Moriwaki, K. (1989). New evidence for trans-species evolution of the H-2 class I polymorphism. *Immunogenetics* **30:** 89–98.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Sharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239:** 487–491.

Saitou, N., and Imanishi, T. (1989). Relative efficiencies of the Fitch–Margoliash, maximum-parsimony, maximum-likelihood, minimum-evolution, and neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. *Mol. Biol. Evol.* **6:** 514–525.

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

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chainterminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74:** 5463–5467.

Smith, G. P. (1976). Evolution of repeated DNA sequences by unequal crossover. *Science* **191:** 528–535.

Vogelstein, B., and Kinzler, K. W. (1992). p53 function and dysfunction. *Cell* **70:** 523–526.

Zakut-Houri, R., Oren, M., Bienz, B., Lavie, V., Hazum, S., and Givol, D. (1983). A single gene and a pseudogene for the cellular tumor antigen p53. *Nature* **306:** 594–597.