Genes Genet. Syst. (2006) **81**, p. 201–209

Further evidence for recombination between mouse hemoglobin beta b1 and b2 genes based on the nucleotide sequences of intron, UTR, and intergenic spacer regions

Jun J. Sato[1], Yoshiharu Tsuru[1], Kyoko Hirai[1], Yasunori Yamaguchi[1*], Kazuyuki Mekada[2], Naoyuki Takahata[3] and Kazuo Moriwaki[2]

[1]*Laboratory of Animal Cell Technology, Faculty of Life Science and Technology, Fukuyama University, Higashimura-cho, Aza, Sanzo, Fukuyama 729-0292, Japan*
[2]*RIKEN Tsukuba Institute, RIKEN Bioresource Center, Kouyadai 3-1-1, Tsukuba 305-0074, Japan*
[3]*Department of Biosystems Science, Graduate University for Advanced Studies, Hayama, Kanagawa 240-0193, Japan*

(Received 11 May 2006, accepted 10 June 2006)

Nucleotide sequences of the intron regions and UTRs (Untranslated regions) of the hemoglobin beta adult genes, b1 and b2, and of the intergenic spacer region were determined for mouse strains representing the *d*, *p*, and *w1* hemoglobin haplotypes defined by protein electrophoretic analyses. The hypothesis of recombination of the b1 and b2 genes between the *d* and *w1* haplotypes previously reported in the cDNA nucleotide sequences was confirmed by neighbor-joining analyses of the intron regions and UTRs within the b1 and b2 genes, suggesting that all of the structures of hemoglobin beta adult genes support the hypothesis that the *p* haplotype was established by hybridization between *d* and *w1* haplotype mice. The resultant recombinant of the *p* haplotype was found to have a *d*-like b1 gene and a *w1*-like b2 gene. In addition to the possible recombination, a break point was suggested around 2–3 kb downstream of the b1 gene within the intergenic spacer region, despite the absence of clear properties that could stimulate the recombination machinery. Some large insertions or deletions (indels) specific to the *p* or *d* haplotypes were located within the intergenic spacer region, in which the 1010-bp indel specific to the *p* haplotype was shared by all examined strains representing the *p* haplotype.

Key words: hemoglobin beta genes, hybridization, *Mus musculus*, recombination, subspecies groups.

INTRODUCTION

Hybridization among genetically and biochemically defined subspecies groups within the house mouse (*Mus musculus*) in the wild has often been proposed (Hunt and Selander, 1973; Ferris et al., 1983a; Sage et al., 1986; Yonekawa et al., 1986, 1988; Vanlenberghe et al., 1988; Bonhomme et al., 1989; Boursot et al., 1989; Frisman et al., 1990; Moriwaki, 1994; Mezhzherin et al., 1998). The differentiation among the major lineages, i.e., *domesticus*, *musculus*, and *castaneus* subspecies groups, have been inferred to occur around 1–2 million years ago (Moriwaki et al., 1979; Yonekawa et al., 1980, 1981; Ferris et al., 1983a; Suzuki et al., 2004), and it is possible that their genes are partly exchanged by hybridization where the populations overlap (Hunt and Selander, 1973; Ferris et al., 1983b; Yonekawa et al., 1986, 1988; Bonhomme et al., 1989; Boursot et al., 1989).

The hypothesis of recombination detected in the wild-derived mouse has been proposed for the hemoglobin beta genes b1 and b2 based on cDNA sequences (Ueda et al., 1999), in which the three hemoglobin haplotypes *d*, *p*, and *w1* (determined by protein electrophoretic analysis; Miyashita et al., 1985; Kawashima et al., 1991; Miyashita et al., 1994) were examined. In Ueda et al. (1999), the authors hypothesized the possible recombination between hemoglobin b1 and b2 genes due to hybridization between subspecies groups, i.e., *musculus* with the *w1* haplotype and *castaneus* with the *d* haplotype (Fig. 4 in Ueda et al.,

Edited by Yoichi Matsuda
* Corresponding author. E-mail: yamaguti@bt.fubt.fukuyama-u.ac.jp

1999), suggesting that the *p* haplotype was the resultant recombinant carrying the b1 gene of the *d* haplotype and the b2 gene of the *w1* haplotype. It is not surprising that this event is likely to occur, considering the close geographical distributions of these subspecies groups (Kawashima et al., 1995) and the phylogenetic evidence that *M. m. musculus* and *M. m. castaneus* are more closely related to the exclusion of *M. m. domesticus* (Lundrigan et al., 2002). The idea, however, was created based only on cDNA sequences representing the exon parts of each gene. The hemoglobin beta adult genes, both b1 and b2, consist of three exons intervened by two intron regions with UTRs flanking both sides of each gene (Konkel et al., 1978, 1979; Shehee et al., 1989). Therefore, there still remain two intron regions and 5' and 3' UTRs of b1 and b2 genes for further testing of this recombination hypothesis. In addition, ca. 14 kb of the intergenic spacer region between b1 and b2 genes are available in which to locate the possible break point of the recombination.

Here we present further evidence for the recombination hypothesis using the nucleotide sequences of intron regions and UTRs; additionally, we searched for the possible recombination break point within the intergenic spacer region between hemoglobin b1 and b2 genes.

MATERIALS AND METHODS

Mice examined in this study Mice with the *p* haplotype from which the inbred strain MSM (*M. m. molossinus*) was developed were captured at Mishima, Japan, in 1978. Mice with the *w1* haplotype were captured at Jiayuguang, Gansu Province, China, in 1981. BALB/cAnN congenic strains carrying either Hbb^p of the MSM strain or Hbb^{w1} of the Gansu mouse have been established by repeated backcrosses of eight generations in the former and five in the latter case. The congenic strains BALB/c-Hbb^{w1} and BALB/c-Hbb^p thus developed have been maintained in the National Institute of Genetics (NIG; Shizuoka, Japan). BALB/cAnN mice were purchased from Japan SLC, Inc. (Shizuoka, Japan). The AU/SsJ strain purchased from the Jackson Laboratory (Bar Harbor, ME, USA) and the JF1, SWN/Ms, and KJR/Ms strains were also maintained at NIG. All mouse strains used in this study are listed in Table 1. The Car strain (*Mus caroli*, BRC No. 00823) was provided by RIKEN BRC with the support of National BioResource Project of Ministry of Education, Culture, Sports, Science and Technology, Japan.

Isolation, amplification, and sequencing of DNA Total genomic DNA was extracted from tissues by the conventional phenol–chloroform method. Amplification was performed via polymerase chain reactions (PCRs) using an automated thermal cycler (PTC-200, Peltier Thermal Cycler; MJ Research, Watertown, MA, USA). For the intron and UTR sequences and the 14 kb of the intergenic spacer region, PCR was first performed with 2.5 mM $MgCl_2$, 0.4 mM dNTP mix, 0.2 μM of each pair of primers (HBB1-Fw/HBB1-Rv for the b1 gene, HBB-b2-UTR-F/HBB-b2-UTR-R for the b2 gene, and F1/R9, F9/R11, F11/R21, F18/R26 for the intergenic spacer region; Appendix), 2.5 units of LA Taq DNA polymerase (Takara, Tokyo, Japan), appropriate buffer for the enzyme, and 100 ng of total genomic template DNA in a total reaction volume of 50 μl. Thermal cycling parameters were as follows: 35 cycles of denaturation at 94°C for 30 sec, annealing at 55°C for 30 sec, and extension at 72°C for 2 or 5 min according to the amplified fragment length. After purification of the DNA fragments by ethanol precipitation, 50 ng of the first PCR product was used as a template for the second PCR in a 50-μl reaction with 0.2 mM dNTP mix, 0.8 μM of each pair of primers, 1.25 units of Pyrobest DNA polymerase (Takara), and appropriate buffer for the enzyme including $MgCl_2$. The primer pairs for the second PCR were as follows: HBB1-Fw/HBB① for the 5'-UTR and first intron region of the b1 gene, HBB②/HBB③ for the second intron region of the b1 gene, Hb1b2-F1/HBB1-Rv for the 3'-UTR of the b1 gene, HBB-b2-UTR-F/HBB① for the 5'-UTR and first intron region of the b2 gene, HBB②/HBB④ for the second intron region of the b2 gene, HBB⑨/HBB-b2-UTR-R for the 3'-UTR of the b2

Table 1. Samples used in this study

strain	Hbb haplotype*	Description
BALB/c	*d*	Inbred strain established as an European laboratory mouse, belonging to *domesticus* subspecies group
BALB/c-Hbb^{w1}	*w1*	BALB/c congenic strain carrying Hbb^{w1} of the mouse captured at Jiayuguang, Gansu Province in China
BALB/c-Hbb^p	*p*	BALB/c congenic strain carrying Hbb^p of the MSM strain captured at Mishima, Japan
AU/SsJ	*p*	Inbred strain established as an European laboratory mouse
JF1	*p*	Inbred strain established as a fancy mouse, belonging to *musculus* subspecies group
SWN/Ms	*p*	Wild derived inbred strain from Suewon, Korea, belonging to *musculus* subspecies group
KJR/Ms	*p*	Wild derived inbred strain from Kojuri, Korea, belonging to *musculus* subspecies group
Car	–	Wild derived strain from Okinawa, Japan (maintained by closed colony), belonging to *Mus caroli*

* *Hbb* haplotypes are defined by cellulose acetate membrane electrophoresis.

gene, and F1/R1, F2/R2, F3/R3, F4/R4, F5/1R, F6/R6, F7/R7, F8/R8, F9/R9, F10/R10, F11/R11, F12/R12, F13/R13, F14/R14, F15/R15, F16/R16, F18/R17, F19/R19, F20/R20, F21/R21, F22/3R, F23/R23, F24/R24, F25/R25, and F26/R26 for the intergenic spacer region (Appendix). Thermal cycling parameters were as follows: 35 cycles of denaturation at 94°C for 30 sec, annealing at 55°C for 30 sec, and extension at 72°C for 30 or 40 sec according to the amplified fragment length. The sequence reaction with the purified products of the second PCR was performed according to the manufacturer's instructions (Big Dye Terminator cycle sequencing kit; PE Biosystems, Foster City, CA, USA), and sequenced products were detected on an ABI 377, ABI 310, or ABI 3100 Avant automated sequencer (Applied Biosystems, Foster City, CA, USA). Because there were some repeated sequences within the intergenic spacer region, GeneScan (Freiburg, Germany) was performed to determine the length of these repetitive sequences. PCR was conducted with the primer pairs F9.5CT.FAM/R9.5CT.FAM, F23GA.TET/R23GA.TET, and F24GA.HEX/R24GA.HEX. PCR products were then subjected to GeneScan analyses on ABI 377. The nucleotide sequence data reported in this paper have been deposited in the DNA Data Bank of Japan (DDBJ) International Nucleotide Sequence Database with accession numbers AB189402–AB189428, AB219809, and AB219811.

Data analyses Multiple alignment of the nucleotide sequences for three haplotypes was conducted by visual inspection. Phylogenetic analyses were performed for the sequences of the intron regions and UTRs to represent the relationships among the sequences of hemoglobin b1 and b2 genes for each haplotype, *d*, *p*, and *w1*. The neighbor-joining method (NJ; Saitou and Nei, 1987), implemented using PAUP* version 4.0b10 (Swofford, 2001), was performed with matrices of *p*-distances for the combined sequences of first and second introns and those of 5' and 3' UTRs. For the sequences of the intergenic spacer region, we searched sites and insertions or deletions (indels) supporting the associations [(*d*, *p*), *w1*] or [*d*, (*p*, *w1*)] to find the possible recombination break point. In addition, because a part of the spacer region of *M. caroli* was available, this species was used as the outgroup to determine whether these similarities were derivative or primitive.

RESULTS

Nucleotide variations and phylogenetic analyses

Intron regions Nucleotide variations within the intron regions were found among the three strains of BALB/c, BALB/c-*Hbb*^{p}, and BALB/c-*Hbb*^{w1}, representing the *d*, *p*, and *w1* haplotypes, respectively. Length variation occurred in which a single 3-bp indel was observed at sites 38466–38468 (site numbers according to the mouse reference sequences [Shehee et al., 1989] throughout this study) within the first intron region of the b1 gene. The *w1* haplotype had a shorter sequence than the *d* and *p* haplotypes because of the 3-bp indel. Except for the indel, there was only one variable site, which corresponded to site 38505. Within the second intron of the b1 gene, a single 1-bp indel was detected, again resulting in a shorter sequence of the *w1* haplotype. Excluding one indel, there were 14 variable sites, 11 of which showed the affinity of the *d*–*p* haplotypes.

No length polymorphism was evident within both the first and second intron of the b2 gene, although all haplotypes had an extra base (G) between sites 54103 and 54104. There was only one variable site within the first intron, which corresponded to site 53678, representing the *p*–*w1* affinity. Within the second intron of the b2 gene, there were nine variable sites, eight of which showed the affinity of the *p*–*w1* haplotypes.

Phylogenetic trees that resulted from the NJ analyses for the combined sequence of first and second introns (b1 gene, 769 bp; b2 gene, 745 bp) showed different topologies between b1 and b2 genes (Fig. 1a and 1c). While the *d* haplotype was more closely related to the *p* haplotype than the *w1* haplotype for the b1 gene, the *w1* haplotype showed close affinity with the *p* haplotype for the b2 gene.

5' and 3' UTRs Nucleotide variations within 5' and 3' UTRs were found among the three strains, as in the intron regions, representing the *d*, *p*, and *w1* haplotypes. There was no length variation for the 5'-UTR of the b1 gene and the 5' and 3' UTRs of the b2 gene, but a single 2-bp indel was observed at site 39632–39633 within the 3'-UTR of the b1 gene, in which the *w1* haplotype possessed a longer sequence than the *d* or *p* haplotypes. There was only one variable site (site 39621) within the 3'-UTR of the b1 gene, at which the affinity of the *d*-*p* haplotypes was evident. Four variable sites were observed within UTRs of the b2 gene (one in the 5'-UTR, three in the 3'-UTR), all of which showed the affinity of the *p*–*w1* haplotypes.

Phylogenetic trees that resulted from the NJ analyses for the combined sequence of 5' and 3' UTRs (b1 gene, 217 bp; b2 gene, 214 bp) showed different topologies between b1 and b2 genes, as was the case for the intron regions (Fig. 1b and 1d). While the sequence of the *d* haplotype was the same as that of the *p* haplotype for the b1 gene, the *w1* haplotype had the same sequence as the *p* haplotype for the b2 gene.

Intergenic spacer region We observed nucleotide variations within the intergenic spacer region between the b1 and b2 genes. There were 59 indels within the sequences of three haplotypes aligned visually; 35 of these were associated with the repeated sequences, the sizes of which were considered to fluctuate readily. Some indels

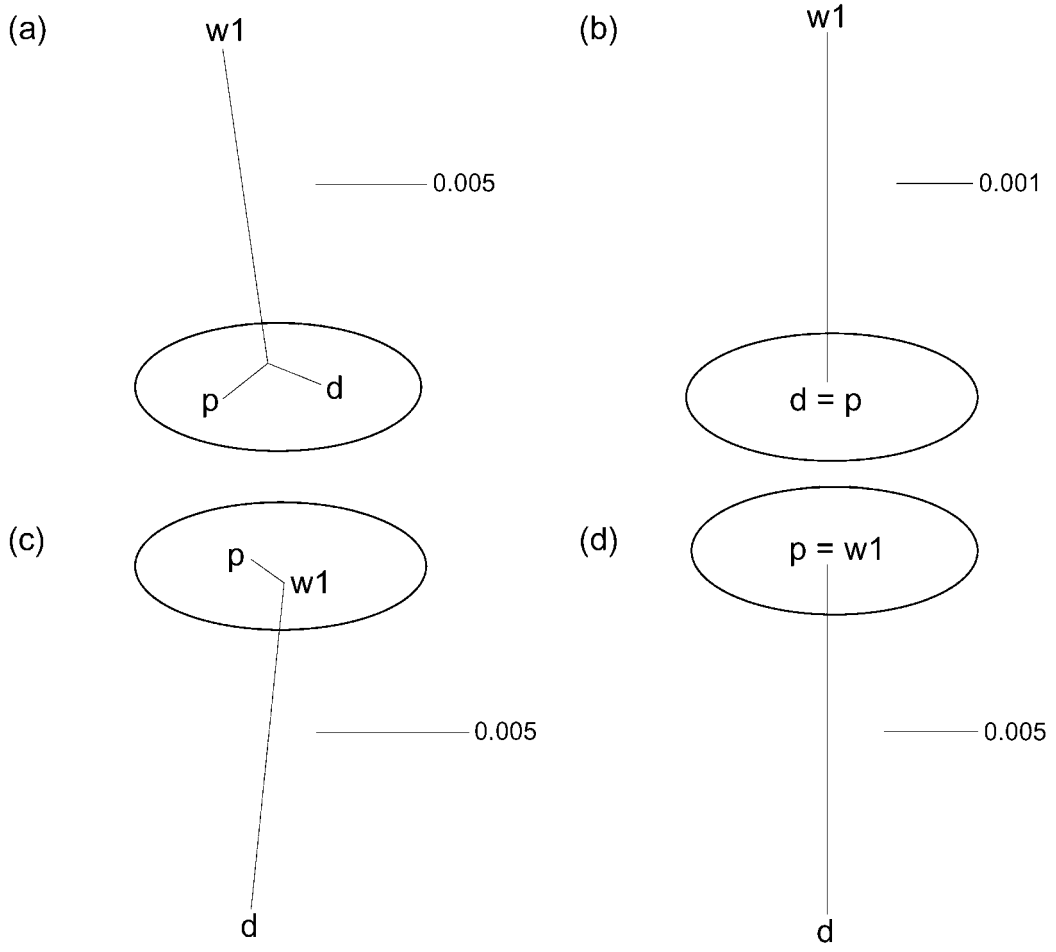

Fig. 1. Neighbor-joining trees for the *d*, *p*, and *w1* haplotypes based on the intron regions and UTRs of hemoglobin b1 and b2 genes. (a) Tree resulting from the combined sequence of the intron regions of the b1 gene (769 bp). (b) Tree resulting from the combined sequence of the 5' and 3'-UTRs of the b1 gene (217 bp). (c) Tree resulting from the combined sequence of the intron regions of the b2 gene (745 bp). (d) Tree resulting from the combined sequence of the 5' and 3'-UTRs of the b2 gene (214 bp). Bars indicate criteria of distances for each tree (substitutions/site). Figures attached by bars are based on the *p*-distance.

were not related to the repeated sequences and were very large. The sequence of the *p* haplotype had a specific 1010-bp sequence at site 43515–43516, while the sequence of the *d* haplotype had 2007-bp and 127-bp sequences at sites 46220–48226 and 52501–52627, respectively (Fig. 3). The 1010-bp sequence specific to the *p* haplotype was detected from all examined strains representing the *p* haplotype, MSM, AU/SsJ, JF1, SWN/Ms, and KJR/Ms by agarose electrophoretic analyses (data not shown). Apparently, these dramatic differences were not associated with the repeated sequences. Three indels greater than 10 bp were also not related to the repeated region, i.e., a 10-bp indel at sites 43644–43653 with a shorter sequence of the *p* haplotype, and 22-bp and 13-bp indels at sites 44639–44640 and 45679–45680, respectively, with a longer sequence of the *p* and *w1* haplotypes. While the *d–w1* affinity was indicated by the 1010-bp and 10-bp indels, the *p–w1* affinity was evidenced by the 22-bp, 13-bp, 2007-bp, and 127-bp indels. Excluding these indels, there were 205 variable sites, each showing affinity as in Fig. 2. The sites representing the *p–w1* affinity were overwhelmingly abundant within the latter region (e.g., downstream of sites around 43000), whereas a relatively large number of sites showing the *d–p* affinity was found within the region near the b1 gene (e.g., at sites around 39000–42000) although there were some sites supporting the *d–w1* or *p–w1* affinity. To investigate whether these similarities were characterized by synapomorphic or plesiomorphic states, we compared the partial sequences of *M. musculus* to the outgroup *M. caroli* (Fig. 2). Although almost all similarities resulted from ancestral characters including all sites showing the *d–w1* and *p–w1* affinities, four independent sites represented derived similarities (sites 40102, 40398, 40403, and 41470), all of which support the *d–p* affinity.

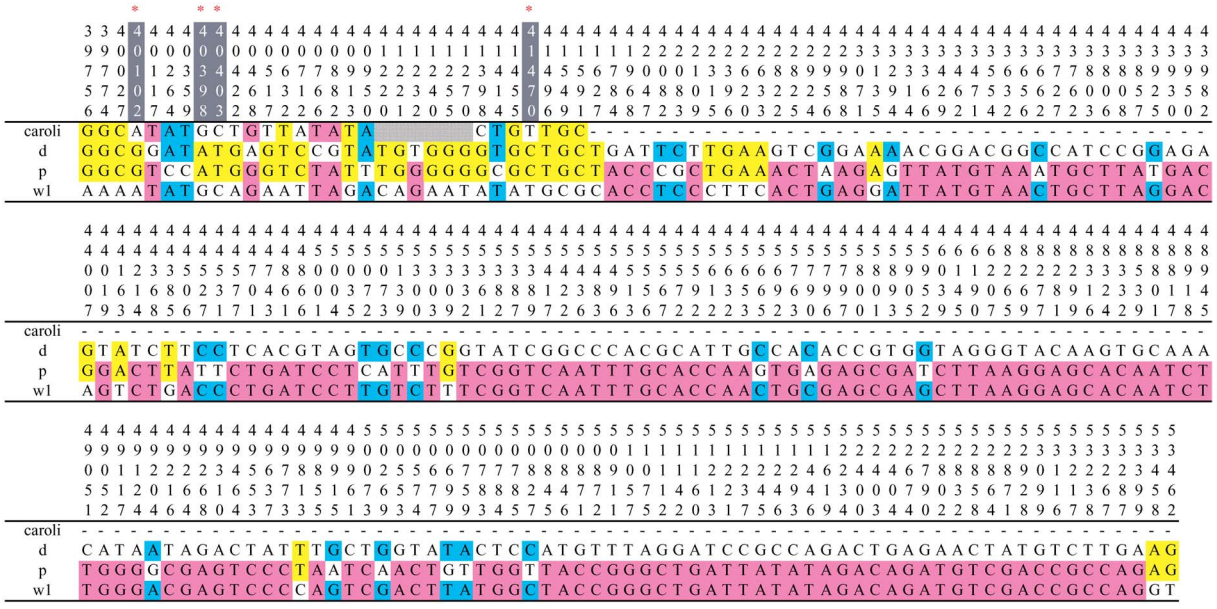

Fig. 2. Variable sites within the intergenic spacer region that agree with the close affinity of each pattern, *d–p* (yellow), *d–w1* (blue), and *p–w1* (red), are represented with the partial sequence of *Mus caroli* for outgroup comparison. Red asterisks above sites on a gray background and white site numbers indicate parsimonious informative sites. Figures on these nucleotides correspond to the site numbers according to sequences of the hemoglobin complex determined by Shehee et al. (1989). Shaded regions lacked an indel for the sequences of *M. caroli*. Because the spacer region of *M. caroli* has not been determined completely, missing sites are represented by dashes (–).

DISCUSSION

The recombination hypothesis proposed based on the cDNA sequences of hemoglobin b1 and b2 genes (Ueda et al., 1999) was confirmed by the results obtained from the NJ analyses for both of the intron regions and UTRs (Fig. 1). While the *d* haplotype was closely related to the *p* haplotype for the b1 gene, the *w1* haplotype alternatively indicated an affinity to the *p* haplotype for the b2 gene, supporting the hypothesis that the *p* haplotype is the recombinant of the *d* and *w1* haplotypes (Fig. 3). The evidence suggests that all of the structures of b1 and b2 genes, exons, introns, and UTRs, possess the same information to support the recombination hypothesis, which then raises the question: Where is the recombination point? The sites that agree with the *d–p* or *p–w1* affinity were searched within the intergenic spacer region (Fig. 2), and the possible recombination break point was suggested at a site around 42000–43000. Although there was some support for the *d–w1* or *p–w1* affinities in the region upstream of the hypothetical recombination point, these similarities were found to be ancestral by outgroup comparison with *Mus caroli*. Besides, four parsimonious informative characters in this region represented the derived nature of the *d–p* similarity, which doesn't contradict with the position of the inferred recombination point.

Focusing on indels greater than 10-bp, the 22-bp (44639–44640), 13-bp (45679–45680), 2007-bp (46220–48226), and 127-bp (52501–52627) indels represented the *p–w1* affinity and the consistency with the hypothetical recombination point. However, the *d–w1* affinity, i.e., a *p*-specific property, was represented by 1010-bp (43515–43516) and 10-bp indels (43644–43653), proposing two alternative hypotheses. One is that these fragments have been specifically inserted and deleted, respectively, within the sequence of the *p* haplotype after the establishment of this haplotype (drawn in Fig. 3). The other is that these fragments have been deleted and inserted, respectively, within the sequence of only *w1* haplotype after the establishment of the *p* haplotype (figure not shown). To determine which hypothesis is more plausible, further studies are required. Nevertheless, the observed indels do not seem to contradict the hypothesis of recombination irrespective of the direction of indel events.

Assuming based on the above evidence that the recombination hypothesis is true, we can estimate the time scale of this event. Since the differences between the *d* and *p* haplotypes within the b1 gene and the region from the end of the b1 gene to the possible recombination point, where the sequence of the *d* and *p* haplotypes are similar to each other, can be explained by the substitutions obtained after the establishment of the *p* haplotype, these variations allow us to infer the time estimate since the possible recombination event. Using a calibration point previously estimated for the divergence among the three major species groups within the subgenus *Mus* (2.4–5.4

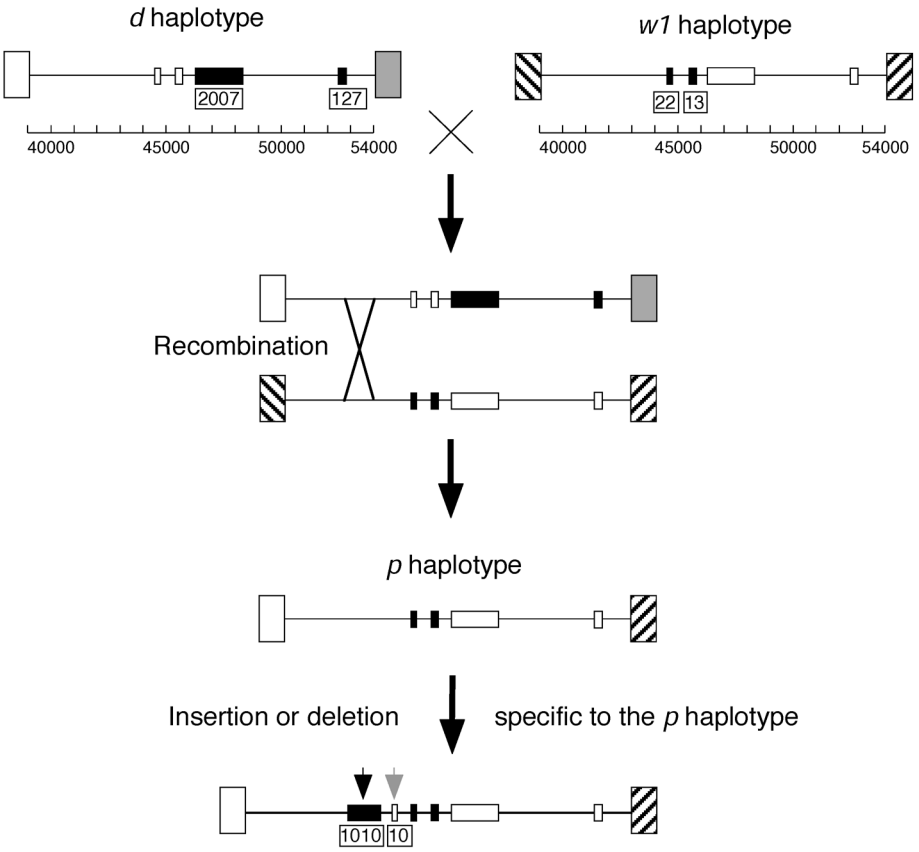

Fig. 3. Possible recombination hypothesis focusing on large indels (greater than10 bp). Filled rectangles on the intergenic spacer regions are longer sequences, and hollow rectangles represent no sequences in those regions. The numbers within boxes beneath indels show the length of those indels (see text) and the number of measures refers to the sequence reported previously (Shehee et al., 1989).

megannum [Ma]; Suzuki et al., 2004) as a divergence time between lineages of *M. musculus* and *M. caroli*, the possible recombination event was here estimated to have occurred at 0.23–0.52 Ma with the molecular clock assumption (Zuckerkandl and Pauling, 1965) by employing sequence variations of the whole b1 gene sequences (1430 bp) and about 1.7-kb sequences downstream of the b1 gene (in total, 3155 bp; excluding gaps, 202 bp) obtained in this study. Therefore, the hypothetical recombination seems to be a relatively recent event in the evolution of *M. musculus*.

To date, recombination events have been studied extensively for their machinery during meiosis, although the factors that could stimulate the meiotic recombination events remain unclear (Petes, 2001; Massy, 2003). Our results are consistent with the suggestion derived from studies of yeasts, humans, and mice that recombination events are likely to occur more frequently within the regions between genes than within genes (Petes, 2001; Massy, 2003). One of the proposed causes for a high level of recombination events is a high GC-content of the sequences (Petes, 2001; Massy, 2003). It has been suggested that high GC-content may result in replication fork blockage, which leads to local modification of histones recognized by the recombination factors (Petes, 2001). When the base composition within the intergenic spacer region was scanned with the 1-kb sliding window shifted at 100-bp intervals (Fig. 4), the fluctuation in the GC-content was observed within the region. However, a high GC-content is not likely to be directly associated with the region around the possible recombination point. Meanwhile, there are some homopolymers, such as the poly (dA:dT) or poly (dG:dC) tract that could stimulate the recombination by excluding nucleosomes (Iyer and Struhl, 1995; Kirkpatrick et al., 1999; Petes, 2001), e.g., the perfect poly T tract at site 41645–41660, the perfect poly C tract at site 42527–42536, the imperfect poly A tract at site 42537–42550, and the perfect poly A tract at site 43486–43495. Although these repetitive single bases may be associated with the recombination event, we found no direct evidence for the association in this study. Such homopolymers were also found in other parts within the intergenic spacer region.

Despite the absence of clear properties promoting recombination events within sequences around the possible break point, the hypothesis of recombination between

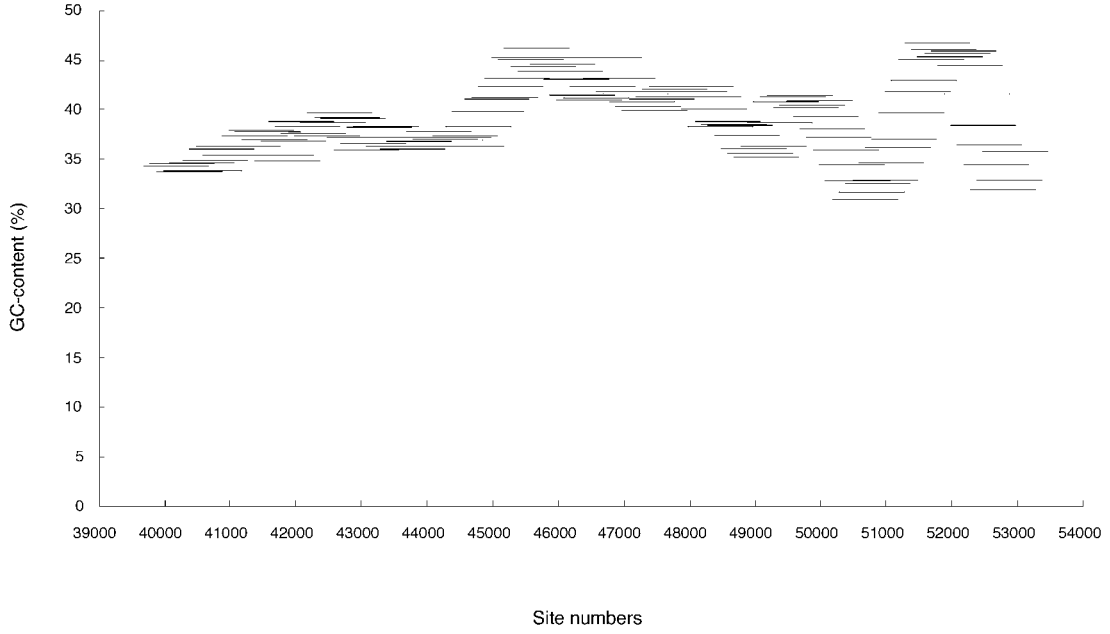

Fig. 4. Variations in GC-content (%) within the intergenic spacer region between b1 and b2 genes; site numbers correspond to sequences of the hemoglobin complex of Shehee et al. (1989). The 1-kbp sliding window analyses were performed by shifting the frame at 100-bp intervals. Bars in the graphics indicate the 1-kbp sliding windows.

the *d* and *w1* haplotypes is the most probable to explain the evidence derived from all structures of b1 and b2 genes and sequences within the intergenic spacer region among the three haplotypes. The proximity of the natural geographic distribution of the *musculus* and *castaneus* subspecies groups, representing the *w1* and *d* hemoglobin haplotypes, respectively, and the close phylogenetic relationship between *M. m. musculus* and *M. m. castaneus* (Lundrigan et al., 2002), in the sense of genetically and biochemically circumscribed subspecies (Bonhomme et al., 1984; Moriwaki et al., 1986, 1990), could be the evidence that would allow the recombination hypothesis to be justifiable. But it is unclear in this study how such a recombination could have occurred in the phylogeographical point of view. Considering the current distribution that the places inhabited by the *p* haplotype seems to exist between that of the *d* and *w1* haplotypes (Miyashita et al., 1994), secondary contact between the current populations of the *musculus* (including the *w1* haplotype) and *castaneus* (including the *d* haplotype) subspecies groups might cause the recombination. However it is also probable that the recombinants as ancestral polymorphisms have been maintained to present, taking into account that the polymorphisms of the *d*, *p* and *w1* haplotypes within one locality were observed in many places (Kawashima et al., 1991, 1995; Miyashita et al., 1994). Clearly, population genetic approaches with more sampling of individuals are required to clarify that the recombinant (the *p* haplotype) is the result of the recent secondary contact or maintenance of old polymorphisms of ancestral population including the original lineages of the *musculus* and *castaneus* subspecies groups. Nevertheless, the possible recombination event caused by the hybridization between at least ancestral original lineages of these subspecies groups could plausibly be explained by the current available circumstantial evidence (Kawashima et al., 1991, 1995; Miyashita et al., 1994; Ueda et al., 1999; this study).

We thank Sonoko Kawano, Chisa Sugihara, and Kozue Hiyama (Fukuyama University) for help with the experimental work. This study was supported in part by a Grant-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

REFERENCES

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

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

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

Frisman, L. V., Korobitsina, K. V., Yakimenko, L. V., Bokshtein, F. M., and Muntyanu, A. I. (1990) Genetic differentiation of U.S.S.R. house mice: Electrophoretic study of proteins. Biol. J. Linn. Soc. **41**, 65–72.

Hunt, W. G., and Selander, R. K. (1973) Biochemical genetics of hybridization in European house mice. Heredity **31**, 11–33.

Iyer, V., and Struhl, K. (1995) Poly (dA: dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. EMBO. J. **14**, 2570–2579.

Kawashima, T., Miyashita, N., Wang, C-H., He, X-Q., Jin, M-L., Wu, Z-A., and Moriwaki, K. (1991) A new haplotype of the globin gene complex, *Hbb*^{w1}, in Chinese wild mouse. Jpn. J. Genet. **66**, 491–500.

Kawashima, T., Miyashita, N., Tsuchiya, K., Li, H., Wang, F., Wang, C-H., Wu, X-L., Wang, C., Jin, M-L., He, X-Q., Kryukov, A. P., Yakimenko, L. V., Frisman, L. V., and Moriwaki, K. (1995) Geographical distribution of the *Hbb* haplotypes in the *Mus musculus* subspecies in Eastern Asia. Jpn. J. Genet. **70**, 17–23.

Kirkpatrick, D. T., Wang, Y-H., Dominska, M., Griffith, J. D., and Petes, T. D. (1999) Control of meiotic recombination and gene expression in yeast by a simple repetitive DNA sequence that excludes nucleosomes. Mol. Cell. Biol. **19**, 7661–7671.

Konkel, D. A., Tilghman, S. M., and Leder, P. (1978) The sequence of the chromosomal mouse β-globin major gene: Homologies in capping, splicing and poly (A) sites. Cell **15**, 1125–1132.

Konkel, D. A., Maizel, J. V. Jr., and Leder, P. (1979) The evolution and sequence comparison of two recently diverged mouse chromosomal globin genes. Cell **18**, 865–873.

Lundrigan, B-L., Jansa, S. A., and Tucker, P. K. (2002) Phylogenetic relationships in the genus *Mus*, based on paternally, maternally, and biparentally inherited characters. Syst. Biol. **51(3)**, 410–431.

Massy, B. (2003) Distribution of meiotic recombination sites. Trend in Genetics **19(9)**, 524–522.

Mezhzherin, S. V., Kotenkova, E. V., and Mikhailenko, A. G. (1998) The house mice, *Mus musculus* s. l., hybrid zone of Transcaucasus. Z. Saug **63**, 154–168.

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

Miyashita, N., Kawashima, T., Wang, C-H., Jin, M-L., Wang, F., Gotoh, H., Yakimenko, L. V., Kryukov, A., Frisman, L. V., Akabarzadeh, J., and Moriwaki, K. (1994) Genetic polymorphisms of *Hbb* Haplotypes in wild mice. In: Genetics in wild mice: Its application to biomedical research (eds.: K. Moriwaki, T. Shiroishi, and H. Yonekawa), pp. 85–93. Japan Science Society Press, Tokyo.

Moriwaki, K., Shiroishi, T., Minezawa, M., Aotsuka, T., and Kondo, K. (1979) Frequency distribution of histocompatibility-2 antigenic specificities in the Japanese wild mouse genetically remote from the European subspecies. J. Immunogenet. **6**, 99–113.

Moriwaki, K., Miyashita, N., Suzuki, H., Kurihara, Y., and Yonekawa, H. (1986) Genetic features of major geographical isolates of *Mus musculus*. Curr. Top. Microb. Immunol. **127**, 55–61.

Moriwaki, K., Sagai, T., Shiroishi, T., Bonhomme, F., Wang, C-H., He, X-Q., Jin, M-L., and Wu, Z-A. (1990) Mouse subspecies differentiation and H-2 polymorphism. Biol. J. Linn. Soc. **41**, 125–139.

Moriwaki, K. (1994) Wild mouse from geneticist's viewpoint. Pages In: Genetics in wild mice: Its application to biomedical research (eds.: K. Moriwaki, T. Shiroishi, and H. Yonekawa), pp. xiii-xxv. Japan Science Society Press, Tokyo.

Petes, T. D. (2001) Meiotic recombination hot spot and cold spot. Nature genetics **2**, 360–369.

Sage, R. D., Heyneman, D., Kim, K-C., and Wilson, A. C. (1986) Wormy mice in a hybrid zone. Nature **324**, 60–63.

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

Shehee, W. R., Loeb, D. D., Adey, N. B., Burton, F. H., Casavant, N. C., Cole, P., Davies, C. J., McGraw, R. A., Schichman, S. A., Severynse, D. M., Voliva, C. F., Weyter, F. W., Wisely, G. B., Edgell, M. H., and Hutchison III, C. A. (1989) Nucleotide sequence of the BALB/c mouse globin complex. J. Mol. Biol. **205**, 41–62.

Suzuki, H., Shimada T., Terashima, M., Tsuchiya, K., and Aplin, K. (2004) Temporal, spatial, and ecological modes of evolution of Eurasian *Mus* based on mitochondrial and nuclear gene sequences. Mol. Phylogenet. Evol. **33(3)**, 626–646.

Swofford, D. L. (2001) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland.

Tsuchiya, K., Miyashita, N., Wang, C-H., Wu, X-L., He, X-Q., Jin, M-L., Li, H., Wang, F., Shi, L., and Moriwaki, K. (1994) Taxonomic study of the genus *Mus* in China, Korea, and Japan-Morphologic identification. In: Genetics in wild mice: Its application to biomedical research (eds.: K. Moriwaki, T. Shiroishi, and H. Yonekawa), pp3–12. Japan Science Society Press, Tokyo.

Ueda, Y., Miyashita, N., Imai, K., Yamaguchi, Y., Takamura, K., Notohara, M., Shiroishi, T., Kawashima, T., Ning, L., Wang, C., Wu, X., and Moriwaki, K. (1999) Nucleotide sequences of the mouse globin beta gene cDNAs in a wild derived new haplotype *Hbb*^{w1}. Mammalian Genome **10**, 879–882.

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

Yonekawa, H., Moriwaki, K., Gotoh, O., Watanabe, J., Hayashi, J-I., Miyashita, N., Petras, M. L., and Tagashira, Y. (1980) Relationship between laboratory mice and the subspecies *Mus musculus domesticus* based on restriction endonuclease cleavage patterns of mitochondrial DNA. Jpn. J. Genet. **55**, 289–296.

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**, 801–806.

Yonekawa, H., Gotoh, O., Tagashira, Y., Matsushima, Y., Shi, L-I, Cho, W-S., Miyashita, N., and Moriwaki, K. (1986) A hybrid origin of Japanese mice "*Mus musculus molossinus*". Curr. Top. Microb. Immunol. **127**, 62–67.

Yonekawa, H., Moriwaki, K., Gotoh, O., Miyashita, N., Matsushima, Y., Shi, L., Cho, W-S., Zhen, X., 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.

Zuckerkandle, E., and Pauling, L. (1965) Evolutionary divergence and convergence in proteins. In: Evolving Genes and Proteins (eds.: V. Bryson, and H. J. Vogel), pp. 97–166. Academic Press, New York.

Appendix. Primers used to amplify the sequences of introns, UTRs, and the intergenic spacer region

Primer name	Primer sequences	Site numbers*	Primer name	Primer sequences	Site numbers*
HBB1-Fw	aga gct gag act cct aag	38127	Hb1b2-F13	agt ttg cat gtc caa tgg gc	46388
HBB1-Rv	ttg gct tga gaa ctg tga ag	39975	Hb1b2-R13	aca atg gag tac tac tca gc	47027
HBB-b2-UTR-F	aag aac aga cac tac t	53308	Hb1b2-F14	tac ctc act cag gat gat gc	46975
HBB-b2-UTR-R	act atc tca cca ctt a	55088	Hb1b2-R14	tgt gaa gtt cct tga tcc ac	47554
HBB①	tcc aaa gct atc aaa gta	3,857,853,786	Hb1b2-F15	aac aat tgg tgc tgg cac ag	47484
HBB②	gtg gat cct gag aac ttc	3,876,653,975	Hb1b2-R15	ttc cag agt ggt tgt aca ag	48113
HBB③	aat cac gat cat att gcc	39430	Hb1b2-F16	cac cag tca gaa tgg cta ag	48031
HBB④	cag cac aat cac gat cgc	54619	Hb1b2-R16	atg tgg agt aga ggt gtg ca	48701
HBB⑨	att gat gcg tct tct gtc	54597	Hb1b2-F18	atc tgc tgt cat tgc ttg tc	49236
Hb1b2-F1	taa gcc tgc agt atc tgg ta	39378	Hb1b2-R17	tat gca atg gtg tca gcg tt	49736
Hb1b2-R1	agc agt cca act gta tga ag	40313	Hb1b2-F19	agt gtg gac act atg cca ct	49639
Hb1b2-F2	ttg tca tac cat gcc tgc ac	40165	Hb1b2-R19	aag ctg tct gaa agc tgt gg	50329
Hb1b2-R2	gga ttc agt cga gga atg ca	41125	Hb1b2-F20	acc tcc act aat cca ggt ta	50224
Hb1b2-F3	ctg aat gca atg tcc aat gg	40890	Hb1b2-R20	ctg tgt act act tca ctg gc	50847
Hb1b2-R3	ctg tta agt ccc ttc ctg tt	41614	Hb1b2-F21	ggc tga tat ggc cac ttc tt	50737
Hb1b2-F4	gct gac tgt ctc atc gtg at	41496	Hb1b2-R21	cca gaa gtg aag agg tct ca	51357
Hb1b2-R4	aca gta gaa aca gcc tga tg	42080	Hb1b2-F22	tgc aca cat ata tgg atg tg	51269
Hb1b2-F5	gca agc taa tct gtt cat gc	41992	Hb1-b2/3R	tct ctg tgt ctg ctc cta tc	51663
Hb1-b2/1R	gtg atc ttg gaa tga gtc tg	42663	Hb1b2-F23	cta gag acc cat gat tga ac	51537
Hb1b2-F6	agc cat caa tca tca cag tg	42588	Hb1b2-R23	agt ttc tgt gtg aga tga tg	52237
Hb1b2-R6	gac aaa ctc aca aac cac ct	43224	Hb1b2-F24	ctc agc agt gca tgt tgc tt	52132
Hb1b2-F7	cag act aga gaa cca gac at	43148	Hb1b2-R24	gga tcc aga caa gga aac at	52802
Hb1b2-R7	gga gca gaa aga cat agt tg	43761	Hb1b2-F25	gtt tac ctg tca cta gga aa	52689
Hb1b2-F8	act ccc agt gtg agc ata ca	43677	Hb1b2-R25	gct tca gag atg aca acc at	53368
Hb1b2-R8	gaa cta caa cat tgg att gg	44338	Hb1b2-F26	aca gaa tta gct gcg agg at	53292
Hb1b2-F9	atg gat gca ctt cat tag gc	44265	Hb1b2-R26	agg ttg agc aga ata gcc ag	54014
Hb1b2-R9	caa aag gta agg aca act gg	44847	Hb1b2-F9.5CT.FAM	gcc tcc agt ttc tcc tgt at	44485
Hb1b2-F10	caa cag agg aat gga tac ag	44721	Hb1b2-R9.5CT.FAM	atg agt tcc aga aca gcc ag	44602
Hb1b2-R10	aac acc ttg gtc ctg cag ag	45343	Hb1b2-F23GA.TET	agt aag cag gac tca tag gg	51804
Hb1b2-F11	ctc caa cct gcc tat atg tg	45263	Hb1b2-R23GA.TET	cat gca ctg ctg agt agt ga	52107
Hb1b2-R11	cca tca cat aat cat cca cc	45879	Hb1b2-F24GA.HEX	tag gtg ggg act atc aag ta	52494
Hb1b2-F12	act tct gtg ttt gcc agg ca	45796	Hb1b2-R24GA.HEX	agt gaa ggg agc ttg aaa ga	52648
Hb1b2-R12	ctt gac tct agc tgc ata tg	46429			

* Site numbers designate the position of the 3' end of the primer in the mouse reference sequence (Shehee et al., 1989)