Gene Diversity of Chimpanzee ABO Blood Group Genes Elucidated from Intron 6 Sequences

T. Kitano, R. Noda, K. Sumiyama, R. E. Ferrell, and N. Saitou

The human and nonhuman primate ABO blood group gene shows relatively large numbers of nucleotide differences around the exon 7 region. In this study we determined intron 6 sequences for 9 alleles of common chimpanzee and for 3 alleles of bonobo to estimate nucleotide diversities among them. Sequence length polymorphisms are observed in this region as a repeat appears one to five times. From a phylogenetic network of intron 6 sequences of ABO blood group genes for humans, common chimpanzee, and bonobo, parallel substitutions and/or some kinds of convergent events are predicted in the chimpanzee lineage. We also estimated nucleotide diversities for common chimpanzee and bonobo ABO blood group genes; these values were 0.219% and 0.208%, respectively.

The human ABO blood group was discovered by Karl Landsteiner in 1900, and its mode of inheritance as multiple alleles at a single genetic locus was established by Bernstein 25 years later (Crow 1993). The chemical basis of ABO blood group specificities was shown to be carbohydrate structures of glycoproteins and glycolipids through studies performed in the 1950s and early 1960s [see Yamamoto (1995) for a review]. Based on this finding, ABO alleles A and B were considered to code glycosyltransferases, which transfer GalNAc and galactose, respectively. Yamamoto et al. (1990a) determined the cDNA sequences of three common alleles A^1, B, and O, and Yamamoto et al. (1995) determined the genomic organization of the gene. Critical sites for the distinction between A and B activities of the glycosyltransferases have also been identified (Yamamoto et al. 1990a,b, 1991, 1992, 1995). These sites are located in the exon 7 region, and these regions of many other nonhuman ABO blood group genes were sequenced (e.g., Kermarrec et al. 1999; Kominato et al. 1992; Martinko et al. 1993). Saitou and Yamamoto (1997) compared nucleotide sequences of primate ABO blood group genes, and relatively large numbers of nucleotide differences were found among them. These differences were unusually large for human allelic divergence under neutral evolution. In this study we determine and analyze the nucleotide sequences of intron 6 for two chimpanzee species to compare and elucidate the gene diversity among the ABO blood group genes.

From the Laboratory of Evolutionary Genetics, National Institute of Genetics, Mishima, 411-8540 Japan (Kitano, Noda, Sumiyama, and Saitou) and the Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania (Ferrell). K. Sumiyama is currently at the Department of Biology, Yale University, New Haven, Connecticutt. This study was partially supported by grants in aid for scientific studies from the Ministry of Education, Science, Sport, and Culture to N.S. We thank Ingrid Jakobsen for her comments on an earlier version of this article. Address correspondence to N. Saitou at the address above or e-mail: nsaitou@genes.nig.ac.jp. This paper was delivered at a symposium entitled "Genetic Diversity and Evolution" sponsored by the American Genetic Association at the Pennsylvania State University, University Park, PA, USA, June 12–13, 1999.

 91:211–214

Materials and Methods

Sequencing of Genomic DNA

Genomic DNAs of nine common chimpanzees (*Pan troglodytes*) and three bonobo (*Pan paniscus*) were used. PCR reaction was performed using 1× Gene Taq Universal Buffer (Mg^{2+} free) (Nippon Gene), 1.5 mM $MgCl_2$, 0.2 mM dNTP, 10 pmol of upper primer (SN-12; 5′-ACCCCCAGCCAAAGGTGCTGACA-3′ or SN-13; 5′-ACCCTGCCAGCTCCATGTGAC-3′) and lower primer (SN-10; 5′-CGATGCCGTTGGCCTGGTC-3′), and 1 unit of AmpliTaq Gold (Perkin-Elmer). Amplification was carried out in DNA GeneAmp PCR System 2400 (Perkin-Elmer) with the following temperature parameters: 10 min at 95°C followed by 40 cycles of 95°C for 30 s, 65°C for 15 s, and 72°C for 1 min. PCR products were purified using MicroSpin Columns S-300 HR (Pharmacia Biotech). At first, PCR products were sequenced by using the direct sequencing method. Then heterogeneous PCR products were cloned in the TA cloning vectors pCRII (Invitrogen). DNA sequencing was performed on double-stranded plasmid DNA and PCR products using the Dye Terminator Cycle Sequencing Kit and ABI prism 377 DNA sequencer (Perkin-Elmer). A progressive sequencing strategy was carried out with design of further primers to both strands of the DNA.

Table 1. Variant positions of ABO blood group genes in intron 6

Sequence	Nucleotide position	Repeat[a]	Exon 7[b]
	`                                11`		
	`     1122222222245556677777888889900`		
	`2448957457888999057480234501330122`		
	`5299041219017564914435738804582779`		
Common chimpanzee (P. troglodytes)			
c-1	CGATACCCCCTCACCTTTAACGGGATAGTCAGAG	5	II
c-2	CGATACCCCCTCACCTTTAATGGGATAGTCAGAG	5	IV
c-3	CGATACCCCCTCACCTTTAACGGGACAGTCAGAG	5	V
c-4	CGATACCCCCTCACCTTTAACGGGATAGTCAGAG	4	I, II
c-5	CGATACCCCCTCACCTTTAACGGGATAGCCAGAG	4	III
c-6	CGATACCCCCTCAACTTTAACGGGATAGCCAGAG	4	III
c-7	CGATACCCCCTCACCTTTAACGGGATAGCCAGAG	3	III
c-8	CGATACCCCCTCACCCTCGACGAGACAGTCAGAG	3	V
c-9	CGATACCCCCTCACCCTCGACGAGACAGTCCGAG	3	V
Bonobo (P. paniscus)			
b-1	CGATACTCC---ACCTTCAACGGGACAGCCAGAG	3	bI
b-2	CGATACCCCCTCACCTTCAACGGGACAGCCAGAG	3	bII
b-3	CGTTATCCCCTCACCTTCAACGGGATAACCAGAG	1	bIII
Human (H. sapiens)			
h-A	TTATGCCCCCTCACTTTCGGCAGGGCGGCTAGGA	3	A
h-B	TTATGCCCCCTCGCTTTCGGCGGGGCGGCCAAAG	3	B
h-0	TGAAGCCTGCTCACCTCCGACGGAGCGGCCAGGA	3	0

[a] Number of copies of repeat in intron 6 (see Figure 1).
[b] Allele type of exon 7 (Sumiyama et al., unpublished data).

Sequence Analyses

CLUSTAL W version 1.6 (Thompson et al. 1994) was used for constructing the multiple alignments. Phylogenetic networks were constructed following the procedure of Bandelt (1994) and Saitou and Yamamoto (1997). The nucdiv program of the ODEN package (Ina 1994) was used for estimation of nucleotide diversities.

Results and Discussion

Comparison of Sequences

We determined nine alleles for common chimpanzee and three alleles for bonobo of intron 6 sequences of ABO blood group genes (DDBJ/EMBL/GenBank international nucleotide sequence database accession numbers AB031235–AB031246). Table 1 shows the variant sites of ABO blood group genes in intron 6. Three human sequences (AF016622–AF016624; Olsson and Chester 1998) are also shown. O'hUigin et al. (1997) sequenced the intron 6 region for humans, chimpanzees, and gorillas but those were partial sequences. Therefore those data are not included in this table. One bonobo gene (b-1) had a gap (279–281) of three nucleotides. Sequence lengths were 1036–1068 bp. These sequence length differences were caused by different numbers of repeats of the sequence TGGGCTCG. Figure 1 shows a sequence comparison of repeat regions (160–240) of ABO blood group genes in intron 6. There are one to five copies of the repeat.

The Phylogenetic Network of Hominoid ABO Blood Group Genes

Figure 2 shows the phylogenetic network of intron 6 sequences of ABO blood group genes for humans, common chimpanzee, and bonobo. This phylogenetic network was constructed from the sequence data of Table 1. Four sites (25, 90, 748, 800) divide genes of humans and chimpanzees. O'hUigin et al. (1997) presented partial sequences (sites 1–398 and 837–1068) for gorilla, and we determined the branching point of gorilla genes using the consensus sequence of two gorilla B genes. From this phylogenetic network, the branching point of gorilla B genes was located between hu-

```
     160                                                                             240
c-1  TCACGTGGGCTCGTGGGCTCGTGGGCTCGTGGGCTCGTGGGCTCGCCAGGTCGGTAAAACCCAGCTCCTTCTCCAGAGGCT
c-2  TCACGTGGGCTCGTGGGCTCGTGGGCTCGTGGGCTCGTGGGCTCGCCAGGTCGGTAAAACCCAGCTCCTTCTCCAGAGGCT
c-3  TCACGTGGGCTCGTGGGCTCGTGGGCTCGTGGGCTCGTGGGCTCGCCAGGTCGGTAAAACCCAGCTCCTTCTCCAGAGGCT
c-4  TCACGTGGGCTCGTGGGCTCGTGGGCTCGTGGGCTCG--------CCAGGTCGGTAAAACCCAGCTCCTTCTCCAGAGGCT
c-5  TCACGTGGGCTCGTGGGCTCGTGGGCTCGTGGGCTCG--------CCAGGTCGGTAAAACCCAGCTCCTTCTCCAGAGGCT
c-6  TCACGTGGGCTCGTGGGCTCGTGGGCTCGTGGGCTCG--------CCAGGTCGGTAAAACCCAGCTCCTTCTCCAGAGGCT
c-7  TCACGTGGGCTCGTGGGCTCGTGGGCTCG----------------CCAGGTCGGTAAAACCCAGCTCCTTCTCCAGAGGCT
c-8  TCACGTGGGCTCGTGGGCTCGTGGGCTCG----------------CCAGGTCGGTAAAACCCAGCTCCTTCTCCAGAGGCT
c-9  TCACGTGGGCTCGTGGGCTCGTGGGCTCG----------------CCAGGTCGGTAAAACCCAGCTCCTTCTCCAGAGGCT
b-1  TCACGTGGGCTTGTGGGCTCGTGGGCTCG----------------CCAGGTCGGTAAAACCCAGCTCCTTCTCCAGAGGCT
b-2  TCACGTGGGCTCGTGGGCTCGTGGGCTCG----------------CCAGGTCGGTAAAACCCAGCTCCTTCTCCAGAGGCT
b-3  TCACGTGGGCTCG--------------------------------CCAGGTCGGTAAAACCCAGCTCCTTCTCCAGAGGCT
h-A  TCACGTGGGCTCGTGGGCTTGTGGGCTCG----------------CCAGGTCGGTAAAACCCAGCTCCTTCTCCAGAGGCT
h-B  TCACGTGGGCTCGTGGGCTTGTGGGCTCG----------------CCAGGTCGGTAAAACCCAGCTCCTTCTCCAGAGGCT
h-0  TCACGTGGGCTCGTGGGCTCGTGGGCTCA----------------CCAGGTCGGTAAAACCCAGCTCCTTCTCCAGAGGCT
     *********** *                                ************************************
          <======><======><======><======><======>
```

Figure 1. Sequence comparison around the repeat region (sites 160–240) of intron 6 for human and chimpanzee ABO blood group genes. Hyphens and asterisks designate gap and invariant site, respectively. Equal signs surrounded by angled brackets indicate the repeat unit of eight nucleotides.

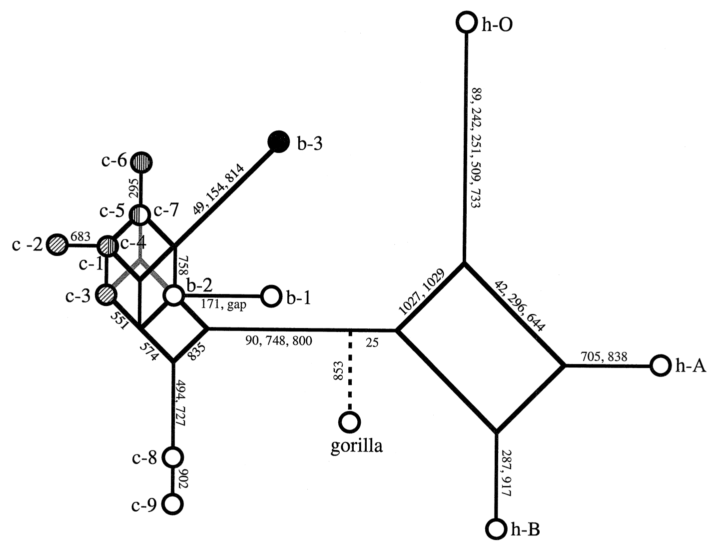

Figure 2. The phylogenetic network of intron 6 sequences for human and two chimpanzee ABO blood group genes. Numbers are nucleotide positions responsible for corresponding edges and edge lengths are proportional to the number of nucleotide differences. Because sites 173–204 include gaps by repeat units, these sites were not used for this phylogenetic network. Alleles are marked by circles, and black, white, vertical stripes, and slant stripes designate 1, 3, 4, and 5 repeats in intron 6 (see also Figure 1), respectively. A gap on the branch connecting to b-1 means a gap (279–281) of three nucleotides. Abbreviations of alleles are the same as in Table 1.

mans and chimpanzee. Saitou and Yamamoto (1997) proposed that B-type alleles evolved independently on human and gorilla lineages. Our result for intron 6 sequences is compatible with theirs.

There is one big rectangle among the human sequences of this phylogenetic network. Sites 1027 and 1029 divide A/O alleles from the B allele. Because these sites are contiguous, there is a possibility of recombination in this region among these alleles.

Four small parallelograms were created for chimpanzee sequences. Therefore parallel substitutions and/or some kinds of convergent events (recombination) must be assumed. For example, c-8, c-9, and c-3 are identical in exon 7 (Sumiyama et al., unpublished data), but there are four nucleotide differences between c-8 and c-3 in intron 6. In these site differences, site changes 551 and 574 are embedded in parallelograms. Because these two sites are contiguous, they may have been transferred between alleles by recombination.

As mentioned above, chimpanzee sequences have variations of repeat number (see Figure 1). It is clear that ABO blood group genes for humans and gorilla have three repeats from O'hUigin et al.'s (1997) and Olsson and Chester's (1998) data. Therefore the common ancestor of humans, chimpanzees, and gorillas probably had three repeats. After the speciation of humans and chimpanzees, the number of repeats has changed in the chimpanzee lineage.

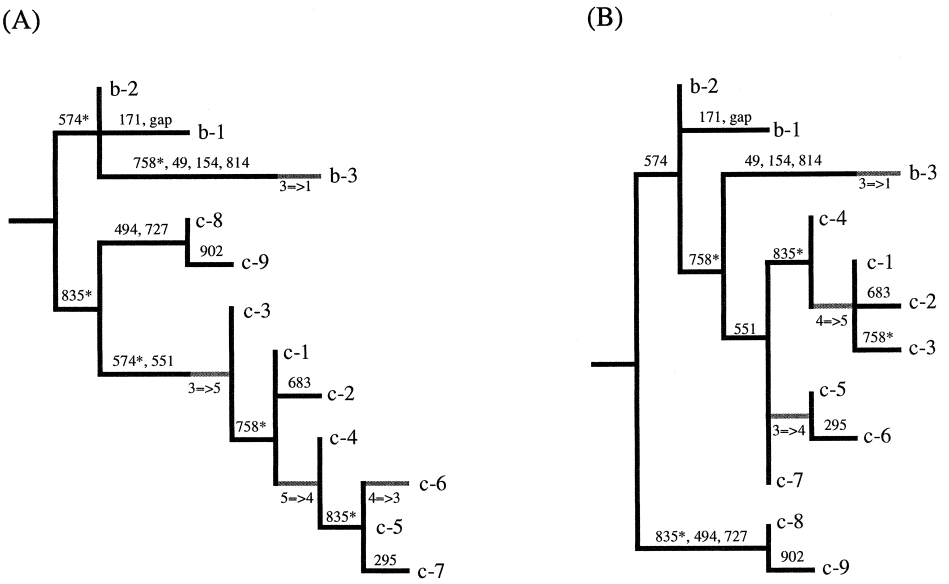

Figure 3. Two possible trees (A and B) deduced from the phylogenetic network of intron 6 sequences (Figure 2) of chimpanzee ABO blood group genes. Numbers on each branch denote the nucleotide positions in which substitutions occurred. Numbers with asterisks signify nucleotide position in which parallel substitutions occurred. Branches with equal signs and angled bracket designate change in the number of repeats.

The Phylogenetic Relationship Among Chimpanzee ABO Blood Group Genes

In the phylogenetic network of intron 6 sequences of ABO blood group genes for humans, common chimpanzee, and bonobo (Figure 2), genes of common chimpanzee and bonobo do not form a clear treelike structure. Therefore we deduced two possible trees for common chimpanzee and bonobo (Figure 3). The tree of Figure 3A is deduced based on species relationships. In this tree, genes for common chimpanzee and bonobo form different clusters defined by sites 835 and 574, respectively. Three genes of common chimpanzee (c-5, c-6, and c-7) form a cluster. The numbers of nucleotide changes and repeat changes in this tree were 16 and 4, respectively.

The tree of Figure 3B is the maximum parsimony tree. This tree was constructed by using the branch and bound method of PAUP version 3.1.1 (Illinois Natural History Survey). Numbers of nucleotide changes and repeat changes in this tree were 15 and 3, respectively. From the parsimony principle, tree (B) is more parsimonious than tree (A). However, we have to assume ancestral polymorphism among common chimpanzee and bonobo in tree (B).

Therefore, we believe that tree (A) is more plausible. In this tree, the number of repeats first increased from three to five, then decreased from five to four, and finally decreased from four to three. In the b-3 gene, repeats decreased from three to one. On the other hand, three genes of common chimpanzee (c-3, c-8, and c-9) have a deletion of 9 bp in exon 7 (Kermarrec et al. 1999; Sumiyama et al., unpublished data). It is difficult to imagine that all the other common chimpanzee ABO genes evolved from a gene having a deletion of 9 bp in exon 7. Therefore it is possible that the c-3 gene is a recombinant between a gene with 9 bp deletion and another common allele (probably a type I or II gene).

Rates of Nucleotide Substitution and Nucleotide Diversities in Chimpanzee ABO Blood Group Genes

To estimate rates of nucleotide substitution (λ) of intron 6 sequences of hominoid ABO blood group genes, we estimated the single lineage number of nucleotide substitutions per site (K) for each species by applying Ishida et al.'s (1995) method. The number of nucleotide substitutions in the common chimpanzee lineage (Figure 3A) is 3.65. The number of nucleotide substitutions in the bonobo lineage (Figure 3A) is 2.67. The number of nucleotide substitutions in the human lineage (Figure 2) is 7.33. Therefore K values of common chimpanzee, bonobo, and human lineages are estimated to be 0.00353, 0.00258, and 0.00711 by Jukes and Cantor's (1969) method. Divergence times (T) between humans and chimpanzee and between common chimpanzee and bonobo are thought to be 5.5 (Kumar and Hedges 1998) and 2.5 (Horai et al. 1992) million years ago, respectively. Thus λ values of common chimpanzee, bonobo, and human lineages are estimated to be 1.41×10^{-9}, 1.03×10^{-9}, and 1.29×10^{-9} by the formula $\lambda = K/T$, respectively.

Li and Graur (1991) estimated λ values in the 5′ and 3′ untranslated regions and at fourfold degenerate sites of protein-coding genes, based on comparisons between humans and rodents for 16 genes. Our estimated values were similar to those for the 5′ and 3′ untranslated regions (averages are 1.96×10^{-9} and 2.10×10^{-9}, respectively). In the case of human ABO blood group genes, three alleles (i.e., A, B, and O alleles) have been observed, but it is known that chimpanzees have two alleles (i.e., A and O alleles). We do not know the actual phenotypes of the chimpanzee ABO blood group genes used in this study. Nevertheless, the λ value of the common chimpanzee lineage is higher than that of the human lineage.

We computed nucleotide diversities of common chimpanzee and bonobo ABO

Table 2. The list of chimpanzees and their genotypes for ABO blood group gene intron 6

Individual	Genotype
Common chimpanzee (*P. troglodytes*)	
CH-46	c-5/c-7
CH-90	c-2/-[a]
CH-206	c-3/c-8
CH-220	c-3/c-8
CH-235	c-6/c-6
CH-75	c-4/c-4
CH-83	c-1/c-4
CH-80	c-1/c-9
CH-76	c-4/c-4
CH-86	c-1/c-1
Bonobo (*P. paniscus*)	
BO-3	b-2/b-2
BO-4	b-1/b-1
BO-5	b-2/b-3
BO-8	b-2/b-3

[a] The chimpanzee CH-90 is heterozygous, but another allele could not be determined.

blood group gene intron 6 sequences. Table 2 shows the genotype of the ABO blood group genes for each individual. The chimpanzee CH-90 has a heterozygous genotype, but another allele could not be determined. Table 3 shows the comparison of nucleotide diversities. Nucleotide diversities of common chimpanzee and bonobo were 0.219% and 0.208%, respectively.

Nachman et al. (1998) analyzed nucleotide diversity at X-linked loci in humans and compared those with previous studies (e.g., Li and Sadler 1991) of nucleotide diversity in humans. To compare directly with variation at autosomal loci they used standardized values, and they concluded that the average level of nucleotide diversity in humans for silent and noncoding DNA is 0.1%, and the highest levels of nucleotide diversity at any locus is about 2.5 times the average for all loci. Therefore, nucleotide diversity among alleles of chimpanzee ABO blood group genes is higher than that of most human genes.

In this study we determined common chimpanzee and bonobo ABO blood group gene intron 6 sequences. We observed sequence length polymorphisms, as a repeat appears one to five times in this intron 6 sequences. From a phylogenetic network of intron 6 sequences of ABO blood group genes for humans, common chimpanzee, and bonobo, parallel substitutions and/or some kinds of convergent events are predicted in the chimpanzee lineage. We also estimated nucleotide diversities among alleles of chimpanzee ABO blood group genes to be about 0.2%. This value is higher than that of most human genes. Wise et al. (1997) compared nuclear and mitochondrial gene diversities of humans and chimpanzees. They proposed that chimpanzees appear to have more mitochondrial gene diversity than humans, whereas humans have more nuclear gene diversity than chimpanzees. Our result on the ABO gene does not agree with their study. In any case, there are not enough nucleotide sequence data for chimpanzee nuclear genes to allow comparison with human nucleotide diversity. Therefore we need more nuclear sequence data for chimpanzees to estimate nucleotide diversity reliably.

Table 3. Nucleotide diversities and rates of nucleotide substitution in chimpanzee ABO blood group genes

Sample size	Number of sites compared (bp)	Nucleotide diversity (%)
Common chimpanzee (*P. troglodytes*)		
19	1052	0.219
Bonobo (*P. paniscus*)		
8	1033	0.208

References

Bandelt HJ, 1994. Phylogenetic networks. Verh Naturewiss Ver Hamburg (NF) 34:51–71.

Crow JF, 1993. Felix Bernstein and the first human marker locus. Genetics 133:4–7.

Horai S, Satta Y, Hayasaka K, Kondo R, Inoue T, Ishida T, Hayashi S, and Takahata N, 1992. Man's place in Hominoidea revealed by mitochondrial DNA genealogy. J Mol Evol 35:32–43.

Ina Y, 1994. ODEN: a program package for molecular evolutionary analysis and database search of DNA and amino acid sequences. Comput Appl Biosci 10:11–12.

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 41:180–188.

Jukes TH and Cantor CR, 1969. Evolution of protein molecules. In: Mammalian protein metabolism (Munro HN, eds). New York: Academic Press; 21–132.

Kermarrec N, Roubinet F, Apoil PA, and Blancher A, 1999. Comparison of allele O sequences of the human and non-human primate ABO system. Immunogenetics 49:517–526.

Kominato Y, McNeill PD, Yamamoto M, Russell M, Hakomori S, and Yamamoto F, 1992. Animal histo-blood group ABO genes. Biochem Biophys Res Commun 189: 154–164.

Kumar S and Hedges SB, 1998. A molecular timescale for vertebrate evolution. Nature 392:917–920.

Li WH and Graur D, 1991. Fundamentals of molecular evolution. Sunderland, MA: Sinauer.

Li WH and Sadler LA, 1991. Low nucleotide diversity in man. Genetics 129:513–523.

Martinko JM, Vincek V, Klein D, and Klein J, 1993. Primate ABO glycosyltransferases: evidence for trans-species evolution. Immunogenetics 37:274–278.

Nachman MW, Bauer VL, Crowell SL, and Aquadro CF, 1998. DNA variability and recombination rates at X-linked loci in humans. Genetics 150:1133–1141.

O'hUigin C, Sato A, and Klein J, 1997. Evidence for convergent evolution of A and B blood group antigens in primates. Hum Genet 101:141–148.

Olsson ML and Chester MA, 1998. Heterogeneity of the blood group Ax allele: genetic recombination of common alleles can result in the Ax phenotype. Transfus Med 8:231–238.

Saitou N and Yamamoto F, 1997. Evolution of primate ABO blood group genes and their homologous genes. Mol Biol Evol 14:399–411.

Thompson JD, Gibson TJ, and Higgins DG, 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680.

Wise CA, Sraml M, Rubinsztein DC, and Easteal S, 1997. Comparative nuclear and mitochondrial genome diversity in humans and chimpanzees. Mol Biol Evol 14:707–716.

Yamamoto F, 1995. Molecular genetics of the ABO histo-blood group system. Vox Sang 69:1–7.

Yamamoto F, Clausen H, White T, Marken J, and Hakomori S, 1990a. Molecular genetic basis of the histo-blood group ABO system. Nature 345:229–233.

Yamamoto F, Marken J, Tsuji T, White T, Clausen H, and Hakomori S, 1990b. Cloning and characterization of DNA complementary to human UDP-GalNAc: Fuc alpha 1 ⇒ 2Gal alpha 1 ⇒ 3GalNAc transferase (histo-blood group A transferase) mRNA. J Biol Chem 265:1146–1151.

Yamamoto F, McNeill PD, and Hakomori S, 1991. Identification in human genomic DNA of the sequence homologous but not identical to either the histo-blood group ABH genes or alpha 1 ⇒ 3 galactosyltransferase pseudogene. Biochem Biophys Res Commun 175:986–994.

Yamamoto F, McNeill PD, and Hakomori S, 1992. Human histo-blood group A2 transferase coded by A2 allele, one of the A subtypes, is characterized by a single base deletion in the coding sequence, which results in an additional domain at the carboxyl terminal. Biochem Biophys Res Commun 187:366–374.

Yamamoto F, McNeill PD, and Hakomori S, 1995. Genomic organization of human histo-blood group ABO genes. Glycobiology 5:51–58.

Corresponding Editor: Masatoshi Nei