Evolution of Primate ABO Blood Group Genes and Their Homologous Genes[1]

Naruya Saitou and Fumi-ichiro Yamamoto†*

*Laboratory of Evolutionary Genetics, National Institute of Genetics, Mishima, Japan; and †The Burnham Institute, La Jolla, California

There are three common alleles (A, B, and O) at the human ABO blood group locus. We compared nucleotide sequences of these alleles, and relatively large numbers of nucleotide differences were found among them. These differences correspond to the divergence time of at least a few million years, which is unusually large for a human allelic divergence under neutral evolution. We constructed phylogenetic networks of human and nonhuman primate ABO alleles, and at least three independent appearances of B alleles from the ancestral A form were observed. These results suggest that some kind of balancing selection may have been operating at the ABO locus. We also constructed phylogenetic trees of ABO and their evolutionarily related α-1,3-galactosyltransferase genes, and the divergence time between these two gene families was estimated to be roughly 400 MYA.

Introduction

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 Felix Bernstein a quarter century later (Crow 1993). Immunodominant ABH antigens were chemically characterized to be carbohydrate structures of glycoproteins and glycolipids through the studies performed in the 1950s and early 1960s (see Yamamoto 1995 for review). Based on this finding, ABO alleles A and B were hypothesized to code for glycosyltransferases which transfer GalNAc and galactose, respectively, while O was hypothesized to be a null allele incapable of coding for a functional glycosyltransferase. Yamamoto et al. (1990*a*, 1990*b*) determined the cDNA sequences of three common alleles, A^1, B, and O, and Yamamoto, McNeill, and Hakomori (1995) determined the genomic organization of the gene.

The purpose of this paper is to analyze the nucleotide and amino acid sequences for primate ABO blood group genes and their homologous genes. Biological significance of the ABO blood group will be discussed based on the clues obtained from phylogenetic analyses of ABO blood group genes.

Materials and Methods

For human ABO blood group genes (hereafter called "ABO genes"), sequence data of Yamamoto et al. (1990*a*, 1990*b*, 1992, 1993*a*, 1993*b*, 1993*c*, 1993*d*) and those of Ogasawara et al. (1996*a*) were used. Table 1 lists the variant nucleotide sites for the human alleles. It should be noted that allele O^2, named by Grunnet et al. (1994), has the identical sequence with O-3, previously reported by Yamamoto et al. (1993*b*). The coordinate system of the nucleotide and amino acid positions of the human and nonhuman primate ABO gene sequences follows that of Yamamoto (1995).

As for nonhuman primate ABO gene sequences, those for two chimpanzees (*Pan troglodytes*), one gorilla (*Gorilla gorilla*), two orangutans (*Pongo pygmaeus*), one crab-eating macaque (*Macaca fascicularis*), and two yellow baboons (*Papio cynocephalus*) are from Kominato et al. (1992), and sequences for three chimpanzees, three gorillas, and two orangutans are from Martinko et al. (1993). Table 2 lists the variant nucleotide sites for the primate sequences. Amino acid sequence differences between positions 152–355 among those human and nonhuman primate sequences are shown in table 3.

As for the primate ABO-related sequences, the hgt4 pseudogene found from the human genome (Yamamoto, McNeill, and Hakomori 1991) and the mouse ABO gene (unpublished data) were used, as well as the α-1,3-galactosyltransferase functional gene cDNA sequences of mice (Larsen et al. 1989), cow (Joziasse et al. 1989), and pig (Strahan et al. 1995). The two human pseudogenes for the α-1,3-galactosyltransferase gene were also included; an ordinary type (Larsen et al. 1990) and a processed type (Joziasse et al. 1991).

Galili and Swanson (1991) determined partial nucleotide sequences of the α-1,3-galactosyltransferase functional gene for squirrel monkey (*Saimiri sciureus*), spider monkey (*Ateles geoffroyi*), and howler monkey (*Alouatta caraya*), and its pseudogenes for chimpanzee, gorilla, orangutan, rhesus monkey (*Macaca mulatta*), green monkey (*Cercopithecus aethiops*), and patas monkey (*Erythrocebus patas*). Henion et al. (1994) also sequenced the complete coding region of this gene for marmoset (*Callithrix* sp.). Those sequences were used to construct a primate-specific tree for this gene.

We searched the latest version of the DDBJ/EMBL/GenBank international nucleotide sequence database by using BLAST (Altschul et al. 1990) and FASTA (Pearson and Lipman 1988), and did not find any sequence homologous to the ABO blood group gene.

Phylogenetic trees were constructed by using the neighbor-joining method (Saitou and Nei 1987) and the maximum-likelihood method (Felsenstein 1981) for relatively distantly related sequences. When the neighbor-joining method was applied, Kimura's (1980) two-pa-

[1] Dedicated to the memory of the late Dr. Motoo Kimura.

Key words: ABO blood group, glycosyltransferase, primates, polymorphism, overdominant selection, phylogenetic network.

Address for correspondence and reprints: Naruya Saitou, Laboratory of Evolutionary Genetics, National Institute of Genetics, 1111 Yata, Mishima-shi, Shizuoka-ken, 411, Japan. E-mail: nsaitou@genes.nig.ac.jp.

Mol. Biol. Evol. 14(4):399–411. 1997

Table 1
Sequence Comparison of the Human ABO Blood Group Locus Alleles

Allele	Nucleotide Position 11 11122245556667778888900 09906962674580279002735 5 91231776496713116239104 9 abcbccbaccaccaacaabaaccb	Reference
A^1-1 (A101)	GGCCGACCCTTCGGCCCGGGGGCC	Yamamoto et al. (1990*a*)
A^1-2 (A102)	......T.....	Yamamoto et al. (1990*a*)
A^1-3 (A103)	****..T.T.....	Ogasawara et al. (1996*a*)
A^1-4 (A104)	****.G.....	Ogasawara et al. (1996*a*)
A^2	****..T.....-	Yamamoto, McNeill, and Hakomori (1992)
A^X	****.....A.....	Yamamoto et al. (1993*c*)
A^3-1	****.....A...	Yamamoto et al. (1993*d*)
cis-AB	****..T.....C.....	Yamamoto et al. (1993*a*)
O-1 (O101)	....-.....	Yamamoto et al. (1990*a*)
O-2 (O201)	TATT-G....A.A..T...A....	Yamamoto et al. (1990*a*)
O-3	****.G.G.....A.....	Yamamoto et al. (1993*b*)
O-4 (O102)	****-....C.....	Ogasawara et al. (1996*a*)
O-5 (O103)	****-G.....	Ogasawara et al. (1996*a*)
O-6 (O202)	****-.....A.A..T...A....	Ogasawara et al. (1996*a*)
O-7 (O203)	****-G....A.A.TT...A....	Ogasawara et al. (1996*a*)
B-1 (B101)	.....G.G...T.A..A.C..A..	Yamamoto et al. (1990*a*)
B-2 (B102)	****.G.G...T.A..A.C.....	Ogasawara et al. (1996*a*)
B-3 (B103)	****.G.G.....A..A.C..A..	Ogasawara et al. (1996*a*)
$B^{(A)}$	****.G.G.....A.C..A..	Yamamoto et al. (1993*c*)
B^3-1	****.G.G...T.A..A.C..AT.	Yamamoto et al. (1993*d*)

Note.—Dots, asterisks, and hyphens mean identical nucleotides as the human A^1-1 sequence, nonexamined positions, and gaps, respectively. Position 1059 can be either 1059, 1060, or 1061. Allele names in parentheses were used in Ogasawara et al. (1996*a*). Letters (a, b, or c) below nucleotide position numbers mean first, second, or third nucleotide positions, respectively.

Table 2
Sequence Comparison of the Primate ABO Blood Group Locus Genes (Positions 435–1003)

Allele	Nucleotide Position 11 44444444444555555555566666777777778888899 00 35566677888011223788922458001167890122479 00 80778914079639174959319211241957363356709 02 ccabcacccbcbccbbcccabcbccccbcacccabccabbcac	Reference
A^1-1 (human)	TTACGGCGCAGAAGAGCTCGGGTGCGCGCGCCCCGGAGGTCGC	1
Chimpanzee-1	....A..A....G......C......T....T...A.....T.	2
Chimpanzee-2	....A..A....G......C......T.....A.....T.	2
Chimpanzee-3 (Patr-1)	**.....G......C.....T...A...****	3
Chimpanzee-4 (Patr-2)	**.T.....G......C......T.....A...****	3
Chimpanzee-5 (Patr-3)	**...C....C.G......C......T.....A...****	3
Gorilla-1	.....G.....AC.....A.	2
Gorilla-2,4,5 (Gogo-1,3,4)	**.....G.....AC....****	3
Gorilla-3 (Gogo-2)	**.....G.....AC..A.****	3
Gorilla-6 (Gogo-5)	**.....G.....T.....AC....****	3
Orangutan-1	.....G.G......A..T....CT....A..A....	2
Orangutan-2	.....G.G......A..T....CT....A.....	2
Orangutan-3 (Popy-1)	**T.....TT.GT..G..A..T....C.....****	3
Macaque	GC....T.GG..GTG.TC..A...TA.CT...TT..G..CT.G	2
Baboon-1	CC....T.GG..GTG.TC....C..A.C....TT.AG..CT.G	2
Baboon-2	CC....T.GG..GTG.TC....C..A.CT...TAC.G..CT.G	2

Note.—Dots and asterisks mean identical nucleotides as the human A^1-1 sequence and nonexamined positions, respectively. Allele names in parentheses were used in Martinko et al. (1993). Letters (a, b, or c) below nucleotide position numbers mean first, second, or third nucleotide positions, respectively. References are as follows: 1, Yamamoto et al. (1990*a*); 2, Kominato et al. (1992); 3, Martinko et al. (1993).

Table 3
Amino Acid Sequence Comparison of the Primate ABO Blood Group Enzymes (Positions 152–355)

Allele	Amino Acid Position 1111111111222222222333 5556677999111346678935 5 3673946578046506863142 5
Human A^1-1	TPATQERFERVMFGSLGEADAR#
Human A^1-2	.L.....
Human A^2	.L.....E
Human A^X	.....A.....
Human A^3-1	.....N...
Human *cis*-AB	.L.....A.....
Human B-1	.....G.....S.MA.....
Human $B^{(A)}$	.....G.....MA.....
Human B^3-1	.....G.....S.MA....W.
Human O-3	.....G.....R.....
Chimpanzee-1,2	.....Q.....S**
Chimpanzee-3	..P.....Q.....****
Chimpanzee-4	.L.....Q.....****
Chimpanzee-5	.....Q.....****
Gorilla-1	.....MA...T**
Gorilla-2, 4, 5, 6	.....MA..****
Gorilla-3	.....MAL.****
Orang-1	.....G.....T...T..**
Orang-2	.....G.....T.....**
Orang-3	S...LGLL...I..T....****
Macaque	...A.G...Q...A.....**
Baboon-1	...A.G....A..A.....**
Baboon-2	...A.G....A..A.MA....**

NOTE.—"#" at position 355 means stop codon.

rameter method was used for estimating numbers of nucleotide substitution, while Kimura's (1983) formula for approximating Dayhoff matrix distances was used for estimating numbers of amino acid substitutions. CLUSTAL W (Thompson, Higgins, and Gibson 1994) was used for multiple alignment and the neighbor-joining method for amino acid sequences, NJBOOT2 (kindly provided by Koichiro Tamura) was used for the neighbor-joining method for nucleotide sequences, and DNAML of PHYLIP 3.5c (Felsenstein 1993) was used for the maximum-likelihood method. After the construction of a tree, each branch length (number of amino acid or nucleotide substitutions per site) was recomputed to estimate the integer number of substitutions in the sequences compared, applying Ishida et al.'s (1995) method.

The evolutionary history of a gene should be presented as a tree. When we analyze real sequence data, however, this tree structure may not be clearly observed. In this case, construction of phylogenetic networks is useful for delineating anomalies in the history of gene trees. When two nucleotide positions show incongruent partition (or configuration) patterns, a discordancy diagram (Fitch 1977) appears. A phylogenetic network can be considered as a generalization of this discordancy diagram (Saitou 1996). Bandelt (1994) described the mathematical properties of the phylogenetic network method, and Bandelt et al. (1995) proposed to construct a "median network" that contains all the equally parsimonious trees. Because of these reasons, the phylogenetic-network method and the maximum-parsimony method (Fitch 1977) were applied for closely related sequences. In this study, constructed networks are not median networks, for all the possible nodes are connected. PAUP 3.1.1 (Illinois Natural History Survey) was used for some maximum-parsimony analyses.

Results and Discussion

Comparison of the Five Human ABO Gene Sequences

We first compared five cDNA sequences for the human ABO gene presented by Yamamoto et al. (1990*a*, 1990*b*). A^1-1, A^1-2, and B alleles cover the complete coding sequences, while O-1 and O-2 allele sequences lack three nucleotides corresponding to the first codon (Yamamoto et al. 1990*b*). Polymorphic nucleotide sites are shown in table 1. Because we dealt with closely related sequences, the phylogenetic network method and the maximum-parsimony method were used. Figure 1 shows the unique phylogenetic network (*A*) and four equally parsimonious trees (*B–E*). Because a partition defined by a single gap (position 261) and one defined by a nucleotide configuration at position 297 are incompatible, there is one loop in network *A*. If we cut one of those branches that form this loop, four alternative trees are produced (trees *B–E*). Since allele A^1-1 was assumed to be identical with the ancestral sequence of human ABO genes (see below), this information was used to locate the root (designated by a broken line) in those equally parsimonious trees. It should be noted that topologies of trees *B* and *C* are identical and only some branch lengths are different.

Although 18 changes are required in all four trees, two insertion/deletion events are involved in trees *D* and *E*, while only one deletion is required for trees *B* and *C*. Saitou and Ueda (1994) showed that the evolutionary rate of insertions and deletions in primates was about one order slower than that of nucleotide substitutions. Therefore, trees *B* and *C* are much more probable than trees *D* and *E*. In the former two trees, however, the number of nucleotide substitutions along the branches leading to O-1 and O-2 are quite different. Interestingly, allele O-1 is identical to allele A^1-1 except for the single nucleotide deletion, while allele O-2 is different from allele O-1 with nine nucleotides. It is possible that allele O-1 might be a product of intragenic recombinations or gene conversions, because those events homogenize different alleles. However, we have to assume at least two such events to explain the observed sequence pattern. It suggests that intragenic recombinations or gene conversions occur rather frequently at this locus. In fact, Ogasawara et al. (1996*b*) did find a probable recombinant between an O and a B allele.

We estimated divergence times among human ABO alleles based on nucleotide differences. Nucleotide difference per site between A^1-2 and B-1 is 0.008 (=8/1,062), while that between B-1 and O-2 is 0.013 (=14/1,059), respectively. These values are much larger than the average number of nucleotide differences among different nucleotide sequences within a locus in human (Li and Sadler 1991). Saitou (1991) estimated the average number of nucleotide differences of noncoding regions between human and chimpanzee to be 0.014/site. If we

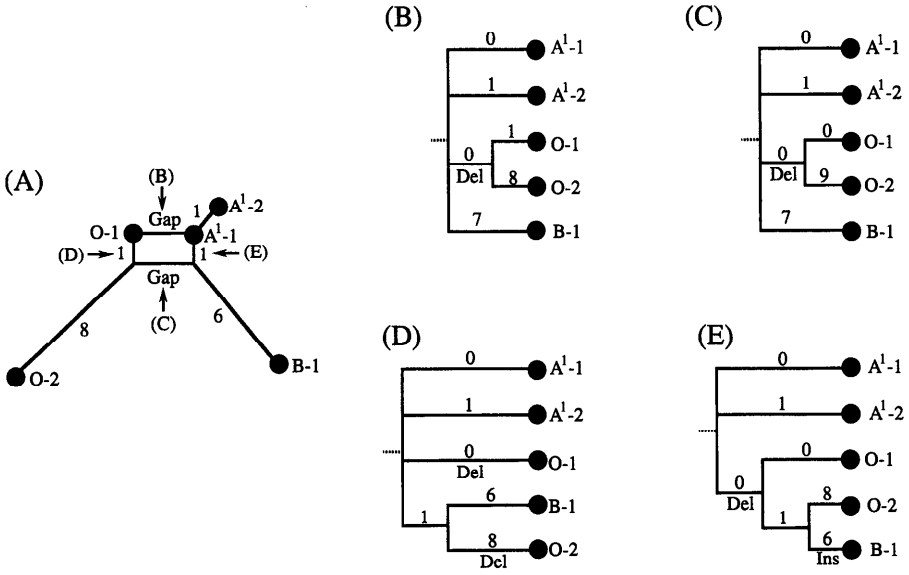

FIG. 1.—The phylogenetic network (*A*) and four possible maximum-parsimony trees (*B–E*) for five human ABO blood group alleles. Almost-complete coding region (1,059 bp) sequences excluding the first three nucleotides were compared. Data are from Yamamoto et al. (1990*a*, 1990*b*). Numbers on each branch are estimated numbers of nucleotide substitutions. Del and Ins mean deletion and insertion of a single nucleotide, respectively.

use 5 Myr for the divergence time between human and chimpanzee, the rate of nucleotide substitution becomes 1.4×10^{-9}/site/year, and those allelic nucleotide differences for the human ABO genes correspond to 2.7–4.7 Myr. Those values are unusually large for different alleles of a typical human locus (see discussion below), and yet those divergence times may still be underestimations. Because the ABO gene is functional, use of an evolutionary rate for noncoding-region DNA is expected to give underestimations of the divergence time. Gene conversion and recombination are also possible causes for underestimation, for they homogenize different alleles.

The Phylogenetic Network of Human ABO Gene Alleles

We used all of the 20 sequence data shown in table 1. Because many of the sequences analyzed did not identify the nucleotides including variant positions 109, 191, 192, and 203, those positions were not used for the following analysis. The maximum-parsimony method was first used, and 39 equally maximum parsimonious trees were produced by using PAUP 3.1.1 with the branch-and-bound option (trees not shown). Clearly, the maximum-parsimony method is not appropriate for delineating the complex nature of the ABO allele polymorphism.

We then used the phylogenetic-network method, and the resulting network is presented in figure 2. Ogasawara et al.'s (1996*a*) phylogenetic network of ABO common alleles can be considered to be a subset of this network. There are multiple loops in that network, indicating the existence of mutually incompatible sites. For example, position 803 separates all the five B alleles and *cis*-AB alleles from the remaining A and O alleles, but this partition pattern is incompatible with that for position 526, where all the B alleles and the O-3 allele are separated from the remaining alleles. Because the *cis*-AB allele is rare, it is probable that the G-to-C change at position 803 occurred independently in the B allele lineage and in the *cis*-AB formation.

Let us consider the rectangle in the B allele group. When we use the maximum-parsimony method, there is only one solution in this sequence group. In this case allele $B^{(A)}$ is assumed to be identical with an ancestral sequence of the B allele group. However, this allele is very rare and it is possible that the allele appeared in the human population only recently. If so, we have to consider subparsimonious solutions that are embedded in the rectangular structure of figure 2. This again shows the superiority of the network to the maximum-parsimony analysis.

The O-3 allele with no deletion is clearly in a different lineage from the remaining O alleles with a single nucleotide deletion at position 261. It is likely that the nucleotide change at position 802 that caused the amino acid change from glycine to arginine at position 268 (see tables 1 and 3) was responsible for the nonfunctionalization of the O-3 allele (Yamamoto et al. 1993*b*). It should be mentioned that the amino acid position 268 is one of the two crucial positions that determine the donor nucleotide-sugar substrate specificity differences between A and B transferases (Yamamoto and Hakomori 1990; Yamamoto and McNeill 1996).

The Phylogenetic Network of Primate ABO Alleles

Figure 3 is the phylogenetic network for human and nonhuman primate sequences. This network was produced by combining sequence data of tables 1 and 2, except that six rare alleles observed in human populations were excluded. When the same nucleotide change is shared with different species groups (three species groups are as follows: human-chimpanzee-gorilla, orangutan, and macaque-baboon), this change is as-

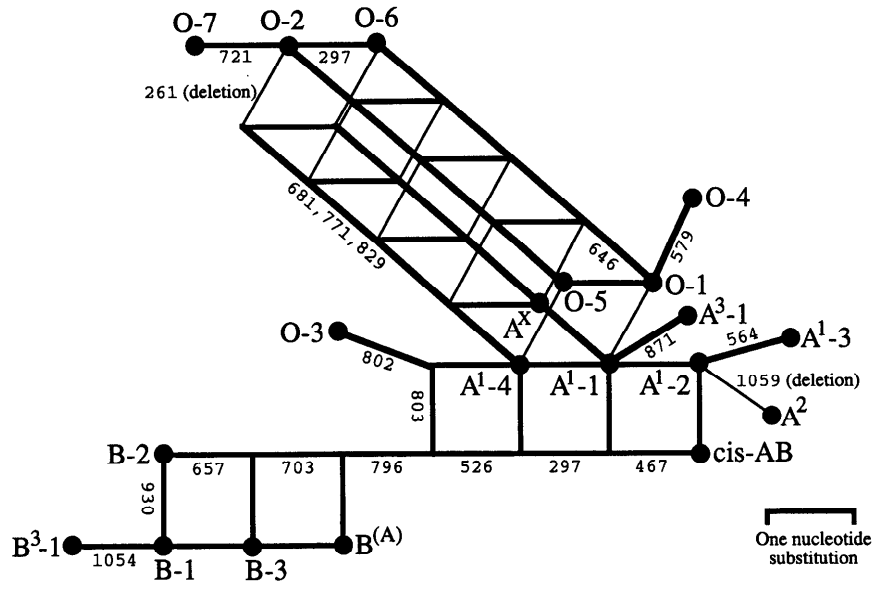

FIG. 2.—The phylogenetic network for 20 human ABO blood group alleles. See table 1 for the sequence data used. Thick lines and thin lines denote nucleotide substitutions and deletions, respectively. Numbers on some branches are nucleotide positions responsible for those branches. Full circles denote observed human alleles.

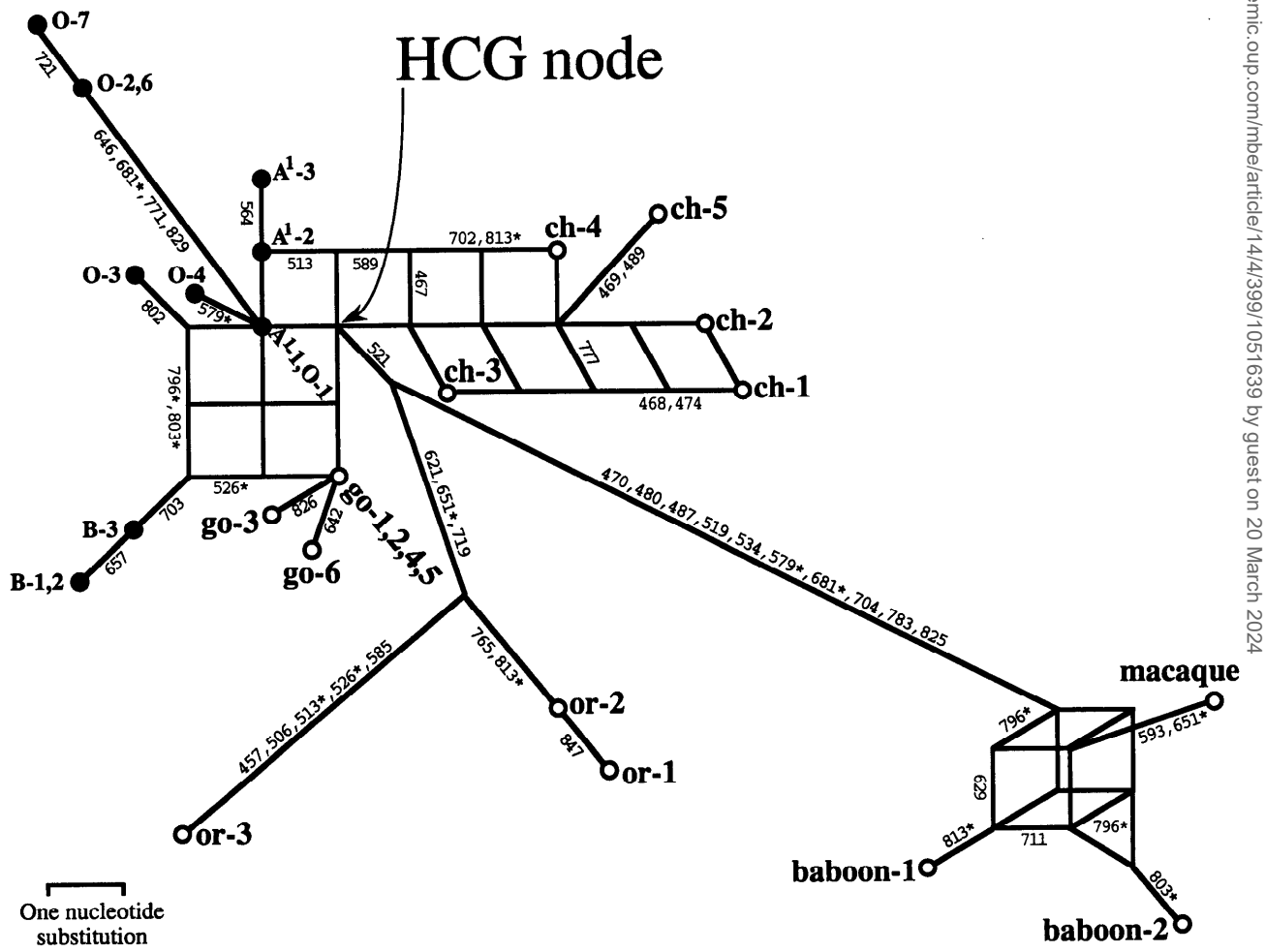

FIG. 3.—The phylogenetic network for common alleles of the primate ABO blood group locus. Abbreviations: ch = chimpanzee, go = gorilla, or = orangutan. See tables 1 and 2 for the sequence data used. Numbers on some branches are nucleotide positions responsible for those branches. Numbers with asterisks signify nucleotide positions in which parallel substitutions occurred. Full and open circles denote human and nonhuman ABO alleles, respectively. "HCG node" is the position of the common ancestral sequence for human, chimpanzee, and gorilla.

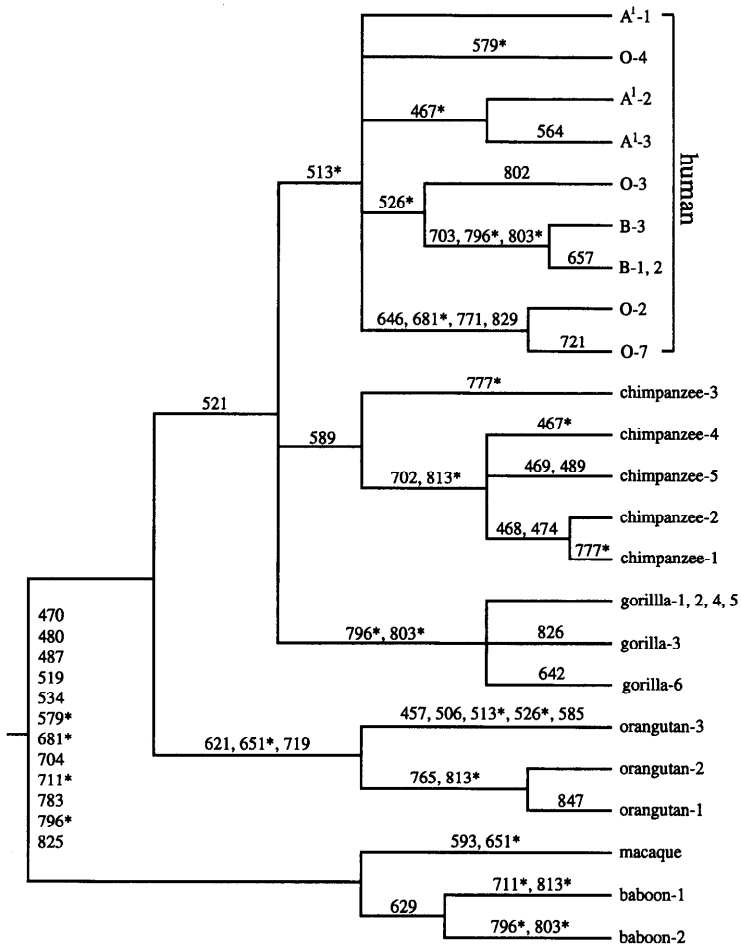

FIG. 4.—The estimated phylogenetic tree for primate ABO blood group alleles based on the phylogenetic network of fig. 3. Numbers on each branch denote the nucleotide positions in which substitutions occurred. Numbers with asterisks signify nucleotide position in which parallel substitutions occurred.

sumed to be caused by parallel substitutions, and no link was produced. This is because it is unlikely for a pair of sequences from different species groups to share a mutation by descent. That is, all these cases were assumed to be parallel substitutions. For example, three human O alleles (O-2, O-6, and O-7) and three Old World monkey sequences (macaque, baboon-1, and baboon-2) share nucleotide A at position 681. This pattern is assumed to be the result of two parallel substitutions, as indicated by an asterisk after the position number.

Let us examine the characteristics of the network in figure 3. First of all, sequences for hominoids and Old World monkeys are clearly distinguished. There are 10 unequivocal substitutions accumulated in the branch connecting these two groups. Within hominoids, three orangutan sequences form a distinct cluster; three substitutions are allocated to the branch diverged from the common ancestor of the human-chimpanzee-gorilla lineage and the orangutan lineage. These clusterings are compatible with the established primate phylogeny by molecular data (e.g., Horai et al. 1995).

Within hominoids, we now observe some incompatible partitions that are responsible for loops. For example, nucleotide configurations at positions 796 and 803 are incompatible with those at positions 513 and 526 (see tables 1 and 2). These relationships produced four consecutive rectangles connecting human and gorilla sequences in figure 3. We designated the upper right node of those two rectangles as the HCG node, because species-specific clusters of human, chimpanzee, and gorilla alleles can be converged onto this node, and the HCG node sequence is considered to be the common ancestor for all the observed alleles of human, chimpanzee, and gorilla.

The Phylogenetic Tree of Primate ABO Gene Alleles

Figure 4 is the phylogenetic tree of the primate ABO genes based on the phylogenetic network of figure 3. All the parallelograms that existed in the network were eliminated by cutting some of those edges. In the case of the four rectangles around human and gorilla sequences, for example, the edge connecting a gorilla sequence (go-1,2,4,5) and a node neighboring the human B-3 sequence was eliminated, because an extant gorilla sequence is unlikely to be identical with an ancestral sequence of human alleles. The resulting tree of figure 4 is one of 10 equally parsimonious trees (not shown) that were found by using PAUP with the branch-and-bound option. All of the 60 substitution events were

Table 4
Observed Nucleotide Substitution Matrix

(A)

	New				
Old	A	C	T	G	Sum
A.....	—	0	2	0	2
C.....	2	—	13	2	17
T.....	2	3	—	0	5
G.....	15	7	2	—	24
Sum	19	10	17	2	48

(B)

Transitions		Transversions	
A ↔ G....	18	A ↔ C	2
C ↔ T	22	A ↔ T	4
Sum	40	G ↔ C	11
		G ↔ T	3
		Sum	20

unambiguously located at one of the branches of the tree.

It should be noted that the topology of the tree in figure 4 is different from that of Martinko et al. (1993), where human B allele was clustered with gorilla B alleles. Although this relationship was also one of most parsimonious trees in our data set, we believe that the topology of figure 4 is more probable than their tree, according to our argument above on the network of figure 3.

This tree indicates that the common ancestral gene for the hominoid and Old World monkey ABO blood group is A type, and three B alleles evolved independently on the human, gorilla, and baboon lineages. Those changes correspond to the two amino acid substitutions (266: L→M and 268: G→A; see table 3) that are responsible for the change of substrate specificity (Yamamoto and Hakomori 1990).

It has been known that the human ABO-like blood group also exists in nonhuman primates. Moor-Jankowski, Wiener, and Rogers (1964) summarized the data, and they are presented in table 5 in an abbreviated form. Because orangutan and gibbon both possess A and B alleles, it is possible that B alleles evolved independently in those lineages too. Macaque species and New World monkeys also have both A and B alleles. However, this prediction (repeated emergence of the B allele) awaits the determination of nucleotide sequences for these species in the future.

Because all the nucleotide changes were inferred in the gene tree of figure 4, directions of changes could be determined except for 12 substitutions on the branch connecting hominoid and Old World monkeys. This process involved the reconstruction of the nucleotides in all the interior nodes, and those reconstructions are considered to be reliable because the compared sequences are closely related (see Yang, Kumar, and Nei 1995). It has enabled us to estimate the pattern of nucleotide substitutions, as shown in table 4. Transitions occurred two times as frequently as transversions, so the transition parameter (α) for the two-parameter model (Kimura 1980) is roughly four times higher than the transversion parameter (β). When we consider the direction of substitutions, it is clear that there is a bias toward AT richness; G-to-A and C-to-T changes are much more abundant than A-to-G and T-to-C changes.

Possibility of Natural Selection on ABO Genes

We found unusually large coalescence times for human ABO alleles. It is of interest to compare those with coalescence times for nonhuman primate ABO genes. We thus estimated coalescence times of the ABO alleles in each species in figure 4 as follows. The numbers of nucleotide substitutions from the most recent ancestor sequence to the extant sequences in each species were first estimated applying Ishida et al.'s (1995) method. The resultant values are 0.0072, 0.0085, 0.0023, 0.0132, and 0.0070 for human, chimpanzee, gorilla, orangutan, and baboon, respectively. We then used for the calibration the evolutionary rate (1.4×10^{-9}) based on the human–chimpanzee comparison as used in the previous section. The results are 2.6, 3.0, 0.8, 4.7, and 2.5 Myr for human, chimpanzee, gorilla, orangutan, and baboon, respectively. It is interesting that not only human, but also chimpanzee, orangutan, and baboon showed large coalescence times.

Table 5
The ABO Blood Groups in Nonhuman Primates (Adapted from Moor-Jankowski, Wiener, and Rogers 1964)

Common Name	Latin Binomen[a]	Observed Phenotype[b]
Chimpanzee.....	*Pan troglodytes*	A (113), O (17)
Gorilla	*Gorilla gorilla*	B (2)
Orangutan.....	*Pongo pygmaeus*	A (22), B (1), AB (3)
Gibbon	*Hylobates lar*	A (2), B (5), AB (4)
Baboons	*Papio anubis, P. cyanocephalus*	A (36), B (17), AB (34)
Rhesus macaque	*Macaca mulatta*	B (10)
Pigtailed macaque.....	*Macaca nemestrina*	B (5)
Java macaque.....	*Macaca irus*	A (8), B (1), AB (3), O (1)
Sulawesi crested macaque[c].....	*Macaca nigra*	A (7), B (2)
Squirrel monkey	*Samimiri sciurea*	A (3), O (1)
Cebus monkey.....	*Cebus albifrans*	B (3), O (1)

[a] Names currently used in primate taxonomy are listed.

[b] Numbers in parentheses are observed numbers.

[c] Listed as "Celebes black ape" in Moor-Jankowski, Wiener, and Rogers (1964).

We applied the same method to the 238-bp HOX2 intergenic sequence data of Ruano et al. (1992) to see if the long coalescence times are special for the ABO gene. The numbers of nucleotide substitutions from the most recent ancestor sequence to the extant sequences were estimated to be 0.008 and 0.011 for chimpanzee and gorilla, respectively. The coalescence times are thus estimated to be 3 Myr for chimpanzee and 3.8 Myr for gorilla. The HOX2 intergenic region is assumed to be under neutral evolution (Kimura 1983). Although only short nucleotide sequences were compared, it is thus suggested that nonhuman primates have large coalescence times even for a DNA region under neutral evolution. It is possible that nonhuman primates are highly geographically isolated and this caused the coalescence times to be much longer than that for human. Such long coalescence times for nonhuman primates are a clear contrast to human, where the nucleotide difference per site between randomly chosen genes of a locus was estimated to be only 0.0004 (Li and Sadler 1991). This situation is similar to that of primate MHC genes, where the coalescence time of the class II DRB1 locus was estimated to be more than 30 MYA, probably caused by a strong balancing selection (Takahata 1993*a*). Therefore, long coalescence times estimated for the human ABO gene both from the complete cDNA region and from a partial cDNA region suggest the possibility of the existence of some kind of balancing selection at this gene, at least for human.

Higher rates of nonsynonymous substitutions to synonymous ones were observed for the antigen recognition sites of human and mice MHC class I and class II proteins (Hughes and Nei 1988; Nei and Hughes 1991). Those higher rates are considered to be clear evidence of the existence of positive selection on those MHC genes. We, therefore, estimated numbers of synonymous and nonsynonymous substitutions between human ABO A^1-1 allele and B-1 allele by using Nei and Gojobori's (1986) method. The ODEN computer package (Ina 1994) was used for computation. Estimated numbers of synonymous and nonsynonymous sites out of the complete cDNA sequences are 258.2 and 800.8, respectively (the initiation codon ATG was eliminated from the comparison, the sum being 1,059). Because the numbers of synonymous and nonsynonymous nucleotide differences between the two alleles are 3 and 4, respectively (see table 1), the proportions of synonymous and nonsynonymous differences were 0.0116 (=3/258.2) and 0.0050 (=4/800.8), respectively. The numbers of synonymous and nonsynonymous substitutions were thus estimated to be 0.0117 ± 0.0068 and 0.0050 ± 0.0025, respectively. The ratio of nonsynonymous/synonymous substitutions became 0.43. Takahata (1993*b*) estimated this ratio for the human ABO gene, and it turned out to be 2.0. That comparison was, however, based on partial ABO sequences (405 bp long) corresponding to the region sequenced by Martinko et al. (1993) for nonhuman primate ABO genes. Therefore, it is not clear whether that high ratio can be considered evidence for the existence of positive selection on the ABO gene.

Table 6
Goodness of Fit Between Observed and Expected Numbers of Nucleotide Substitutions for the ABO Genes

	Numbers of Sites		
N	Observed	Expected	χ^2
0	363	351.8	0.36
1	33	49.5	5.50
2	6	3.5	1.79
≥3	3	0.2	39.20
Sum	405	405.0	46.85

NOTE.—$\lambda = 0.14$, $P < 0.001$ (df = 3). *N*: number of nucleotide substitutions per site.

We saw that B alleles evolved independently at least three times in primate evolution in the previous section. To examine whether this phenomenon is statistically significant or not, we performed a goodness-of-fit test between the observed and the expected numbers of substitutions per site for the ABO gene based on the tree of figure 4 (table 6). Expected numbers of substitutions were computed applying a Poisson distribution under the assumption of equal substitution rate at every site. The mean number (λ) of nucleotide substitutions per site per whole tree was estimated to be 0.14. A highly significant difference ($P < 0.001$) between observed and expected numbers was observed (table 6). This difference was mainly caused by unusually high substitutions at the nucleotide sites (796 and 803) responsible for the functional difference between A and B transferases. Unless those positions are mutational hot spots, this recurrent occurrence of B alleles does not seem to reflect the pattern of mutations. Because the neutrally evolving genes are expected to accumulate nucleotide changes according to their mutation pattern (Kimura 1983), we have to consider the existence of some kind of positive selection on the ABO blood group locus. The best candidate may be the overdominant selection, for the emergence of new alleles produces heterozygotes.

Natural selection on the ABO gene has long been studied (e.g., Chung and Morton 1961; Hiraizumi 1990). However, all of the studies were based on nonmolecular data, and only a short time span was considered to detect any type of selection. It is now increasingly clear that analysis of long-term evolution is more powerful than that of short-term evolution. Therefore, we believe that our present study, based on the accumulation of mutations over a long evolutionary time period, opened a new aspect for the study of natural selection on the ABO gene.

Phylogenetic Relationship of the ABO Genes and Their Related Genes

It has been known that α-1,3-galactosyltransferase genes (hereafter abbreviated as GAL) are homologous to the ABO genes (e.g., Joziasse 1992). We retrieved four amino acid sequences translated from GAL gene nucleotide sequences (DDBJ/EMBL/GenBank international nucleotide sequence database accession numbers were M85153, J04989, L36152, S71333, and J05175 for mouse GAL, bovine GAL, pig GAL, marmoset GAL,

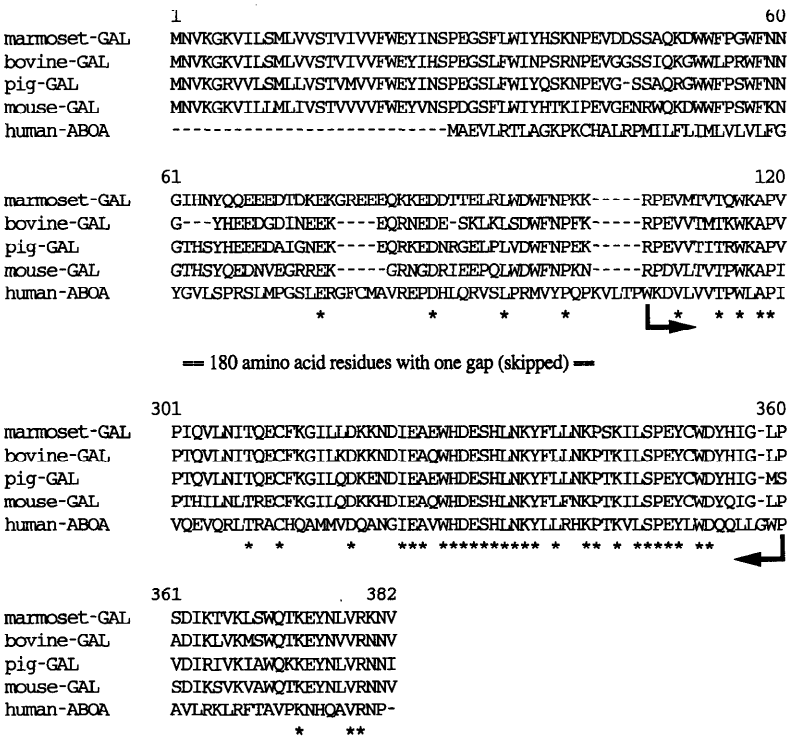

FIG. 5.—A partial multiple alignment of four α-1,3-galactosyltransferases and allele A of the human ABO amino acid sequences. Asterisks denote invariant amino acid positions, and arrows denote the region of amino acid positions used for phylogenetic analysis.

and human ABO-A, respectively), and multiple–aligned them (see fig. 5). We used only the conserved 254 amino acid positions for the phylogenetic analysis. There was only one amino acid gap in this region.

Figure 6 shows a phylogenetic tree of this gene constructed by using the neighbor-joining method. Amino acid sequences were used, and the product of human ABO A allele was used as the outgroup. Mouse GAL sequence and the remaining three mammalian sequences are clearly separated with 100% bootstrap probability, while bovine and pig GAL sequences are clustered with a slightly lower bootstrap value (81%). The numbers of amino acid substitutions from the common ancestor (node A) of the four mammalian species to the present day proteins vary greatly; the smallest value (11) is for the mouse lineage and the largest one (49) is for the pig lineage. If we assume the divergence time between order rodentia and other mammalian orders to be roughly within the range of 115–129 MYA as suggested by Easteal, Collet, and Betty (1995), the evolutionary rate of GAL proteins for those two lineages can be estimated as follows. Because we consider single lineages, the rates of amino acid substitution (per site per year) are estimated to be 0.34×10^{-9} to 0.38×10^{-9} (=[11/253]/[115–129 MYA]) for the mouse lineage and 1.5×10^{-9} to 1.7×10^{-9} (=[49/253]/[115–129 MYA]) for the pig lineage. Both rates are within the range of evolutionary rates for typical proteins (e.g., see Nei 1987).

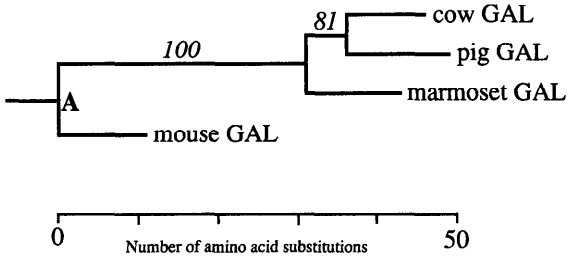

FIG. 6.—A phylogenetic tree for four α-1,3-galactosyltransferases based on 253 amino acid residue sequences. The product of the human ABO A allele was used to locate the root (node A) of the tree. The neighbor-joining method was used for Kimura's (1983) distances. Numbers on two interior branches are bootstrap probabilities (%).

We also reconstructed phylogenetic trees of ABO and GAL genes using 455-bp-long nucleotide sequence data (fig. 7). These pseudogenes and mouse ABO gene are included in this tree. Because mouse sequences separated first among the mammalian sequences in both ABO and GAL genes, we equated the location of the nodes (S1 for GAL and S2 for ABO in fig. 7) for speciation of the mouse lineage and the other mammalian lineages in the tree. An ABO pseudogene (hgt4) found from the human genome seems to diverge before the rodent lineage diverged, although the bootstrap value (50%) supporting this pattern is relatively low.

Kominato et al. (1992) observed hybridization of human ABO gene with a wide range of mammalian genomic DNA. Human ABO-like blood groups were found in various mammals, and it is now clear that probably all the mammalian species have the homologue of the human ABO gene. In fact, Yamamoto et al. (unpublished data) found a human ABO homologue from the mouse genome (see fig. 7), and Ellegren et al. (1994) suggested that the pig blood group gene EAA is homologous to the human ABO gene based on the chro-

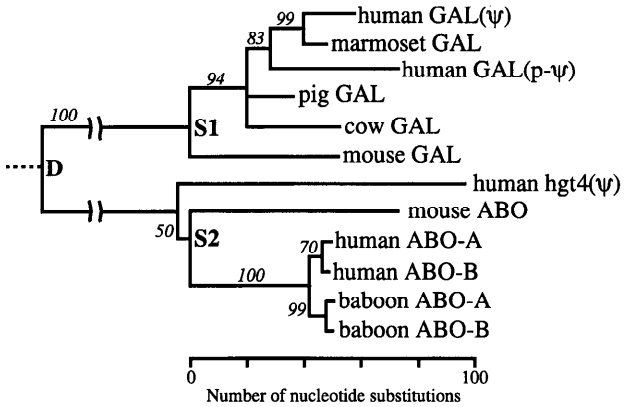

FIG. 7.—A phylogenetic tree for the human ABO gene and its homologous nucleotide sequences. The neighbor-joining method was used for Kimura's (1980) distances based on the 455-bp sequence data. Numbers on internal branches are bootstrap probabilities (%). Node D denotes the duplication event that produced ABO and GAL genes, while nodes S1 and S2 denote the common ancestors of the rodent lineage and other mammalian lineage for GAL and ABO genes, respectively.

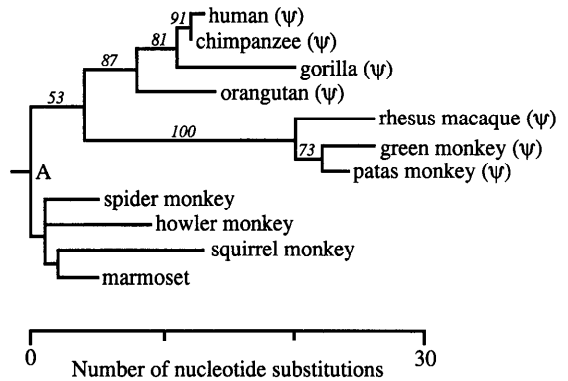

FIG. 8.—A phylogenetic tree for the primate α-1,3-galactosyltransferase functional genes and its pseudogenes based on the 369-bp sequence data. Bovine GAL gene was used as the outgroup. The maximum-likelihood method was used under the assumption of monophyly of New World monkeys. Bootstrap probabilities (%) on interior branches are for the neighbor-joining tree.

mosome comparative map between pig and human. Furthermore, the homologue of the hgt4 pseudogene should also exist in all the mammalian species genomes, if the tree topology of figure 7 is correct.

We estimated the rate of nucleotide substitutions for ABO and GAL genes under the assumption of constancy of the evolutionary rate after the divergence between the rodent and the other mammalian lineages at 115–129 MYA (Easteal, Collet, and Betty 1995). The average numbers of nucleotide substitutions per site from node S1 to extant functional GAL genes and from node S2 to extant functional ABO genes were estimated to be 0.105 (=47.8/455) and 0.134 (=61.6/455), respectively, applying Ishida et al.'s (1995) method. Thus, the rates within the mammalian lineage for ABO and GAL functional genes are estimated to be 1.0×10^{-9} to 1.2×10^{-9} (=0.134/[115–129 MYA]) and 0.8×10^{-9} to 0.9×10^{-9} (=0.105/[115–129 MYA]), respectively. These evolutionary rate estimates are somewhat lower than that (ca. 5×10^{-9}) for mammalian pseudogenes (Nei 1987). If we assume the evolutionary rate of those genes to be similar (ca. 1.0×10^{-9}) throughout their evolution, the time of the gene duplication (node D) that produced ABO and GAL genes is estimated as follows. Because the branch connecting nodes S1 and S2 in figure 7 was estimated to have 238 nucleotide substitutions out of 455 nucleotides compared, the divergence time between those two nodes is estimated to be 260 MYA (=[238/455]/[$2 \times 1.0 \times 10^{-9}$]). Thus, by adding the divergence time estimate (115–129 MYA) of nodes S1 and S2, the divergence time of the node D from the present time becomes 375–389 MYA, or roughly 400 MYA. It seems that the time of gene dupulication producing ABO and GAL genes may be around the emergence of vertebrates (ca. 500 MYA).

ABO and GAL genes share characteristics other than sequence homology. The chromosomal location of the human ABO gene has been mapped to 9q34.2, the tip of the long arm of chromosome 9 (Povey et al. 1994). The human GAL gene was also mapped to the same region (Shaper et al. 1992). Therefore, ABO and GAL genes of many vertebrates may be located on the same chromosomal region, if both of them coexist. In contrast to the close affinity of the chromosomal locations of those two genes, the intron and exon structures are somewhat different from each other. Although substantial proportions of the coding regions reside in the last exons in both genes, the whole coding regions are dispersed over seven and six exons for the ABO gene (Yamamoto, McNeill, and Hakomori 1995) and the GAL gene (Joziasse et al. 1992), respectively, with only the last two exons being homologous.

Evolution of Primate GAL Genes

Partial sequences of several primate GAL genes were reported by Galili and Swanson (1991), and we constructed the phylogenetic trees of the primate GAL genes based on sequence data of 369-bp region. We first made neighbor-joining and maximum-likelihood trees, but New World monkeys did not form a monophyletic cluster in both trees that have the identical topology (not shown). However, the monophyletic clustering of New World monkeys has been clearly demonstrated by using much more molecular data (Schneider et al. 1993). Therefore, a submaximum-likelihood tree with the same branching pattern within the New World monkey species as estimated by Schneider et al. (1993) was constructed, and branch lengths were estimated by applying Ishida et al.'s (1995) method (see fig. 8).

All the Catarrhini (Old World monkeys and hominoids) have GAL pseudogenes, while all the four New World monkeys examined have functional GAL genes (Galili and Swanson 1991). We estimated the average number of nucleotide substitutions (per site) between the common ancestor (node A of fig. 8) and the extant sequences for functional genes (New World monkey lineages) and pseudogenes (Catarrhine lineages) separately. Those values were 0.021 and 0.053 for functional genes and pseudogenes, respectively. If we assume the divergence time of node A to be roughly 40–50 MYA, the rate of nucleotide substitutions for functional GAL

genes and GAL pseudogenes becomes 0.4×10^{-9} to 0.5×10^{-9} and 1.1×10^{-9} to 1.3×10^{-9}, respectively. The evolutionary rate for the primate GAL functional gene is slightly lower than that estimated for mammalian GAL functional genes based on the tree of figure 7, while the rate for the pseudogene lineage is more or less the same as that (1.4×10^{-9}) used in previous sections.

Under the standard framework of molecular evolution, pseudogenes are considered to undergo neutral evolution (Kimura 1983). Galili and Andrews (1995) suggested the following evolutionary scenario on the GAL gene inactivation; an infectious microbial agent manifesting the Gal alpha 1,3 Gal carbohydrate epitope was endemic to the Old World around the speciation of Catarrhini and Platyrrhini (New World monkeys), and individuals without functional GAL genes could survive the endemic due to antibodies against this epitope that existed on the cell wall of that infectious agent. This means that individuals with GAL pseudogenes had higher fitness than those with GAL functional genes. It may be necessary to conduct some experiments to examine this interesting scenario. In this context, it should be noted that type C retroviruses produced in the murine or canine cells manifesting this Gal alpha 1,3 Gal epitope were shown to be neutralized by the human serum which contains antibodies against this epitope (Takeuchi et al. 1996). In any case, it seems that we now have to consider the evolutionary outcome of the emergence of a pseudogene not only from the neutral viewpoint but also from the selective viewpoint.

Perspectives

We have presented various phylogenetic analyses of primate ABO genes and its related genes in this paper. We found some unusual evolutionary patterns for the ABO genes; however, those phylogenetic analyses inevitably remain to show only indirect evidence for the possibility of natural selection. More direct evidence is needed from experimental studies. For example, Borén et al. (1993) found that a gram-negative bacterium *Helicobacter pylori,* a possible causative agent in gastritis and gastric ulcers, binds to the carbohydrate structure Le^b, while it does not bind to ALe^b determinant antigens. Borén et al. thus suggested that the availability of receptors for this bacterium may be reduced in individuals with A and B phenotypes compared to those with O phenotype. More of such experimental studies directly connecting microorganisms and specific blood group types will be definitely necessary to clarify the biological mechanism of the existence of the ABO blood group.

Natural antibodies against A and B determinants are present in individuals who do not express these antigens (Landsteiner's Law). High titers of these antibodies have been attributed to constant stimulation by bacterial flora in intestines, some of which share the epitopes of A and B antigens (Springer, Horton, and Forbes 1959). If natural antibodies are important for fighting against parasites, then individuals with O phenotype may be selectively more advantageous than those with A, B, or AB phenotypes, for O individuals are expected to have both anti-A and anti-B antibodies. If so, the argument of Galili and Andrews (1995) on the GAL pseudogene may also apply to the ABO gene, because the O allele is by definition nonfunctional. In fact, it has been one of the major mysteries of the ABO gene that the nonfunctional O allele is one of the common alleles. The O allele is even fixed in some South American populations (see table 141 of Roychoudhury and Nei 1988). This pattern of allele frequency distribution is clearly out of the scope of the standard mutation-drift balance model, where null (nonfunctional) alleles are expected to remain rare.

It should be remembered that several glycosyltransferases are involved in the production of complex carbohydrate structures to manifest the blood group antigens. When one transferase happens to lose its function, the precursor carbohydrate structure (H determinant in the case of the ABO blood group system) will accumulate without further modification. Clearly, more studies are needed to grasp the whole story of the evolution of the ABO blood group gene.

Acknowledgments

The authors thank Dr. Uri Galili for providing us with various information on the GAL gene, and Dr. Hans-Jürgen Bandelt for consulting on the construction of phylogenetic networks. This study was partially supported by grants-in-aid for scientific studies from Ministry of Education, Science, Sport, and Culture, Japan to N.S. and funds from The Burnham Institute to F.Y.

LITERATURE CITED

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS, and D. J. LIPMAN. 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410.

BANDELT, H.-J. 1994. Phylogenetic networks. Verh. Naturwiss. Ver. Hamburg (NF) **34**:51–71.

BANDELT, H.-J., P. FORSTER, B. C. SYKES, and M. B. RICHARDS. 1995. Mitochondrial portraits of human populations using median networks. Genetics **141**:743–753.

BORÉN, T., P. FALK, K. A. ROTH, G. LARSON, and S. NORMARK. 1993. Attachment of Helicobacter pylori to human gastric epithelium mediated by blood group antigens. Science **262**: 1892–1895.

CHUNG, C. S., and N. E. MORTON. 1961. Selection at the ABO locus. Am. J. Hum. Genet. **13**:9–27.

CROW, J. F. 1993. Felix Bernstein and the first human marker locus. Genetics **133**:4–7.

EASTEAL, S., C. COLLET, and D. BETTY. 1995. The mammalian molecular clock. Springer-Verlag, Heidelberg.

ELLEGREN, H., B. P. CHOWDHARY, M. FREDHOLM, B. HØYHEIM, M. JOHANSSON, P. B. NIELSEN, P. D. THOMSEN, and L. ANDERSSON. 1994. A physically anchored linkage map of pig chromosome 1 uncovers sex- and position-specific recombination rates. Genomics **24**:342–350.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17**:368–376.

———. 1993. PHYLIP: phylogeny inference package. Version 3.5c. University of Washington, Seattle, Wash.

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

GALILI, U., and P. ANDREWS. 1995. Suppression of α-galactosyl epitopes synthesis and production of the natural anti-Gal antibody: a major evolutionary event in ancestral Old World primates. J. Hum. Evol. **29**:433–442.

GALILI, U., and K. SWANSON. 1991. Gene sequences suggest inactivation of α-1,3-galactosyltransferase in catarrhines after the divergence of apes from monkeys. Proc. Natl. Acad. Sci. USA **88**:7401–7404.

GRUNNET, N., R. STEFFENSEN, E. P. BENNETT, and H. CLAUSEN. 1994. Evaluation of histo-blood group ABO genotying in a Danish population: frequency of a novel O allele defined as $O^{(2)}$. Vox Sang. **67**:210–215.

HENION, T. R., B. A. MACHER, F. ANARAKI, and U. GALILI. 1994. Defining the minimal size of catalytically active primate α 1,3 galactosyltransferase: structure-function studies on the recombinant truncated enzyme. Glycobiology **4**:93–201.

HIRAIZUMI, Y. 1990. Selection at the ABO locus in the Japanese population. Jpn. J. Genet. **65**:95–108.

HORAI, S., K. HAYASAKA, R. KONDO, K. TSUGANE, and N. TAKAHATA. 1995. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc. Natl. Acad. Sci. USA **92**:532–536.

HUGHES, A. L., and M. NEI. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature **352**:595–600.

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., T. OYUNSUREN, S. MASHIMA, H. MUKOYAMA, and N. SAITOU. 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.

JOZIASSE, D. H. 1992. Mammalian glycosyltransferases: genomic organization and protein structure. Glycobiology **2**: 271–278.

JOZIASSE, D. H., J. H. SHAPER, D. H. V. DEN EIJINDEN, A. J. V. TUNEN, and N. L. SHAPER. 1989. Bovine α1→3-galactosyltransferase: isolation and characterization of a cDNA clone. J. Biol. Chem. **264**:14290–14297.

JOZIASSE, D. H., J. H. SHAPER, E. W. JABS, and N. L. SHAPER. 1991. Characterization of an α1→3-galactosyltransferase homologue on human chromosome 12 that is organized as a processed pseudogene. J. Biol. Chem. **266**:6991–6998.

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.

———. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

KOMINATO, Y., P. D. MCNEILL, M. YAMAMOTO, M. RUSSELL, S. HAKOMORI, and F. YAMAMOTO. 1992. Animal histo-blood group ABO genes. Biochem. Biophys. Res. Com. **189**:154–164.

LARSEN, R. D., V. P. RAJAN, M. M. RUFF, J. KUKOWSKA-LATALLO, R. D. CUMMINGS, and J. B. LOWE. 1989. Isolation of a cDNA encoding a murine UDPgalactose:β-D-galactosyl-1,4-N-acetyl-D-glucosamine α-1,3-galactosyltransferase: expression cloning by gene transfer. Proc. Natl. Acad. Sci. USA **86**:8227–8231.

LARSEN, R. D., C. A. RIVERA-MARRERO, L. K. ERNST, R. D. CUMMINGS, and J. B. LOWE. 1990. Frameshift and nonsense mutations in a human genomic sequence homologues to a murine UDP-Gal:β-D-Gal(1,4)-D-GlcNAc α(1,3)-galactosyltransferase cDNA. J. Biol. Chem. **265**:7055–7061.

LI, W. H., and L. A. SADLER. 1991. Low nucleotide diversity in man. Genetics **129**:513–523.

MARTINKO, J. M., V. VINCEK, D. KLEIN, and J. KLEIN. 1993. Primate ABO glycosyltransferases: evidence for trans-species evolution. Immunogenetics **37**:274–278.

MOOR-JANKOWSKI, J., A. S. WIENER, and C. R. ROGERS. 1964. Human blood group factors in non-human primates. Nature **202**:663–665.

NEI, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

NEI, M., and T. GOJOBORI. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.

NEI, M., and A. L. HUGHES. 1991. Polymorphism and evolution of the major histocompatibility complex loci in mammals. Pp. 222–247 *in* R. K. SELEANDER, A. G. CLARK, and T. S. WHITTAM, eds. Evolution at the molecular level. Sinauer, Sunderland, Mass.

OGASAWARA, K., M. BANNAI, N. SAITOU et al. (11 co-authors). 1996*a*. Extensive polymorphism of ABO blood group gene: three major lineages for the common ABO phenotypes. Hum. Genet. **97**:777–783.

OGASAWARA, K., R. YABE, M. UCHIKAWA et al. (11 co-authors). 1996*b*. Molecular genetic analysis of variant phenotypes of the p ABO blood group system. Blood **88**:2732–2737.

PEARSON, W. R., and D. J. LIPMAN. 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85**:2444–2448.

POVEY, S., J. ARMOUR, P. FARNDON, J. L. HAINES, M. KNOWLES, F. OLOPADE, A. PILZ, J. A. WHITE, and D. J. KWIATKOWSKI. 1994. Report on the third international workshop on chromosome 9. Ann. Hum. Genet. **58**:177–250.

ROYCHOUDHURY, A. K., and M. NEI. 1988. Human polymorphic genes: a world distribution. Oxford University Press, Oxford.

RUANO, G., J. ROGERS, A. C. FERGUSON-SMITH, and K. K. KIDD. 1992. DNA sequence polymorphism within hominoid species exceeds the number of phylogenetically informative characters for a HOX2 locus. Mol. Biol. Evol. **9**:575–586.

SAITOU, N. 1991. Reconstruction of molecular phylogeny of extant hominoids from DNA sequence data. Am. J. Phys. Anthropol. **84**:75–85.

———. 1996. Reconstruction of gene trees from sequence data. Methods Enzymol. **266**:427–449.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for constructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SAITOU, N., and S. UEDA. 1994. The evolutionary rate of insertions and deletions in noncoding nucleotide sequences of primates. Mol. Biol. Evol. **11**:504–512.

SCHNEIDER, H., M. P. SCHNEIDER, I. SAMPAIO, M. L. HARADA, M. STANHOPE, J. CZELUSNIAK, and M. GOODMAN. 1993. Molecular phylogeny of the New World Monkeys (Platyrrhini, Primates). Mol. Phylogenet. Evol. **2**:225–242.

SHAPER, N. L., S.-P. LIN, D. H. JOZIASSE, D. Y. KIM, and T. L. YANG-FENG. 1992. Assignment of two human α-1,3-galactosyltransferase gene sequences (GGTA1 and GGTA1P) to chromosomes 9q33-q34 and 12q14-q15. Genomics **12**: 613–615.

SPRINGER, G. F., R. E. HORTON, and M. FORBES. 1959. Origin of anti-human blood group B agglutinins in white leghorn chicks. J. Exp. Med. **110**:221–244.

STRAHAN, K. M., F. GU, A. F. PREECE, I. GUSTAVSSON, L. ANDERSSON, and K. GUSTAFSSON. 1995. cDNA sequence and chromosome localization of pig α-1,3 galactosyltransferase. Immunogenetics **41**:101–105.

TAKAHATA, N. 1993*a*. Allelic genealogy and human evolution. Mol. Biol. Evol. **10**:2–22.

———. 1993*b*. Relaxed natural selection in human populations during the Pleistocene. Jpn. J. Hum. Genet. **68**:539–547.

TAKEUCHI, Y., C. D. PORTER, K. M. STRAHAN, A. F. PREECE, K. GUSTAFSSON, F.-L. COSSET, R. A. WEISS, and M. K. L. COLLINS. 1996. Sensitization of cells and retroviruses to human serum by (α1-3) galactosyltransferase. Nature **379**: 85–88.

THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.

YAMAMOTO, F. 1995. Molecular genetics of the ABO histo-blood group A System. Vox Sang. **69**:1–7.

YAMAMOTO, F., H. CLAUSEN, T. WHITE, J. MARKEN, and S. HAKOMORI. 1990*a*. Molecular genetic basis of the histo-blood group ABO system. Nature **345**:229–233.

YAMAMOTO, F., and S. HAKOMORI. 1990. Sugar-nucleotide donor specificity of histo-blood group A and B transferase is based on amino acid substitutions. J. Biol. Chem. **265**: 19257–19262.

YAMAMOTO, F., J. MARKEN, T. TSUJI, T. WHITE, H. CLAUSEN, and S. HAKOMORI. 1990*b*. 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., and P. D. MCNEILL. 1996. Amino acid residue at codon 268 determines both activity and nucleotide-sugar donor substrate specificity of human histo-blood group A and B transferase. J. Biol. Chem. **271**:10515–10520.

YAMAMOTO, F., P. D. MCNEILL, and S. HAKOMORI. 1991. Identification in human genomic DNA of the sequence homologous but not identical to either the histo-blood group ABH genes or α1,3-galactosyltransferase pseudogene. Biochem. Biophys. Res. Com. **175**:986–994.

———. 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. Com. **187**:366–374.

———. 1995. Genomic organization of histo-blood group ABO genes. Glycobiology **5**:51–58.

YAMAMOTO, F., P. D. MCNEILL, Y. KOMINATO, M. YAMAMOTO, S. HAKOMORI, S. ISHIMOTO, S. NISHIDA, M. SHIMA, and Y. FUJIMURA. 1993*a*. Molecular genetic analysis of the ABO blood group system: 2. cis-AB alleles. Vox Sang. **64**:120–123.

YAMAMOTO, F., P. D. MCNEILL, M. YAMAMOTO, S. HAKOMORI, I. M. BROMILOW, and J. K. M. DUGUID. 1993*b*. Molecular genetic analysis of the ABO blood group system: 4. Another type of O allele. Vox Sang. **64**:175–178.

YAMAMOTO, F., P. D. MCNEILL, M. YAMAMOTO, S. HAKOMORI, and T. HARRIS. 1993*c*. Molecular genetic analysis of the ABO blood group system: 3. $A^{(X)}$ and $B^{(A)}$ alleles. Vox Sang. **64**:171–174.

YAMAMOTO, F., P. D. MCNEILL, M. YAMAMOTO, S. HAKOMORI, T. HARRIS, W. J. JUDD, and R. D. DAVENPORT. 1993*d*. Molecular genetic analysis of the ABO blood group system: 1. Weak subgroups: $A^{(3)}$ and $B^{(3)}$ alleles. Vox Sang. **64**:116–119.

YANG, Z., S. KUMAR, and M. NEI. 1995. A new method of inference of ancestral nucleotide and amino acid sequences. Genetics **141**:1641–1650.

NAOYUKI TAKAHATA, reviewing editor

Accepted December 6, 1996