J Mol Evol (1999) 49:615–626

JOURNAL OF MOLECULAR EVOLUTION

Evolution of Rh Blood Group Genes Have Experienced Gene Conversions and Positive Selection

Takashi Kitano, Naruya Saitou

Laboratory of Evolutionary Genetics, National Institute of Genetics, Mishima, 411-8540, Japan

Received: 16 March 1999 / Accepted: 17 June 1999

Abstract. There are two tightly linked loci (D and CE) for the human Rh blood group. Their gene products are membrane proteins having 12 transmembrane domains and form a complex with Rh50 glycoprotein on erythrocytes. We constructed phylogenetic networks of human and nonhuman primate Rh genes, and the network patterns suggested the occurrences of gene conversions. We therefore used a modified site-by-site reconstruction method by using two assumed gene trees and detected 9 or 11 converted regions. After eliminating the effect of gene conversions, we estimated numbers of nonsynonymous and synonymous substitutions for each branch of both trees. Whichever gene tree we selected the branch connecting hominoids and Old World monkeys showed significantly higher nonsynonymous than synonymous substitutions, an indication of positive selection. Many other branches also showed higher nonsynonymous than synonymous substitutions; this suggests that the Rh genes have experienced some kind of positive selection.

Key words: Rh blood group — gene conversion — positive selection

Introduction

Most genes are evolving under the neutral mutation pressure (Kimura 1983). However, some genes are evolving through positive selection. For example, Hughes and Nei (1988) showed that the rate of nonsynonymous substitution was significantly higher than that synonymous substitution in the antigen recognition sites of the major histocompatibility complex class I loci. Ina and Gojobori (1994) showed that positive selection operated on antigenic sites of the hemagglutinin 1 gene of human influenza A virus. Positive selection of these types are considered to be caused by interaction between host defense systems and pathogens.

Blood types were originally distinguished by the different molecular structure on erythrocytes. Therefore, products of blood group genes may cause interactions with other organisms, and there is the possibility of positive selection on those genes as shown for the ABO blood group genes by Saitou and Yamamoto (1997). We analyzed the Rh blood group gene sequences for seeking another possibility of positive selection.

The human Rh blood type is one of the major blood group systems and plays important roles in transfusion and clinical medicine. Landsteiner and Wiener (1940) detected an antibody that agglutinates blood cells from rhesus macaques, and it was named Rh. This antibody had a similar feature with the antibody discovered by Levine and Stetoson (1939) from transfusion incompatibilities. Although the two antibodies had similar specificity, these antibodies were shown to detect distinct antigens. Therefore, the antibody detected by Landsteiner and Wiener was renamed LW. There were historically two hypotheses about the Rh blood group system; Wiener's (1943) one-locus theory and Fisher-Race's) three linked loci (C, D, E) theory (Race 1944). Therefore, two different nomenclatures for loci and alleles have been used.

Rh polypeptides were observed as phosphorylated 30–32 kD membrane proteins by using SDS-PAGE and

Correspondence to: N. Saitou; *e-mail:* nsaitou@genes.nig.ac.jp

immunoprecipitation (Moore et al. 1982; Gahmberg 1982). Nucleotide sequences of Rh genes were determined independently by Cherif-Zahar et al. (1990) and Avent et al. (1990). The Rh blood group system was shown to be composed of two closely linked D and CE loci (Mouro et al. 1993) as predicted by Tippett (1986). In human, D and CE loci are located on chromosome 1p34–p36 (Ruddle et al. 1972; Cherif-Zahar et al. 1991), and individuals are divided into Rh-positive and Rh-negative according to the presence or absence of the D antigen. C/c and E/e specificities are distinguished by four and one amino acid differences, respectively (Mouro et al. 1993). Rh gene products were estimated to have 12 transmembrane domains through hydropathy analysis (Avent et al. 1990) and immunological studies using an antipeptide antibody (Avent et al. 1992).

Nucleotide sequences of Rh-like blood group genes in nonhuman primates were also reported (Salvignol et al. 1994, 1995; Mouro et al. 1994a). Genomic DNA analysis by Southern blot using the human Rh genes as probes have shown that chimpanzee possesses three Rh-like loci (Salvignol et al. 1993, 1994), though only two types of genes for chimpanzee were so far sequenced (Salvignol et al. 1995). Gorillas carry two Rh-like genes, whereas orangutans, gibbons, Old World monkeys, and New World monkeys carry a single Rh-like gene (Blancher et al. 1992).

Salvignol et al. (1995) constructed a neighbor-joining tree of primate Rh blood group genes. If we accept their tree, three gene duplications and four gene losses must be assumed. Blancher and Socha (1997) constructed a maximum likelihood tree of primate Rh blood group genes. Four gene duplications and nine gene losses must be assumed in their tree. Klein et al. (1997) constructed a UPGMA tree of primate Rh blood group genes. In their tree, four gene duplications and five gene losses must be assumed. Because D and CE loci are tightly linked and they are quite similar (96.4% nucleotide sequence identity), gene conversions seem to affect the phylogenetic relationships of those genes. Occurrences of gene conversions and unequal crossing-overs in the Rh loci have been detected (e.g., Cherif-Zahar et al. 1994; Mouro et al. 1994b; Beckers et al. 1996; Huang et al. 1996; Kemp et al. 1996; Carritt et al. 1997), but there has been no detailed analysis on those events. The first objective of this study is thus a detailed analysis of the gene conversion events in hominid Rh blood group genes so as to infer their true phylogenetic relationship. The second objective is the examination of existence of positive selection on the Rh blood group gene. This analysis is possible only after eliminating the effect of gene conversion.

Materials and Methods

Nucleotide sequence data for Rh blood group genes were retrieved from the DDBJ/EMBL/GenBank international nucleotide sequence database and from published papers. Ten human (*Homo sapiens*) sequences, five chimpanzee (*Pan troglodytes*) sequences, three gorilla (*Gorilla gorilla*) sequences, two crab-eating macaque (*Macaca fascicularis*) sequences, and one rhesus macaque (*Macaca mulatta*) sequence were used in this study (Table 1). All the sequences were complete cDNA with 1,251 bp.

Table 1. Rh blood group genes and references used in this study

Genes (accession number)	References
Human D-1 (X63097)	Le Van Kim et al. (1992)
Human D-2 (X63094)	Le Van Kim et al. (1992)
Human D-3 (S57971)	Kajii et al. (1993)
Human D-4 (L08429)	Arce et al. (1993)
Human D-5 (S78509)	Huang et al. (1995)
Human D-6	Huang et al. (1996)
Human cE-1 (M34015)[a]	Cherif-Zahar et al. (1990)
Human cE-1 (X54534)[a]	Avent et al. (1990)
Human cE-2 (S577967)	Kajii et al. (1993)
Human Ce	Huang et al. (1996)
Chimpanzee 1-1 [317-IIR] (L37050)	Salvignol et al. (1995)
Chimpanzee 1-2 [394-2G]	Salvignol et al. (1995)
Chimpanzee 1-3 [211-6E]	Salvignol et al. (1995)
Chimpanzee 1-4 [317-IA] (L37049)	Salvignol et al. (1995)
Chimpanzee 2 [211-IIF] (L37048)	Salvignol et al. (1995)
Gorilla 1-1 [IC] (L37052)	Salvignol et al. (1995)
Gorilla 1-2 [IIA2b]	Salvignol et al. (1995)
Gorilla 2 [ID] (L37053)	Salvignol et al. (1995)
Crab-eating macaque 1 (L37054)	Salvignol et al. (1995)
Crab-eating macaque 2	Salvignol et al. (1995)
Rhesus macaque (S70343)	Mouro et al. (1994a)

[a] Nucleotide sequences of M34015 and X54534 (both human cE-1) are identical.

Gene names given in square brackets are those used by Salvignol et al. (1995).

CLUSTAL W version 1.6 (Thompson et al. 1994) was used for multiple alignment. Phylogenetic networks were constructed following the procedure of Bandelt (1994) and Saitou and Yamamoto (1997). To identify gene conversion events, the site-by-site reconstruction method (Slightom et al. 1987) was used. The original method was based on the parsimony method; however, we used the maximum likelihood method to determine nucleotide sequences of ancestral nodes. PAML version 1.3 (Yang 1997) was used for the maximum likelihood analysis. We used Ina's (1995) method to estimate numbers of synonymous and nonsynonymous substitutions, and statistical tests were performed following Zhang et al. (1997).

Results and Discussion

Gene Conversions Confuse the Rh Gene Tree

Because the two Rh loci are tightly linked, it is possible that they have experienced gene conversions or crossing-overs. Those events can confuse the phylogenetic relationship of the linked loci. We thus constructed phylogenetic networks for human, chimpanzee, and gorilla Rh blood group genes. Because the network for the whole sequence data has many dimensions, we constructed networks for five regions of the coding region of this gene. For simplicity, we selected one allele of each locus by examining phylogenetic networks of human and

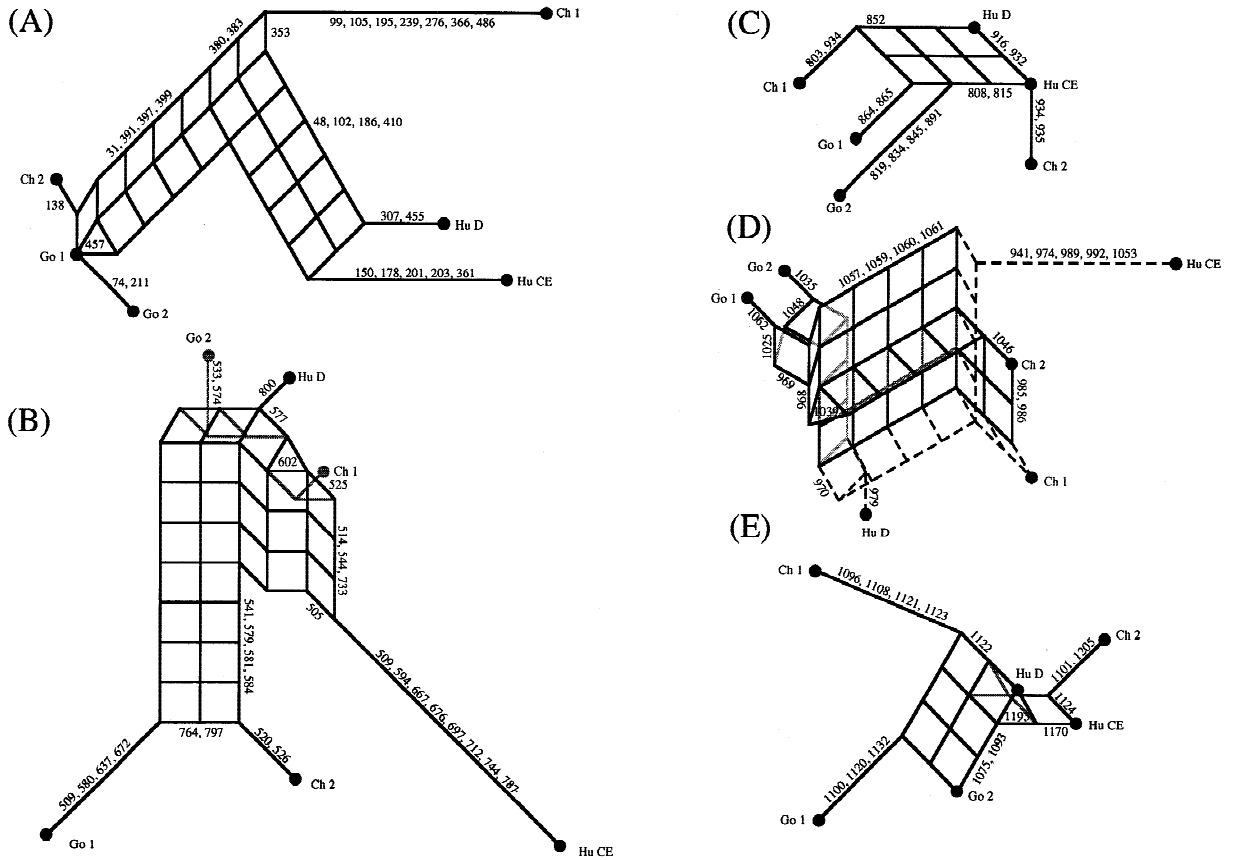


Fig. 1. Phylogenetic networks for hominoid Rh blood group gene exons 1, 2, and 3 **(A);** exons 4 and 5 **(B);** exon 6 **(C);** exon 7 **(D);** and exons 8, 9, and 10 **(E).** Numbers are nucleotide positions responsible for corresponding edges, and edge lengths are proportional to number of nucleotide differences. Full circles denote observed sequences. Hu, Ch, and Go mean human, chimpanzee, and gorilla, respectively.

chimpanzee Rh blood group genes (figures not shown). We thus used human D (a consensus sequence of human D alleles), human CE (human cE-1 in Table 1), chimpanzee 1 (chimpanzee 1–3 in Table 1), chimpanzee 2, gorilla 1 (gorilla 1-1 in Table 1), and gorilla 2 genes for the following analyses. Nucleotide identity between human D and human CE, that between chimpanzee 1 and chimpanzee 2, and that between gorilla 1 and gorilla 2 were 96.4%, 95.9%, and 97.0%, respectively.

Figure 1 shows phylogenetic networks for exons 1–3 (A), exons 4–5 (B), exon 6 (C), exon 7 (D), and exons 8–10 (E). All the networks contain parallelograms that suggest parallel substitutions or some kind of convergent changes. If the relationship of genes is not affected by gene conversion and/or crossing-over, the network may not contain so many parallelograms. Moreover, some sites of those parallelograms in Fig. 1 are contiguous (e.g., 391, 397, and 399 of Fig. 1A), suggesting the existence of conversion-like events.

Because chimpanzee possesses three Rh-like loci, at least two gene duplications occurred in the hominoid lineage. There are three possible gene trees under this situation (Figs. 2A–2C). Figure 2A assumes that the first gene duplication occurred in the common ancestor of human, chimpanzee, and gorilla, and the second gene duplication occurred on one duplicated gene of the chimpanzee lineage. Figure 2B assumes that one gene duplication occurred in the common ancestor of human, chimpanzee, and gorilla, and the other gene duplication occurred in one duplicated gene of the common ancestor of human and chimpanzee. Figure 2C assumes that two gene duplications occurred in the common ancestor of human, chimpanzee, and gorilla. The established phylogeny (human and chimpanzee are clustered first) for the three species (e.g., Horai et al. 1995) is adopted for each orthologous gene group.

To identify orthologous genes, we first classified sites based on phylogenetic networks (Table 2). For example, sites 380 and 383 of network A (Fig. 1) divide the human D–chimpanzee 1 pair from the remaining genes. Because two genes from the same species can no longer form a closest cluster in either trees, sites indicating those clusters are not shown.

If we assume tree 2A, two genes of chimpanzee (α-1 and α-2) are probably quite similar to each other, and it is not easy to identify them. Because the nucleotide identity between chimpanzee 1 and chimpanzee 2 sequences is similar to that between human D and human CE and that between gorilla 1 and gorilla 2, we assume that these two genes probably correspond to chimpanzee α-1 (or

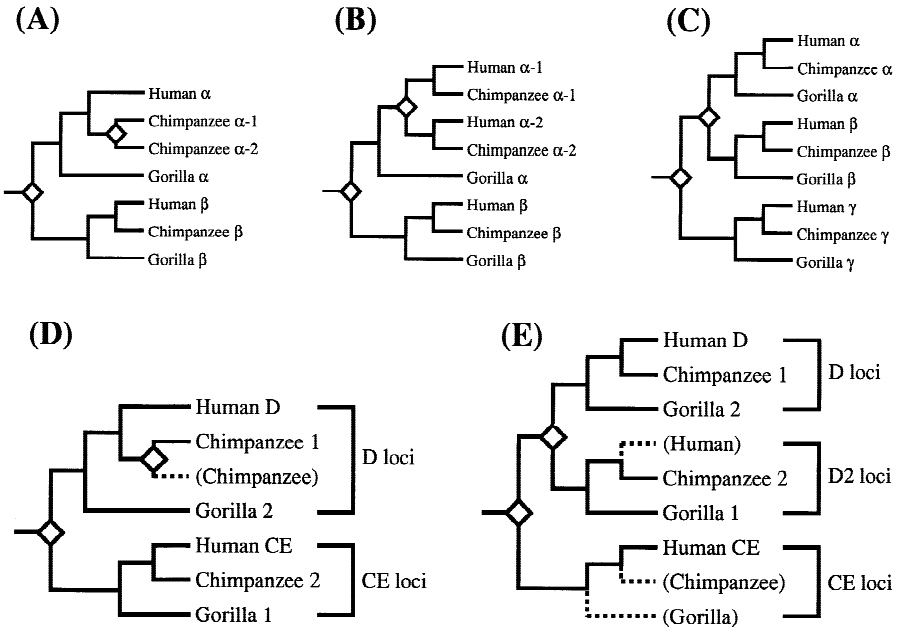


Fig. 2. Three model trees **(A–C)** and two assumed trees **(D** and **E).** We must assume that at least two gene duplications (represented by diamonds) occurred in the hominoid lineage because chimpanzee possesses three Rh-like loci (trees **A–C**). Names of genes in these model trees are arbitrary. Two assumed trees **(D** and **E)** were used in the following analyses. Names of genes in these trees are actual ones. Gene names in parentheses indicate undetected or deleted ones. The cluster including chimpanzee 2 and gorilla 1 is named the D2 loci cluster in tree E.

Table 2. Classification of sites to identify orthologous genes from phylogenetic networks

Pairs[a]	No. of sites	Sites	Network[b]
(Hu D, Ch 1, Go 2)–(Hu CE, Ch 2, Go 1)	3	514, 544, 733	B
(Hu D, Ch 2, Go 2)–(Hu CE, Ch 1, Go 2)	1	577	B
(Hu D, Ch 1)–(others)	6	380, 383	A
		916, 932	C
		985, 986	D
(Ch 2, Go 1)–(others)	4	541, 579, 581, 584	B
(Ch 1, Go 1)–(others)	2	852	C
		1122	E
(Hu D, Go 2)–(others)	1	1048	D
(Hu CE, Ch 1)–(others)	1	505	B
(Hu CE, Ch 2)–(others)	1	1170	E
(Hu CE, Go 1)–(others)	1	1025	D

[a] Hu: human, Ch: chimpanzee, Go: gorilla.
[b] See Fig. 1 for networks A–E.

α-2) and β, not chimpanzee α-1 and α-2. In this case we can extract orthologous trios from phylogenetic networks. Because sites 514, 544, and 733 show clusters of human D–chimpanzee 1–gorilla 2 and human CE–chimpanzee 2–gorilla 1, and the other 13 sites (380, 383, 916, 932, 985, 986, 541, 579, 581, 584, 1,048, 1,170, and 1,025) are compatible with these clusters, we determined the topology of hominoid Rh blood group genes as tree D in Fig. 2.

When we consider either trees 2B or 2C, we have to assume two and three gene losses (or genes not yet identified), respectively. Because we don't know which genes are lost (or genes not yet identified), examination under trees 2B or 2C is more difficult than that under tree 2A. Because the cluster for human D and chimpanzee 1 is supported with six sites (380, 383, 916, 932, 985, and 986), and the cluster for chimpanzee 2 and gorilla 1 is supported with four sites (541, 579, 581, and 584), these clusters are plausible. Moreover, three sites (514, 544, and 733) show clusters of human D–chimpanzee 1–gorilla 2 and human CE–chimpanzee 2–gorilla 1. Therefore we can choose the unrooted topology as follows: ([(human D, chimpanzee 1), gorilla 2], [human cE, (chimpanzee 2, gorilla 1)]). To determine the root, we eliminated sites indicating two genes from the same species form a cluster from the multiply aligned sequence data, and constructed a neighbor-joining tree (Saitou and Nei 1987) using CLUSTAL W of Thompson et al. (1994). Three sequences of Old World monkeys were used as outgroups. The topology of the NJ tree was compatible with tree E of Fig. 2. The maximum likelihood analysis using NucML of Adachi and Hasegawa (1994) also supported this topology. It is interesting to note that the trees D and E both showed the same cluster of human D–chimpanzee 1–gorilla 2. Only the position of human CE is different between the two trees. Because four sites (541, 579, 581, and 584) in the network B of Fig. 1 supports clusters of chimpanzee 2 and gorilla 1, tree E seems to be more plausible than tree D. Recently, Apoil and Blancher (personal communication) studied the evolutionary relationship of the primate Rh genes using the intron 4 sequences. They suggested that two gene duplications occurred in the common ancestor of human, chimpanzee, and gorilla. This scenario is compatible with our tree E of Fig. 2. However, tree E requires three gene losses (or genes not yet identified) compared to one gene loss for tree D, and it is more parsimonious. We thus used those two assumed trees for the following analyses.

We used the site-by-site reconstruction method to identify regions of gene conversion events assuming the

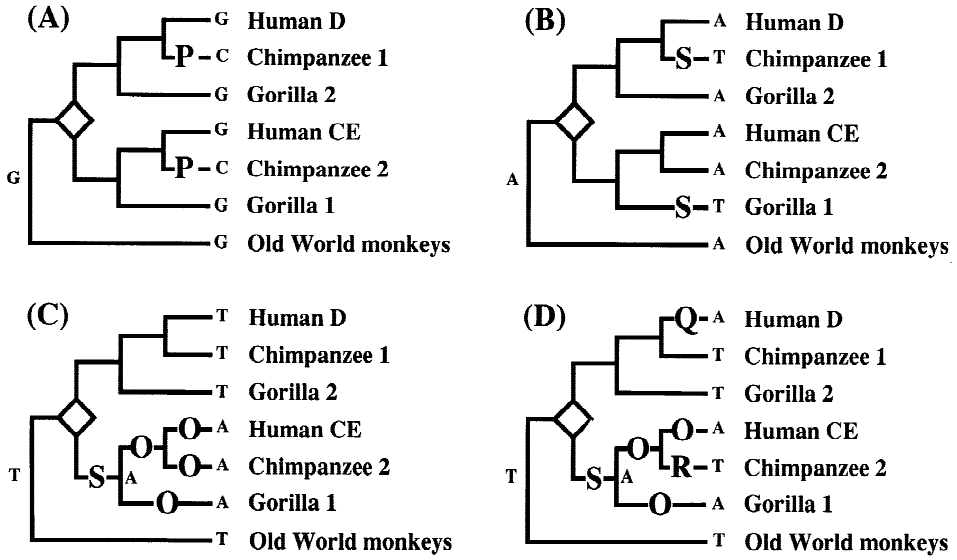


Fig. 3. Explanation of the site-by-site reconstruction method. To classify variant sites, we used five symbols. Tree D of Fig. 2 was assumed. In case **A,** the root nucleotide is G, thus variants are chimpanzee 1 and 2 sequences. These variants are caused by shared nucleotide changes (form G to C) between the duplicated genes in a single species, suggesting a gene conversion in chimpanzee over this site, and they are presented by P. In case **B,** the root nucleotide is A and changes from A to T occurred in the chimpanzee 1 and the gorilla 1 lineages. These nucleotide changes are not shared by the duplicated genes of the same species, thus, they are true parallel substitutions, and designated by S. In case **C,** the root nucleotide is T and a change (from T to A) occurred in the common ancestor of orthologous human CE, chimpanzee 2, and gorilla 1 sequences. This change is not shared with its paralogous counterparts, and is indicated by S. There are no changes in its descendants, suggesting no gene conversion after the hominoid divergence over this site. Therefore, we give O to those branches. Case **D** is a kind of sequel to case C, and additional nucleotide changes occurred in the human D and chimpanzee 2 genes. These additional changes resulted in the same nucleotide with their paralogous counterparts, and it suggests that directions of gene conversions were from the human CE to the human D genes and from chimpanzee 1 to chimpanzee 2. Therefore, Q is given to the human D gene lineage and R is given to the chimpanzee 2 gene one.

two trees (Figs. 2D and 2E). First, we obtained the maximum likelihood estimates of the nucleotide sequences of ancestral nodes by using PAML program (Yang 1997), and then substitution patterns were plotted on each branch of the assumed tree. Let us explain the actual procedure of the site-by-site reconstruction method using Fig. 3, where tree D of Fig. 2 was assumed. The two chimpanzee loci both had the same nucleotide (C) in case A, and the event causing this change is indicated by the letter P. They look like parallel substitutions, but this pattern could also be produced by a gene conversion after a substitution. In case B, two independent substitutions designated by S occurred in different species. In case C, a branch of one cluster experienced a substitution (indicated by the letter S) after the gene duplication, and no change in its descendants is indicated with O. Case D is a next step of case C. Substitutions occurred in the descendants. These additional changes result in the same nucleotide with their paralogue, suggesting a gene conversion from human CE to human D and a gene conversion from chimpanzee 1 to chimpanzee 2 over this site. Therefore, letter Q is given to human D and the letter R is given to chimpanzee 2. If cases A or D are contiguous for two or more variant sites, a gene conversion event over the region containing those sites is inferred based on the maximum parsimony principle. In case D, we can infer the direction of gene conversion. We also performed the same procedure under the assumption of tree E of Fig. 2.

Let us compare rates of gene conversion with nucleotide substitutions to see if our parsimonious argument is valid. Hogstrand and Bohme (1998) estimated cis gene conversion frequency between *Abk* and *Ebk*, and that between *Abd* and *Ebd* in mouse MHC class II genes, and obtained values 1.2×10^{-6}–9.7×10^{-5} per locus (about 200 bp region) per generation. Rates of synonymous substitutions in various mammalian protein-coding genes is estimated to be 3.5×10^{-9} per site per year (Li 1997). If we assume that the generation time of wild mice is half a year (Tsuyoshi Koide, personal communication), the rate of cis gene conversion frequency becomes 2.4×10^{-6}–1.9×10^{-4} per 200 nucleotide sites per year, and the corresponding rate of synonymous substitution is 7.0×10^{-7} per 200 nucleotide sites per year. It suggests that the rate of gene conversion seems to be one to three orders higher than that of nucleotide substitutions. Of course, the rate of gene conversion may depend on the copy number of corresponding gene loci in a genome and sequence similarities among corresponding genes. Furthermore, we cannot directly compare rates between nucleotide substitutions and gene conversions. However, at least we can say that the rate of gene conversion seems to be much higher than that of nucleotide substitutions. This justifies the site-by-site reconstruction procedure.

Table 3. Patterns of site changes for each branch of tree D estimated from the maximum liklihood analysis

ML	Site	31	48	74	86	97	99	102	105	138
		<=						=>		
	MPU[a]	1	1	1	1	1	1	1	1	1
1st ML	H–D	O	P					P		
	H–CE	Q	P					P		
	C–1	O					S		S	
	C–2									S
	G–2			S						
	G–1									
	HC–D	S								
	HC–CE									
	HCG–D									
	HCG–CE									
	OH		S		S	S				S
2nd ML	H–D									
	H–CE									
	C–1									
	C–2									
	G–2									
	G–1									
	HC–D									
	HC–CE									
	HCG–D									
	HCG–CE									
	OH									
3rd ML	H–D									
	H–CE									
	C–1									
	C–2									
	G–2									
	G–1									
	HC–D									
	HC–CE									
	HCG–D									
	HCG–CE									
	OH									

ML	Site	150	151	176	178	186	193	195	201	203	207	211	214	239	247	276	304	307	326
	MPU[a]	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1st ML	H–D					P												S	
	H–CE	S			S	P			S	S									
	C–1							S						S		S			
	C–2											S							
	G–2																		
	G–1																		
	HC–D																		
	HC–CE																		
	HCG–D																		
	HCG–CE																		
	OH	S	S	S	S	S	S			S	S		S		S		S		S
2nd ML	H–D																		
	H–CE																		
	C–1																		
	C–2																		
	G–2																		
	G–1																		
	HC–D																		
	HC–CE																		
	HCG–D																		
	HCG–CE																		
	OH																		
3rd ML	H–D																		
	H–CE																		
	C–1																		
	C–2																		
	G–2																		
	G–1																		
	HC–D																		
	HC–CE																		
	HCG–D																		
	HCG–CE																		
	OH																		

ML	Site	353	357	358	361	366	367	379	380	382	383	391	397	399	410	428	439	446	447	450	455	457	465	473	486
									<=				=>												
												<=										=>			
	MPU[a]	1	1	1	1	1	1	1	2	1	2	1	3	3	1	1	1	1	1	1	1	1	1	1	1
1st ML	H–D								O		O	O	O	O	P						S	P			
	H–CE				S							Q	Q	Q	P							P			
	C–1	P				S			O		O	O	O	O								S			S
	C–2	P																							
	G–2																								
	G–1																								
	HC–D								S		S	S	S	S											
	HC–CE																								
	HCG–D																								
	HCG–CE																								
	OH	S	S	S	S		S	S	S	S	S		S	S		S	S	S	S	S			S	S	S
2nd ML	H–D																					P			
	H–CE								O		O											P			
	C–1																					O			
	C–2								O		O		S	S											
	G–2								Q		Q		P	P											
	G–1								O		O		P	P											
	HC–D																					S			
	HC–CE								O		O														
	HCG–D																								
	HCG–CE								S		S														
	OH																								
3rd ML	H–D																					O			
	H–CE												R	R								Q			
	C–1																					S			
	C–2												O	O											
	G–2												Q	Q											
	G–1												O	O											
	HC–D																					S			
	HC–CE												O	O											
	HCG–D																								
	HCG–CE												S	S											
	OH																								

ML	Site	488	496	497	505	509	514	517	520	525	526	533	541	544	574	577	579	580	581	584	585	594	595	602	604	611	612	623
																	<=			=>								
	MPU[a]	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	2	1	1	1	1	1	1	1	1
1st ML	H–D													O		S												
	H–CE				S	S	O															S		S				
	C–1				S					S				O										Q				
	C–2						O		S		S		S			S	S		S	S				O				
	G–2											S		O	S													
	G–1					S	O						S			S	S	S	S	S				O				
	HC–D													O														
	HC–CE						O																	O				
	HCG–D													S														
	HCG–CE						S																	S				
	OH	S	S	S				S													S		S		S	S	S	S
2nd ML	H–D															S												
	H–CE												R			S	R		R	R								
	C–1																											
	C–2												O			O	O		O	O								
	G–2																											
	G–1												O			O	O		O	O								
	HC–D																											
	HC–CE												O			O	O		O	O								
	HCG–D																											
	HCG–CE												S			S	S		S	S								
	OH																											
3rd ML	H–D																											
	H–CE																											
	C–1																											
	C–2																											
	G–2																											
	G–1																											
	HC–D																											
	HC–CE																											
	HCG–D																											
	HCG–CE																											
	OH																											

ML	Site	637	665	667	672	676	685	686	689	697	712	719	733
	MPU[a]	1	1	1	1	1	1	1	1	1	1	1	1
1st ML	H–D												O
	H–CE			S		S				S	S		
	C–1												O
	C–2												
	G–2												O
	G–1	S			S								
	HC–D												O
	HC–CE												
	HCG–D												S
	HCG–CE												
	OH	S	S				S	S	S			S	
2nd ML	H–D	O											
	H–CE	O											
	C–1	O											
	C–2	O											
	G–2	O											
	G–1												
	HC–D	O											
	HC–CE	S											
	HCG–D	S											
	HCG–CE												
	OH												
3rd ML	H–D												
	H–CE												
	C–1												
	C–2												
	G–2												
	G–1												
	HC–D												
	HC–CE												
	HCG–D												
	HCG–CE												
	OH												

Table 3 shows estimated substitution patterns of all the variable sites under tree D of Fig. 2 by using the site-by-site reconstruction method. Ancestral sequences were estimated by using the maximum likelihood method. According to likelihood values, patterns are arranged from the top to the bottom. We inferred regions of gene conversion events from these results. For example, sites 31–102 are inferred to have experienced a gene conversion because letters P and Q were contiguous. Human D and chimpanzee 1 sequences are identical on site 31 (indicated by O) and human CE is indicated by Q, thus, the direction of this gene conversion is inferred from the human D gene to the human CE gene.

Two possible patterns were inferred for sites 579–584. One pattern was that three parallel substitutions occurred in chimpanzee 2 and gorilla 1 genes on sites 579, 581, and 584, and one substitution occurred in the gorilla 1 gene on site 580. Another pattern involved three substitutions in the ancestral CE gene of human, chimpanzee, and gorilla, followed by three backward substitutions in human CE on sites 579, 581, and 584, and one substitution in gorilla 1 gene on site 580. Either patterns require seven substitution events. If we assume a gene conversion on this region on the latter case, however, only five events are necessary: three substitutions in the ancestral CE gene of human, chimpanzee, and gorilla; one gene conversion in human CE; and one substitution on the gorilla 1 gene on site 580. Therefore, we inferred that a gene conversion occurred on this region.

If the total numbers of events (substitution and gene conversion) is decreased by taking into account the gene conversions, we selected that pattern in spite of lower likelihood estimation of ancestral nodes as shown in the above example (Table 3). All other gene conversion regions were inferred in the same fashion. Table 4 shows substitution patterns of each variable sites under tree E of Fig. 2 by using the same procedure as in the case for Table 3.

Table 3. Continued

		Site	MPU[a]	Human	Chimpanzee	Gorilla	Common ancestor of human and chimpanzee	Common ancestor of HC and gorilla	Branch between hominoids and Old World monkeys
1st	H–D	7777778 4456890 4604770	1111111	O OS	QQ OO	O O S OO O S	P O		
	H–CE			S OSO	OO	S Q S SS SOOOO	P O O		
	C–1			O O	S S OOS	O O OO O P P QQQQ	S S SSS O		
	C–2			O O	OO SS	Q Q P P OOOO	S O OS		
	G–2				S S S S	P OS S	P P S P		
	G–1				S SS	P S S	P P S S S S		
	HC–D			P P	S S	S S O	S		
ML	HC–CE			P P	S S	S S S S	S S		
	HCG–D					S			
	HCG–CE								
	OH			SS	S S S SS S SSS	SS S SSSS SSSS S S SSS S S SSSS	SS S SS S SSSS	SSS SS SSS	
2nd	H–D	7777778 4456890 4604770	1111111		O Q	O O S S	O Q O	O O	
	H–CE				O O	Q	O O Q		
	C–1				O	P R	O R	O O	
	C–2				O O	P	O O		
	G–2				S	O O P P	O S O	O P	
	G–1				O	P P		Q P	
	HC–D				Q	O O	O O	O S	
ML	HC–CE				O S		Q S		
	HCG–D					S S	S S	S	
	HCG–CE				S				
	OH								
3rd	H–D	7777778 4456890 4604770	1111111		O	O OO	O	O	
	H–CE				O	Q	O		
	C–1				R	R R	S	O	
	C–2				O	O	O		
	G–2				O	O OO	O	O	
	G–1					QQ		Q	
	HC–D				O	O OO	O	S	
ML	HC–CE				Q	S	Q		
	HCG–D				S	S S	S	S	
	HCG–CE								
	OH								

Vertical lines divide each exon. Accoding to likelihood values, patterns are arranged from top to bottom.

[a] MPU means ML patterns used. Five patterns (O, P, Q, R, and S) listed in this table are explained in Fig. 2 and text.

Equal signs surrounded with angled brackets designate regions of gene conversions.

H: human, C: chimpanzee, G: gorilla, HC: common ancestor of human and chimpanzee, HCG: common ancestor of HC and gorilla, OH: the branch between hominoids and Old World monkeys.

Table 5 shows the inferred gene conversion events in the hominoid phylogeny. Eleven and nine events were estimated for trees D and E of Fig. 2, respectively. The direction of gene conversion was not identified in some cases. Because we cannot detect break points of gene conversions and analyzed only cDNA sequences, these ranges of gene conversions are minimum estimates.

The Phylogenetic Tree and Synonymous/ Nonsynonymous Substitutions

We reconstructed the two sets of nonconverted sequences from the results shown in Tables 3 and 4. In the case of gene conversion 1 for tree D (see Table 5), for example, nucleotides of sites 31–102 of the human CE gene were substituted to those of the CE gene of the common ancestor of human and chimpanzee because of the direction of gene conversion (D $\rightarrow$ CE). As directions of gene conversions five and nine for trees D and E and gene conversion six for tree E were not determined, the sequence data of those sites were not used in the later analysis.

Figure 4 shows two phylogenetic trees (A and B) for the reconstructed primate Rh blood group gene sequences assuming tree D (Fig. 4A) and tree E (Fig. 4B), respectively. All the branch lengths were estimated from sequence data of extant nodes and ancestral nodes estimated above, and the midpoint rooting was used. Num-

Table 4. Patterns of site changes for each branch of tree E estimated from the maximum liklihood analysis

		Site	1 11	11111112222222233 3	33333333 33 3 3 3 44444444 444	444555 555555 5 55 5 55 5 5556 6666	66666666677 77
		t	3 4 7 8 9 9 0 03	55778 9900011347002	55566678 88 9 9 9 12344555 678	899001 122234 4 77 7 88 8 8990 0112	3667788891134 4
		e	1 8 4 6 7 9 2 58	01686 3513714976476	37816790 23 1 7 9 089670 57 536	867594 705631 4 47 9 01 4 5452 4123	7572656972934 6
			<═══>		<═══════>		
					< >		
		MPU	1 1 1 1 1 1 1 11	11111111111111111	11111111 11 1 1 1 11111111 111	111111 111111 1 11 1 11 1 1111 1111	11111111111 11
		H–D	OR P	R S	O OO P SP	O O S	O
		H–CE	Q P	S	Q P P	SS S S	S S SS S
1		C–1	O S S	S S S	P S O OO S S	S O S O Q	O
s		C–2	S		P OO	S S O OO OO O	
t		G–2	S	S	QQ	O S OS	O
		G–1			OO	S O OOSOO O	S S
		HC–D	S		S SS	O O	O
M		HCG–D				S S	S
L		CG–D2			S S	S S SS SS S	
		D–D2	S	S SS S	S		
		OH	S S S	SS S S S S S S S	SS SSSS S SSSSS SSS	SSS S S S S S S S	S SSS S S
		H–D			OO P		
		H–CE			P	S	
2		C–1			OO O		
n		C–2				O	
d		G–2					
		G–1				O	
		HC–D			S S S		
M		HCG–D					
L		CG–D2				S	
		D–D2			S S		
		OH					
		H–D			O		
		H–CE			Q		
3		C–1			S		
r		C–2					
d		G–2					
		G–1					
		HC–D			S		
M		HCG–D					
L		CG–D2					
		D–D2					
		OH					

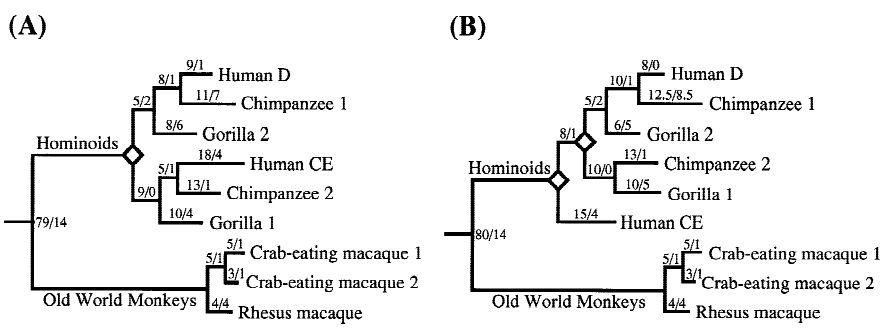
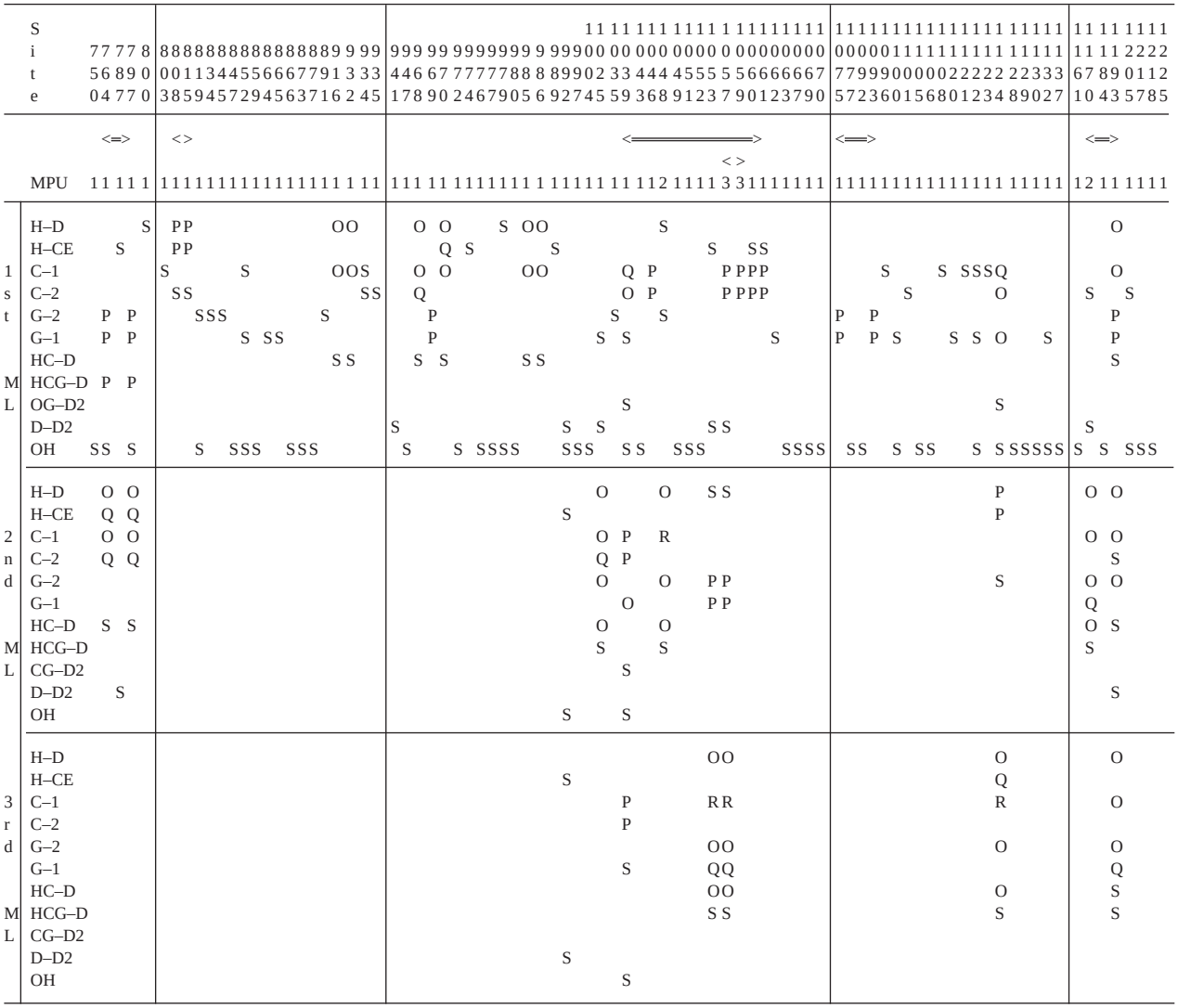


Fig. 4. The two possible gene trees **(A** and **B)** for the primate Rh blood group genes after eliminating the effect of gene conversions, assuming tree D and tree E of Fig. 2, respectively. Diamonds designate gene duplications. Numbers of nonsynonymous and synonymous substitutions per sequence per branch are presented above each branch before and after slash, respectively.

bers of synonymous and nonsynonymous substitutions were estimated by Ina's (1995) method. In this method, the proportion (R) of transition/transversion of the third codon is estimated from proportions of base changes that are observed in the entire phylogenetic tree. R values for trees D and tree E were estimated to be 1.79 and 1.87, respectively. Numbers of nonsynonyomous substitutions are higher than those of synonymous substitutions on almost all branches of both trees (data not shown). Application of Fisher's exact test (Zhang et al. 1997) showed that the nonsynonymous substitution was significantly higher than synonymous ones ($p = 0.0041$) in the branch that connected hominoids and Old World monkeys for tree A of Fig. 4, where 79 nonsynonymous and 14 synonymous substitutions were estimated. We also considered an alternative, less likely substitution patterns

Table 4. Continued



Designations are the same as those in Table 3.
CG-D2: common ancestor of chimpanzee 2 and gorilla 1 genes, D-D2: common ancestor of HCG-D and CG-D2.

Table 5. Gene conversion events occurred in the hominoid phylogeny

			Tree D[b]		Tree E[b]	
ID	Branch[a]	Exon	Sites	Direction	Sites	Direction
1	HC-Human	Exon 1	31–102	D ➡ CE	48–102	CE ➡ D
2	HCG-Gorilla	Exon 3	380–399	CE ➡ D	397–399	D2 ➡ D
3	HC-Human	Exon 3	391–457	D ➡ CE	391–457	D ➡ CE
4	HC-Human	Exon 4	579–584	D ➡ CE	—	—
5	HCG-HC	Exon 5	764–797	Undetermined	764–797	Undetermined
6	HC-Human	Exon 6	808–852	CE ➡ D	808–815	Undetermined
7	HC-Chimpanzee	Exon 7	1,039–1,061	CE ➡ D	1,039–1,061	D2 ➡ D
8	HCG-Gorilla	Exon 7	1,057–1,059	D ➡ CE	1,057–1,059	D ➡ D2
9	HCG-Gorilla	Exon 8	1,075–1,093	Undetermined	1,075–1,093	Undetermined
10	HC-Human	Exon 8	1,122–1,124	CE ➡ D	—	—
11	HCG-Gorilla	Exon 9	1,170–1,193	D ➡ CE	1,170–1,193	D ➡ D2

[a] HC: common ancestor of human and chimpanzee, HCG: common ancestor of HC and gorilla.
[b] See Fig. 2.

indicated in Table 3 for this branch. Numbers of nonsynonyomous and synonymous substitutions for the branch became 76 and 14, respectively, but the difference was still highly significant ($p = 0.0067$). The same branch that connects hominoids and Old World monkeys was also significant ($p = 0.0031$) for tree B of Fig. 4, where 80 nonsynonymous and 14 synonymous substitutions were estimated. We also considered an alternative, less likely substitution patterns indicated in Table 4. Numbers of nonsynonyomous and synonymous substitutions for the branch became 73 and 14, respectively, and the difference was still statistically significant ($p = 0.0100$). In any case, whichever gene tree we selected, we found the possibility of positive selection in the branch that connects hominoid and Old World monkey clusters.

We also estimated the total numbers of nonsynonyomous and synonymous substitutions for hominoid branches to examine the overall evolutionary pattern within hominoids. Nonsynonymous substitutions were not significantly higher than synonymous ones in hominoids for tree A ($p = 0.1055$), where 96 nonsynonymous and 27 synonymous substitutions were estimated, nor for tree B ($p = 0.0889$), where 97.5 nonsynonymous and 27.5 synonymous substitutions were estimated. Because these values were estimated from nonconverted sequence data, however, we also considered the situation that all observed sequence data were results of parallel substitutions. The number of nonsynonymous substitutions (132) now became significantly higher than that of synonymous ones (30) ($p = 0.0046$) within hominoids for tree A. The nonsynonymous substitutions for tree B (124.5) was also much higher than synonymous ones (30.5) in hominoids ($p = 0.0117$). Therefore, estimates of nonsynonymous substitutions after eliminating effects of gene conversion are probably underestimation.

We also estimated the average rates of synonymous and nonsynonymous substitutions for the Rh blood group gene, under the assumption of constancy of the evolutionary rate and the divergence between the Old World monkey and hominoid lineages to be 23 million years ago (MYA) (Kumar and Hedges 1998). Average numbers of synonymous and nonsynonymous substitutions per site between Rh blood group genes of Old World monkeys and those of hominoids were estimated to be 0.068 and 0.124, respectively, for tree A of Fig. 4, applying Ishida et al.'s (1995) method. Average numbers of synonymous and nonsynonymous substitutions per site between Rh blood group genes of Old World monkeys and those of hominoids for tree B of Fig. 4 were almost the same (0.067 and 0.125 for synonymous and nonsynonymous substitutions, respectively) as those of tree A of Fig. 4. Therefore, the rates of synonymous and nonsynonymous substitutions (per site per year) for the Rh blood group gene of Old World monkeys and hominoids are estimated to be $1.46–1.48 \times 10^{-9}$ ($0.067–0.068$ / [2×23 MYA]) and $2.70–2.72 \times 10^{-9}$ ($0.124–0.125$ / [2×23 MYA]), respectively. The evolutionary rate of synonymous substitution for the Rh gene is somewhat lower than that for other primate genes (2.3×10^{-9}; Li and Tanimura 1987). In any case, it is clear that the nonsynonymous substitution is on average higher than the synonymous one in primate Rh genes.

Because the Rh blood group gene products are membrane proteins, there is a possibility that these products of blood group genes are affected by interactions with other organisms or cells on surface regions. Endo et al. (1996) searched the nucleotide sequence database and found that 17 gene groups were the candidates for the genes on which positive selection may operate. Nine of those 17 groups were surface antigens of parasites or viruses. Eder and Spitalnik (1997) suggested that blood group antigens such as ABO, MN, and Lewis play a key role in pathogenesis of some diseases. The high rate of nonsynonymous substitutions for the primate Rh blood group also suggests the existence of positive selection on this gene, and this might be caused by some kind of interaction with pathogens.

Kitano et al. (1998) showed the long-term evolutionary tree of Rh blood group genes and their homologous Rh50 genes, and concluded that Rh blood group genes had a higher evolutionary rate than Rh50 genes. Similar results were obtained independently by Matassi et al. (1999). Therefore, Rh-specific antigenicities might be gained additionally after the gene duplication on the primate lineage of the Rh blood group gene.

Because we don't know the actual gene tree topology, we assumed two plausible trees from nucleotide sequence data in this study. Whichever gene tree we selected we found evidence of gene conversions on primate Rh blood group genes by using the modified site-by-site reconstruction method. We also showed the possibility of positive selection on the primate lineage (especially the branch that connects hominoid and Old World monkey clusters) of the Rh blood group gene by using a statistical test. In any case, we should be very careful when we analyze the evolutionary history of tandemly duplicated genes, for there is always possibility of gene conversions.

Acknowledgments. We thank Dr. Kenta Sumiyama for various discussions throughout the preparation of this manuscript and Dr. Yasuo Ina for providing his programs. This study was partially supported by grants in aid for scientific studies from Ministry of Education, Science, Sport and Culture, Japan to NS.

References

Adachi J, Hasegawa M (1994) MOLPHY. Programs for molecular phylogenetics, version 2.2. Institute of Statistical Mathematics, Tokyo

Arce M, Thompson ES, Wagner S, Coyne KE, Ferdman BA, Lublin DM (1993) Molecular cloning of Rh D cDNA derived from a gene

present in Rh D-positive but not Rh D-negative individuals. Blood 82:651–655

Avent ND, Ridgwell K, Tanner MJ, Anstee DJ (1990) cDNA cloning of a 30 kDa erythrocyte membrane protein associated with Rh (Rhesus)-blood-group-antigen expression. Biochem J 271:821–825

Avent ND, Butcher SK, Liu W, Mawby WJ, Mallinson G, Parsons SF, Anstee DJ, Tanner MJ (1992) Localization of the C termini of the Rh (rhesus) polypeptides to the cytoplasmic face of the human erythrocyte membrane. J Biol Chem 267:15134–15139

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

Beckers EA, Faas BH, Ligthart P, Simsek S, Overbeeke MA, von dem Borne AE, van Rhenen DJ, van der Schoot CE (1996) Characterization of the hybrid RHD gene leading to the partial D category IIIc phenotype. Transfusion 36:567–574

Blancher A, Socha WW (1997) The Rhesus system. In: Blancher A, Klein J, Socha WW (eds) Molecular biology and evolution of blood group and MHC antigens in primates. Springer, New York, NY, pp 147–218

Blancher A, Calvas P, Ruffie J (1992) Etude des equivalents des antigenes Rhesus chez les primates non hominiens. CR Soc Biol 186: 682–695

Carritt B, Kemp TJ, Poulter M (1997) Evolution of the human RH (rhesus) blood group genes: a 50 year old prediction (partially) fulfilled. Hum Mol Genet 6:843–850

Cherif-Zahar B, Bloy C, Le Van Kim C, Blanchard D, Bailly P, Hermand P, Salmon C, Cartron JP, Colin Y (1990) Molecular cloning and protein structure of a human blood group Rh polypeptide. Proc Natl Acad Sci USA 87:6243–6247

Cherif-Zahar B, Mattei MG, Le Van Kim C, Bailly P, Cartron JP, Colin Y (1991) Localization of the human Rh blood group gene structure to chromosome region 1p34.3–1p36.1 by in situ hybridization. Hum Genet 86:398–400

Cherif-Zahar B, Raynal V, D'Ambrosio AM, Cartron JP, Colin Y (1994) Molecular analysis of the structure and expression of the RH locus in individuals with D--Dc-, and DCw- gene complexes. Blood 84:4354–4360

Eder AF, Spitalnik SL (1997) Blood group antigens as receptors for pathogens. In: Blancher A, Klein J, Socha WW (eds) Molecular biology and evolution of blood group and MHC antigens in primates. Springer, New York, NY, pp 268–304

Endo T, Ikeo K, Gojobori T (1996) Large-scale search for genes on which positive selection may operate. Mol Biol Evol 13:685–690

Gahmberg CG (1982) Molecular identification of the human Rho (D) antigen. FEBS Lett 140:93–97

Hogstrand K, Bohme J (1998) Intrachromosomal gene conversion frequency in the H2 differs between haplotypes. Immunogenetics 48: 47–55

Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc Natl Acad Sci USA 92: 532–536

Huang CH, Reid ME, Chen Y (1995) Identification of a partial internal deletion in the RH locus causing the human erythrocyte D-phenotype. Blood 86:784–790

Huang CH, Chen Y, Reid M, Ghosh S (1996) Genetic recombination at the human RH locus: a family study of the red-cell Evans phenotype reveals a transfer of exons 2–6 from the RHD to the RHCE gene. Am J Hum Genet 59:825–833

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167–170

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

Ina Y, Gojobori T (1994) Statistical analysis of nucleotide sequences of the hemagglutinin gene of human influenza A viruses. Proc Natl Acad Sci USA 91:8388–8392

Ishida N, Oyunsuren T, Mashima S, Mukoyama H, 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

Kajii E, Umenishi F, Iwamoto S, Ikemoto S (1993) Isolation of a new cDNA clone encoding an Rh polypeptide associated with the Rh blood group system. Hum Genet 91:157–162

Kemp TJ, Poulter M, Carritt B (1996) A recombination hot spot in the Rh genes revealed by analysis of unrelated donors with the rare D-phenotype. Am J Hum Genet 59:1066–1073

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Kitano T, Sumiyama K, Shiroishi T, Saitou N (1998) Conserved evolution of the Rh50 gene compared to its homologous Rh blood group gene. Biochem Biophys Res Comm 249:78–85

Klein J, O'hUigin C, Blancher A (1997) Evolution of blood group antigen polymorphism. In: Blancher A, Klein J, Socha WW (eds) Molecular biology and evolution of blood group and MHC antigens in primates. Springer, New York, NY, pp 305–321

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

Landsteiner K, Wiener AS (1940) An agglutinable factor in human blood recognized by immune sera for blood. Proc Soc Exp Med 43:223

Le Van Kim C, Mouro I, Cherif-Zahar B, Raynal V, Cherrier C, Cartron JP, Colin Y (1992) Molecular cloning and primary structure of the human blood group RhD polypeptide. Proc Natl Acad Sci USA 89:10925–10929

Levine P, Stetoson RE (1939) An unusual case of intragroup agglutination. J Am Med Assoc 133:126–127

Li WH (1997) Molecular evolution. Sinauer Associates, Sunderland, MA

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

Matassi G, Cherif-Zahar B, Pesole G, Raynal V, Cartron JP (1999) The members of the RH gene family (RH50 and RH30) underwent different evolutionary pathways. J Mol Evol 48:151–159

Moore S, Woodrow CF, McClelland DB (1982) Isolation of membrane components associated with human red cell antigens Rh(D), (c), (E) and Fy. Nature 295:529–531

Mouro I, Colin Y, Cherif-Zahar B, Cartron JP, Le Van Kim C (1993) Molecular genetic basis of the human Rhesus blood group system. Nat Genet 5:62–65

Mouro I, Le Van Kim C, Cherif-Zahar B, Salvignol I, Blancher A, Cartron JP, Colin Y (1994a) Molecular characterization of the Rh-like locus and gene transcripts from the rhesus monkey (*Macaca mulatta*). J Mol Evol 38:169–176

Mouro I, Le Van Kim C, Rouillac C, van Rhenen DJ, Le Pennec PY, Bailly P, Cartron JP, Colin Y (1994b) Rearrangements of the blood group RhD gene associated with the DVI category phenotype. Blood 83:1129–1135

Race RR (1944) An "incomplete" antibody in human serum. Nature 153:771–772

Ruddle F, Ricciuti F, McMorris FA, Tischfield J, Creagan R, Darlington G, Chen T (1972) Somatic cell genetic assignment of peptidase C and the Rh linkage group to chromosome A-1 in man. Science 176:1429–1431

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

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

Salvignol I, Blancher A, Calvas P, Socha WW, Colin Y, Cartron JP, Ruffie J (1993) Relationship between chimpanzee Rh-like genes and the R-C-E-F blood group system. J Med Primatol 22:19–28

Salvignol I, Blancher A, Calvas P, Clayton J, Socha WW, Colin Y, Ruffie J (1994) Molecular genetics of chimpanzee Rh-related genes: their relationship with the R-C-E-F blood group system, the chimpanzee counterpart of the human rhesus system. Biochem Genet 32:201–221

Salvignol I, Calvas P, Socha WW, Colin Y, Le Van Kim C, Bailly P, Ruffie J, Cartron JP, Blancher A (1995) Structural analysis of the RH-like blood group gene products in nonhuman primates. Immunogenetics 41:271–281

Slightom JL, Theisen TW, Koop BF, Goodman M (1987) Orangutan fetal globin genes. J Biol Chem 262:7472–7483

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

Tippett P (1986) A speculative model for the Rh blood groups. Ann Hum Genet 50:241–247

Wiener AS (1943) Genetic theory of the Rh blood types. Proc Soc Exp Biol NY 54:316–319

Yang Z (1997) Phylogenetic analysis by maximum likelihood (PAML), version 1.3. Department of Integrative Biology, University of California at Berkeley, CA

Zhang J, Kumar S, Nei M (1997) Small-sample tests of episodic adaptive evolution: a case study of primate lysozymes. Mol Biol Evol 14:1335–1338