Immunogenetics (2007) 59:427–431
DOI 10.1007/s00251-007-0204-z

BRIEF COMMUNICATION

Tempo and mode of evolution of the Rh blood group genes before and after gene duplication

Takashi Kitano · Kazuo Umetsu · Wei Tian · Kentaro Yamazaki · Naruya Saitou

Received: 2 November 2006 / Accepted: 20 February 2007 / Published online: 3 March 2007

Abstract The Rh blood group genes became duplicated in a common ancestor of human–chimpanzee–gorilla. We compared the evolutionary rates of the Rh blood group genes for each exon for branches connecting to humans, having duplicated Rh loci, and to orangutan, gibbon, and Old World monkeys, species having a single Rh locus. Our results show that evolutionary rates of nonsynonymous substitutions at exon 7 became accelerated in the human lineage. Furthermore, we surveyed the sequence variation in the region surrounding exon 7 of gibbons to clarify whether the diversity of the human exon 7 was introduced after the duplication or had been maintained before it. Two amino acid polymorphisms in white-handed gibbons were observed in the immediate vicinity of the D-specific motif in the human exon 7. Although the evolutionary rate of exon 7 was accelerated after the gene duplication, our results suggest that exon 7 had the potential for change even before the gene duplication.

Keywords Rh · Blood · Gibbon · Primate · Nucleotide diversity · Evolutionary rate

Electronic supplementary material The online version of this article (doi:10.1007/s00251-007-0204-z) contains supplementary material, which is available to authorized users.

T. Kitano (✉) · K. Umetsu · W. Tian · K. Yamazaki
Department of Experimental and Forensic Pathology,
Yamagata University School of Medicine,
2-2-2 Iidanishi,
Yamagata 990-9585, Japan
e-mail: tkitano@med.id.yamagata-u.ac.jp

N. Saitou
Division of Population Genetics, National Institute of Genetics,
Mishima 411-8540, Japan

The human Rh blood group plays important roles in transfusion and clinical medicine, being involved in hemolytic diseases of newborns, autoimmune diseases, and mild hemolytic anemia. The Rh blood group system is composed of two closely linked loci, D and CE (Mouro et al. 1993). Each locus has ten exons (Cherif-Zahar et al. 1994), and the D and CE polypeptides have only an 8.4% amino acid difference (Le van Kim et al. 1992). These two loci are located on chromosome 1p34–p36 in humans (Ruddle et al. 1972; Cherif-Zahar et al. 1991), have opposite orientation, and face each other with their 3′ ends (Suto et al. 2000; Wagner and Flegel 2000). It has been suggested that gene duplication occurred after the divergence of the human–African ape clade from Asian apes (orangutan and gibbon; Salvignol et al. 1993; Blancher and Socha 1997), giving rise to the two loci, by southern blot analyses.

Innan (2003) analyzed polymorphisms in human D and CE loci, and observed very high amino acid divergence between the two loci only in exon 7. According to his hypothesis, immediately after the gene duplication that produced D and CE loci, gene conversion occurred quite frequently at both loci, at a rate high enough to keep the genes of the two loci nearly identical. Then an advantageous mutation was introduced (probably in exon 7) and became fixed in one locus. Selection might have been so strong that this fixation state was nearly stable and continued for quite a long time. During this state, additional mutations accumulated near the target site of selection, creating a high level of sequence divergence between the two loci around exon 7. As an alternative evolutionary process, however, one can assume that those polymorphisms of exon 7 had been maintained in one locus for a long time before the duplication, such as in the case of the Duffy blood group gene (Hamblin et al. 2002) and glycophorin A gene (Baum et al. 2002). The Duffy blood

group locus has an extreme pattern of geographic differentiation of its three major alleles in humans because of resistance to malaria. Allelic variations of the glycophorin A gene within human populations have been maintained by balancing selection because the product of this gene is one of the components on the erythrocyte surface.

To determine whether the diversity at exon 7 of the Rh blood group genes was introduced after the duplication or had been maintained before the duplication, we surveyed the sequence variation in the region surrounding exon 7 (approx. 1 kb) in nine white-handed gibbons (*Hylobates lar*) and five siamangs (*Hylobates syndactylus*), both of which have a single Rh locus. To compare the level of polymorphisms in exon 7 with that in another exon, we also analyzed the region surrounding exon 5 (approx. 1 kb) in the same samples because exon 5 is the second longest exon in the Rh blood group gene and is a neighbor of exon 7. There has been no earlier study that tried to analyze polymorphisms for primate species before the gene duplication of the Rh blood group genes. The DDBJ/EMBL/GenBank International Nucleotide Sequence Database accession numbers are AB289827–AB289856. A list of primers used in this study is available from Kitano. Haplotypes were estimated by using the PHASE 2.0.2 program (Stephens et al. 2001).

Table 1 shows polymorphisms in hominoids at exons 5 and 7 of the Rh blood group genes. The average number (π) of nucleotide differences per site between two sequences for exon 7 of white-handed gibbons was 0.009. It is no surprise that this value is four to six times lower than the π value for exon 7 of humans, chimpanzees, and gorillas, all of which have duplicated Rh loci. However, π of exon 7 in white-handed gibbons was three times higher than that of exon 5. This ratio of π for exon 7 to exon 5 (π [ex7/ex5]) in the white-handed gibbon was comparable to that for other hominoids. Thus, this finding suggests that exon 7 of the Rh blood group gene had greater potential for change than other exons before the gene duplication. Tajima's (1989) *D* was calculated for exons 5 and 7 of white-handed gibbons. These values were positive (1.378 for exon 5 and 1.175 for exon 7) but did not differ significantly from the neutral expectation of zero.

We further analyzed the distributions of π for two species of gibbons at the regions surrounding exons 5 (Fig. 1a) and 7 (Fig. 1b). Nucleotide differences (*d*) between the two gibbon species and between the gibbon and the orangutan at the regions surrounding exons 5 and 7 are also shown in this figure. Interestingly, the nucleotide diversity was clustered at exon 7 in white-handed gibbons (Fig. 1b). Three polymorphic sites were observed within exon 7 of the white-handed gibbons (Table 1). Two of them were nonsynonymous differences (V-352-F and G-355-S), and the remaining one was a synonymous site at the 330th amino acid residue. These two nonsynonymous polymorphisms in the white-handed gibbons were in the immediate vicinity of the D-specific motif D-350, G-353, and A-354 in exon 7 of humans, and this motif is located on the sixth outer membrane of the protein (Cartron and Agre 1993). This finding also suggests that exon 7 of the Rh blood group gene had the potential for change before the gene duplication. Most genes are evolving under the neutral mutation pressure (Kimura 1983). However, some genes are evolving through positive selection. 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). Thus, it is possible that the amino acids of the sixth outer membrane of the Rh protein may be involved in some kind of host–parasite interaction.

To compare the evolutionary rate before and after the gene duplication of the Rh blood group genes, we calculated the evolutionary rate of synonymous (λ_S) and nonsynonymous (λ_N) substitutions for each exon for branches connecting to Old World monkeys (OWM), gibbon, orangutan, and human species (Table 2). The

Table 1 Polymorphisms in hominoids at exons 5 and 7 of the Rh blood group genes

	Exon 5 (167 bp)				Exon 7 (134 bp)				π (ex7/ex5)	Sources
	N	*S*	π	θ_w	*N*	*S*	π	θ_w		
Human	80	17	0.024	0.021	67	20	0.054	0.031	2.29	From DNA databank[a]
Chimpanzee	11	7	0.012	0.014	11	25	0.066	0.064	5.63	From DNA databank[a] and Kitano et al. (unpublished)
Gorilla	4	5	0.017	0.016	4	10	0.044	0.041	2.57	From DNA databank[a] and Kitano et al. (unpublished)
White-handed gibbon	18	1	0.003	0.002	18	3	0.009	0.007	3.08	This study
Siamang	10	0	0.000	0.000	10	0	0.000	0.000	–	This study

N Number of chromosomes, *S* number of segregating sites, π average number of nucleotide differences per site between two sequences (Nei 1987), θ_w nucleotide diversity based on the proportion of segregating sites in a sample (Watterson 1975), π (*ex7/ex5*) ratio of πs for exon 7 to exon 5, – not estimated

[a] Accession numbers are listed in Supplementary Table 1.

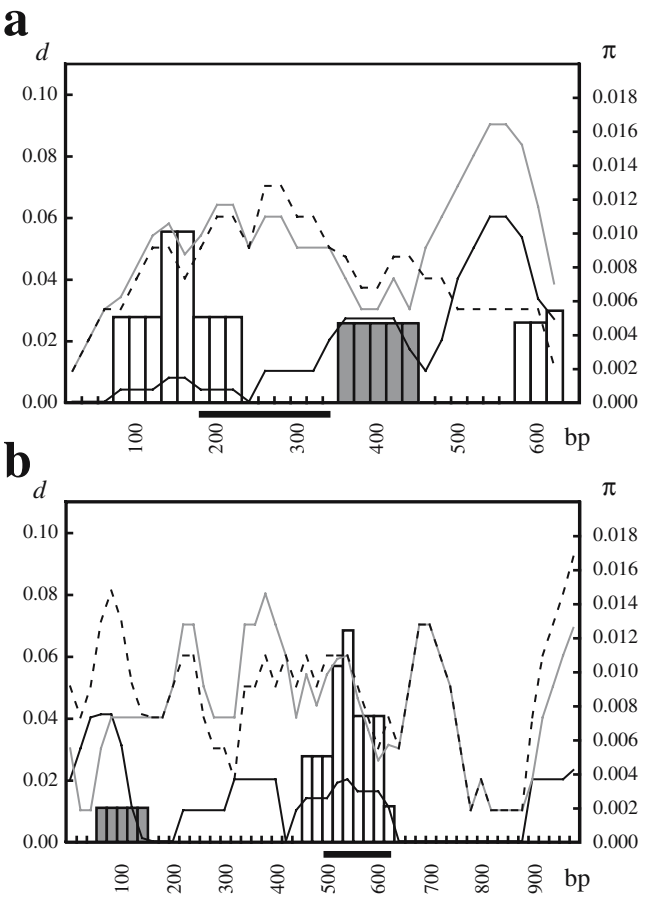

Fig. 1 Sliding window analyses of π and average nucleotide differences between two species for regions surrounding exon 5 (**a**) and exon 7 (**b**). Window size of 100 bp and sliding range of 20 bp were used. Exons are shown by thick horizontal bars. *White bar* π for the white-handed gibbon, *gray bar* π for the siamang, *black line* the average number of nucleotide differences (d) between the white-handed gibbon and siamang, *broken line d* between the siamang and the orangutan, *grey line d* between the white-handed gibbon and the orangutan

established phylogeny for primate species (e.g., Horai et al. 1995; Goodman et al. 1998) was adopted for primate Rh blood group genes (Cherif-Zahar et al. 1990; Le van Kim et al. 1992; Mouro et al. 1994; Salvignol et al. 1994; Apoil and Blancher 1999), and divergence times (T) of the human lineage from the orangutan, gibbon, and Old World monkey were assumed to be 13, 15, and 23 million years ago (Glazko and Nei 2003; Fig. 2). The Rh blood group genes of the chimpanzee and gorilla were not included in this tree because gene conversions in them interrupt the orthologous relationship among their duplicated genes (Kitano and Saitou 1999; see also Supplementary Fig. 1). For simplicity, human D and CE loci were used as representatives of duplicated Rh genes in the phylogenetic tree. The nucleotide sequences of ancestral nodes were reconstructed by the Bayesian method (Yang et al. 1995) using the program BASEML in the PAML 3.15 package (Yang 1997). d_N (the number of nonsynonymous substitutions per nonsynonymous site) and d_S (the number of synonymous substitutions per synonymous site) were estimated (Nei and Gojobori 1986) for each branch, and λ_S and λ_N were estimated by the formula $\lambda = d/T$. d_N and d_S for the branch of Node *Z*—Human were estimated as [((branch *g* + branch *h*)/2)+branch *f*], and d_N and d_S for the branch of Node *X*—OWM were estimated as [(((branch *d*+branch *e*)/2) + branch *c*+branch *b*)/2 + branch *a*].

λ_S values for all (exons 1–9) were 1.7×10^{-9}, 2.4×10^{-9}, 1.5×10^{-9}, and 1.8×10^{-9} for branches of Old World monkeys (Node *X*—OWM), gibbons (Node *Y*—Gibbon), orangutans (Node *Z*—Orangutan), and humans (Node *Z*—Human), respectively. These λ_S values were lower than the average value (4.6×10^{-9}) of other mammalian genes (Li and Graur 1991). Exon 9 of the orangutan showed an extremely high λ_S value. This is probably a stochastic effect

Table 2 Evolutionary rate (λ, $\times10^{-9}$) of synonymous (λ_S) and nonsynonymous (λ_N) substitutions for each exon for each branch connecting to OWM, gibbon, orangutan, and human species

Exon	bp	Node *X*—OWM		Node *Y*—gibbon		Node *Z*—orangutan		Node *Z*—human	
		λ_S	λ_N	λ_S	λ_N	λ_S	λ_N	λ_S	λ_N
1	146 (144)	0.9 (0.9)	0.0 (0.0)	1.8 (1.9)	1.9 (0.7)	2.1 (2.2)	0.0 (0.0)	4.2 (4.4)	2.2 (0.8)
2	184 (183)	0.0 (0.0)	2.5 (2.5)	1.5 (1.6)	1.0 (0.5)	0.0 (0.0)	4.3 (4.4)	1.8 (1.8)	2.0 (1.5)
3	150 (149)	1.2 (1.2)	4.6 (4.1)	0.0 (0.0)	1.8 (1.3)	2.1 (2.1)	1.4 (1.5)	2.1 (2.1)	4.2 (4.3)
4	146 (143)	2.0 (2.1)	3.2 (2.8)	1.0 (1.1)	2.8 (2.3)	2.3 (2.4)	2.2 (0.8)	0.0 (0.0)	5.4 (3.8)
5	167 (164)	0.3 (0.3)	1.7 (1.2)	7.2 (5.7)	2.2 (1.1)	2.0 (2.2)	0.6 (0.6)	1.0 (1.1)	5.0 (4.0)
6	138 (135)	3.8 (2.8)	3.0 (2.2)	0.0 (0.0)	0.7 (0.0)	0.0 (0.0)	1.6 (0.8)	2.0 (2.2)	2.4 (2.5)
7	129 (126)	2.0 (0.7)	3.0 (2.2)	6.3 (4.5)	2.9 (2.3)	0.0 (0.0)	0.0 (0.0)	2.3 (1.2)	*15.2* (*13.9*)
8	76 (74)	2.3 (2.5)	3.5 (1.9)	0.0 (0.0)	5.3 (4.3)	0.0 (0.0)	4.5 (5.0)	0.0 (0.0)	4.6 (3.3)
9	69 (69)	7.9 (7.9)	0.0 (0.0)	4.6 (4.6)	1.3 (1.3)	*11.5* (*11.5*)	1.5 (1.5)	2.7 (2.7)	2.2 (2.2)
All	1205 (1187)	1.7 (1.5)	2.4 (2.0)	2.4 (2.0)	2.1 (1.3)	1.5 (1.6)	1.7 (1.5)	1.8 (1.7)	4.7 (3.9)

Sequence data AB289846 and L37051 are used for exon 7 and other exons, respectively, for the gibbon. λ_S and λ_N greater than 10.0×10^{-9} are shown in italics. The numbers in parentheses indicate values estimated from data including ambiguously reconstructed sites (probability less than 0.95). See Fig. 2 for locations of nodes *X*–*Z*.

bp Compared nucleotide sites

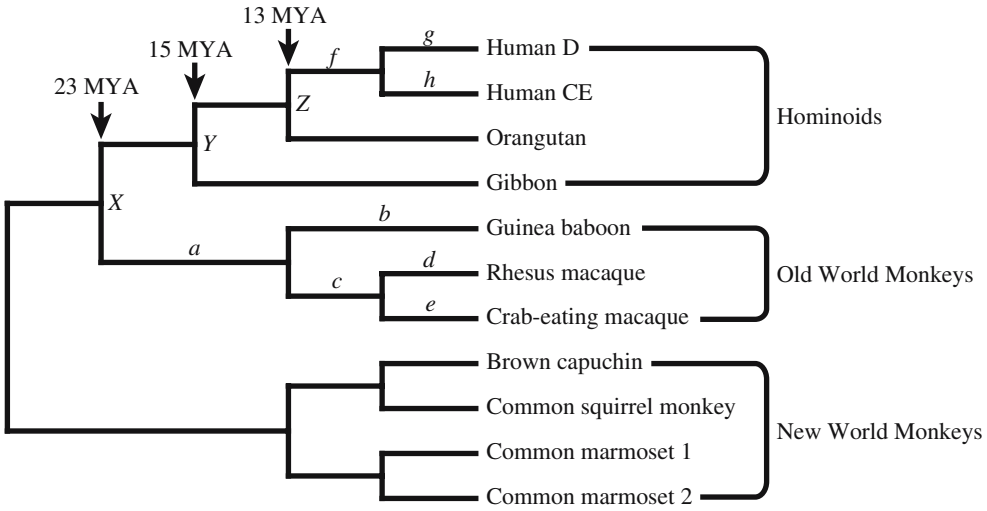

Fig. 2 A phylogenetic tree for primate Rh blood group genes. Nucleotide sequences were retrieved from the DDBJ/EMBL/GenBank International Nucleotide Sequence Database: human D (X63097), human CE (M34015), orangutan (AF012425), gibbon (L37051 and AB289846), guinea baboon (AF012426), rhesus macaque (S70343), crab-eating macaque (L37054), brown capuchin (AF101479), common squirrel monkey (AF012428), and two common marmosets (AF012429 and AF052588). The nucleotide sequences ancestral nodes (*X*–*Z*) were reconstructed by the Bayesian method (Yang et al. 1995). Some branches are indicated by *a*–*h*

because of the small number of compared sites (two synonymous substitutions in 69 compared sites). In contrast, λ_N for all (exons 1–9) were 2.4×10^{-9}, 2.1×10^{-9}, 1.7×10^{-9}, and 4.7×10^{-9} for branches of Old World monkeys (Node *X*—OWM), gibbons (Node *Y*—Gibbon), orangutans (Node *Z*—Orangutan), and humans (Node *Z*—Human), respectively. These λ_N values were somehow higher than the average value (0.9×10^{-9}) of other mammalian genes (Li and Graur 1991), suggesting positive selection on the Rh blood group genes (Kitano et al. 1998; Kitano and Saitou 1999). In particular, λ_N of the human exon 7 showed an extremely high value. When we eliminated sites showing a probability of accuracy for reconstructed nucleotides at each ancestral node of less than 0.95 and estimated λ_N, exon 7 of humans still showed an extremely high value (13.9×10^{-9}). We also carried out the maximum parsimony method to reconstruct ancestral sequences by the PAMP program in the PAML package. There were some incompatible sites between the Bayesian and the maximum parsimony methods, but all of these sites belonged to ambiguously reconstructed sites (probability less than 0.95; see Table 2). Ideally, estimation of λ_S and λ_N for the human branch should be divided into that before and that after the gene duplication, but we did not do so because it is not easy to estimate the time of the gene duplication because of positive selection and gene conversions. At least we can say that λ_N of exon 7 became higher after the divergence between the orangutan and the common ancestor of human–chimpanzee–gorilla.

To compare the evolutionary rate between human D and CE lineages, we applied the relative rate test (Tajima 1993) for translated amino acid sequences of the Rh blood group genes (Table 3). The orangutan sequence was used as an outgroup. Amino acid changes tended to have accumulated in human D (19 in total) rather than in human CE (12 in total). However, this difference was not significant (p= 0.209). In exon 7, five and three amino acid changes were observed in human D and CE, respectively, but the difference was not statistically significant (p=0.480). When the gibbon sequence was used as an outgroup, similar results were obtained (data not shown). The result indicates that both D and CE loci accumulated amino acid changes after the gene duplication at similar levels.

Table 3 Results of relative rate test of human D- and CE-translated amino acid sequences

Exon	AA	Human D	Human CE	Orangutan	*p*
1	49	0	0	3	–
2	60	3	0	9	0.083
3	50	3	1	6	0.317
4	48	2	3	4	0.655
5	56	3	5	5	0.480
6	46	2	0	4	0.157
7	45	5	3	8	0.480
8	24	0	0	6	–
9	23	1	0	2	0.317
All	401	19	12	47	0.209

AA Number of amino acid residues compared, *Human D* number of amino acid changes on the human D branch, *Human CE* number of amino acid changes on the human CE branch, *Orangutan* number of amino acid changes on the branch between orangutan and the common ancestor of human D and CE

In conclusion, our results show that the evolutionary rate for nonsynonymous substitutions at exon 7 became accelerated in the human lineage. The results of the relative rate test suggested that both D and CE loci accumulated amino acid changes in exon 7 after the gene duplication rather than that one gene maintained its original function and the other gene accumulates amino acid changes by relaxation of functional constraints after the gene duplication. From the polymorphism analysis of gibbons, which species has a single Rh locus, a clear pattern of balancing selection was not observed by Tajima's *D* test. However, the higher nucleotide diversity at exon 7 than at exon 5 in the white-handed gibbons suggests that, in contrast with other exons, exon 7 had greater potential for change than other exons. Thus, it is likely that exon 7 of the Rh blood group genes had a changeable nature beforehand, and then this tendency might have been enhanced by the gene duplication.

Acknowledgments We are grateful to the Great Ape Information Network (GAIN) for providing us an orangutan DNA sample and to Dr. Osamu Takenaka (now deceased) for providing us gibbon DNA samples. This study was supported by a grant-in-aid for scientific studies from the Ministry of Education, Science, Sports, and Culture of Japan to TK.

References

Apoil P, Blancher A (1999) Sequences and evolution of mammalian RH gene transcripts and proteins. Immunogenetics 49:15–25

Baum J, Ward RH, Conway DJ (2002) Natural selection on the erythrocyte surface. Mol Biol Evol 19:223–229

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-Verlag, Heidelberg, pp 147–218

Cartron JP, Agre P (1993) Rh blood group antigens: protein and gene structure. Semin Hematol 30:193–208

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

Glazko CV, Nei M (2003) Estimation of divergence times for major lineages of primate species. Mol Biol Evol 20:424–434

Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, Shoshani J, Gunnell G, Groves CP (1998) Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. Mol Phylogenet Evol 9:585–598

Hamblin MT, Thompson EE, Di Rienzo A (2002) Complex signatures of natural selection at the Duffy blood group locus. Am J Hum Genet 70:369–383

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

Innan H (2003) A two-locus gene conversion model with selection and its application to the human RHCE and RHD genes. Proc Natl Acad Sci USA 100:8793–8798

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

Kitano T, Saitou N (1999) Evolution of Rh blood group genes have experienced gene conversions and positive selection. J Mol Evol 49:615–626

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 Commun 249:78–85

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

Li WH, Graur D (1991) Fundamentals of molecular evolution. Sinauer, Massachusetts

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 (1994) Molecular characterization of the Rh-like locus and gene transcripts from the rhesus monkey (*Macaca mulatta*). J Mol Evol 38:169–176

Nei M (1987) Molecular evolutionary genetics. Columbia Univ. Press, New York

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

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, 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

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989

Suto Y, Ishikawa Y, Hyodo H, Uchikawa M, Juji T (2000) Gene organization and rearrangements at the human Rhesus blood group locus revealed by fiber-FISH analysis. Hum Genet 106:164–171

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Tajima F (1993) Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135:599–607

Wagner FF, Flegel WA (2000) RHD gene deletion occurred in the Rhesus box. Blood 95:3662–3668

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256–276

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13:555–556

Yang Z, Kumar S, Nei M (1995) A new method of inference of ancestral nucleotide and amino acid sequences. Genetics 141:1641–1650