BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **249,** 78–85 (1998)
ARTICLE NO. RC989074

Conserved Evolution of the Rh50 Gene Compared to Its Homologous Rh Blood Group Gene

Takashi Kitano,* Kenta Sumiyama,*[,1] Toshihiko Shiroishi,† and Naruya Saitou*
**Laboratory of Evolutionary Genetics and †Laboratory of Mammalian Genetics, National Institute of Genetics, Mishima 411-8540, Japan*

Received July 3, 1998

We have sequenced the complete coding region of the Rh blood group gene for mouse and rat and that of Rh-related 50 kD glycoprotein (Rh50) for mouse, rat, and crab-eating macaque. Phylogenetic analyses of Rh and Rh50 amino acid sequences indicate that the Rh50 gene has been evolving about two times more slowly than the Rh blood group gene in both primates and rodents. This conservative nature of the Rh50 gene suggests its relative importance to the Rh blood group gene. The time of gene duplication that produced the Rh and Rh50 genes was estimated to be about 240-310 million years ago. We also conducted window analyses of synonymous and nonsynonymous nucleotide substitutions for those two genes. Some peaks where nonsynonymous substitutions are higher than synonymous ones were located on outer membrane regions. This suggests the existence of positive Darwinian selection on Rh and Rh50 genes through host-parasite interactions.

Key Words: **evolutionary rate; phylogenetic tree; window analysis; positive selection.**

The human Rh blood group plays important roles in transfusion and clinical medicine, including haemolytic diseases of newborns, autoimmune diseases, and mild haemolytic anemia. Nucleotide sequences of human Rh genes were determined (1-5), and their products were estimated to have 12 transmembrane domains through hydropathy analysis (2) and immunological studies using an antipeptide antibody (6). Nucleotide sequences of Rh blood group genes in nonhuman primates were also reported (7-9). We recently analyzed those published Rh blood group genes of primates, and found a higher nonsynonymous substitutions than synonymous ones, a clear evidence of positive Darwinian evolution (Kitano and Saitou, unpublished). Therefore, it is interesting if a similar evolutionary pattern also exists in other mammalian groups.

A protein was obtained together with the Rh gene product on immunoprecipitation with anti-Rh antibodies from human, and named as 50 kD glycoprotein (10). This glycoprotein was considered to form heterotetramer with Rh blood group gene products on erythrocyte membranes (11). The nucleotide sequence of the human 50 kD glycoprotein was determined, and its amino acid sequence was homologous with that of the human Rh gene (12). That protein was also predicted to have the 12 trans-membrane domains which are similar to those of the Rh blood group gene product. There are several names for this gene such as RHAG, but we call this gene as Rh50 and the Rh blood group gene as Rh hereafter for simplicity. It has been shown that the Rh_{null} regulator and the Rh_{mod} phenotypes are suppressed by the Rh50 product (13), and a splicing mutant of this gene was shown to cause an Rh_{null} phenotype (14). These observations clearly indicate that the Rh50 gene is essential for expression of Rh antigens on erythrocytes.

These Rh gene and Rh related gene products seem to play an important role for erythrocytes. In this study, we determined nucleotide sequences for Rh genes and Rh50 genes of two mouse subspecies and rat as well as the Rh50 gene for crab-eating macaque, and compared these evolutionary relationships.

MATERIALS AND METHODS

PCR-direct sequencing of cDNA. Total RNAs were extracted from bone marrow of two mouse (*Mus musculus*) subspecies (*M. m. domesticus* and *M. m. brevirostris*), rat (*Rattus norvegicus*), and crab-eating macaque (*Macaca fascicularis*), using the AGPC (Acid Guanidinium-Phenol-Chloroform) method. After DNase reactions, reverse transcription was performed by using AMV (Avian Myeloblastosis Virus) reverse transcriptase and oligo dT-adaptor primer of RNA PCR Kit AMV Ver. 2.1 (TaKaRa). Degenerate PCR was performed and a partial product was obtained. We then performed

[1] Present address: Department of Biology, Yale University, New Haven, CT 06520-8103.

0006-291X/98 $25.00

5′ RACE (rapid amplification of the 5′ cDNA ends) using 5′RACE System for Rapid Amplification of cDNA Ends version 2.0 (Gibco-BRL). PCR was performed in a 20 μl reaction containing 0.5-1 μl of the first-strand cDNA, 1 × Gene Taq Universal Buffer (Mg^{2+} free) (Nippon Gene), 1.5 mM $MgCl_2$, 0.2 mM dNTP, 10 pmol of each primer (designed on sites of 5′ and 3′ ends), and 1 units of AmpliTaq Gold (Perkin-Elmer). Amplification was carried out in DNA GeneAmp PCR System 2400 (Perkin-Elmer) with the following temperature parameters: 10 min at 95°C followed by 40 cycles of 95°C for 30 sec, 65°C for 15 sec, and 72°C for 1 min. PCR products were purified using MicroSpin Columns S-300 HR (Pharmacia Biotech) and cloned in the TA cloning vectors pCRII (Invitrogen). DNA sequencing was performed on double-stranded plasmid DNA and PCR products using Dye Terminator Cycle Sequencing Kit and ABI prism 377 DNA sequencer (Perkin-Elmer). A progressive sequencing strategy was carried out with design of further primers to complete the sequence for coding region of both strands of the cDNA.

Sequence analyses. CLUSTAL W version 1.6 (15) was used for multiple alignments and tree analyses. The neighbor-joining method (16) was used for constructing phylogenetic trees. Numbers of amino acid substitutions were estimated by using Kimura's method (17). The PredictProtein server (EMBL) was used for analyses of transmembrane helix location of these proteins. ODEN package (18) was used to estimate numbers of synonymous and nonsynonymous substitutions (19), and the WINA program (20) was used for window analyses.

RESULTS AND DISCUSSION

PCR-cloning and sequencing of Rh and Rh50 cDNA coding regions. We sequenced Rh and Rh50 gene cDNAs for two mouse subspecies and rat. Those newly determined rodent sequences (DDBJ/EMBL/GenBank international nucleotide sequence database accession numbers are AB015189 - AB015194) were compared with the human Rh gene (1, 2) and the human Rh50 gene (12). Figure 1 shows the multiple alignment of nucleotide sequences of Rh genes. Nucleotide sequence lengths of human, crab-eating macaque, mouse, and rat are 1254, 1254, 1257, and 1269 bp, respectively. Four gaps (3, 15, 3, and 6 nucleotide long) were observed between primate and rodent sequences, and the rat Rh gene had extra 12 nucleotides (positions 337-348). Lengths of all gaps were multiplication of 3 and there is no frame shift. We also obtained an incomplete sequence for the rat Rh cDNA which lacks sites 149-661. These sites correspond to exons 2-4 of the human Rh gene, and this incomplete cDNA were probably produced by a splicing error.

Figure 2 shows the multiple alignment of nucleotide sequences of Rh50 genes. We obtained a Rh50 gene cDNA for crab-eating macaque (DDBJ/EMBL/GenBank accession number is AB015467), and it was also compared. Nucleotide sequence lengths of human, crab-eating macaque, mouse, and rat are 1230, 1287, 1317, and 1353 bp, respectively. The location of the stop codon of the human Rh50 gene is different from that of others, and its protein is 19 amino acids shorter corresponding to this region. There are repeats of 15 nucleotides around positions 100-150 (see Fig. 2), and its consensus sequence is AATGCTTCCCAGCAG. Rat and mouse have 5 and 3 repeats, respectively, while the two primate species have single repeat. Because all gaps were multiple of 3, they did not alter codon frames.

Sequence similarities (both for nucleotide and amino acid) are shown in Table 1. Nucleotide and amino acid sequence similarities between Rh genes and Rh50 genes are 47-49 and 34-38%, respectively. The GC contents of Rh and Rh50 genes were 53-55 and 45-47%, respectively (shown on the diagonal of Table 1). These values were similar to those previously reported (21), and may be related to gene locations on gnomes; the Rh gene is located on chromosome 1p34-36 (22-23), while the Rh50 gene is on chromosome 6p21-qter (12). We would like to note that similar sequencing results for mouse and macaque were recently obtained independently by G. Matassi *et al.* (personal communication).

Evolutionary rates and the phylogenetic tree. We estimated numbers of synonymous (d_S) and nonsynonymous (d_N) substitutions for Rh and Rh50 genes (Table 2). d_S and d_N values between primates and rodents were estimated by averaging pairwise values. Numbers of synonymous substitutions (d_S) were similar between Rh and Rh50 genes, and they are more or less similar to those for other genes (24). Branching pattern of the Rh and Rh50 genes are also compatible with the established mammalian phylogeny. This indicates that we did orthologous comparison both for Rh and Rh50 genes.

Numbers of nonsynonymous substitutions (d_N) are about two times higher for the Rh gene than for the Rh50 gene; the ratios of Rh-d_N and Rh50-d_N are 2.0, 1.7, and 2.0 for human-macaque, mouse-rat, and primates-rodents comparisons, respectively (we neglected the comparison of the two mouse subspecies, for standard errors are so large). This evolutionary conservation of the Rh50 gene suggests that it may have more important function than the Rh gene. A relatively uniform ratio of Rh-d_N and Rh50-d_N for three different levels of divergence also suggests that a molecular clock (constancy of evolutionary rate) exists both for Rh and Rh50 genes.

Majority of genes are known to undergo neutral evolution, and number (d_S) of synonymous substitutions are expected to be higher than those (d_N) for nonsynonymous substitutions under this situation (17). We compared d_S and d_N values to see if there is any unusual pattern deviated from neutrality in Rh and Rh50 genes. d_N of both Rh and Rh50 genes were higher than d_S when human and macaque sequences were compared, while the situation is reversed for other comparisons (Table 2). We previously found that many branches of a phylogenetic tree of primate

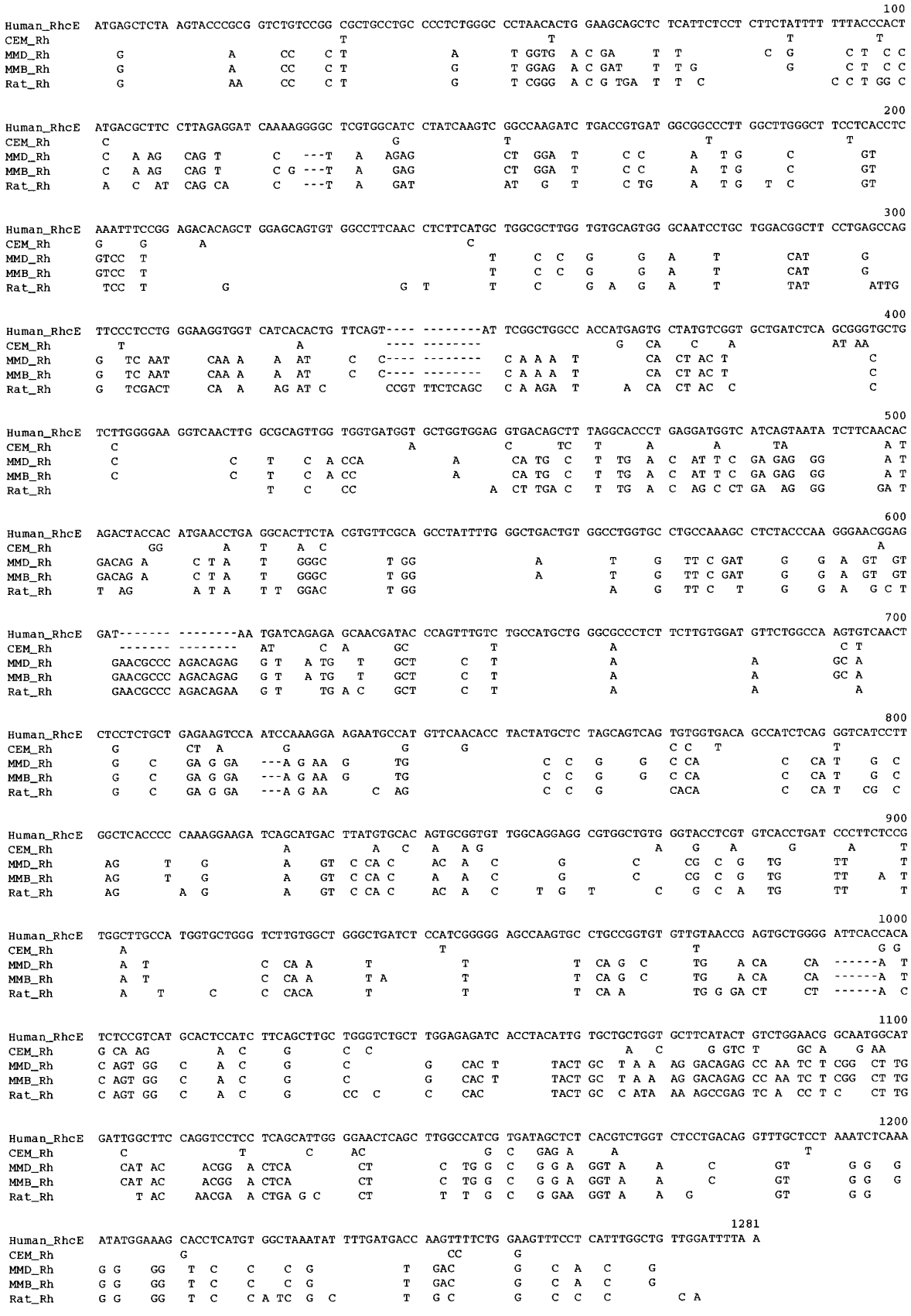

FIG. 1. The multiple alignment of nucleotide sequences of Rh genes. Nucleotide sequences for human RhcE (M34015 or X54534; they are identical) and for crab-eating macaque (L37054) were also included for comparison. Gaps are denoted by hyphens, and only nucleotides different from those of the human sequence are shown. MMD, MMB, and CEM denote *M. m. domesticus* and *M. m. brevirostris*, and crab-eating macaque, respectively.

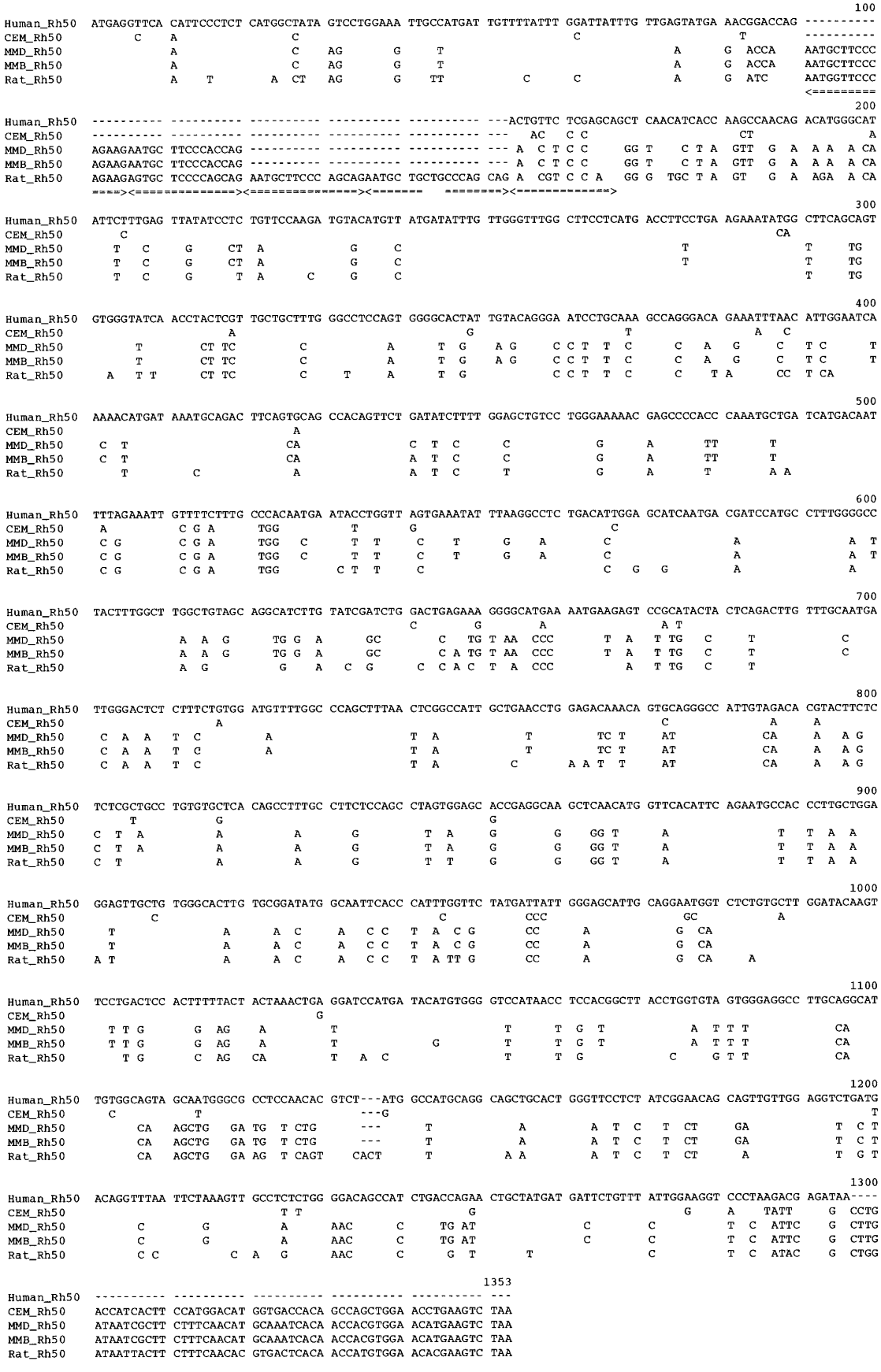

FIG. 2. The multiple alignment of nucleotide sequences of Rh50 genes. Human Rh50 (X64594) was also included. Gaps are denoted by hyphens, and only nucleotides different from those of the human sequence are shown. Equal signs surrounded with angled brackets designate the repeat unit of 15 nucleotides. Abbreviations of species are the same as Fig. 1.

TABLE 1

Similarities (%) of Rh and Rh50 Nucleotide Sequences (above Diagonal) and Amino Acid Sequences (below Diagonal) and GC Content (%) of Each Gene (on the Diagonal in Parentheses)

	1	2	3	4	5	6	7	8	9	10
1. Human RhcE	(53.7)	90.4	71.4	71.3	70.8	48.6	48.5	47.3	47.4	47.3
2. CEM Rh	79.1	(52.5)	71.9	71.8	70.9	48.8	48.9	48.4	48.5	48.2
3. MMD Rh	57.9	59.1	(55.2)	99.1	88.3	48.6	48.9	47.2	47.1	46.9
4. MMB Rh	58.1	59.3	98.3	(55.3)	88.3	48.5	48.8	47.1	47.0	46.8
5. Rat Rh	56.7	58.6	81.6	81.8	(54.0)	48.8	48.5	47.7	47.6	47.4
6. Human Rh50	35.2	35.7	37.1	36.8	35.3	(46.7)	94.6	80.0	79.8	79.4
7. CEM Rh50	35.4	37.7	37.8	37.5	36.3	88.8	(47.4)	79.6	79.4	78.6
8. MMD Rh50	34.4	35.4	35.0	35.0	35.5	77.0	74.3	(45.0)	99.8	91.6
9. MMB Rh50	34.7	35.7	34.8	34.8	35.3	76.8	74.1	99.8	(45.0)	91.7
10. Rat Rh50	33.8	35.1	34.7	34.7	34.4	75.8	73.1	88.8	89.0	(45.4)

Note. MMD, MMB, and CEM denote *M. m. domesticus, M. m. brevirostris,* and crab-eating macaque, respectively.

Rh genes showed higher d_N than d_S (Kitano and Saitou, unpublished), and this is compatible with a higher d_N for human and macaque Rh gene shown in Table 2. It is interesting that the Rh50 gene also showed a similar evolutionary pattern for primates, but not for rodents. If the heterotetramer structure of the Rh and Rh50 gene products is correct, it is possible that this erythrocyte membrane protein complex is under some kind of positive selection in primates but not in rodents.

We constructed a multiple alignment of amino acid sequences of Rh, Rh50, and their related genes (alignment not shown). Two genes of *C. elegans* (25) and an Rh-like gene of sponge (26) found by database searches were also included. The 12 predicted hydrophobic membrane-spanning regions did not include gaps and are relatively conserved. We thus used only the 216 amino acid sites for membrane-spanning regions for construction of the neighbor-joining tree (Fig. 3). The root was located by assuming the Rh-like protein of sponge as an outgroup. There are three clusters in this tree; Rh50 genes of mammals, Rh genes of mammals, and two genes of *C. elegans*. The branch lengths of Rh50 genes is much shorter than those of Rh genes, indicating a lower evolutionary rate in the Rh50 gene than in the Rh gene. This pattern is consistent with the result of d_N in Table 2 where all the coding regions were compared. It is interesting that after the gene duplication (node D in Fig. 3) which produced Rh and Rh50 genes, the Rh gene lineage started to evolve more rapidly than the Rh50 lineage.

Numbers of amino acid substitutions (d_A) are also estimated (Table 3). The phylogenetic tree of Fig. 3 was used and single-lineage d_A values were obtained applying Ishida *et al.*'s (27) method. Because d_A values for Rh were consistently two - three times higher than those for Rh50, a rough molecular clock exists for both genes. This is consistent with the result of Table 2. Therefore, we estimated evolutionary rates of Rh and Rh50 genes by using the regression through origin. Divergence times between human and macaque, between mouse and rat, and between primates and rodents were assumed to be 23.3, 40.7, and 112 million years (28), and they were used for calibration of the molecular clock. The rate of amino acid substitutions (per site per year) were thus obtained as 2.12×10^{-9} and 0.94×10^{-9} for Rh and Rh50 genes, respectively. If we use these rates, the time of gene duplication (node D in Fig. 3) producing Rh and Rh50 genes was estimated to be about 240 or 310 million years ago from the data for Rh or Rh50, respectively. This period roughly corresponds to the late Paleozoic where the mammalian lin-

TABLE 2

Numbers of Synonymous (d_S) and Nonsynonymous (d_N) Substitutions

	MMD vs MMB[a]	Human vs macaque[a]	Mouse vs rat[a]	Primates vs rodents[b]
d_S of Rh	0.013 ± 0.007	0.071 ± 0.016	0.226 ± 0.031	0.595
d_S of Rh50	0.007 ± 0.005	0.049 ± 0.013	0.200 ± 0.028	0.620
d_N of Rh	0.007 ± 0.003	0.115 ± 0.011	0.098 ± 0.011	0.302
d_N of Rh50	0.001 ± 0.001	0.057 ± 0.008	0.058 ± 0.008	0.153

Note. MMD and MMB designate *M. m. domesticus* and *M. m. brevirostris,* respectively.
[a] Pairwise values with standard errors.
[b] Averages of pairwise values.

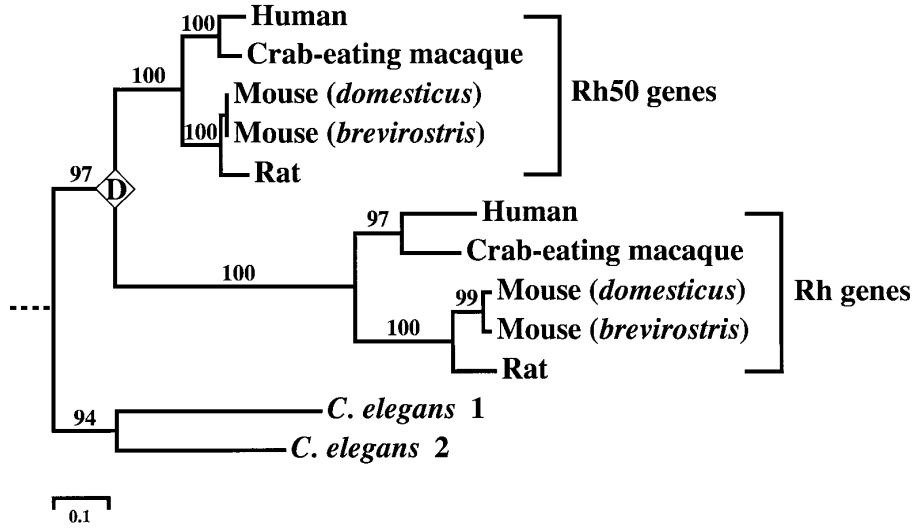

FIG. 3. The phylogenetic tree of mammalian Rh blood group genes and their homologous genes. Amino acid sequences of only membrane-spanning regions (216 sites) are used. Bootstrap probabilities (%) are shown on interior branches. The root was located by assuming the Rh-like protein of sponge as an outgroup. Accession numbers and gene IDs for *C. elegans* 1 and 2 are Z74026-B0240.1 and U64847-F08F3.3, respectively.

eage started to diverge from an ancestral reptilian lineage.

The Rh and its related genes have probably been existing as essential membrane proteins in many animal phyla. Because its evolutionary rate is lower than that for the Rh protein gene, the Rh50 protein may be closer to the ancestral form before gene duplication of Rh and Rh50 genes. In this context, it may be interesting to note a similarity between the Rh50 protein and NH_4^+ transporters (29). However, two coding regions of *C. elegans* (C05E11.4 in the DDBJ/EMBL/GenBank accession number U53338 and M195.3 in Z66498) are more similar to NH_4^+ transporters. Therefore, the actual function of Rh and Rh50 gene products remain to be found.

Window analyses of synonymous and nonsynonymous substitutions. We performed window analyses for synonymous (d_S) and nonsynonymous (d_N) nucleotide substitutions to investigate their possible correlation with the protein structure (Fig. 4). The twelve predicted hydrophobic membrane-spanning regions are shown by black boxes with numbers. There are several peaks (depicted by arrows) where nonsynonymous substitutions are higher than synonymous ones on putative outer membrane regions on primate Rh genes (Fig. 4A). One peak (designated as long arrows) is observed at the cell surface region between membrane-spanning regions 3 and 4 in all four comparisons, and four and two peaks were also observed for other cell surface regions in primate Rh and Rh50 comparisons, respectively (indicated by short arrows with asterisks). Whether an intermembrane region is on the cell surface or in the cytoplasm is based on the currently accepted model of the Rh and Rh50 protein structure (2, 30).

TABLE 3

Numbers (d_A) of Amino Acid Substitutions and Divergence Times

Diverging node	Human/ macaque	Mouse/ rat	Primates/ rodents	Rh/Rh50
Single lineage d_A of Rh[a]	0.120	0.072	0.228	0.663
Single lineage d_A of Rh50[a]	0.042	0.029	0.104	0.225
Divergence time (MYA)	23.3[b]	40.7[b]	112[b]	240–310[c]

[a] Based on phylogenetic tree of Fig. 3.
[b] Taken from Ref. 28.
[c] Estimated from d_A values.

We examined Rh blood group gene and its homologous Rh50 genes in the present study. Blood group antigens may play a key role in pathogenesis of diseases (31), and there is a possibility of positive Darwinian selection caused by interaction between organisms (host mammals and parasites such as bacteria). In fact, Saitou and Yamamoto (32) and Kitano and Saitou (unpublished) found evidences of positive selection in the ABO and Rh blood group genes of primates, respectively. Comparison of synonymous and nonsynonymous substitutions for the Rh50 gene also revealed a possibility of existence of positive selection for this gene in primates. Because primates showed more clear sign of positive selection than rodents both for Rh and Rh50 genes, it is possible that the pattern of host-parasite interaction is different between primates and rodents.

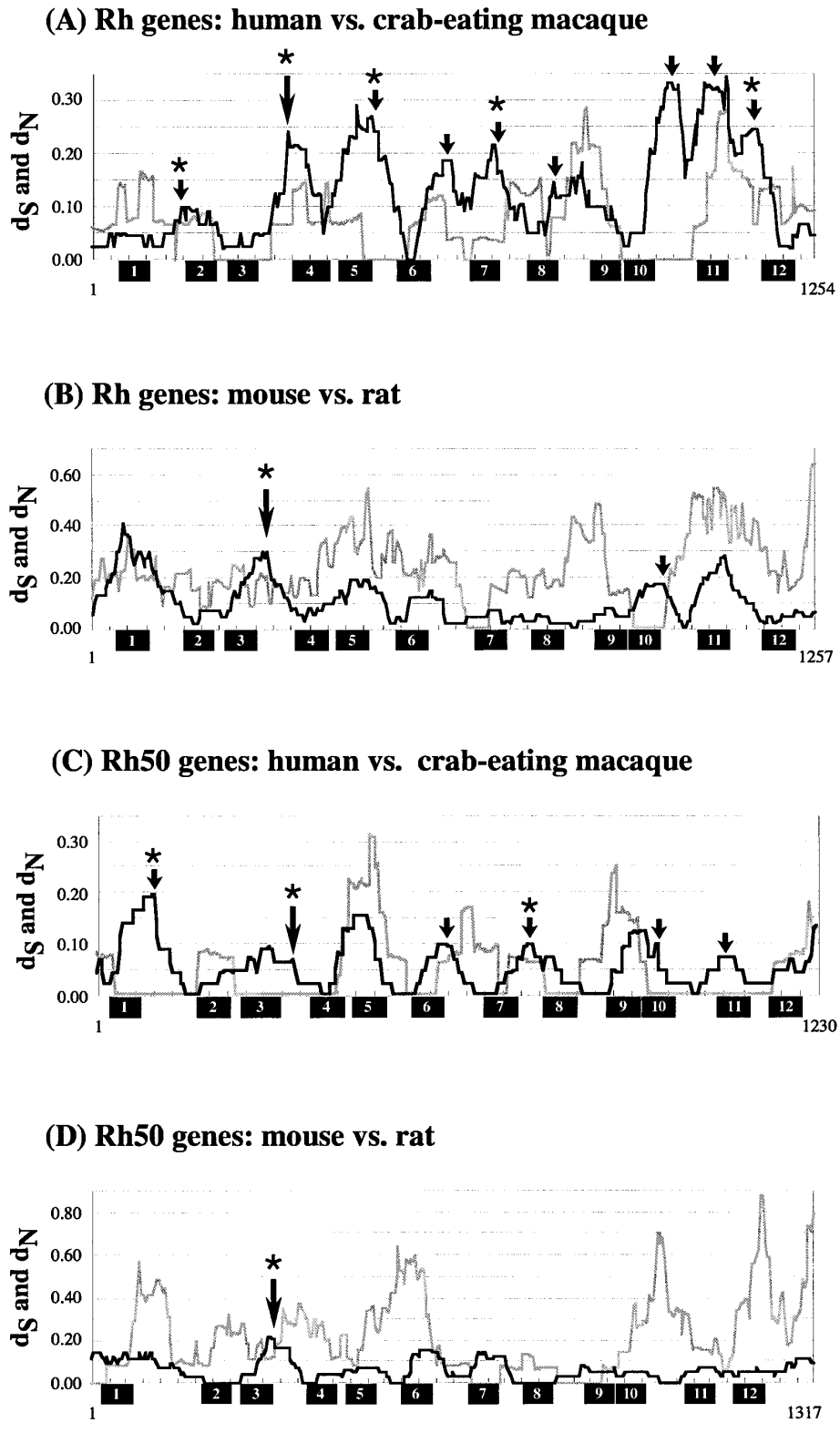

FIG. 4. Window analyses for synonymous (d_S: gray lines) and nonsynonymous (d_N: black lines) nucleotide substitutions for Rh genes between human and crab-eating macaque (A), for Rh genes between *M. m. domesticus* and rat (B), for Rh50 genes between human and crab-eating macaque (C), and for Rh50 genes between *M. m. domesticus* and rat (D). The 12 predicted hydrophobic membrane-spanning regions are shown by black boxes. Horizontal axes indicate numbers of nucleotide sites. Arrows indicate peaks of d_N higher than d_S, and arrows with asterisks are in the putative cell surface regions. Arrows in the same cell surface region for all the A-D comparisons are drawn longer than others.

ACKNOWLEDGMENTS

We thank Dr. Y. Ikehara for preparation of the crab-eating macaque tissues and Dr. T. Sagai for preparation of the mouse tissues. This study was partially supported by grants-in-aid for scientific studies from the Ministry of Education, Science, Sport, and Culture, Japan, to N. S.

REFERENCES

1. Cherif-Zahar, B., Bloy, C., Le Van Kim, C., Blanchard, D., Bailly, P., Hermand, P., Salmon, C., Cartron, J.-P., and Colin, Y. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 6243–6247.
2. Avent, N. D., Ridgwell, K., Tanner, M. J., and Anstee, D. J. (1990) *Biochem. J.* **271,** 821–825.
3. Le van Kim, C., Mouro, I., Cherif-Zahar, B., Raynal, V., Cherrier, C., Cartron, J.-P., and Colin, Y. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 10925–10929.
4. Kajii, E., Umenishi, F., Iwamoto, S., and Ikemoto, S. (1993) *Hum. Genet.* **91,** 157–162.
5. Arce, M., Thompson, E. S., Wagner, S., Coyne, K. E., Ferdman, B. A., and Lublin, D. M. (1993) *Blood* **82,** 651–655.
6. Avent, N. D., Butcher, S. K., Liu, W., Mawby, W. J., Mallinson, G., Parsons, S. F., Anstee, D. J., and Tanner, M. J. (1992) *J. Biol. Chem.* **267,** 15134–15139.
7. Salvignol, I., Blancher, A., Calvas, P., Clayton, J., Socha, W. W., Colin, Y., and Ruffie, J. (1994) *Biochem. Genet.* **32,** 201–221.
8. Salvignol, I., Calvas, P., Socha, W. W., Colin, Y., Le Van Kim, C., Bailly, P., Ruffie, J., Cartron, J.-P., and Blancher, A. (1995) *Immunogenetics* **41,** 271–281.
9. Mouro, I., Le Van Kim, C., Cherif-Zahar, B., Salvignol, I., Blancher, A., Cartron, J.-P., and Colin, Y. (1994) *J. Mol. Evol.* **38,** 169–176.
10. Moore, S., and Green, C. (1987) *Biochem. J.* **244,** 735–741.
11. Eyers, S. A. C., Ridgwell, K., Mawby, W. J., and Tanner, M. J. A. (1994) *J. Biol. Chem.* **269,** 6417–6423.
12. Ridgwell, K., Spurr, N. K., Laguda, B., MacGeoch, C., Avent, N. D., and Tanner, M. J. (1992) *Biochem. J.* **287,** 223–228.
13. Cherif-Zahar, B., Raynal, V., Gane, P., Mattei, M. G., Bailly, P., Gibbs, B., Colin, Y., and Cartron, J.-P. (1996) *Nature Genet.* **12,** 168–173.
14. Kawano, M., Iwamoto, S., Okuda, H., Fukuda, S., Hasegawa, N., and Kajii, E. (1998) *Ann. Hum. Genet.* **62,** 107–113.
15. Thompson, J. D., Gibson, T. J., and Higgins, D. G. (1994) *Nucleic Acids Res.* **22,** 4673–4680.
16. Saitou, N., and Nei, M. (1987) *Mol. Biol. Evol.* **4,** 406–425.
17. Kimura, M. (1983) The Neutral Theory of Molecular Evolution, Cambridge Univ. Press, Cambridge.
18. Ina, Y. (1994) *Comput. Appl. Biosci.* **10,** 11–12.
19. Nei, M., and Gojobori, T. (1986) *Mol. Biol. Evol.* **3,** 418–426.
20. Endo, T., Ikeo, K., and Gojobori, T. (1996) *Mol. Biol. Evol.* **13,** 685–690.
21. Matassi, G., Cherif-Zahar, B., Raynal, V., Rouger, P., and Cartron, J.-P. (1998) *Genomics* **47,** 286–293.
22. Ruddle, F., Ricciuti, F., McMorris, F. A., Tischfield, J., Creagan, R., Darlington, G., and Chen, T. (1972) *Science* **176,** 1429–1431.
23. Cherif-Zahar, B., Mattei, M. G., Le Van Kim, C., Bailly, P., Cartron, J.-P., and Colin, Y. (1991) *Hum. Genet.* **86,** 398–400.
24. Li, W.-H., and Tanimura, M. (1987) *Nature* **326,** 93–96.
25. Wilson, R., Ainscough, R., Anderson, K., *et al.* (53 co-authors) (1994) *Nature* **368,** 32–38.
26. Seack, J., Pancer, Z., Mueller, I. M., and Mueller, W. E. G. (1997) *Immunogenetics* **47,** 493–498.
27. Ishida, N., Oyunsuren, T., Mashima, S., Mukoyama, H., and Saitou, N. (1995) *J. Mol. Evol.* **41,** 180–188.
28 Kumar, S., and Hedges, S. B. (1998) *Nature* **392,** 917–920.
29. Marini, A. M., Urrestarazu, A., Beauwens, R., and Andre, B. (1997) *Trends Biochem. Sci.* **22,** 460–461.
30. Avent, N. D., Butcher, S. K., Lie, W., Mawby, W. J., Mallinson, G., Parsons, S. F., Anstee, D. J., and Tanner, M. J. A. (1992) *J. Biol. Chem.* **267,** 15134–15139.
31. Eder, A. F., and Spitalnik, S. L. (1997) *in* Molecular Biology and Evolution of Blood Group and MHC Antigens in Primates (Blancher, A., Klein, J., and Socha, W. W., Eds.), pp. 268–304, Springer.
32. Saitou, N., and Yamamoto, F. (1997) *Mol. Biol. Evol.* **14,** 399–411.