# No distinction of orthology/paralogy between human and chimpanzee Rh blood group genes

Takashi Kitano<sup>1,\*</sup>, Choong-Gon Kim<sup>1,#</sup>, Antoine Blancher<sup>2</sup>, and Naruya Saitou<sup>1</sup>

<sup>1</sup>Division of Population Genetics, National Institute of Genetics, Mishima, Japan

<sup>2</sup>Laboratoire d'Immunoge'ne'tique Mole'culaire (LIMT, EA3034), Faculte' de Me'decine

Purpan, Universite' Paul Sabatier, Toulouse III, France

\*Current address: Department of Biomolecular Functional Engineering, College of Engineering, Ibaraki University, Hitachi, Japan

<sup>#</sup>Current address: Marine Ecosystem Research Division, Korea Institute of Ocean Science and Technology, Ansan, Korea

Author for Correspondence: Takashi Kitano, Department of Biomolecular Functional Engineering, College of Engineering, Ibaraki University, 4-12-1 Nakanarusawa-cho, Hitachi 316-8511, Japan, Tel: +81-294-38-5050, Fax: +81-294-38-5078, E-mail: tkitano@mx.ibaraki.ac.jp

DNA sequences reported in this study are available at the DDBJ/EMBL/GenBank International Nucleotide Sequence Database with the following accession numbers: LC032056-LC032100.

© The Author(s) 2016. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

On human (Homo sapiens) chromosome 1, there is a tandem duplication encompassing Rh blood group genes (Hosa\_RHD and Hosa\_RHCE). This duplication occurred in the common ancestor of humans, chimpanzees (Pan troglodytes), and gorillas, after splitting from their common ancestor with orangutans. Although several studies have been conducted on ape Rh blood group genes, the clear genome structures of the gene clusters remain unknown. Here, we determined the genome structure of the gene cluster of chimpanzee Rh genes by sequencing five BAC clones derived from chimpanzees. We characterized three complete loci (Patr RHa, Patr RHb, and Patr RHy). In the Patr RHb locus, a short version of the gene, which lacked the middle part containing exons 4–8, was observed. The Patr RH $\alpha$  and Patr RHß genes were located on the locations corresponding to Hosa RHD and Hosa RHCE, respectively, and Patr\_RHy was in the immediate vicinity of Patr\_RH<sub>\beta</sub>. Sequence comparisons revealed high sequence similarity between Patr RHB and Hosa RHCE, while the chimpanzee Rh gene closest to Hosa RHD was not Patr RHa but rather Patr RHy. The results suggest that rearrangements and gene conversions frequently occurred between these genes and that the classic orthology/paralogy dichotomy no longer holds between human and chimpanzee Rh blood group genes.

Keywords: gene duplication, phylogeny, rearrangement

#### Introduction

The Rh blood group is one of the major blood group systems in humans, and bears important practical considerations in blood transfusions and clinical medicine. The nucleotide sequences of the Rh blood group genes were determined independently by Cherif-Zahar et al. (1990) and Avent et al. (1990). This blood group system was shown to be composed of two closely linked RHD and RHCE loci (Mouro et al. 1993), as predicted by Tippett (1986). RHD and RHCE loci are located on chromosome 1p34-p36 in humans (Ruddle et al. 1972; Cherif-Zahar et al. 1991), and individuals are classified in practice as Rh-positive and Rh-negative according to the presence or absence of the D antigen. Nucleotide sequences of the Rh blood group genes in nonhuman primates have also been reported (Salvignol et al. 1994; 1995; Mouro et al. 1994). Genomic DNA analyses by Southern blot have shown that chimpanzees (Pan troglodytes) possesses three Rh loci (Salvignol et al. 1993, 1994). Gorillas carry two Rh loci (Apoil et al. 1999), whereas orangutans, gibbons, Old World monkeys, and New World monkeys carry a single Rh locus (Blancher et al. 1992). This suggests that the Rh blood group genes were duplicated in the common ancestor of humans, chimpanzees, and gorillas, after splitting from their common ancestor with orangutans.

Suto et al. (2000) found that the two human Rh loci have opposite orientations by using fiber-FISH analysis, a useful approach for analyzing genome sequences containing duplicated genes. Wagner and Flegel (2000) went on to report that the 3' ends of these genes

Downloaded from http://gbe.oxfordjournals.org/ at Idenkagaku Kenkyujo on February 16, 2016

face each other and are separated by a ~30-kb segment that contains the TMEM50A (transmembrane protein 50A) gene. In addition, Suto et al. (2003) later discovered the existence of two-Rh-gene and four-Rh-gene types in chimpanzees, again using fiber-FISH analysis. The two-Rh-gene type is arranged as Rh  $(5'\rightarrow 3')$  – Rh  $('3\leftarrow 5')$ , with the distance between the two genes about 50 kb longer than that in humans, while the arrangement of the four-Rh-gene type is Rh  $(5'\rightarrow 3')$  – Rh  $('3\leftarrow 5')$  – Rh  $('3\leftarrow 5')$  – Rh  $('3\leftarrow 5')$  within a region spanning about 300 kb. The fact that the Rh blood group genes have been duplicated in humans, chimpanzees, and gorillas suggests that their evolutionary history is complex (Kitano and Saitou 1999). As no authoritative chimpanzee reference genome is yet available in Ensembl, GenBank Map Viewer or UCSC genome browser (supplementary fig. S1B) with which to examine the gene cluster of chimpanzee Rh blood group genes, we examined a ~400-kb sequence region of the chimpanzee genome to carry out detailed analyses on the evolution of this gene cluster and its constituent genes.

#### **Materials and Methods**

#### Genomic DNA sequencing

Five chimpanzee BAC clones for the Rh blood group genes (PTB-081F20, PTB-149J04, RP43-004E10, RP43-024B01, and RP43-026A23) were screened by

BLAST-searches using BAC-end sequences (Fujiyama et al. 2002). The PTB and RP43 BAC clone libraries are from the chimpanzee individuals "Gon" and "Donald", respectively. "Gon" is Pan troglodytes verus (http://www.shigen.nig.ac.jp/pgdb2/greatApe.html?individualId=338), while "Donald", a second-generation captive, is a hybrid between *troglodytes* and the *verus* subspecies (Prado-Martinez et al. 2013). Nucleotide sequences were determined primarily using the shotgun method following Watanabe et al. (2004). The Phred/Phrap software package (Ewing et al. 1998) was used for base-calling and assembly and to obtain quality scores for assembled data. Editing to identify all low-quality bases and to check the correct assembly on the basis of linking information was performed using Consed (Gordon et al. 1998). Finishing was then performed by primer walking and PCR-coupled primer walking. We completely sequenced (finished) the two BAC clones (RP43-004E10 and RP43-026A23), while the remaining three (PTB-081F20, PTB-149J04, and RP43-024B01) were retained as ordered piece (i.e., phase 2) sequences.

#### Haplotypes of exon 7 sequences

We sequenced the region surrounding exon 7 (approximately 860 bp) from 12 chimpanzees (*Pan troglodytes verus*), three bonobos (*Pan paniscus*), three gorillas (*Gorilla gorilla*), one orangutan (*Pongo pygmaeus*), and two siamangs (*Symphalangus syndactylus*). Five ("Akina", "Aluku", "Chiko", "Kanao", and "Mikota") of the 12 chimpanzees were the

Analyzer (Applied Biosystems).

same individuals used in the Suto et al. (2003) fiber-FISH analysis. Two primers on intron 6 (forward: 5'-TTT GCA TAT GTG TCC ACA TCT G-3') and on intron 7 (reverse: 5'-GCC AAC AAA TAT TCA CCG AAG C-3') were used for each PCR experiment and sequencing. KOD DNA polymerase (TOYOBO), which has 3'→5' exonuclease activity, was used for all PCR experiments to ensure high fidelity. Following PCR, one extension reaction (10 min, 72°C) using Ex Taq (TaKaRa) was carried out with one nucleotide addition for TA-cloning. PCR products were confirmed by 1.5% agarose gel electrophoresis, purified using Micro Spin Columns (Amersham Biosciences), and inserted into a vector using the TOPO TA Cloning Kit (Invitrogen). At least two clones for each sequence type from each individual were sequenced using a BigDye Terminator Cycle Sequencing Kit and an ABI PRISM 3100 DNA Genetic

To confirm the presence of Patr\_RHβS in each individual, two primers on exon 3 (forward: 5'-GTG CTG GTG GAG GTG ACA GCT TTA-3') and intron 8 (reverse: 5'-CAT CCA CCT CCT GCT TAG GGA TAC C-3'), able to amplify the ~650-bp product from Patr\_RHβS, were used. This primer set cannot amplify products from complete Rh blood group genes obtained by ordinary PCR because the expected size reaches ~28 kbp.

The five sequence datasets for "Clint" available in GenBank (NW\_003481311.1, NW\_003481846.1, NW\_003460690.1, NW\_003456511.1, and NW\_003460689.1) were used.

#### Sequence analyses

Multiple alignments were performed using CLUSTAL W (Thompson et al. 1994) implemented in the MEGA5 software (Tamura et al. 2011). To select substitution models, the Find Best DNA/Protein Models (ML) option in the MEGA5 software package was used for nucleotide sequence data. Phylogenetic trees were inferred by using the Maximum Likelihood method (Felsenstein 1981) with nearest-neighbor-interchange heuristic search from the initial neighbor-joining tree (Saitou and Nei 1987). Phylogenetic networks were constructed using the Neighbor-Net method (Bryant and Moulton 2004) with p-distance implemented in the SplitsTree4 software (Huson and Bryant 2006). The number of synonymous and nonsynonymous substitutions (p-distance) was estimated following Nei and Gojobori (1986).

#### **Results and Discussion**

Genome sequence of the chimpanzee Rh blood group genes

BAC clones containing Rh genes were characterized from the two chimpanzees "Gon" and "Donald". By analyzing the BAC sequences, we reconstructed the three chromosomes (RP43-1, RP43-2, and PTB-1) shown in **fig. 1**. Chromosomes RP43-1 and RP43-2 were characterized in a single individual ("Donald"), and differ by 1,403 out of 167,165 bp (0.008). RP43-1 consists of two completely sequenced BAC clones

(RP43-026A23 and RP43-004E10) that are strictly identical in their overlapping region (3,801 bp), which encompasses the 3' end of TMEM50A. The BAC clone (RP43-024B01) defines chromosome RP43-2. PTB-1 consists of two BAC clones (PTB-149J04 and PTB-081F20) which overlap from a *Rhesus box* (Wagner and Flegel 2000; a DNA segment of approximately 9 kb that contains a nearly complete remnant of a transposon-like human element, THE-1B) on the 5' end of Patr\_RH $\alpha$ 3 to exon 1 of Patr\_RH $\beta$ S3. These two BAC clones were derived from the same individual ("Gon") and differ only by four nucleotide changes and two length differences consisting of 4- and 2-base repeats throughout the 137,844-bp overlap region. Because the nucleotide difference between these clones (4 / 137,844 = 2.9×10<sup>-5</sup>) is too small to assume they originate from different chromosomes of the same individual, we conclude that they most probably derive from the same chromosome, and that the observed differences were most likely caused by mutations during DNA replication in bacterial cells.

We characterized three Rh loci (Patr\_RHα, Patr\_RHβ, and Patr\_RHγ) in the gene cluster of chimpanzee Rh blood group genes (**fig. 1**). The Patr\_RHα locus was located between C1orf63 and TMEM50A, while Patr\_RHβ and Patr\_RHγ loci were located between TMEM50A and TMEM57. In the Patr\_RHβ locus, a short version of the gene, which lacked the middle part containing exons 4–8, was observed in PTB-1; we refer to this as Patr\_RHβS (Patr\_RHβS3). In the human genome sequence data (GRCh38.p2, Release 107), Hosa\_RHD (NG\_007494.1) and Hosa\_RHCE (NG\_009208.3) loci are located between C1orf63 and

TMEM50A and between TMEM50A and TMEM57, respectively. Thus, the location of the Patr\_RH $\alpha$  locus corresponds to that of Hosa\_RHD, while Patr\_RH $\beta$  (and Patr\_RH $\gamma$ ) has the same location as Hosa\_RHCE. Two *Rhesus boxes* are located between the 5' and 3' ends of the Patr\_RH $\alpha$  locus, similar to Hosa\_RHD. In addition, chimpanzees were found to have another *Rhesus box*, between Patr\_RH $\beta$  and Patr\_RH $\gamma$ .

As mentioned before, Suto et al. (2003) reported the existence of two-Rh-gene and four-Rh-gene types in chimpanzees. It is likely that PTB-1 (fig. 1) is a two-Rh-gene type. Because Suto et al. used two genomic DNA probes that contained introns 3 and 7 in the fiber-FISH analysis, they did not detect Patr RHBS, but instead observed that the two-Rh-gene type of chimpanzees have a longer interval (approximately 120 kb) than in humans and gorillas. Another interesting difference between our results is that whereas they did not observe any three-Rh-gene types from five chimpanzee individuals, we did discovered one: RP43-1 (fig. 1). This incongruity could be caused by subspecies differences. All samples examined by Suto et al. were from Pan troglodytes verus; in contrast, "Donald", in whom the three-Rh-gene type was detected, is a hybrid between troglodytes and the verus subspecies (Prado-Martinez et al. 2013). Some studies have reported that some functional genes are highly diversified among subspecies of chimpanzees because of region-specific natural selection (Hvilsom et al. 2008; Hayakawa et al. 2012; Groeneveld et al. 2012; Ferreira et al. 2013). The uniqueness of RP43-1 could derive from the *troglodytes* subspecies.

#### Sequence differences between human and chimpanzee Rh blood group genes

To examine the overall sequence differences between the chimpanzee Rh blood group genes, we constructed a phylogenetic tree using genome sequences (~62 kb in length from start to stop codons) of five chimpanzee Rh blood group genes (Patr\_RHa1, Patr\_RHa3, Patr RH $\beta$ 1, Patr RH $\gamma$ 1, and Patr RH $\gamma$ 2; fig. 1) and the human genes Hosa RHD and Hosa\_RHCE (fig. 2). The GTR+G+I model was selected as the best by the model selection option in MEGA5. Initially, Patr\_RHa1 and Patr\_RHa3, and Patr\_RHy1 and Patr\_RHy2 were formed two clusters, both with bootstrap values of 100%. Moreover, Hosa\_RHD and Hosa\_RHCE inexplicably formed a cluster with a bootstrap value of 100%. When we constructed a phylogenetic network (supplementary fig. S2A), an alternative branch supporting a Patr\_RH\beta1-Hosa\_RHCE cluster and an alternative branch supporting a Patr\_RHy1-Patr\_RHy2-Hosa\_RHD cluster were observed. These alternative findings demonstrate the advantages of using phylogenetic networks that can show second-best alternative clustering using reticulations. Patr\_RHB and Hosa\_RHCE might be true orthologues because they share the same genomic locations (fig. 1). In contrast, the genomic locations of Patr\_RHy and Hosa\_RHD differ. Although Patr\_RHa is located on the genomic location corresponding to Hosa\_RHD's, no branching signal supports their clustering, even in the phylogenetic network. Supplementary fig. S2B shows a phylogenetic network of amino

acid sequences using the same genes as supplementary fig. S2A. Similar phylogenetic relationships were observed in this phylogenetic network. These inconsistencies between phylogenetic relationships and their genomic locations are likely caused by sequence rearrangements such as gene conversions and recombination, as suggested by Kitano and Saitou (1999).

> To analyze the distributions of phylogenetically informative sites, we first searched for sites that differed between the Hosa\_RHD and Hosa\_RHCE genomic sequences, and found 307 of them. Then, we checked whether each of the five chimpanzee genes (Patr RHa1, Patr RHa3, Patr RHb1, Patr RHy1, and Patr RHy3) had identical nucleotide sites with Hosa\_RHD or Hosa\_RHCE, and categorized these locations as shared sites. Furthermore, we defined sets of more than three contiguous shared sites as shared blocks (fig. 3). For example, sites 2178, 2180, 2184, 2193, 2203, and 2210 are shared sites between Patr RHa1 and Hosa RHCE, and they comprise one shared block between Patr RHa1 and Hosa RHCE. Patr RHa1 had 137 shared sites with Hosa RHD and 172 with Hosa RHCE, and Patr\_RHa3 had 132 shared sites with Hosa\_RHD and 177 with Hosa\_RHCE. In terms of shared blocks, Patr RHa1 had 16 with Hosa RHD and 19 with Hosa RHCE, and Patr RHa3 had 17 with Hosa\_RHD and 24 with Hosa\_RHCE. There is no clear distribution bias of shared sites and blocks between Patr\_RHa1/Patr\_RHa3 and Hosa\_RHD or between it and Hosa\_RHCE. Thus, it is not clear whether Patr\_RH $\alpha$ 1/Patr\_RH $\alpha$ 3 is phylogenetically

homologous to Hosa\_RHD or to Hosa\_RHCE. Patr\_RHB1 had 209 shared sites and 31 shared blocks with Hosa\_RHCE, respectively about 2 times as many and 2.8 times as many as with Hosa\_RHD (which had 100 shared sites and 11 shared blocks). The nucleotide sequence similarity between Patr\_RH\beta1 and Hosa\_RHCE is also observed in the phylogenetic network (supplementary fig. S2A) as the alternative clustering. Furthermore, Patr\_RH $\beta$  and Hosa RHCE share the same genomic locations (fig. 1). Thus, one can assume that Patr RH $\beta$ and Hosa\_RHCE are orthologous genes, and that some sequence rearrangements such as gene conversions and recombination have occurred on some genomic regions of Patr\_RHB and/or Hosa RHCE. Patr RHy1 had 228 shared sites and 28 shared blocks with Hosa RHD, respectively about 2.8 times as many and 5.6 times as many as with Hosa\_RHCE (which had 81 shared sites and 5 shared blocks). Similarly, Patr\_RHy2 had 230 shared sites and 24 shared blocks with Hosa\_RHD, respectively about 2.9 times as many and 8 times as many as with Hosa\_RHCE (which had 79 shared sites and 3 shared blocks). The nucleotide sequence similarities between Patr\_RHy1/Patr\_RHy2 and Hosa\_RHD are also observed in the phylogenetic network (supplementary fig. S2A). However, the genomic location of Patr RHy1 and Patr RHy2 differs from that of Hosa RHD. One interpretation is that in the chimpanzee lineage, a gene that had been located between C1orf63 and TMEM50A moved to the current Patr\_RHy genomic location and became Patr\_RHy. Alternatively, Hosa\_RHD moved to a location between C1orf63 and TMEM50A in the human lineage. When a

phylogenetic tree was constructed by adding two genome sequences of the RH genes obtained from the gorilla draft genome data (**supplementary fig. S1C**), the two gorilla genes formed a cluster with a bootstrap value of 99% and the two-gorilla-gene cluster formed a cluster with Patr\_RH $\beta$  with a bootstrap value of 93% (**supplementary fig. S3**). From the parsimonious standpoint, the former; that is a gene that had been located between C1orf63 and TMEM50A moved to the current Patr\_RH $\gamma$  genomic location and became Patr\_RH $\gamma$  in the chimpanzee lineage, might be possible. Since the two gorilla sequences are incomplete because of the draft genome data, it calls for further investigation. In either case, the result suggests the volatility of the genomic location of the genes in the region.

Apoil and Blancher (2000) reported that the length difference in intron 4 between Hosa\_RHD and Hosa\_RHCE (longer) in humans caused by a 645-bp segment that includes *Alu-Sx* is also observed in chimpanzees. We therefore decided to analyze the distribution of *Alu* in the chimpanzee Rh blood group genes from the five BAC clones (**supplementary table S1**). Our findings showed that none of the chimpanzee genes used in this study had the 645-bp segment including *Alu-Sx* in intron 4. However, this does not necessarily signify that the intron including *Alu-Sx* does not exist in other chimpanzees. Moreover, because we could not sequence the whole gene region of Patr\_RH $\delta$  in this study, it is possible that Patr\_RH $\delta$  has the 645-bp segment in which *Alu-Sx* is present in intron 4.

We confirmed that two partially sequenced genes (Patr\_RH\beta2 and Patr\_RH\gamma3)

region of Patr\_RHy.

Predicting the haplotypes of the chimpanzee Rh blood group genes

To determine the haplotypes of the gene cluster of chimpanzee Rh blood group genes, we compared the region surrounding exon 7 (~860 bp) between 15 chimpanzee individuals. This region was selected because exon 7 is the most divergent exon between the human Hosa\_RHD and Hosa\_RHCE loci (Westhoff and Wylie 1996; Innan 2003; Kitano et al. 2007). We sequenced 12 chimpanzees, three bonobos, three gorillas, one orangutan, and two siamangs. In addition, sequences from "Gon" (PTB-1), "Donald" (RP43-1 and RP43-2), and "Clint" (genome shotgun data in the database) were used. We then constructed a phylogenetic tree using the sequences for the region surrounding exon 7 obtained from humans and these non-human primate species (fig. 4). The Kimura 2-parameter (Kimura 1980) with a gamma distribution model was selected as the best by the model selection option in MEGA5. In chimpanzees and bonobos, four (Patr\_RHa, Patr\_RHB, Patr\_RHy, and Patr\_RHb) and three (Papa\_RHα, Papa\_RHβ, and Papa\_RHγ) clusters were observed, respectively. Patr\_RHα1 of

formed respective clusters with other Patr\_RH $\beta$  and Patr\_RH $\gamma$  genes (supplementary fig.

S4A and S4B). Since Patr\_RHBS3 formed a cluster with Patr\_RHB1 in the phylogenetic

network (supplementary fig. S4C), Patr RHBS might be a deletion type of Patr RHB,

instead of a combined type consisting of a downstream region of Patr\_RHB and an upstream

RP43-1 (fig. 1) did not form a cluster with other Patr\_RH $\alpha$  genes in the phylogenetic tree, but did form a cluster with Patr\_RH\delta genes (fig. 4) with a bootstrap value of 93%. Two Patr\_RHy genes (Patr RHy1 of RP43-1 and Patr RHy2 of RP43-2, fig. 1) did not form a cluster with other Patr\_RHy genes, but did form a cluster with Patr\_RHa genes (fig. 4). We assigned cluster names in the phylogenetic tree (fig. 4) by following Patr\_RH $\alpha$ 3 and Patr\_RH $\gamma$ 3 on the PTB-1 chromosome (fig. 1), since the RP43-1 chromosome came from a hybrid individual. Thus, the cluster of eight chimpanzee sequences including Patr RH $\alpha$ 3 and four bonobo sequences was designated as Patr\_RHa-Papa\_RHa, and the cluster of five chimpanzee sequences including Patr RHy3 as Patr RHy. A cluster containing five bonobo sequences (Papa RH $\gamma$ 1–5) was designated as Papa RH $\gamma$ , because although this cluster did not group together with Patr\_RHy, they were located close to each other in the phylogenetic tree. The cluster of five chimpanzee sequences including Patr\_RHB1 and Patr\_RHB2, and one bonobo sequence (Papa RH\beta1) was designated as Patr RH\beta-Papa RH\beta. Inexplicably, Patr RH\beta2 of RP43-2 (**fig. 1**) from the hybrid individual formed a cluster with Papa RHβ1 rather than with other Patr\_RHß genes. A remaining chimpanzee cluster including Patr\_RHa1 of RP43-1 was designated as Patr RHS. Assuming that four is the maximum number of loci in chimpanzees (Suto et al. 2003), the four clusters  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  might correspond to the chimpanzee Rh loci. Three clusters (Papa\_RH $\alpha$ , Papa\_RH $\beta$ , and Papa\_RH $\gamma$ ) for bonobos and two clusters  $(Gogo_RH\alpha \text{ and } Gogo_RH\beta)$  for gorillas were found in the phylogenetic tree (fig. 4),

suggesting the presence of three and two loci, respectively, for these species. These findings are compatible with the results of Suto et al. (2003).

From the results of the phylogenetic tree (fig. 4), we extrapolated the haplotypes of the Rh genes for various chimpanzee individuals (fig. 5). Individuals "Gon", "Mikota", "Akina", "Aluku", "Rimi", "Journey", and "Hitomi" had sequence types categorized only in Patr RH $\alpha$  and Patr RH $\gamma$ . We also confirmed the presence of Patr RH $\beta$ S, a short version of Patr RHB lacking the middle part (exons 4–8) of the gene, in these individuals using PCR. Fiber-FISH analysis (Suto et al. 2003) has shown that "Mikota", "Akina", and "Aluku" have two-Rh-gene types; the genome sequence PTB-1 determined in this study suggested that this is also true for "Gon". The genomic location of Patr RH\delta is still unclear, as its location beside Patr\_RH $\gamma$  appears arbitrary. The haplotype [ $\alpha$ 3 - S -  $\gamma$ 3] (which corresponds to PTB-1 in fig. 1) is most probably present in double dose in "Mikota" (fig. 5). "Chiko" and "Mary" have the haplotype  $[\alpha 3 - S - \gamma 3]$  as a single dose, but more complex haplotypes on their other chromosome: namely,  $[\alpha 4 - \beta 4 - \gamma 6 - \delta 2]$  and  $[\alpha 5 - \beta 6 - \gamma 3 - \delta 3]$ , respectively. The other individuals ("Akina", "Aluku", "Rimi", "Journey", and "Hitomi") most likely have the haplotype  $[\alpha 3 - S - \gamma 3]$  (which corresponds to PTB-1), and the Rh haplotypes of these individuals on other chromosomes were deduced without issue (fig. 5).

Synonymous and nonsynonymous substitutions in the chimpanzee Rh blood group genes

Both variations in the numbers of loci of the chimpanzee Rh blood group genes and frequent arrangements among the genes themselves may contribute to the extant diversity of chimpanzee Rh proteins. It has been suggested that the Rh blood group genes have experienced some type of positive selection (Kitano et al. 1998; Kitano and Saitou 1999; Innan 2003). Blood types were originally distinguished by different molecular structures on the cell surface of erythrocytes. Therefore, products of blood group genes may interact with other organisms such as bacteria and viruses to protect against them, leading to the possibility of positive selection on these genes. To attempt to detect positive selection, we estimated synonymous (dS) and nonsynonymous (dN) substitutions for each exon of the chimpanzee Rh blood group genes (fig. 6). Significant dN and dS differences were observed in exon 2 (higher dSthan dN), exon 5 (no dS at all), and exon 7 (higher dN than dS). The latter was particularly notable: sixteen sites of different amino acids were observed in exon 7. In exon 7 of humans, the D-specific motif of amino acid residues D-350, G-353, and A-354 is located on the sixth outer membrane loop of the RHD protein (Cartron and Agre 1993). These amino acid residues may be directly correlated with antigenic properties, because their position with respect to the membrane makes them accessible to antibodies (Blancher and Socha 1997). Thus, it is possible that this protein region may be involved in some kinds of host-parasite interactions. Interestingly, two (G-353 and A-354) of the D-specific motif were present in the Patr\_RHß locus, although the Patr\_RHβ locus did not share the same genomic location as Hosa\_RHCE.

This finding also supports the theory of partial rearrangements frequently reoccurring in the chimpanzee Rh blood group genes.

#### Acknowledgments

We are grateful to the Primate Park of Sanwa Kagaku Kennkyusho Co., Ltd. (Kumamoto, Japan) for providing chimpanzee samples. We thank Asao Fujiyama of the Comparative Genomics Laboratory, National Institute of Genetics for helping with preparation of BAC clones, and Yuji Kohara of the Genome Biology Laboratory, National Institute of Genetics for helping with sequencing. We thank Akiko Aoshima, Atsuko Ide, and Yoshimi Noaki for technical assistance. We thank anonymous reviewers and the editor for their valuable comments. This study was partially supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology, Japan to NS and TK, and a JSPS Grant-in-Aid for Scientific Research to TK.

#### References

Apoil, P.A., and Blancher, A. 2000. Rh gene evolution in primates: study of intron sequences.

Mol. Biol. Evol. 17: 127-136.

- Apoil, P.A., Roubinet, F., and Blancher, A. 1999. Gorilla RH-like genes and antigens. *Immunogenetics* **49**: 125-133.
- Avent, N.D., Ridgwell, K., Tanner, M.J., and Anstee, D.J. 1990. cDNA cloning of a 30 kDa erythrocyte membrane protein associated with Rh (Rhesus)-blood-group-antigen expression. *Biochem. J.* **271:** 821-825.
- Blancher, A., Calvas, P., and Ruffie, J. 1992. Study of equivalents of rhesus antigens in non-human primates. *C. R. Seances Soc. Biol. Fil.* **186:** 682-695.
- Blancher, A., and Socha, W.W. 1997. The Rhesus system. In: Blancher, A., Klein, J., Socha,
  W.W., editor. *Molecular biology and evolution of blood group and MHC antigens in primates*. Springer-Verlag, Heidelberg. p. 147–218.
- Bryant, D., and Moulton, V. 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol. Biol. Evol.* **21:** 255-265.
- Cartron, J.P., and 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, J.P., and 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, M.G., Le Van Kim, C., Bailly, P., Cartron, J.P., and 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.

- Ewing, B., Hillier, L., Wendl, M.C., and Green, P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8: 175-185.
- Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17: 368-376.
- Ferreira, Z., Hurle, B., Andrés, A.M., Kretzschmar, W.W., Mullikin, J.C., Cherukuri, P.F.,
  Cruz, P., Gonder, M.K., Stone, A.C., Tishkoff, S., Swanson, W.J.; NISC Comparative
  Sequencing Program, Green, E.D., Clark, A.G., Seixas, S. 2013. Sequence diversity of
  Pan troglodytes subspecies and the impact of WFDC6 selective constraints in
  reproductive immunity. *Genome Biol. Evol.* 5: 2512-2523.
- Fujiyama, A., Watanabe, H., Toyoda, A., Taylor, T.D., Itoh, T., Tsai, S.F., Park, H.S., Yaspo,
  M.L., Lehrach, H., Chen, Z. et al. 2002. Construction and analysis of a human-chimpanzee comparative clone map. *Science* 295: 131-134.
- Gordon, D., Abajian, C., and Green, P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res.* 8: 195-202.
- Groeneveld, L.F., Atencia, R., Garriga, R.M., Vigilant, L. 2012. High diversity at PRDM9 in chimpanzees and bonobos. *PLoS One* **7**: e39064.
- Hayakawa, T., Sugawara, T., Go, Y., Udono, T., Hirai, H., Imai, H. 2012. Eco-geographical diversification of bitter taste receptor genes (TAS2Rs) among subspecies of chimpanzees

(Pan troglodytes). PLoS One 7: e43277.

- Huson, D.H., and Bryant, D. 2006. Application of Phylogenetic Networks in Evolutionary Studies. *Mol. Biol. Evol.* 23: 254-267.
- Hvilsom, C., Carlsen, F., Siegismund, H.R., Corbet, S., Nerrienet, E., Fomsgaard, A. 2008.
  Genetic subspecies diversity of the chimpanzee CD4 virus-receptor gene. *Genomics* 92: 322-328.
- 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. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16:** 111-120.
- Kitano, T., and 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., and Saitou, N. 1998. Conserved evolution of the Rh50 gene compared to its homologous Rh blood group gene. *Biochem. Biophys. Res. Commun.* 249: 78-85.
- Kitano, T., Umetsu, K., Tian, W., Yamazaki, K., and Saitou, N. 2007. Tempo and mode of evolution of the Rh blood group genes before and after gene duplication. *Immunogenetics* 59: 427-431.

Mouro, I., Colin, Y., Cherif-Zahar, B., Cartron, J.P., and Le Van Kim, C. 1993. Molecular

genetic basis of the human Rhesus blood group system. Nature Genet. 5: 62-65.

Mouro, I., Le Van Kim, C., Cherif-Zahar, B., Salvignol, I., Blancher, A., Cartron, J.P., and 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., and Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3:** 418-426.
- Nei, M., and Kumar, S. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, New York. p. 52-57.
- Prado-Martinez, J., Sudmant, P.H., Kidd, J.M., Li, H., Kelley, J.L., Lorente-Galdos, B.,
  Veeramah, K.R., Woerner, A.E., O'Connor, T.D., Santpere, G., Cagan, A., Theunert, C.,
  Casals, F., Laayouni, H., Munch, K., Hobolth, A., Halager, A.E., Malig, M.,
  Hernandez-Rodriguez, J., Hernando-Herraez, I., Prüfer, K., Pybus, M., Johnstone, L.,
  Lachmann, M., Alkan, C., Twigg, D., Petit, N., Baker, C., Hormozdiari, F.,
  Fernandez-Callejo, M., Dabad, M., Wilson, M.L., Stevison, L., Camprubí, C., Carvalho,
  T., Ruiz-Herrera, A., Vives, L., Mele, M., Abello, T., Kondova, I., Bontrop, R.E., Pusey,
  A., Lankester, F., Kiyang, J.A., Bergl, R.A., Lonsdorf, E., Myers, S., Ventura, M.,
  Gagneux, P., Comas, D., Siegismund, H., Blanc, J., Agueda-Calpena, L., Gut, M., Fulton,
  L., Tishkoff, S.A., Mullikin, J.C., Wilson, R.K., Gut, I.G., Gonder, M.K., Ryder, O.A.,
  Hahn, B.H., Navarro, A., Akey, J.M., Bertranpetit, J., Reich, D., Mailund, T., Schierup,

M.H., Hvilsom, C., Andrés, A.M., Wall, J.D., Bustamante, C.D., Hammer, M.F., Eichler,
E.E., Marques-Bonet, T. 2013. Great ape genetic diversity and population history. *Nature*499: 471-475.

- Ruddle, F., Ricciuti, F., McMorris, F.A., Tischfield, J., Creagan, R., Darlington, G., and 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.
- Salvignol, I., Calvas, P., Socha, W.W., Colin, Y., Le Van Kim, C., Bailly, P., Ruffie, J., Cartron, J.P., and Blancher, A. 1995. Structural analysis of the RH-like blood group gene products in nonhuman primates. *Immunogenetics* **41**: 271-281.
- Salvignol, I., Blancher, A., Calvas, P., Clayton, J., Socha, W.W., Colin, Y., and 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., Blancher, A., Calvas, P., Socha, W.W., Colin, Y., Cartron, J.P., and Ruffie, J. 1993. Relationship between chimpanzee Rh-like genes and the R-C-E-F blood group system. *J. Med. Primatol.* **22**: 19-28.

Suto, Y., Ishikawa, Y., Hyodo, H., Ishida, T., Kasai, F., Tanoue, T., Hayasaka, I., Uchikawa,

M., Juji, T., and Hirai, M. 2003. Gene arrangement at the Rhesus blood group locus of chimpanzees detected by fiber-FISH. Cytogenet. Genome Res. 101: 161-165.

- Suto, Y., Ishikawa, Y., Hyodo, H., Uchikawa, M., and Juji, T. 2000. Gene organization and rearrangements at the human Rhesus blood group locus revealed by fiber-FISH analysis. Hum. Genet. 106: 164-171.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28: 2731-2739.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22: 4673-4680.
- Tippett, P. 1986. A speculative model for the Rh blood groups. Ann. Hum. Genet. 50: 241-247.
- Watanabe, H., Fujiyama, A., Hattori, M., Taylor, T.D., Toyoda, A., Kuroki, Y., Noguchi, H., BenKahla, A., Lehrach, H., Sudbrak, R., et al. 2004. DNA sequence and comparative analysis of chimpanzee chromosome 22. Nature 429: 382-388.
- Wagner, F.F., and Flegel, W.A. 2000. RHD gene deletion occurred in the Rhesus box. Blood **95:** 3662-3668.

### Westhoff, C.M., and Wylie, D.E. 1996. Investigation of the RH locus in gorillas and

chimpanzees. J. Mol. Evol. 42: 658-668.

#### **Figure legends**

**FIG. 1.** – Genome structure of the gene cluster of chimpanzee Rh blood group genes. Three chromosomes (RP43-1, RP43-2, and PTB-1) and three loci (Patr\_RH- $\alpha$ , Patr\_RH- $\beta$ , and Patr\_RH- $\gamma$ ) are shown. Vertical bars with numbers 1–10 show the exons for each Rh gene. Black boxes denote the SYF2, C1orf63, TMEM50A, and TMEM57 genes. Gray boxes represent *Rhesus boxes*, and an arrow in a box indicates its direction. The gap indicated by broken lines in PTB-1 is a large deletion that produced a short version of the Rh gene (Patr\_RH- $\beta$ S3) in chimpanzees. Light gray bars below the genome structure show the BAC clones for the sequencing. The genome structure of the human Rh blood group genes is provided for reference at the bottom of the figure. A scale bar (kb) is shown between the chimpanzee and human Rh blood group genes.

**FIG. 2.** – A maximum likelihood tree based on the GTR model constructed from genomic sequences for human and chimpanzee Rh blood group genes of **fig. 1**. Orangutan (Popy\_RH) sequence is used as an outgroup. A discrete gamma distribution was used to model evolutionary rate differences among sites (8 categories (+G, parameter = 0.4117)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 55.4082% sites). A scale bar representing nucleotide substitutions per site is shown. Bootstrap values (1000

replications) are shown on each branch. Black, gray, and horizontal-striped circles indicate Patr\_RH $\alpha$ , Patr\_RH $\beta$ , and Patr\_RH $\gamma$ , respectively. Black boxes indicate Hosa\_RHD and Hosa\_RHCE.

**FIG. 3.** – Distributions of phylogenetically informative sites for the chimpanzee Rh blood group genes (Patr\_RH $\alpha$ 1, Patr\_RH $\alpha$ 3, Patr\_RH $\beta$ 1, Patr\_RH $\gamma$ 1, and Patr\_RH $\gamma$ 2). Sites that differ between Hosa\_RHD and Hosa\_RHCE in their genome sequences are shown. Black boxes with numbers indicated exons. When a nucleotide of the chimpanzee Rh gene is identical to the corresponding Hosa\_RHD nucleotide, the site is shown as a red box, when identical to the corresponding Hosa\_RHCE nucleotide, it is shown as a blue box.

**FIG. 4.** – A maximum likelihood tree based on the Kimura 2-parameter model constructed using the region surrounding exon 7 obtained from humans and the listed primate species. Orangutan (Popy\_RH) and siamang (Sysy\_RH1 and Sysy\_RH2) sequences are used as outgroups. A discrete gamma distribution is used to model evolutionary rate differences among sites (8 categories (+G, parameter = 0.2057)). A scale bar representing nucleotide substitutions per site is shown. Bootstrap values (1000 replications) are shown on each branch. Sequences obtained from genomic data in **fig. 1** are noted in parentheses.

FIG. 5. – Predicted haplotypes of the chimpanzee Rh blood group genes based on comparisons of the region surrounding exon 7 (~860 bp) from 15 chimpanzee individuals. Patr\_RHa, Patr\_RH $\beta$ , Patr\_RH $\gamma$ , and Patr\_RH $\delta$  are shown as  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$ , respectively, with the relevant variant number. Patr\_RHBS is denoted by "S", when its presence in the individual was confirmed by PCR and by "s" when it was supposed but not confirmed through PCR experiments. Although the genome location of Patr RHS is unknown, it was placed beside Patr RHy for this analysis.

FIG. 6. - Distributions of nonsynonymous (dN) and synonymous (dS) substitutions for each exon of the chimpanzee Rh blood group genes. Proportions of dN and dS within each exon are indicated by grey and white bars with standard error bars, respectively. These standard errors were obtained using a bootstrap procedure (Nei and Kumer 2000). The dN and dS for exons 8, 9, and 10 were estimated together due to the short length of these exons.



## fig. 1



0.01





Gon	<mark>α3 - S - γ3</mark>	(PTB-1)
Mikota	α3 - S - γ3 α3 - S - γ3	
Akina	<mark>α3 - S - γ3</mark> α3 - S - γ6	
Aluku	<mark>α3 - S - γ3</mark> α3 - S - γ4	
Rimi	<mark>α3 - S - γ3</mark> α8 - S - γ3	
Journey	<mark>α3 - S - γ3</mark> α7 - S - γ3	
Hitomi	<mark>α3 - s - γ3</mark> α3 - s - γ5	
Chiko	<mark>α3 - S - γ3</mark> α4 - β4 - γ6 ·	- δ2
Mary	<mark>α3 - S - γ3</mark> α5 - β6 - γ3 ·	- δ3
Clint	<mark>α7 - S - γ3</mark> α4 - β4 - γ3	- δ2
Kanao	<mark>α3 - S - γ6</mark> α4 - β6 - γ7 -	- δ2
Niko	α <mark>3 - s - γ6</mark> - γ7 -	δ2
Reiko	- β2 - γ2 - γ6	
Mayumi	α6 - β5 -	- δ2
Donald	α1 - β1 - γ1 - β2 - γ2	(RP43-1) (RP43-2)

