Genetics and Molecular Biology, 27, 3, 342-349 (2004)
Copyright by the Brazilian Society of Genetics. Printed in Brazil
www.sbg.org.br

Research Article

Divergent evolution and purifying selection of the H (FUT1) gene in New World monkeys (Primates, Platyrrhini)

Bárbara do Nascimento Borges and Maria Lúcia Harada

Universidade Federal do Pará, Centro de Ciências Biológicas, Departamento de Genética, Laboratório de Biologia Molecular "Francisco Mauro Salzano", Belém, Pará, Brazil.

Abstract

In the present study, the coding region of the H gene was sequenced and analyzed in fourteen genera of New World primates (*Alouatta*, *Aotus*, *Ateles*, *Brachyteles*, *Cacajao*, *Callicebus*, *Callithrix*, *Cebus*, *Chiropotes*, *Lagothrix*, *Leontopithecus, Pithecia, Saguinus*, and *Saimiri*), in order to investigate the evolution of the gene. The analyses revealed that this coding region contains 1,101 nucleotides, with the exception of *Brachyteles*, the callitrichines (*Callithrix*, *Leontopithecus*, and *Saguinus*) and one species of *Callicebus* (*moloch*), in which one codon was deleted. In the primates studied, the high GC content (63%), the nonrandom distribution of codons and the low evolution rate of the gene (0.513 substitutions/site/MA in the order Primates) suggest the action of a purifying type of selective pressure, confirmed by the Z-test. Our analyses did not identify mutations equivalent to those responsible for the H-deficient phenotypes found in humans, nor any other alteration that might explain the lack of expression of the gene in the erythrocytes of Neotropical monkeys. The phylogenetic trees obtained for the H gene and the distance matrix data suggest the occurrence of divergent evolution in the primates.

Key words: H gene, molecular evolution, purifying selection, divergent evolution, New World monkeys.

Received: March 3, 2003; Accepted: February 16, 2004.

Introduction

The human H antigen is synthesized through the action of two α1,2 fucosyltransferases (α 1,2 FT) encoded by two distinct genes, H (FUT1) and Se (FUT2), both tissue- and stage-specific. The H enzyme is responsible for the expression of the H antigen in tissues derived from mesoderm and ectoderm, such as erythrocytes and vascular endothelial cells, and the Se enzyme for the expression of the same antigen in tissues derived from endoderm, such as epithelial cells of the digestive tract and exocrine glands (Watkins, 1980; Oriol *et al.*, 1981, 1986). The H and Se genes, and the Sec1 pseudogene constitute a cluster on the long arm of human chromosome 19 (19q13.3), where H is telomeric (Ball *et al.*, 1991; Rouquier *et al.*, 1994, 1995; Reguigne-Arnould *et al.*, 1995, 1996).

The H human gene consists of eight exons and three introns. The last exon is a coding one, whereas the others have a regulatory function (Koda *et al.*, 1997, 1998). Its product is a typical type II transmembrane protein, with 365 amino acids (Larsen *et al.,* 1990).

Send correspondence to Bárbara do Nascimento Borges. Universidade Federal do Pará, Centro de Ciências Biológicas, Departamento de Genética, Laboratório de Biologia Molecular "Francisco Mauro Salzano", R. Augusto Correa 01, Belém, Pará, Brazil. E-mail: harada@ufpa.br.

All nonhuman primates studied to date present the H antigen in their secretions, although only the great apes present it on the surface of erythrocytes and in the vascular endothelial cells (Ruffié, 1974; Socha *et al.*, 1984; Blancher and Socha, 1997). Few studies of the structure of this gene in New World primates (Platyrrhini) have been conducted. Apoil *et al.* (2000) isolated a fragment corresponding to the H gene in *Saimiri* and *Callithrix*, which potentially encodes an enzyme with 365 amino acids, equivalent to the human enzyme. However, these authors did not find mutations that might explain the lack of expression of the gene in the erythrocytes of platyrrhines and prosimians, and suggested that the insertion of an *Alu*-Y sequence in the promoter region of the H gene, in intron 1, has led to red cell expression in men and great apes.

In this study, we describe the complete sequence of the coding region of the H gene in New World primates, the occurrence of selection and the type of evolution affecting it, in order to clarify the evolutionary pathway followed by FUT1.

Materials and Methods

Samples used

Thirty-five samples from fourteen genera of New World primates were investigated (Table 1). All the sam-

Table 1 - Samples used in the present study, their respective codes, access numbers, number of base pairs sequenced and origin, where known.

Species	Code	Access number	Origin[1]	Base pairs sequenced
Alouatta belzebul	Ab1	AY219625	Rio Grande do Sul, Brazil	865
Alouatta belzebul	Ab2	AY219626	Tocantins River, Pará, Brazil	870
Alouatta belzebul	Ab3	AY219627	Tocantins River, Pará, Brazil	1101
Alouatta caraya	Ac1	AY219623	Goiás, Brazil	861
Alouatta caraya	Ac2	AY219624	Goiás, Brazil	1101
Alouatta seniculus	Ase	AY219628	Jarí River, Amapá, Brazil	862
Aotus azarae	Aaz	AY219622	Jamari River, Rondônia, Brazil	1101
Aotus nancymai	Ana	AY219621	Leticia, Colombia	1101
Ateles belzebuth	Abm	AY219641	CENP, Brazil	1101
Ateles paniscus	Apa	AY219640	Trombetas River, Pará, Brazil	876
Brachyteles arachnoides	Bar	AY219642	CPRJ, Brazil	1002
Cacajao melanocephalus	Cme	AY219650	Unknown	1101
Callicebus brunneus	Cbr	AY219643	Jamari River, Rondônia, Brazil	1101
Callicebus moloch	Cmo	AY219645	Tocantins River, Pará, Brazil	1098
Callicebus personatus	Cpe	AY219646	CPRJ, Brazil	869
Callicebus torquatus	Cto	AY219644	CENP, Brazil	1101
Callithrix emiliae	Cem	AY219630	Jamari River, Rondônia, Brazil	859
Callithrix humeralifer	Chu	AY219629	Unknown	1098
Callithrix jacchus	Cja1	AY219631	Rio Grande do Norte, Brazil	857
Cebus apella	Caa	AY219620	Jamari River, Rondônia, Brazil	1101
Cebus olivaceus	Cni	AY219619	Rio Grande do Sul, Brazil	857
Chiropotes satanas	Cs1	AY219647	Unknown	868
Chiropotes satanas	Cs2	AY219648	Tocantins River, Pará, Brazil	1101
Lagothrix lagotricha	Lla	AY219639	Unknown	1101
Leontopithecus chrysomelas	Lch	AY219632	Unknown	1098
Pithecia irrorata	Pir	AY219649	Jamari River, Rondônia, Brazil	1101
Saguinus bicolor ochraceus	Sbo	AY219638	Amazonas, Brazil	856
Saguinus fuscicollis weddelli	Sf1	AY219635	Jamari River, Rondônia, Brazil	851
Saguinus fuscicollis weddelli	Sf2	AY219636	Jamari River, Rondônia, Brazil	1098
Saguinus midas niger	Sm1	AY219633	Tocantins River, Pará, Brazil	858
Saguinus midas niger	Sm2	AY219634	Tocantins River, Pará, Brazil	854
Saguinus mystax	Smy	AY219637	CPRJ, Brazil	856
Saimiri boliviensis boliviensis	Sbb	AY219618	Bolívia	1101
Saimiri sciureus macrodon	Ss2	AY219616	Iquitos, Peru	1101
Saimiri sciureus macrodon	Ss3	AY219617	Iquitos, Peru	1101

1CENP = National Primate Center; CPRJ: Rio de Janeiro Primatology Center.

ples were obtained from the primate sample bank of the Genetics Department of the Federal University of Pará.

Isolation, amplification and sequencing of genomic DNA

DNA was isolated according to the phenolic extraction protocol (Sambrook *et al.,* 1989). The coding exon was amplified using the external primers F1U-103 (5' TTCGCCTTTCCTCCCCTGCA 3') and F1L-1264 (5' TGAAGCCACGTACTGCTGGC 3'), described by Yu *et al.* (1997). We constructed the internal primers F1U-714 (5'CAACAGCGCCTACCTCCG 3') and F1L-1005 (5' TGTGTGAGCAGGGCAAAGTC 3'), based on the sequences described by Apoil *et al.* (2000). PCR was performed using 1x reaction buffer, 100 ng of DNA, 0.4 mM of each primer, 0.03 U/µL of Taq DNA polymerase, 1.4 mM of $MgCl_2$, 0.1 µg/µL of BSA and 10.0 mM of each dNTP. Thirty-five amplification cycles were carried out under the following conditions: 94 °C for 50 s, 60 °C for 50 s and 72 °C for 90 s. Amplified fragments were purified

with the *Wizard® PCR Preps* kit (Promega) and sequenced by the dideoxyterminal method (Sanger *et al.*, 1977) using the *Big Dye Cycle Sequencing Standard* kit and an *ABI 377* (Perkin Elmer) automatic sequencer. Fragments were sequenced in both strands in order to clarify ambiguous sites. A site was interpreted as polymorphic only when double peaks were exactly overlapping.

Analysis of the nucleotide sequences

The following sequences from the literature were included in the database: *Homo sapiens* (Hsa - NM000148), *Pan troglodytes* (Ptr - AF080603), *Pongo pygmaeus* (Ppy - AF111935), *Gorilla gorilla* (Gog - AF080605), *Hylobates lar* (Hla - AF045545), *Cercopithecus aethiops* (Cae - D87932), *Macaca fascicularis* (Mfas - AF112474), *Macaca mulatta* (Mmul - AF080607), *Eulemur fulvus* (Eul - AF045546), *Callithrix jacchus* (Cjalit - AF111936) and *Saimiri sciureus* (Ss1 - AF136647). Alignment was carried out using XESEE, version 3.0 (Cabot and Beckenbach, 1989). All sequences obtained were analyzed. MEGA, version 2.1 (Kumar *et al.,* 2001) was used to compute transition (Ti) and transversion (Tv) rates among nucleotides, relative nucleotide and amino acid frequencies, the codon usage, the number of synonymous (dS) and nonsynonymous (dN) substitutions per site, and the test for selection. The modified Nei-Gojobori method (Zhang *et al*., 1998) was used to estimate dN and dS, while selection was tested using the Z-test (codon-based), validated by the *t*-test (Zhang *et al.*, 1997). The saturation test (transitions and transversions over divergence) was carried out by DAMBE, version 4.0.75 (Xia and Xie, 2001), using Kimura's distance (Kimura, 1980).

Phylogenetic analyses were carried out by PAUP, version 4.0b10 (Swofford, 1998), using maximum parsimony, with heuristic search and random addition of taxa, and neighbor-joining methods. The confidence level of each node was verified using the bootstrap method, with 2000 replications. Neighbor-joining analysis and divergence matrix were developed according to the evolutionary model and parameters of Modeltest, version 3.06 (Posada and Crandall, 1998). The evolution rate was calculated by r = K/2T (Li, 1997), where "r" was the rate of nucleotide substitution, "K" was the genetic distance between two or more taxa and "T" was the divergence time between taxa, in years. The divergence times used were those proposed by Goodman *et al.* (1998). The relative rate test was developed as proposed by Takezaki *et al*. (1995).

Results and Discussion

The coding region of the H gene was completely sequenced in 18 samples, including at least one representative of each of the fourteen genera studied, except *Brachyteles*, for which the coding region was partly sequenced (1,002 base pairs). The other samples (17 in total), which included the majority of the genera, were only partly sequenced (Table 1).

In most samples, the region alignment presented 1,101 nucleotides. As compared to the human sequence, the Old World monkeys and all Neotropical monkeys except *Callicebus moloch* and *Brachyteles* present an insertion of three base pairs (bp) between positions 67 and 69 (codon 23), which corresponds to that described by Apoil *et al.* (2000) for *Hylobates* and humans. The 3-bp deletion at positions 139 to 141 (codon 47), described by Apoil *et al*. (2000) in *Callithrix*, was also found in the other callitrichines (*Saguinus* and *Leontopithecus*). So, in most platyrrhines, H encodes a protein with 366 amino acids, one more than in humans, in which codon 23 of the transmembrane domain was lost. Like humans and *Hylobates*, the callitrichines, *Callicebus moloch* and *Brachyteles* have lost a codon, resulting in a protein with 365 amino acids.

A predominance of pyrimidine bases (55.5%) over purine bases (44.5%) was found. The mean GC content was 63%, and the GC:AT ratio was 1.70. This proportion is similar to that found by Epstein *et al.* (2000) in the phosphofructokinase gene and in other genes related to the development of tumors, which exhibit great gene expression activity. On the other hand, the GC content was higher than that described by Kitano *et al.* (1998) for the Rh and Rh50 genes (between 45% and 55%), which also encode surface antigens.

A highly heterogeneous distribution of amino acids was observed, ranging from a minimum of 1.65% (lysine) to a maximum of 12.03% (leucine). Similarly, specific codons were predominant in a given amino acid family, containing GC third bases.

As compared to human sequences, the mean similarity of amino acids in platyrrhines was 90.83%, an intermediate value between those found by Apoil *et al*. (2000) for the Old World monkeys (96.0%) and the prosimians (87.0%), and the mean similarity between the nucleotide sequences of the platyrrhine H gene and the human H, Se and Sec1 genes was 91.2%, 67.28%, and 65.71%, respectively. The high level of similarity is certainly due to the conservation of the coding region of the gene along primate evolution. This becomes even more evident when we consider the similarity between the human protein and those of other vertebrates, such as rats, rabbits and pigs (80%), and mice (75%) (Piau *et al.*, 1994; Hitoshi *et al.*, 1995, 1999; Cohney *et al.*, 1996; Domino *et al.*, 1997).

All platyrrhines shared 23 exclusive mutations**,** in addition to those shared with other nonhuman primates. A number of alterations exclusive to different platyrrhine genera and families were also identified. As compared to the human sequence, the H gene of the platyrrhines presented two amino acid alterations in the transmembrane domain and one in the second conserved motif (Figure 1). *Aotus* and *S. b. ochraceus* were the only taxa which presented more than one amino acid substitution in the second

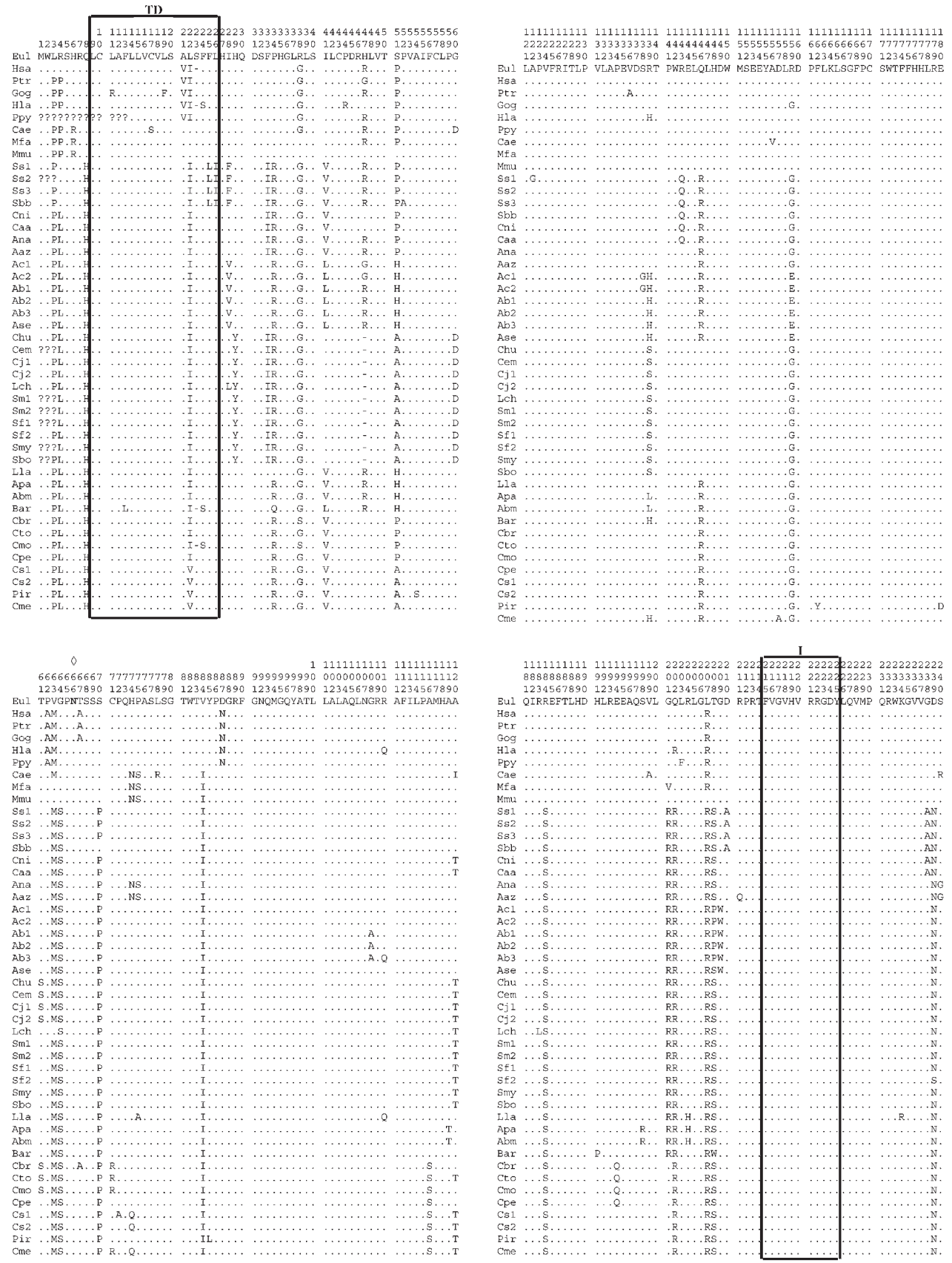

Figure 1 - Amino acid sequence of FUT1 gene. The lozenges (◊) indicate the glycosylation sites described by Larsen *et al.* (1990) and Apoil *et al.* (2000); the box indicates the position of the transmembrane domain (TD) and of the three conserved motifs described by Oriol *et al*. (1999).

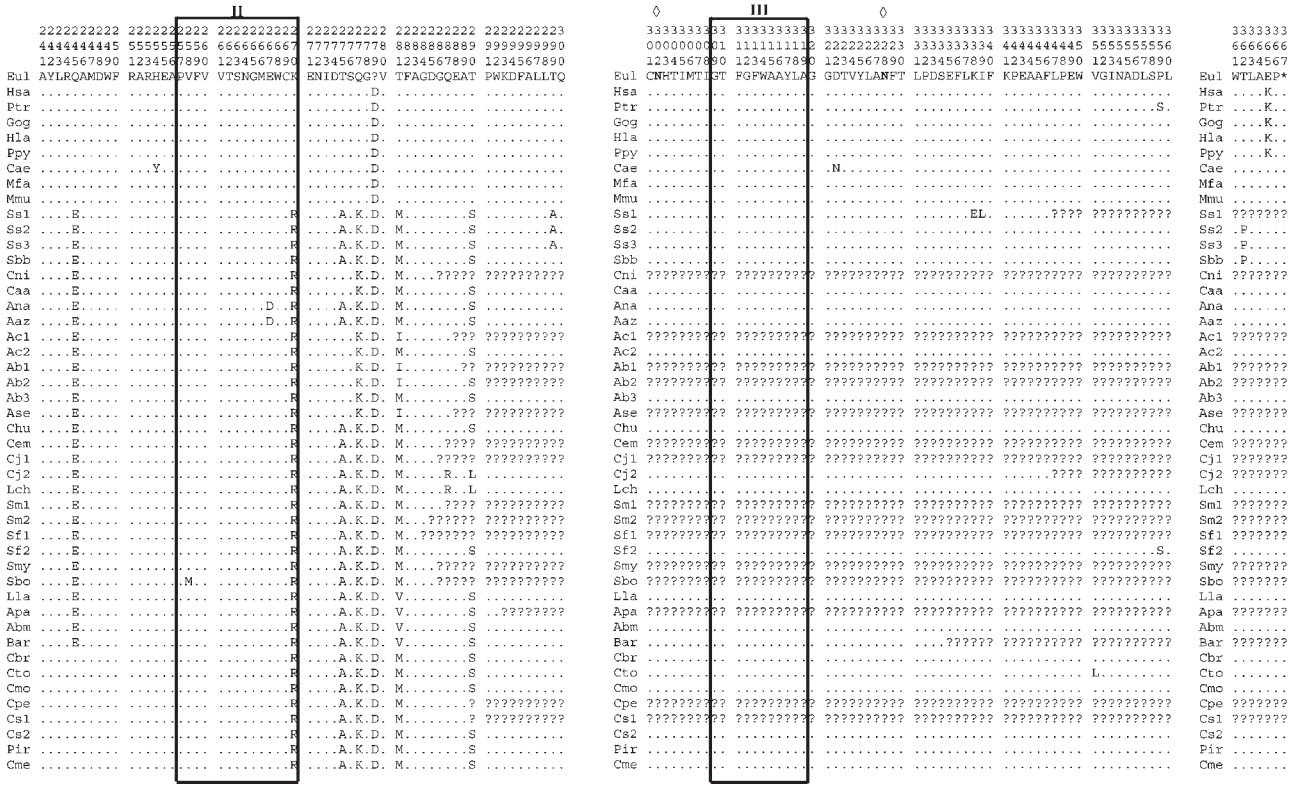

Figure 1 (cont.) - Amino acid sequence of FUT1 gene. The lozenges (◊) indicate the glycosylation sites described by Larsen *et al.* (1990) and Apoil *et al.* (2000); the box indicates the position of the transmembrane domain (TD) and of the three conserved motifs described by Oriol *et al*. (1999).

protein motif. None of the mutations was equivalent to those responsible for the H-deficient phenotypes found in humans, nor were any other alterations identified that might explain the lack of expression of the gene in the erythrocytes of Neotropical monkeys. These data and the high GC content reflect the functionality of the H gene in the platyrrhines, supported by the absence of mutations at the glycosylation sites and by the conservation of the amino acid serine (position 5 of the cytoplasmatic tail) and of the C-terminal region of the gene, where the majority of the nonsense and frameshift mutations occur (Wagner *et al.*, 2001). The glycosylation sites must be intact for full enzyme activity, and the amino acid Ser^5 is crucial for locating the enzyme in the Golgi complex (Christensen *et al.*, 2000; Milland *et al.*, 2001). These findings support the hypothesis of Apoil *et al.* (2000) that a factor outside the coding region of the gene is responsible for its expression in red blood cells and that this event is recent in the evolution of the gene.

Nucleotide substitution patterns revealed a predominance of transitions over transversions, with a mean ratio of 2.095. In pairwise comparisons, there was a slight predominance of C ⇔ T transitions, which were 1.5 times more frequent than A ⇔ G. There was a clear predominance (2.5 times) of C ⇔ G transversions.

To calculate dS and dN, the primates were divided into seven groups: the order Primates, apes and humans, Old World monkeys, New World primates, and the families Atelidae, Pitheciidae and Cebidae (according to Goodman *et al*., 1998). Synonymous substitutions were clearly predominant, even when considering the three distinct parts of the coding region separately: the N-terminal region (NT), the transmembrane domain (TD) and the C-terminal region (CT). In fact, only in the TD region of Cebidae dN is higher than dS (Table 2), although not significantly according to the *t*-test (ts < 1.65 and p > 0.05 in the one-tailed test). In the platyrrhines the dN:dS ratio was 0.34, a value similar to that found in primates by Zhang (2000) for 47 genes with moderate evolution rates (mean dN:dS ratio of 0.28), suggesting the occurrence of purifying selection. To test this hypothesis, we applied a Z-test, which resulted in a value equal to zero, thus rejecting the null hypothesis (dS = dN) and accepting the alternative hypothesis (dN < dS). This result is different from that of Kitano *et al*. (1998), who found more nonsynonymous substituitions than synonymous substitutions in Rh blood group genes of primates, which is clear evidence of positive selection. This could be due to an interaction between organisms (parasites) and host mammal blood group antigens in the case of the Rh blood groups, which does not occur in the H antigen (Kitano *et al*., 1998).

The estimated evolution rate of H in the platyrrhines was 0.462×10^{-9} substitutions/site/year (Table 3). In the families, rates varied between 0.384×10^{-9} in Cebidae and

Table 2 - Rates of synonymous (dS) and nonsynonymous (dN) substitutions per site of the H gene for the different primate groups analyzed.

	dS - dN for:			
Group[1]	Exon 8	N-terminal region (NT)	Transmembrane Domain (TD)	C-terminal region (CT)
Primates	0.110-0.029	0.080-0.065	0.053-0.035	0.115-0.028
HOM	0.023-0.008	0.115-0.057	0.037-0.036	0.022-0.006
OWM	0.020-0.014	0.000-0.000	0.038-0.019	0.019-0.014
NWM	0.061-0.021	0.042-0.014	0.022-0.023	0.065-0.021
Family Atelidae	0.035-0.014	0.000-0.000	0.030-0.012	0.034-0.015
Family Pitheciidae	0.039-0 010	0.000-0.000	0.028-0.025	0.041-0.010
Family Cebidae	0.058-0.014	0.056-0.028	0.000-0.021	0.063-0.014

[1]HOM = Apes and humans; OWM = Old World monkeys; NWM = New World monkeys.

0.668×10^{-9} in Atelidae. These results indicate, according to the relative rate test, that the H gene is evolving at a constant rate, both in the platyrrhines and in the order Primates. These values are in agreement with the suggestion of Barreaud *et al.* (2000) and Bureau *et al.* (2001), who observed that the evolution rate of H was intermediate among the α1,2 FT genes. This finding supports the hypothesis of selective pressure, given that selection, structural and functional requirements are the main factors which determine the evolution rate of a protein (Duret and Mouchiroud, 2000; Tourasse and Li, 2000).

Because of its broad expression, the H gene should be under relatively high selective pressure, given that the type of protein it encodes tends to be more conserved than tissue-specific ones (Hastings, 1996). If a protein, or part of a protein, has strict structural or functional requirements, the coding gene must be under strong selective pressure, which limits the alterations in the gene product. As a consequence, this gene will evolve more slowly, which explains why certain functionally critical regions, such as catalytic sites or ligation domains, are better conserved in the molecule (Tourasse and Li, 2000). This characteristic is clearly apparent in the present study, where the number of synonymous substitutions was almost four times greater than that of nonsynonymous ones, especially in the conserved motifs of the α1,2 FT protein. This corroborates the theory of Duret and Mouchiroud (2000), who suggested that a reduction in dN could be related to an increase in selective pressure on the amino acid sequence of the protein. No evidence of saturation was found for the H gene in any of the Primates analyzed (Figure 2).

Maximum parsimony analyses resulted in a single most parsimonious tree, that was identical with that obtained by neighbor-joining analysis, with similar bootstrap values. Therefore, we present only the parsimony arrangement (Figure 3). The phylogenetic relationships are similar to those proposed for New World monkeys (Schneider *et al*., 1996; Goodman *et al*., 1998; Schneider, 2000).

The genetic distance matrix shows low divergence rates, with the highest intrageneric value found in *Callicebus* (1.23%), and the highest intergeneric value for Pir x Ss1 (5.56%). The considerable overall similarities reinforce once again the highly conserved nature of the sequences.

The results of the genetic distance matrix, which shows low substitution rates, the agreement between the gene tree and the proposed phylogeny of New World monkeys (Schneider *et al*., 1996; Goodman *et al*., 1998; Schneider, 2000), the absence of saturation and the common nucleotide alterations shared by all Neotropical primates

Table 3 - Estimates of the evolution rates (r) of the H gene in the different groups of primates analyzed in the present study. Primate groups as in Table 2. Divergence times (T) are given in millions of years, according to Goodman *et al*. (1998).

Group	K	T	2T	Substitutions/site/year (r)
Family Cebidae	0.0169	22	44	0.384×10^{-9}
Family Atelidae	0.0214	16	32	0.668×10^{-9}
Family Pitheciidae	0.0202	17	34	0.593×10^{-9}
NWM	0.0369	25	50	0.739×10^{-9}
OWM	0.0160	14	28	0.571×10^{-9}
HOM	0.0126	18	36	0.350×10^{-9}
Primates	0.0646	63	126	0.513×10^{-9}

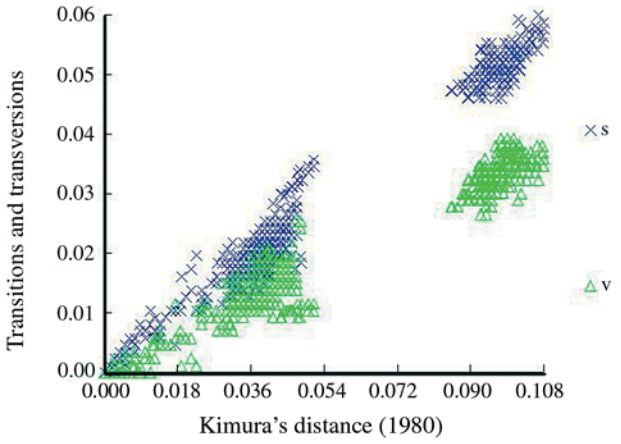

Figure 2 - Plot of the saturation test of the H gene. The graphic shows the absence of saturation in the studied samples.

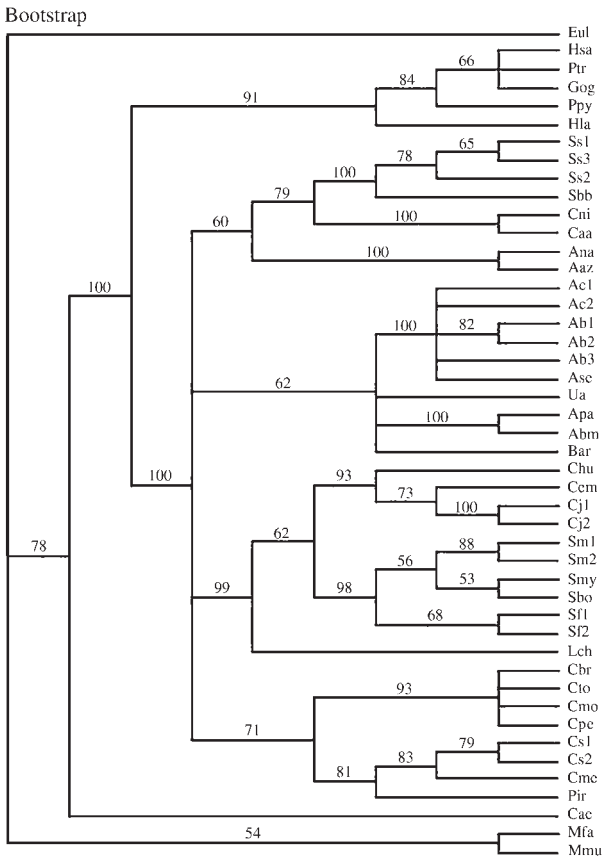

Figure 3 - Maximum parsimony tree (L = 396) obtained for the H gene of 45 primate samples. The numbers above each node are the bootstrap values for 2000 replications. See text and Table 1 for the identification of the specimens.

suggest the occurrence of divergent evolution of the H gene in New World monkeys. This idea is supported by the mean similarity values of the platyrrhine nucleotide sequences and the human H genes, which are higher than those found between Se and Sec1 genes, and the presence of the three previously described conserved motifs (Figure 1), shared by all α1,2 fucosyltransferases studied so far (Breton *et al.*, 1998; Oriol *et al.*, 1999). The divergent evolution model proposed here is in agreement with the evolution model proposed by other studies that suggest that the common shared motifs represent evidence that the α1,2 fucosyltransferases, the H gene included, have a common genetic origin by duplication events, followed by divergent evolution of the species (Breton *et al.*, 1998; Oriol *et al.*, 1999; Barreaud *et al.*, 2000; Bureau *et al*., 2001).

Acknowledgments

We would like to thank Dr. Evonnildo Gonçalves and MSc. Wilsea Figueiredo for their help with the statistical analyses, Dr. Stephen F. Ferrari for critical review of the manuscript and Arlindo Pinto de Souza Junior for technical support. B.N.B was supported by CAPES. This study was supported by grants from Universidade Federal do Pará (UFPA) and CNPq.

References

Apoil PA, Roubinet F, Despiau S, Mollicone R, Oriol R and Blancher A (2000) Evolution of α2-fucosyltransferases genes in primates: Relation between an intronic *Alu*-Y element and red cell expression of ABH antigens. Mol Biol Evol 17:337-351.

Ball SP, Tongue N, Gibaud A, Le Pendu J, Mollicone R, Gerard G and Oriol R (1991) The human chromosome 19 linkage group FUT1 (H), FUT2 (Se), LE, Lu, PEPD, C3, APOC2, DS19S7 and DS19S9. Ann Hum Genet 55:225-233.

Barreaud JP, Saunier K, Souchaire J, Delourme D, Oulmouden A, Oriol R, Levéziel H, Julien R and Petit JM (2000) Three bovine α2-fucosyltransferase genes encode enzymes that preferentially transfer fucose on Galβ1-3GalNAc acceptor substrates. Glycobiology 10:611-621.

Blancher A and Socha WW (1997) The ABO, Hh and Lewis blood group in humans and nonhuman primates. In: Blancher AA, Klein J and Socha WW (eds) Molecular Biology and Evolution of Blood Group and MHC Antigens in Primates. Springer-Verlag, Berlin, pp 30-92.

Breton C, Oriol R and Imberty A (1998) Conserved structural features in eukaryotic and prokaryotic fucosyltransferases. Glycobiology 8:87-94.

Bureau V, Marionneau S, Caileau-Thomas A, Le Moullac-Vaidye B, Liehr T and Le Pendu J (2001) Comparison of the three rat GDP-L-fucose: β-D-galactoside 2-α-L-fucosyltransferases FTA, FTB and FTC. Eur J Biochem 268:1006-1019.

Cabot ET and Beckenbach AT (1989) Simultaneous editing of multiple nucleic acid and protein sequence with ESSE. Comp Applic Biosc 5:233-234.

Christensen LL, Jensen UB, Bross P and Ørntoft TF (2000) The C-terminal N-glycosylation sites of the human α1,3/4- fucosyltransferase III, -V, and -VI (hFucTIII, -V and -VI) are necessary for the expression of full enzyme activity. Glycobiology 10:931-939.

Cohney S, Mouhtouris E, Mckenzie IFC and Sandrin MS (1996) Molecular cloning of the gene coding for pig α1 → 2 fucosyltransferase. Immunogenetics 44:76-79.

Domino SE, Hiraiwa N and Lowe JB (1997) Molecular cloning, chromosomal assignment and tissue-specific expression of a murine α(1,2)fucosyltransferase expressed in thymic and epididymal epithelial cells. Biochem J 327:105-115.

Duret L and Mouchiroud D (2000) Determinants of substitution rates in mammalian genes: Expression pattern affects selection intensity but not mutation rate. Mol Biol Evol 17:68-74.

Epstein RJ, Lin K and Tan TW (2000) A functional significance for codon third bases. Gene 245:291-298.

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

Hastings KEM (1996) Strong evolutionary conservation of broadly expressed protein isoforms in the tropoin I gene family and other vertebrate gene families. J Mol Evol 42:631-640.

Hitoshi S, Kusunoki S, Kanazawa I and Tsuji S (1995) Molecular cloning and expression of two types of rabbit β-galactoside α1,2 fucosyltransferase. J Biol Chem 270:8844-8850.

Hitoshi S, Kusunoki S, Kanazawa I and Tsuji S (1999) Dorsal root ganglia neuron-specific promoter activity of the rabbit β-galactoside α1,2-fucosyltransferase gene. J Biol Chem 274:389-396.

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, 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.

Koda Y, Soejima M and Kimura H (1997) Structure and expression of H-type GDP-L-fucose: β-D-galactoside 2-α-L-fucosyltransferase gene (FUT1). Two transcription start sites and alternative splicing generate several forms of FUT1 mRNA. J Biol Chem 272:7501-7505.

Koda Y, Soejima M and Kimura H (1998) Changing transcription start sites in H-type α(1,2) fucosyltransferase gene (FUT1) during differentiation of the human erythroid lineage. Eur J Biochem 256:379-387.

Kumar S, Tamura K, Jakobsen IB and Nei M (2001) MEGA 2: Molecular evolutionary genetics analysis software. Bioinformatics 17:1244-1245.

Larsen RD, Ernst LK, Nair RP and Lowe JB (1990) Molecular cloning, sequence and expression of a human GDP-L-fucose: β-D-galactoside 2-α-L-fucosyltransferase cDNA that can form the H blood group antigen. Proc Natl Acad Sci USA 87:6674-6678.

Li WH (1997) Molecular Evolution. 2nd edition. Sinauer Associates Inc, Canada, 487 pp.

Milland J, Taylor SG, Dodson HC, Mckenzie IFC and Sandrin MS (2001) The cytoplasmatic tail of α1,2-fucosyltransferase contains a sequence for Golgi localization. J Biol Chem 276:12012-12018.

Oriol R, Danilovs J and Hawkins BR (1981) A new genetic model proposing that the *Se* gene is a structural gene closely linked to the *H* gene. Am J Hum Genet 33:421-431.

Oriol R, Le Pendu J and Mollicone R (1986) Genetics of ABO, H, Lewis X and related antigens. Vox Sang 51:161-171.

Oriol R, Mollicone R, Cailleau A, Balanzino L and Breton C (1999) Divergent evolution of fucosyltransferase genes from vertebrates, invertebrates and bacteria. Glycobiology 9:323-334.

Piau JP, Labarriere N, Dabouis G and Denis MG (1994) Evidence for two distinct α(1,2) fucosyltransferase genes differentially expressed throughout the rat colon. Biochem J 300:623-626.

Posada D and Crandall KA (1998) MODELTEST: Testing the model of DNA substitution. Bioinformatics 14:817-818.

Reguigne-Arnould I, Couillin P, Mollicone R, Faué S, Fletcher A, Kelly RJ, Lowe JB and Oriol R (1995) Relative positions of two clusters of human α-L-fucosyltransferases in 19q (FUT1-FUT2) and 19p (FUT6-FUT3-FUT5) within the microsatellite genetic map of chromosome 19. Cytogenet Cell Genet 72:158-162.

Reguigne-Arnould I, Faure S, Chery M, Mota-Vieira L, Mollicone R, Candelier JJ, Oriol R and Couillin P (1996) Physical mapping of 49 microsatellite markers on chromosome 19 and correlation with the genetic linkage map. Genomics 32:458-461.

Rouquier S, Giorgi D, Bergman A, Brandriff B and Lennon G (1994) The fucosyltransferase locus FUT1 maps distal to apolipoprotein loci E and C2 on human chromosome 19. Cytogenet Cell Genet 66:70-71.

Rouquier S, Lowe JB, Kelly RJ, Fertitta AL, Lennon GG and Giorgi D (1995) Molecular cloning of a human genomic region containing the H blood group α(1,2) fucosyltransferase gene and two H locus-related DNA restriction fragments. Isolation of a candidate for the human secretor blood group locus. J Biol Chem 270:4632-4639.

Ruffié J (1974) Immunogenetics of primates. Adv Behav Biol 9:177-200.

Sambrook J, Fritsch EF and Maniatis T (1989) Molecular Cloning: A Laboratory Manual. 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sanger F, Nichlen S and Coulson AR (1977) DNA sequencing with chain-terminator inhibitors. Proc Natl Acad Sci USA 74:5463-5468.

Schneider H (2000) The current status of the New World monkey phylogeny. An Acad Bras Cienc 72:165-172.

Schneider H, Sampaio I, Harada ML, Barroso CM, Schneider MP, Czelusniak J and Goodman M (1996) Molecular phylogeny of the New World monkeys (Platyrrhini, primates) based on two unlinked nuclear genes: IRBP intron 1 and epsilon-globin sequences. Am J Phys Anthropol 100:153-179.

Socha WW, Moor-Jankowski J and Ruffié J (1984) Blood groups of primates: Present studies, theoretical implications and pratical applications: A review. J Med Primatol 13:11-40.

Swofford DL (1998) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Takezaki N, Rzhetsky A and Nei M (1995) Phylogenetic test of the molecular clock and linearized trees. Mol Biol Evol 12:823-833.

Tourasse NJ and Li WH (2000) Selective constraints, amino acid composition, and the rate of protein evolution. Mol Biol Evol 17:656-664.

Wagner T, Vadon M, Staudacher E, Schmarda A, Gassner C, Helmberg W, Lanzer G, Flegel WA and Wagner FF (2001) A new h allele detected in Europe has a missense mutation in (1,2)-fucosyltransferase motif II. Transfusion 41:31-38.

Watkins WM (1980) Biochemistry and genetics of the ABO, Lewis and P blood group systems. Adv Hum Genet 10:1-136.

Xia X and Xie Z (2001) DAMBE: Software package for data analysis in molecular biology and evolution. J Hered 92:371-373.

Yu LC, Yang YH, Broadberry RE, Chen YH and Lin M (1997) Heterogeneity of the human H blood group α(1,2) fucosyltransferase gene among para-bombay individuals. Vox Sang 72:36-40.

Zhang J (2000) Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. J Mol Evol 50:56-68.

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

Zhang J, Rosenberger HF and Nei M (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc Natl Acad Sci USA 95:3708-3713.

Associate Editor: Sérgio Furtado dos Reis