Evolution of α2-Fucosyltransferase Genes in Primates: Relation Between an Intronic *Alu*-Y Element and Red Cell Expression of ABH Antigens

Pol-André Apoil, Francis Roubinet,* Stéphanie Despiau,* Rosella Mollicone,† Rafael Oriol,† and Antoine Blancher**

*Laboratoire d'Immunogénétique Moléculaire, Hôpital Purpan, Toulouse, France; and †Institut National de la Santé et de la Recherche Midicale U504, Glycobiology, Université de Paris Sud (XI), Villejuif, France

Coding sequences of the paralogous *FUT1* (*H*), *FUT2* (*Se*), and *Sec1* α2-fucosyltransferase genes were obtained from different primate species. Analysis of the primate *FUT1*-like and *FUT2*-like sequences revealed the absence of the known human inactivating mutations giving rise to the *h* null alleles of *FUT1* and the *se* null alleles of *FUT2.* Therefore, most primate *FUT1*-like and *FUT2*-like genes potentially code for functional enzymes. The *Sec1*-like gene encodes for a potentially functional α2-fucosyltransferase enzyme in nonprimate mammals, New World monkeys, and Old World monkeys, but it has been inactivated by a nonsense mutation at codon 325 in the ancestor of humans and African apes (gorillas, chimpanzees). Human and gorilla *Sec1*'s have, in addition, two deletions and one insertion, respectively, 5′ of the nonsense mutation leading to proteins shorter than chimpanzee Sec1. Phylogenetic analysis of the available H, Se, and Sec1 mammalian protein sequences demonstrates the existence of three clusters which correspond to the three genes. This suggests that the differentiation of the three genes is rather old and predates the great mammalian radiation. The phylogenetic analysis also suggests that *Sec1* has a higher evolutionary rate than *FUT2* and *FUT1.* Finally, we show that an *Alu*-Y element was inserted in intron 1 of the *FUT1* ancestor of humans and apes (chimpanzees, gorillas, orangutans, and gibbons); this *Alu*-Y element has not been found in monkeys or nonprimate mammals, which lack ABH antigens on red cells. A potential mechanism leading to the red cell expression of the H enzyme in primates, related to the insertion of this *Alu*-Y sequence, is proposed.

Introduction

Antigens of ABO and Lewis systems are oligosaccharides synthesized by the sequential action of glycosyltransferases. These enzymes add specific carbohydrate residues onto precursor lactosamine sugar chains or onto the H blood group antigen. In humans, this H antigen can be synthesized by two α2-fucosyltransferases (EC 2.4.1.69), named H and Se and encoded by *FUT1* and *FUT2* genes, respectively (Oriol, Danilovs, and Hawkins 1981). These H and Se enzymes possess a high degree of homology to each other (67% identity), but they differ in substrate specificities and tissue expression (Costache et al. 1997*b*). The H enzyme (encoded by *FUT1*) is responsible for the expression of H antigen in the red cell lineage and vascular endothelium, whereas the Se enzyme (encoded by *FUT2*) is responsible for the synthesis of H antigen in secretory epithelia of individuals referred to as ''secretors'' because they have soluble ABH substances in exocrine secretions (Oriol et al. 1992). The secretors possess at least one functional *FUT2* allele (*Se*). About 20% of European individuals display a ''nonsecretor'' phenotype and are homozygous for *FUT2* null alleles (*se*/*se*). Various types of *se* null alleles have been identified in human populations (Kelly et al. 1995; Henry et al. 1996*a*, 1996*b*; Koda et al. 1996; Yu et al. 1996, 1999; Liu et al. 1998). Contrary to the relatively high frequency of null alleles at the human *FUT2* locus (*se*), the null alleles at the *FUT1* locus (*h*) are rare. When present in double doses, they are responsible for the absence of H antigen at the surfaces of red blood cells (*Bombay* and *para-Bombay* phenotypes). Twenty-three different inactivating mutations in *h* null alleles of *FUT1* have been characterized. Each one has a very low frequency and either introduces a stop codon, alters the reading frame, or induces a single amino acid replacement in the catalytic domain (Johnson et al. 1994; Kelly et al. 1994; Kaneko et al. 1997; Koda et al. 1997*a*; Wagner and Flegel 1997; Wang et al. 1997; Yu et al. 1997; Fernandez-Mateos et al. 1998).

Human *FUT1*, *FUT2*, and *Sec1* constitute a gene cluster located in 19q13.3 (Reguigne-Arnould et al. 1995). *FUT1* and *FUT2* genes are separated by 35.5 kb, and *FUT1* is in a telomeric position. The *Sec1* human pseudogene belonging to the α2-fucosyltransferase gene family is at 12 kb and centromeric from *FUT2* (Rouquier et al. 1995).

The human H and Se enzymes are anchored in the Golgi membrane by only 25 (H) or 28 (Se) N-terminal amino acid residues, which compose the intracytoplasmic and transmembrane segments and are not indispensable for the activity of the enzyme (Larsen et al. 1990; Kelly et al. 1995). This *Sec1* gene is characterized by two adjacent deletions responsible for a frameshift and a premature stop codon. This truncated Sec1 polypeptide exhibits no α2-fucosyltransferase activity when tested by transfection/expression in COS cells (Kelly et al. 1995). Therefore, it is a pseudogene, although its promoter is functional because Sec1 transcripts are present in human tumor cell lines (Koda et al. 1997*b*).

Among nonhuman primates, only anthropoid apes (chimpanzees, gorillas, orangutans, and gibbons) ex-

Abbreviations: H, *FUT1* encoded human α2-fucosyltransferase; PCR, polymerase chain reaction; Se, *FUT2* encoded human α2-fucosyltransferase; *Sec1,* human pseudogene of the α2-fucosyltransferase family; USF, upstream stimulation factor.

Key words: α2-fucosyltransferase, primate, evolution, *Alu,* ABO blood group, H, Se.

Address for correspondence and reprints: Antoine Blancher, Laboratoire d'Immunogénétique Moléculaire, Université Paul Sabatier, Pavillon Charles Lefebvre, Hôpital Purpan, 31059 Toulouse Cedex, France. E-mail: blancher@easynet.fr.

Mol. Biol. Evol. 17(3):337–351. 2000

Table 1
PCR Primers Used for Amplification of Homologs of Human *FUT1, FUT2,* and *Sec1* genes

Primer	Sequence (5′–3′)	Position of Primer 5′ End
FUT1.dir.1	TCAGCCCCAGGAGTCTAAAC	150 bp up ATG
FUT1.dir.2	CATGTGGCTCCGGAGCCATCGTCAGCTCTGC	1 bp up ATG
FUT1.rev.1	GGCAGGGTGATGCGGAAT	389 bp dw ATG
FUT1.rev.2	CTGCCCGAGTGGGTGGGCATTAATGCAGAC	54 bp up TGA
FUT1.rev.3	GAGCCAGCAGTACGTGGCTTCAGAGGCCTG	38 bp dw TGA
FUT2.dir.1	CACACACCCACACTATGCCTGCACACCAC	105 bp up ''first'' ATG
FUT2.dir2	GACGATCAATGCAATAGGC	192 bp dw ''first'' ATG
FUT2.rev	GCACTAATGCTGGCCCGTCCTTTGAGACCT	23 bp dw TAA
Sec1.dir.1	CATGTCCTCTCTCCTCCCC(A/G)CAGCCG	1 bp up putative ATG
Sec1.dir.2	TGACCATGTCCTCTCTCCT(C/T)CC	5 bp up putative ATG
Sec1.rev.1	GCCAGCACCCCCTTCCATAC	641 bp dw putative ATG
Sec1.rev.2	CTAGAGGCCGTTCTGTCCAGCCTGTCCAAGG	1,045 bp dw putative ATG
FUT1.intron1.dir	GGAGGCTGGGTATTATCTAG	+936 of *FUT1* intron 1
FUT1.exon2.rev	CACTACCATCAAGACGCAGAGACCG	+48 of *FUT1* exon 2

NOTE.—Positions of intron 1 and exon 2 *FUT1* primers are given by reference to the genomic sequence available in GenBank/EBI (I34189). Positions of *FUT2* primers are indicated by reference to the first initiation codon (Kelly et al. 1995). up = upstream; dw = downstream.

press ABH antigens on red cells. All apes studied so far secrete, in addition, soluble ABH substances in saliva, with the exception of one orangutan (Moor-Jankowski and Wiener 1964). In contrast, monkeys (Old and New World) lack ABH on red cells and are secretors of ABH in saliva (Blancher and Socha 1997; Oriol 1995).

The purpose of this work was to identify potential inactivating mutations in the normal nonhuman primate counterparts of *FUT1* and *FUT2,* especially in the species which do not express H antigen on erythrocytes. Such inactivating mutations could explain the lack of ABH antigens on primate red cells. We also compared our sequences with those of humans, primates, and mammals present in databases in order to study the evolutionary pathway followed by *FUT1*, *FUT2,* and *Sec1* genes.

Materials and Methods

Origin of Samples

Human DNA samples were from Toulouse (Etablissement de Transfusion Sanguine Pyrénées Garonne, Site de Toulouse, France). Chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*), gibbon (*Hylobates lar*), rhesus monkey (*Macaca mulatta*), cynomolgus (*Macaca fascicularis*), and marmoset (*Callithrix jacchus*) blood samples were obtained from animals maintained at the Laboratory for Experimental Medicine and Surgery in Primates (LEMSIP, New York Medical Center, New York University). Squirrel monkey (*Saimiri sciureus*) and brown lemur (*Eulemur fulvus*) blood samples were obtained from the Centre de Primatologie de Strasbourg, Nierderhausbergen, France. Orangutan total RNA was from an animal living at the Bronx Zoo (New York).

Southern Blot

*Bam*HI, *Eco*RI, or *Hin*dIII restriction enzyme-digested DNA samples (10 μg) were separated by 0.8% agarose gel electrophoresis. Southern blot analysis was carried out using ^{32}P-labeled human *FUT1* and *FUT2* probes (whole coding regions). Nylon membranes (Hybond N+, Amersham Pharmacia Biotech, Buckinghamshire, England) were hybridized first with the *FUT1* probe. Hybridizations were performed at 65°C in 5 × SSPE (1 × SSPE = 0.18 M NaCl, 10 mM NaH_2PO4, 1 mM EDTA) supplemented with 0.02% Ficoll, 0.02% bovine serum albumin, 0.5% SDS, and 100 μg/ml of salmon sperm DNA. After each hybridization, filters were washed, first at low stringency (2 × SSPE, 20°C, 15 min) and then at high stringency (2 × SSPE, 0.1% SDS, 50°C for 15 min). Autoradiographies were carried out after the two types of washings (Biomax film, Kodak, Rochester, N.Y.). After removal of the *FUT1* probe in boiling 0.5% SDS, the same membranes were rehybridized with the *FUT2* probe. The hybridization, washing, and autoradiography conditions for the *FUT2* probe were the same as those used for the *FUT1* probe.

Polymerase Chain Reaction and Sequencing

All polymerase chain reactions (PCRs) were carried out using the Expand High Fidelity enzyme mix (Boerhinger, Indianapolis, Ind.). Primers were deduced from the human sequences of *FUT1*, *FUT2,* and Sec1 and from their available nonhuman primate counterparts (table 1). The pairs of primers used for the various species studied were as follows: for chimpanzee, gorilla, gibbon, rhesus monkey *FUT1*-like (DNA)—FUT1.dir.1/FUT1.rev.1 and FUT1.dir.2/FUT1.rev.3; for marmoset and lemur *FUT1*-like (DNA)—FUT1.dir.1/FUT1.rev.1 and FUT1.dir.2/FUT1.rev.2; for orangutan *FUT1*-like (cDNA)—FUT1.dir.2/FUT1.rev.3; for chimpanzee, gorilla, and gibbon *FUT2*-like (DNA)—FUT2.dir/FUT2.rev; for rhesus monkey, marmoset, and lemur *FUT2*-like (DNA)—FUT2.dir.2/FUT2.rev; for rhesus monkey *Sec1*-like (DNA)—Sec1.dir.1/Sec1.rev.2; for squirrel monkey *Sec1*-like (DNA)—Sec1.dir.2/Sec1.rev.1; for marmoset *Sec1*-like (DNA)—Sec1.dir.2/Sec1.rev.2; for chimpanzee, gorilla, orangutan, gibbon, rhesus monkey, cynomolgus, and baboon *FUT1* intron 1—FUT1.intron1.dir/FUT1.exon2.rev.

Genomic DNA was used as a template for PCR for all sequences, with the exception of the orangutan *FUT1*-like sequence, which was obtained from a total RNA sample. The *FUT1*-like cDNA was amplified by PCR after reverse transcription using oligo-dT and Superscript II (Life Technologies Inc., Gaithersburg, Md.). When necessary, amplified fragments were cloned into pCR 2.1.TOPO plasmid vector (TOPO TA cloning kit, Invitrogen, Leek, the Netherlands). PCR primers and species-specific internal primers were used for fully sequencing both strands by means of a fluorescent dye terminator cycle sequencing kit (Amplitaq FS, Perkin Elmer, Foster City, Calif.).

Sequence Analysis and Phylogeny

Alignment of all the available α2-fucosyltransferase nucleotide and peptide sequences was carried out using CLUSTAL W, version 1.7 (Thompson, Higgins, and Gibson 1994), on the basis of the *FUT1*/*FUT2* and *FUT1*/*Sec1* alignments of human sequences (Kelly et al. 1995). Calculated distances were either gamma distances which were calculated between amino acid sequences (parameter a = 2), or distances between nucleotide sequences which were estimated by using the two-parameter method (Kimura 1980). Phylogenetic trees were constructed with the neighbor-joining method (MEGA software; Kumar, Tamura, and Nei 1993). Bootstrap analysis was performed on 500 data sets.

The new primate sequences reported in this paper have been submitted to GenBank/EBI under the following accession numbers: *FUT1*-like genes—chimpanzee, AF080603; gorilla, AF080605; orangutan, AF111935; gibbon, AF045545; rhesus monkey, AF080607; cynomolgus, AF112474; squirrel monkey, AF136647; marmoset, AF111936; brown lemur, AF045546. *FUT2*-like genes—chimpanzee, AF080604; gorilla, AF080606; gibbon, AF136648; rhesus monkey, AF136644; marmoset, AF136645; brown lemur, AF136647. *Sec1*-like genes—rhesus monkey, AF080608; cynomolgus, AF112475; squirrel monkey, AF111937; marmoset, AF111938.

In addition to the primate α2-fucosyltransferase genes described in this article, the GenBank/EBI accession numbers of fucosyltransferase genes shown in the phylogenetic tree but cloned by other teams are as follows: *FUT1*—human, M35531; green monkey, D87932; pig, U70883, L50534; rabbit, X80226; mouse, U90553, Y09883, AF113533; rat, AB015637, AB006137, AF131237; *FUT2*—human, U17894; chimpanzee, AB015634; gorilla, AB015635; orangutan, AB015636; green monkey, D87934; cow, X99620; pig, U70881, AF027304; rabbit, X91269; mouse, AF064792; rat, AB006138, AF131238, AF042743; *Sec1*—human U17895; chimpanzee, AB006612; gorilla, AB006611; orangutan, AB006610; gibbon, AB006609; green monkey, D87933; pig, U70882; rabbit, X80225; mouse, Y09882, AF113532; rat, AF131239. For cases in which several sequences of the same gene for a given species were found in data banks, all of the available accession numbers are listed, but only one sequence appears in the tree.

Detection of repetitive elements and *Alu* subfamily classification were performed with the CENSOR Web server (http://www.girinst.org/) (Jurka et al. 1996). Analysis of the promoter region located 5′ of *FUT1* exon 2 was performed with the TFSEARCH program (http://pdap1.trc.rwcp.or.jp) (Akiyama 1995).

Results

Southern Blot Analysis of α2-Fucosyltransferase Genes

Genomic DNA from man and nonhuman primates was digested with *Bam*HI, *Eco*RI, or *Hin*dIII restriction enzymes and hybridized with the human *FUT1* probe (fig. 1*A–C*, left) and the human *FUT2* probe (fig. 1*A–C*, right). The numbers of bands revealed with distinct couples of restriction enzymes and probes are summarized in table 2.

For humans, the sizes of *Eco*RI restriction fragments were in accordance with previously published results (Rouquier et al. 1995). The *FUT1* probe recognized a band of 6.4 kb, corresponding to the *FUT1* gene. The *FUT2* probe revealed two *Eco*RI fragments: a strong band of 18 kb, which corresponds to *FUT2*, and a weak band of 8.2 kb, which contains *Sec1*.

Hybridization patterns obtained with human, chimpanzee, and gorilla samples were nearly identical for all combinations of enzyme and probe. For these three species, the *FUT2* probe recognized two bands of different intensities. The more intense band corresponds to *FUT2* in humans and to *FUT2*-like genes in African apes. The less intense band contains *Sec1* (humans) or *Sec1*-like (African apes) genes. Hybridization patterns obtained with gibbons, Old World monkeys (baboons, rhesus monkeys, cynomolgus monkeys), and New World monkeys (squirrel monkeys, marmosets) differed from those observed for humans and African apes (chimpanzees, gorillas) in the numbers of bands and the lengths of hybridized fragments.

Amplification and Sequencing of *FUT1*-like Genes

The whole coding sequence of *FUT1*-like genes was amplified by PCR from genomic DNA using primers described in table 1. In humans, the coding part of *FUT1* is fully included in exon 8 (Koda, Soejima, and Kimura 1998). The whole coding sequence was obtained for chimpanzees, gorillas, gibbons, rhesus monkeys, and cynomolgus monkeys. The 3′ end of the *FUT1*-like coding exon (potentially encoding amino acids 346–365) could not be amplified in samples from New World monkeys (squirrel monkeys and marmosets) and brown lemurs (two animals studied).

Nonhuman *FUT1*-like sequences potentially encode proteins of 365 aa. The predicted H-like enzymes (fig. 2*A*) did not exhibit premature stop codons and are highly similar to the human H sequence, with 96% amino acid identity between humans and rhesus monkeys, and 87% identity between humans and brown lemurs.

Nucleotides flanking *FUT1*-like coding sequences, which correspond in humans to the 3′ end of *FUT1* in-

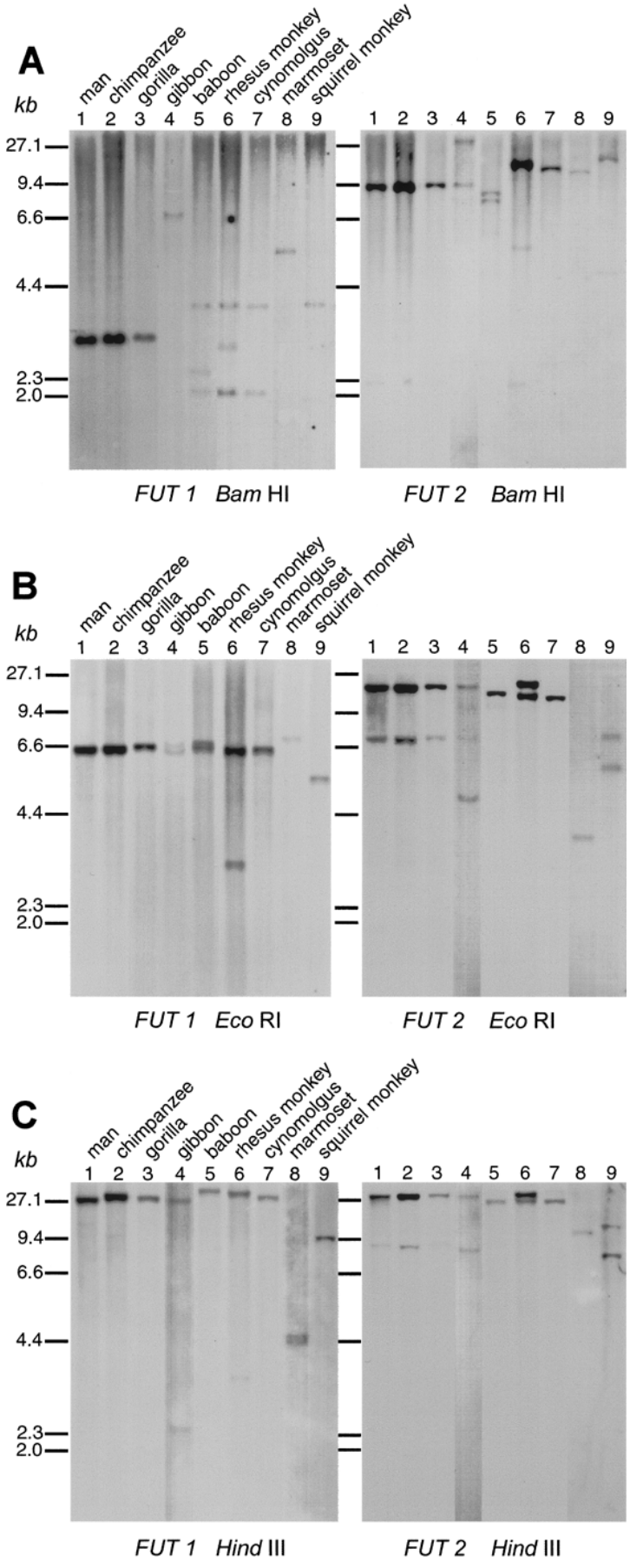


Table 2
Number of Bands Revealed with *FUT1* or *FUT2* Probes on Southern Blots

DNA species	*FUT*1 Probe			*FUT2* Probe		
	*Bam*HI	*Eco*RI	*Hind*III	*Bam*HI	*Eco*RI	*Hind*III
Human	1	1	1	2	2	2
Chimpanzee	1	1	1	2	2	2
Gorilla	1	1	1	2	2	2
Gibbon	1	2	2	2	3	2
Baboon	3?	2?	1	2	1	2
Rhesus monkey	3*	2	2	3	2	2
Cynomolgus	2*	1	1	1	1	2
Marmoset	1	1	2?	1	1	1
Squirrel monkey	2*	1	1	2	2	2

NOTE.—Genomic DNA was digested with *Bam*HI, *Eco*RI, or *Hind*III restriction enzymes and hybridized with cDNA probes corresponding to the coding parts of human *FUT1* or *FUT2* genes. Relative intensities of the bands are shown in figure 1*A–C.* DNA samples from two chimpanzees and two gorillas gave identical bands (data not shown). An asterisk indicates that a *Bam*HI site was found in the coding sequence of *FUT1* (rhesus monkeys, cynomolgus monkeys, and marmosets). A question mark indicates that the *FUT1* coding sequence was not fully determined (marmosets) or that *FUT1* was not sequenced in a species (baboons).

tron 7, were amplified and sequenced from DNA samples of chimpanzees, gorillas, gibbons, rhesus monkeys, cynomolgus monkeys, marmosets, and lemurs (fig. 3).

Amplification and Sequencing of *FUT2*-like Genes

We characterized chimpanzee, gorilla, and gibbon genomic sequences of *FUT2*-like coding regions, which are, in humans, part of the *FUT2* exon 2 (Koda et al. 1997*b*). Only the 3′ part of *FUT2* (corresponding to amino acids 58–330) could be amplified from samples from rhesus monkeys, marmosets, and brown lemurs. In addition, a short segment of the cynomolgus *FUT2*-like gene was amplified from nucleotide 192 to nucleotide 591 (data not shown). All sequences were translatable into 332 (short) or 343 (long) Se-like polypeptides (fig. 2*B*), with a degree of identity with the human Se enzyme ranging from 89–90% (marmosets and lemurs) to more than 99% (only three amino acid replacements between humans and chimpanzees). The chimpanzee *FUT2*-like sequence reported here was identical to a previously characterized chimpanzee *FUT2*-like sequence (AB015634). In humans, two distinct translation initiation codons have been described for *FUT2* cDNA (Kelly et al. 1995), and the two corresponding polypeptides are shown in figure 2*B.*

Amplification and Sequencing of *Sec1*-like Genes

Human *Sec1* is a pseudogene with a deletion of 2 nt (GG between positions 668 and 671) leading to a frameshift and predicting a truncated polypeptide of 246 aa, with residues 223–246 being unrelated to fucosyltransferase polypeptides (Kelly et al. 1995). The gorilla *Sec1* has an insertion of one C in position 612 or 613 inducing a frameshift. Its predicted polypeptide is truncated to 247 aa, and residues 205–247 are unrelated to H enzymes. The chimpanzee *Sec1* has a nonsense C→T mutation at position 976, inducing a premature TAA stop codon and the loss of the 23 C-terminal amino acids. The human and gorilla frame 2 sequences had the same C→T mutation (fig. 2*C*).

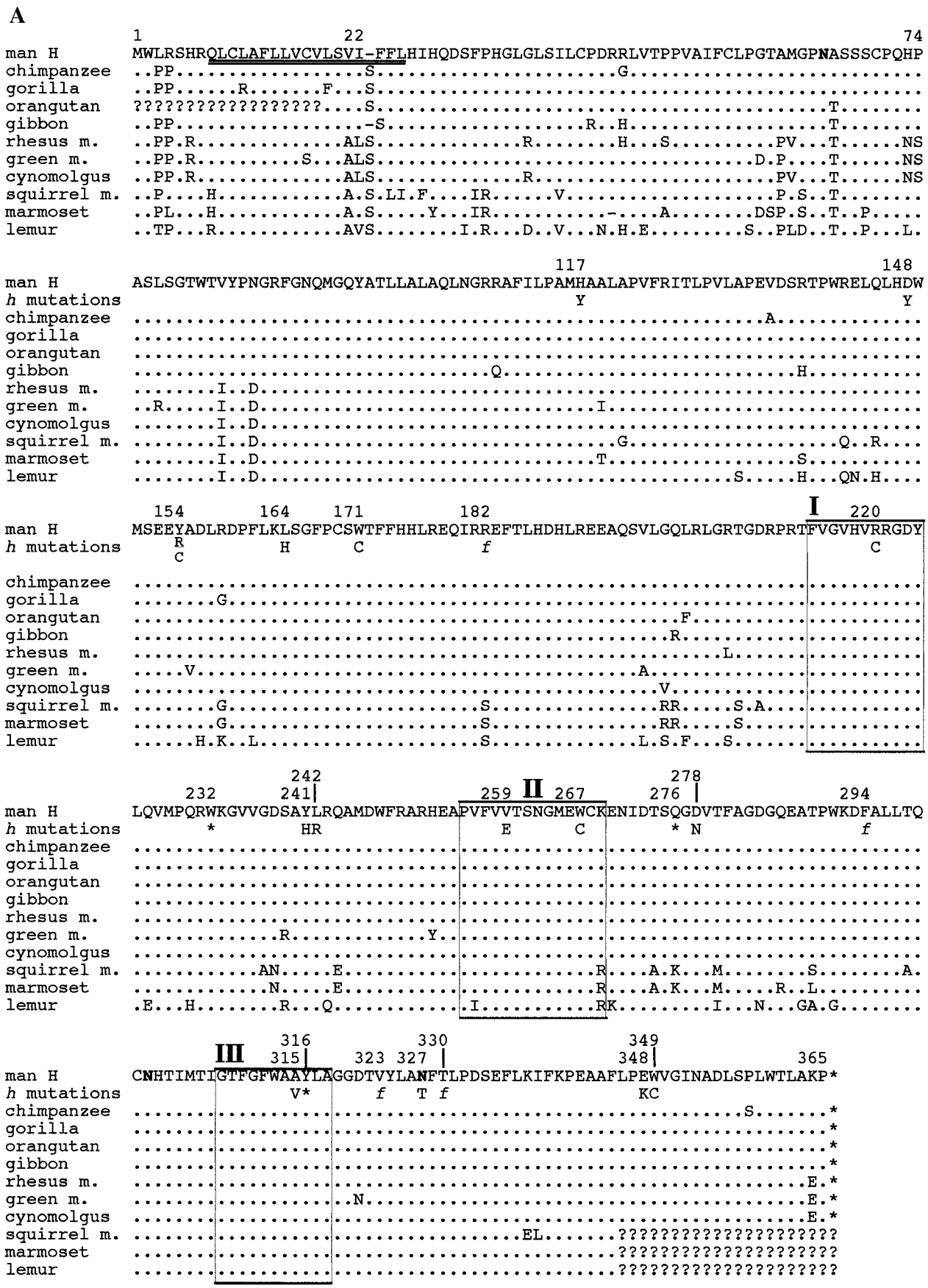
We also characterized genomic *Sec1*-like sequences of rhesus monkeys, cynomolgus monkeys, squirrel monkeys, and marmosets. Unlike human, gorilla, and chimpanzee *Sec1*, rhesus monkey and cynomolgus *Sec1*-like

←

FIG. 1.—Southern blot analysis with human *FUT1* (left) and *FUT2* (right) probes on primate genomic DNA. *A, Bam*HI restriction enzyme. *B, Eco*RI restriction enzyme. *C, Hin*dIII restriction enzyme. Under the stringency conditions used, the human *FUT1* probe reveals *FUT1*-like primate bands, and the *FUT2* human probe reveals *FUT2*-like and *Sec1*-like primate bands.

→

FIG. 2.—Amino acid multiple alignments. H-like, Se-like, and Sec1-like polypeptides are aligned by reference to H and Se human proteins or rabbit Sec1, respectively. Dots represent identity, dashes represent amino acid deletions, and question marks represent unknown positions. Residue numbers are indicated above the reference sequences. Amino acid positions associated to silent phenotypes are individually numbered. Amino acid residues within the proposed transmembrane domains in reference polypeptides are double underlined. The α2-fucosyltransferase conserved peptide motifs I, II, and III (Oriol et al. 1999) are boxed. Potential Asn-linked glycosylation sites are indicated in bold type. *A,* H polypeptides (*FUT1* gene). Replacement amino acids in silent *h* alleles are shown under the human H sequence; dashes indicate deletions, asterisks indicate that silencing mutations induced the introduction of a stop codon, and f's (*f*) indicate that mutations led to frameshifts in the reading frame. *B,* Se polypeptides (*FUT2* gene). Replacement amino acids or stop codons (*) associated with the nonsecretor phenotype are shown below the Se sequence. *C,* Sec1 polypeptides are aligned by reference to the translation of the rabbit *Sec1* coding sequence, whose functionality has been assessed by in vitro experiments (Hitoshi et al. 1995, 1996). Amino acid residues in human and gorilla Sec1, when unrelated to rabbit Sec1, are indicated with crosses. Human *Sec1* was made translatable (human frame 2) by addition of two nucleotides at positions 670 and 671. The gorilla *Sec1*-like sequence is missing one nucleotide at position 613 to obtain the translatable segment of gorilla frame 2. Chimpanzees and reading frames 2 of gorillas and humans had the 976 C→T mutation, inducing a premature TAA stop codon. Segments of human and gorilla polypeptides which were translatable from modified *Sec1* genes are shown in lowercase letters.



B

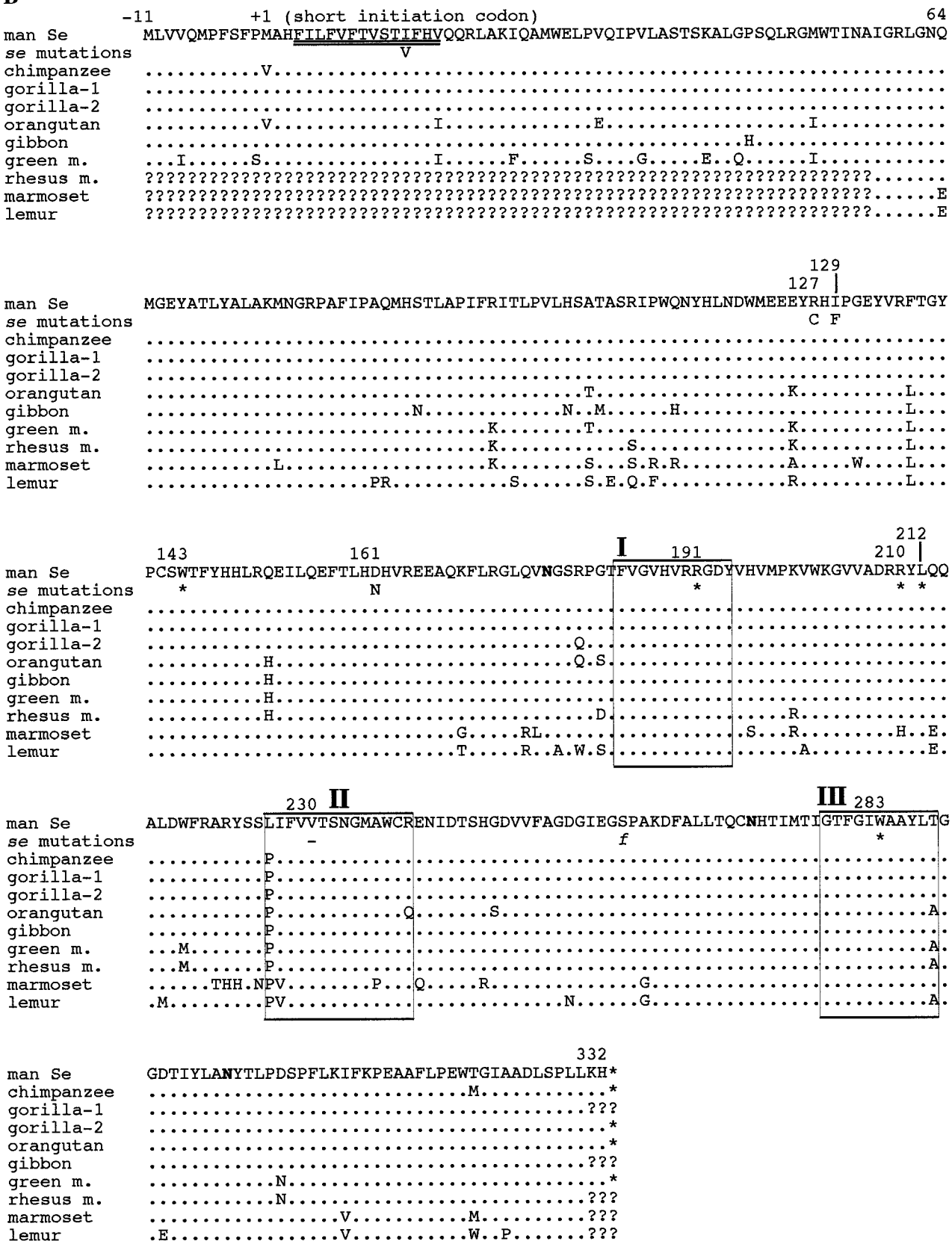


FIG. 2 *(Continued)*

C

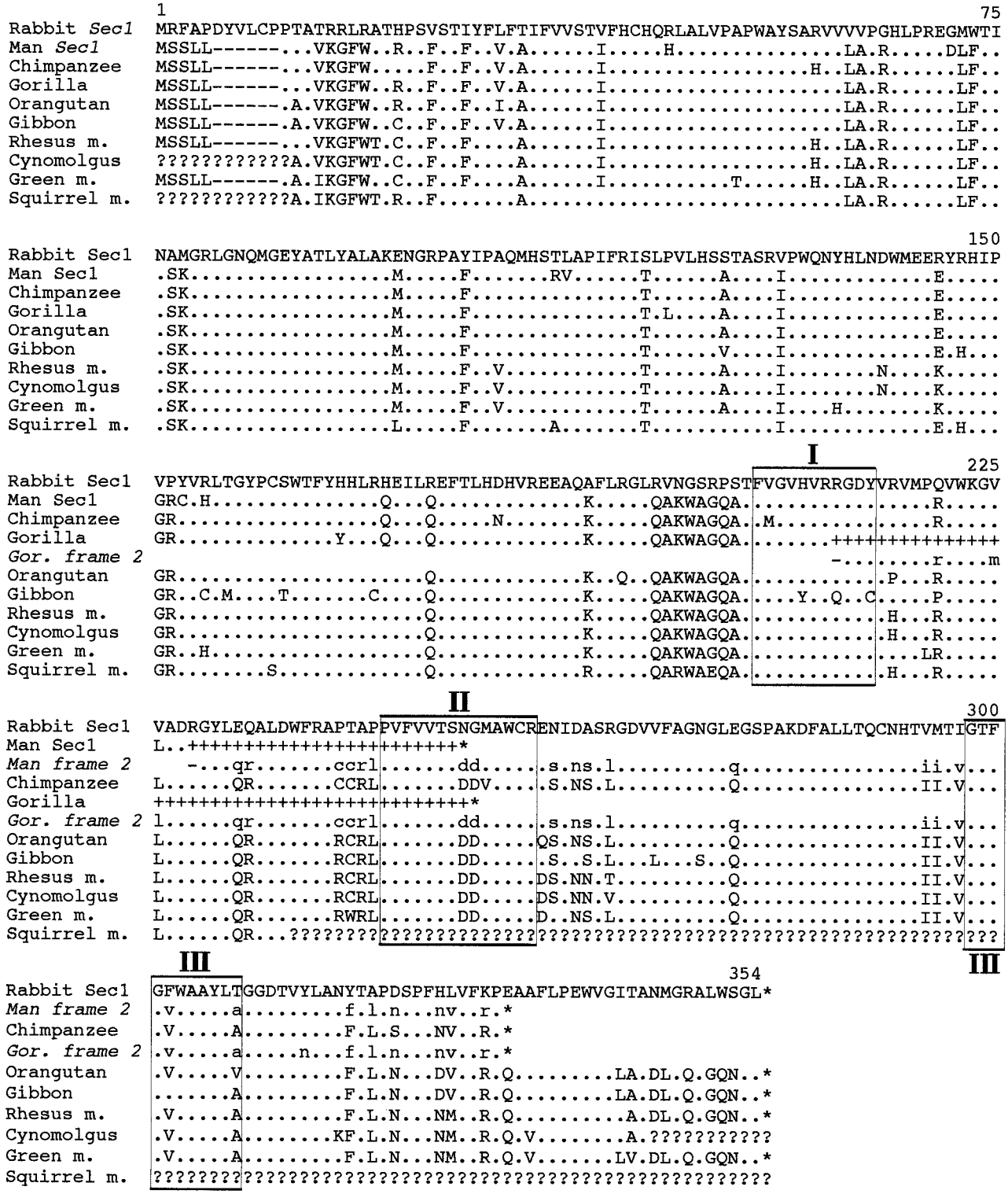


FIG. 2 *(Continued)*

genes had the same open reading frame as those of orangutans, gibbons, and green monkeys (fig. 2*C*). The squirrel monkey *Sec1*-like gene could be amplified only between nucleotides 40 and 693 (by reference to the human Sec1 sequence) and potentially encodes a Sec1 polypeptide without premature stop codons. From two marmoset genomic DNA samples, we amplified fragments which were homologous to human *Sec1* but displayed multiple deletions (one base in position 170, 23 bases between nucleotides 193 and 216, 210 bases between nucleotides 326 and 536, and a deletion of one base in position 610). Excluding the deleted segments,

```
        -105
man         TCCCTGGGAC TAAGGAGTGC TGCACCCCAG GCGCCTCCCT TACCCC-ACA TCCCTCCTCA
chimpanzee  .......... .......... .......... .......... ......-... ..........
gorilla     .......... .......... .......... .......... ......-... ..........
gibbon      .......... .......... .......... .......... .C....-... ..........
rhesus      .......... .......... .......... ..A....... .C....-... ..........
cynomolgus  ...-...... .......... .......... ..A....... .C....C... ..........
marmoset    ...-..CA.T GG...GAAAA ..A..T..C. ..CT.C.T.. CCT...-C.. ..........
lemur       ?????????? ?????????? ??????.... ..--...T.C .C..TG-..C .T.......G

                    -43                          intron 7                  -1  exon 8
man         GCCTCCCCTC CWCAGCCTCA GTGCATTTGC TAATTCGCCT TTCCTCC--C CTGCAG  CCATG
chimpanzee  ...-------- -------... .......... .......... .......--. ......
gorilla     .......... .......... .......... .......... .......--. ......
gibbon      ...-------- --------.C .......... .......T.. .......--. ......
rhesus      .......... .CT..T...T ...T...... ......A... .......--. ......
cynomolgus  .......... .CT..T..TT ...T...... ......A... .......--. ......
marmoset    ...G.....G ....A....T .C....C... ......T... .......--. ......
lemur       .G.------- --------.T .A.T..C... ...C...... .......TC. ......
```

FIG. 3.—Alignment of intronic sequences 5′ upstream of the coding exon of *FUT1*. *FUT1*-like sequences of nonhuman primates are shown by reference to the human *FUT1* intron 3 sequence (Wagner and Flegel 1997) (Z69587). Dots represent nucleotide identities and dashes represent deletions. Nucleotide positions of intron 3 are numbered backward, taking as position −1 the G of the 3′ splicing consensus site of intron 7, which is underlined. Position −43 is polymorphic in humans (A/T). The ATG initiation codon within exon 8 is double underlined.

the marmoset *Sec1* sequence exhibited 90% and 84% nucleotide identities when compared with human *Sec1* or human *FUT2*, respectively.

Phylogeny

In order to study the phylogenetic relationships of the three α2-fucosyltransferase families, we made a protein alignment of the available sequences from the products of *FUT1*, *FUT2*, and *Sec1* genes. Three primate sequences for which large segments remained uncharacterized (squirrel monkey Sec1, rhesus monkey Se, and marmoset Se) and polypeptides encoded by pseudogenes (human, gorilla, and marmoset Sec1) were not included. The phylogenetic tree showed two major clusters, one encompassing the H-related (*FUT1*) proteins and the other encompassing all the Se-related (*FUT2*) and Sec1-related proteins (fig. 4). Therefore, the tree suggests that a common α2-fucosyltransferase ancestor gene duplicated to give rise to the *FUT1* ancestor gene and an ancestor gene common to *FUT2* and *Sec1*. In the *FUT2-Sec1* cluster, two distinct subfamilies can be identified, corresponding to Se and Sec1 proteins, respectively. This suggests that the *FUT2-Sec1* common ancestor gene duplicated before the great mammalian radiation and gave rise to *FUT2* and *Sec1* genes. Moreover, distances in the cluster of Sec1 sequences are longer than distances in the Se (FUT2) cluster. This is particularly evident for primate sequences (fig. 4).

Amplification of Intron 1 of *FUT1*-like Genes

A contig of 11.5 kb was created by aligning three *FUT1* DNA segments available in databases: (1) exon 1 and its 5′ flanking segment (Koda, Soejima, and Kimura 1998), (2) a large segment encompassing exon 1 to the beginning of exon 8 (direct submission by J. B. Lowe; GenBank accession number I34189), and (3) a genomic fragment from exon 8 to the end of the 3′ untranslated region (Larsen et al. 1990). In this contig, which encompasses most of the *FUT1* gene, we identified 11 *Alu* elements which belong to various *Alu* subfamilies. On the basis of the known age of these *Alu* subfamilies, one can infer the relative date of insertion of these *Alu* elements into *FUT1* (fig. 5). One of these *Alu* elements lies in close proximity to a region suggested to be involved in the red cell expression of *FUT1* in humans (Koda, Soejima, and Kimura 1998) and belongs to the *Alu*-Y/Sb2 subfamily (Jurka et al. 1996). Therefore, we checked for the presence of this *Alu* element in species known to express the H antigen at the surfaces of red blood cells (humans, chimpanzees, gorillas, orangutans, and gibbons) and in species which are known to lack the H antigen on red blood cells (macaques and baboons). Primers were deduced from available human sequences for the amplification of fragments including this *Alu* element (see table 1 and *Materials and Methods*).

Primers *intron1-dir* and *exon2-rev* allowed amplification of 0.75-kb fragments from human, chimpanzee, gorilla, and orangutan DNA samples. Partial sequencing and analysis of the fragments demonstrated the presence of the *Alu*-Y element in these species. For the gibbon, the amplified fragments were 1.05 kb long, but two *Alu*-Y elements were present, with the upstream one being inserted in the 3′ oligo-dA tail of the *Alu* in 3′ position. In Old World monkeys (rhesus monkeys, cynomolgus monkeys, and baboons) PCR products of 0.45 kb were obtained, but direct sequencing of these fragments did not allow the study of intron 1, because multiple genomic targets were amplified with primers *intron1-dir* and *exon2-rev* in these species.

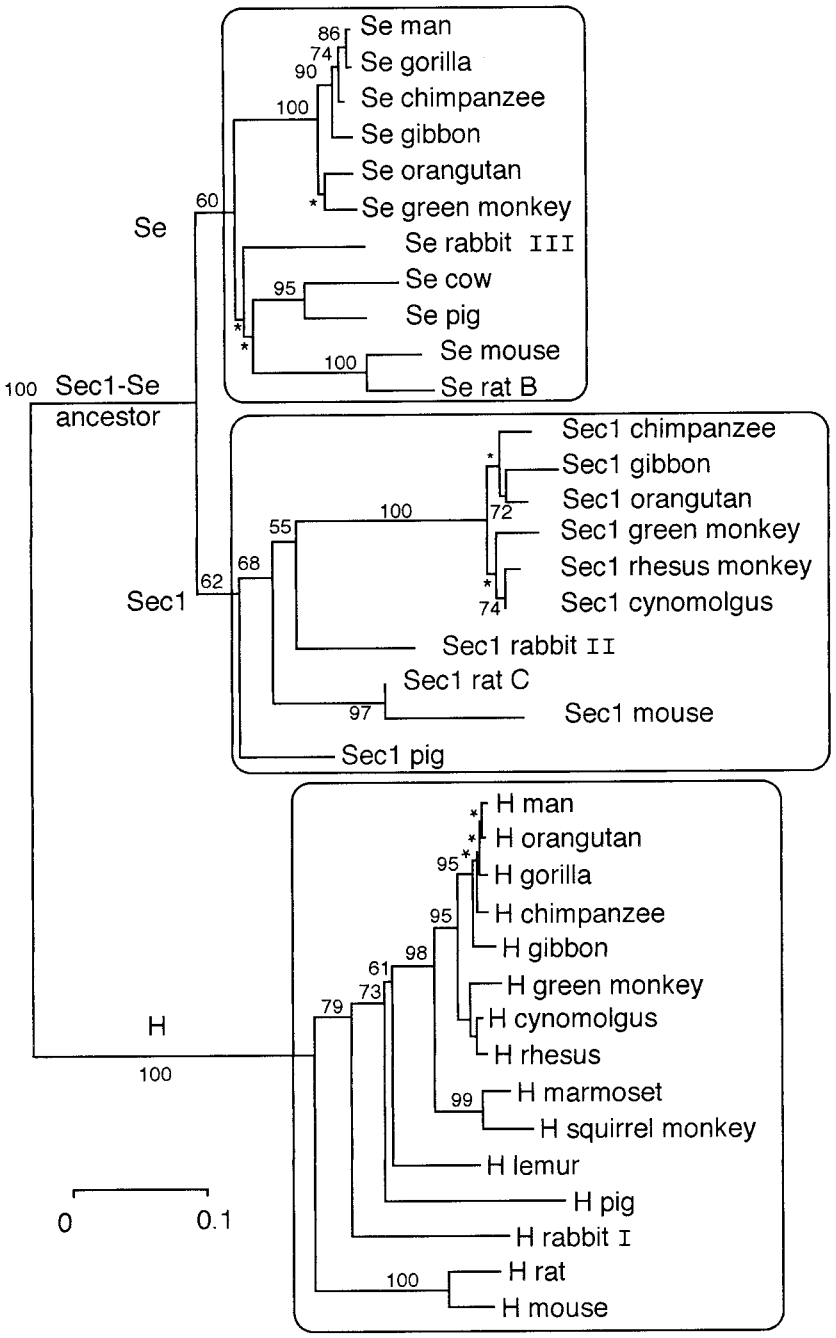


FIG. 4.—Phylogenetic tree of amino acid sequences deduced from *FUT1*-like, *FUT2*-like, and *Sec1*-like coding sequences. Multiple alignments were made with CLUSTAL W taking into account the peptide conserved motifs I, II, and III (Oriol et al. 1999). Human, gorilla, and marmoset Sec1's were excluded from the protein alignment because frameshifts and deletions would have given aberrant distances. Likewise, rhesus monkey Se, marmoset Se, and squirrel monkey Sec1 were not included because these sequences were not fully characterized. Gamma distances were calculated (parameter a = 2), and 500 data sets were generated by bootstrap (MEGA software; Kumar, Tamura, and Nei 1993). The resulting tree was constructed by neighbor joining and is shown with the bootstrap frequencies on the branching points. Asterisk indicates bootstrap frequencies <50%.

Discussion

Counterparts of *FUT1*, *FUT2*, and *Sec1* in Nonhuman Primates

Southern patterns revealed by the FUT1 probe in human, chimpanzee, and gorilla genomic DNA digested either by *Bam*HI, *Eco*RI, or *Hin*dIII were identical and consisted of only one band (table 2). This confirmed that chimpanzees and gorillas, like humans, possess only one *FUT1* gene per haploid genome.

The sequence of *FUT1* cynomolgus and squirrel monkey amplicons revealed the presence of one *Bam*HI site, which explains the presence of two bands on the *Bam*HI Southern blot. Since the cynomolgus and squirrel monkey *Eco*RI and *Hin*dIII patterns displayed only

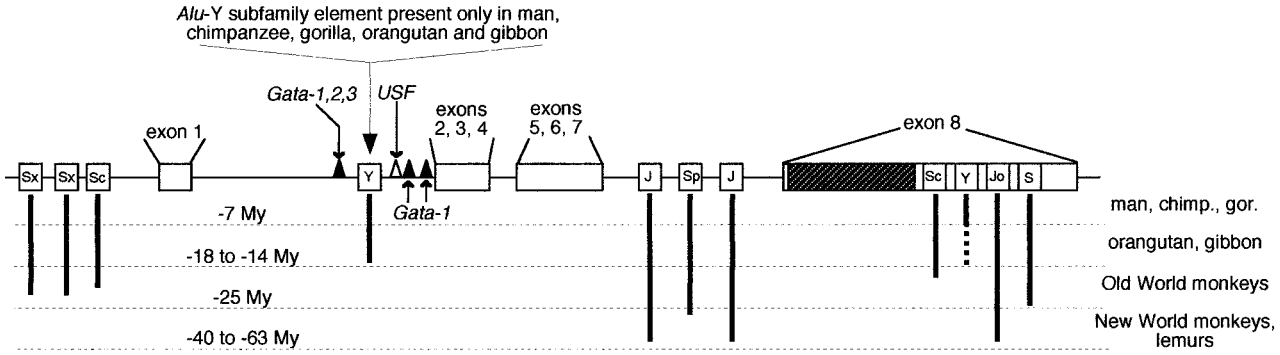


FIG. 5.—Proposed organization of the *FUT1* genomic region in humans and nonhuman primates. The genomic organization of human *FUT1* was deduced from the alignment of genomic sequences available in databases (see text). Each box represents either exons or *Alu* elements inserted in noncoding portions of the gene. *Alu* subfamilies (Jurka et al. 1996) are indicated in the squares: *Alu*-J/Jo, *Alu*-Sp, *Alu*-Sx, *Alu*-Sc, and *Alu*-Y. The coding part of the *FUT1* gene is indicated by a hatched box. Relative dates of insertion for *Alu* elements are proposed on the basis of the literature (Kapitonov and Jurka 1996) and are shown with vertical bars. The presence of an *Alu*-Y element in intron 1 was established for humans and anthropoid apes (this paper). Proposed sites for the binding of GATA and USF transcription factors in intron 1 are indicated with black or white triangles, respectively.

one *FUT1* band, the results are compatible with the presence of a single *FUT1* gene per haploid genome in these two species.

In gibbons, the presence two *Eco*RI and *Hin*dIII *FUT1* bands and the absence of *Eco*RI and *Hin*dIII sites in *FUT1* amplicons suggest that gibbons could possess two copies of the *FUT1* gene per haploid genome.

In rhesus monkeys and baboons, the *FUT1* probe revealed three bands (*Bam*HI) or two bands (*Eco*RI) (table 2). In the case of rhesus monkey, the sequence of the *FUT1* amplicon displayed one *Bam*HI site, but no *Eco*RI or *Hin*dIII restriction sites. Therefore, we conclude that the number of bands revealed by Southern blot is compatible with at least two genomic targets. The absence of any ambiguity by direct sequencing of rhesus *FUT1* amplicons led us to suppose that only one *FUT1* gene was amplified. Accordingly, one of the two rhesus *FUT1* genes is probably too mutated and cannot be amplified by using primer pairs described in this study. In the case of baboons, the whole DNA sample was used for the Southern blot, and no firm conclusion could be drawn for the number of *FUT1* genes in this species. However, the similarity of the Southern pattern with that of rhesus monkeys suggests that the baboon also possesses two copies of *FUT1* per haploid genome.

The study of marmoset *FUT1* amplicons revealed the absence of *Bam*HI, *Eco*RI, and *Hin*dIII restriction sites. The presence of two *FUT1 Hin*dIII bands is therefore not explained by the presence of a *Hin*dIII restriction site in the region hybridized by the *FUT1* probe. This could signify the presence of at least two *FUT1* copies per haploid genome. However, the sequence of the marmoset *FUT1* is incomplete, and one cannot exclude the presence of a *Hin*dIII site in the missing portion of the sequence.

Both *FUT2* and *Sec1* genes are detected in humans with the *FUT2* probe (Rouquier et al. 1995). On the basis of the band patterns revealed by this *FUT2* probe, one can infer the presence of counterparts of these two genes in African apes, gibbons, rhesus monkeys, and squirrel monkeys (fig. 2*B*). On the other hand, the *FUT2* probe only revealed patterns of one or two bands in cynomolgus and marmoset samples (table 2). Since the *Sec1*-like coding sequence of the marmoset had large deletions, it is possible that the *FUT2* probe could not hybridize with marmoset *Sec1*. Likewise, only a short segment of the cynomolgus *FUT2*-like gene could be amplified (nucleotides 192–591, data not shown), and all other combinations of *FUT2*-specific PCR primers failed to amplify *FUT2* in this species. Therefore, in the same way as the marmoset *Sec1*-like gene, the cynomolgus *FUT2*-like gene could be inactivated by deletions inside its coding region. Another possibility is the presence of additional restriction sites in marmoset and cynomolgus *Sec1* or *FUT2* genes, leading to the appearance of short DNA fragments which could not be detected by Southern blot analysis.

Most Nonhuman Primates Potentially Express Three α2-Fucosyltransferases

None of the primate *FUT1*-like coding sequences exhibit mutations leading to the appearance of stop codons or altering the reading frame, and none of the mutations restricted to red cell H-deficient human phenotypes were found in the nonhuman primate H-like polypeptides analyzed. Furthermore, 10 highly conserved amino acids distributed within three conserved peptide motifs have been identified in all the functional α2-fucosyltransferases of vertebrates, invertebrates, and bacteria (Oriol et al. 1999). These amino acids are expected to participate in the GDP-fucose donor recognition of the enzyme (Breton, Oriol, and Imberty 1996, 1998), and none of them was modified in any of the *FUT1*-like primate sequences analyzed. Regions upstream of exon 8 of primate *FUT1* genes were also sequenced, and no defect was found in the 3′ splicing consensus site (fig. 3).

Similar to *FUT1*, none of the *FUT2* inactivating mutations identified in humans were found in the nonhuman *FUT2*-like sequences studied (fig. 2*B*), and all of the highly conserved amino acids of α2-fucosyltrans-

ferases were conserved in all of the *FUT2*-like primate sequences analyzed.

The predicted human Sec1 protein has no enzyme activity and lacks the α2-fucosyltransferase conserved motifs II and III (Oriol et al. 1999). The gorilla Sec1-like sequence also lacks these two conserved motifs and half of the α2-fucosyltransferase conserved motif I (fig. 2*C*). Consequently, no enzyme activity can be expected from this gorilla peptide. The predicted chimpanzee Sec1 protein has the three conserved α2-fucosyltransferase peptide motifs, but it is shortened to 325 aa by a nonsense mutation and is also probably nonfunctional (fig. 2*C*). Although the truncation effects on the activity of this particular enzyme have not been determined, studies on other fucosyltransferases have shown that deletion of one or more amino acids on the C-terminus (catalytic domain) induces a dramatic loss of enzyme activity (Xu, Vo, and Macher 1996), whereas truncation of as much as 60 aa on the N-terminus (intracytosolic, transmembrane, and stem domains) did not modify the enzyme activity (DeVries et al. 1995). Careful analysis of the human and gorilla *Sec1* sequences showed that the same stop codon inducing the *Sec1* chimpanzee premature truncation is also present, but in reading frame 2 of human and gorilla *Sec1* pseudogenes (fig. 2*C*), suggesting that the nonsense mutation had occurred in the common ancestor of humans and African apes (chimpanzees and gorillas). The other mutations (GG deletions in humans and C insertion in gorillas) differ in the two species and therefore postdated the speciation of humans and gorillas. For all the other primate species, with the exception of chimpanzees, gorillas, and marmosets, *Sec1*-like genes encode potential functional Sec1-like enzymes of 348 aa (fig. 2*C*). Therefore, Asian apes, Old World monkeys, and some New World monkeys might express three distinct α2-fucosyltransferases like the rabbit (Hitoshi et al. 1995, 1996).

Divergent Evolution of Fucosyltransferase Genes

No strong sequence homology has been detected between α2-fucosyltransferase (*FUT1, FUT2, Sec1*) and α3-fucosyltransferase (*FUT3, FUT4, FUT5, FUT6, FUT7,* and *FUT9*) genes, but within each of these two gene families pairwise comparison of homologous genes of different mammals always gives strong homology. This suggests that the α2-fucosyltransferase ancestor gene and the α3-fucosyltransferase ancestor gene differentiated before the mammalian radiation (Costache et al. 1997*a*). We previously proposed that these two ancestor genes could have derived, by duplication, from a single common ancestor gene (Oriol et al. 1999).

The present study confirms that before the mammal radiation, the α2-fucosyltransferase ancestor gene duplicated to give rise to two genes: the *FUT1* gene and another gene which duplicated again, giving rise to *FUT2* and *Sec1*. As a matter of fact, the three human α2-fucosyltransferase genes are located in a segment of less than 100 kb in 19q13.3. *FUT2* and *Sec1* are separated by only 12 kb (Rouquier et al. 1995). This is also the case for the rabbit, for which the loci of *RFT-I, RFT-II,* and *RFT-III* genes, which are the respective counterparts of *FUT1, Sec1,* and *FUT2,* also extend over less than 100 kb (Hitoshi et al. 1995, 1996).

Genetic exchange between the three α2-fucosyltransferase genes must have been infrequent, since there is no clustering of *FUT1, FUT2,* and *Sec1* sequences by species in the tree presented in figure 4. On the contrary, the differentiation between the three α2-fucosyltransferase genes has been maintained in all of the mammals studied here. The phylogenetic tree of amino acid sequences shows that the evolutionary rates of the three α2-fucosyltransferase genes are different, with the highest rate being observed for Sec1 sequences.

Red Cell–Specific Transcription of *FUT1* in Anthropoids: A Possible Role for the Insertion of an *Alu*-Y Element in Intron 1 of the Anthropoid Ancestor

Although our experiments did not formally demonstrate the absence of the *Alu*-Y element in Old World monkeys, it is very unlikely that an *Alu* element belonging to the Y subfamily is inserted in the genomes of these species (Kapitonov and Jurka 1996). Analysis of the *Alu*-Y sequence in gibbons and orangutans (data not shown) confirmed that this element is similar to other *Alu* elements, proposed to be precursors of the *Alu*-Y subfamily, which expanded mostly in humans and African apes (Shaikh and Deininger 1996; Arcot et al. 1998).

In humans, the transcription start site of *FUT1* changes during differentiation of the red cell lineage. The preferred transcription start site in the leukemia cell line K562 is upstream of exon 1, but in erythroleukemia cells (HEL), transcription starts preferentially within intron 1, upstream of exon 2. In bone marrow cells, transcription of *FUT1* starts in intron 7, upstream of the coding part of exon 8 (Koda, Soejima, and Kimura 1997, 1998). Interestingly, we found an *Alu*-Y element inside intron 1, close to the transcription initiation site (Koda, Soejima, and Kimura 1998). An analysis of the promoter region upstream of exon 2 allowed us to identify potential sites for the fixation of transcription factors: a red cell–specific (W)GATA(R) box is present in position −555, upstream of nucleotide +1 of exon 2, and is flanked downstream (−102 bp) by an ''E'' box, which is a target for the DNA binding of upstream stimulation factors (USF 1 and 2). These two sites are separated by the *Alu*-Y (fig. 5). Proximity of GATA-1 and USF binding sites is crucial for the transcriptional regulation of other red cell–specific genes such as the glycophorins A and B (Vignal et al. 1990; Rahuel et al. 1992). On the other hand, Koda, Soejima, and Kimura (1998) transfected K562 and HEL cells with constructs harboring a reporter gene coupled with intron 1 segments of variable lengths. The USF binding site was identified as the main DNA segment necessary for the intron 1 promoter activity. Therefore, the presence of the *Alu*-Y could either inactivate (by disruption of the GATA/USF sites) or activate (by making the USF binding site more accessible) the transcription initiated by the intron 1 promoter (fig. 5).

The presence of *Alu* elements in promoter regions of human genes has previously been related to gene transcription modulation. For example, an *Alu* element (Sz subfamily) is inserted in the promoter region of an allele of the myeloperoxydase (MPO) gene and contains target sequences for the binding of SP1 transcription factors (Piedrafita et al. 1996). The MPO allele which possesses this *Alu* is 25-fold more transcribed than the allele which misses the *Alu* element. On the other hand, the deletion of an *Alu*-Sz element in the promoter of a potassium channel gene (*Kir6.2*), which is involved in the regulation of insulin secretion, increases the promoter activity of *Kir6.2* (Ashfield and Ashcroft 1998). A third example is the presence of estrogen receptor-dependent enhancers in *Alu* elements which belong to a peculiar subcategory of *Alu*-S elements (Norris et al. 1995). These *Alu* elements have to be considered as important contributors to the regulation of gene transcription in estrogen receptor-containing cells.

In conclusion, we show that an *Alu*-Y element is inserted in an orthologous position in intron 1 of human, chimpanzee, gorilla, gibbon, and orangutan *FUT1* genes. This *Alu*-Y element, which lies in a region crucial for the regulation of gene expression, might be responsible for the expression of the H enzyme in the erythroid cell lineage. The expression of the H activity in this lineage determines the possibility to express ABO antigens on red cells in humans and anthropoid apes.

Other expression regulation mechanisms will have to be found in order to explain the differential expression of ABH antigens on vascular endothelium, since, in addition to humans and apes, Old World monkeys such as macaques (Socha et al. 1987) and baboons (Oriol et al. 1984) express ABH in this tissue location, while New World monkeys such as marmosets (Mollicone et al. 1986) and all nonprimate mammals do not express ABH on vascular endothelial cells. This tissue is of particular clinical importance, since the vascular endothelial cell is the first target for the hyperacute vascular rejection of discordant organ xenotransplants. Understanding the addressing mechanisms responsible for the expression of ABH on vascular endothelium might help, for instance, to obtain pig organs expressing ABH on vascular endothelium, which could consequently serve as xenogeneic donor organs not rejected by human anti-αGal antibodies, as normal pig organs are.

Acknowledgments

This work was supported by funds from Agence Française du Sang (contract 97002551231); by private funds from the Etablissement de Transfusion Sanguine Pyrénées Garonne, Site de Toulouse, France; by the network ''GT-rec''; and by the Immunology Biotechnology program DG XII (contract IO4-CT97-2242) of the European Union (EU). We thank N. Herrenschmidt for monkey blood samples and O. Blumenfeld for the gift of orangutan DNA and RNA samples.

LITERATURE CITED

AKIYAMA, Y. 1995. TFSEARCH: searching transcription factor binding sites. Kyoto University, Kyoto, Japan.

ARCOT, S. S., A. W. ADAMSON, G. W. RISCH, J. LAFLEUR, M. B. ROBICHAUX, J. E. LAMERDIN, A. V. CARRANO, and M. A. BATZER. 1998. High-resolution cartography of recently integrated human chromosome 19-specific *Alu* fossils. J. Mol. Biol. **281**:843–856.

ASHFIELD, R., and S. J. ASHCROFT. 1998. Cloning of the promoters for the beta-cell ATP-sensitive K-channel subunits Kir6.2 and SUR1. Diabetes **47**:1274–1280.

BLANCHER, A., and W. W. SOCHA. 1997. The ABO Hh and Lewis blood group in humans and nonhuman primates. Pp. 30–92 *in* A. BLANCHER, J. KLEIN, and W. W. SOCHA, eds. Molecular biology and evolution of blood group and MHC antigens in primates. Springer-Verlag, Berlin.

BRETON, C., R. ORIOL, and A. IMBERTY. 1996. Sequence alignment and fold recognition of fucosyltransferases. Glycobiology **6**:vii–xii.

———. 1998. Conserved structural features in Eukaryotic and Prokaryotic fucosyltransferases. Glycobiology **8**:87–94.

COSTACHE, M., P. A. APOIL, A. CAILLEAU, A. ELMGREN, G. LARSON, S. HENRY, A. BLANCHER, D. IORDACHESCU, R. ORIOL, and R. MOLLICONE. 1997*a*. Evolution of fucosyltransferase genes in vertebrates. J. Biol. Chem. **272**:29721–29728.

COSTACHE, M., A. CAILLEAU, P. FERNANDEZ-MATEOS, R. ORIOL, and R. MOLLICONE. 1997*b*. Advances in molecular genetics of fucosyltransferases: H, secretor, Lewis, plasma and leukocyte gene products. Transfus. Clin. Biol. **4**:365–380.

DEVRIES, T., C. A. SRNKA, M. M. PALCIC, S. J. SWIEDLER, D. H. VAN DEN EIJNDEN, and B. A. MACHER. 1995. Acceptor specificity of different length constructs of human recombinant α 1,3/4-fucosyltransferases—replacement of the stem region and the transmembrane domain of fucosyltransferase V by protein A results in an enzyme with GDP-fucose hydrolyzing activity. J. Biol. Chem. **270**:8712–8722.

FERNANDEZ-MATEOS, P., A. CAILLEAU, M. COSTACHE, S. HENRY, A. ELMGREN, G. LARSON, B. E. SAMUELSSON, R. ORIOL, and R. MOLLICONE. 1998. Point mutations and deletion responsible for Bombay H null and Reunion H weak blood groups. Vox Sang. **75**:37–46.

HENRY, S., R. MOLLICONE, P. FERNANDEZ, B. SAMUELSSON, R. ORIOL, and G. LARSON. 1996*a*. Molecular basis for erythrocyte Le(a+b+) and salivary ABH partial-secretor phenotypes: expression of a *FUT2* secretor allele with an A→T mutation at nucleotide 385 correlates with reduced α(1,2)fucosyltransferase activity. Glycoconj. J. **13**:985–993.

HENRY, S., R. MOLLICONE, J. B. LOWE, B. SAMUELSSON, and G. LARSON. 1996*b*. A second nonsecretor allele of the blood group α(1,2)fucosyltransferase gene (*FUT2*). Vox Sang. **70**: 21–25.

HITOSHI, S., S. KUSUNOKI, I. KANAZAWA, and S. TSUJI. 1995. Molecular cloning and expression of two types of rabbit β-galactoside α1,2-fucosyltransferase. J. Biol. Chem. **270**: 8844–8850.

———. 1996. Molecular cloning and expression of a third type of rabbit GDP-L-fucose: β-D-galactoside 2-α-L-fucosyltransferase. J. Biol. Chem. **271**:16975–16981.

JOHNSON, P. H., M. K. MAK, S. LEONG et al. (12 co-authors). 1994. Analysis of mutations in the blood group H gene in donors with H-deficient phenotypes. Vox Sang. **67**(Suppl. 2):25 [abstract].

JURKA, J., P. KLONOWSKI, V. DAGMAN, and P. PELTON. 1996. CENSOR—a program for identification and elimination of repetitive elements from DNA sequences. Comput. Chem. **20**:119–122.

KANEKO, M., S. NISHIHARA, N. SHINYA, T. KUDO, H. IWASAKI, T. SENO, Y. OKUBO, and H. NARIMATSU. 1997. Wide variety of point mutations in the H gene of Bombay and para-Bom-

bay individuals that inactivate H enzyme. Blood **90**:839–849.

KAPITONOV, V., and J. JURKA. 1996. The age of *Alu* subfamilies. J. Mol. Evol. **42**:59–65.

KELLY, R. J., L. K. ERNST, R. D. LARSEN, J. G. BRYANT, J. S. ROBINSON, and J. B. LOWE. 1994. Molecular basis for H blood group deficiency in Bombay (Oh) and Para-Bombay individuals. Proc. Natl. Acad. Sci. USA **91**:5843–5847.

KELLY, R. J., S. ROUQUIER, D. GIORGI, G. G. LENNON, and J. B. LOWE. 1995. Sequence and expression of a candidate for the human secretor blood group α(1,2)fucosyltransferase gene (*FUT2*)—homozygosity for an enzyme-inactivating nonsense mutation commonly correlates with the non-secretor phenotype. J. Biol. Chem. **270**:4640–4649.

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.

KODA, Y., M. SOEJIMA, P. H. JOHNSON, E. SMART, and H. KIMURA. 1997*a*. Missense mutation of *FUT1* and deletion of *FUT2* are responsible for Indian Bombay phenotype of ABO blood group system. Biochem. Biophys. Res. Commun. **238**:21–25.

KODA, Y., M. SOEJIMA, and H. KIMURA. 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.

———. 1998. Changing transcription start sites in H-type α(1,2)fucosyltransferase gene (*FUT1*) during differenciation of the human erythroid lineage. Eur. J. Biochem. **256**: 379–387.

KODA, Y., M. SOEJIMA, Y. LIU, and H. KIMURA. 1996. Molecular basis for secretor type α(1,2)fucosyltyransferase gene deficiency in Japanese population: a fusion gene generated by unequal crossover responsible for a nonsecretor phenotype. Am. J. Hum. Genet. **59**:343–350.

KODA, Y., M. SOEJIMA, B. J. WANG, and H. KIMURA. 1997*b*. Structure and expression of the gene encoding secretor-type galactoside 2-α-L-fucosyltransferase (*FUT2*). Eur. J. Biochem. **246**:750–755.

KUMAR, S., K. TAMURA, and M. NEI. 1993. MEGA: molecular evolutionary genetics analysis. Version 1.01. Pennsylvania State University, University Park.

LARSEN, R. D., L. K. ERNST, R. P. NAIR, and J. B. LOWE. 1990. Molecular cloning, sequence and expression of a human GDP-L-fucose: β-D-galactoside α2-L-fucosyltransferase cDNA that can form the H blood group antigen. Proc. Natl. Acad. Sci. USA **87**:6674–6678.

LIU, Y., Y. KODA, M. SOEJIMA, H. PANG, T. SCHLAPHOFF, E. D. DU TOIT, and H. KIMURA. 1998. Extensive polymorphism of the *FUT2* gene in an African (Xhosa) population of South Africa. Hum. Genet. **103**:204–210.

MOLLICONE, R., D. R. DAVIES, B. EVANS, A. M. DALIX, and R. ORIOL. 1986. Cellular expression and genetic control of ABH antigens in primary sensory neurons of marmoset, baboon and man. J. Neuroimmunol. **10**:255–269.

MOOR-JANKOWSKI, J., and A. WIENER 1964. Human blood group factors in non-human primates. Nature **202**:663–665.

NORRIS, J., D. FAN, C. ALEMAN, J. R. MARKS, P. A. FUTREAL, R. W. WISEMAN, J. D. IGLEHART, P. L. DEININGER and D. P. MCDONNELL. 1995. Identification of a new subclass of Alu DNA repeats which can function as estrogen receptor-dependent transcriptional enhancers. J. Biol. Chem. **270**: 22777–22782.

ORIOL, R. 1995. ABO, Hh, Lewis and secretion. Serology, genetics and tissue distribution. Pp. 37–73 *in* J. P. CARTRON and P. ROUGER, eds. Molecular basis of major blood group antigens. Vol. 6. Plenum Press, London.

ORIOL, R., J. E. COOPER, D. R. DAVIES, and P. W. N. KEELING. 1984. ABH antigens in vascular endothelium and some epithelial tissues of baboons. Lab. Invest. **50**:514–518.

ORIOL, R., J. DANILOVS, and B. R. HAWKINS. 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., R. MOLLICONE, A. CAILLEAU, L. BALANZINO, and C. BRETON. 1999. Divergent evolution of fucosyltransferase genes from vertebrates, invertebrates and bacteria. Glycobiology **9**:323–334.

ORIOL, R., R. MOLLICONE, P. COUILLIN, A. M. DALIX, and J. J. CANDELIER. 1992. Genetic regulation of the expression of ABH and Lewis antigens in tissues. APMIS **100**:28–38.

PIEDRAFITA, F. J., R. B. MOLANDER, G. VANSANT, E. A. ORLOVA, M. PFAHL, and W. F. REYNOLDS. 1996. An *Alu* element in the myeloperoxidase promoter contains a composite SP1-thyroid hormone-retinoic acid response element. J. Biol. Chem. **271**:14412–14420.

RAHUEL, C., M. A. VINIT, V. LEMARCHANDEL, J. P. CARTRON, and P. H. ROMÉO. 1992. Erythroid-specific activity of the glycophorin B promoter requires GATA-1 mediated displacement of a represor. EMBO J. **11**:4095–4102.

REGUIGNE-ARNOULD, I., P. COUILLIN, R. MOLLICONE, S. FAURE, A. FLETCHER, R. J. KELLY, J. B. LOWE, and R. ORIOL. 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. **71**:158–162.

ROUQUIER, S., J. B. LOWE, R. J. KELLY, A. L. FERTITTA, G. C. LENNON, and D. GIORGI. 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.

SHAIKH, T. H., and P. L. DEININGER. 1996. The role and amplification of the HS *Alu* subfamily founder gene. J. Mol. Evol. **42**:15–21.

SOCHA, W. W., C. C. MARBOE, R. E. MICHLER, E. A. ROSE, and J. MOOR-JANKOSWKI. 1987. Primate animal model for the study of ABO incompatibility in organ transplantation. Transplant. Proc. **6**:4448–4455.

THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 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.

VIGNAL, A., J. LONDON, C. RAHUEL, and J. P. CARTRON. 1990. Promoter sequence and chromosomal organization of the genes encoding glycophorins A, B, and E. Gene **95**:289–293.

WAGNER, F. F., and W. A. FLEGEL. 1997. Polymorphism of the *h* allele and the population frequency of sporadic nonfunctional alleles. Transfusion **37**:284–290.

WANG, B., Y. KODA, M. SOEJIMA, and H. KIMURA. 1997. Two missense mutations of H type α(1,2) fucosyltransferase gene (*FUT1*) responsible for para-Bombay. Vox Sang. **72**: 31–35.

XU, X. H., L. VO, and B. A. MACHER. 1996. Structure-function analysis of human α1,3-fucosyltransferase—amino acids involved in acceptor substrate specificity. J. Biol. Chem. **271**: 8818–8823.

YU, L. C., R. E. BROADBERRY, Y. H. YANG, Y. H. CHEN, and M. LIN. 1996. Heterogeneity of the human *secretor*

α(1,2)fucosyltransferase gene among Lewis(a+b−) non-secretors. Biochem. Biophys. Res. Commun. **222**:390–394.

YU, L. C., H. L. LEE, C. C. CHU, R. E. BROADBERRY, and M. LIN. 1999. A newly identified nonsecretor allele of the human histo-blood-group α(1.2)fucosyltransferase gene (*FUT2*). Vox Sang. **76**:115–119.

YU, L. C., Y. H. YANG, R. E. BROADBERRY, Y. H. CHEN, and M. LIN. 1997. Heterogeneity of the human *H* blood group α(1,2)fucosyltransferase gene among para-Bombay individuals. Vox Sang. **72**:36–40.

PIERRE CAPY, reviewing editor

Accepted November 1, 1999