J Mol Evol (2009) 69:22–31
DOI 10.1007/s00239-009-9239-0

Widespread Gene Conversion of Alpha-2-Fucosyltransferase Genes in Mammals

Joana Abrantes · David Posada · Patrice Guillon · Pedro J. Esteves · Jacques Le Pendu

Received: 13 December 2008 / Accepted: 15 April 2009 / Published online: 16 June 2009

Abstract The alpha-2-fucosyltransferases (α2FTs) are enzymes involved in the biosynthesis of α2fucosylated glycan structures. In mammalian genomes, there are three α2FT genes located in tandem—*FUT*1, *FUT*2, and *Sec*1—each contained within a single exon. It has been suggested that these genes originated from two successive duplications, with *FUT*1 being generated first and *FUT*2 and *Sec*1 second. Despite gene conversion being considered the main mechanism of concerted evolution in gene families, previous studies of primates α2FTs failed to detect it, although the occurrence of gene conversion between *FUT*2 and *Sec*1 was recently reported in a human allele. The primary aim of our work was to initiate a broader study on the molecular evolution of mammalian α2FTs. Sequence comparison leads us to confirm that the three genes appeared by two rounds of duplication. In addition, we were able to detect multiple gene-conversion events at the base of primates and within several nonprimate species involving *FUT*2 and *Sec*1. Gene conversion involving *FUT*1 and either *FUT*2 or *Sec*1 was also detected in rabbit. The extent of gene conversion between the α2FTs genes appears to be species-specific, possibly related to functional differentiation of these genes. With the exception of rabbits, gene conversion was not observed in the region coding the C-terminal part of the catalytic domain. In this region, the number of amino acids that are identical between *FUT*1 and *FUT2,* but different in *Sec1,* is higher than in other parts of the protein. The biologic meaning of this observation may be related to functional constraints.

Keywords Fucosyltransferases · Histo-blood group antigens · Gene conversion

Abbreviations

α2FTs	Alpha1,2fucosyltransferases
GDP-Fuc	Guanosyldiphosphate fucose
Fuc	Fucose
Gal	Galactose
GlcNAc	*N*-Acetylglucosamine
GalNAc	*N*-Acetylgalactosamine

Introduction

Alpha-2-fucosyltransferases (α2FTs) are enzymes required for the biosynthesis of the terminal glycan motif Fucα2-Galβ-R found in ABH and Lewis histo-blood group

Electronic supplementary material The online version of this article (doi:10.1007/s00239-009-9239-0) contains supplementary material, which is available to authorized users.

J. Abrantes · P. J. Esteves
CIBIO-UP, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, Rua Padre Armando Quintas, Vairao, Portugal

J. Abrantes · P. Guillon · J. Le Pendu (✉)
INSERM, U892, Institut de Biologie, Université de Nantes, 9 Quai Moncousu, 44093 Nantes, Cedex 01, France
e-mail: jlependu@nantes.inserm.fr

J. Abrantes
Departamento de Zoologia e Antropologia, Faculdade de Ciências da Universidade do Porto, Porto, Portugal

D. Posada
Departamento de Bioquímica, Genética e Inmunología, Facultad de Biología, Campus Lagoas-Marcosende, Universidad de Vigo, Vigo 36310, Spain

P. J. Esteves
CITS, Centro de Investigação em Tecnologias da Saúde, IPSN, CESPU, Gandra, Portugal

antigens. The biologic functions of such carbohydrate motifs are not completely clear yet, but their main expression at the surface of epithelial cells that constitute doors of entry for pathogens, as well as on soluble mucins present in these epithelia, suggests that they might have functions related to interactions with microorganisms, pathogenic or not (Marionneau et al. 2001). Consistent with this view, *Helicobacter pylori* strains, *Campilobacter pilori*, uropathogenic strains of *Escherichia coli*, lactobacilli strains, and several strains of *Caliciviruses* are known to attach to α2fucosylated glycan structures (Boren et al. 1993; Le Pendu et al. 2006; Ruiz-Palacios et al. 2003; Uchida et al. 2006). Additional cellular functions—such as involvement in the development of the olfactory system, angiogenesis, interaction between dendritic cells and the vascular endothelium, and regulation of apoptosis (Garcia-Vallejo et al. 2008; Halloran et al. 2000; Moehler et al. 2008; St. John et al. 2006)—recently have been suggested. Like other glycosyltransferases, α2FTs are type II membrane proteins anchored in the Golgi apparatus. They present a short intracytoplasmic tail and a transmembrane domain in N-terminal location, followed by a stem region and the catalytic domain, which can be subdivided into two subdomains, the N- and C-terminal subdomains.

In mammalian genomes, three α2FT genes are located in tandem and designated as *FUT*1, *FUT2,* and *Sec*1 (Oriol et al. 2000). The coding sequence of each of the three genes is comprised within a single exon, and it has been suggested that this monoexonic structure results from an L1-retrotransposition event that occurred within the α2FT mammalian ancestor gene (Saunier et al. 2001). Their tandem localization and earlier sequence comparisons using only primates suggested that the *FUT*1, *FUT*2, and *Sec*1 genes originated from two successive duplications. The first one would have given rise to *FUT*1 and to the ancestor of both *FUT*2 and *Sec*1, whilst the second duplication event would have generated *FUT*2 and *Sec*1.

Gene duplication is generally considered important for adaptation because it allows advantageous mutations in one of the duplicates to promote a new role without impairing the original function exerted by the other duplicate (Ohno 1970). However, theoretical studies suggest that one of the duplicates has a high probability to become silenced rapidly (Walsh 1995), which would be consistent with the silencing of *Sec*1 in Catharrinians (Apoil et al. 2000). In addition, it has been proposed that functional diversification is a rare event because gene conversion tends to homogenize the variation between duplicated genes (Walsh 1987). For that reason, gene conversion is considered the main mechanism of concerted evolution of gene families. Previous studies have not detected gene conversion in primates because each of the *FUT*1, *FUT*2, and *Sec*1 genes appeared as separated clusters (Apoil et al. 2000); however, very recently the occurrence of gene conversion between *FUT*2 and *Sec*1 was reported in a *Sec1-FUT2-Sec*1 human allele (Soejima et al. 2008).

Here, with the aim to improve our understanding of the evolution and the biologic role of the α2FTs gene family, we addressed two main questions. First, what are the evolutionary relations between *FUT*1, *FUT*2, and *Sec*1 when looking at a broader phylogenetic context? Second, is there evidence of gene-conversion events between these three genes?

Materials and Methods

Complete coding sequences of α2FTs *FUT*1, *FUT*2, and *Sec*1 genes were retrieved from GenBank and aligned with Clustal W (Thompson et al. 1994), followed by visual inspection (see Table 1 for a list of species used, abbreviations, and GenBank accession numbers).

Phylogenetic relations between mammalian *FUT*1, *FUT*2, and *Sec*1 genes were analyzed using the entire catalytic domain, corresponding to nucleotide positions 235 to 1083, 184 to 1026, and 193 to 1033 of the human *FUT*1, *FUT*2, and *Sec*1 sequences, respectively. Only the catalytic domains of these enzymes were used because the three genes are highly divergent for the transmembrane and stem regions, which could not be aligned with confidence. The optimal model of sequence evolution was estimated using the ModelTest web server (Posada 2006). This model was then used to estimate a maximum likelihood (ML) tree using Phyml (Guindon and Gascuel 2003). The software GARD (Kosakovsky-Pond et al. 2006a, b) was used to detect possible phylogenetic incongruences, such as those due to gene conversion. ML trees were estimated for each of the segments identified by GARD and compared using the Shimodaira–Hasegawa (SH) (Shimodaira and Hasegawa 1999) test implemented in PAUP* (Swofford 2000).

In addition, the program Geneconv (Sawyer 1989) was used to confirm the conversion events inferred by visual inspection of the phylogenetic trees recovered for each fragment. Geneconv looks for aligned segments in which pairs of sequences are similar enough to be suggestive of past gene conversion. The programs finds and ranks the highest-scoring fragments globally for the entire alignment ("global" fragments), and if specified, also for each sequence pair ("pairwise" fragments). *p* values are obtained by permutation (in this case 10,000); however, whereas global *p*-values compare each fragment with all possible fragments for the entire alignment, pairwise *p*-values compare each fragment with the maximum that might have been expected for that sequence pair in the absence of gene conversion. Global fragments have

Table 1 List of the coding sequences of the α2FTs genes, *FUT*1, *FUT*2 and *Sec*1, retrieved from GenBank and used in this study

Species name	Common name	Gene	GenBank accession no.	Abbreviation
Homo sapiens	Human	*FUT*2	U17894	Human_FUT2
		*FUT*1	M35531	Human_FUT1
		*Sec*1	U17895	Human_Sec1
Pan troglodytes	Common chimpanzee	*FUT*2	AF080604	Chimpanzee_FUT2
		*FUT*1	AF080603	Chimpanzee_FUT1
		*Sec*1	AB006612	Chimpanzee_Sec1
Pongo pygmaeus	Orangutan	*FUT*2	AB015636	Orangutan_FUT2
		*Sec*1	AB006610	Orangutan_Sec1
Gorilla gorilla	Gorilla	*FUT*2	AF080606	Gorilla_FUT2
		*FUT*1	AF080605	Gorilla_FUT1
Hylobates lar	Lar gibbon	*FUT*2	AF136648	Gibbon_FUT2
H. agilis	Agile gibbon	*Sec*1	AB006609	Gibbon_Sec1
Chlorocebus aethiops sabaeus	Green monkey	*FUT*2	D87934	Green_monkey_FUT2
		*FUT*1	D87932	Green_monkey_FUT1
		*Sec*1	D87933	Green_monkey_Sec1
Macaca fascicularis	Cynomolgus	*Sec*1	AF080608	Cynomolgus_Sec1
Bos taurus	Cow	*FUT*2	X99620	Cow_FUT2
		*FUT*1	NM_177499	Cow_FUT1
		*Sec*1	AF187851	Cow_Sec1
Mus musculus	Mouse	*FUT*2	AF064792	Mouse_FUT2
		*FUT*1	U90553	Mouse_FUT1
		*Sec*1	Y09882	Mouse_Sec1
Rattus norvegicus	Rat	*FUT*2	AB006138	Rat_FUT2
		*FUT*1	AB015637	Rat_FUT1
		*Sec*1	AF131239	Rat_Sec1
Sus scrofa	Pig	*FUT*2	U70881	Pig_FUT2
		*FUT*1	U70883	Pig_FUT1
		*Sec*1	U70882	Pig_Sec1
Oryctolagus cuniculus	Rabbit	*FUT*2	X91269	Rabbit_FUT2
		*FUT*1	X80226	Rabbit_FUT1
		*Sec*1	X80225	Rabbit_Sec1
Xenopus tropicalis	Frog	*FUT*1	NM_001004772	Frog_FUT1
Monodelphis domestica	Opossum	*FUT2-like*	XM_001362239	Opossum_FUT2_like

p-values that are multiple comparison–corrected for all possible sequence pairs, whereas pairwise fragments have a built-in multiple comparison–correction for the length of the alignment. The program also distinguishes between "inner" fragments, i.e., gene-conversion events between ancestors of two sequences in the alignment, and "outer" fragments, i.e., evidence of past-gene conversion events that may have originated from outside of the alignment. A mismatch penalty was allowed (gscale = 1); therefore, conversion fragments did not have to be identical. ML trees were also obtained with Phyml from the amino acid sequences for each of the segments identified by GARD using the best-fit model suggested by ProtTest (Abascal et al. 2005).

Results

Phylogenetic analysis of the nucleotide sequences corresponding to the catalytic domain (Supplementary data) yielded evolutionary relations somewhat different from those previously published based on full protein sequences (Apoil et al. 2000; Barreaud et al. 2000; Bureau et al. 2001). The *FUT*1 sequences appeared as a well-supported basal clade. However, *FUT*2 and *Sec*1 sequences did not form clearly separated groups. For example, rabbit, rat, and mouse *FUT*2 sequences clustered, with high support, with their corresponding *Sec*1 sequences. This was also the case for the pig *FUT*2 and *Sec*1 sequences, but here the pair was embedded inside the main *FUT*2 group.

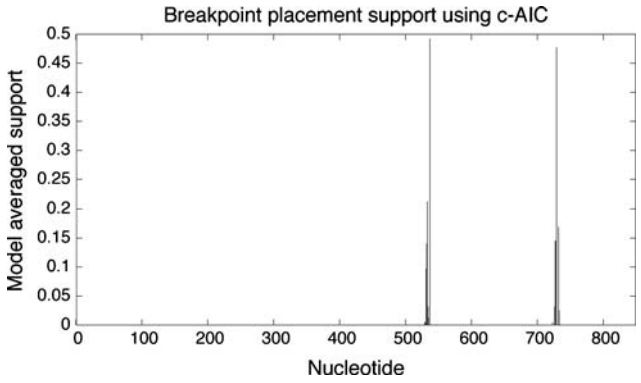

Fig. 1 cAIC model-averaged support for recombination break points as detected by GARD. Nucleotide position 1 in the graph corresponds to nucleotide position 184 of the human *FUT*2 sequence

GARD analyses indicated 2 highly supported recombination break points at positions 720 and 912 (Fig. 1). The resulting fragments were named segment A (nucleotide position 184–720); segment B (nucleotide position 721–912); and segment C (913–1026 human FUT2 nucleotide positions). According to SH test, the resulting phylogenetic trees for each segment were significantly different ($p < 0.001$). In segment A, such as for the complete sequences, *FUT*1 appears as a highly supported basal clade. Interestingly, the *FUT*2 and *Sec*1 sequences clustered by gene in primates and by species in nonprimates (Fig. 2a). For segment B, the *FUT*1 sequences did not form a clade (Fig. 2b). Although the major *FUT*2 and *Sec*1 groups were established, all of the rabbit sequences (*FUT*1, *FUT*2 and *Sec*1) formed a well-supported group, and the pig *Sec*1 sequence clustered with the pig *FUT*2 sequence inside the main *FUT*2 group. In segment C, the α2FTs formed three distinct clades (*FUT*1, (*FUT*2, *Sec*1)), although they did so with low bootstrap values (Fig. 2c).

Geneconv found 52 globally significant global inner fragments (Table 2) and 6 additional pairwise inner fragments (Table 3). The length of the estimated gene conversion tracts ranged from 177 to 521 base pairs (bp). Inclusion/exclusion of the outgroup resulted in similar inferences. Among the 52 inner fragments, Geneconv detected many significant conversion events that were quite consistent with the phylogenetic partition suggested by GARD and the corresponding trees (Fig. 2). Taking into account both sources of information, i.e., the phylogenetic incongruences and the Geneconv output, several gene-conversion events appear to have occurred between *FUT*2 and *Sec*1 in segment A. We can infer a gene-conversion event in segment A (nucleotides [nt] 208 to 559) before the diversification of primates. Note that all of the *FUT*2/*Sec*1 primate pairs in the Geneconv output (Table 2) group together. Additional but independent *FUT*2/*Sec*1 conversion events also seem to have occurred in this segment for cow (nt 193–608), rat (nt

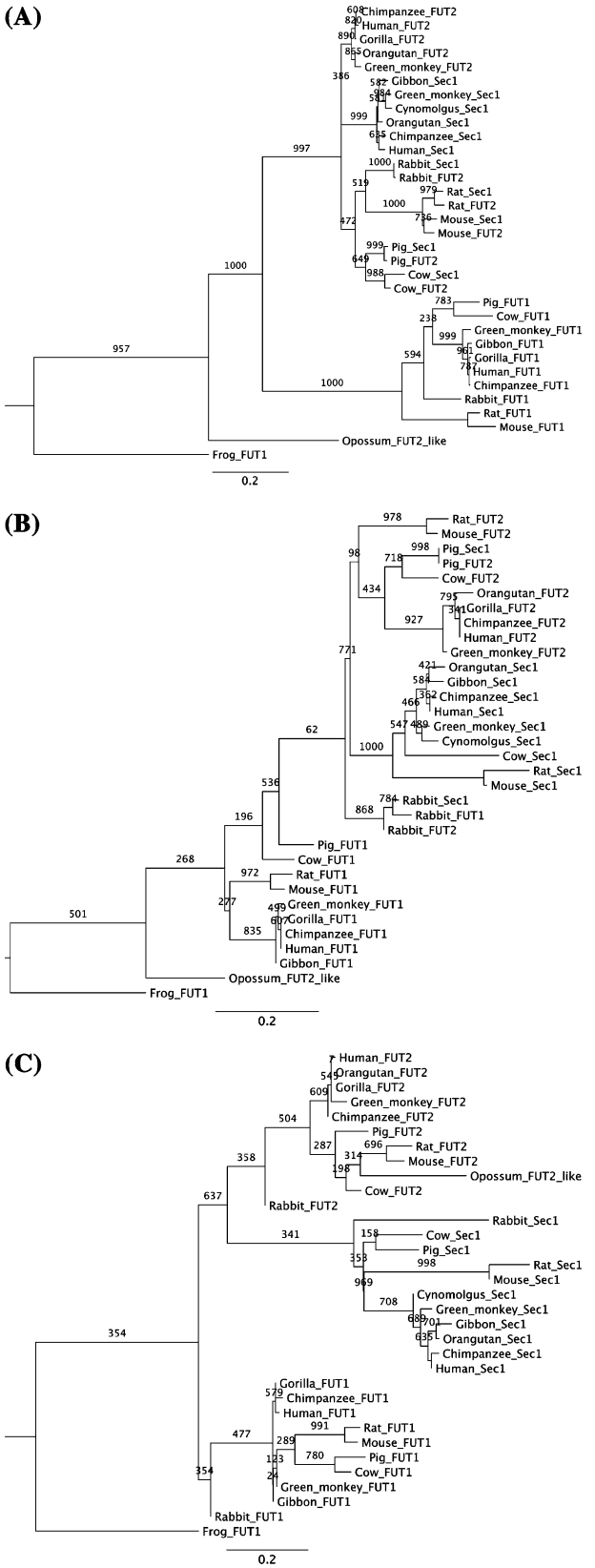

Fig. 2 ML trees for (**a**) segment A (nucleotides 184–720), (**b**) segment B (nucleotides 721–912), and (**c**) segment C (nucleotides 913–1032). Nucleotide positions are according to human *FUT*2 sequence

Table 2 List of global inner fragments (484 polymorphisms and 849 aligned bases) obtained with Geneconv where inner fragments are runs of matching sites with penalties

Sequences compared	Sim p[a]	BC KA p[b]	Aligned offsets			No. polymorphic[c]	No. differences[d]	Total differences[e]	Mismatch penalty[f]
			Begin	End	Length				
Cow_FUT2; Cow_Sec1	0.0000	0.00000	193	608	415	242	3	112	5
Rabbit_FUT1; Rabbit_Sec1	0.0000	0.00000	755	931	177	93	3	198	3
Rabbit_FUT1; Rabbit_FUT2	0.0000	0.00000	755	1023	269	152	14	180	3
Rat_FUT2; Rat_Sec1	0.0000	0.00000	208	725	521	291	18	112	5
Orangutan_FUT2; Orangutan_Sec1	0.0000	0.00000	208	559	352	201	11	134	4
Mouse_FUT2; Mouse_Sec1	0.0000	0.00000	307	728	422	241	9	110	5
Orangutan_FUT2; Gibbon_Sec1	0.0000	0.00000	208	559	352	201	16	140	4
Gorilla_FUT2; Orangutan_Sec1	0.0000	0.00000	208	559	352	201	11	131	4
Orangutan_FUT2; Green_monkey_Sec1	0.0000	0.00000	208	559	352	201	18	143	4
Orangutan_FUT2; Human_Sec1	0.0000	0.00000	208	559	352	201	14	135	4
Orangutan_FUT2; Chimpanzee_Sec1	0.0000	0.00000	208	559	352	201	14	135	4
Human_FUT2; Orangutan_Sec1	0.0000	0.00000	208	559	352	201	13	133	4
Green_monkey_FUT2; Green_monkey_Sec1	0.0000	0.00000	208	559	352	201	18	142	4
Green_monkey_FUT2; Gibbon_Sec1	0.0000	0.00000	208	559	352	201	18	142	4
Green_monkey_FUT2; Orangutan_Sec1	0.0000	0.00000	208	559	352	201	13	132	4
Gorilla_FUT2; Gibbon_Sec1	0.0000	0.00000	208	559	352	201	18	141	4
Gorilla_FUT2; Human_Sec1	0.0000	0.00000	208	559	352	201	12	130	4
Gorilla_FUT2; Chimpanzee_Sec1	0.0000	0.00000	208	559	352	201	12	130	4
Human_FUT2; Human_Sec1	0.0000	0.00000	208	559	352	201	14	132	4
Human_FUT2; Chimpanzee_Sec1	0.0000	0.00000	208	559	352	201	14	132	4
Mouse_FUT2; Rat_Sec1	0.0000	0.00000	241	728	488	279	29	131	4
Chimpanzee_FUT2; Orangutan_Sec1	0.0000	0.00000	232	559	328	193	12	131	4
Human_FUT2; Gibbon_Sec1	0.0000	0.00000	208	559	352	201	20	143	4
Green_monkey_FUT2; Human_Sec1	0.0000	0.00000	208	559	352	201	16	134	4
Green_monkey_FUT2; Chimpanzee_Sec1	0.0000	0.00000	208	559	352	201	16	134	4
Green_monkey_FUT2; Cynomolgus_Sec1	0.0000	0.00001	208	559	352	201	18	136	4
Orangutan_FUT2; Cynomolgus_Sec1	0.0000	0.00001	208	559	352	201	18	136	4
Gorilla_FUT2; Green_monkey_Sec1	0.0000	0.00001	208	559	352	201	21	143	4
Chimpanzee_FUT2; Chimpanzee_Sec1	0.0000	0.00001	208	559	352	201	14	129	4
Chimpanzee_FUT2; Human_Sec1	0.0000	0.00002	208	559	352	201	15	130	4
Chimpanzee_FUT2; Gibbon_Sec1	0.0000	0.00002	232	559	328	193	19	141	4
Human_FUT2; Green_monkey_Sec1	0.0000	0.00003	208	559	352	201	23	147	4
Chimpanzee_FUT2; Green_monkey_Sec1	0.0000	0.00011	232	559	328	193	22	145	4
Human_FUT2; Cynomolgus_Sec1	0.0000	0.00012	208	559	352	201	22	140	4
Gorilla_FUT2; Cynomolgus_Sec1	0.0000	0.00024	208	559	352	201	22	138	4
Chimpanzee_FUT2; Cynomolgus_Sec1	0.0000	0.00044	232	559	328	193	21	138	4
Gorilla_FUT2; Cow_Sec1	0.0000	0.00044	207	569	363	207	26	145	4
Human_FUT2; Cow_Sec1	0.0000	0.00059	207	569	363	207	27	148	4
Orangutan_FUT2; Cow_Sec1	0.0000	0.00099	207	551	345	196	25	147	4
Rat_FUT2; Mouse_Sec1	0.0000	0.00181	307	728	422	241	23	121	4
Cow_FUT2; Orangutan_Sec1	0.0001	0.00335	208	551	344	195	31	162	3
Green_monkey_FUT2; Cow_Sec1	0.0002	0.00554	365	569	205	118	12	153	4
Chimpanzee_FUT2; Cow_Sec1	0.0002	0.00657	232	569	338	198	27	147	4
Cow_FUT2; Chimpanzee_Sec1	0.0005	0.01069	208	551	344	195	33	164	3
Pig_FUT2; Cow_Sec1	0.0009	0.01627	208	569	362	206	27	137	4
Pig_FUT2; Chimpanzee_Sec1	0.0032	0.03161	253	551	299	176	30	166	3

Table 2 continued

Sequences compared	Sim p[a]	BC KA p[b]	Aligned offsets			No. polymorphic[c]	No. differences[d]	Total differences[e]	Mismatch penalty[f]
			Begin	End	Length				
Pig_FUT2; Orangutan_Sec1	0.0056	0.05048	208	551	344	195	35	165	3
Pig_FUT2; Human_Sec1	0.0093	0.10173	208	551	344	195	35	162	3
Cow_FUT2; Gibbon_Sec1	0.0112	0.11739	208	551	344	195	37	170	3
Cow_FUT2; Human_Sec1	0.0117	0.12367	208	426	219	127	17	160	4
Pig_FUT2; Gibbon_Sec1	0.0126	0.14306	208	551	344	195	37	169	3
Cow_FUT2; Green_monkey_Sec1	0.0211	0.22221	349	551	203	122	19	168	3

Only fragments with Sim $p \leq 0.05$ are listed

[a,b] Corrected for multiple comparisons

[b] Bonferroni-corrected KA (BLAST-like) p, where KA p is not corrected for multiple pairwise comparison. BC p is KA $p \times 465$

[c] Number of polymorphic sites in the fragment

[d] Number of mismatches within the fragment

[e] Total number of mismatches between two sequences

[f] Penalty per mismatch for the two sequences

Table 3 Additional pairwise fragments obtained by Geneconv with BC Pairwise SimPval <0.05 or listed global fragments with significantly better BC SimPval (≤3 pairwise fragments considered per pair)

Sequences compared	BC Sim p[a]	BC KA p[b]	Aligned offsets			No. polymorphic[c]	No. differences[d]	Tot differences[e]	Mismatch penalty[f]
			Begin	End	Length				
Orangutan_FUT2; Pig_Sec1	0.0465	>1.0	207	479	273	154	19	127	4
Rabbit_FUT2; Rabbit_Sec1	0.0465	>1.0	184	881	698	393	5	39	13
Pig_FUT2; Gibbon_Sec1	0.0465	0.14306	208	551	344	195	37	169	3
Pig_FUT2; Pig_Sec1	0.0465	0.26279	208	929	722	397	2	42	12
Cow_FUT2; Human_Sec1	0.0465	0.12367	208	426	219	127	17	160	4
Cow_FUT2; Gibbon_Sec1	0.0465	0.11739	208	551	344	195	37	170	3

[a] Pairwise Sim $p \times 465$

[b] Bonferroni-corrected KA (BLAST-like) p where KA p is not corrected for multiple pairwise comparisons. BC p is KA $p \times 465$

[c] Number of polymorphic sites in the fragment

[d] Number of mismatches within the fragment

[e] Total number of mismatches between two sequences

[f] Penalty per mismatch for the two sequences

208–728), and mouse (nt 307–728). Another *FUT*2/*Sec*1 conversion could have occurred in segment A before the rat and mouse split (nt 241–728). It is possible also to infer two events that imply both segments A and B in pig (nt 208–929) and in segments A, B and C in the rabbit (nt 184–881). These two events are obvious from the trees but only appear in the pairwise inner fragment list provided by Geneconv (Table 3), probably as a consequence of their large mismatch penalties or because overlapping events occurred in segment B. Therefore, we can infer two more events in rabbit, a *FUT*1-to-*Sec*1 (nt 755–931) and a *FUT*1-to-*FUT*2 (nt 755–1023) conversion. Some of these events are highlighted in grey in the amino acid alignment shown in Fig. 3.

ML trees were estimated at the peptide level for each of the A, B, and C segments of the catalytic domain (amino acids 62–342 of the human *FUT*2 enzyme). The trees obtained for each of the segments (data not shown) were consistent with the ones obtained at the nucleotide level. However, when examining the amino acid alignment (Fig. 3), we noted that the number of sites that are identical between *FUT*1 and *FUT*2, but that differ in *Sec*1, is greater in the C-terminal (which includes both B and C segments from amino acid position 237–342) than in the N-terminal subdomain (positions 62–236), with the exception of the *Sus* and *Oryctolagus* sequences that suffered gene conversion.

Discussion

Our results confirm the evolutionary scenario for the origin of α2FT genes previously reported for primates (Apoil et al. 2000). As proposed by Apoil et al., two duplication events could explain the emergence of these genes: An ancestral duplication event originated *FUT*1 and the ancestor of *FUT*2 and *Sec*1, and the ancestor of *FUT*2 and *Sec*1 duplicated and originated the *FUT*2 and *Sec*1 genes.

The idea that in mammals gene conversion between α2FT genes is rare (Apoil et al. 2000) is severely challenged by our results. When only primates are considered, gene conversion is not apparent because *FUT*1, *FUT*2, and *Sec*1 cluster by gene, creating three independent clusters (Apoil et al. 2000); however, gene conversion still could be detected with statistical methods as those implemented in Geneconv. Indeed, when we include other α2FT mammal genes and use GARD and Geneconv, we can readily see that gene conversion between *FUT*2 and *Sec*1 has been common. An inspection of the trees, with distinct clades formed by conspecific *FUT*2 and *Sec*1 sequences, and the different segments detected by Geneconv, suggests that multiple independent events of gene conversion occurred in the evolution of the α2FT gene family in mammals. The conversion events have different lengths and can span the three different segments previously detected with the GARD software. Note that the two phylogenetic break points detected maximize the phylogenetic disagreement and not the exact limits of the conversion events.

In our analyses, multiple gene-conversion fragments involving primates were detected by Geneconv in segment A (30 of 52). Given that all of the *Sec*1 and *FUT*2 primate sequences form a single group in the tree, although defined by a very short branch with a small bootstrap value, all of these fragments might be parsimoniously explained by a single gene-conversion event in the ancestor of primates. Evolution after this conversion event, with accumulation of specific mutations in each gene, would explain why they cluster by gene. The clustering of all the other mammals by species and not by gene suggests an ongoing gene-conversion process between *FUT*2 and *Sec*1 within species. In addition, the position of the opossum sequence in segment C within the nonprimate *FUT*2 clade was not expected, although the use of a small segment (120 nt) and the low bootstrap values suggest that this particular result may not be reliable.

Indeed, the fact that the three genes are located within <80 kb in the same chromosome prompts gene conversion. In addition, the gene-conversion events may be related to the biologic role of the α2FTs. The *Sec*1 gene is inactivated in many primate species, both in Old World and New World lineages, by a premature stop codon (Apoil et al. 2000; Borges et al. 2008). This gene has also been shown to be inactivated in pig and mouse (Iwamori and Domino 2004), and we recently observed similar evidences in rabbits, where although some *Sec*1 alleles show residual enzyme activity, most are inactive (Guillon et al. 2009). In these species, however, no premature stop codon was observed. Altogether, these observations suggest that *Sec*1 is either a pseudogene or that it is on the way to pseudogenisation. At variance, both *FUT*1 and *FUT*2 are active in all mammalian species tested so far (Oriol et al. 2000). The fact that the proportion of sites identical between *FUT*1 and *FUT*2, but different in *Sec*1, is higher in the C-terminal subdomain than in the N-terminal domain, and the fact that for most of the species, gene conversion is limited to the N-terminal, suggests that the enzymes must maintain the ancestral characteristics for this particular region (this part of the enzymes most likely resembles the ancestral enzyme that gave rise to this protein family), probably because they require some structural identity to preserve their functionality. The structure and mechanisms of fucosyltransferases are as yet unknown. Nevertheless, based on comparisons with many other glycosyltransferases, some predictions have been made (Breton et al. 1998, 2006). According to the models, the C-terminal region would correspond to the nucleotide binding domain, whereas the N-terminal part would correspond to the acceptor-binding domain. The latter is generally more variable than the former becasue it should accommodate a number of acceptor substrates much larger than the number of donor substrates. In the case of α2FTs, there is a single possible donor substrate, GDP-Fuc, whereas the number of acceptor substrates can be quite large. Indeed, albeit with different affinities, these enzymes use various acceptor substrates such as Galβ3GlcNAcβ-R, Galβ4GlcNAcβ-R, Galβ3GalNAcα-R, Galβ3GalNAcβ-R, and Galβ4Glcβ-R, where R represents the highly variable subjacent chains of glycolipids and of O-linked or N-linked glycan chains of glycoproteins. The redundant nature of *Sec*1 and the different functional constraints on the two regions of the catalytic domain of *FUT*1 and *FUT*2 would explain a higher similarity between these enzymes in the C-terminal part. A comparison of the synonymous and nonsynonymous divergences in both domains indicated that for the three proteins, dN/dS ratios were <1, suggesting that they are under purifying selection. Nevertheless, dN/dS ratios for the C-terminal domain are lower than for the N-terminal domain, consistent with our hypothesis of a higher functional constraint on the C-terminal domain. Concerning Sec1, dN/dS ratios were also <1, but they were higher than for either Fut1 and Fut2, which is at odds with the idea of *Sec*1 being a pseudogene. Indeed, if *Sec*1 were a pseudogene, it would be evolving neutrally, and dN/dS should be close to 1. Its deviation from neutrality suggests some functional constraints caused by the action of purifying selection. Nevertheless, these constraints appear

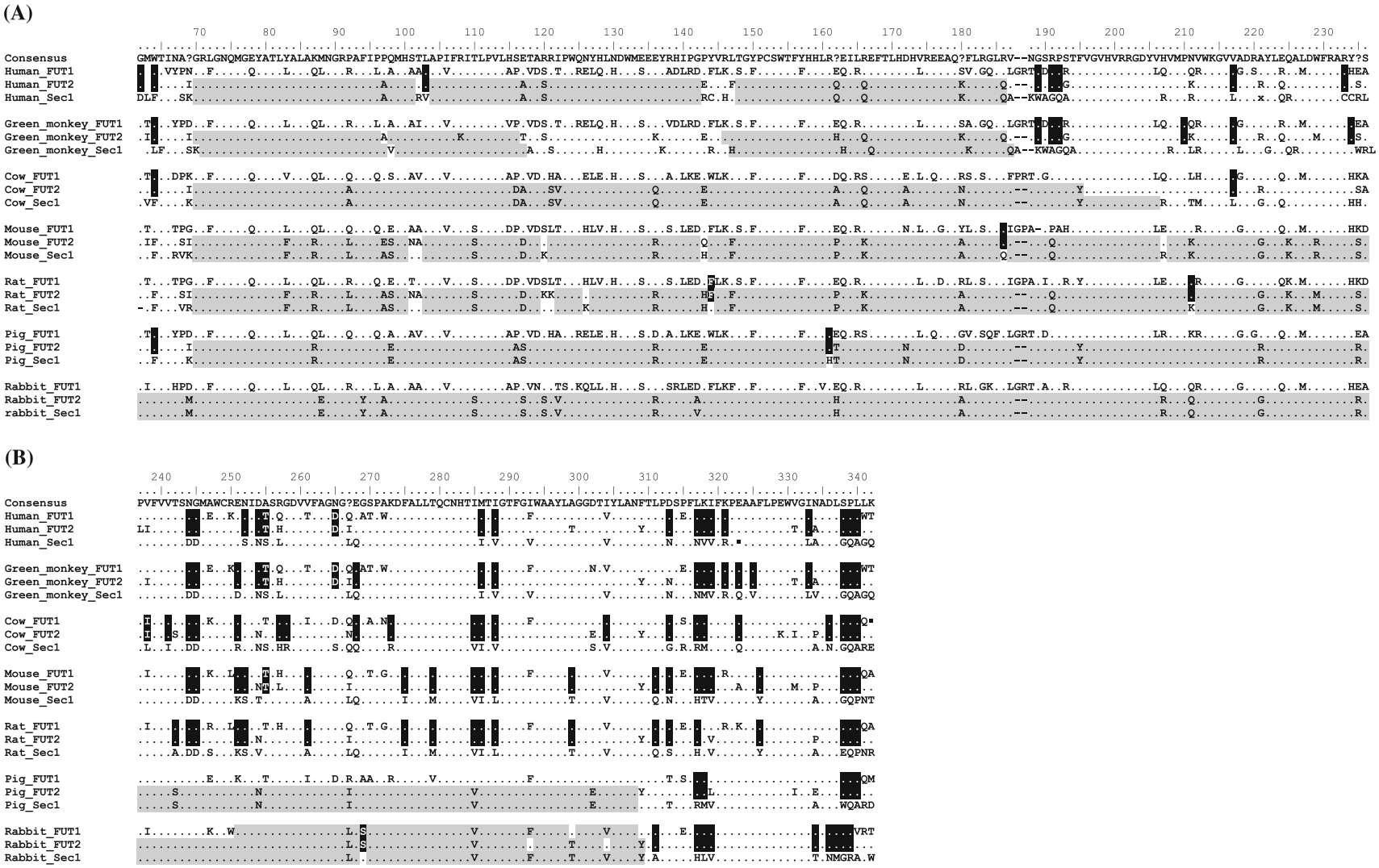

Fig. 3 Amino acid alignment of the catalytic domain of the α2FTs proteins of **a** the N-terminal region and **b** the C-terminal region. The homologous regions between *FUT*1, *FUT*2, and *Sec*1 that may have resulted by gene conversion are highlighted in light grey. Amino acid substitutions shared between *FUT*1 and *FUT*2, but not shared with *Sec*1, are highlighted in black. Closed square = stop codon; dashes = alignment gaps; dots = identity with the consensus sequence; and question marks = consensus sequence represents nonconsensual amino acids. The consensus sequence is based on the amino acid sequences of the α2FTs proteins of the mammals presented

lower than those for Fut1 and Fut2 (Table 1 Supplementary material).

In humans, both *FUT*1 and *FUT*2 present polymorphisms with null alleles encoding inactive or nearly inactive enzymes, responsible for Bombay and the nonsecretor phenotypes, respectively (Oriol et al. 2000). However, the frequency of these alleles can be different. Likewise, the cell types expressing each enzyme can vary in a species-specific manner. In humans, *FUT*1 null alleles are extremely rare (Wagner and Flegel 1997), and *FUT*1 is expressed in many cell types, including erythrocytes, the vascular endothelium, some neurons, and epithelial cells (Ravn and Dabelsteen 2000). In contrast, in humans, *FUT*2 null alleles are almost as frequent as functional alleles, and it has been shown that the gene undergoes balanced selection to maintain both types of alleles at high frequency in various human populations (Koda et al. 2001). *FUT*1 has been shown to be involved in some cellular functions, such as adhesion of leukocytes to the vascular endothelium, angiogenesis, and development of the olfactory bulb (Amin et al. 2008; Garcia-Vallejo et al. 2008; Moehler et al. 2008; St John et al. 2006). In contrast, *FUT*2 is mainly expressed in epithelial cells lining the surface of the digestive tract, the upper respiratory tract, and the lower urinary and genital tracts, i.e., in cells in contact with the external environment and potential pathogens (Marionneau et al. 2001). The secretor/nonsecretor polymorphism determined by *FUT*2 has been shown to be associated with sensitivity or resistance to various pathogens, including uropathogenic strains of *E. coli*, BabA-expressing strains of *H. pylori*, and various strains of norovirus (Azevedo et al. 2008; Le Pendu et al. 2006; Stapleton et al. 1995). The involvement of *FUT*1 in cellular functions and that of *FUT*2 in interactions with pathogens may explain the high frequency of *FUT*2 null alleles in contrast to the rare occurrence of such *FUT*1 alleles.

Classical studies on the evolution of duplicated genes indicate that the persistence of both duplicates requires their functional differentiation. In the absence of such differentiation, one of the duplicate should rapidly become a pseudogene (Teshima and Innan 2004; Walsh 2003). This is consistent with the inactivation of *Sec*1 in most primate species and with our observation of a limited and most likely ancient gene-conversion event in this lineage. As discussed previously, *FUT*1 and *FUT*2 have become functionally differentiated. Gene conversion involving *Sec*1 after its inactivation may no longer be observed because it would be deleterious to *FUT*2 or *FUT*1. In other species, inactivation of *Sec*1 may be recent or as yet not complete; therefore, many gene-conversion events between *FUT*2 and *Sec*1 can still be detected. The situation is clearly different in rabbit in which the three genes are involved in gene-conversion events. In such a species, *Sec*1 might have acquired a function distinct from those of the two other α1,2fucosyltransferases genes; however, that remains to be defined.

In conclusion, the gross evolutionary history of α2FTs (*FUT*1, *FUT*2, and *Sec*1) seems clear, but the evolution of these genes involved many gene-conversion events that can only be partially characterized and enumerated. It will be difficult to describe the exact phylogenetic relations for each species and gene because these gene conversions differ in position and in length size and because the several histories embedded in the sequences alignment obscures true evolutionary relations. The degree of concerted evolution of the three α1,2fucosyltransferases genes appears to be species-specific, possibly related to the functional differentiation of these genes.

Acknowledgements This work was supported by the Foundation for Science and Technology Portugal: Project POCTI/BIA-BDE/61553/2004 and grants SFRH/BD/31048/2006, SFRH/BPD/27021/2006 to J.A. and P.J.E., respectively, by the Spanish Ministry of Science and Education (grant number BIO2007-61411 to D.P.) and by the Région des Pays de la Loire (project Calilago to J.L.P.).

References

Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21:2104–2105

Amin MA, Ruth JH, Haas CS, Pakozdi A, Mansfield PJ, Haghshenas J, Koch AE (2008) H-2g, a glucose analog of blood group H antigen, mediates mononuclear cell recruitment via Src and phosphatidylinositol 3-kinase pathways. Arthritis Rheum 58:689–695

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

Azevedo M, Eriksson S, Mendes N, Serpa J, Figueiredo C, Resende LP, Ruvöen-Clouet N, Haas R, Borén T, Le Pendu J, David L (2008) Infection by *Helicobacter pylori* expressing the BabA adhesin is influenced by the secretor phenotype. J Pathol 215:308–316

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

Boren T, Falk P, Roth KA, Larson G, Normak S (1993) Attachment of *Helicobacter pylori* to human gastric epithelium mediated by blood group antigens. Science 262:1892–1895

Borges BN, Paiva TS, Harada ML (2008) Evolution of the SEC1 gene in New World monkey lineages (Primates, Platyrrhini). Genet Mol Res 7:663–678

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

Breton C, Snajdrovà L, Jeanneau C, Koca J, Imberty A (2006) Structures and mechanisms of glycosyltransferases. Glycobiology 16:29R–37R

Bureau V, Marionneau S, Cailleau-Thomas A, Le Moullac-Vaydie B, Liehr T, 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

Garcia-Vallejo JJ, van Liempt E, da Costa Martins P, Beckers C, van het Hof B, Gringhuis SI, Zwaginga JJ, van Dijk W, Geitjenbeek TB, van Kooyk Y, van Die I (2008) DC-SIGN mediates adhesion and rolling of dendritic cells on primary human umbilical vein endothelial cells through Lewis Y antigen expressed on ICAM-2. Mol Immunol 45:2359–2369

Guillon P, Ruvöen-Clouet N, Le Moullac-Vaidye B, Marchandeau S, Le Pendu J (2009) Association between expression of the H histo-blood group antigen α1,2fucosyltransferases polymorphism of wild rabbits, and sensitivity to rabbit hemorrhagic disease virus. Glycobiology 19:21–28

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704

Halloran MM, Carley WW, Polverini PJ, Haskell CJ, Phan S, Anderson BJ, Woods JM, Campbell PL, Volin MV, Bäcker AE, Koch AE (2000) Le^y/H: an endothelial-selective, cytokine-inducible, angiogenic mediator. J Immunol 164:4868–4877

Iwamori M, Domino SE (2004) Tissue-specific loss of fucosylated glycolipids in mice targeted deletion of alpha(1,2)fucosyltransferase genes. Biochem J 380:75–81

Koda Y, Tachida H, Wang H, Liu Y, Soejima M, Ghaderi AA, Takenada O, Kimura H (2001) Contrasting patterns of polymorphisms at the ABO-secretor gene (FUT2) and plasma (1, 3)fucosyltrasnferase gene (FUT6) in human populations. Genetics 158:747–756

Kosakovsky-Pond SL, Posada D, Gravenor MB, Woelk CH, Frost S (2006a) Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol 23:1891–1901

Kosakovsky-Pond SL, Posada D, Gravenor MB, Woelk CH, Frost S (2006b) GARD: a genetic algorithm for recombination detection. Bioinformatics 22:3096–3098

Le Pendu J, Ruvöen-Clouet N, Kindberg E, Svensson L (2006) Mendelian resistance to human norovirus infections. Semin Immunol 18:375–386

Marionneau S, Cailleau-Thomas A, Rocher J, Le Moullac-Vaidye B, Ruvoën-Clouet N, Clément M, Le Pendu J (2001) ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world. Biochimie 83:565–573

Moehler TM, Sauer S, Witzel M, Andrulis M, Garcia-Vallejo JJ, Grobholz R, Willhauck-Fleckenstein M, Greiner A, Goldschmidt H, Schwartz-Albiez R (2008) Involvement of alpha 1–2-fucosyltransferase (FUT1) and surface-expressed Lewis(y) (CD174) in first endothelial cell-cell contacts during angiogenesis. J Cell Physiol 215:27–36

Ohno S (1970) Evolution by gene duplication. Springer, New York

Oriol R, Candelier JJ, Mollicone R (2000) Molecular genetics of H. Vox Sang 78:105–118

Posada D (2006) ModelTest Server: a web-based tool for the statistical selection of models of nucleotide substitution online. Nucleic Acids Res 34:W700–W703

Ravn V, Dabelsteen E (2000) Tissue distribution of histo-blood group antigens. APMIS 108:1–28

Ruiz-Palacios GM, Cervantes LE, Ramos P, Chavez-Munguia B, Newburg DS (2003) *Campylobacter jejuni* binds intestinal H(O) antigen and fucosyloligosaccharides of human milk inhibit its binding and infection. J Biol Chem 278:14112–14120

Saunier K, Barreaud JP, Eggen J, Oriol R, Levéziel H, Julien R, Petit JM (2001) Organization of the bovine alpha2-fucosyltransferase gene cluster suggests that the Sec1 gene might have been shaped through a nonautonomous L1-retrotransposition event within the same locus. Mol Biol Evol 18:2083–2091

Sawyer SA (1989) Statistical tests for detecting gene conversion. Mol Biol Evol 6:526–538

Shimodaira H, Hassegawa M (1999) Multiple Comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol 16:1114–1234

Soejima M, Fujihara J, Takeshita H, Koda Y (2008) Sec1-FUT2-Sec1 hybrid allele generated by interlocus gene conversion. Transfusion 48:488–492

St John JA, Claxton C, Robinson MW, Yamamoto F, Domino SE, Key B (2006) Genetic manipulation of blood group carbohydrates alters development and pathfinding of primary sensory axons of the olfactory systems. Dev Biol 298:470–484

Stapleton A, Hooton TM, Fennell C, Roberts PL, Stamm WE (1995) Effect of secretor status on vaginal and rectal colonization with fimbriated *Escherichia coli* in women with and without recurrent urinary tract infection. J Infect Dis 171:717–720

Swofford DL (2000) Phylogenetic analysis using Parsimony (*and other methods). Sinauer Associates, Sunderland

Teshima KM, Innan H (2004) The effect of conversion on the divergence between duplicated genes. Genetics 166:1553–1560

Thompson J, Higgins D, Gibson T (1994) ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucl Acid Res 22:4673–4680

Uchida H, Kinoshita H, Kawai Y, Kitazawa H, Miura K, Shiiba K, Horii A, Kimura K, Taketomo N, Oda M, Yajima T, Saito T (2006) Lactobacilli binding human A-antigen expressed in intestinal mucosa. Res Microbiol 157:659–665

Wagner FF, Flegel WA (1997) Polymorphism of the h allele and the population frequency of sporadic nonfunctional alleles. Transfusion 37:284–290

Walsh B (2003) Population-genetic models of the fates of duplicated genes. Genetica 118:279–294

Walsh J (1987) Sequence-dependent gene conversion: can duplicate genes diverge fast enough to escape gene conversion? Genetics 117:543–557

Walsh J (1995) How often do duplicated genes evolve new functions? Genetics 139:421–428