Copyright © 2001 by the Genetics Society of America

Contrasting Patterns of Polymorphisms at the ABO-Secretor Gene (*FUT2*) and Plasma α(1,3)Fucosyltransferase Gene (*FUT6*) in Human Populations

Yoshiro Koda,* Hidenori Tachida,† Hao Pang,* Yuhua Liu,* Mikiko Soejima,* Abbas A. Ghaderi,‡ Osamu Takenaka§ and Hiroshi Kimura*

**Division of Human Genetics, Department of Forensic Medicine, Kurume University School of Medicine, Kurume 830-0011, Japan, †Department of Biology, Faculty of Sciences, Kyushu University, Fukuoka 812-8581, Japan, ‡Department of Immunology, Shiraz University of Medical Sciences, Shiraz, Iran and §Department of Biochemistry, Primate Research Institute, Kyoto University, Inuyama 484-0002, Japan*

Manuscript received August 25, 2000
Accepted for publication February 15, 2001

ABSTRACT

The coding sequences (~1 kb) of *FUT2* [ABO-Secretor type α(1,2)fucosyltransferase] and of *FUT6* [plasma α(1,3)fucosyltransferase] were analyzed for allelic polymorphism by direct sequencing in five populations. The nucleotide diversities of *FUT2* estimated from pairwise sequence differences were 0.0045, 0.0042, 0.0042, 0.0009, and 0.0008 in Africans, European-Africans, Iranians, Chinese, and Japanese, respectively. The nucleotide diversities of *FUT6* were 0.0024, 0.0016, 0.0015, 0.0017, and 0.0020 in Africans, European-Africans, Iranians, Chinese, and Japanese, respectively. At *FUT2*, excesses in pairwise sequence differences compared to the number of polymorphic sites as indicated by a significantly positive Tajima's *D* were observed in European-Africans and in Iranians. The data do not fit expectations of the equilibrium neutral model with an infinite number of sites. On the other hand, Tajima's *D*'s at *FUT6* in each of the five populations and at *FUT2* in Africans, Chinese, and Japanese were not significantly different from zero. F_{ST} between the Asians and the others measured at *FUT2* was higher than at *FUT6*. These results suggest that natural selection was responsible for the generation of the *FUT2* polymorphism in European-Africans and in Iranians.

THE human α(1,2)- and α(1,3)-fucosyltransferase genes constitute a multigene family, and seven genes encoding human fucosyltransferases (*FUT1*–*FUT7*) have been isolated (Kukowska-Latallo *et al.* 1990; Larsen *et al.* 1990; Lowe *et al.* 1991; Weston *et al.* 1992a,b; Sasaki *et al.* 1994; Kelly *et al.* 1995). The α(1,2)fucosyltransferase plays a key role for tissue expression of the H antigen—a common precursor for the blood group A and B antigens. Two human α(1,2)fucosyltransferase genes (*FUT1* and *FUT2*) have been isolated (Larsen *et al.* 1990; Kelly *et al.* 1995; Rouquier *et al.* 1995). *FUT1* encodes H type α(1,2)fucosyltransferase (H enzyme) that regulates the expression of the H antigen and thereafter A and/or B antigens in the erythroid lineage and in the vascular endothelia. *FUT2* encodes Secretor type α(1,2)fucosyltransferase (Se enzyme) that regulates the expression of these antigens in the secretory glands and in the digestive mucosa. About 20% of individuals are nonsecretors who fail to express the ABO antigen in saliva, because they are homozygous for the null allele of *FUT2*.

Recent studies have indicated that the *FUT2* polymorphism showed an ethnic group-specific pattern (Kelly *et al.* 1995; Koda *et al.* 1996; Liu *et al.* 1998). A common null allele of *FUT2* in East Asians has a missense mutation in codon 129 (Ile-Phe; se^{385}, B6 and B7 alleles in this study), whereas a common null allele of *FUT2* in Europeans and Africans has a nonsense mutation in codon 143 (Trp-ter; se^{428}, B3 and B4 alleles in this study). The enzyme encoded by se^{385} shows weak activity (2–5% activity of the wild-type enzyme), while the enzyme encoded by se^{428} shows no activity (Kelly *et al.* 1995; Koda *et al.* 1996). As compared to the wild-type allele (*Se*, A1 allele in this study), seven base differences were observed within the 1-kb DNA fragment encompassing the protein coding region in the se^{428} allele (see Figure 1). On the other hand, only two base changes from the wild-type allele were observed in the se^{385} allele.

FUT3, *FUT5*, and *FUT6* encode α(1,3)fucosyltransferases and are organized in a cluster, within a 40-kb region, in the short arm of chromosome 19 (19p13.3; McCurley *et al.* 1995). These three genes maintain intronless coding segments with nucleic acid sequences that share ~90% identity to each other. *FUT3* encodes the Lewis α(1,3/4)fucosyltransferase that synthesizes the Lewis a and Lewis b epitopes, whereas the tissue expression of the *FUT5*-encoding enzyme is not yet known. *FUT6* encodes the plasma α(1,3)fucosyltransferase and is known to be polymorphic (Mollicone *et al.* 1994; Larson *et al.* 1996).

In this study, we analyzed the *FUT2* allelic variation by direct DNA sequencing in five populations: Africans (Xhosa), European-Africans of South Africa, Iranians,

Corresponding author: Hiroshi Kimura, Department of Forensic Medicine, Kurume University School of Medicine, Kurume 830-0011, Japan. E-mail: hkimura@med.kurume-u.ac.jp

Chinese, and Japanese. We found that the nucleotide variation at *FUT2* did not fit the expectations of the equilibrium neutral model in European-Africans and in Iranians. The allelic variation of *FUT6* was also examined for comparison. We found contrasting patterns of variations at these two loci, suggesting the action of selection at *FUT2*.

MATERIALS AND METHODS

Population samples: Individuals from five populations were sampled: Africans (Xhosa) from Cape Town, European-Africans from Cape Town, Iranians from Shiraz, Chinese from Guangzhou, and Japanese from Fukuoka (Koda *et al.* 1996; Liu *et al.* 1998, 1999; Pang *et al.* 1999). In addition, 15 chimpanzees were sampled. Chimpanzee genomic DNA was obtained from one of the authors (O.T.).

PCR amplification and sequencing of *FUT2* and *FUT6*: Genomic DNA was prepared from peripheral blood leukocytes by an organic solvent method. The *FUT2* and *FUT6* genes were amplified from genomic DNA by PCR for sequencing. *FUT2* (16- to 999-bp coding region and 1000- to 1029-bp noncoding region) was amplified and sequenced as described previously (Koda *et al.* 1996; Liu *et al.* 1998). A DNA fragment (1.2 kb) including the full coding sequence (1080 bp) of *FUT6* was amplified and sequenced as described previously (Pang *et al.* 1999). DNA sequence was determined by the dideoxynucleotide chain termination method using an ABI PRISM Big Dye Terminator cycle sequencing ready reaction kit and an ABI PRISM 310 genetic analyzer (Perkin-Elmer-Japan ABI). Numbers of individuals sequenced in this study in respective populations are as follows: $n = 30$ for *FUT2* in Africans; $n = 36$ for *FUT2* in European-Africans; $n = 53$ for *FUT2* and $n = 52$ for *FUT6* in Iranians; $n = 30$ for *FUT2* and *FUT6* in Chinese; and $n = 30$ for *FUT2* in Japanese populations. The *FUT6* sequences of Africans ($n = 56$), European-Africans ($n = 52$), and Japanese ($n = 53$) from our previous study (Pang *et al.* 1999) were added for data analysis. Individuals examined for *FUT6* include all individuals examined for *FUT2*. In addition, the *FUT2* region was sequenced in 15 chimpanzee individuals.

Haplotype determination: For haplotyping of the alleles, PCR products from selective individuals heterozygous at more than one variant site in *FUT2* ($n = 17$) or *FUT6* ($n = 21$) were cloned into plasmid pGEM using the pGEM-T vector system I (Promega, Madison, WI). At least six clones from each subject were analyzed by sequencing to determine the haplotypes (alleles). Probable alleles in other individuals heterozygous at more than one variant site were estimated by the results of linkage of the base substitutions in all individuals homozygous at all sites and heterozygous at one site, and of haplotyping of the alleles in selective individuals as mentioned above.

Statistical analysis: To measure diversity within a population, the number of segregating sites S for sample size n and the average pairwise sequence difference k were calculated. Under the equilibrium neutral model, these values $S/(\sum_{i=1}^{n-1} 1/i)$ and k provide alternative estimators of $\theta = 4N_e\mu$, where N_e is the effective population size and μ is the mutation rate per sequence per generation (Watterson 1975; Nei and Tajima 1981). From the two estimates of θ, a test statistic D proposed by Tajima (1989) was computed. In addition, with the number of singleton sites η_s, test statistics D^* and F^* proposed by Fu and Li (1993) were computed. Nucleotide diversity π was estimated by dividing the average pairwise sequence difference by the numbers of sites. F_{ST} between populations was estimated from the sequence data using the method of Hudson *et al.* (1992). All analyses were carried out using DNASP3.1.4 (Rozas and Rozas 1999).

RESULTS

Sequence variation of *FUT2* and *FUT6*: In previous studies, we sequenced only selective individuals to determine the null alleles of *FUT2* and frequencies of identified polymorphic sites were determined by PCR-restriction fragment length polymorphism or allele-specific PCR analysis in Africans, European-Africans, Chinese, and Japanese (Koda *et al.* 1996; Liu *et al.* 1998, 1999). To analyze DNA sequence variation of *FUT2* in detail, we sequenced the coding region of *FUT2* for 355 chromosomes from 179 randomly selected individuals drawn from the five populations and 15 chimpanzees in this study (we deleted 3 chromosomes for calculation of nucleotide diversity because of the presence of three fusion genes in *FUT2*; Koda *et al.* 1996). Concerning *FUT2* sequence analysis, we include the results of one Japanese, seven Africans, and one European-African (none of Chinese) from previous studies (Koda *et al.* 1996; Liu *et al.* 1998). We also sequenced the coding region of *FUT6* for 164 chromosomes from 82 individuals from the Iranian and Chinese populations. In addition, the *FUT6* sequences of Africans, European-Africans, and Japanese obtained from Pang *et al.* (1999) were reanalyzed.

We found 21 polymorphic sites at *FUT2* (1014 bp) and 18 polymorphic sites at *FUT6* (1080 bp; Figure 1). At these polymorphic sites, 1 site was a biallelic 1-base deletion and 20 sites were single nucleotide polymorphism (10 replacement, 8 synonymous, and 2 noncoding changes) at *FUT2*, and 1 site was a biallelic 1-base insertion and 17 sites were single nucleotide polymorphism (9 replacement and 8 synonymous changes) at *FUT6*. Three of the 21 sites (162 G to A, 278 C to T, and 543 G to A) at *FUT2* and none of the 18 sites at *FUT6* were newly identified in this study. Six of the 21 sites at *FUT2* had only a single copy of the rarer nucleotide (singletons), whereas none of the 18 sites was a singleton at *FUT6*.

Variations in *FUT2* and *FUT6* showed different patterns and were distributed unevenly among the five populations. Only 1 of the 21 sites of *FUT2* and 7 of the 18 sites of *FUT6* showed biallelic segregating variation in all five populations. The African population was found to have the largest numbers of population-specific polymorphic sites both at *FUT2* (5 sites) and *FUT6* (8 sites). The European-African, Iranian, and Japanese populations each had 1 population-specific polymorphic site at *FUT2*, while the Chinese population had 3 population-specific polymorphic sites at *FUT2*. Two population-specific polymorphic sites were found in European-Africans at *FUT6*, whereas no population-specific polymorphic site was observed in the Japanese, Chinese, and Iranian populations at *FUT6*.

FUT2

```
                                                  11
          1112223333344445556667778990 0
         4678117577880288347227379456 01
         0219068759575801131897986990 91
         RSSSSSRSSRRSSRSRSSRRSRR-RRRSNN

A1       AGACACCCACACCGCGCGCCGTGCCGCAAT
A2       G****************************
A3       ********G*********************G*
A4       ***************A**************
A5       *******T***********************
A6       *******T******T****************
A7       *******T*T****T****************
A8       *******T**************A****GGC
B1       ***********************T**********
B2       *************************A*****
B3       **G**T*******A********A****GGC
B4       **G**T*******A***A****A****GGC
B5       ******TT***********************
B6       *******T**T*********************
B7       *A*****T**T*********************
B8       *******T**********T************
B9       *******T******T********-******
C1       **GTG*******T********C****TGGC
C2       **G*G*******T***T***AC*****GGC
C3       **GTGT******T********C****TGGC
C4       **GTG******AT********C****TGGC
C5       **GTG*******T********C**T*TGGC
```

FUT6

```
                    1
          1333457778899990
         1673779223357047 70
         8326069798959751 72
         SSRSRRiRSSRSSRRRRS

A1       GGGCCC-GTCGGCCCCGG
A2       *A****************
A3       *AA**T************
A4       *AA**T*T**********
A5       ********C*********
A6       *********T******A*
A7       ****************C
A8       ****T***********C
A9       ****T********G***C
A10      ***********T**T*C
A11      **********A*****C
B1       ****T*****A******C
B2       ****T*****A***A**C
B3       A**A**C****A*****C
B4       A**A**C****A****AC
```

FIGURE 1.—Polymorphic base positions at 17 *FUT2* alleles and at 15 *FUT6* alleles. Functional alleles were indicated as A and nonfunctional alleles as B in both *FUT2* and *FUT6*. C1–C5 alleles were found in 15 chimpanzee individuals and all chimpanzee alleles were functional. Vertical labels on the top indicate the positions of nucleotides. An asterisk indicates the sequence matched the reference allele (A1) at that position. A dash indicates a single base deletion at that position at *FUT2*. The 499i in *FUT6* means a single base insertion at that position. Sites are also classified depending on whether the change is replacement (R), synonymous (S), or in the noncoding region (N). The B10 allele, not described here, is the fusion gene allele found in the Japanese population.

From the results of direct sequencing and haplotyping of some alleles, we estimated the probable alleles and their frequencies. We found 18 alleles (including a fusion gene; KODA *et al.* 1996) at *FUT2* and 15 alleles at *FUT6*. Functional and null alleles are designated as A's and B's, respectively. We also found 5 alleles at chimpanzee *FUT2* and these alleles are designated as C's. Figure 1 shows the allelic variations found at *FUT2* and *FUT6* in the five populations and chimpanzees. We found 9 alleles in the Africans, 6 in European-Africans, 5 in Iranians, 6 in Chinese, and 5 alleles in Japanese at *FUT2*, whereas we found 13 alleles in Africans, 8 in European-Africans, 7 in Iranians, 6 in Chinese, and 5 alleles in Japanese at *FUT6*. Figure 2 summarizes the allele frequencies in the five populations. To make phylogenetic analyses of *FUT2* and *FUT6*, neighbor-joining trees were drawn using orangutan sequences (AB015636 at *FUT2* and AB051859 at *FUT6*) as outgroups (Figure 3, A and B). Human alleles are monophyletic at both loci, although the origination of the *FUT2* B3/B4 alleles seemed to be very old (KODA *et al.* 2000). The results also suggested that nonfunctional alleles (B alleles) have multiple origins at both loci.

There were apparently different allelic distributions in the five populations. At *FUT2*, A1–A8 are functional alleles (*Se*), and B1–B10 are null alleles (*se*). The B10 allele of *FUT2* is a fusion gene, which was found in the Japanese population (KODA *et al.* 1996; LIU *et al.* 1999). Although we cannot rule out the possibility that the A8 allele is simply intermediate to connect the A5 and B3 alleles possibly generated by point mutations, a more parsimonious interpretation is that A8 was generated by crossing over between the A5 (or A6) and the B3 alleles. A8 was found only in the European population where the A5 and B3 frequencies were high, suggesting that recombination had occurred in this population. The crossover region is within a 310-bp stretch corresponding to the region between nucleotides 429 and 738 of *FUT2*. Consistent with the results from previous DNA sequencing of a few se^{428} alleles (KELLY *et al.* 1995; LIU *et al.* 1998), all 177 se^{428} alleles sequenced in this study had at least seven base substitutions within the 1.1-kb DNA fragment as compared to the A1 allele, and the sequence variation within the se^{428} alleles was very low (only one allele, B4, showed one additional polymorphic site at position 543). In addition, we found a novel allele (B5) with a missense mutation at position 278 in Iranians that may result in an amino acid change from alanine to valine. Determination of enzyme activity from a transient expression study indicated that this allele product has 4% enzyme activity as compared with the wild-type allele products (not shown). Accordingly, we classified this allele into an Se enzyme-deficient allele. The frequencies of functional alleles of *FUT2* alleles were evenly distributed in the African population, whereas A5 (Se^{357}) was a common functional allele in European-African and Iranian populations (Figure 2). A1 and A5 were common in the two Asian populations.

At *FUT6*, A1–A11 are functional alleles and B1–B4 are nonfunctional alleles. As for *FUT2* functional alleles, all *FUT6* alleles were much more evenly distributed in the African population, and almost half of the alleles found in the Africans were absent in non-African populations (Figure 2). The null allele B2 was detected in the European and Asian populations but not in the Africans. A9 and A11 constituted the two main alleles in the Asian populations, whereas A11 was a common allele in the European and Iranian populations. The

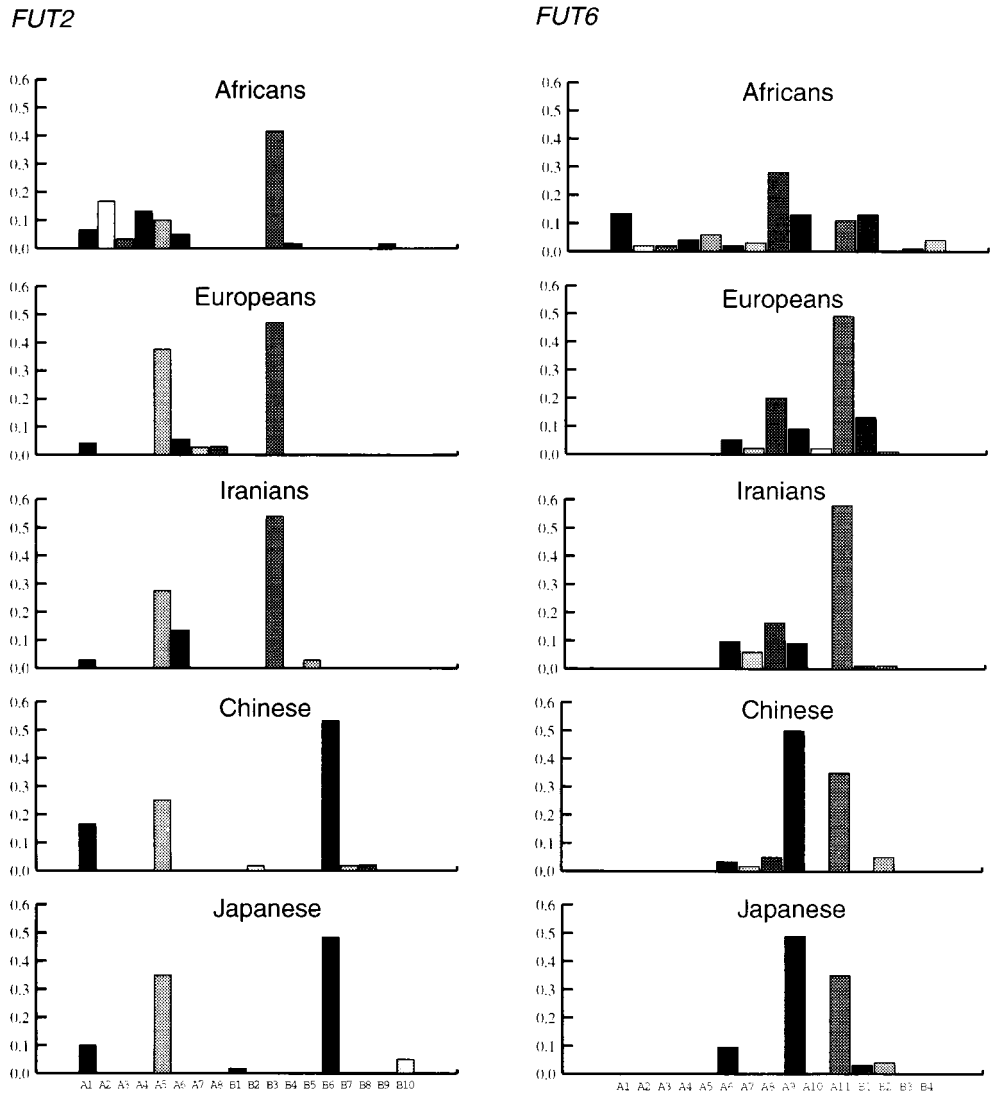

FIGURE 2.—Histogram representations of 18 *FUT2* and 15 *FUT6* allele frequencies for the five populations.

genotype frequencies with regard to the functional and null alleles in each population do not deviate significantly from the Hardy-Weinberg expectation (data not shown).

Statistical analysis: The average number of pairwise nucleotide differences (k), nucleotide diversity (π), θ, Tajima's D, number of singleton sites (η_s), and Fu and Li's F^* and D^* were estimated (Tables 1 and 2). Nucleotide diversity in the coding region of *FUT2* (π = 0.0033 ± 0.00007, 985 bp) and that of *FUT6* (π = 0.0020 ± 0.00002, 1080 bp) was much higher than in the lipoprotein lipase (LPL) coding region (π = 0.0005, 998 bp; NICKERSON *et al.* 1998) and other coding regions of 49 loci encompassing 54,193 bp (the π values ranged from a maximum of 0.0011 for fourfold degenerate sites to a minimum of 0.0003 for nondegenerate sites; LI and SADLER 1991). Nucleotide diversity at *FUT2* but not at *FUT6* was also higher than in the 2.67-kb region of the β-globin gene (π = 0.0018; HARDING *et al.* 1997), the 9.7-kb region of the LPL gene (π = 0.002), and the four degenerate sites in melanocortin 1 receptor locus (π = 0.0021; CLARK *et al.* 1998; NICKERSON *et al.* 1998; RANA *et al.* 1999). *FUT2* has two polymorphic sites at 10 and 12 bp downstream of the termination codon in addition to those in the coding region. If we add these sites, the diversity of the 1014-bp region at *FUT2* is increased (π = 0.0041 ± 0.00007). Nucleotide diversity of replacement sites was 0.0018 and 0.0016 in *FUT2* and *FUT6*, respectively, whereas that of synonymous and noncoding sites was 0.0105 and 0.0035 in *FUT2* and *FUT6*, respectively. In addition, nucleotide diversity within the functional *FUT2* alleles was 0.0009 (replacement sites, 0.0003). Thus, high nucleotide diversity is due to large silent-site variation and does not seem to be due to less constraint on replacement changes at the *FUT2* locus.

Tajima's D values at *FUT2* were estimated to be 2.86 ($P < 0.01$) and 3.17 ($P < 0.01$) in European-Africans and Iranians, respectively, whereas the D values in the other three populations at *FUT2* were estimated as −0.35 to 1.52 (Table 1). Tajima's D values at *FUT6* were estimated as −0.27 to 1.02 in the five populations (Table

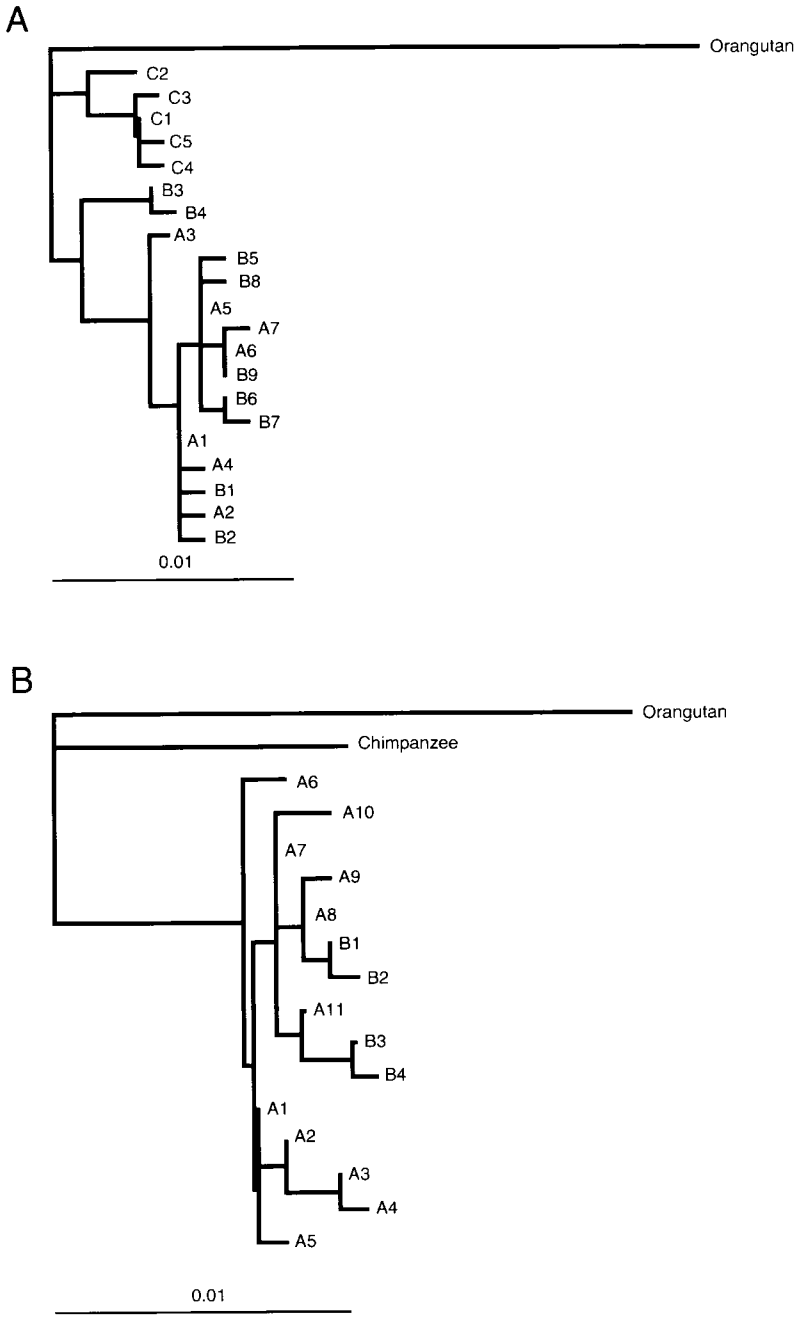

FIGURE 3.—The neighbor-joining trees constructed from the human *FUT2* alleles and primate *FUT2* homologues (A) and from the human *FUT6* alleles and primate *FUT6* homologues (B). We exclude the A8 allele of *FUT2* from the tree because A8 seemed to be generated by crossing over between the A5 (or A6) and the B3 alleles.

2). We cannot reject the null hypothesis of neutrality at *FUT6* in any of the five populations and at *FUT2* in the African, Chinese, and Japanese populations. However, the polymorphism at *FUT2* is difficult to explain by the equilibrium neutral model with constant effective population size in the European-Africans and Iranians. Fu and Li's test statistics at the *FUT2* locus were estimated with or without using chimpanzee sequences (Figure 1) as outgroup. The values with or without outgroup sequences did not differ significantly. Table 1 shows the results obtained from the method with outgroup sequences. Fu and Li's F^* also showed a similar pattern. F^* was significantly positive in Iranians and European-Africans.

To quantify population differentiation, we computed F_{ST} in all pairs of populations (Table 3). First, note that F_{ST}'s between Iranians and Europeans, and between Chinese and Japanese, were very small at both loci. Both pairs of populations are considered to be not genetically differentiated. The estimate of F_{ST} between Africans and Europeans based on the data at *FUT6* was 0.081 and comparable to that (0.065 ± 0.014) estimated in the LPL gene region (CLARK *et al.* 1998). F_{ST} values between Africans and non-Africans were not significantly different at the two loci (0.069 at *FUT2* and 0.072 at *FUT6*). However, F_{ST} estimated at *FUT2* showed a peculiar pattern. First, F_{ST}'s between any pairs of African, European, and Iranian populations were small. Second, F_{ST} be-

TABLE 1

Population genetic statistics for the *FUT2* variation

Population	n	S	θ	k	π	η_S	Tajima's D	Fu and Li's F^*	Fu and Li's D^*
Africans	60	14	3.00	4.55	0.0045	2	1.52 NS	1.07 NS	0.54 NS
Europeans	72	10	2.06	4.26	0.0042	0	2.86 $P < 0.01$	2.30 $P < 0.02$	1.43 NS
Iranians	106	10	1.91	4.30	0.0042	0	3.17 $P < 0.01$	2.42 $P < 0.02$	1.40 NS
Chinese	60	5	1.07	0.91	0.0009	3	−0.35 NS	−1.71 NS	−1.96 NS
Japanese	57	3	0.65	0.76	0.0008	1	0.35 NS	−0.26 NS	−0.47 NS
Total	355	21	3.26	4.19	0.0041	6	0.73 NS	−0.71 NS	−1.50 NS
Chimpanzee	30	7	1.77	1.96	0.0019	1	0.32 NS	0.58 NS	0.58 NS

S, number of segregating sites; n, number of chromosomes; $\theta = S/(\sum_{i=1}^{n-1} 1/i)$. Fu and Li's F^* and D^* values were estimated using chimpanzee sequences as outgroup. NS, not significant.

tween the African/European/Iranian and Asian populations was large (>0.42). The F_{ST} analysis shows that the differentiation pattern is very different at the two loci.

Chimpanzee *FUT2* homologue variation: Since the sequence variation of human *FUT2* was out of the scope of the standard mutation-selection balance model or the equilibrium neutral model with random mating, we examined the sequence variation of the chimpanzee *FUT2* homologue. We found 7 polymorphic sites (2 replacement and 5 synonymous single nucleotide changes) and 5 alleles in the chimpanzee *FUT2* homologue (Figure 1). In contrast to human, we did not find a null allele in the chimpanzee *FUT2* homologue by transient expression study (Koda *et al.* 1996). The average number of pairwise nucleotide differences (k), nucleotide diversity (π), θ, Tajima's D, number of singleton sites (η_S), and Fu and Li's F^* and D^* of the chimpanzee *FUT2* homologues were also estimated (Table 1). We cannot reject the null hypothesis of neutrality at the *FUT2* homologue in the chimpanzee population. We can test the levels of divergence and polymorphism at silent *vs.* replacement sites by the McDonald and Kreitman (1991) test. Within respective species 12 sites were silent and 13 were replacement polymorphisms in total, whereas the human/chimpanzee comparison revealed 2 silent and 1 replacement fixed differences. This test did not yield a significant result (for Fisher's exact test, $P = 1.0$). We also performed the McDonald and Kreitman test at the *FUT6* locus using a chimpanzee sequence (DDBJ/EMBL/GenBank accession no. Y14035). Within each species 8 sites were silent and 10 were replacement polymorphisms in total, whereas the human/chimpanzee comparison revealed 6 silent and 12 replacement fixed differences. As at the *FUT2* locus, the McDonald and Kreitman test did not yield a significant result (for Fisher's exact test, $P = 0.73$).

DISCUSSION

In this study, we examined DNA polymorphisms at the *FUT2* and *FUT6* loci in five human populations. We found contrasting patterns of DNA polymorphisms at the two loci. First, Tajima's D and Fu and Li's F^* were significantly positive in European-African and Iranian populations but not so in the other populations at *FUT2*. These statistics are not significant in any populations at *FUT6*. Second, there was strong differentiation between the Asian (Chinese and Japanese) and non-Asian (African, European-African, and Iranian) populations at *FUT2* as indicated by large F_{ST}'s between these two groups of populations but F_{ST} values at *FUT6* were between 0.1 and 0.16. The F_{ST} values at *FUT6* are close to those estimated at 84 protein and 33 blood group loci

TABLE 2

Population genetic statistics for the *FUT6* variation

Population	n	S	θ	k	π	η_S	Tajima's D	Fu and Li's F^*	Fu and Li's D^*
Africans	112	15	2.83	2.55	0.0024	0	−0.27 NS	1.10 NS	1.63 $P < 0.05$
Europeans	104	10	1.92	1.77	0.0016	1	−0.20 NS	0.43 NS	0.67 NS
Iranians	104	8	1.53	1.63	0.0015	1	0.16 NS	0.41 NS	0.45 NS
Chinese	60	8	1.72	1.85	0.0017	0	0.21 NS	1.13 NS	1.33 NS
Japanese	106	8	1.53	2.18	0.0020	0	1.02 NS	1.41 NS	1.27 NS
Total	486	18	2.66	2.18	0.0020	0	−0.43 NS	0.98 NS	1.65 NS

S, number of segregating sites; n, number of chromosomes; $\theta = S/(\sum_{i=1}^{n-1} 1/i)$. Fu and Li's F^* and D^* values were estimated without outgroup. NS, not significant.

TABLE 3

F_{ST}'s in all pairs of populations estimated from sequence diversity at *FUT2* and *FUT6*

	Africans	Europeans	Iranians	Chinese	Japanese
Africans		0.017	0.026	0.422	0.437
Europeans	0.081		−0.006	0.438	0.450
Iranians	0.130	0.021		0.483	0.494
Chinese	0.100	0.103	0.157		−0.011
Japanese	0.084	0.090	0.129	−0.006	

F_{ST}'s estimated at *FUT2* (above diagonal) and at *FUT6* (below diagonal) are shown.

(Nei and Livshits 1990). The pattern observed at *FUT6* is similar to those at LPL (Clark *et al.* 1998; Nickerson *et al.* 1998), β-globin (Harding *et al.* 1997), and dystrophin (Zietkiewicz *et al.* 1998) genes. Thus, the pattern at *FUT6* seems to reflect the past demographic history of human populations. However, the pattern at *FUT2* was very different from those at other loci, suggesting action of some locus-specific factors.

The contrasting pattern at *FUT2* results mainly from high frequencies of two null alleles, B3 (se^{428}) in the non-Asian populations and B6 (se^{385}) in the Asian populations. The B3 allele diverged greatly from the functional alleles (A alleles). Indeed, the divergence time between B3 and the functional alleles was estimated to be about 3 million years ago (95% confidence limits were 1.25 and 4.72 million years ago assuming the divergence time of the human and chimpanzee genes to be 5 million years ago; Koda *et al.* 2000). The ancient origin of B3 was difficult to explain by the neutrality of the null allele and we suggest it to be due to balancing selection. This ancient origin and low variability both within functional alleles and within null alleles result in significant positive *D* values in the European-African and Iranian populations. The frequency of B3 is also high in the African population, leading to a positive *D*, but *D* was not significant because of a modest amount of variability among the functional alleles. Since Tajima's *D* (and all other neutrality statistics) is not very sensitive to detect selection (Simonsen *et al.* 1995), the irregular pattern found in the European-African and Iranian populations needs some explanation. One simple explanation is balancing selection acting on functional alleles and the B3 allele as suggested previously (Koda *et al.* 2000). However, we also need to explain large F_{ST}'s between the Asian and non-Asian populations.

Large F_{ST}'s between the Asian and non-Asian populations were caused by high frequencies of the B6 allele in the Asian populations. B6 most likely originated from the functional allele, A5 (Se^{357}), that is separated from B6 by only one nucleotide change. Because of the small number of nucleotide changes, Tajima's *D* is significantly positive in contrast to *D* values in the European-African and Iranian populations. Balancing selection makes Tajima's *D* significantly positive if it persisted long (Simonsen *et al.* 1995). Since B6 exists only in the Asian populations, the increase of its frequency occurred after the separation of the Asian and non-Asian populations, the time of which is thought to be less than 100,000 years ago (Nei and Livshits 1990). In fact, a recent study based on 53 complete mtDNA sequences suggests this time to be less than *ca.* 50,000 years ago (Ingman *et al.* 2000). Such a large differentiation between races is also found at other loci. Harris and Hey (1999) reported that F_{ST} between African and non-African populations is 0.617 at the X-linked locus coding for PDHA1. Also, Rana *et al.* (1999) found an allele whose frequency is 0.66 in a Chinese population but that is absent in non-Asian populations at the locus coding for melanocortin 1 receptor. The authors of both articles suggested the action of selection on these two loci. Our case is similar to that of Rana *et al.* (1999) and we can apply their method of inferring selection utilizing the formula of Li (1975). The frequency of B6 is 0.55 in the Chinese population. It takes 140,000 years on average for a neutral allele to reach that frequency if we assume that effective population size is 5000, initial frequency is $\frac{1}{2N}$, and one generation is 20 years as were assumed by Rana *et al.* (1999). Although the time depends on the presumed parameter values such as effective population size and the initial frequency, the action of selection is suspected for the increase of B6 in the Asian populations.

To summarize, the contrasting pattern at *FUT2* is due mainly to the long persistence of B3 in the non-Asian populations and a rapid increase of B6 in the Asian populations. Both observations are difficult to explain by the simple neutral model and suggest the action of selection on this locus.

One possible selection model that explains both observations is a two-class overdominance model for the functional and null alleles. Assume that there are two classes of alleles A_i and B_i and that heterozygotes of alleles belonging to different classes (A_iB_i) have higher fitness than homozygotes and heterozygotes of alleles belonging to the same class. In this system, only two alleles A_i and B_i, each from the two classes, are maintained in a population at a time. Occasionally a new allele rapidly invades the population, replacing the old

allele of the same class. In the case of the *FUT2* locus, the functional and null alleles can be regarded as A_i and B_i, respectively. Under this scenario, B3 was maintained for a long time because of overdominance and B6 rapidly replaced B3 after the separation of the Asian and non-Asian populations.

To be more concrete and investigate turnovers, consider a simple model with only three alleles, A representing all A's, B3, and B6. Assume that an AB heterozygote has fitness of 1 and a B3B6 heterozygote has fitness of $1 - (t_3 + t_6)/2$. Also assume that fitnesses of homozygotes AA, B3B3, and B6B6 are $1 - s$, $1 - t_3$, and $1 - t_6$, respectively. Let the frequencies of A, B3, and B6 be $1 - y$, $y(1 - x)$, and yx, respectively. If we assume that selection is strong, then the frequency, x, of B6 in the null alleles satisfies the following differential equation:

$$\frac{dx}{dt} = \frac{y}{2}x(1 - x)(t_3 - t_6). \quad (1)$$

Thus, if B6B6 has higher fitness than B3B3 ($t_3 > t_6$), the B6 allele rapidly increases its frequency as a semidominant advantageous mutation with a selection coefficient $y(t_3 - t_6)$. The turnover rate of B's depends on the selective advantage, y (frequency of null alleles), and the mutation rate. This simple analysis shows that the long persistence of B3 and rapid increase of B6 can be explained by the two-class overdominance model if we assume that the mutation rate to new B alleles with higher homozygote fitness is low. One drawback of this explanation is that it cannot explain the low diversity within se^{428} alleles including B3 unless some modification such as time-dependent change of selection coefficients is incorporated.

Alternatively, low nucleotide diversities within se^{428} alleles (B3 and B4, $\pi = 0.00002$) and within se^{385} alleles (B6 and B7, $\pi = 0.00003$) suggest that frequencies of the se^{428} alleles and the se^{385} alleles might have increased relatively recently. Thus we might speculate that the B6 allele is on its way to fixation in the Asian populations and the B3 allele is on its way to fixation in the non-Asian populations by diversifying selection. The similarity of the null allele frequencies among the populations could be explained if we assume that the putative selective agent(s) had started to work almost simultaneously in the populations after the separation of Asian populations from the other populations, although the long persistence of B3 in the non-Asian populations is difficult to explain only by this model.

If polymorphism of *FUT2* is due to two-class overdominance for the functional and null alleles or diversifying selection of the null allele, what would be the selective advantage for heterozygotes of the functional and null alleles or for homozygotes of null alleles? This question needs attention because the null alleles are usually considered deleterious or neutral. One hint comes from studies on the α(1,3)galactosyltransferase gene. The α(1,3)galactosyltransferase gene had changed to a pseudogene in the lineage of Old World monkeys and hominoids after the separation from the lineage of New World monkeys (Galili and Swanson 1991). Old World monkeys and hominoids have a natural antibody against the Galα1-3Gal epitope. Galili and Andrews (1995) speculated that an infectious microbial agent containing the Galα1-3Gal epitope was endemic to the Old World after the speciation of Old World monkeys and New World monkeys, and individuals with the α(1,3)galactosyltransferase pseudogene had higher fitness than those with the α(1,3)galactosyltransferase functional gene. Thus, a null allele can have a selective advantage over functional alleles.

In the case of the *FUT2* locus, secretors who have at least one functional allele can express ABO(H) and Lewis b antigens in their secretions and digestive and respiratory mucosa, while nonsecretors with two null alleles do not. A recent study suggested that the Lewis b antigen was a possible receptor for *Helicobacter pylori* and that this microorganism was not bound to gastric mucosa lacking Lewis b expression (Boren *et al.* 1993). Most microorganisms may invade through the respiratory and digestive mucosa via ABO(H) and related carbohydrate structures (Ramphal *et al.* 1991; Plotkowski *et al.* 1993; Kauffmann *et al.* 1996). Accordingly, the absence of some carbohydrate structures in the respiratory and digestive mucosa may be advantageous for protecting mammals from microorganism infections and secretor status may play a role in susceptibility to some microorganisms (Raza *et al.* 1991). Although these arguments suggest just a possible advantage of the null allele over functional alleles, we can build various models such as marginal overdominance to make the two-allele model overdominant.

The two-class overdominance model and recent diversifying selection model are just two of the possible explanations for the contrasting pattern observed at *FUT2*. We need to investigate the physiological and epidemiological significance of the null alleles in future studies to show that selection is really operating on this locus. In addition, some complex demographic scenarios may also explain the pattern, although we did not explore them here. For example, Wall (2000) showed that admixture after a long time of separation of two populations is still compatible with the current DNA data. This or some other population structure may explain the long persistence of B3 without invoking balancing selection (see Tajima 1993). Also, expansion of human populations after bottleneck events was suggested by microsatellite data (Kimmel *et al.* 1998) and such bottleneck events can explain positive D values (see Tajima 1993) although only a few are significant in currently available data of nuclear polymorphisms observed at nuclear loci (Hey 1997; Nachman *et al.* 1998; Fay and Wu 1999; Hey and Harris 1999; Przeworski *et al.* 2000). Therefore, we also need to explore possible neutral models

to explain the contrasting pattern at the *FUT2* locus as the accumulation of data of human DNA polymorphism will make more accurate modeling of the past demographic history of human populations possible. Although our conclusion for the contrasting pattern at the *FUT2* locus is thus not definite, collecting DNA polymorphism data with random sampling certainly will enhance our understanding of the past population structure and give some hints about locus-specific factors such as selective agents acting on human populations.

We thank Dr. Yun-Xin Fu for his suggestion of the two-allele balancing selection model and helpful discussion and Dr. Ernette D. du Toit for providing blood samples from Cape Town, South Africa. This work was supported in part by grants-in-aid for Scientific Research from the Ministry of Education, Science and Culture of Japan, and grants from Uehara Memorial Foundation. H.T. was also supported in part by a grant from the Program for Promotion of Basic Research Activities for Innovative Biosciences (PROBRAIN).

LITERATURE CITED

Boren, T., P. Falk, K. A. Roth, G. Larson and S. Normark, 1993 Attachment of *Helicobacter pylori* to human gastric epithelium mediated by blood group antigens. Science **262:** 1892–1895.

Clark, A. G., K. M. Weiss, D. A. Nickerson, S. L. Taylor, A. Buchanan *et al.*, 1998 Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. Am. J. Hum. Genet. **63:** 595–612.

Fay, J., and C. I. Wu, 1999 A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. Mol. Biol. Evol. **16:** 1003–1005.

Fu, Y. X., and W. H. Li, 1993 Statistical tests of neutrality of mutations. Genetics **133:** 693–709.

Galili, U., and P. Andrews, 1995 Suppression of α-galactosyl epitopes synthesis and production of the natural anti-Gal antibody: a major evolutionary event in ancestral Old World primates. J. Hum. Biol. **29:** 433–442.

Galili, U., and K. Swanson, 1991 Gene sequences suggest inactivation of α-1,3-galactosyltransferase in catarrhines after the divergence of apes from monkeys. Proc. Natl. Acad. Sci. USA **88:** 7401–7404.

Harding, R. M., S. M. Fullerton, R. C. Griffiths, J. Bond, M. J. Cox *et al.*, 1997 Archaic African and Asian lineages in the genetic ancestry of modern humans. Am. J. Hum. Genet. **60:** 772–789.

Harris, E. H., and J. Hey, 1999 X chromosome evidence for ancient human histories. Proc. Natl. Acad. Sci. USA **96:** 3320–3324.

Hey, J., 1997 Mitochondrial and nuclear genes present conflicting portraits of human origins. Mol. Biol. Evol. **14:** 166–172.

Hey, J., and E. Harris, 1999 Population bottlenecks and patterns of human polymorphism. Mol. Biol. Evol. **16:** 1423–1426.

Hudson, R. R., M. Slatkin and W. P. Maddison, 1992 Estimation of levels of gene flow from DNA sequence data. Genetics **132:** 583–589.

Ingman, M., H. Kaessmann, S. Paabo and U. Gyllensten, 2000 Mitochondrial genome variation and the origin of modern humans. Nature **408:** 708–713.

Kauffmann, F., C. Frette, Q-T. Pham, S. Nafissi, J. P. Bertrand *et al.*, 1996 Associations of blood group-related antigens to EFV1, wheezing, and asthma. Am. J. Respir. Crit. Care Med. **153:** 76–82.

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.

Kimmel, M., R. Chakraborty, J. P. King, M. Bamshad, W. S. Watkins *et al.*, 1998 Signatures of population expansion in microsatellite repeat data. Genetics **148:** 1921–1930.

Koda, Y., M. Soejima, Y. Liu and H. Kimura, 1996 Molecular basis for secretor type α(1,2)-fucosyltransferase gene deficiency in a Japanese population: a fusion gene generated by unequal crossover responsible for the enzyme deficiency. Am. J. Hum. Genet. **59:** 343–350.

Koda, Y., H. Tachida, M. Soejima, O. Takenaka and H. Kimura, 2000 Ancient origin of the null allele se^{428} of the human ABO-secretor locus (*FUT2*). J. Mol. Evol. **50:** 243–248.

Kukowska-Latallo, J. F., R. D. Larsen, R. P. Nair and J. B. Lowe, 1990 A cloned human cDNA determines expression of a mouse stage-specific embryonic antigen and the Lewis blood group α(1,3/1,4)fucosyltransferase. Genes Dev. **4:** 1288–1303.

Larsen, R. D., L. K. Ernst, R. P. Nair and J. B. Lowe, 1990 Molecular cloning, sequence, and expression of human GDP-L-fucose: β-D-galactoside 2-α-L-fucosyltransferase cDNA that can form the H group antigen. Proc. Natl. Acad. Sci. USA **87:** 6674–6678.

Larson, G., C. Borjeson, A. Elmgren, A. Kernholt, S. Henry *et al.*, 1996 Identification of a new plasma α(1,3)fucosyltransferase (*FUT6*) allele requires an extended genotyping strategy. Vox Sang. **71:** 233–241.

Li, W.-H., 1975 The first arrival time and mean age of a deleterious mutant gene in a finite population. Am. J. Hum. Genet. **27:** 274–286.

Li, W. H., and L. A. Sadler, 1991 Low nucleotide diversity in man. Genetics **129:** 513–523.

Liu, Y.-H., Y. Koda, M. Soejima, H. Pang, T. Schlaphoff *et al.*, 1998 Extensive polymorphism of the *FUT2* gene in an African (Xhosa) population of South Africa. Hum. Genet. **103:** 204–210.

Liu, Y.-H., Y. Koda, M. Soejima, H. Pang, B.-J. Wang *et al.*, 1999 The fusion gene at the ABO-Secretor locus (*FUT2*): absence in Chinese populations. J. Hum. Genet. **44:** 181–184.

Lowe, J. B., J. F. Kukowska-Latallo, R. P. Nair, R. D. Larsen, R. M. Marks *et al.*, 1991 Molecular cloning of a human fucosyltransferase gene that determines expression of the $Lewis^x$ and VIM-2 epitopes but not ELAM-1-dependent cell adhesion. J. Biol. Chem. **266:** 17467–17477.

McCurley, R. S., A. Recinos III, A. S. Olsen, J. C. Gingrich, D. Szczepaniak *et al.*, 1995 Physical maps of human α(1,3)fucosyltransferase genes *FUT3-FUT6* on chromosomes 19p13.3 and 11q21. Genomics **26:** 142–146.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in Drosophila. Nature **351:** 652–654.

Mollicone, R., I. Reguigne, A. Fletcher, A. Aziz, M. Rustam *et al.*, 1994 Molecular basis for plasma α(1,3)-fucosyltransferase gene deficiency (*FUT6*). J. Biol. Chem. **269:** 12662–12671.

Nachman, M. W., V. L. Bauer, S. L. Crowell and C. F. Aquadro, 1998 DNA variability and recombination rates at X-linked loci in humans. Genetics **150:** 1133–1141.

Nei, M., and G. Livshits, 1990 Evolutionary relationships of Europeans, Asians, and Africans at the molecular level, pp. 251–265 in *Population Biology of Genes and Molecules*, edited by N. Takahata and J. F. Crow. Baifukan, Tokyo.

Nei, M., and F. Tajima, 1981 DNA polymorphism detectable by restriction endonucleases. Genetics **97:** 145–163.

Nickerson, D. A., S. L. Taylor, K. M. Weiss, A. G. Clark, R. G. Hutchinson *et al.*, 1998 DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. Nat. Genet. **19:** 233–240.

Pang, H., Y. Koda, M. Soejima, T. Schlaphoff, E. D. du Toit *et al.*, 1999 Allelic diversity of the human plasma α(1,3)fucosyltransferase gene (*FUT6*). Ann. Hum. Genet. **63:** 277–284.

Plotkowski, M. C., O. Bajolet-Laudinat and E. Puchelle, 1993 Cellular and molecular mechanisms of bacterial adhesion to respiratory mucosa. Eur. Respir. J. **6:** 903–916.

Przeworski, M., R. R. Hudson and A. D. Rienzo, 2000 Adjusting the focus on human variation. Trends Genet. **16:** 296–302.

Ramphal, R., C. Carnoy, S. Fievre, J. C. Nichalski, N. Houdret *et al.*, 1991 *Pseudomonas aeruginosa* recognizes carbohydrate chains containing type 1 (Galβ1-3GlcNAc) or type 2 (Galβ1-4GlcNAc) disaccharide units. Infect. Immun. **59:** 700–704.

Rana, B. K., D. Hewett-Emmett, L. Jin, B. H.-J. Chang, N. Sambuughin *et al.*, 1999 High polymorphism at the human melanocortin 1 receptor locus. Genetics **151:** 1547–1557.

Raza, M. W., C. C. Blackwell, P. Molyneaux, V. S. James, M. M. Ogilvie *et al.*, 1991 Association between secretor status and respiratory viral illness. Br. Med. J. **303:** 815–818.

Rouquier, S., J. B. Lowe, R. J. Kelly, A. L. Fertitta, G. G. Lennon *et al.*, 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.

Rozas, J., and R. Rozas, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

Sasaki, K., K. Kurata, K. Funayama, M. Nagata, E. Watanabe *et al.*, 1994 Expression cloning of a novel α1,3fucosyltransferase that is involved in biosynthesis of the sialyl Lewis^x carbohydrate determinants in leukocytes. J. Biol. Chem. **269:** 14730–14737.

Simonsen, K. L., G. A. Churchill and C. F. Aquadro, 1995 Properties of statistical tests of neutrality for DNA polymorphism data. Genetics **141:** 413–429.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Tajima, F., 1993 Statistical analysis of DNA polymorphism. Jpn. J. Genet. **68:** 567–595.

Wall, J. D., 2000 Detecting ancient admixture in humans using sequence polymorphism data. Genetics **154:** 1271–1279.

Watterson, G. A., 1975 On the number of segregating sites in genetical model without recombination. Theor. Popul. Biol. **7:** 256–276.

Weston, B. W., R. P. Nair, R. D. Larsen and J. B. Lowe, 1992a Isolation of a novel human α(1,3)fucosyltransferase gene and molecular comparison to the human Lewis blood group α(1,3/1,4)fucosyltransferase gene. Syntenic, homologous, nonallelic genes encoding enzymes with distinct acceptor substrate specificities. J. Biol. Chem. **267:** 4152–4160.

Weston, B. W., P. L. Smith, R. J. Kelly and J. B. Lowe, 1992b Molecular cloning of a fourth member of a human α(1,3)fucosyltransferase gene family. Multiple homologous sequences that determine expression of the Lewis^x, sialyl Lewis^x, and difucosyl sialyl Lewis^x epitopes. J. Biol. Chem. **267:** 24575–24584.

Zietkiewicz, E., V. Yotova, M. Jarnik, M. Korab-Laskowska, K. K. Kidd *et al.*, 1998 Genetic structure of the ancestral populations of modern humans. J. Mol. Evol. **47:** 146–155.

Communicating editor: N. Takahata