Human-Specific Amino Acid Changes Found in 103 Protein-Coding Genes

Takashi Kitano, Yu-Hua Liu,*[1] Shintaroh Ueda,† and Naruya Saitou**

*Division of Population Genetics, National Institute of Genetics, Mishima, Japan; and †Department of Biological Sciences, Graduate School of Science, University of Tokyo, Tokyo, Japan

We humans have many characteristics that are different from those of the great apes. These human-specific characters must have arisen through mutations accumulated in the genome of our direct ancestor after the divergence of the last common ancestor with chimpanzee. Gene trees of human and great apes are necessary for extracting these human-specific genetic changes. We conducted a systematic analysis of 103 protein-coding genes for human, chimpanzee, gorilla, and orangutan. Nucleotide sequences for 18 genes were newly determined for this study, and those for the remaining genes were retrieved from the DDBJ/EMBL/GenBank database. The total number of amino acid changes in the human lineage was 147 for 26,199 codons (0.56%). The total number of amino acid changes in the human genome was, thus, estimated to be about 80,000. We applied the acceleration index test and Fisher's synonymous/nonsynonymous exact test for each gene tree to detect any human-specific enhancement of amino acid changes compared with ape branches. Six and two genes were shown to have significantly higher nonsynonymous changes at the human lineage from the acceleration index and exact tests, respectively. We also compared the distribution of the differences of the nonsynonymous substitutions on the human lineage and those on the great ape lineage. Two genes were more conserved in the ape lineage, whereas one gene was more conserved in the human lineage. These results suggest that a small proportion of protein-coding genes started to evolve differently in the human lineage after it diverged from the ape lineage.

Introduction

We now have a clear picture of the phylogenetic constellation of human (*Homo sapiens*): chimpanzee (*Pan troglodytes*) and bonobo (*Pan paniscus*) are equally closely related organisms to human (e.g., Sibley and Ahlquist [1984], Saitou [1991], and Horai et al. [1995]). Nucleotide substitution difference between human and chimpanzee was estimated to be 1.23% based on 19-Mb BAC end sequence comparison (Fujiyama et al. 2002). This difference corresponds to 3.7 million bp for the whole genome under the assumption that the human and chimpanzee genomes are both approximately 3 billion nucleotides. Many of those differences were probably caused by mutations occurred in so-called junk DNA (Ohno 1972) and had no effect on phenotypic difference between human and chimpanzee. Some proportion of nucleotide changes, however, must be responsible for human-specific characteristics, such as large brain size and bipedalism. King and Wilson (1975) proposed that genetic changes at the gene expression control region are more important than changes in the protein-coding region. However, there are more than 32,000 genes in the human genome (International Human Genome Sequencing Consortium 2001), and a considerable number of protein-coding genes must produce proteins that slightly differ in amino acid sequences between human and chimpanzee. Some of those amino acid differences may be responsible for human-specific characteristics.

It is our interest to determine whether amino acid changes occurred in the human lineage after the last common ancestor diverged from the chimpanzee lineage. Such changes are the candidates for the genetic basis of human-specific characteristics. Gene trees of human and great apes are necessary for extracting those genetic changes that occurred in the human lineage. There are three possible gene trees for human, chimpanzee, and gorilla (see figure 1). Because the speciation period of human and chimpanzee is difficult to infer, the "human lineage" in this paper is defined as branch connecting the present-day human and the last branching point designated as a circle in figure 1.

The tree topology of human, chimpanzee, and gorilla differed from gene to gene. Satta, Klein, and Takahata (2000) compared 34 genes and found that about 60% of loci supported the human-chimpanzee clade, and the remaining 40% supported one of the two alternative trees. Chen and Li (2001) and O'hUigin et al. (2002) compared 53 segments and 51 genes, respectively, and they also found a similar tendency. Because the time span between the human-chimpanzee common ancestor and gorilla speciation is short (approximately 1 to 2 Myr), gene genealogies might differ from gene to gene.

The majority of genes is evolving under neutral fashion, and natural selection plays mainly a conservative role as negative or purifying selection (Kimura 1983; Nei 1987). Nevertheless, a small portion of genes is under positive selection, and evidence of positive selection at the molecular level has been accumulated through comparison of synonymous and nonsynonymous substitutions since it was first found for MHC genes (Hughes and Nei 1988, 1989). Even if we restrict our attention to primates, 15 genes were so far shown to experience positive selection (table 1). We, therefore, also compared synonymous and nonsynonymous substitutions to identify human lineage–specific positive selection.

Materials and Methods

DNA Sequencing

We determined nucleotide sequences of the 18 protein-coding loci (A4GALT, B3GALT1, B3GALT5, CHRM2, CHRM3, CX36, HRH1, HRH2, HTR1A, HTR1E, HTR1F, HTR2A, NGFB, NPPB, OTX1, OTX2, SCN8A, and SIX6) for human, chimpanzee, gorilla, and

[1] Present address: The Jackson Laboratory, Bar Harbor, Maine.

Key words: humanness, positive selection, hominoids, gene tree.

E-mail: nsaitou@genes.nig.ac.jp.

Mol. Biol. Evol. 21(5):936–944. 2004
DOI:10.1093/molbev/msh100
Advance Access publication March 10, 2004

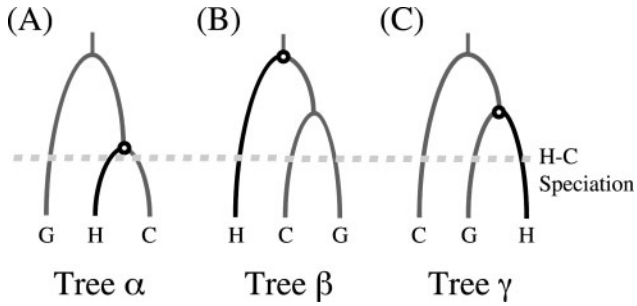

FIG. 1.—Three possible gene trees (α, β, and γ) for human, chimpanzee, and gorilla. Gene tree α (*A*) has the same topology with the species tree. (*B*) shows a gene tree of a topology ((chimpanzee, gorilla), human, orangutan) (β). (*C*) shows a gene tree of a topology ((human, gorilla), chimpanzee, orangutan) (γ).

Table 1
List of Genes Found to be Under Positive Selection in Primate Evolution

Gene	Reference
BRCA1	Huttley et al. (2000)
RNASE3 (ECP)	Zhang, Rosenberg, and Nei (1998)
RNASE2 (EDN)	Zhang, Rosenberg, and Nei (1998)
FOXP2	Enard et al. (2002)
IGHA[a]	Sumiyama, Saitou, and Ueda (2002)
NPIP[a] (LCR16a)	Johnson et al. (2001)
LYZ	Messier and Stewart (1997)
PRM1	Wyckoff, Wang, and Wu (2000)
PRM2	Wyckoff, Wang, and Wu (2000)
RHAG	Kitano et al. (1998)
RH[a]	Kitano and Saitou (1999)
TNP2	Wyckoff, Wang, and Wu (2000)

[a] These genes were not used in this study, because gorilla sequence was not available for NPIP, and gene duplication occurred after orangutan divergence for IGHA and RH.

orangutan. Because we needed to compare coding sequences (CDS) of human and great apes, we arbitrarily chose genes that can be easily amplified (relatively long CDS region [>300 bp]) from genomic DNA. In fact, 10 out of 18 genes were single CDS genes. DNA was extracted from peripheral blood sample of a certain Japanese individual, with informed consent for human (*Homo sapiens*) samples. DNA samples for chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*), and orangutan (*Pongo pygmaeus*) were also used. Each PCR reaction mixture contained 0.2 mM dNTP, 1.5 mM $MgCl_2$, 1×GeneTaq Mg^{2+} free Universal Buffer (Nippon Gene), 10 pmol of each primer, and 1 unit AmpliTaq Gold (PE Biosystems). A list of primers used in this study is available in the Supplementary Material online. The typical PCR condition used in this study consisted of 40 cycles of 30 s denaturation at 95°C, followed by 15 s primer annealing at 60°C and 1 min extension at 72°C (PE GeneAmp PCR system 2400 and 9700). Immediately proceeding these cycles, a 10 min hot-start step at 95°C was included. All PCR primers were designed based on human sequences in the database. PCR products were confirmed by 1.5% agarose gel electrophoresis and purified using Micro Spin Columns (Amersham Biosciences). The purified PCR products were sequenced by using BigDye Terminator Cycle Sequencing Kit and ABI PRISM 377/310 DNA sequencer (PE Biosystems). When long sequences were read, both strands were read using PCR primers and inner primers.

Sequence Data Retrieval and Analyses

Eighty-five protein-coding gene sequences were retrieved from the DDBJ/EMBL/GenBank International Nucleotide Sequence Database (Supplementary Material online). This data set contains human, chimpanzee, gorilla, and orangutan sequences longer than 100 bp. To retrieve those sequences, we used orangutan sequences as queries for Blast search to obtain homologous sequences for human, chimpanzee, and gorilla. For human sequences, we used sequences that were cited in the NCBI Reference Sequences (RefSeq) as representative ones. When more than two sequences were found from one ape species, we used the sequence that showed the shortest branch in the neighbor-joining tree (Saitou and Nei 1987). ClustalW version 1.8 (Thompson, Gibson, and Higgins 1994) was used for multiple alignments. Tree topologies were determined by counting numbers of informative sites (Supplementary Material online). Tree α shows human-chimpanzee cluster, tree β shows chimpanzee-gorilla cluster, and tree γ shows human-gorilla cluster. Genes with unclear topology (trichotomy or same number of informative sites support different trees) was categorized into group δ. Genes belonging to group δ were recategorized into δ-α, δ-β, and δ-γ by using the UPGMA method (Sneath and Sokal 1973). Program pamp in PAML package (Yang 1997) was used for reconstruction of internal nodes of sequences. ODEN package (Ina 1994) was used for estimation of synonymous and nonsynonymous substitutions (Nei and Gojobori 1986).

Statistical Tests

We used two kinds of statistical tests for detecting human-specific natural selection. One test is the acceleration index test for nonsynonymous substitutions of human and apes. This test analogous to the test of Zhang, Webb, and Podlaha (2002), in which an acceleration index for the human lineage in comparison to the mammalian lineage before the human-chimpanzee split is defined by the equation $(h/5.5)/[m/(2 \times 90 - 5.5)] = 31.7h/m$. The variables h and m are numbers of amino acid substitutions in the human lineage and the mouse lineage, respectively. Zhang, Webb, and Podlaha (2002) used divergence times between human and chimpanzee (5.5 MYA) and between primates and rodents (90 MYA). They also computed the tail probability in a binomial distribution of $B(h+m, 0.03056)$ for testing the statistical significance of rate enhancement in the human lineage. The value 0.03056 is from 5.5/180, the time span for human branch, relative to that for primates and rodents branches. We applied this test as B(n-Human + n-Ape, 0.13268) to determine the statistical significance of rate enhancement in the human lineage in contrast to ape lineage (n-Human and n-Ape are numbers of nonsynonymous changes at the human lineage and ape lineage, respectively). In our test, $0.13268 = 5.4/$

40.7, is the ratio of the time span for human branch (5.4 MYA) to the time span for ape lineages (40.7 MYA). We used divergence times estimated by Chen and Li (2001). Taking the orangutan speciation date as approximately 12 to 16 MYA (midpoint is 14 MYA) (Goodman et al. 1998), they obtained an estimate of 4.6 to 6.2 MYA (midpoint is 5.4 MYA) for the human and chimpanzee divergence and an estimate of 6.2 to 8.4 MYA (midpoint is 7.3 MYA) for the gorilla speciation date, suggesting that the gorilla lineage branched off approximately 1.6 to 2.2 MYA (midpoint is 1.9 MYA) earlier than did the human and chimpanzee divergence. The time span between the ancestor of human-chimpanzee-gorilla-orangutan and the ancestor of human-chimpanzee-gorilla can be estimated to 6.7 MYA (= 14 MYA to 7.3 MYA). For simplicity, we took midpoint values and assumed the species tree. Therefore, the total divergence time of hominoid lineages is 40.7 MYA = 5.4 MYA + 5.4 MYA + 1.9 MYA + 7.3 MYA + 6.7 MYA + 14.0 MYA.

We also used Fisher's exact test (two tails) for synonymous and nonsynonymous substitutions of human and apes. This test is analogous to the test of McDonald and Kreitman (1991), in which silent and amino acid replacement changes for polymorphic and fixed differences were compared. We compared silent and amino acid replacement changes for human and ape branches in this study.

Results and Discussion

Topologies of Gene Trees for Human, Chimpanzee, and Gorilla

Orthology for each gene was checked by examining the total number of synonymous substitutions for all branches for a gene. The average value for 103 genes was 0.06, and the maximum value was 0.14 ± 0.05 for the PRM2 gene. Because the amount of synonymous substitution of PRM2 was not significantly larger than the average, we can expect that all compared genes are orthologous.

We divided the 103 protein-coding genes into four groups by its tree topology using the parsimony method. Human and chimpanzee are clustered in tree α (fig. 1*A*), and it is the same as the species tree. Chimpanzee and gorilla are clustered in tree β (fig. 1*B*), and human and gorilla are clustered in tree γ (fig. 1*C*). Genes with unclear topology (trichotomy, or same number of informative sites, support different trees) was categorized into group δ. Numbers of nucleotide sites supporting each tree for each gene are shown in Supplementary Material online. Thirty-four genes were group α, 10 genes were group β, 14 genes were group γ, and the remaining 45 genes were group δ (table 2). The 45 genes in trifurcating tree (group δ) were further classified by using UPGMA, under the assumption of approximate constancy of the evolutionary rate.

If we consider the proportion of tree topology by using 58 genes (45 genes in group δ were excluded), 59% of the coding genes supported tree α, 17% supported tree β, and 24% supported tree γ. Satta, Klein, and Takahata (2000) showed that from 34 nuclear loci, 59% of loci supported tree α, 21% of loci supported tree β, and 21% of loci supported tree γ (they used both coding and noncoding regions). Chen and Li (2001) reported that 58% of loci supported tree α, 23% of loci supported tree β, and 19% of loci supported tree γ from a comparison of 53 autosomal intergenic noncoding DNA segments. Proportions of three categories (α, β, and γ) estimated from the present study with a larger number of genes were similar to these previous studies. When we consider the proportion of tree topology by using 94 genes, where 36 genes in group δ were reclassified to α, β, or γ by using UPGMA, similar proportions were observed (52% for tree α, 26% for tree β, and 22% for tree γ).

Table 2
Distribution of Genes According to Gene Tree Topology and Numbers of Synonymous (s) and Nonsynonymous (n) Substitutions on Human and Ape Branches

Tree	Number of Genes[a]	Human Branch[b]		Ape Branch[b]	
		s	n	s	n
α	34 (49)	36.5 (52.5)	44.5 (58.5)	377.5 (485.0)	371.5 (495.0)
β	10 (24)	24.0 (56.0)	26.0 (43.0)	84.0 (168.0)	73.0 (121.0)
γ	14 (21)	24.0 (28.0)	44.0 (46.0)	181.0 (225.0)	235.0 (251.0)
δ	45 (9)	55.0 (2.0)	36.0 (2.0)	260.0 (24.5)	239.0 (50.5)

[a] Numbers in parentheses are values when genes initially clustered to group δ were reassigned by assuming rate constancy.

[b] Numbers in parentheses are values, including group δ genes, reassigned by assuming rate constancy.

We also classified the total number (182) of informative sites into those supporting the three possible trees: 55% of sites supported tree α, 21% supported tree β, and 24% supported tree γ. These proportions are similar to those estimated from numbers of genes. O'hUigin et al. (2002) estimated that the 53% of the informative nucleotide sites supported tree α, 31% supported tree β, and 16% supported tree γ from 87 informative sites found in 51 genes. The result of the present study showed more uniform distribution of two alternative informative sites.

If we compare different gene trees, the branch length of the human lineage for tree β is expected to be longer than those of tree α and tree γ under the assumption of the molecular clock. We, thus, compared the number of synonymous substitutions (*dS*) for each branch of three gene trees (table 3). As expected, *dS* of the human branches for tree β were longer than those of tree α and tree γ, with clear statistical significance. This is consistent with the topological difference between tree β and the remaining two trees. Human forms a cluster with chimpanzee or gorilla in tree α and γ, whereas human is an outgroup to the chimpanzee-gorilla clade in tree β (see figure 1). Similarly, the branch length of the chimpanzee lineage for tree γ is expected to be longer than those of tree α and tree β, and the branch length of the gorilla lineage for tree α is expected to be longer than those of tree β and tree γ. However, clear results were not obtained. This finding may be caused by a smaller number of compared genes.

We also expect that the internal branch of tree α is the longest among the three gene trees because the topology of tree α is the same as that of the species tree. In fact, the internal branch length of tree α was about two times longer than those of tree β and tree γ (table 3). When we consider *dS* by including group δ genes reassigned by assuming the

Table 3
Number of Total Synonymous Substitutions per Synonymous Sites (*d*S) for Each Tree

Branch	α	β	γ	Significance
Human	0.0053 ± 0.0009	0.0157 ± 0.0032	0.0059 ± 0.0012	α-β**, β-γ**
	(0.0059 ± 0.0008)	(0.0146 ± 0.0020)	(0.0056 ± 0.0011)	α-β**, β-γ**
Chimpanzee	0.0089 ± 0.0011	0.0085 ± 0.0024	0.0052 ± 0.0011	α-γ*
	(0.0076 ± 0.0009)	(0.0067 ± 0.0013)	(0.0068 ± 0.0012)	—
Gorilla	0.0104 ± 0.0012	0.0105 ± 0.0026	0.0104 ± 0.0016	—
	(0.0112 ± 0.0011)	(0.0083 ± 0.0015)	(0.0097 ± 0.0014)	—
Orangutan	0.0290 ± 0.0021	0.0318 ± 0.0046	0.0258 ± 0.0026	—
	(0.0298 ± 0.0018)	(0.0271 ± 0.0027)	(0.0260 ± 0.0023)	—
Internal	0.0075 ± 0.0010	0.0045 ± 0.0017	0.0037 ± 0.0010	α-γ**
	(0.0058 ± 0.0008)	(0.0018 ± 0.0007)	(0.0030 ± 0.0008)	α-β**, α-γ*

NOTE.—Numbers in parentheses are values, including group δ genes, reassigned by assuming rate constancy. Statistical significance was tested for tree α versus tree β and tree α versus tree γ. * Indicates significant at 5% level. ** Indicates significant at 1% level.

rate constancy, similar results were observed (table 3). These results suggest that even if gene tree topologies differ from the species tree topology, these genes are considered to be orthologous when polymorphism of an ancestral population is assumed.

Synonymous and Nonsynonymous Changes on the Human Branch

We compared numbers of synonymous and nonsynonymous substitutions for human and ape branches (see table 4). The ape branch denotes the sum of all branch lengths of the tree except for the human branch. We applied the acceleration index test (Zhang, Webb, and Podlaha 2002) for nonsynonymous substitutions of human and apes, and six genes (APOE, BRCA1, FOXP2, HCR, PRM2, and ZFY) showed acceleration at the human lineage with statistical significance at the 5% level (table 5). Zhang, Webb, and Podlaha (2002) applied this test to amino acid changes of 120 genes among human, chimpanzee, and mouse, and identified FOXP2 and PRM2, with significantly enhanced evolutionary rates in the human lineage (table 5). FOXP2 and PRM2 genes were also reported to be under positive selection on the human branch using different tests by Enard et al. (2002) and Wyckoff, Wang, and Wu (2001), respectively.

Both FOXP2 and PRM2 genes were also determined to have experienced human-specific acceleration in the present study. However, probabilities for these two genes were increased in this study, from 0.003 to 0.048 for FOXP2 and from 0.001 to 0.005 for PRM2 (table 5). Interestingly, the other four genes (APOE, BRCA1, HCR, and ZFY) showed the reversed tendency; *P*(A.I.-ape) is much lower than *P*(A.I.-mouse). This shows the stronger power of the test used in this study.

Zhang, Webb, and Podlaha (2002) used human, chimpanzee, and mouse sequence data to analyze human-specific selection of genes. Outgroup species are necessary to estimate the human lineage–specific changes; however, mouse may be too far removed to be used as an outgroup. Recently, Clark et al. (2003) compared coding regions of mouse, human, and chimpanzee but did not find many genes with significantly higher nonsynonymous substitutions in the human lineage. More closely related species, such as gorilla and orangutan, are appropriate as outgroups for the kind of analysis we conducted in the present study.

The BRCA1 gene was determined to be under positive selection on human and chimpanzee branches (Huttley et al. 2000). APOE codes apolipoproteins involved in cholesterol metabolism. Three major isoforms are known for human APOE (Weisgraber, Rall, and Mahley 1981), and these alleles differ in their association with hyperlipoproteinemia (Rall et al. 1982) and Alzheimer disease risk (Corder et al. 1993). It is possible that the cholesterol metabolism underwent different evolutionary pressures between the human and the ape branches. The HCR gene locates near the HLA-C locus and is a candidate gene for psoriasis (Asumalahti et al. 2000). However, there is so far no report of positive selection on this gene. The ZFY gene encodes a zinc finger–containing protein that may function as a transcription factor. Differential rates of evolution of the ZFY-related genes were recently observed in mice species (Tucker, Adkins, and Rest 2003).

We also applied Fisher's exact test (two tails) for synonymous and nonsynonymous substitutions of human and apes. Three genes (BRCA1, FOXP2, and DAF) showed statistical significance at the 5% level (table 5). BRCA1 and FOXP2 genes showed significant enhancement of nonsynonymous substitutions in the human lineage, as also found by using the acceleration test, whereas the DAF (decay accelerating factor) gene showed significant reduction at the human lineage. Generally speaking, *P*(n/s) values are higher than *P*(A.I.-ape) values except for the FOXP2 gene.

There are two equally parsimonious trees for the DAF gene. Figure 2 shows these two trees (shown with bold lines in A and B) on the phylogenetic network. This gene was categorized into group δ-α, corresponding to tree A of figure 2. Four synonymous and two nonsynonymous substitutions on the human branch and six synonymous and 28 nonsynonymous substitutions on ape branches were assumed for tree A. The number of nonsynonymous substitutions on the human branch is significantly smaller than those of other branches (table 5). When the alternative maximum-parsimonious tree (fig. 2*B*) is considered,

Table 4
Number of Synonymous and Nonsynonymous Substitutions on Human and Ape Branches

	Human Branch				Ape Branch			
Gene	s	*dS*	n	*dN*	s	*dS*	n	*dN*
Tree α								
ACAT2	0	0.000	2	0.019	1	0.025	2	0.019
ADRB2	1	0.004	2	0.002	13	0.050	11	0.013
APOB	1	0.007	4	0.009	4	0.030	12	0.027
C5R1	0	0.000	0	0.000	21.5	0.084	18.5	0.025
CCR5	1	0.004	2	0.003	16	0.066	2	0.003
CD209	1	0.004	2	0.002	19	0.071	38	0.042
CHRM2	1	0.003	0	0.000	15	0.050	2	0.002
COX8	0	0.000	1	0.007	4	0.071	1	0.007
CSTB	0	0.000	0	0.000	5	0.081	1	0.004
CXCR4	1	0.004	0	0.000	6	0.024	1	0.001
CXCR6	0	0.000	1	0.003	6	0.058	1	0.003
DMP1	1	0.006	2	0.003	7	0.045	19	0.027
EKN1	3	0.011	2	0.002	5	0.019	7	0.007
F9	0	0.000	0	0.000	4	0.036	8	0.022
FPRL1	0	0.000	1	0.001	21	0.085	16	0.021
FPRL2	2.5	0.010	4.5	0.006	18	0.077	17	0.022
FUT2	3	0.012	1	0.001	19.5	0.082	15.5	0.020
HRH2	0	0.000	0	0.000	22	0.084	0	0.000
HTR1E	1	0.004	0	0.000	16.5	0.066	2.5	0.003
HTR1F	2	0.008	0	0.000	3	0.012	6	0.007
IL3	0	0.000	1	0.004	6	0.091	4	0.017
IL8RB	3	0.011	0	0.000	16	0.061	19	0.024
LCAT	1	0.008	0	0.000	9	0.074	4	0.011
MEFV	1	0.009	0	0.000	14	0.131	13	0.037
ODC1	0	0.000	1	0.002	8	0.059	3	0.007
OTX1	0	0.000	0	0.000	9	0.049	1	0.002
RHBG	1	0.003	3	0.003	18	0.054	13	0.013
RNASE1	2	0.018	2	0.006	6	0.056	8	0.023
RPS4Y	1	0.005	0	0.000	16	0.091	9	0.015
TAF1L	2	0.004	5	0.003	20	0.036	34	0.017
TGIF2LX	1	0.006	1	0.002	6	0.037	32	0.060
TNP2	0	0.000	3	0.010	8	0.102	17	0.056
TWIST1	4	0.027	2	0.005	6	0.042	8	0.018
ZNF80	2	0.011	2	0.003	9	0.052	26	0.042
Total	36.5	0.005	44.5	0.002	377.5	0.057	371.5	0.017
Tree β								
APOE	3	0.013	6	0.008	7	0.030	7	0.010
B3GALT5	3	0.015	3	0.004	18	0.096	13	0.019
CX36	2	0.011	0	0.000	5	0.029	0	0.000
LEP	2	0.019	0	0.000	3	0.029	6	0.018
OXTR	4	0.018	3	0.005	9	0.040	6	0.009
POMC	2	0.017	1	0.003	9	0.081	8	0.022
PRM1	0	0.000	3	0.026	3	0.088	15	0.141
PRM2	3	0.043	8	0.035	7	0.106	14	0.062
SCG2	4	0.020	2	0.003	8	0.040	4	0.005
UBB	1	0.006	0	0.000	15	0.098	0	0.000
Total	24	0.016	26	0.005	84	0.057	73	0.015
Tree γ								
AFP	0	0.000	0	0.000	3	0.101	0	0.000
BRCA1	2	0.003	17	0.006	24	0.035	40	0.015
CD209L1	0	0.000	2	0.005	6	0.050	6	0.015
CD22	0	0.000	4	0.005	6	0.027	35	0.047
CHRM3	3	0.007	0	0.000	17	0.042	11	0.008
FOXP2	1	0.002	2	0.001	19	0.042	1	0.001
FUT5	3	0.011	2	0.002	21.5	0.082	29.5	0.036
FUT6	3	0.011	3	0.004	23	0.093	29	0.037
HCR	5	0.009	7	0.004	20	0.035	17	0.010
INS	3	0.035	0	0.000	5	0.059	4	0.017
NPPB	1	0.010	1	0.003	1	0.010	9	0.031
RHAG	0	0.000	3	0.003	14.5	0.051	31.5	0.035
SRY	0	0.000	2	0.004	7	0.055	16	0.034
TYR	3	0.009	1	0.001	14	0.041	6	0.005
Total	24	0.006	44	0.003	181	0.046	235	0.018

Table 4
Continued

	Human Branch				Ape Branch			
Gene	s	*dS*	n	*dN*	s	*dS*	n	*dN*
Tree δ–α								
A4GALT	2	0.013	0	0.000	12	0.081	7	0.015
C1orf9	0	0.000	0	0.000	4	0.033	2	0.005
DAF	4	0.016	2	0.003	6.5	0.027	28.5	0.038
DEFB1	0	0.000	0	0.000	3	0.062	2	0.013
DRD4	0	0.000	1	0.005	5	0.064	4	0.021
FPR1	4	0.015	3	0.004	15	0.060	15	0.020
FUT1	1	0.004	0	0.000	6	0.023	7	0.009
IL16	1	0.004	2	0.003	14	0.061	12	0.017
LPL	0	0.000	0	0.000	4	0.083	0	0.000
NPPA	1	0.012	0	0.000	4	0.051	4	0.016
RNASE2	2	0.018	0	0.000	11	0.107	14	0.039
RNASE6	1	0.010	0	0.000	6	0.063	4	0.011
SIGLECL1	0	0.000	4	0.007	14	0.075	15	0.026
TNF	0	0.000	1	0.003	1	0.008	5	0.015
ZNF75	0	0.000	1	0.002	2	0.016	4	0.008
Total	16	0.007	14	0.002	107.5	0.050	123.5	0.018
Tree δ–β								
ADRB3	3	0.011	3	0.005	10	0.037	7	0.011
B3GALT1	5	0.023	0	0.000	9	0.042	0	0.000
CMAH	0	0.000	1	0.007	0	0.000	1	0.007
COX4I1	2	0.024	0	0.000	2	0.024	6	0.018
EPO	2	0.020	1	0.003	7	0.073	7	0.024
HRH1	3	0.009	4	0.004	16	0.049	13	0.012
HTR1A	2	0.006	3	0.003	11	0.035	4	0.004
LYZ	2	0.019	0	0.000	2	0.019	4	0.012
NGFB	4	0.024	2	0.004	7	0.042	4	0.007
PABPC5	1	0.004	0	0.000	0	0.000	0	0.000
SIX6	3	0.025	1	0.003	7	0.061	0	0.000
ZFX	1	0.013	0	0.000	0	0.000	0	0.000
ZFY	1	0.013	2	0.006	6	0.080	1	0.003
ZNFN1A1	3	0.023	0	0.000	7	0.055	1	0.002
Total	32	0.014	17	0.002	84	0.039	48	0.006
Tree δ–γ								
APOA1	0	0.000	0	0.000	2	0.015	4	0.009
HTR2A	2	0.012	0	0.000	8	0.050	2	0.004
IFNG	0	0.000	1	0.003	10	0.126	2	0.006
MSH2	0	0.000	0	0.000	2	0.050	2	0.016
OTX2	0	0.000	0	0.000	3	0.038	0	0.000
PRNP	1	0.006	1	0.002	14	0.084	5	0.009
SCN8A	1	0.004	0	0.000	5	0.020	1	0.001
Total	4	0.004	2	0.001	44	0.048	16	0.005
Tree δ–δ								
ANG	1	0.009	0	0.000	2.5	0.024	11.5	0.035
B2M	0	0.000	0	0.000	1	0.012	4	0.015
COX7C	0	0.000	0	0.000	3	0.070	1	0.007
HBE1	0	0.000	0	0.000	3	0.030	2	0.006
OPN1SW	0	0.000	0	0.000	4	0.075	1	0.005
RET	0	0.000	1	0.007	3	0.062	4	0.026
RNASE3	0	0.000	1	0.003	4	0.035	27	0.078
TCP1	1	0.028	0	0.000	2	0.056	0	0.000
TH	0	0.000	0	0.000	2	0.050	0	0.000
Total	2	0.003	2	0.001	24.5	0.039	50.5	0.025

however, two synonymous and five nonsynonymous substitutions on human branch and six synonymous and 27 nonsynonymous substitutions on other branches were observed. In this case, number of substitutions between human and other branches is not statistically significant. The DAF gene codes a glycoprotein and it is related to Cromer blood group system (CR) (Reid et al. 1996). Kuttner-Kondo et al. (2000) mentioned that number of amino acid changes differ region by region in primate DAF genes. This gene may have a human-specific nucleotide substitution pattern, but it depends on a topology to be analyzed. More detailed analyses might be needed for this gene.

A total of 147 amino acid changes were observed in the human lineage for 26,199 codons (0.56%). About 60% of amino acid changes were radical changes. If we assume that the number of genes in human genome is 32,000 and

Table 5
Genes Showing Significantly Different Nonsynonymous Changes Between Human Lineage and Ape Lineages

Gene	n-Human	n-Ape	*P*(A.I.-ape)	*P*(A.I.-mouse)	*P*(n/s)
BRCA1	17	40	0.001**	0.568[a]	0.027*
APOE	6	7	0.004**	0.106[a]	0.670
PRM2	8	14	0.005**	0.001[a]**	1.000
HCR	7	17	0.032*	0.352[b]	0.520
FOXP2	2	1	0.048*	0.003[a]**	0.034*
ZFY	2	1	0.048*	0.267[b]	0.183
DAF tree A	2	28	0.926	0.944[b]	0.026*
DAF tree B	5	27	0.422	0.944[b]	0.611

NOTE.—n-Human and n-Ape are numbers of nonsynonymous changes at the human lineage and ape lineage, respectively. *P*(A.I.-ape) and *P*(A.I.-mouse) are probabilities of acceleration index for the human lineage compared with the ape (chimpanzee, gorilla, and orangutan) lineage and compared with the mouse lineage, respectively. *P*(n/s) is the probability of Fisher's exact test (two tails) for synonymous and nonsynonymous substitutions between the human and ape lineages. * Indicates significant at 5% level. ** Indicates significant at 1% level.

[a] Probability was estimated from data presented in Zhang, Webb, and Podlaha (2002).

[b] Probability was estimated using the mouse orthologs given in NCBI reference sequences (HCR: BC031416, ZFY: M24401, and DAF: L41366).

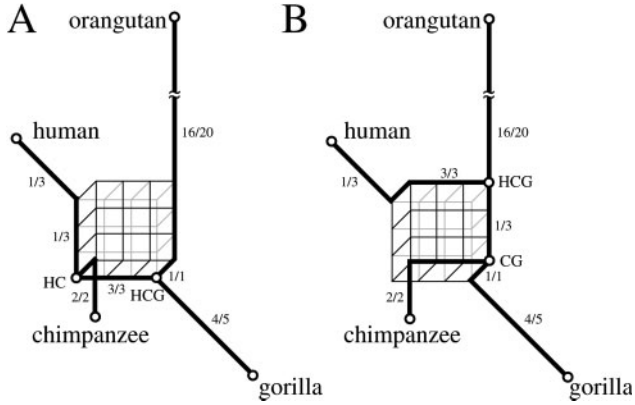

FIG. 2.—Maximum-parsimony trees embedded in a phylogenetic network of DAF. Numbers indicate number of nonsynonymous substitutions/total number of nucleotide substitutions. Bold lines show the maximum-parsimony pathway in each network. (*A*) is a topology ((human, chimpanzee), gorilla, orangutan), and HC and HCG nodes are the human-chimpanzee common ancestor and the human-chimpanzee-gorilla common ancestor. (*B*) is a topology ((chimpanzee, gorilla), human, orangutan), and CG and HCG nodes are the chimpanzee-gorilla common ancestor and the human-chimpanzee-gorilla common ancestor.

the mean number of amino acid residues is 447 (International Human Genome Sequencing Consortium 2001), the number human genome-wide amino acid changes is estimated to be 80,258. Considerable numbers of those amino acid differences may be responsible for human-specific characteristics.

Differences of Nonsynonymous Substitutions for Human and Ape Branch

Comparison of synonymous and nonsynonymous substitutions for each coding region is a standard way of detecting the pattern of natural selection. However, the number of synonymous substitutions may undergo stochastic changes, and it can be rather small for one gene but become large in another gene. We therefore decided to compare the number of human and ape nonsynonymous substitutions (*dN*) with the number of synonymous substitutions (*dS*) for each gene. Human and ape *dN*s were positively correlated with high statistical significance ($R^2 = 0.33$, $P = 2.00 \times 10^{-10}$). This result is compatible with that of Wildman et al. (2003). Figure 3 shows a plot of human *dN* − ape *dN* for each gene. We multiplied human *dN* by 6.54, because of the difference of human divergence time (5.4 MYA) and ape divergence times (35.3 MYA). If *dN* is constant for human and ape branches, human *dN* − ape *dN* is expected to be zero. In fact, majority of the genes are located around the zero line in figure 3. Substitution rates (*dS*) varied among genes. For example, PRM2 showed the highest substitution rate by *dS*. However, there was no correlation between the difference of human *dN* − ape *dN* and total *dS*.

Two genes (ACAT2 and PRM2) showed higher rates of *dN* for human branch than ape branches. In ACAT2, no synonymous substitution and two nonsynonymous substitutions were observed on the human branch, and one synonymous and two nonsynonymous substitutions were observed on the ape branch. However, because the compared number of nucleotides (147 bp) was small, a further analysis is necessary to determine whether this difference is significant. The RNASE3 gene was shown to have higher rate of *dN* for ape branches than human branch (figure 3). Zhang, Rosenberg, and Nei (1998) analyzed RNASE3 (eosinophil cationic protein or ECP) and suggested the existence of positive selection on this gene. However, our result on figure 3 suggests that positive selection operates only on the ape branch. It is possible that the selective constraint became strong after the human lineage diverged from the remaining hominoid lineage for the RNASE3 gene.

In conclusion, we conducted a systematic analysis of 103 protein-coding genes for human, chimpanzee, gorilla, and orangutan. We showed that gene genealogies differ from gene to gene, because the time span between the human-chimpanzee common ancestor and gorilla speciation is short. We conducted three types of analyses for detecting the human-specific pattern in nonsynonymous

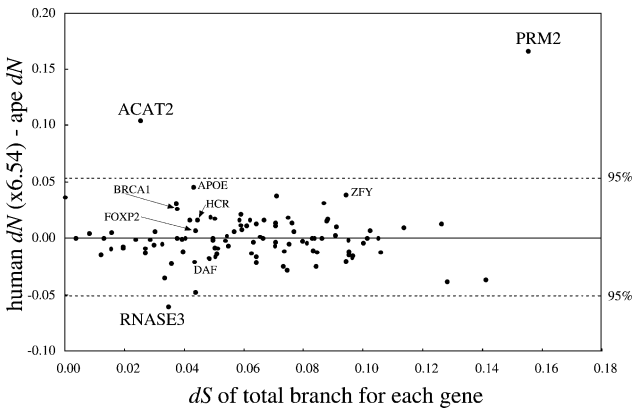

FIG. 3.—A plot of human *dN* − ape *dN* and total *dS* for each gene. Human *dN* was multiplied by 6.54 because of the difference of human divergence time (5.4 MYA) and ape divergence times (35.3 MYA). The total *dS* was estimated from all branches of each gene. Broken lines denote 95% level of the variation of human *dN* − ape *dN*.

changes. Comparison of each coding region is a standard way of detecting the pattern of natural selection. However, it is sometimes difficult to detect the pattern of natural selection because of a few numbers of changes. We conducted comparison of *dNs* by using a large number of genes. This kind of analysis may help to find candidate genes that caused human-specific phenotypic changes.

Acknowledgments

We are grateful for Robert E. Ferrell for providing chimpanzee DNA sample. We thank Hidemi Kobayakawa for technical assistance and Kenta Sumiyama for suggestions and discussion. This study was supported by a grant in aid for scientific studies from Ministry of Education, Science, Sport, and Culture, Japan to N.S. and U.S., Joint Research Project (Soken/K99-1) of the Graduate University for Advanced Studies, Japan to N.S. T.K. was supported by COE research fellowship of the National Institute of Genetics and JSPS postdoctoral fellowship.

Literature Cited

Asumalahti, K., T. Laitinen, R. Itkonen-Vatjus, M.-L. Lokki, S. Suomela, E. Snellman, U. Saarialho-Kere, and J. Kere. 2000. A candidate gene for psoriasis near HLA-C, HCR (Pg8), is highly polymorphic with a disease-associated susceptibility allele. Hum. Mol. Genet. **9**:1533–1542.

Chen, F.-C., and W.-H. Li. 2001. Genomic divergence between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. Am. J. Hum. Genet. **68**:444–456.

Clark, A. G., S. Glanowski, R. Nielsen et al. (17 co-authors). 2003. Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. Science **302**:1960–1963.

Corder, E. H., A. M. Saunders, W. J. Strittmatter, D. E. Schmechel, P. C. Gaskell, G. W. Small, A. D. Roses, J. L. Haines, and M. A. Pericak-Vance. 1993. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science **261**:921–923.

Enard, W., M. Przeworski, S. E. Fisher, C. S. Lai, V. Wiebe, T. Kitano, A. P. Monaco, and S. Pääbo. 2002. Molecular evolution of FOXP2, a gene involved in speech and language. Nature **418**:869–872.

Fujiyama, A., H. Watanabe, A. Toyoda et al. (14 co-authors). 2002. Construction and analysis of a human-chimpanzee comparative clone map. Science **295**:131–134.

Goodman, M., C. A. Porter, J. Czelusniak, S. L. Page, H. Schneider, J. Shoshani, G. Gunnell, and C. P. Groves. 1998. Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. Mol. Phylogenet. Evol. **19**:585–598.

Horai, S., K. Hayasaka, R. Kondo, K. Tsugane, and N. Takahata. 1995. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc. Natl. Acad. Sci. USA **92**:532–536.

Hughes, A. L., and M. Nei. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature **335**:167–170.

———. 1989. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc. Natl. Acad. Sci. USA **86**:958–962.

Huttley, G. A., S. Easteal, M. C. Southey, A. Tesoriero, G. G. Giles, M. R. E. McCredie, J. L. Hopper, and D. J. Venter. 2000. Adaptive evolution of the tumour suppressor BRCA1 in humans and chimpanzees. Nat. Genet. **25**:410–413.

Ina, Y. 1994. ODEN: a program package for molecular evolutionary analysis and database search of DNA and amino acid sequences. Comput. Appl. Biosci. **10**:11–12.

International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. Nature **409**:860–921.

Johnson, M. E., L. Viggiano, J. A. Bailey, M. Abdul-Rauf, G. Goodwin, M. Rocchi, and E. E. Eichler. 2001. Positive selection of a gene family during the emergence of humans and African apes. Nature **413**:514–519.

Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge, UK.

King, M. C., and A. C. Wilson. 1975. Evolution at two levels in humans and chimpanzees. Science **188**:107–116.

Kitano, T., and N. Saitou. 1999. Evolution of the Rh blood group genes has experienced gene conversions and positive selection. J. Mol. Evol. **49**:615–626.

Kitano, T., K. Sumiyama, T. Shiroishi, and N. Saitou. 1998. Conserved evolution of the Rh50 gene compared to its homologous Rh blood group gene. Biochem. Biophys. Res. Comm. **249**:78–85.

Kuttner-Kondo, L., V. B. Subramanian, J. P. Atkinson, J. Yu, and M. E. Medof. 2000. Conservation in decay accelerating factor (DAF) structure among primates. Dev. Comp. Immunol. **24**:815–827.

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

Messier, W., and C. B. Stewart. 1997. Episodic adaptive evolution of primate lysozymes. Nature **385**:151–154.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.

Ohno, S. 1972. So much "junk" DNA in our genome. Brookhaven Symp. Biol. **23**:366–370.

O'hUigin, C., Y. Satta, N. Takahata, and J. Klein. 2002. Contribution of homoplasy and of ancestral polymorphism to the evolution of genes in anthropoid primates. Mol. Biol. Evol. **19**:1501–1513.

Rall, S. C. Jr., K. H. Weisgraber, T. L. Innerarity, and R. W. Mahley. 1982. Structural basis for receptor binding heterogeneity of apolipoprotein E from type III hyperlipoproteinemic subjects. Proc. Natl. Acad. Sci. USA **79**:4696–4700.

Reid, M. E., V. Chandrasekaran, L. Sausais, J. Pierre, and R. Bullock. 1996. Disappearance of antibodies to Cromer blood group system antigens during mid pregnancy. Vox Sang. **71**: 48–50.

Saitou, N. 1991. Reconstruction of molecular phylogeny of extant hominoids from DNA sequence data. Am. J. Phys. Anthropol. **84**:75–85.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

Satta, Y., J. Klein, and N. Takahata. 2000. DNA archives and our nearest relative: the trichotomy problem revisited. Mol. Phyl. Evol. **14**:259–275.

Sibley, C. G., and J. E. Ahlquist. 1984. The phylogeny of the hominoid primates, as indicated by DNA-DNA hybridization. J. Mol. Evol. **20**:2–15.

Sneath, P. H. A., and R. R. Sokal. 1973. Numerical taxonomy. Freeman, San Francisco.

Sumiyama, K., N. Saitou, and S. Ueda. 2002. Adaptive evolution of the IgA hinge region in primates. Mol. Biol. Evol. **19**: 1093–1099.

Thompson, J. D., T. J. Gibson, and D. G. Higgins. 1994. CLUSTAL W: improving the sensitivity of progressive

multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.

Tucker, P. K., R. M. Adkins, and J. S. Rest. 2003. Differential rates of evolution for the ZFY-related zinc finger genes, Zfy, Zfx, and Zfa in the mouse genus *Mus*. Mol. Biol. Evol. **20**:999–1005.

Weisgraber, K. H., S. C. Rall, and R. W. Mahley. 1981. Human E apoprotein heterogeneity: cysteine-arginine interchanges in the amino acid sequence of the apo-E isoforms. J. Biol. Chem. **256**:9077–9083.

Wildman, D. E., M. Uddin, G. Liu, L. I. Grossman, and M. Goodman. 2003. Implications of natural selection in shaping 99.4% nonsynonymous DNA identity between humans and chimpanzees: enlarging genus *Homo*. Proc. Natl. Acad. Sci. USA **100**:7181–7188.

Wyckoff, G. J., W. Wang, and C.-I. Wu. 2000. Rapid evolution of male reproductive genes in the descent of man. Nature **403**:304–309.

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. CABIOS **13**:555–556.

Zhang, J., H. F. Rosenberg, and M. Nei. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. USA **95**:3708–3713.

Zhang, J., D. M. Webb, and O. Podlaha. 2002. Accelerated protein evolution and origins of human-specific features: Foxp2 as an example. Genetics **162**:1825–1835.

Pekka Pamilo, Associate Editor

Accepted January 8, 2004