THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 267, No. 11, Issue of April 15, pp. 7359-7367, 1992
Printed in U.S.A.

Concerted Evolution of the Primate Immunoglobulin α-Gene through Gene Conversion*

(Received for publication, September 17, 1991)

Shohji Kawamura, Naruya Saitou‡, and Shintaroh Ueda§

From the Department of Anthropology, Faculty of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan

We determined four nucleotide sequences of the hominoid immunoglobulin α (Cα) genes (chimpanzee Cα2, gorilla Cα2, and gibbon Cα1 and Cα2 genes), which made possible the examination of gene conversions in all hominoid Cα genes. The following three methods were used to detect gene conversions: 1) phenetic tree construction; 2) detection of a DNA segment with extremely low variability between duplicated Cα genes; and 3) a site by site search of shared nucleotide changes between duplicated Cα genes. Results obtained from method 1 indicated a concerted evolution of the duplicated Cα genes in the human, chimpanzee, gorilla, and gibbon lineages, while results obtained from method 2 suggested gene conversions in the human, gorilla, and gibbon Cα genes. With method 3 we identified clusters of shared nucleotide changes between duplicated Cα genes in human, chimpanzee, gorilla, and gibbon lineages, and in their hypothetical ancestors. In the present study converted regions were identified over the entire Cα gene region excluding a few sites in the coding region which have escaped from gene conversion. This indicates that gene conversion is a general phenomenon in evolution, that can be clearly observed in nonfunctional regions.

The mammalian immunoglobulin heavy chain constant region (C_H)[1] gene cluster has evolved through multiple gene duplications and resulting specializations into its present classes (Cμ, Cδ, Cγ, Cε, and Cα). The C_H gene organization of human is 5′-Cμ-Cδ-Cγ3-Cγ1-Cε2-Cα1-ΨCγ-Cγ2-Cγ4-Cε1-Cα2-3′, while that of mouse is 5′-Cμ-Cδ-Cγ3-Cγ1-Cγ2b-Cγ2a-Cε-Cα-3′(Ref. 1, and references therein). Comparing the nucleotide sequences of the human immunoglobulin Cα1 gene and two alleles [A2m(1) and A2m(2)] of the Cα2 genes, Flanagan *et al.* (2) found that the nucleotide sequence of a contiguous stretch in the 3′ end of the Cα1 gene is identical with that of the A2m(1) allele of the Cα2 gene, but is different from that of the A2m(2) allele. Thus, they argued that there was a localized transfer of the genetic information from the 3′ end of the Cα1 gene to the A2m(1) allele of the Cα2 gene through gene conversion.

Gene conversion is broadly defined as a nonreciprocal superposition of the information of one piece of DNA onto another (3). Although its molecular mechanism is not well understood, segmental homology between related genes is often attributed to gene conversion. It can occur between alleles on homologous chromosomes or between related genes on the same, sister, or nonhomologous chromosomes of an organism at mitosis or meiosis (4). Gene conversion has been suggested as a process of concerted evolution which maintains or creates sequence homogeneity in the related genes (5). The existence of concerted evolution can be shown when duplicated genes share the same nucleotide changes within a species but different changes between species, and gene conversion is suggested when the nucleotide positions with such changes are contiguous (6).

We previously studied the organization of the C_H gene cluster in non-human primates (7–10). Chimpanzee and gorilla were found to have the Cε2-Cα1 and Cε1-Cα2 regions as does human, although all exons and introns of the chimpanzee Cε2 gene are deleted (7). Orangutan, gibbon, and Old World monkey have one, two, and one Cα genes, respectively (8), and have one Cε gene in their C_H gene clusters (9). Our recent study shows that duplication, including the Cγ-Cγ-Cε-Cα genes, occurred in the common ancestor of hominoids (human, chimpanzee, gorilla, orangutan, and gibbon), followed by the deletion of the gibbon Cε gene upstream from the Cα1 gene and also of one of the Cε-Cα regions in orangutan. Support for this gene phylogeny can also be found when the evolution of the Cα hinge region is considered. Human, chimpanzee, and gorilla share different hinge structures in their Cα1 and Cα2 genes (four tandem repeats of a 15-bp unit with a 6-bp overlap in the Cα1 genes and only one 15-bp unit in the Cα2 genes) (7, 10). The hinge region of the orangutan Cα gene is a Cα1 type of human, chimpanzee, and gorilla, while that of the gibbon Cα1 gene consists of two tandem repeats of a 15-bp unit with a 6-bp deletion. The hinge region of the gibbon Cα2 gene is a Cα2 type of human, chimpanzee, and gorilla (8). The most parsimonious explanation for the evolutionary history of the hominoid Cα hinge region is a gene phylogeny in which duplication of the Cα gene occurred in the common ancestor of hominoids and the deletion of the Cα2 gene occurred in the orangutan lineage (Fig. 1).

On the basis of the Cα gene phylogeny described above, we examined in the present study whether gene conversion has occurred in the duplicated Cα genes of primates. We sequenced chimpanzee and gorilla Cα2 genes as well as Cα1 and Cα2 genes of gibbon. These new sequence data made it possible to compare all the nucleotide sequences of hominoid and Old World monkey Cα genes. Our phylogenetic analysis

* This work was supported by the Ministry of Education, Science and Culture of Japan and by a cooperative research grant from the Primate Research Institute, Koyoto University. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Nucleotide sequence data reported in this paper have been deposited in the DDBJ/EMBL/GenBank International Nucleotide Sequence data bases under accession numbers X53706 (chimpanzee Cα2), X53707 (gorilla Cα2), X53708 (gibbon Cα1), and X53709 (gibbon Cα2).

‡ Present address: National Institute of Genetics, Mishima 4111, Japan.

§ To whom correspondence should be addressed.

[1] The abbreviations used are: C_H, immunoglobulin heavy chain constant region; bp, base pair(s); CH, domain of immunoglobulin heavy chain constant region.

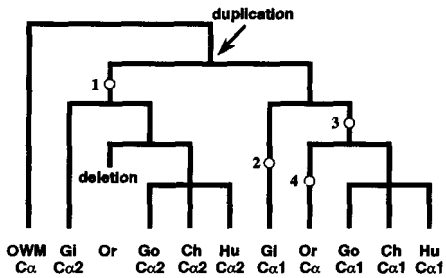

FIG. 1. **Cα gene phylogeny in higher primates.** A gene duplication in the common ancestor of hominoids and a deletion of the Cα2 gene in the orangutan lineage are indicated. For convenience sake we used a tricotomous topology for the relationship among human, chimpanzee, and gorilla. *Hu*, *Ch*, *Go*, *Or*, *Gi*, and *OWM* indicate human, chimpanzee, gorilla, orangutan, gibbon, and Old World monkey, respectively. This tree topology is most parsimonious for the evolution of the Cα hinge region in hominoids. Since the hinge region of the ancestral Cα gene of hominoids is thought to consist of two tandem repeats of a 15-bp unit (8), four duplication or deletion events are required in its evolution: 1) a deletion of the first 15-bp unit of the Cα2 gene; 2) a 6-bp deletion from the first unit of the gibbon Cα1 gene; 3) a duplication of 2 units resulting in 4 units with 6-bp overlap; and 4) a 6-bp deletion from the first unit of the orangutan Cα gene.

showed that gene conversions occurred on the duplicated Cα genes of the chimpanzee, gorilla, gibbon, and their ancestors. We also found that human Cα gene conversion occurred not only between the Cα1 and the Cα2[A2m(1)] genes but also between the Cα1 gene and the ancestral Cα2 gene of the two alleles.

MATERIALS AND METHODS

DNA Manipulations—Recombinant phage clones containing chimpanzee (*Pan troglodytes*) Cα2, gorilla (*Gorilla gorilla*) Cα2, and gibbon (*Hylobates lar*) Cα1 and Cα2 genes were isolated previously from their genome libraries (7, 8). A restriction fragment from each phage clone containing the Cα gene was subcloned into pUC119 or Bluescript plasmid vectors in both directions. Using exonuclease III and mung bean nuclease (Toyobo Co., Ltd.), varying degrees of unidirectional deletions were introduced into the insert of each plasmid. Single-stranded DNAs were rescued using a helper phage and their nucleotide sequences were determined by the dideoxynucleotide chain termination method. All sequencing runs were performed on both sense and antisense strands. These DNA manipulations were performed using standard procedures (11).

DNA Sequence Analyses—Sequence alignment was done by visual examination for human Cα1, Cα2[allele A2m(1)], and Cα2[allele A2m(2)], chimpanzee Cα1 and Cα2, gorilla Cα1 and Cα2, orangutan (*Pongo pygmaeus*) Cα, gibbon Cα1 and Cα2, and crab-eating macaque (*Macaca fascicularis*; an Old World monkey) Cα genes (Fig. 2). The

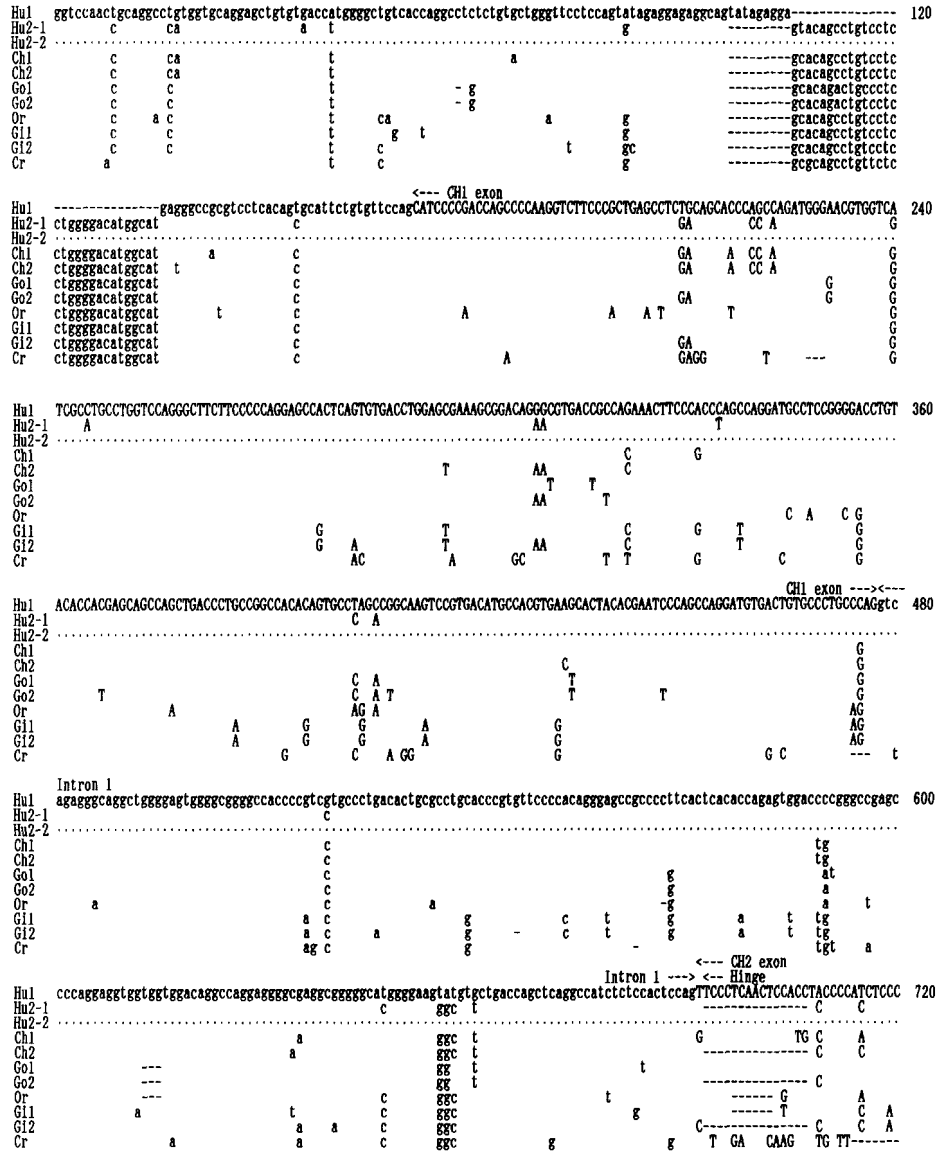

FIG. 2. **Aligned Cα gene sequences.** Sequences of the human Cα1, Cα2[allele A2m(1)], and Cα2[allele A2m(2)], the chimpanzee Cα1 and Cα2, the gorilla Cα1 and Cα2, the orangutan Cα, the gibbon Cα1 and Cα2, and the crab-eating macaque Cα genes (designated by *Hu1*, *Hu2-1*, *Hu2-2*, *Ch1*, *Ch2*, *Go1*, *Go2*, *Or*, *Gi1*, *Gi2*, and *Cr*, respectively) are shown. For clarity, the only complete sequence shown is the human Cα1 gene and only nucleotides different from it are shown in other Cα genes. Gaps are designated by *hyphens*. *Dots* represent positions for which no data are available. Noncoding and coding sequences are denoted by *lower case* and *capital letters*, respectively. Landmarks for CH exons, introns, and the hinge region are indicated *above* the human Cα1 sequence, and those for termination codon, putative polyadenylation signal (aataaa), and alternating purine and pyrimidine sequence are represented by *Ter*, *poly-A*, and *Pu/Py*, respectively. Human Cα1 and Cα2 gene sequences are from Flanagan *et al.* (2), and the nucleotide sequences of the chimpanzee Cα1, the gorilla Cα1, the orangutan Cα, and the crab-eating macaque Cα genes are from Kawamura *et al.* (28).

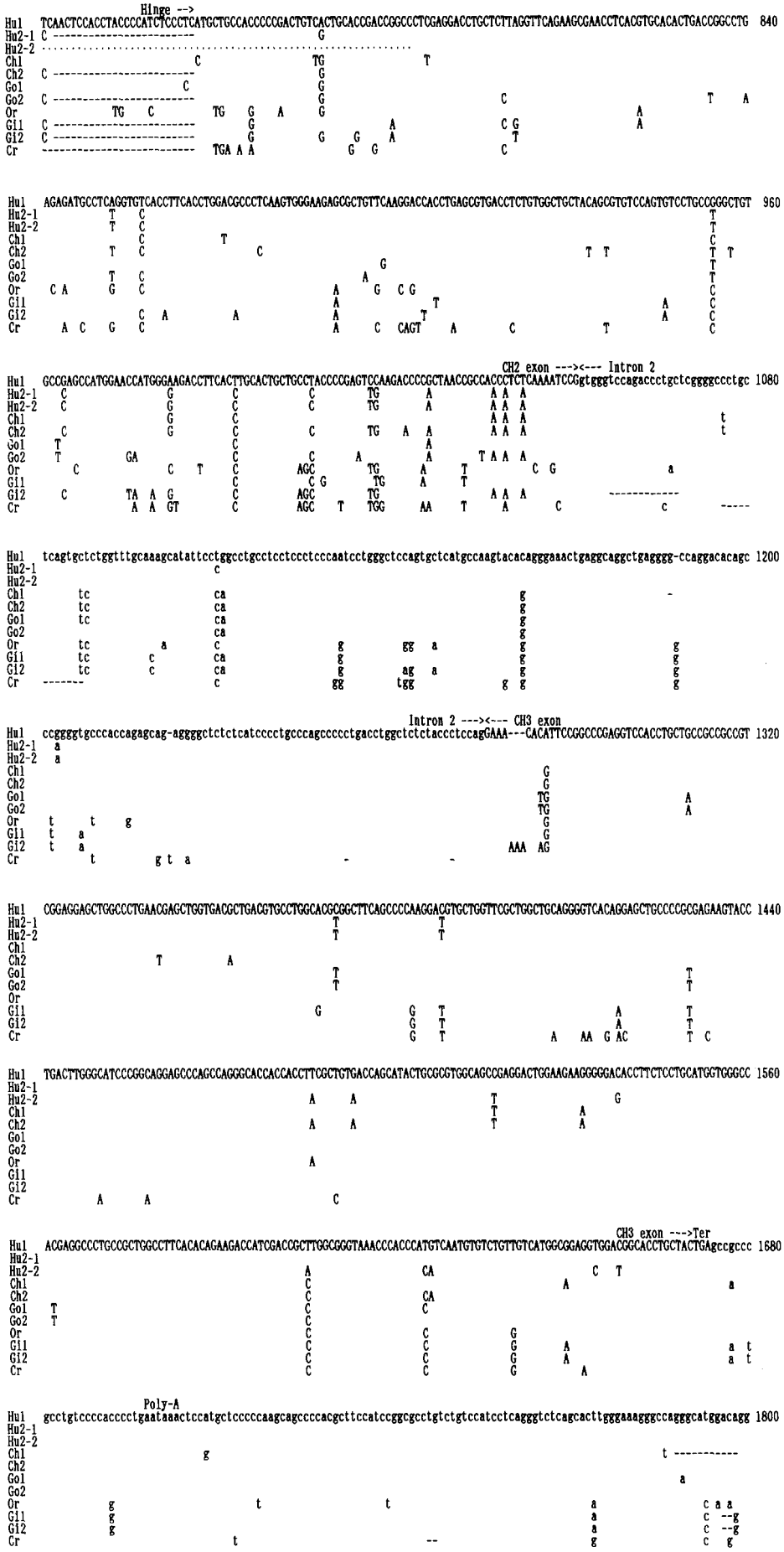

FIG. 2.—*continued*

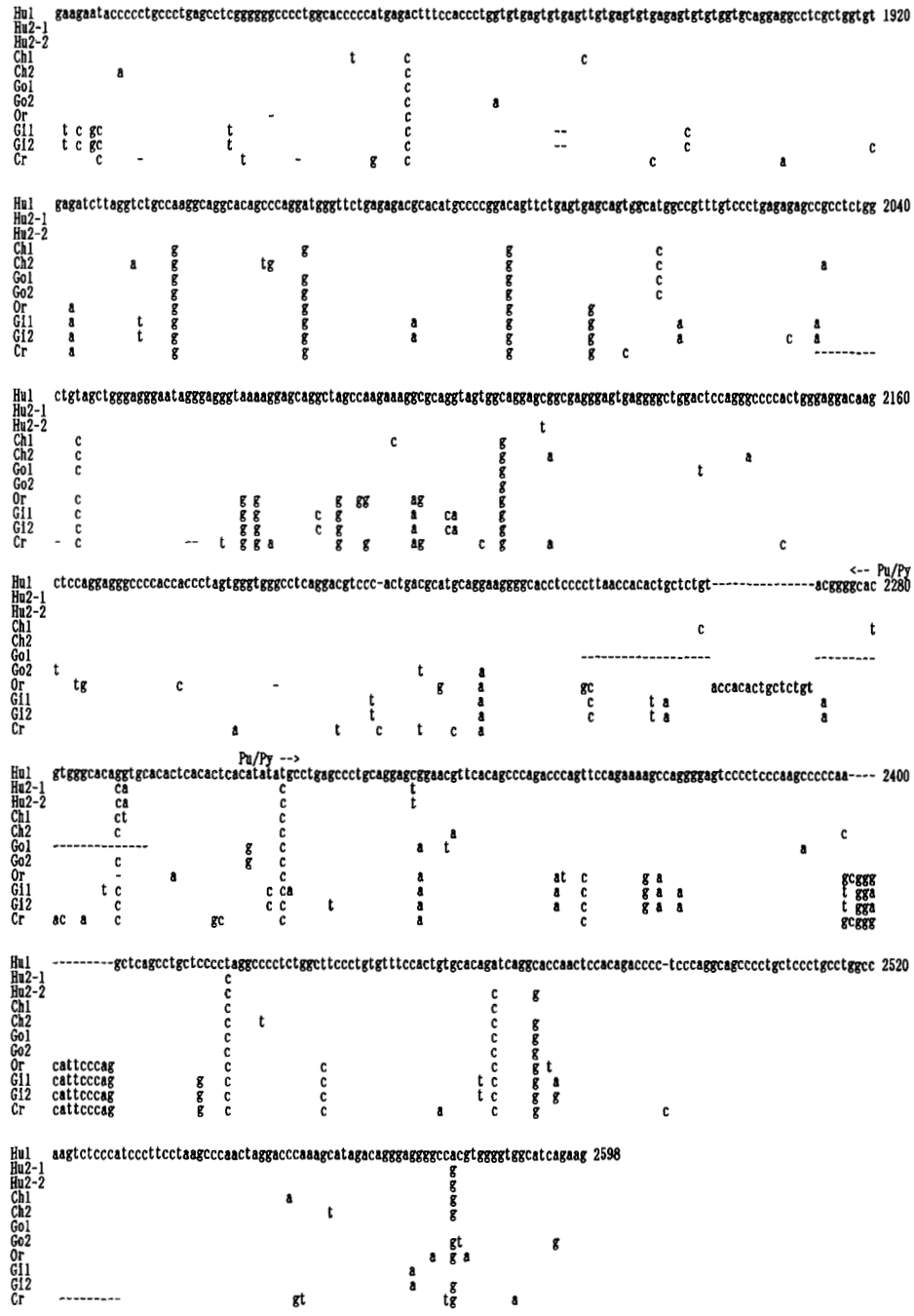

FIG. 2.—*continued*

estimated number of nucleotide substitutions per site (K^c) between two DNA sequences was computed using the formula: $K^c = -(3/4)\ln[1-(4/3)K]$, where K is the number of nucleotide differences per site (12). The variance of the K^c value is given by $V(K^c) = K(1-K)/n(1-4K/3)^2$, where n is the number of nucleotides compared (13). The standard error (S.E.) of the K^c value is given by the square root of $V(K^c)$. The human Cα2[allele A2m(2)] was excluded from this computation because its nucleotide sequence in the upstream portion from position 783 of Fig. 2 has not yet been published. In computation, gap sites were excluded and the remaining 2354 bp were used.

A relative rate test (14) was performed by comparing the evolutionary distances (K^c) from the crab-eating macaque Cα gene to hominoid Cα genes applying the statistical method of Wu and Li (15). When K^c between hominoid gene i and crab-eating macaque gene (m) is referred to as K^c_{im}, a variance of $K^c_{1m} - K^c_{2m}$ is given by $V(K^c_{1m}) + V(K^c_{2m}) - 2V(K^c_{0m})$, where $K^c_{0m} = (K^c_{1m} + K^c_{2m} - K^c_{12})/2$, $V(x)$ denotes the variance of x (15).

A phenetic gene tree was constructed from the K^c distance matrix using the unweighted pair group method with arithmetic mean (16) and the S.E. value of each branching point of this tree was calculated using the method described by Nei *et al.* (17). Using these S.E. values we tested the statistical significance of the tree topology (see Ref. 17 for a detailed explanation of the method).

Regions of extremely low variability between Cα1 and Cα2 genes of a species were identified using Stephens' (18) method under an assumption of random distribution of varied and unvaried sites. The probability (p) of observing at least one unvaried segment as long as or longer than an unvaried segment observed, is given by $p = 1 - [1 - \sum_{k=g_o}^{r} \{_{(s+r-k-3)}C_{(r-k)/(s+r-2)}C_r\}]^{s-1}$, where s is the number of varied sites, r is the total number of unvaried sites placed between varied sites, and g_o is the length of consecutive unvaried sites (18). An insertion/deletion site was regarded as one varied site regardless of its length.

A site by site reconstruction of gene conversion was performed using Fitch *et al.*'s method (6). At each aligned position (Fig. 2), a parsimony reconstruction was made of the nucleotide changes which occurred, assuming the Cα gene phylogeny shown in Fig. 1. Each

insertion/deletion was regarded as 1 nucleotide change regardless of its length. Each variant sequence was classified into four categories by every varied site: (*a*) variants shared between the Cα1 and Cα2 genes of a single species (shown as *open circles* in Figs. 5*a* and 6); (*b*) variants shared among orthologous Cα1 or Cα2 genes of different species, but not between paralogous Cα1 and Cα2 within a species, or those shared between two alleles of the human Cα2 gene but not with the human Cα1 gene (shown as *open squares* in Figs. 5*b* and 6); (*c*) variants in which paralogous genes of the same species had the same nucleotide (shown as *triangles* when the Cα1 gene changed, or *inverted triangles* when the Cα2 gene changed (Figs. 5*c* and 6); and (*d*) all other variants (shown as *full circles* in Figs. 5*d* and 6). (Orthologous genes are related genes which are derived by speciation from a single common ancestral gene, whereas paralogous genes are those which arose from gene duplication events within a species.) In cases where there were multiple equally parsimonious solutions, we chose category *d* because this provides a conservative answer concerning the occurrence of gene conversion (*e.g.* position 47 of Fig. 2). When either category *a* or *c* was possible, category *a* was chosen as a conservative determination of gene conversion direction (*e.g.* position 18 of Fig. 2). When two equally parsimonious solutions fell into the same category, both solutions were indicated by a pair of *bars* (left and right, up and down, left-up and right-down, or left-down and right-up) (see Fig. 6). If two or more *a* or *c* sites were contiguous, a gene conversion event over the region containing those sites was inferred. On the other hand, if *b* sites were contiguous, a gene conversion over the region was not inferred (6).

It should be noted that the boundaries of the converted regions were defined as the outermost *a* or *c* sites lacking two or more consecutive *b* sites. In the original method (6, 19–21), boundaries were defined as *b* sites which flank a cluster of *a* or *c* sites, but this definition gives an upperlimit for a converted region and there is a risk of overestimating the span of conversions.

RESULTS AND DISCUSSION

We determined the nucleotide sequences of chimpanzee Cα2, gorilla Cα2, and gibbon Cα1 and Cα2 genes, and compared them with human Cα1 and Cα2[A2m(1) and A2m(2) alleles], chimpanzee Cα1, gorilla Cα1, orangutan Cα, and crab-eating macaque Cα genes (Fig. 2). All sequences have the same exon-intron organization and have neither a termination codon nor frame shift mutation within their open reading frames. The coupled deletion-duplication in the 5′-flanking region of the human Cα1 gene (positions 78–139 of Fig. 2) reported by Flanagan *et al.* (2) was not found in any non-human Cα gene. Therefore, this probably occurred only in the human lineage.

Phenetic Tree Construction—Table I shows the K^c distance matrix for primate Cα genes. Other estimation methods (22, 23) for superimposed nucleotide substitution gave essentially the same K^c values as did the Jukes-Cantor method used in the present study (data not shown). A relative rate test (14) was performed for the hominoid genes with the crab-eating macaque gene as an outgroup reference using Wu and Li's (15) statistical method (Table II). No significant variation was observed for evolutionary distances from the crab-eating macaque gene among hominoid genes (see Table II), indicating the existence of an almost constant evolutionary rate for hominoid Cα genes. Therefore, we used the unweighted pair group method with arithmetic mean (16) to construct a phenetic gene tree (see Fig. 3) because this method produces a good evolutionary tree when the expected rate of nucleotide substitution is constant and because clusterings can be easily tested using S.E. values for each branching point (17).

In the phenetic tree, the two Cα genes of human, chimpanzee, gorilla, and gibbon formed statistically significant clusters in every species (clusters A, C, B, and D of Fig. 3, respectively) at the 5, 5, 0.5, and 0.01% levels, respectively, when the one-tailed *t* test was used with infinite degrees of freedom. A clustering of the two paralogous Cα genes in every species was also observed in gene trees constructed using the neighbor-joining method (24) and the maximum parsimony method (25), both of which do not assume rate constancy of nucleotide substitution (data not shown). Because of the Cα gene duplication in the common ancestor of hominoids and the deletion of the orangutan Cα2 gene (see Introduction and Fig. 1), these results indicate that a concerted evolution of the Cα genes occurred in human, chimpanzee, gorilla, and gibbon lineages.

Statistical Test for Detecting a Cluster of Unvaried Sites—To examine whether gene conversion is the underlying mechanism for the concerted evolution of hominoid Cα genes and to elucidate where it occurred in the gene region, we searched for consecutive unvaried regions between paired human, chimpanzee, gorilla, and gibbon Cα sequences. Observation of such regions is expected when gene conversion has occurred between two gene loci. Fig. 4 shows the distribution of the varied sites between the Cα1 and Cα2 genes. Varied sites include sites of nucleotide substitution and insertion/deletion. An insertion/deletion site was regarded as one varied site regardless of its length.

TABLE I

Estimated numbers of nucleotide substitutions per 100 sites of primate immunoglobulin Cα genes

A total of 2354 nucleotides were used for comparison. Values are $K^c \times 100$. Standard errors are given in parentheses.

	Human Cα1	Human Cα2[A2m(1)]	Chimpanzee Cα1	Chimpanzee Cα2	Gorilla Cα1	Gorilla Cα2	Orangutan Cα	Gibbon Cα1	Gibbon Cα2
Human Cα2[A2m(1)]	2.07 (0.30)								
Chimpanzee Cα1	2.81 (0.35)	2.42 (0.32)							
Chimpanzee Cα2	3.52 (0.39)	2.37 (0.32)	2.29 (0.32)						
Gorilla Cα1	2.33 (0.32)	2.77 (0.35)	2.73 (0.35)	3.21 (0.38)					
Gorilla Cα2	3.12 (0.37)	2.51 (0.33)	2.99 (0.36)	3.21 (0.38)	1.81 (0.28)				
Orangutan Cα	5.28 (0.49)	5.28 (0.49)	5.55 (0.50)	5.78 (0.51)	4.92 (0.47)	5.69 (0.51)			
Gibbon Cα1	5.19 (0.48)	5.74 (0.51)	5.23 (0.48)	5.87 (0.52)	4.92 (0.47)	5.83 (0.51)	5.42 (0.49)		
Gibbon Cα2	6.24 (0.53)	5.46 (0.50)	5.33 (0.49)	5.65 (0.50)	5.83 (0.51)	5.78 (0.51)	5.78 (0.51)	2.42 (0.32)	
Crab-eating macaque Cα	7.45 (0.58)	7.22 (0.57)	7.50 (0.59)	7.74 (0.60)	7.45 (0.58)	7.50 (0.59)	7.03 (0.57)	7.50 (0.59)	7.50 (0.59)

TABLE II

Evolutionary distance differences between crab-eating macaque Cα gene and hominoid Cα genes

Gene 1/gene 2	K^c_{1m}-K^c_{2m}[a]	S.E.	p[b]
Human Cα1/human Cα2[A2m(1)]	0.24	0.32	>0.4
Human Cα1/chimpanzee Cα1	−0.05	0.37	>0.8
Human Cα1/chimpanzee Cα2	−0.28	0.41	>0.4
Human Cα1/gorilla Cα1	0	0.34	1
Human Cα1/gorilla Cα2	−0.05	0.39	>0.9
Human Cα1/orangutan Cα	0.42	0.50	>0.4
Human Cα1/gibbon Cα1	−0.05	0.50	>0.9
Human Cα1/gibbon Cα2	−0.05	0.55	>0.9
Human Cα2[A2m(1)]/chimpanzee Cα1	−0.28	0.34	>0.4
Human Cα2[A2m(1)]/chimpanzee Cα2	−0.52	0.34	>0.1
Human Cα2[A2m(1)]/gorilla Cα1	−0.24	0.37	>0.5
Human Cα2[A2m(1)/gorilla Cα2	−0.28	0.35	>0.4
Human Cα2[A2m(1)]/orangutan Cα	0.19	0.50	>0.7
Human Cα2[A2m(1)]/gibbon Cα1	−0.28	0.52	>0.5
Human Cα2[A2m(1)]/gibbon Cα2	−0.28	0.51	>0.5
Chimpanzee Cα1/chimpanzee Cα2	−0.24	0.33	>0.4
Chimpanzee Cα1/gorilla Cα1	0.05	0.36	>0.8
Chimpanzee Cα1/gorilla Cα2	0	0.38	1
Chimpanzee Cα1/orangutan Cα	0.47	0.51	>0.3
Chimpanzee Cα1/gibbon Cα1	0	0.50	1
Chimpanzee Cα1/gibbon Cα2	0	0.51	1
Chimpanzee Cα2/gorilla Cα1	0.28	0.39	>0.4
Chimpanzee Cα2/gorilla Cα2	0.24	0.39	>0.5
Chimpanzee Cα2/orangutan Cα	0.70	0.53	>0.1
Chimpanzee Cα2/gibbon Cα1	0.24	0.53	>0.6
Chimpanzee Cα2/gibbon Cα2	0.24	0.52	>0.6
Gorilla Cα1/gorilla Cα2	−0.05	0.30	>0.8
Gorilla Cα1/orangutan Cα	0.42	0.48	>0.3
Gorilla Cα1/gibbon Cα1	−0.05	0.49	>0.9
Gorilla Cα1/gibbon Cα2	−0.05	0.53	>0.9
Gorilla Cα2/orangutan Cα	0.47	0.52	>0.3
Gorilla Cα2/gibbon Cα1	0	0.53	1
Gorilla Cα2/gibbon Cα2	0	0.53	1
Orangutan Cα/gibbon Cα1	−0.47	0.51	>0.3
Orangutan Cα/gibbon Cα2	−0.47	0.52	>0.3
Gibbon Cα1/gibbon Cα2	0	0.34	1

[a] K^c_{im} (i = 1 or 2) is the number of nucleotide substitutions per 100 sites between hominoid Cα gene i and crab-eating macaque Cα gene.
[b] Probability that there is no difference between K^c_{1m} and K^c_{2m}, which is evaluated using a two-tailed t test with degrees of freedom = ∞.

Under the assumption of random distribution of varied and unvaried sites, regions of extremely low variability can be identified using Stephens' (18) method. Therefore, we applied this method to our data. In computation we excluded the hinge region because this region is known to have evolved through frequent recombinational events (8). The longest unvaried segment was significantly long in human, gorilla, and gibbon Cα genes (see Fig. 4). Since the next longest segments in human, gorilla, and gibbon were not significantly long, we excluded the longest segment and recalculated the probability of the next longest segment. In this procedure, r is substituted by $\{r - \text{longest } g_o\}$ and s by $\{s - 1\}$ (18). The next longest segment was significantly long only in gibbon (Fig. 4). Thus we identified the regions which have extremely low variability under the assumption of random distribution of variable sites (regions 2, 4, 6, 8, and 9 in Fig. 4; see the figure legends for details).

Site by Site Reconstruction of Gene Conversions—Stephens' (18) method is effective in detecting gene conversions spanning long stretches of sequences, but we cannot rule out the possibility that low variability may be attributed to a higher selection pressure or lower mutation rate than in other regions. Therefore, in the next step, we applied the site by site reconstruction method which was used to identify γ^1-/γ^2-globin conversions and δ-/β-globin conversions during primate evolution (6, 19–21). This method, 1) is effective in detecting gene conversions for both short stretches of sequences and long ones with nucleotide changes after conversion, both of which Stephens' method fails to detect; 2) is capable of examining whether low variability between the two Cα genes of a species is due to lack of nucleotide changes or due to shared nucleotide changes between them; and 3) is capable of detecting gene conversions in the ancestral species.

Once a gene tree is determined, we can identify where a nucleotide change occurred in the tree and the nature of the sequence change on the basis of the maximum parsimony principle. Each insertion/deletion was regarded as 1 nucleotide change regardless of its length. Each variant site was classified into four categories, *a*, *b*, *c*, and *d* (see "Materials and Methods" and Fig. 5), assuming the gene tree of Fig. 1. By mapping these sites along the Cα gene region for the paralogous gene pair of each species including ancestors, we identified clusters of shared nucleotide changes between paralogous Cα genes of each species (Fig. 6).

In gorilla and gibbon, the regions of extremely low variability identified in Fig. 4 overlap the clusters of shared nucleotide changes categorized as *a* (Fig. 6, *V* and *VI*). This indicates that the low variability between the duplicated Cα genes is not because of scarce nucleotide changes in these regions, but because of shared nucleotide changes, indicating the gene conversions in these regions suggested by using Stephens' (18) method. Besides this large cluster of *a* sites in the downstream region from intron 2, small stretches of possible gene conversions were observed in the region upstream from the CH2 exon (Fig. 6, *V* and *VI*).

In human genes, however, the conversion in region 2 of Fig. 4, which was identified using allele A2m(1) as a Cα2 gene, is indicated only by a short cluster of two *a* sites (Fig. 6*I*). Furthermore, conversion in region 4 of Fig. 4, which was identified using allele A2m(2) as a Cα2 gene, is not indicated by any such cluster (Fig. 6*II*). Instead, we found a long stretch of *a* sites in the gene pair of human Cα1 and ancestral Cα2 gene of the two alleles, which overlaps regions 2 and 4 of Fig. 4 (Fig. 6*III*). Therefore, conversions in regions 2 and 4 of Fig. 4 are considered to have occurred not independently but

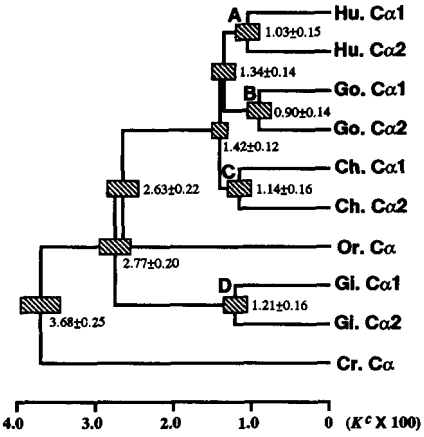

FIG. 3. **A unweighted pair group method with arithmetic mean phenetic tree for the Cα gene.** One S.E. for each branching point is shown as a *hatched box*. The cluster consisting of the Cα1 and Cα2 genes of the human, gorilla, chimpanzee, and gibbon are shown as *A*, *B*, *C*, and *D*, respectively. The scale represents the number of nucleotide substitutions per 100 sites (K^c value × 100). *Hu*, *Ch*, *Go*, *Or*, *Gi*, and *Cr* indicate human, chimpanzee, gorilla, orangutan, gibbon, and crab-eating macaque, respectively.

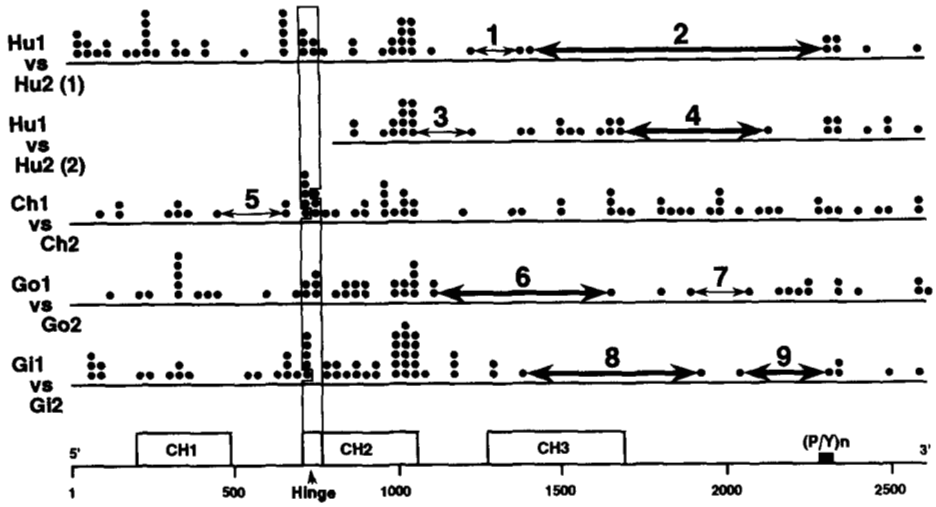

FIG. 4. **Distribution of the different sites between paralogous Cα genes of the human, chimpanzee, gorilla, and gibbon.** Different sites between the Cα1 and Cα2 genes of each species are indicated by *full circles*. Positions of CH exons, the hinge region, and alternating purine and pyrimidine sequence are illustrated at the *bottom* with the scale presenting nucleotide positions of aligned Cα sequences of Fig. 2. *Lines* which have *arrows* at both ends represent the longest or the second longest DNA segments of consecutive unvaried sites for each species. The lengths (g_o) of segments 1–9 are 163, 886, 180, 454, 201, 532, 179, 549, and 244 nucleotides, respectively. Their probabilities (p) are 0.83, 1.53×10^{-8}, 0.59, 1.38×10^{-3}, 0.34, 7.80×10^{-4}, 0.79, 2.30×10^{-5}, and 0.12, respectively. Probabilities for the second longest segments recalculated after excluding the longest segments, *i.e.* segments 1, 3, 7, and 9, are 0.29, 0.21, 0.45, and 2.15×10^{-2}, respectively. Significantly long segments are represented by *thicker lines*. Hinge regions (positions 693–746 of Fig. 2) are excluded from the calculation. Human Cα1, Cα2[allele A2m(1)], and Cα2[allele A2m(2)], chimpanzee Cα1 and Cα2, gorilla Cα1 and Cα2, and gibbon Cα1 and Cα2 genes are denoted by *Hu1*, *Hu2(1)*, *Hu2(2)*, *Ch1*, *Ch2*, *Go1*, *Go2*, *Gi1*, and *Gi2*, respectively. The numbers (s) of varied sites in Hu1-Hu2(1), Hu1-Hu2(2), Ch1-Ch2, Go1-Go2, and Gi1-Gi2 pairs are 51, 34, 59, 45, and 59, respectively. The numbers (r) of unvaried sites among varied sites are 2431, 1659, 2339, 2342, and 2378, respectively.

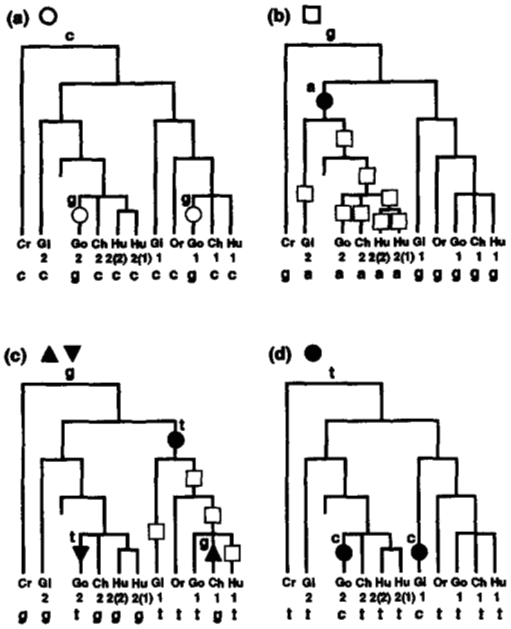

FIG. 5. **An explanatory figure for classification of variant sites.** To help visualize the way we classified the variant sites in order to find gene conversions, we use four hypothetical varied sites. In case *a*, the root nucleotide is *c*, thus variants are the gorilla Cα1 and Cα2 genes. These variants are caused by shared nucleotide changes (*c* to *g*) between the paralogous genes in a single species, suggesting a gene conversion in gorilla over this site, and they are presented by *open circles*. In case *b*, the root nucleotide is *g* and a change (*g* to *a*) occurs in the common ancestral Cα2 gene of hominoids (this change is not shared with its paralog and is indicated by a *full circle*). This varied nucleotide is retained during hominoid evolution among the Cα2 genes and no additional change occurred in the Cα1 genes of hominoids as well, suggesting no gene conversion before the divergence of the human Cα2 gene into the present two alleles. In the Cα1 and Cα2[allele A2m(1)] gene pair, three possible conversions with short stretches were found (Fig. 6*I*).

Although the Stephens' (18) method detected no significant stretch of unvaried sites in chimpanzee (Fig. 4), we identified three possible regions for gene conversion with the site by site method (Fig. 6*IV*). Distribution of these regions is similar to those observed in human, gorilla, and gibbon, *i.e.* the largest one in the region downstream from intron 2 and smaller ones in the region upstream from the CH2 exon. Therefore, lack of a significantly long stretch of unvaried sites in this downstream portion in chimpanzee is considered to be due to nucleotide changes after conversion. Possible converted segments were also observed in the hypothetical gene pair of the common ancestor of human, chimpanzee, and gorilla, and in that of orangutan and human, chimpanzee, and gorilla (see Fig. 6, *VII* and *VIII*), distributions of which are similar to

after the hominoid divergence over this site. Therefore, we give *open squares* to the Cα2 genes of human, chimpanzee, gorilla, gibbon, and hypothetical ancestors at this varied site. Case *c* is different from case *b* in that additional nucleotide changes occur in the chimpanzee Cα1 and the gorilla Cα2 genes after a nucleotide change occurred in the common ancestral Cα1 gene of hominoids. These additional changes result in the same nucleotide with their paralogs, suggesting a gene conversion of the chimpanzee Cα1 gene by the Cα2 gene and that of the gorilla Cα2 gene by its Cα1 gene over this site. Therefore, *triangles* are given to the chimpanzee Cα1 gene and *inverted triangles* are given to the gorilla Cα2 gene. In case *d*, the root is *t* and changes from *t* to *c* occur in the gibbon Cα1 and the gorilla Cα2 genes. These nucleotide changes are not shared either by paralogs or orthologs, and they are given *full circles*. *Hu*, *Ch*, *Go*, *Or*, *Gi*, and *Cr* indicate human, chimpanzee, gorilla, orangutan, gibbon, and crab-eating macaque, respectively. *1* and *2* indicate Cα1 and Cα2 genes, respectively. The A2m(1) and A2m(2) alleles of the human Cα2 gene are shown as *2(1)* and *2(2)*, respectively.

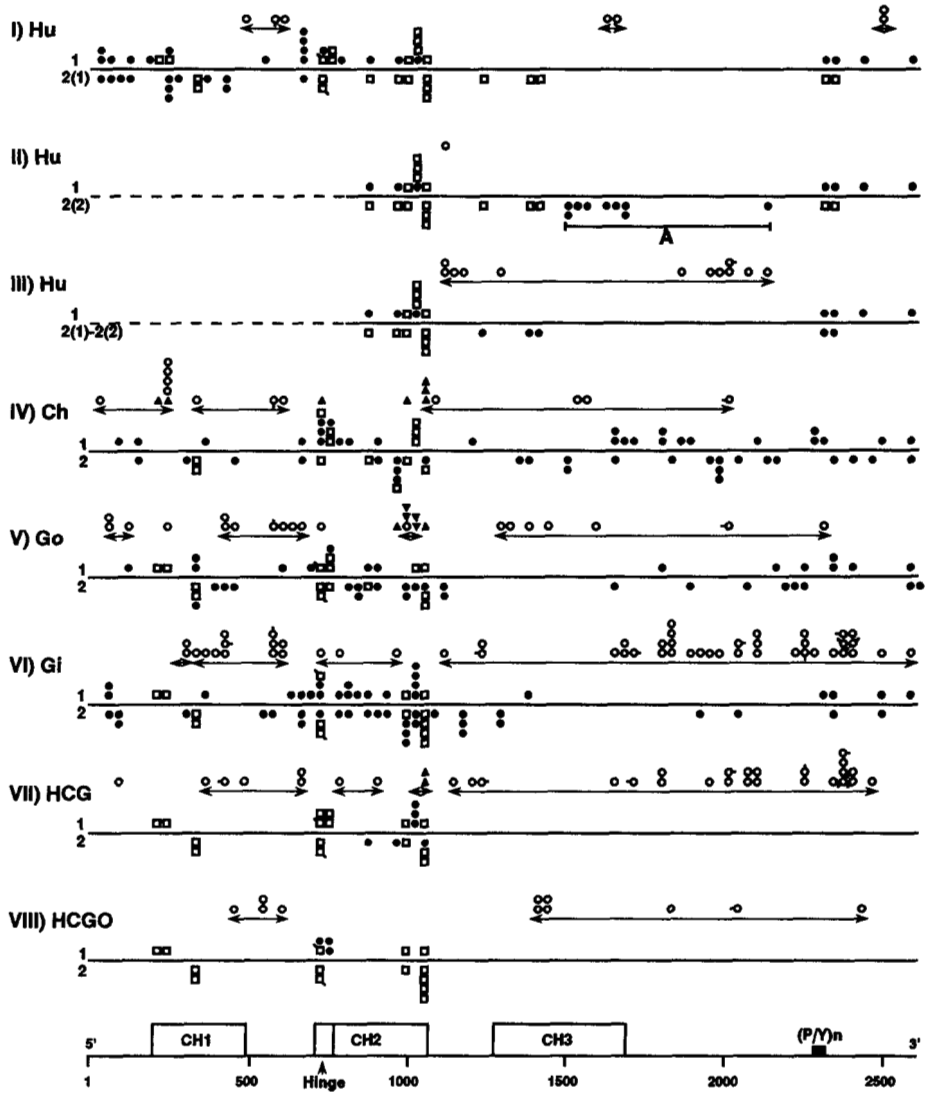

FIG. 6. **Site by site reconstruction of Cα gene conversions in hominoids.** Distribution maps of the categorized varied sites in: *I–III*, human (*Hu*); *IV*, chimpanzee (*Ch*); *V*, gorilla (*Go*); *VI*, gibbon (*Gi*); *VII*, the common ancestor (*HCG*) of the human, chimpanzee, and gorilla; and *VIII*, the common ancestor (*HCGO*) of the HCG and orangutan are shown. An organizational map is shown as in Fig. 4. A *horizontal line* represents the DNA region examined for each species and hypothetical ancestor. Varied sites classified into categories *b* (*open squares*) and *d* (*full circles*) in the Cα1 genes are mapped *above* the *horizontal lines* and those in the Cα2 genes are mapped *below* the *lines*. *Symbols* for varied sites classified into categories *a* (*open circles*) and *c* (*triangles* and *inverted triangles*) are shown *above* each row of symbols for categories *b* and *d* sites of the Cα1 sequences. *1* and *2* represent the Cα1 and Cα2 genes, respectively. The A2m(1) and A2m(2) alleles and their ancestral one of the human Cα2 gene are shown as *2(1)*, *2(2)*, and *2(1)-2(2)*, respectively. Suspected regions of gene conversion are indicated by *lines* with *arrows* at both ends (see "Materials and Methods"). A pair of *bars* (left and right, up and down, left-up and right-down, or left-down, and right-up) indicate alternative assignments of the variant sequence when two parsimonious solutions fall into the same category. *Dashed lines* represent positions for which no sequence data of the human Cα2[A2m(2)] gene is available.

those in extant species with the feature described above.

Both the largest segments of possible gene conversions suggested by the site by site method and the significantly long segments of unvaried sites detected by Stephens' (18) method were always found downstream from intron 2. In addition, no category *b* site, which does not support the occurrence of gene conversion, was observed in this downstream region in all cases except for two human cases using two Cα2 alleles (Fig. 6). These results suggest that the frequency of gene conversion is higher in this downstream portion than in the remaining upstream portion. As Flanagan *et al.* (2) pointed out, an alternating purine and pyrimidine sequence (positions 2277–2316 of Fig. 2) might trigger frequent gene conversions in the downstream portion.

Directions of some conversions were estimated based on category *c* sites as shown in Fig. 6. The conversions of the Cα1 gene by the Cα2 gene were identified in three clusters containing sites marked with *triangles*; two in chimpanzee and one in HCG. In gorilla, conversion of the Cα2 gene by the Cα1 gene was identified in a short cluster containing sites marked with *inverted triangles*. However, it is not clear whether a preferred direction of gene conversion exists in Cα gene conversions, unlike those observed in the γ^1-/γ^2-globin and δ-/β-globin gene regions (6, 19–21).

One drawback of the site by site reconstruction method is that the parsimonious solution is not always true, and we might miss or erroneously identify gene conversions. Human Cα2[allele A2m(2)] has nine changes of category *d* (see region *A* in Fig. 6*II*), which are regarded as substitutions after divergence of A2m(1) and A2m(2) alleles using the site by site method, while the A2m(1) allele has no such change in this region. Therefore, there is a possibility that, as Flanagan *et al.* (2) proposed, the A2m(1) allele was converted by the Cα1 gene in region *A*, after divergence into the two alleles, while the A2m(2) allele remains unconverted. Nevertheless, the site by site method gives more information about gene conversion using a phylogenetic approach than when merely comparing duplicated genes of a single species, as shown in the present study.

Gene conversion has been proposed as a process of concerted evolution of related genes. When the conversions involve a donor gene that is a more distantly related family member, they might generally result in short regions with high diversity within a family of closely related genes (see Refs. 5 and 26, for reviews). In the Cα gene loci, gene conversion apparently causes the concerted evolution. The phenomenon of concerted evolution has been considered an efficient mechanism for fixing selectively advantageous mutations among all of the members of a multigene family (5). However, category *b* positions 309 and 310 in the CH1 exon and 693–707 in the hinge region have escaped from gene conversion in all the hominoid species with two Cα2 genes. It has been

suggested that a long hinge region of IgA1 is effective for allowing antigen-binding arms to move and thus bind to antigens, whereas a short hinge region of IgA2 is important for protection from proteases produced by pathogenic bacteria in the secretory fluids where IgA is characteristically found (2, 27). Thus, these category *b* sites might be the consequence of some functional difference between IgA1 and IgA2 that has been maintained by selection. Moreover, the converted regions identified in the present study cover a large part of the Cα gene including the downstream region encompassing the poly(A) adenylation signal. These observations suggest that most gene conversions that have spread through a lineage are selectively neutral and that the disadvantageous gene conversions have been eliminated by natural selection. Therefore, we propose that gene conversion is a general phenomenon in evolution and will be found wherever duplicated sequences occur in the genome.

Acknowledgments—We are grateful to Drs. K. Omoto, O. Takenaka, and T. Honjo for their continuous encouragement, and to Mary Shimoyama and anonymous reviewers for improvement of English usage.

REFERENCES

1. Honjo, T., Alt, F. W., and Rabbitts, T. H. (1989) *Immunoglobulin Genes,* Academic Press, London
2. Flanagan, J. G., Lefranc, M.-P., and Rabbitts, T. H. (1984) *Cell* **36,** 681–688
3. Hood, L., Campbell, J. H., and Elgin, S. C. R. (1975) *Annu. Rev. Genet.* **9,** 305–353
4. Baltimore, D. (1981) *Cell* **24,** 592–594
5. Arnheim, N. (1983) in *Evolution of Genes and Proteins* (Nei, M., and Koehn, R. K., eds) pp. 38–61, Sinauer Associates Inc., Sunderland, MA
6. Fitch, D. H. A., Mainone, C., Goodman, M., and Slightom, J. L. (1990) *J. Biol. Chem.* **265,** 781–793
7. Ueda, S., Matsuda, F., and Honjo, T. (1988) *J. Mol. Evol.* **27,** 77–83
8. Kawamura, S., Omoto, K., and Ueda, S. (1990) *J. Mol. Biol.* **215,** 201–206
9. Ueda, S., Takenaka, O., and Honjo, T. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 3712–3715
10. Kawamura, S., Omoto, K., and Ueda, S. (1989) *Nucleic Acids Res.* **17,** 6732
11. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual,* 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
12. Jukes, T. H., and Cantor, C. R. (1969) in *Mammalian Protein Metabolism* (Munro, H. N., ed) pp. 21–132, Academic Press, New York
13. Kimura, M., and Ohta, T. (1972) *J. Mol. Evol.* **2,** 87–90
14. Sarich, V. M., and Wilson, A. C. (1967) *Proc. Natl. Acad. Sci. U. S. A.* **58,** 142–148
15. Wu, C.-I., and Li, W.-H. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 1741–1745
16. Sokal, R. R., and Sneath, P. H. A. (1963) *Principles of Numerical Taxonomy,* W. H. Freeman, San Francisco
17. Nei, M., Stephens, J. C., and Saitou, N. (1985) *Mol. Biol. Evol.* **2,** 66–85
18. Stephens, J. C. (1985) *Mol. Biol. Evol.* **2,** 539–556
19. Slightom, J. L., Theisen, T. W., Koop, B. F., and Goodman, M. (1987) *J. Biol. Chem.* **262,** 7472–7483
20. Slightom, J. L., Koop, B. F., Xu, P., and Goodman, M. (1988) *J. Biol. Chem.* **263,** 12427–12438
21. Koop, B. F., Seimieniac, D., Slightom, J. L., Goodman, M., Dunbar, J., Wright, P. C., and Simons, E. (1989) *J. Biol. Chem.* **264,** 68–79
22. Kimura, M. (1980) *J. Mol. Evol.* **16,** 111–120
23. Tajima, F., and Nei, M. (1984) *Mol. Biol. Evol.* **1,** 269–285
24. Saitou, N., and Nei, M. (1987) *Mol. Biol. Evol.* **4,** 406–425
25. Fitch, W. M. (1977) *Am. Natur.* **111,** 223–257
26. Wysocki, L. J., and Gefter, M. L. (1989) *Annu. Rev. Biochem.* **58,** 509–531
27. Kornfeld, S., and Plaut, A. G. (1981) *Rev. Infect. Dis.* **3,** 521–534
28. Kawamura, S., Tanabe, H., Watanabe, Y., Kurosaki, K., Saitou, N., and Ueda, S. (1991) *Mol. Biol. Evol.* **8,** 743–752