Concerted Evolution of the Primate Immunoglobulin \( \alpha \)-Gene through Gene Conversion*

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 \( \alpha \) (Ca) genes (chimpanzee Ca2, gorilla Ca2, andibbon Ca1 and Ca2 genes), which made possible the examination of gene conversions in all hominoid Ca 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 Ca genes; and 3) a site by site search of shared nucleotide changes between duplicated Ca genes. Results obtained from method 1 indicated a concerted evolution of the duplicated Ca genes in the human, chimpanzee, gorilla, andibbon lineages, while results obtained from method 2 suggested gene conversions in the human, chimpanzee, gorilla, andibbon Ca genes. With method 3 we identified clusters of shared nucleotide changes between duplicated Ca genes in human, chimpanzee, gorilla, andibbon lineages, and in their hypothetical ancestors. In the present study converted regions were identified over the entire Ca 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 non-functional regions.

The mammalian immunoglobulin heavy chain constant region (CH) gene cluster has evolved through multiple gene duplications and resulting specializations into its present classes (Ca, Cb, Cx, Cy, Cz, and Co). The CH gene organization of human is 5'-'CH3-CH2-CH1-CHC-Ca-1-CHC-Ca-1-CHC-Ca-1-CHC-Ca-3' while that of mouse is 5'-'CH1-CH2-CH1-CHC-Cx-2-CHC-Cx-2-CHC-Ca-3' (Ref. 1, and references therein). Comparing the nucleotide sequences of the human immunoglobulin Ca1 gene and two alleles [A2m(1) and A2m(2)] of the Ca2 genes, Flanagan et al. (2) found that the nucleotide sequence of a contiguous stretch in the 3' end of the Ca1 gene is identical with that of the A2m(1) allele of the Ca2 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 Ca1 gene to the A2m(1) allele of the Ca2 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 Ca4 gene cluster in non-human primates (7–10). Chimpanzee and gorilla were found to have the Ca2-Ca1 and Ca1-Ca2 regions as does human, although all exons and introns of the chimpanzee Ca2 gene are deleted (7). Orangutan, gibbon, and Old World monkey have one, two, and one Ca genes, respectively (8), and have one Ca gene in their CH gene clusters (9). Our recent study shows that duplication, including the Cy-Cy-Cy-Ca genes, occurred in the common ancestor of hominoids (human, chimpanzee, gorilla, orangutan, andibbon), followed by the deletion of the gibbon C5 gene upstream from the Ca1 gene and also of one of the Ca-Ca regions in orangutan. Support for this gene phylogeny can also be found when the evolution of the Ca hinge region is considered. Human, chimpanzee, andgorilla share different hinge structures in their Ca1 and Ca2 genes (four tandem repeats of a 15-bp unit with a 6-bp overlap in the Cai genes and only one 15-bp unit in the Ca2 genes) (7, 10). The hinge region of the orangutan Ca gene is a Ca1 type of human, chimpanzee, andgorilla, while that of the gibbon Ca1 gene consists of two tandem repeats of a 15-bp unit with a 6-bp deletion. The hinge region of the gibbon Ca2 gene is a Ca2 type of human, chimpanzee, andgorilla (8). The most parsimonious explanation for the evolutionary history of the hominoid Ca hinge region is a gene phylogeny in which duplication of the Ca gene occurred in the common ancestor of hominoids and the deletion of the Ca2 gene occurred in the orangutan lineage (Fig. 1).

On the basis of the Ca gene phylogeny described above, we examined in the present study whether gene conversion has occurred in the duplicated Ca genes of primates. We sequenced chimpanzee and gorilla Ca2 genes as well as Ca1 and Ca2 genes ofibbon. These new sequence data made it possible to compare all the nucleotide sequences of hominoid and Old World monkey Ca genes. Our phylogenetic analysis...
Fig. 1. Ca gene phylogeny in higher primates. A gene duplication in the common ancestor of hominoids and a deletion of the Ca2 gene in the orangutan lineage are indicated. For convenience, 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 Ca hinge region in hominoids. Since the hinge region of the ancestral Ca 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 gibbon Ca1 gene; 2) a 6-bp deletion from the first unit of the orangutan Ca gene.

Fig. 2. Aligned Ca gene sequences. Sequences of the human Ca1, Ca2(allele A2m(1)), and Ca2(allele A2m(2)), the chimpanzee Ca1 and Ca2, the gorilla Ca1 and Ca2, the orangutan Ca, the gibbon Ca1 and Ca2, and the crab-eating macaque Ca genes (designated by Hu, 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 Ca1 gene and only nucleotides different from it are shown in other Ca 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 Ca1 sequence, and those for termination codon, putative polyadenylation signal (aaataa), and alternating purine and pyrimidine sequence are represented by Ter, poly-A, and Pu/Py, respectively. Human Ca1 and Ca2 gene sequences are from Flanagan et al. (2), and the nucleotide sequences of the chimpanzee Ca1, the gorilla Ca1, the orangutan Ca, and the crab-eating macaque Ca genes are from Kawamura et al. (28).

showed that gene conversions occurred on the duplicated Ca genes of the chimpanzee, gorilla, gibbon, and their ancestors. We also found that human Ca gene conversion occurred not only between the Ca1 and the Ca2[A2m(1)] genes but also between the Ca1 gene and the ancestral Ca2 gene of the two alleles.

MATERIALS AND METHODS

DNA Manipulations—Recombinant phage clones containing chimpanzee (Pan troglodytes) Ca2, gorilla (Gorilla gorilla) Ca2, and gibbon (Hylobates lar) Ca1 and Ca2 genes were isolated previously from their genome libraries (7, 8). A restriction fragment from each phage clone containing the Ca gene was subcloned into pUC119 or Blue-script 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 Ca1, Ca2(allele A2m(1)), and Ca2(allele A2m(2)), chimpanzee Ca1 and Ca2, gorilla Ca1 and Ca2, orangutan (Pongo pygmaeus) Ca, gibbon Ca1 and Ca2, and crab-eating macaque (Macaca fascicularis; an Old World monkey) Ca genes (Fig. 2). The
FIG. 2.—continued

**Fig. 2.** Plots showing the alignment of sequences from various primates for the primate immunoglobulin Ca gene. The sequences are shown as nucleotide alignments with G+C content indicated. The figure includes sequences from Hul, HuZ-1, Hu2-1, Hu2-2, Chl, Ch2, Go1, Go2, Or, GI1, GI2, Cr, H02-1, H02-2, and Poly-A. The alignment shows the conservation of sequence elements across different species, highlighting regions of similarity and divergence. The figure also includes a schematic representation of exon-intron boundaries and potential polyadenylation signals. The legend for the figure is provided at the bottom of the page, explaining the symbols and color coding used in the alignment.
Primate Immunoglobulin Ca Gene Conversion

FIG. 2.—continued

estimated number of nucleotide substitutions per site \((K')\) between two DNA sequences was computed using the formula: \(K' = -(3/4)\ln[1-(4/3)K]\), where \(K\) is the number of nucleotide differences per site (12). The variance of the \(K'\) value is given by \(V(K') = K(1-K)/n\), where \(n\) is the number of nucleotides compared (13). The standard error (S.E.) of the \(K'\) value is given by the square root of \(V(K')\). The variance of \(K'\) value is given by the square root of \(V(K')\). The variance of \(K'\) value is given by the square root of \(V(K')\).

A relative rate test (14) was performed by comparing the evolutionary distances \((K')\) from the crab-eating macaque \(C_0\) gene to hominoid \(C_a\) genes applying the statistical method of Wu and Li (15). When \(K'\) between hominoid gene \(i\) and crab-eating macaque gene \((m)\) is referred to as \(K_{im}\), a variance of \(K_{im} - K_{in}\) is given by \(V(K_{im}) + V(K_{in}) - 2V(K'_{im})\), where \(K_{im} = (K_{im} + K_{in} - K_{in})/2\), \(V(x)\) denotes the variance of \(x\) (15).

A phenetic gene tree was constructed from the \(K'\) 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 \(C_a1\) and \(C_a2\) 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 \(p = 1 - \left[1 - \sum \frac{1}{s + r - 1} C_{s+r-1} \cdot s C_i \cdot (s + r - 1)\right]^{-r}\), where \(s\) is the number of varied sites, \(r\) is the total number of unvaried sites placed between varied sites, and \(g_s\) 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 (6) method. At each aligned position (Fig. 2), a parsimony reconstruction was made of the nucleotide changes which occurred, assuming the \(C_a\) 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 Ca1 and Ca2 genes of a single species (shown as open circles in Figs. 5a and 6); (b) variants shared among orthologous Ca1 or Ca2 genes of different species, but not between paralogous Ca1 and Ca2 within a species, or those shared between two alleles of the human Ca2 gene but not with the human Ca1 gene (shown as open squares in Figs. 5b and 6); (c) variants in which paralogous genes of the same species had the same nucleotide (shown as triangles when the Ca1 gene changed, or inverted triangles when the Ca2 gene changed (Figs. 5c and 6); and (d) all other variants (shown as full circles in Figs. 5d 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 Ca2, gorilla Ca2, and ribbon Ca1 and Ca2 genes, and compared them with human Ca1 and Ca2[A2m(1) and A2m(2) alleles], chimpanzee Ca1, gorilla Ca1, orangutan Ca1, and crab-eating macaque Ca1 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 Ca1 gene (positions 78-139 of Fig. 2) reported by Flanagan et al. (2) was not found in any non-human Ca1 gene. Therefore, this probably occurred only in the human lineage. 

Phenetic Tree Construction—Table I shows the K' distance matrix for primate Ca genes. Other estimation methods (22, 23) for superimposed nucleotide substitution gave essentially the same K' 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 Ca 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 Co genes of human, chimpanzee, gorilla, and ribbon formed statistically significant clusters in every species (clusters A, C, B, and D of Fig. 3, respectively) at the 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 Co 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 Ca gene duplication in the common ancestor of hominoids and the deletion of the orangutan Ca2 gene (see Introduction and Fig. 1), these results indicate that a concerted evolution of the Ca genes occurred in human, chimpanzee, gorilla, and ribbon 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 Co genes and to elucidate where it occurred in the gene region, we searched for consecutive unvaried regions between paired human, chimpanzee, gorilla, and ribbon Ca 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 Ca1 and Ca2 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 Ca genes |
| A total of 2554 nucleotides were used for comparison. Values are K' × 100. Standard errors are given in parentheses. |
| **Human** | **Human** | **Chimpanzee** | **Chimpanzee** | **Gorilla** | **Gorilla** | **Orangutan** | **Gibbon** | **Gibbon** |
| Ca1 | Ca2[A2m(1)] | Ca1 | Ca2 | Ca1 | Ca2 | Ca1 | Ca2 | Ca1 | Ca2 |
|------|-------------|------|------|------|------|------|------|------|------|------|
| Human Ca2[A2m(1)] | 2.07 (0.30) | 2.81 (0.35) | 3.52 (0.39) | 2.33 (0.32) | 3.12 (0.37) | 3.12 (0.37) | 5.19 (0.48) | 5.28 (0.52) |
| Chimpanzee Ca1 | 2.42 (0.32) | 2.37 (0.32) | 2.77 (0.35) | 2.73 (0.32) | 2.51 (0.35) | 2.99 (0.36) | 5.73 (0.50) | 5.55 (0.56) |
| Chimpanzee Ca2 | 2.29 (0.32) | 2.29 (0.32) | 2.73 (0.32) | 2.73 (0.32) | 2.99 (0.36) | 2.99 (0.36) | 5.73 (0.50) | 5.55 (0.56) |
| Gorilla Ca1 | 2.33 (0.32) | 2.33 (0.32) | 2.77 (0.35) | 2.73 (0.32) | 2.51 (0.35) | 2.99 (0.36) | 5.73 (0.50) | 5.55 (0.56) |
| Gorilla Ca2 | 2.42 (0.32) | 2.42 (0.32) | 2.77 (0.35) | 2.73 (0.32) | 2.99 (0.36) | 2.99 (0.36) | 5.73 (0.50) | 5.55 (0.56) |
| Orangutan Ca1 | 1.81 (0.51) | 1.81 (0.51) | 2.51 (0.50) | 2.51 (0.50) | 0.49 (0.49) | 0.49 (0.49) | 0.51 (0.51) | 0.51 (0.51) |
| Orangutan Ca2 | 1.81 (0.51) | 1.81 (0.51) | 2.51 (0.50) | 2.51 (0.50) | 0.49 (0.49) | 0.49 (0.49) | 0.51 (0.51) | 0.51 (0.51) |
| Gibbon Ca1 | 5.19 (0.48) | 5.19 (0.48) | 5.74 (0.51) | 5.74 (0.51) | 5.87 (0.50) | 5.87 (0.50) | 5.83 (0.51) | 5.83 (0.51) |
| Gibbon Ca2 | 5.24 (0.51) | 5.24 (0.51) | 5.46 (0.50) | 5.46 (0.50) | 5.65 (0.50) | 5.65 (0.50) | 5.78 (0.51) | 5.78 (0.51) |
| Crab-eating macaque | 7.45 (0.58) | 7.45 (0.58) | 7.22 (0.57) | 7.22 (0.57) | 7.74 (0.60) | 7.74 (0.60) | 7.50 (0.57) | 7.50 (0.57) |
| Ca1 | 7.50 (0.59) | 7.50 (0.59) | 7.50 (0.59) | 7.50 (0.59) | 7.50 (0.59) | 7.50 (0.59) | 7.50 (0.59) | 7.50 (0.59) |
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 by frequent recombinational events (8). The longest unvaried segment was significantly long in human, gorilla, and gibbon Ca 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 − longest gc| 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 γ1−/γ2−globin conversions and β/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 Ca 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 Ca gene region for the paralogous gene pair of each species including ancestors, we identified clusters of shared nucleotide changes between paralogous Ca 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 Ca 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 Ca2 gene, is indicated only by a short cluster of two a sites (Fig. 6II). Furthermore, conversion in region 4 of Fig. 4, which was identified using allele A2m(2) as a Ca2 gene, is not indicated by any such cluster (Fig. 6III). Instead, we found a long stretch of a sites in the gene pair of human Ca1 and ancestral Ca2 gene of the two alleles, which overlaps regions 2 and 4 of Fig. 4 (Fig. 6III). Therefore, conversions in regions 2 and 4 of Fig. 4 are considered to have occurred not independently but

### Table II

| Gene 1/gene 2 | K_{in} | K_{ev} | S.E. | p* |
|---------------|-------|-------|------|----|
| Human Co1/human Co2[A2m(1)] | 0.24 | 0.32 | >0.4 |
| Human Co1/chimpanzee Co1 | -0.05 | 0.37 | >0.8 |
| Human Co1/chimpanzee Co2 | -0.28 | 0.41 | >0.4 |
| Human Co1/gorilla Co1 | 0 | 0.34 | 1 |
| Human Co1/gorilla Co2 | -0.05 | 0.39 | >0.9 |
| Human Co1/orangutan Co | 0.42 | 0.50 | >0.4 |
| Human Co1/gibbon Co1 | -0.05 | 0.50 | >0.9 |
| Human Co1/gibbon Co2 | 0.05 | 0.55 | >0.9 |
| Human Co2(A2m(1))/chimpanzee Co1 | -0.28 | 0.34 | >0.4 |
| Human Co2(A2m(1))/chimpanzee Co2 | -0.52 | 0.34 | >0.1 |
| Human Co2(A2m(1))/gorilla Co1 | -0.24 | 0.37 | >0.5 |
| Human Co2(A2m(1))/gorilla Co2 | -0.28 | 0.35 | >0.4 |
| Human Co2(A2m(1))/orangutan Co | 0.19 | 0.50 | >0.7 |
| Human Co2(A2m(1))/gibbon Co1 | -0.28 | 0.52 | >0.5 |
| Human Co2(A2m(1))/gibbon Co2 | -0.28 | 0.51 | >0.5 |
| Chimpanzee Co1/chimpanzee Co2 | -0.24 | 0.33 | >0.4 |
| Chimpanzee Co1/gorilla Co1 | 0.05 | 0.36 | >0.8 |
| Chimpanzee Co1/gorilla Co2 | 0 | 0.38 | 1 |
| Chimpanzee Co1/orangutan Co | 0.47 | 0.51 | >0.3 |
| Chimpanzee Co1/gibbon Co1 | 0 | 0.50 | 1 |
| Chimpanzee Co1/gibbon Co2 | 0 | 0.51 | 1 |
| Chimpanzee Co2/gorilla Co1 | 0.28 | 0.39 | >0.4 |
| Chimpanzee Co2/gorilla Co2 | 0.24 | 0.39 | >0.5 |
| Chimpanzee Co2/orangutan Co | 0.70 | 0.53 | >0.1 |
| Chimpanzee Co2/gibbon Co1 | 0.24 | 0.53 | >0.6 |
| Chimpanzee Co2/gibbon Co2 | 0.24 | 0.52 | >0.6 |
| Gorilla Co1/gorilla Co2 | -0.05 | 0.30 | >0.8 |
| Gorilla Co1/orangutan Co | 0.42 | 0.48 | >0.3 |
| Gorilla Co1/gibbon Co1 | -0.05 | 0.49 | >0.9 |
| Gorilla Co1/gibbon Co2 | 0.05 | 0.53 | >0.9 |
| Gorilla Co2/orangutan Co | -0.47 | 0.51 | >0.3 |
| Gorilla Co2/gibbon Co1 | 0.47 | 0.52 | >0.3 |
| Gorilla Co2/gibbon Co2 | 0 | 0.53 | 1 |
| Orangutan Co/gibbon Co1 | -0.47 | 0.51 | >0.3 |
| Orangutan Co/gibbon Co2 | -0.47 | 0.52 | >0.3 |
| Gibbon Co1/gibbon Co2 | 0 | 0.34 | 1 |

a K_{in} (i = 1 or 2) is the number of nucleotide substitutions per 100 sites between hominoid Ca gene i and crab-eating macaque Ca gene.

b Probability that there is no difference between K_{in} and K_{ev}, which is evaluated using a two-tailed t test with degrees of freedom = ∞.

![FIG. 3. A unweighted pair group method with arithmetic mean phenetic tree for the Ca gene.](image-url)
Before the divergence of the human Ca2 gene into the present two alleles, the Ca1 and Ca2 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 Ca 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.) of segments 1-9 are 165, 886, 180, 454, 201, 532, 179, 549, and 244 nucleotides, respectively. Their probabilities (p) are 0.83, 1.53 \times 10^{-8}, 0.59, 1.38 \times 10^{-5}, 0.34, 7.80 \times 10^{-4}, 0.79, 3.26 \times 10^{-3}, 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 \times 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 Ca1, Ca2[allele A2m(1)], and Ca2[allele A2m(2)], chimpanzee Ca1 and Ca2, gorilla Ca1 and Ca2, and gibbon Ca1 and Ca2 genes are denoted by Hu1, Hu2(1), Hu2(2), Ch1, Ch2, Go1, Go2, G1, and G2, respectively. The numbers (r) of varied sites in Hu1-Hu2(1), Hu1-Hu2(2), Ch1-Ch2, Go1-Go2, and G1-G2 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. 4. Distribution of the different sites between paralogous Ca genes of the human, chimpanzee, gorilla, and gibbon. Different sites between the Ca1 and Ca2 genes of each species are indicated by open circles. Positions of CH exons are presented by filled circles. 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.) of segments 1-9 are 165, 886, 180, 454, 201, 532, 179, 549, and 244 nucleotides, respectively. Their probabilities (p) are 0.83, 1.53 \times 10^{-8}, 0.59, 1.38 \times 10^{-5}, 0.34, 7.80 \times 10^{-4}, 0.79, 3.26 \times 10^{-3}, 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 \times 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 Ca1, Ca2[allele A2m(1)], and Ca2[allele A2m(2)], chimpanzee Ca1 and Ca2, gorilla Ca1 and Ca2, and gibbon Ca1 and Ca2 genes are denoted by Hu1, Hu2(1), Hu2(2), Ch1, Ch2, Go1, Go2, G1, and G2, respectively. The numbers (r) of varied sites in Hu1-Hu2(1), Hu1-Hu2(2), Ch1-Ch2, Go1-Go2, and G1-G2 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 Ca1 and Ca2 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 Ca2 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 Ca2 genes and no additional change occurred in the Ca1 genes of hominoids as well, suggesting no gene conversion before the divergence of the human Ca2 gene into the present two alleles. In the Ca1 and Ca2[allele A2m(1)] gene pair, three possible conversions with short stretches were found (Fig. 6).

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. 6IV). 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 Ca2 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 Ca1 and the gorilla Ca2 genes after a nucleotide change occurred in the common ancestral Ca1 gene of hominoids. These additional changes result in the same nucleotide with their paralogs, suggesting a gene conversion of the chimpanzee Ca1 gene by the Ca2 gene and that of the gorilla Ca2 gene by its Ca1 gene over this site. Therefore, triangles are given to the chimpanzee Ca1 gene and inverted triangles are given to the gorilla Ca2 gene. In case d, the root is t and changes from t to c occur in the gibbon Ca1 and the gorilla Ca2 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 Ca1 and Ca2 genes, respectively. The A2m(1) and A2m(2) alleles of the human Ca2 gene are shown as 2(1) and 2(2), respectively.
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 Ca2 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 Ca1 gene by the Ca2 gene were identified in three clusters containing sites marked with triangles; two in chimpanzee and one in HCG. In gorilla, conversion of the Ca2 gene by the Ca1 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 Ca gene conversions, unlike those observed in the \( \gamma^1/\gamma^2 \)-globin and b/\( b \)-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 Ca2(allele A2m(2)) has nine changes of category d (see region A in Fig. 6H), 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 Ca1 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 Ca 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 Ca2 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 Ca 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, 781–783
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. Ecol.* 2, 66–85
18. Stephens, J. C. (1985) *Mol. Biol. Ecol.* 2, 539–556
19. Slightom, J. L., Thenis, 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. Ecol.* 1, 269–285
24. Saitou, N., and Nei, M. (1987) *Mol. Biol. Ecol.* 4, 406–425
25. Fitch, W. M. (1971) *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. Ecol.* 8, 743–752