Dating the Primigenial *C4-CYP21* Duplication in Primates

Yoshihito Horiuchi,*[,1] Hiroshi Kawaguchi,*[,1] Felipe Figueroa,* Colm O'hUigin* and Jan Klein*[,†]

**Max-Planck-Institut für Biologie, Abteilung Immungenetik, D-7400 Tübingen, Germany, and †Department of Microbiology and Immunology, University of Miami School of Medicine, Miami, Florida 33101*

Manuscript received September 10, 1992
Accepted for publication January 29, 1993

ABSTRACT

C4 and *CYP21* are two adjacent, but functionally unrelated genes residing in the middle of the mammalian major histocompatibility complex (*Mhc*). The *C4* gene codes for the fourth component of the complement cascade, whereas the *CYP21* gene specifies an enzyme (cytochrome P450c21) of the glucocorticoid and mineralocorticoid pathways. The genes occur frequently in multiple copies on a single chromosome arranged in the order *C4 . . . CYP21 . . . C4 . . . CYP21*. The unit of duplication (a module) is the *C4-CYP21* gene pair. We sequenced the flanking regions of the C4-CYP21 modules and the intermodular regions of the chimpanzee, gorilla, and orangutan, as well as the intermodular region of an Old World monkey, the pigtail macaque. By aligning the sequences, we could identify the duplication breakpoints in these species. The breakpoint turned out to be at exactly the same position as that found previously in humans. The sequences flanking paralogous genes in the same species were found to be more similar to one another than sequences flanking orthologous genes in different species. We interpret these results as indicating that the original (primigenial) duplication occurred before the separation of apes from Old World monkeys more than 23 million years ago. The nature of the sequence at the breakpoint suggests that the duplication occurred by nonhomologous recombination. Since then, the *C4-CYP21* haplotypes have been expanding and contracting by homologous crossing over which has homogenized the sequences in each species. We speculate that the reason for the concerted evolution of the primate *C4-CYP21* region may be a requirement for the coevolution of certain components of the complement pathway, including the *C4* component. We contrast the evolution of the *C4-CYP21* region with that of other *Mhc* regions.

COMPLEMENT component 4, C4, is one of more than 20 proteins constituting the complement cascade, which is involved in defense of vertebrate bodies against pathogens. Activation of the cascade by antigen-antibody complexes or by other means leads ultimately to the assembly of lytic complexes on the cell surface, perforation of the plasma membrane, and killing of the cell (ROSS 1986). Some of the activated components of the cascade are also involved in a variety of other biological functions. Cytochrome P450c21 (21-hydroxylase), CYP21, is an enzyme participating in the conversion of cholesterol to aldosterone or cortisol in the cortex of the vertebrate adrenal gland (MILLER 1988). In mammals, complement component 4 and cytochrome P450c21 are encoded in adjacent genes, *C4* and *CYP21*, respectively, residing in the middle of the major histocompatibility complex, *Mhc* (KAWAGUCHI, O'HUIGIN and KLEIN 1991). The per haplotype number of *C4* and *CYP21* copies varies from species to species and also among individuals of certain species. Single *C4-CYP21* haplotypes have been found in humans (MCLEAN *et al.* 1988; and others), Syrian hamster (LÉVI-STRAUSS *et al.* 1985), dog (KAY and DAWKINS 1984; DOXIADIS *et al.* 1985), cat (DE KROON *et al.* 1986), guinea pig (BITTER-SUERMANN *et al.* 1977), and several whale species (SPILLIAERT, PALSDOTTIR and ARNASON 1990). Haplotypes with two *C4* and two *CYP21* genes are common in humans (MCLEAN *et al.* 1988; and others), chimpanzees (KAWAGUCHI *et al.* 1990; BONTROP *et al.* 1990; CHRISTIANSEN *et al.* 1991), gorillas (KAWAGUCHI and KLEIN 1992), macaques (MEVAG *et al.* 1983), mice (ROOS, ATKINSON and SHREFFLER 1978), rats (TOSI *et al.* 1985), cattle (YOSHIOKA *et al.* 1986; CHUNG, MATTESON and MILLER 1985, 1986), pigs (KIRSZENBAUM *et al.* 1985), and horses (KAY *et al.* 1987). Haplotypes with three *C4* and three *CYP21* genes have been reported in humans (MCLEAN *et al.* 1988) and orangutans (KAWAGUCHI and KLEIN 1992). Finally, haplotypes with four *C4* and four *CYP21* genes occur occasionally in humans (MCLEAN *et al.* 1988) and orangutans (ZHANG *et al.* 1993). In haplotypes carrying multiple copies of *C4* and *CYP21* genes, the two types of gene always alternate on the linkage map (*i.e.*, *C4 . . . CYP21 . . . C4 . . . CYP21*, etc.), suggesting that the basic unit is a *C4-CYP21* module and that multimodular haplotypes arise by multiplication of this unit (KAWAGUCHI, O'HUIGIN and KLEIN 1991). The human *C4-CYP21* module is about 35 kilobases (kb) long and the distance between the two genes, transcription-

[1] Permanent address: Department of Dermatology, Yokohama City University School of Medicine, Yokohama, Japan.

ally oriented in the same direction, is approximately 3 kb (CARROLL, CAMPBELL and PORTER 1985; DUNHAM *et al.* 1987).

Recent studies indicate, however, that in reality the module contains three genes, the third one being transcribed from the DNA strand complementary to that from which the *C4* and *CYP21* genes are transcribed (MOREL *et al.* 1989). The third gene codes for a protein related to the extracellular matrix protein tenascin (MATSUMOTO *et al.* 1992; GITELMAN, BRISTOW and MILLER 1992) and is either referred to as *MHC-F3* because of its location in the *Mhc* region and the presence of fibronectin type III domains (MATSUMOTO *et al.* 1992), or simply as "X" (MOREL *et al.* 1989). The last exon coding for the 3′-untranslated (3′UT) region of the *CYP21* gene overlaps with the last exon of the *MHC-F3(X)* gene, but the two genes are oriented in opposite directions. In haplotypes with multiple copies of the *CYP21* gene, all but one of the *MHC-F3(X)* copies are truncated (GITELMAN, BRISTOW and MILLER 1992) and most likely pseudogenes.

GITELMAN, BRISTOW and MILLER (1992) sequenced the DNA in the region between the two modules of a bimodular human haplotype and found that the sequence in the 3′ part of the region matches that found upstream from the first module, whereas the sequence in the 5′ part aligns with that found downstream from the second module. The two parts of the intermodular sequences overlap by four nucleotides in the center of the region. This result suggests that the two modules arose by duplication from a single module and that this event occurred by nonhomologous recombination.

In the present study we attempt to answer the following questions: When did the duplication occur? Did nonhomologous recombination occur repeatedly in different primate species and different haplotypes, or did it occur only once, and if so, by what mechanism did the subsequent variation in module number arise? To this end, we sequenced the relevant segments of the *C4-CYP21* region in the chimpanzee, gorilla, orangutan and pigtail macaque.

MATERIALS AND METHODS

Source of cosmid clones and DNA: The relevant DNA segments were obtained from cosmid clones. Three cosmid libraries were used. The first library was constructed using DNA isolated from the Epstein-Barr virus (EBV)-transformed B cell line, Hugo, established from the common chimpanzee (*Pan troglodytes*) at the TNO Institute of Applied Radiobiology and Immunology, Rijswijk, The Netherlands (MAYER *et al.* 1988). The second library was based on DNA of a western lowland gorilla (*Gorilla gorilla*) isolated from the fibroblast line Sylvia, which was established from a skin biopsy sample by KIRBY D. SMITH, The Johns Hopkins University School of Medicine, Baltimore, Maryland. The DNA for the third cosmid library was isolated from the cell line CP81, which was established from monocytic leukemia cells of a 13-year-old female orangutan at the Los Angeles Zoological Garden (RASHEED *et al.* 1977). The pigtail macaque (*Macaca nemestrina*) DNA was derived from the EBV-transformed B cell line 86081 kindly provided to us by LAKSHMI GAUR, HLA Laboratory, Puget Sound Blood Center, Seattle, Washington. Genomic DNA was isolated from the indicated cell lines according to the method described by MANIATIS, FRITSCH and SAMBROOK (1982) and the libraries were constructed and screened according to STEINMETZ *et al.* (1985); for a full description of the libraries, see KAWAGUCHI *et al.* (1990) and KAWAGUCHI and KLEIN (1992).

Analysis of cosmid clones: DNA isolated from cosmid clones following the protocol given in MANIATIS, FRITSCH and SAMBROOK (1982) was digested with restriction endonucleases, and the resulting fragments were separated by agarose gel electrophoresis and transferred to nitrocellulose membranes (Sartorius, Göttingen, Germany). The relevant fragments were subcloned into pBluescript II SK$^+$ phagemid vector (Stratagene, Heidelberg, Germany) according to the standard method (DAVIS, DIBNER and BATTEY 1986).

Sequencing: Subclones were ligated to the pBluescript II SK$^+$ phagemid vectors and the recombinant clones were picked up by the colony hybridization method (DAVIS, DIBNER and BATTEY 1986). Double-stranded DNA was prepared by the plasmid mini-boiling method (HOLMES and QUIGLEY 1981). Five micrograms of DNA were denatured in 0.2 M NaOH, 0.2 mM EDTA for 30 min at 37° and sequenced by the dideoxy chain-termination method (SANGER, NICKLEN and COULSON 1977), using the Sequenase version 2.0 kit (U.S. Biochemicals, Cleveland, Ohio). Fragments were sequenced on both strands two to three times to eliminate or resolve sequencing errors and ambiguities.

Polymerase chain reaction (PCR): Genomic DNA was amplified in the GeneAmp PCR system 9600 (Perkin-Elmer, Überlingen, Germany). Two hundred nanograms of DNA were initially denatured by heating to 94° followed by a 30-cycle profile of 20 sec at 94°, 20 sec annealing at 50°, and 30 sec extension at 72° (using the GeneAmp PCR reagent kit (Perkin-Elmer). The primers used for amplification were 5′-GACTCCTTGATGGATGTTGA-3′ (Tu336) and 5′-AAGGACAGCCTGGCGCCCT-3′ (Tu337), which were specific for sequences located 105 base pairs (bp) upstream and 116 bp downstream of the human recombination breakpoint in the intermodular region, respectively. The amplified products were isolated by electrophoresis on low melting point agarose gel, cloned in Bluescript II SK$^+$ vector, and sequenced.

Construction of dendrograms: The neighbor-joining genetic distance method of SAITOU and NEI (1987) was used for evaluating evolutionary relationships among nucleotide sequences. Genetic distances were calculated by the two-parameter method (KIMURA 1980).

RESULTS

Contig maps of the *C4-CYP21* region in the chimpanzee (Figure 1), gorilla (Figure 2) and orangutan (Figure 3) were constructed previously by KAWAGUCHI *et al.* (1990) and KAWAGUCHI and KLEIN (1992). They revealed the existence of two *C4-CYP21* modules each in the chimpanzee and gorilla, and three modules in the orangutan. To identify the breakpoints of the duplication that produced these multimodular chromosomes, we isolated cosmid fragments located upstream from the most 5′ *C4* gene (region A in Figures

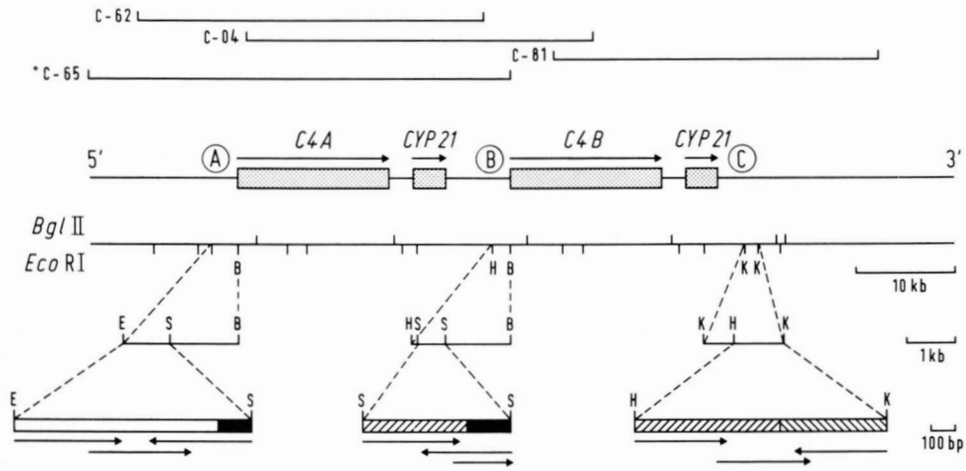

FIGURE 1.—Contig map of the chimpanzee *C4-CYP21* region. Segments at the top indicate the extent of the relevant cosmids (numbered; an asterisk specifies an allelic cosmid). The boxes in the middle portion indicate the genes, arrows the transcriptional orientation of the genes, and the circled letters the sequenced regions. The lower part of the figure gives the restriction map and the blowups of the relevant restriction fragments (B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; S, *Sac*I). The rectangles at the lowest part indicate the fragments sequenced (sequencing strategy is shown by arrows and homology by different shading).

1–3), between the modules (region B in Figures 1 and 2, as well as regions B and B′ in Figure 3), and downstream from the most 3′ *CYP21* gene (region C in Figures 1–3). The chimpanzee (Figure 1) 2.4-kb A-region fragment was derived from the cosmid clone C-62 after digestion with *Eco*RI and *Bam*HI. Further digestion with *Sac*I produced a 0.95-kb fragment which was subcloned and sequenced. Two allelic chimpanzee B-region fragments were derived from cosmids C-65 and C-04 (2.0-kb *Hin*dIII-*Bam*HI fragments which were then used to produce 0.6-kb *Sac*I fragments for sequencing). The chimpanzee C-region fragment was derived from cosmid clone C-81 (a 1.6-kb *Kpn*I fragment produced, upon digestion with *Hin*dIII, a 1-kb fragment, which was sequenced). The gorilla (Figure 2) A-region fragment was obtained from cosmid G-1 upon digestion with *Eco*RI and *Bam*HI (2.3 kb) and subsequent trimming with *Sac*I (0.95 kb). The B-region fragments were derived from cosmids G-21 and G-2 (mapping to identical regions in homologous chromosomes) after digestion, first with *Eco*RI (14.5 kb) and then with *Sac*I (0.6 kb). The gorilla C-region fragment was produced by digestion of the cosmid clone G-18, first with *Eco*RI (8.0 kb) and then with *Sac*I (2.4 kb). The orangutan (Figure 3) A-region fragment came from the cosmid clone O-15, digested first with *Eco*RI/*Bam*HI (2.4 kb) and then with *Sac*I (0.95 kb). The orangutan B- and B′-region fragments were derived from clones O-17 and O-4, respectively, by digestion with *Kpn*I (6.0 kb) and then with *Sac*I (0.6 kb). The orangutan C-region fragment was produced from clone O-6 by digestion with *Eco*RI (14.5 kb), followed by digestion with *Sac*I (2.5 kb). All these fragments were subcloned and sequenced. In the case of the pigtail macaque, the DNA derived from the intermodular region was amplified by PCR using primers based on the human sequence. The amplification produced the expected 263-bp band, which was then sequenced.

In all four primate species, part of the B (B′)-region sequence was found to align with part of the A-region sequence (we refer to the former segment as B-A), while an adjacent part of the B (B′)-region sequence aligned with part of the C-region sequence [this seg-

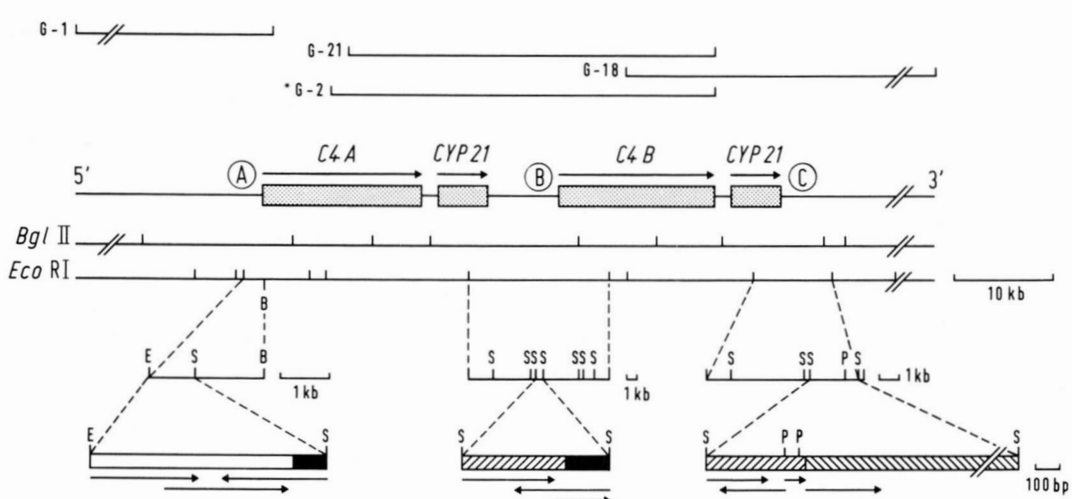

FIGURE 2.—Contig map of the gorilla *C4-CYP21* region. For explanations, see legend to Figure 1.

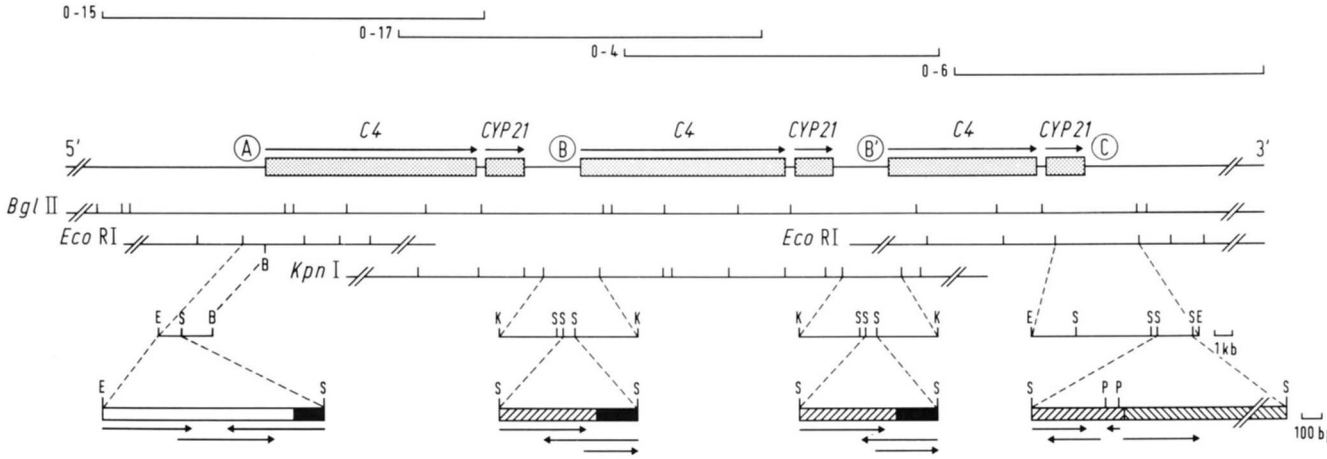

FIGURE 3.—Contig map of the orangutan *C4-CYP21* region. For explanation, see legend to Figure 1.

ment is referred to as B-C; see Figures 4 and 5 in which the human sequences derived from GITELMAN, BRISTOW and MILLER (1992) are also given]. At the site of transition from B-C to B-A, four nucleotides (CAAG) were found that occurred also at the 5′ end of homology in the A region and at the 3′ end of homology in the C region. The only deviations were a deletion of the C in the gorilla A-region sequence, a G → A substitution in the orangutan A-region sequence, and an A → T substitution in the chimpanzee C-region. The CAAG tetranucleotide in the B region is the site of change in homology: the segment 5′ of this site is homologous to the C region, whereas the 3′ segment is homologous to the A region. This homology switch indicates that the breakpoint that resulted in the primigenial duplication of the module lies at this tetranucleotide site. The tetranucleotide overlap between the A and C regions suggests that the duplication occurred by nonhomologous recombination (see DISCUSSION). The presence of the same B-A to B-C segment transition in humans, the great apes, and Old World monkeys (here represented by the pigtail macaque) indicates that the primigenial duplication occurred more than 23 million years (my) ago, the estimated time of separation of the ape and Old World monkey lines (MARTIN 1990). In the orangutan, which has three *C4-CYP21* modules and hence two intermodular regions, the same B-A to B-C and B′-A to B′-C transition indicates that the third module most likely arose by unequal crossing over from misaligned multimodular chromosomes. Had the third module arisen by an independent nonhomologous recombination, one would expect the breakpoint to have occurred by chance at a position different to that between the first and the second module.

Theoretically, it should be possible to estimate how long ago the primigenial duplication occurred from the divergence of the A- from the B-A- and of the C- from the B-C-region sequences. This calculation assumes, however, that the A and B-A, as well as the C and B-C regions, evolved independently. To test whether this assumption is fulfilled, we constructed dendrograms based on the sequences of the three regions (Figures 6 and 7). Had the various regions evolved independently, one would expect the A-region sequences of the different species to cluster together and the different B-A sequences to form another cluster; similarly, the C-region sequences of the different species should form a cluster separately from the B-C region sequences. Figures 6 and 7, however, reveal a different picture: The sequences do not cluster according to regions, but according to species. Thus there is one branch with the chimpanzee A and B-A sequences, another branch with the gorilla A and B-A sequences, and another branch still with the orangutan A and B-A sequences (Figure 6). A similar picture emerges in the dendrogram of the B and B-C sequences (Figure 7). (Because of their shortness, the human and pigtail macaque sequences were not included in these comparisons.) We note also that the orangutan B-C and B′-C sequences are very similar to each other: They differ by a single nucleotide substitution only and share several substitutions and deletions absent in all the other sequences. They must, therefore, have diverged very recently.

Taken together, these data indicate that the A and B-A, as well as C and B-C regions of the *C4-CYP21* modules are *not* evolving independently and hence that their sequences cannot be used for estimating the time the primigenial duplication occurred. The observation that different regions within the same species are more similar to each other than the same regions of different species suggests that concerted evolution (DOVER *et al.* 1982) has been taking place and has led to intraspecific homogenization of sequences. This interpretation is consistent with the conclusion reached in an earlier publication describing the *C4* and *CYP21* genes themselves: Comparison of exonic and intronic human, chimpanzee, gorilla, and orang-

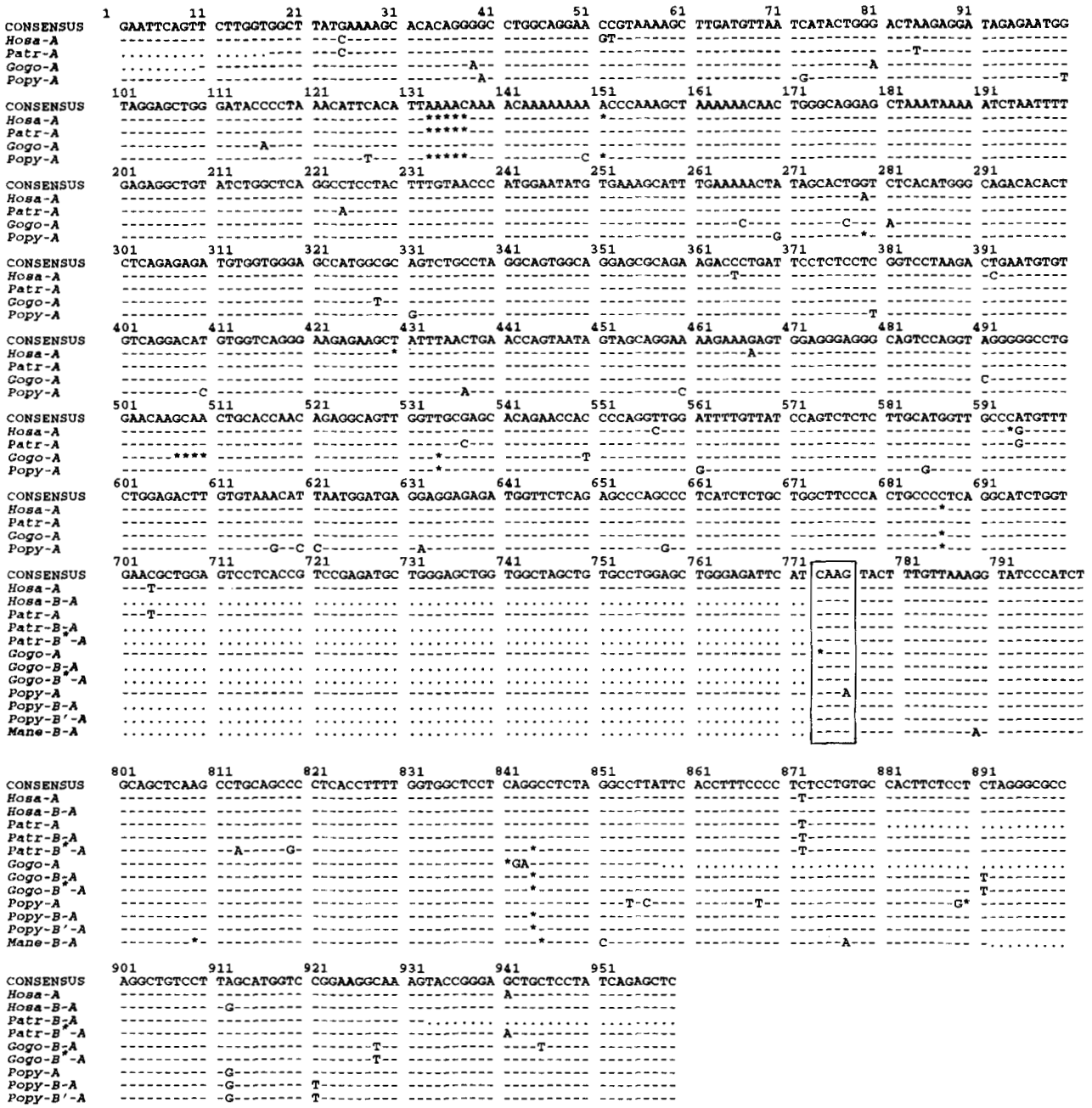

FIGURE 4.—Nucleotide sequence of the A and B-A regions of *C4-CYP21* in different primates: *Hosa*, human; *Patr*, chimpanzee; *Gogo*, gorilla; *Popy*, orangutan; *Mane*, pigtail macaque. Simple majority consensus appears at the top. Dashes (-) indicate identity with the consensus sequence, dots (.) undetermined sequence, asterisks (*) insertions/deletions, and N, sequence ambiguity. The critical tetranucleotide is boxed. The human sequence is from GITELMAN, BRISTOW and MILLER (1992). B*-A is a sequence allelic to B-A; B′-A is a sequence paralogous to B-A.

utan sequences revealed larger genetic distances between orthologous genes of different species than between paralogous genes of the same species (KAWAGUCHI, O'HUIGIN and KLEIN 1992; KAWAGUCHI *et al.* 1992). It appears, therefore, that the entire *C4-CYP21* chromosomal region is evolving in a concerted fashion

DISCUSSION

The answer to the questions that stimulated the present study is that the primigenial duplication probably occurred only once in the Catarrhini line which includes Old World monkeys, apes and humans. The duplication occurred before the separation of the Old World monkey and ape lines more than 23 my ago. Attempts to push this date of occurrence even further back in time have thus far failed for technical reasons: With the set of primers used in this study we have not been able to amplify by PCR the relevant segments of the *C4-CYP21* regions in the New World monkeys and prosimians (our unpublished data). Presumably, these regions have diverged too much from the human

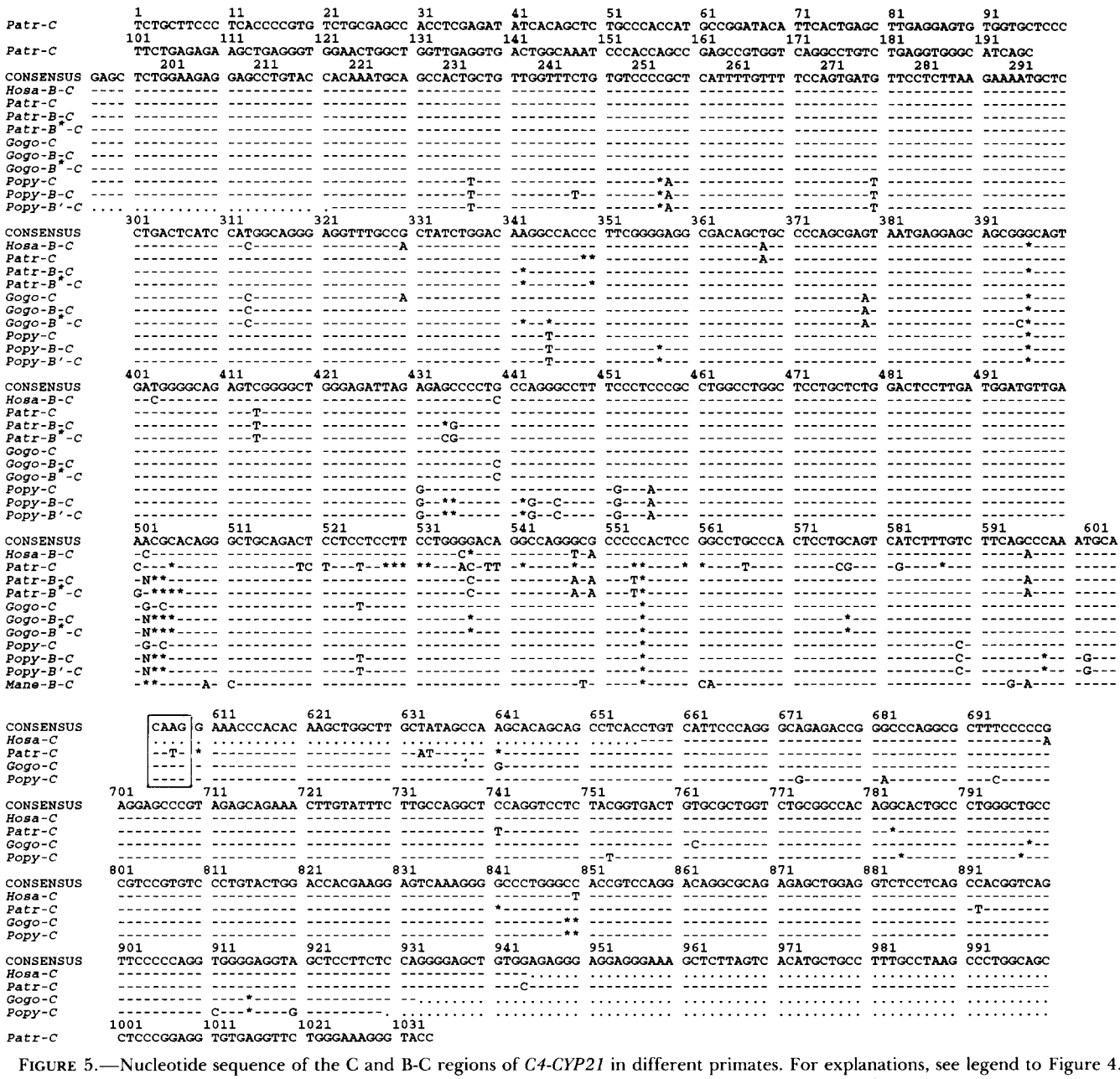

FIGURE 5.—Nucleotide sequence of the C and B-C regions of *C4-CYP21* in different primates. For explanations, see legend to Figure 4. The human sequence is from GITELMAN, BRISTOW and MILLER (1992). B*-C is a sequence allelic to B-C; B′-C is a sequence paralogous to B-C.

region for the primers to work. It may be necessary to construct genomic libraries from representatives of these two primate groups and sequence the relevant clones to find out how far back the primigenial duplication can actually be traced.

An alternative explanation positing repeated, independent duplications is unlikely. We assume that the duplication occurred by nonhomologous recombination. Although the mechanism of nonhomologous recombination is not known (MEUTH 1989; ROTH and WILSON 1986), it is believed to require extensive homology between the exchange partners and to occur at random sites in the genome. Because of the latter condition, the probability of nonhomologous recombination occurring at least four times at exactly the same site is extremely low. One could argue, however, that the duplications were in fact produced by homologous recombination between repetitive elements at the exchange site. However, no repetitive elements could be identified at this site. Furthermore, even if the CAAG tetranucleotide and its flanks were a recombinational hotspot, one would not expect individual, independent recombinations to produce identical breakpoints. In all the current models of homologous recombination, the initial nicking of the DNA is presumed to occur randomly in the aligned region and the subsequent steps of the recombinational process can be expected to introduce further

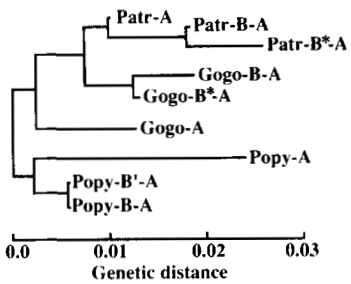

FIGURE 6.—Genetic distance dendrogram constructed from the nucleotide sequences of the A and B-A regions of *C4-CYP21* in different primates. *Patr*, chimpanzee; *Gogo*, gorilla; *Popy*, orangutan. B*-A is a sequence allelic to B-A. B'-A is a sequence paralogous to B-A.

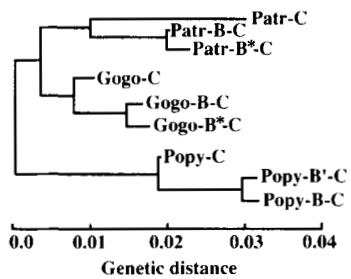

FIGURE 7.—Genetic distance dendrogram constructed from the nucleotide sequences of the C and B-C regions of *C4-CYP21* in different primates. *Patr*, chimpanzee; *Gogo*, gorilla; *Popy*, orangutan. B*-C is a sequence allelic to B-C. B'-C is a sequence paralogous to B-C.

variation in the resulting products so that even if the exchange occurs in the same region several times, the breakpoints do not fall in exactly the same site.

The existence of different primate haplotypes with varying numbers of *C4-CYP21* modules indicates that subsequent to the rare primigenial duplication, frequent secondary rounds of duplication have been occurring repeatedly. Others have documented that the secondary duplications occur by homologous but unequal crossing over between misaligned *C4-CYP21* modules (SINNOTT *et al.* 1990; and others). Unequal crossing over can contract a multimodular haplotype, leaving it monomodular, and then expand it to become multimodular again. The *C4-CYP21* is thus apparently undergoing repeated contractions and expansions with the consequence that the sequences are continually homogenized within each species. Unequal crossing over is probably the main mechanism of concerted evolution in the *C4-CYP21* region (KAWAGUCHI, O'HUIGIN and KLEIN 1991), but additional homogenization of sequences may be achieved by nonreciprocal recombination ("gene conversion"). It has been suggested that the frequency of gene conversion in mammalian cells decreases with increasing nucleotide disparity of homologous sequences (KOOP *et al.* 1989). In the *C4-CYP21* region, the frequent unequal crossing over may help retain high sequence similarity, which favors gene conversion-like processes; the two processes combined could be responsible for the striking intraspecific similarity between paralogous genes.

Why do the *C4* and *CYP21* genes evolve in a concerted fashion? It is difficult to come up with any reason for concerted evolution of the *CYP21* genes since when two or more of these genes are present on a single chromosome, only one of them is usually functional. Moreover, because of homogenization, the defects of the *CYP21* pseudogene are transferred relatively frequently by unequal crossing over or gene conversion onto the functional gene, and when this happens, individuals possessing only inactive *CYP21* genes develop congenital adrenal hyperplasia (CAH), which is a lethal condition (WHITE, NEW and DUPONT 1986). Thus homogenization of *CYP21* genes has fatal consequences and it is unlikely that this process is favored by natural selection. So, the concerted evolution of the *C4-CYP21* region in the Catarrhini (and very likely also in other mammals), must be occurring because of a selection pressure on the *C4* genes, with which the *CYP21* genes accidentally became locked into a permanent partnership by the primigenial duplication.

In contrast to *CYP21*, both *C4* loci on a bimodular chromosome are functional and individuals with complete C4 deficiency are extremely rare [less than 20 individuals of this type have been identified in the human population thus far; see HAUPTMANN *et al.* (1986)]. The complete absence of functional *C4* genes leads to severe systemic lupus erythematosus (SLE), an inflammatory autoimmune disorder which, because of trapping of antigen-antibody complexes in capillaries, may affect multiple organs. The condition, in contrast to CAH, need not be fatal. Partial C4 deficiency, on the other hand, is common in the human population, because some 40% of individuals have fewer than four functioning *C4* genes (MORGAN and WALPORT 1991). The clinical course of the partial C4 deficiency, however, may vary from no detectable symptoms at all to mild episodic disorders and increased susceptibility to SLE. The *C4* loci therefore tolerate homogenization much better than the *CYP21* loci, but this fact in itself does not explain why their homogenization occurs in the first place. We propose that the reason lies in the multifunctionality of the C4 molecule. During its activation and inactivation, the C4 molecule must interact with at least eight other molecules–the antigen, the antibody, complement factors C3, C2, I, C1, as well as the C4 and C4bp complement receptors (PORTER 1985). The molecules specified by the two human *C4* loci, *C4A* and *C4B*, differ in two of the eight interactions (those with the antigen and the antibody) and are identical in the remaining six. The C4A molecules have, in comparison to the C4B molecules, a higher propensity to form

amide bonds with amino groups, whereas the C4B molecules form preferentially ester bonds with hydroxyl groups. It has therefore been suggested that C4A is primarily involved in promoting the physiological disposal of immune complexes, whereas the hemolytically more active C4B molecule participates preferentially in clearance of microrganisms (LAW, DODDS and PORTER 1984). The functional difference between the C4A and C4B molecules resides in the C4d fragment, specifically in the amino acid position 1106 (LAW, DODDS and PORTER 1984; KAWAGUCHI *et al.* 1992); the six other interaction sites are localized in other parts of the C4 molecule. It seems, therefore, that C4 is under pressure to retain the C4A-C4B difference in the C4d fragment and identity in the rest of the molecule (divergence of the two *C4* genes outside the C4d region could lead to a failure in one or several of the six interactions). Indeed, sequence comparisons of the two *C4* genes in bimodular chromosomes of different primates indicate retention of the *C4A-C4B* difference in the different species and striking homogenization in the rest of the *C4* genes (KAWAGUCHI *et al.* 1992), as well as in the flanking regions (this communication). The reason for the concerted evolution of the *C4* genes could thus be the need to maintain identity of six interaction sites in the C4A and C4B molecules, while allowing functional differentiation of the molecules at two other sites.

The *C4-CYP21* region is part of the *Mhc* in all mammals thus far tested but its mode of evolution is not characteristic of the entire complex (KLEIN *et al.* 1993). The *DP* and *DQ* regions, which code for two different families of class II molecules, have remained remarkably stable since the separation of the ancestral Catarrhini and Platyrrhini at least, more than 37 my ago (GRAHOVAC *et al.* 1993; GAUR *et al.* 1992). On the other hand, the *DR* region, which codes for another family of class II proteins, displays a pronounced haplotype polymorphism passed on from species to species (KLEIN *et al.* 1991). The different primate *DR* haplotypes vary in length as well as in the number and composition of their genes. This region is thus marked by instability which seems to be favored by natural selection. Hence there seems to be different selection pressures influencing different regions of the *Mhc*, with the result that each region evolves in its own way.

We thank RYSZARD LORENZ for technical assistance, ANICA MILOSEV for the preparation of the graphics, and LYNNE YAKES for editorial help.

LITERATURE CITED

BITTER-SUERMANN, D., M. KRÖNKE, M. BRADE and U. HADDING, 1977 Inherited polymorphism of guinea pig factor B and C4: evidence for genetic linkage between the C4 and Bf loci. J. Immunol. **118:** 1822–1826.

BONTROP, R. E., L. A. M. BROOS, N. OTTING and M. J. JONKER, 1990 Polymorphism of *C4* and *CYP21* genes in various primate species. Tissue Antigens **37:** 145–151.

CARROLL, M. C., R. D. CAMPBELL and R. R. PORTER, 1985 Mapping of steroid hydroxylase genes adjacent to complement component C4 genes in HLA, the major histocompatibility complex in man. Proc. Natl. Acad. Sci. USA **82:** 521–525.

CHRISTIANSEN, F. T., R. E. BONTROP, M. GIPHART, P. U. CAMERON, W. J. ZHANG, D. TOWNSEND, M. JONKER and R. L. DAWKINS, 1991 Major histocompatibility complex haplotypes in the chimpanzee: identification using C4 allotyping. Hum. Immunol. **31:** 34–39.

CHUNG, B., K. J. MATTESON and W. L. MILLER, 1985 Cloning and characterization of the bovine gene for steroid 21-hydroxylase. DNA **4:** 211–219.

CHUNG, B. C., K. J. MATTESON and W. L. MILLER, 1986 Structure of a bovine gene for P-450c21 (steroid 21-hydroxylase) defines a novel cytochrome P-450 gene family. Proc. Natl. Acad. Sci USA **83:** 4243–4247.

DAVIS, L. G., M. D. DIBNER and J. F. BATTEY, 1986 *Basic Methods in Molecular Biology*. Elsevier, New York.

DE KROON, A. I. P. M., G. DOXIADIS, I. DOXIADIS and E. J. HENSEN, 1986 Structure and polymorphism of the feline complement component C4. Immunogenetics **24:** 202–205.

DOVER, G., S. BROWN, E. COEN, J. DALLAS, T. STRACHAN and M. TRICK, 1982 Dynamics of genome evolution and species differentiation, pp. 343–372 in *Genome Evolution*, edited by G. A. DOVER and R. B. FLAVELL. Academic Press, London.

DOXIADIS, G., V. REBMANN, I. DOXIADIS, K. KRUMBACHER, H. M. VRIESENDORP and H. GROSSE-WILDE, 1985 Polymorphism of the fourth complement component in the dog. Immunobiology **169:** 563–569.

DUNHAM, I., C. A. SARGENT, J. TROWSDALE and R. D. CAMPBELL, 1987 Molecular mapping of the human major histocompatibility complex by pulsed-field gel electrophoresis. Proc. Natl. Acad. Sci. USA **84:** 7237–7241.

GAUR, L. K., E. R. HEISE, P. S. THURTLE and G. T. NEPOM, 1992 Conservation of the HLA-DQB2 locus in nonhuman primates. J. Immunol. **148:** 943–948.

GITELMAN, S. E., J. BRISTOW and W. L. MILLER, 1992 Mechanism and consequences of the duplication on the human C4/P450c21/gene X locus. Mol. Cell. Biol. **12:** 2124–2131.

GRAHOVAC, B., U. BRÄNDLE, C. SCHÖNBACH, W. E. MAYER, F. FIGUEROA, J. TROWSDALE and J. KLEIN, 1993 Conservative evolution of the *Mhc-DP* region in anthropoid primates. Hum. Immunol. (in press).

HAUPTMANN, G., J. GOETZ, B. URING-LAMBERT and E. GROSSHANS, 1986 Complement deficiencies. 2. The fourth component. Progr. Allergy **39:** 232–249.

HOLMES D. S., and M. QUIGLEY, 1981 A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. **114:** 193–197.

KAWAGUCHI, H., and J. KLEIN, 1992 Organization of *C4* and *CYP21* loci in gorilla and orangutan. Hum. Immunol. **33:** 153–162.

KAWAGUCHI, H., C. O'HUIGIN and J. KLEIN, 1991 Evolution of primate *C4* and *CYP21* genes, pp. 357–382 in *Molecular Evolution of the Major Histocompatibility Complex*, edited by J. KLEIN and D. KLEIN. Springer-Verlag, Berlin.

KAWAGUCHI, H., C. O'HUIGIN and J. KLEIN, 1992 Evolutionary origins of mutations in the primate cytochrome P450c21 gene. Am. J. Hum. Genet. **50:** 766–780.

KAWAGUCHI, H., M. GOLUBIC, F. FIGUEROA and J. KLEIN, 1990 Organization of the chimpanzee *C4-CYP21* region: implications for the evolution of human genes. Eur. J. Immunol. **20:** 739–745.

KAWAGUCHI, H., Z. ZALESKA-RUTCZYNSKA, F. FIGUEROA, C. O'HUIGIN and J. KLEIN, 1992 *C4* genes of the chimpanzee,

gorilla, and orangutan: evidence for extensive homogenization. Immunogenetics **35:** 16–23.

KAY, P. H., and R. L. DAWKINS, 1984 Genetic polymorphism of complement C4 in the dog. Tissue Antigens **23:** 151–155.

KAY, P. H., R. L. DAWKINS, A. T. BOWLING and D. BERNOVO, 1987 Heterogeneity and linkage of equine C4 and steroid 21-hydroxylase genes. J. Immunogenet. **14:** 247–253.

KIMURA, M., 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:** 111–120.

KIRSZENBAUM, M., C. RENARD, C. GEFFROTIN, P. CHARDON and M. VAIMAN, 1985 Evidence for mapping pig *C4* gene(s) within the pig major histocompatibility complex (*SLA*). Anim. Blood Groups Biochem. Genet. **16:** 65–68.

KLEIN, J., C. O'hUIGIN, M. KASAHARA, V. VINCEK, D. KLEIN and F. FIGUEROA, 1991 Frozen haplotypes in *Mhc* evolution, pp. 261–286 in *Molecular Evolution of the Major Histocompatibility Complex*, edited by J. KLEIN and D. KLEIN. Springer-Verlag, Berlin.

KLEIN, J., C. O'hUIGIN, F. FIGUEROA, W. E. MAYER and D. KLEIN, 1993 Different modes of *Mhc* evolution in primates. Mol. Biol. Evol. **10:** 48–59.

KOOP, B. F., D. SIEMIENIAK, J. L. SLIGHTOM, M. GOODMAN, J. DUNBART, P. C. WRIGHT and E. L. SIMONS, 1989 Tarsius δ- and β-globin genes: conversions, evolution, and systematic implications. J. Biol. Chem. **264:** 68–79.

LAW, S. K. M., A. W. DODDS and R. R. PORTER, 1984 A comparison of the properties of two classes, C4A and C4B, of the human complement component C4. EMBO J. **3:** 1819–1823.

LÉVI-STRAUSS, M., M. TOSI, M. STEINMETZ, J. KLEIN and T. MEO, 1985 Multiple duplications of complement *C4* gene correlate with *H-2*-controlled testosterone-independent expression of its sex-limited isoform, C4-Slp. Proc. Natl. Acad. Sci. USA **82:** 1746–1750.

MANIATIS, T., E. F. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

MARTIN, R. D., 1990 *Primate Origins and Evolution: A Phylogenetic Reconstruction*. Chapman & Hall, London.

MATSUMOTO, K.-I., M. ARAI, N. ISHIHARA, A. ANDO, H. INOKO and T. IKEMURA, 1992 Cluster of fibronectin type III repeats found in the human major histocompatibility complex class III region shows the highest homology with the repeats in an extracellular matrix protein, tenascin. Genomics **12:** 485–491.

MAYER, W. E., M. JONKER, D. KLEIN, P. IVANYI, G. VAN SEVENTER and J. KLEIN, 1988 Nucleotide sequences of chimpanzee MHC class I alleles: evidence for trans-species mode of evolution. EMBO J. **7:** 2765–2774.

MCLEAN, R. H., P. A. DONOHOUE, N. JOSPE, W. B. BIASE, C. VAN DORP and C. J. MIGEON, 1988 Restriction fragment analysis of duplication of the fourth component of complement (C4A). Genomics **2:** 76–85.

MEUTH, M., 1989 Illegitimate recombination in mammalian cells, pp. 833–859 in *Mobile DNA*, edited by D. E. BERG AND M. H. HOWE. American Society for Microbiology, Washington D.C.

MEVAG, B., B. OLAISEN, P. TEISBERG and D. G. SMITH, 1983 Two *C4* loci in macaca monkeys. Immunobiology **164:** Abstract 1.

MILLER, W. L., 1988 Molecular biology of steroid hormone synthesis. Endocr. Rev. **9:** 295–318.

MOREL, Y., J. BRISTOW, S. E. GITELMAN and W. L. MILLER, 1989 Transcript encoded on the opposite strand of the human steroid 21-hydroxylase/complement component C4 gene locus. Proc. Natl. Acad. Sci. USA **86:** 6582–6586.

MORGAN, B. P., and M. J. WALPORT, 1991 Complement deficiency and disease. Immunol. Today **12:** 301–306.

PORTER, R. R., 1985 The polymorphism of the complement genes in HLA. Ann. Inst. Pasteur Immunol. **136C:** 91–101.

RASHEED, S. A., R. W. RONGEY, J. BRUSZWESKI, W. A. NELSON-REES, H. RABIN, R. H. NEUBAUER, G. ESRA and M. B. GARDNER, 1977 Establishment of a cell line with associated Epstein-Barr-like virus from a leukemic orangutan. Science **198:** 407–409.

ROOS, M. H., J. P. ATKINSON and D. C. SHREFFLER, 1978 Molecular characterization of the Ss and Slp (C4) proteins of the mouse *H-2* complex: subunit composition, chain-size polymorphism, and an intracellular (Pro-Ss) precursor. J. Immunol. **121:** 1106–1115.

ROSS, G. D., 1986 *Immunobiology of the Complement System. An Introduction for Research and Clinical Medicine*. Academic Press, Orlando.

ROTH, D. B., and J. H. WILSON, 1986 Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. Mol. Cell. Biol. **6:** 4295–4304.

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

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SINNOTT, P., S. COLLIER, C. COSTIGAN, P. A. DYER, R. HARRIS and T. STRACHAN, 1990 Genesis by meiotic unequal crossing-over of a *de novo* deletion that contributes to steroid 21-hydroxylase deficiency. Proc. Natl. Acad. Sci. USA **87:** 2107–2111.

SPILLIAERT, R., A. PALSDOTTIR and A. ARNASON, 1990 Analysis of the C4 genes in baleen whales using a human cDNA probe. Immunogenetics **32:** 73–76.

STEINMETZ, M., D. STEPHAN, G. R. DASTRONIKOO, E. GIBB and R. ROMANIUK, 1985 Methods in molecular immunology: chromosome walking in the major histocompatibility complex, pp. 1–19 in *Immunological Methods*, edited by J. LEFKOVITS and B. PERNIS. Academic Press, New York.

TOSI, M., M. LÉVI-STRAUSS, E. GEORGATSON, M. AMOR and T. MEO, 1985 Duplications of complement and noncomplement genes of the H-2S region: evolutionary aspects of the C4 isotypes and molecular analysis of their expression variants. Immunol. Rev. **87:** 141–183.

WHITE, P. C., M. I. NEW and B. DUPONT, 1986 Congenital adrenal hyperplasia. N. Engl. J. Med. **316:** 1519–1524.

YOSHIOKA, H., K. MOROHASHI, K. SOGAWA, M. YAMANE, S. KOMINAMI, S. TAKEMORI, Y. OKADA and Y. FUJI-KURIYAMA, 1986 Structural analysis of cloned cDNA for mRNA of microsomal cytochrome P-450 (c21) which catalyzes steroid 21-hydroxylation in bovine adrenal cortex. J. Biol. Chem. **261:** 4106–4109.

ZHANG, W. J., F. T. CHRISTIANSEN, X. WU, L. J. ABRAHAM, M. GIPHART and R. L. DAWKINS, 1993 Organization and evolution of *C4* and *CYP21* genes in primates. Importance of genomic blocks. Immunogenetics **37:** 170–176.

Communicating editor: W-H. LI