Am. J. Hum. Genet. 50:766–780, 1992

Evolutionary Origin of Mutations in the Primate Cytochrome P450c21 Gene

Hiroshi Kawaguchi,*,[1] Colm O'hUigin,* and Jan Klein*,†

*Max-Planck-Institut für Biologie, Abteilung Immungenetik, Tübingen; and †Department of Microbiology and Immunology, University of Miami School of Medicine, Miami

Summary

The *CYP21* gene codes for the enzyme cytochrome P450c21 (21-hydroxylase), which is critically involved in the synthesis of glucocorticoids and mineralocorticoids. Standard human haplotypes contain two copies of *CYP21*—a functional gene and a pseudogene. Inactivation of the functional gene leads to congenital adrenal hyperplasia (CAH). The pseudogene has three main defects: an 8-bp deletion in exon 3, a T insertion in exon 7, and a stop codon in exon 8. To determine the origin of these defects and to shed light on the evolution of the *CYP21* gene, we sequenced relevant segments of 10 primate *CYP21* genes—three from a chimpanzee, another three from a gorilla, and four from an orangutan. We could show that the 8-bp deletion is present in the chimpanzee and humans, while the other two defects are restricted to humans only. In the gorilla and the orangutan, however, extra *CYP21* copies are inactivated by other defects so that the number of functional copies is reduced in each species. Comparison of the sequences has revealed evidence for intraspecific homogenization (concerted evolution) of the *CYP21* genes, presumably through an expansion-contraction process effected by relatively frequent unequal but homologous crossing-over.

Introduction

The conversion of cholesterol to aldosterone or cortisol in the cortex of the human adrenal gland proceeds via a series of intermediate reactions catalyzed by six different enzymes, four of which have the cytochrome P450 characteristics (Miller 1988). A defect in any of these enzymes could theoretically lead to congenital adrenal hyperplasia (CAH) with or without salt-wasting syndrome (New and Levine 1984; Speiser and New 1985; White et al., 1987). Yet, when CAH is diagnosed in humans, in more than 90% of the cases it is caused by a defect of one particular enzyme, the cytochrome P450c21 (CYP21), or 21-hydroxylase. Moreover, CAH caused by defective CYP21 is rather common, occurring with a frequency of 1/5,000–1/20,000 persons, depending on the clinical criteria used for diagnosis, the population examined, and the type of survey procedure used (Speiser et al. 1985; Wallace et al. 1986). The reason for the high incidence of defective CYP21 seems to lie in the organization of the *CYP21* genes on chromosome 6, in the middle of the *HLA*, the human major histocompatibility complex (*Mhc;* Dupont et al. 1977; Levine et al. 1978). The two genes are arranged in tandem, with two *C4* genes coding for the fourth complement component, an arrangement that presumably arose by duplication of a 35-kb chromosomal segment carrying one *C4* and one *CYP21* gene (Carroll et al. 1985*a;* White et al. 1985; Dunham et al. 1987). One of the two *CYP21* elements (*CYP21P,* also referred to as *CYP21A*) is a pseudogene, and the other *CYP21* (*CYP21B*) is the functional gene. Since the two genes display high sequence similarity, they apparently misalign frequently and are subject to unequal crossing-over (Raum et al. 1984; Carroll et al., 1985*a,* 1985*b;* Donohoue et al. 1986; Rodrigues et al. 1987; Sinnott et al. 1990); they

Received August 9, 1991; final revision received December 5, 1991.

Address for correspondence and reprints: Jan Klein, Ph.D., Max-Planck-Institut für Biologie, Abteilung Immungenetik, Correnstrasse 42, 7400 Tübingen, Germany.

1. Present address: Department of Dermatology, Yokohama City University School of Medicine, Yokohama, Japan.

0002-9297/92/5004-0012$02.00

may also exchange segments by a gene conversion–like mechanism, although this point has not been established unambiguously (Donohoue et al. 1986; Harada et al. 1987; Jospe et al. 1987; Matteson et al. 1987; Amor et al. 1988; Higashi et al. 1988*a*). When these exchanges delete the functional gene or render it defective by a transfer of the deficient sequence from the pseudogene, the person becomes a carrier for CAH.

Three main defects render the *CYP21P* element a pseudogene: an 8-bp deletion in exon 3, a 1-bp substitution in codon 318 of exon 8, and a single-nucleotide (T) insertion in exon 7 (the *CYP21* gene has 10 exons; Higashi et al. 1986; White et al. 1986; Rodrigues et al. 1987). Both the deletion and the insertion destroy the correct reading frame, while the substitution produces a termination signal in the middle of the sequence. Transfer of any of these three defects from *CYP21P* to *CYP21* can cause CAH (Globerman et al. 1988; Higashi et al. 1988*b;* Urabe et al. 1990), although CAH can also result from deletions in the *CYP21* gene itself (White et al. 1988; Collier et al. 1989; Donohoue et al. 1989; Morel et al. 1989; Mornet et al. 1991). In addition to the three mutations and the deletions, which completely prevent protein synthesis, there are other deleterious mutations in the human *CYP21P* gene which adversely affect mRNA processing or enzymatic activity and cause CAH (Amor et al. 1988; Higashi et al. 1988*b,* 1991; Speiser et al. 1988; Chiou et al. 1990; Tusie-Luna et al. 1990, 1991). For example, a C-G nucleotide substitution at position 1655 in intron 2 leads to aberrant mRNA splicing (Higashi et al. 1988*b*), while three missense T-A mutations at positions 1380, 1383, and 1389 in exon 6 (Higashi et al. 1988*b*) and a G-T substitution at position 1683 in exon 7 (Speiser et al. 1988) impair enzyme activity.

It is not uninteresting to inquire how a situation leading to such a high frequency of congenital defects has arisen. We have therefore set out to answer questions concerning the emergence of three *CYP21P*-inactivating mutations and the origin of the two human *CYP21* genes, by studying the *CYP21* region in the nearest relatives of *Homo sapiens,* the great apes. In earlier publications (Kawaguchi et al. 1991; Kawaguchi and Klein, in press), we established that the overall organization of the *C4-CYP21* region in the chimpanzee and the gorilla is similar to that found in humans, with the genes arranged in the order *C4–CYP21–C4–CYP21* (fig. 1). In the orang-utan, on the other hand, we found three C4 and three CYP21 genes arranged in the order *C4–CYP21–C4–CYP21–C4–CYP21* (fig. 1). To determine the point at which the three main defects characterizing the human *CYP21P* gene emerged in the evolution of the three great apes and humans, we isolated all the *CYP21* genes we could detect in the chimpanzee, gorilla, and orangutan and sequenced the regions of the genes in which these defects are located. We sequenced enough DNA of each gene to reveal, by standard statistical methods of phylogenetic analysis, how the individual genes in the three ape species are related to each other and to the human *CYP21* genes.

Material and Methods

Source of DNA

Genomic DNA was isolated from the western lowland gorilla (*Gorilla gorilla*), the chimpanzee (*Pan troglodytes*), and the orangutan (*Pongo pygmaeus*). The source of the gorilla DNA was a fibroblast cell line, Sylvia, established from a skin biopsy sample by Dr. Kirby D. Smith (The Johns Hopkins University School of Medicine, Baltimore). The source of the chimpanzee DNA was the Epstein-Barr virus (EBV)–transformed B-cell line Hugo established from an animal maintained at the Dutch Primate Center at Rijswijk, The Netherlands (Mayer et al. 1988). The orangutan DNA was isolated from the cell line CP81, which was established from monocytic leukemia cells of a 13-year old female born in the wild but housed for 10 years at the Los Angeles Zoological Garden (Rasheed et al. 1977). The cell line was maintained at the Primate Center at Seattle and was supplied to us by Dr. Lakshmi Gaur (Wake Forest University, Bowman Gray School of Medicine, Winston-Salem).

Probes

The 2.0-kb *Bam*HI fragment specific for the human *CYP21* gene was purchased from Appligene (Illkirch, France). The 5′ GAGCAGAGCCCAACGACAGG 3′ oligonucleotide was complementary to the sequence flanking the 8-bp deletion in the human *CYP21* gene (Higashi et al. 1986). The *CYP21* probe was labeled using α^{32}P-CTP (Pharmacia oligolabeling kit, Uppsala, Sweden). Oligonucleotides were labeled using γ^{32}P-ATP and T4 polynucleotide kinase (Conner et al. 1983).

Construction and Screening of Cosmid Library

Genomic DNA was isolated from the indicated cell lines, according to the method of Maniatis et al.

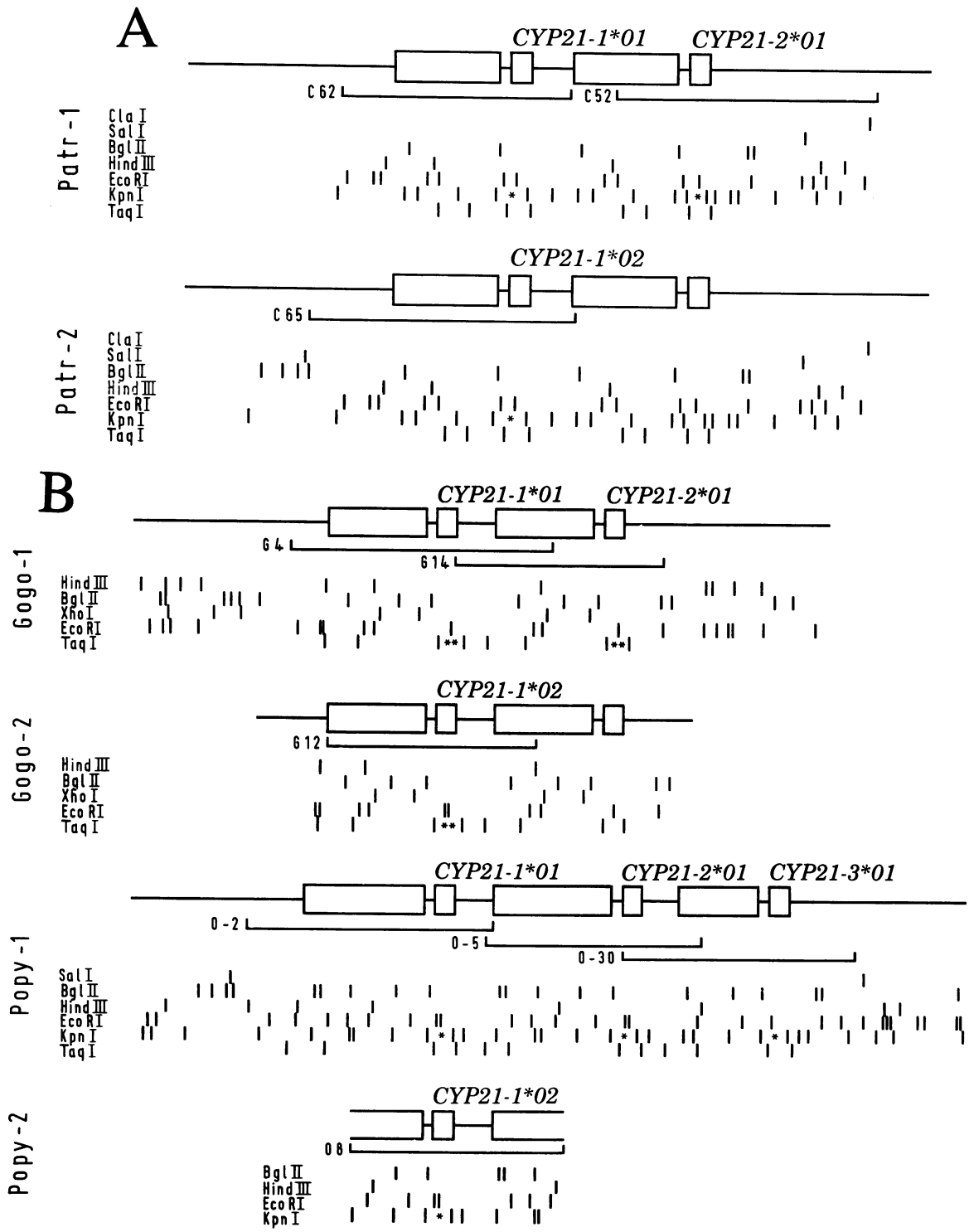

Figure 1 Organization and restriction maps of *C4-CYP21* regions in (*A*) chimpanzee (*Patr*) and (*B*) gorilla (*Gogo*) and orangutan (*Popy*). Restriction endonucleases are listed on the left, and their cutting sites are indicated by vertical lines. Of the second orangutan chromosome only one part is shown; the restriction map of the rest of this chromosome is apparently identical with that of the first chromosome. The segments designated C62, C52, C65, G4, G14, G12, 0-2, and 0-8 indicate the cosmid clones that were subcloned; the fragments of these subclones used for sequencing are *Kpn*I fragments, indicated single asterisks (*), or *Taq*I fragments, indicated by double asterisks (**). The maps were constructed from sets of overlapping cosmids. The scale is in kilobases. For details, see Kawaguchi et al. (1990) and Kawaguchi and Klein (in press).

(1982). Cosmid libraries were constructed and screened according to a method described by Steinmetz et al. (1985). Chimpanzee, gorilla, and orangutan libraries contained 1.5×10^6, 4.1×10^6, and 8.7×10^5 independent clones, respectively. A full description of the libraries is given elsewhere (Kawaguchi et al. 1990; Kawaguchi and Klein, in press). After hybridization, filters were washed with 0.1 × SSC, 0.1% SDS at 60°C.

Isolation and Analysis of Cosmid Clones

DNA was isolated from bacterial colonies according to the method of Maniatis et al. (1982) and was digested with restriction enzymes. Fragments were separated on agarose gel and were transferred to Hybond-N membranes (Amersham, Braunschweig, Germany). Some of the fragments were subcloned into pUC8 or pUC19 plasmid vectors according to the method of Davis et al. (1986).

Subcloning and Sequencing

DNA fragments obtained after digestion with appropriate restriction endonucleases were ligated to the pUC8 plasmid vector according to the procedure described by Davis et al. (1986). Either cosmid clone fragments or pUC8 subclones were ligated to the M13mp18 or M13mp19 phage vectors. The recombinant phage clones were picked up by the colony-hybridization method (Davis et al., 1986), and single-stranded DNA was prepared according to the protocol provided by Boehringer-Mannheim (Mannheim). One microgram of single-stranded DNA and the Sequenase version 2.0 kit (United States Biochemicals, Cleveland) were used for sequencing by the dideoxy chain-termination method (Sanger et al. 1977). Most fragments were sequenced on both strands and at least three times to eliminate or resolve possible sequencing errors and ambiguities.

Construction of Dendrograms

Both maximum-parsimony (Felsenstein 1988) and genetic distance (Saitou and Nei 1987) methods were used for evaluating evolutionary relationships between nucleotide sequences. Genetic distances were calculated by Kimura's two-parameter method (Kimura 1980) for intron sequences and by the Li-Wu-Luo method (Li et al. 1985) for exons. The estimated distances were used to draw trees by using the Saitou and Nei (1987) neighbor-joining method.

Results

To characterize the *CYP21* genes of the chimpanzee (cell line Hugo), the gorilla (cell line Sylvia), and the orangutan (cell line CP81), cosmid clones carrying these genes were isolated from the corresponding genomic libraries (see Kawaguchi and Klein, in press). The clones chosen for further characterization, as well as their restriction maps, are shown in figure 1. The cosmids were digested with either *Kpn*I or *Taq*I endonucleases, and fragments carrying *CYP21* genes were isolated (in fig. 1, the selected *Kpn*I and *Taq*I fragments are marked by a single asterisk (*) and double asterisks (**), respectively). The fragments were subcloned into the pUC8 vector, and the plasmids were then digested with *Sau*3A, *Pst*I, or *Eco*RI endonucleases. Fragments carrying the regions of interest were subcloned into the M13 vector and were sequenced. The position of these fragments is given in figure 2, which also lists the fragments sequenced in the individ-

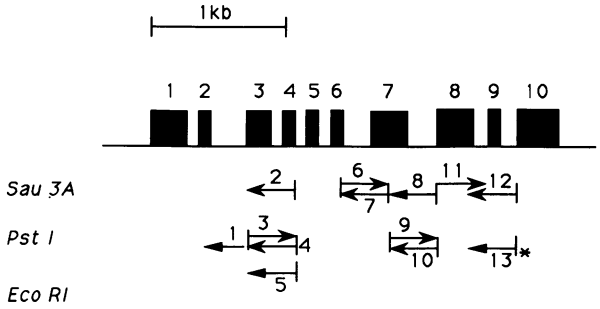

Figure 2 Strategy for sequencing primate *CYP21* genes. The exon-intron organization of a *Hosa-CYP21/H* gene (Higashi et al. (1986) is shown. Exons are indicated by blackened rectangles and numbers, introns by connecting lines. Vertical lines indicate restriction sites of the enzymes listed on the left-hand side; the arrows indicate the sequenced fragments and the direction of sequencing. The cloning sites are those of the *Patr-CYP21*01* gene (but similar sites were found in other genes as well), with the exception of the *Pst*I site marked by a single asterisk (*), which was limited to the *Popy-CYP2-3*01* gene only. Selected fragments were digested with the indicated enzymes and subcloned into M13mp18 or 19. The scale is in kilobases. The fragments used for sequencing of the individual genes were as follows: (numbers 1–13 refer to the arrows): *Patr-1*01*—2–5, 7, 8, 10, and 11; *Patr-1*02*—3, 7, 8, and 10–12; *Patr-2*01*—3 and 7–11; *Gogo-1*01*—1, 3, and 9–12; *Gogo1*02*—1, 3, 9, and 10; *Gogo-2*01*—2, 4, and 9–11; *Popy-1-*01*—2–4 and 8–12; *Popy-1*02*—2–4 and 8–11; *Popy-2*01*—2–4 and 8–11; *Popy-3*01*—3, 5, 8–11, and 13 (for simplicity, the *CYP21* designation is here omitted from the gene symbols). Segment 6 was sequenced but not included in the analysis, because of poor quality of the data.

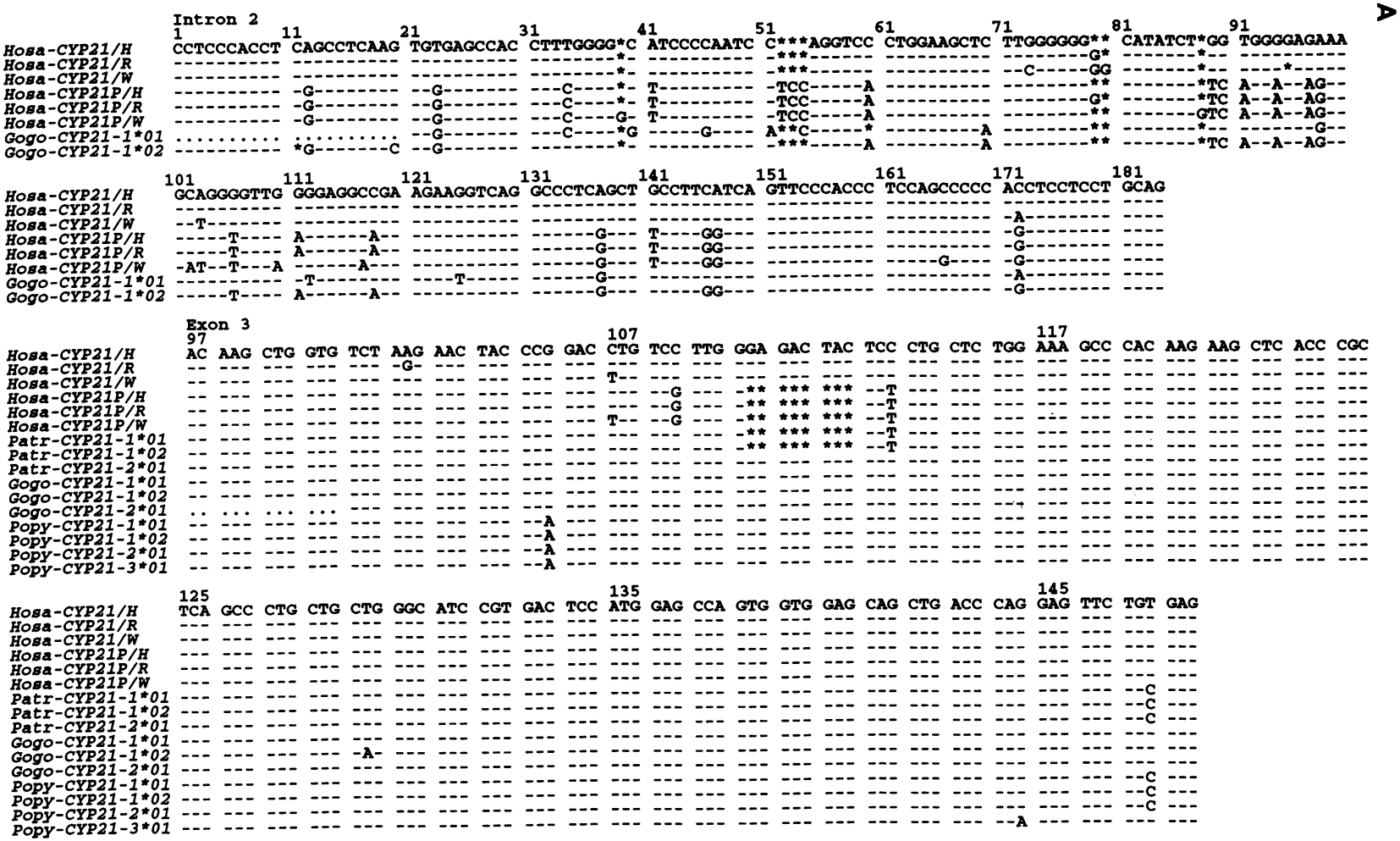

Figure 3 Nucleotide sequences of human (*Hosa*), chimpanzee (*Patr*), gorilla (*Gogo*), and orangutan (*Popy*) *CYP21* genes. The exons/introns which the sequences cover are indicated. Sites in each exon are numbered separately; codons in exons are numbered contiguously using the system of Higashi et al. (1986). The human sequences are from the following sources: *H*, Higashi et al. (1986); *R*, Rodrigues et al. (1987); and *W*, White et al. (1986). A dash (–) indicates identity with the top sequence; a dot (.) indicates unavailability of sequence information; and an asterisk (*) indicates a gap in the sequence.

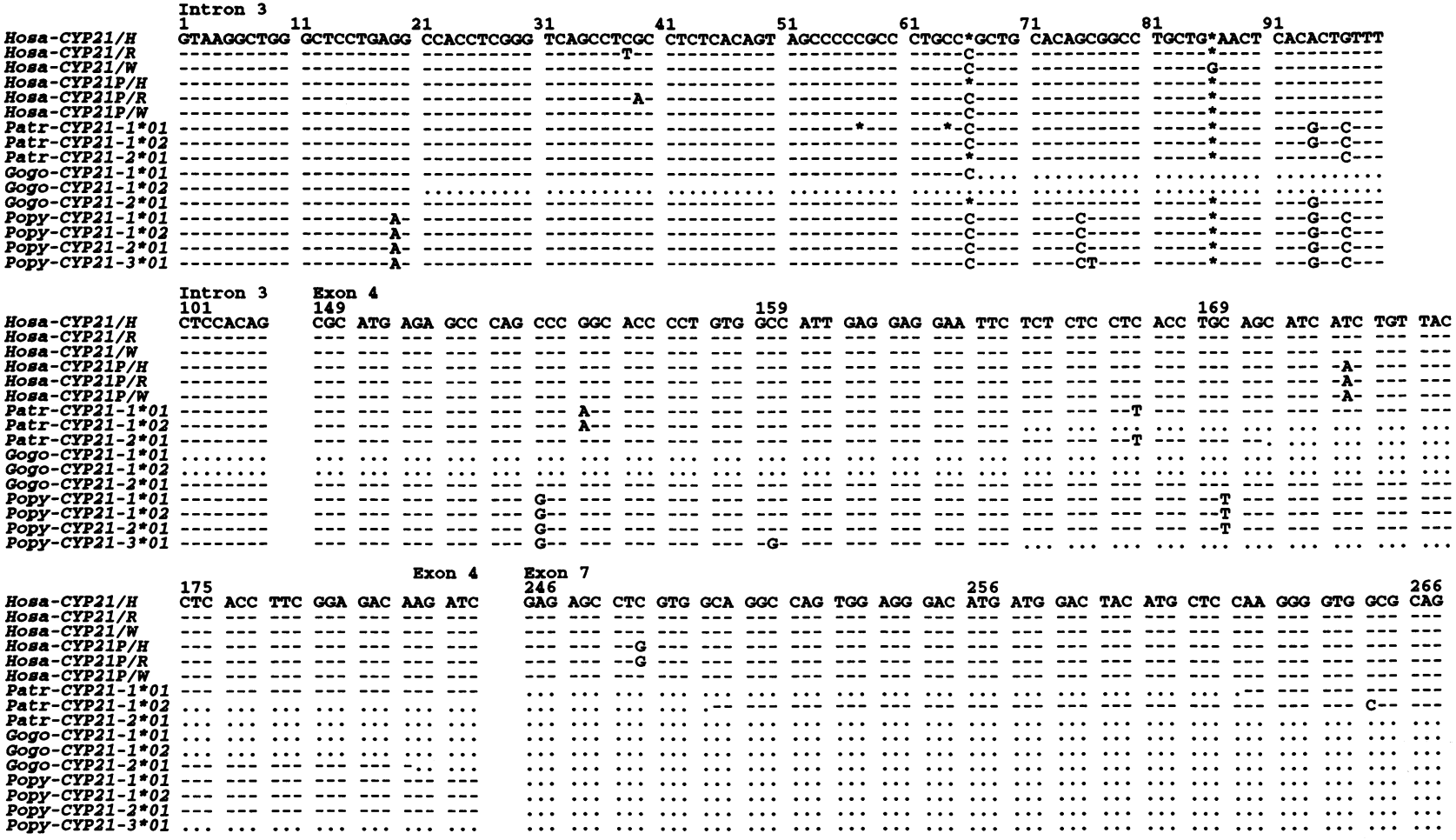

Figure 3 (continued)

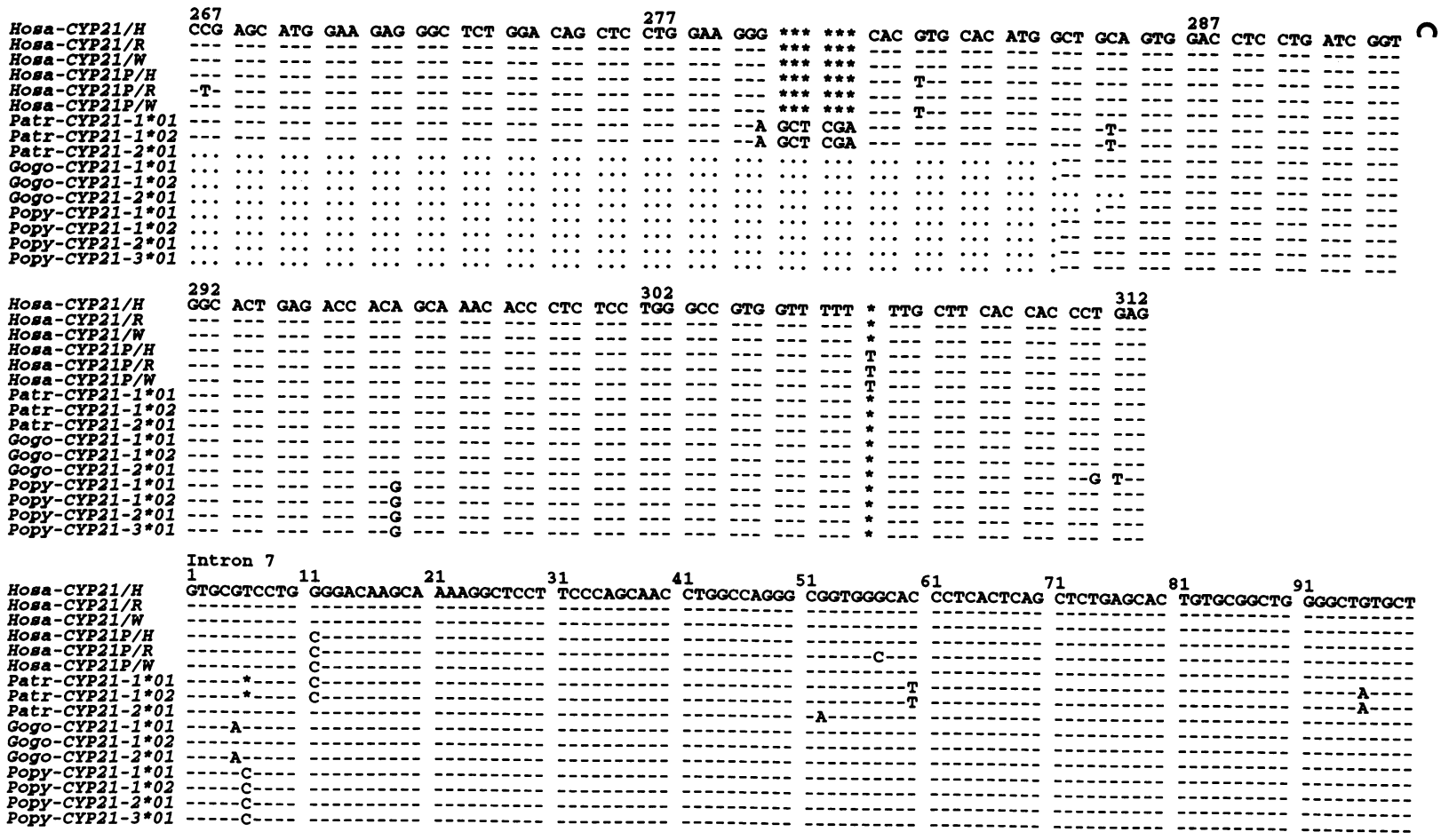

Figure 3 (continued)

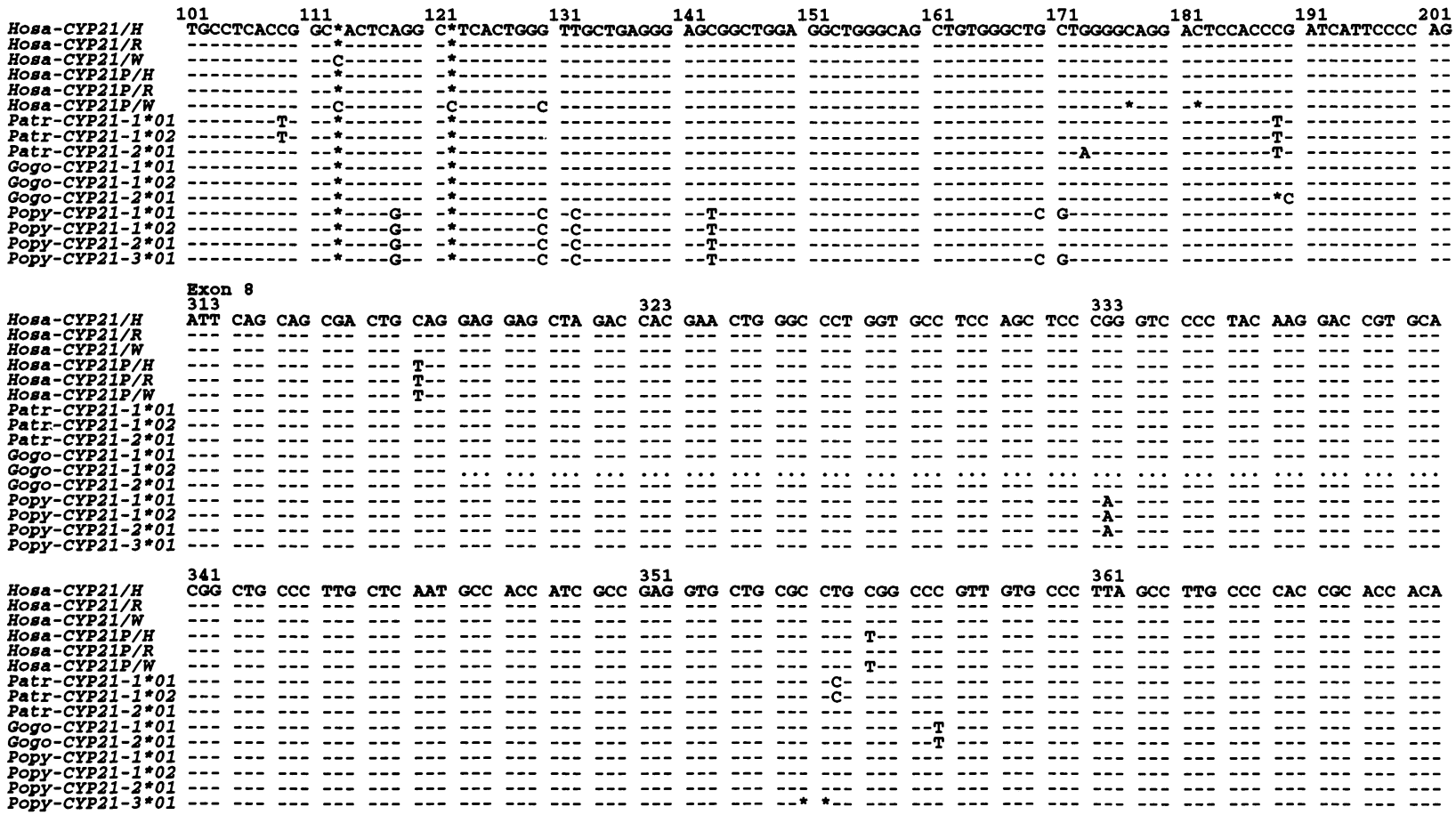

Figure 3 (continued)

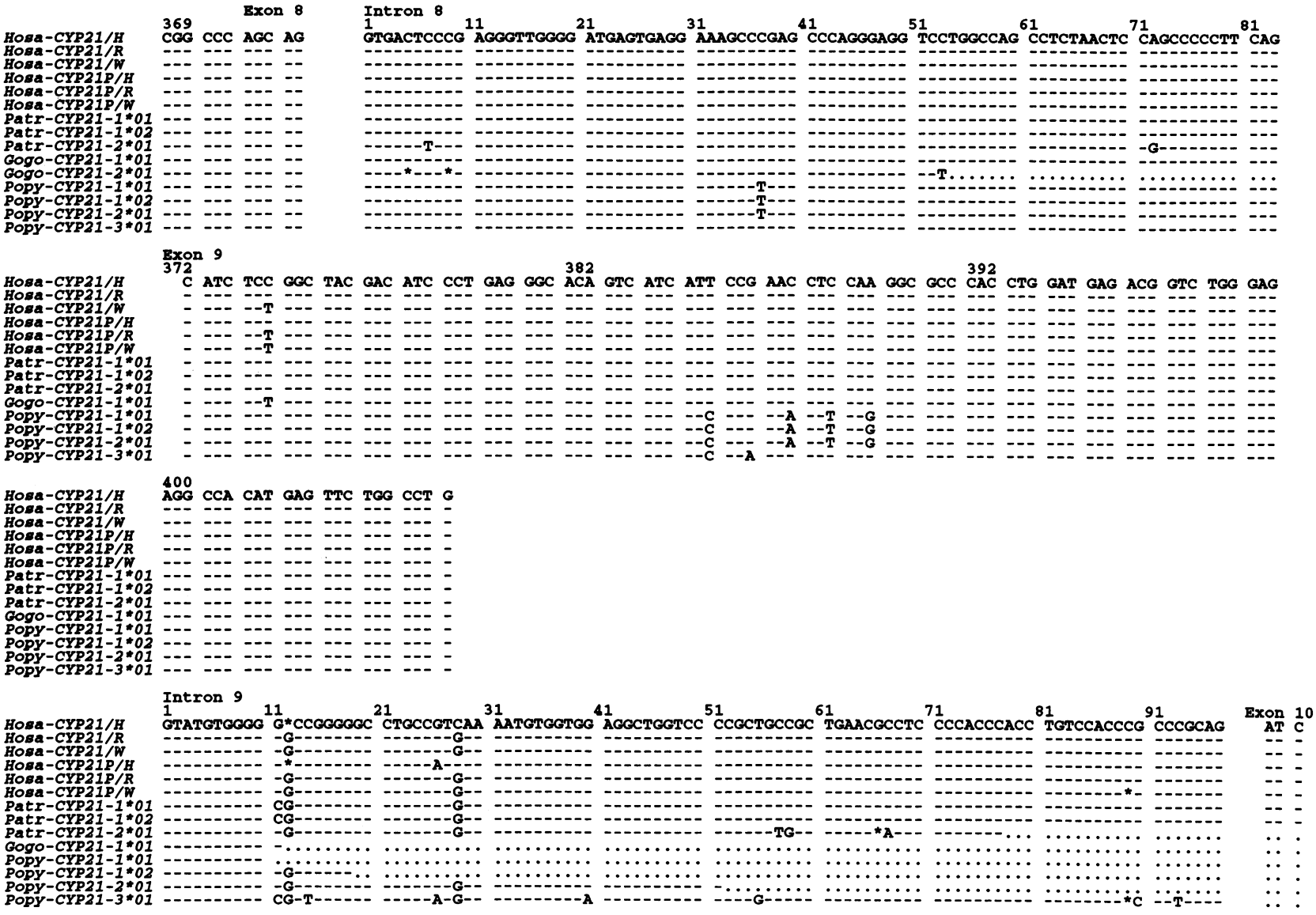

Figure 3 (continued)

ual genes. The choice of the regions for sequencing was dictated by the distribution of the three main defects in the human *CYP21P* gene. Enough sequence was generated for meaningful statistical comparisons of the individual genes; further sequencing would not have increased significantly the amount of information we were interested in and hence was deemed unnecessary.

The nucleotide sequence, obtained together with previously published sequences of human *CYP21* genes, is given in figure 3; a summary of amino acid differences deduced from the nucleotide sequence appears in figure 4. The nucleotide sequence covers most of intron 2, complete exons 3, 4, and 7–9, and complete introns 3, 4, and 7–9, though (for technical reasons) not for all the fragments. Where homologous chromosomes were shown to carry different alleles, these were sequenced, too. Altogether, we partially sequenced three chimpanzee, three gorilla, and four orangutan genes. The longest stretch sequenced from a single gene encompassed 1,421 sites.

To differentiate the sequences, we adopt here the nomenclature used for primate class I and class II *Mhc* genes (Klein et al. 1990). The loci are designated by the *CYP21* symbol prefixed by the first two letters of the scientific genus and species names. Hence the chimpanzee, gorilla, and orangutan loci are designated *Patr-CYP21* (for *Pan troglodytes*), *Gogo-CYP21* (for *Gorilla gorilla*), and *Popy-CYP21* (for *Pongo pygmaeus*), respectively. These symbols are followed, after a hyphen, by the serial locus number and, after an asterisk, by the serial allele number (by "serial" we mean that the loci or alleles are designated in the order of their description). The designations for the 10 partially sequenced genes are as follows: *Patr-CYP21-1*01, Patr-CYP21-1*02, Patr-CYP21-2*01, Gogo-CYP21-1*01, Gogo-CYP21-1*02, Gogo-CYP21-2*01, Popy-CYP21-1*01, Popy-CYP21-1*02, Popy- CYP21-2*01,* and *Popy-CYP21-3*01*. We designate the two human loci *Hosa-CYP21P* and *Hosa-CYP21* (for *Homo sapiens*), which is the nomenclature recommended by the Human Genomic Mapping Conferences (McAlpine et al. 1990), rather than CYP21A and CYP21B, respectively, which is a nomenclature frequently used by other authors. The human sequences reported by different investigators are distinguished by added letters referring to the author's name. The decision as to which of the genes are allelic is based on the consideration of both the locus they occupy and the genetic distance between them.

The relationships among the sequenced genes were evaluated by dendrogram analysis. There are two widely used methods—the distance-matrix and the maximum-parsimony methods (Nei 1987)—of constructing phylogenetic trees (dendrograms) which presumably reflect the evolution of the genes under study. In the neighbor-joining variant of the distance-matrix method (Saitou and Nei 1987), the number of nucleotide substitutions per nucleotide site (genetic distance) is calculated for each pair of compared sequences, and the pairs are then grouped so as to minimize the total length of the tree. The principal of the maximum-parsimony methods is to infer the sequences of the ancestral genes and choose a tree that requires the minimum number of substitutions.

The application of the maximum-parsimony method to our sequence data resulted in 20 most parsimonious trees in all of which the four orangutan sequences formed one cluster, the three chimpanzee sequences a second cluster, and the gorilla and human sequences a third cluster, with the trees differing in the topologies within the clusters, particularly in the

	102	110-2	129	154-5	159	172	265-7	280-1	285	312	318	333	354-6	387
Hosa-CYP21/H	K	GDY	L	PG	A	I	AQP	**HV	A	E	Q	R	RLR	N
Hosa-CYP21/R	R	GDY	L	PG	A	I	AQP	**HV	A	E	Q	R	RLR	N
Hosa-CYP21/W	K	GDY	L	PG	A	I	AQP	**HV	A	E	Q	R	RLR	N
Hosa-CYP21P/H	K	***	L	PG	A	N	AQP	**HL	A	E	!	R	RLW	N
Hosa-CYP21P/R	K	***	L	PG	A	N	AQL	**HV	A	E	!	R	RLR	N
Hosa-CYP21P/W	K	***	L	PG	A	N	AQP	**HL	A	E	!	R	RLW	N
*Patr-CYP21-1*01*	K	***	L	PG	S	I	AQP	ARHV	V	E	Q	R	RPR	N
*Patr-CYP21-1*02*	K	***	L	PG	S	.	PQP	ARHV	V	E	Q	R	RPR	N
*Patr-CYP21-2*01*	K	GDY	L	PG	A	.	...	....	A	E	Q	R	RLR	N
*Gogo-CYP21-1*01*	K	GDY	L	..	.	.	...	....	A	E	Q	R	RLR	N
*Gogo-CYP21-1*02*	K	GDY	Q	..	.	.	...	....	.	E	Q	.	...	.
*Gogo-CYP21-2*01*	K	GDY	L	PG	A	I	...	....	.	!	Q	R	RLR	K
*Popy-CYP21-1*01*	K	GDY	L	AG	A	I	...	....	A	E	Q	Q	RLR	K
*Popy-CYP21-1*02*	K	GDY	L	AG	A	I	...	....	A	E	Q	Q	RLR	K
*Popy-CYP21-2*01*	K	GDY	L	AG	A	I	...	....	A	E	Q	Q	RLR	N
*Popy-CYP21-3*01*	K	GDY	L	AG	G	.	...	....	A	E	Q	R	**R	N

Figure 4 Amino acid sequence of CYP21 proteins that is deduced fron nucleotide sequences in fig. 2. Only positions at which differences in the listed sequences were found are given. A dot (.) indictes unavailability of sequence information; an asterisk (*) indicates a gap in the sequence; and an exlamation mark (!) indicates an in-frame stop codon. Where deletions destroy the reading frame of the nucleotide sequence, the amino acid residues given are those that would appear were the sequence read in the correct reading frame.

gorilla-human cluster (data not shown). As we were unable to choose among these dendrograms, we resorted to genetic-distance analysis and constructed a dendrogram by the neighbor-joining method (fig. 5).

Of the 10 partially sequenced genes, only two of the chimpanzee genes (*Patr-CYP21-1*01* and **02*) contained the 8-bp deletion characterizing the human *CYP21* pseudogene; none of the sequences had either of the two other *Hosa-CYP21P* defects—the T insertion in exon 7 and the stop codon in exon 8. The C-G substitution in intron 2, which in the human *CYP21P* gene leads to a splicing defect (Higashi et al., 1988*b*), is present in the gorilla gene (*Gogo-CYP21-1*02*. Hence this gene, too, is presumably inactive. None of the other defects characterizing the human *CYP21P* gene were found in the ape sequences.

The *Patr-CYP21-1*01* and **02* genes are apparently alleles, because they occupy corresponding positions on homologous chromosomes, they are closely related in their sequence, and they share the 8-bp deletion. Since the *Hosa-CYP21P* and the *Patr-CYP21-1* genes differ on average at 25 of the 1,160 sequenced sites, the pseudogene accumulates mutations at a rate of $1.7/\text{site}/10^9$ years, so that the two *Patr-CYP21-1* alleles presumably diverged less than 1 million years (Myr) ago. This estimate is, however, based on the assumption that no homogenization of sequences via unequal crossing-over or some other, similar mechanism has occurred, which may not be the case (for discussion, see Kawaguchi et al. 1991). The third chimpanzee gene, *Patr-CYP21-2*01*, has no obvious defect in the sequenced part and is presumably functional. Its counterpart on the homologous chromosome is indistinguishable from it, at least by restriction-enzyme analysis (Kawaguchi et al. 1990). Because of its position, the *Patr-CYP21-2*01* gene could have been expected to show more sequence similarity to *Hosa-CYP21* than to *Patr-CYP21-1*, but in fact the opposite is true.

The *Gogo-CYP21-2*01* has a stop codon in exon

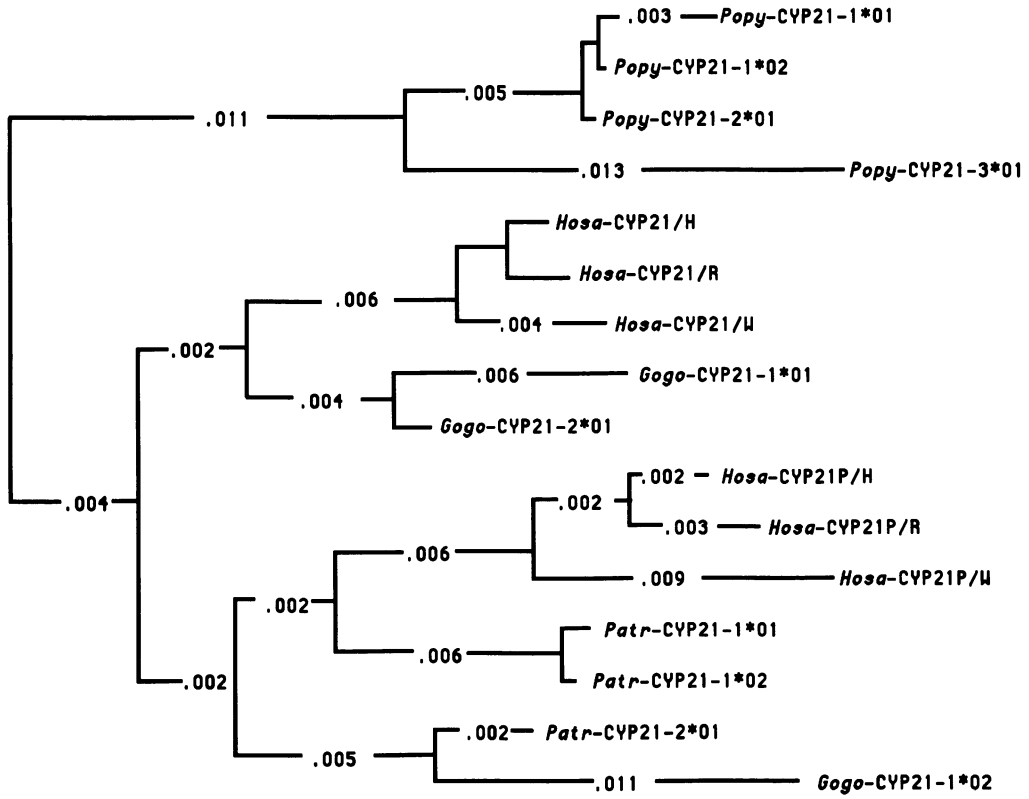

Figure 5 Dendrogram depicting presumed phylogenetic relationships among primate *CYP21* genes for which sequences are given in fig. 3. The tree is based on genetic distances (numbers on individual lines) calculated as indicated in the Material and Methods section and was constructed by the neighbor-joining method. Abbreviations and references are as in fig. 3.

7 at position 312, and we assume therefore that it is a pseudogene. The substitution generating this stop codon is not found in any of the other known *CYP21* sequences and hence represents a new way of silencing one of the cytochrome P450c21 genes. The *Gogo-CYP21-1*02* gene has, as already mentioned, a splicing defect in intron 2, and hence it, too, is presumably nonfunctional. Although the *Gogo-CYP21-1*01* and **02* are alleles by their position, the former sequence is in fact more similar to *Gogo-CYP21-2*01* than to *Gogo-CYP21-1*02*. Furthermore, the *Gogo-CYP21-1*01* and *2*01* genes cluster with the human functional genes, whereas the *Gogo-CYP21*02* gene clusters with the human and chimpanzee pseudogenes. These relationships are particularly evident in intron 2, where *Gogo-CYP21-1*02* shares several substitutions (some of them clustered) with the *Hosa-CYP21P* genes. Hence the order of the functional and nonfunctional genes is reversed in one gorilla haplotype, in comparison with the order in the standard human haplotype. The unusually high variability of intron 2 in comparison with the rest of the determined sequence remains unexplained.

Of the four partially sequenced orangutan *CYP21* genes, two (*Popy-CYP21-1*01* and *1*02*) are classified here as alleles, by virtue of their occupying what appears to be the same locus (Kawaguchi and Klein, in press). The differences between the *Popy-CYP21* genes are quite small, with the exception of *Popy-CYP21-3*01,* which is genetically somewhat more distant from the remaining three genes. The four *Popy* genes are, however, quite distant from the *Hosa, Patr,* and *Gogo* genes. This fact is reflected both in their forming a separate cluster in the dendrogram (fig. 5) and in the existence of several species-specific substitutions scattered along the entire *Popy* sequences.

Discussion

The primary purpose of this study has been to determine the evolutionary origins of the three defects characterizing the human *CYP21P* gene. The study shows that the 8-bp deletion in exon 3 is present in the chimpanzee but not in the gorilla or orangutan genes, whereas the T insertion in exon 7 and the substitution generating the stop codon in exon 8 are restricted to human genes. These findings indicate that the 8-bp deletion was first of the three defects to occur. The chimpanzee, gorilla, and human evolutionary lineages are believed to have separated from each other approximately 6 Myr ago (Martin 1990). Although the manner of separation of the three lineages is still being debated by the experts (Holmes et al. 1989), the majority seems to believe that the gorilla lineage separated first and that subsequently the chimpanzee and human lineages separated from each other. The separation of all three lineages, however, occurred within a relatively short time, probably within less than 0.5 Myr. Our results are consistent with this scenario: the 8-bp deletion apparently occurred after the gorilla lineage split off but before the chimpanzee and human lineages separated from each other. We can thus date the occurrence of the 8-bp deletion rather precisely within a relatively short period of some 6 Myr ago. The deletion was followed, in the human lineage, by the two other defective mutations. We cannot entirely exclude the possibility that the 8-bp deletion will eventually be found in the gorilla as well, when more animals have been tested. We have used, however, oligonucleotide probes specific for the deletion, to test additional DNA samples, and we have obtained evidence for the presence of the deletion in other chimpanzees but not in any of the gorillas or orangutans tested (Kawaguchi et al. 1990; Kawaguchi and Klein, in press). The 8-bp deletion may, however, not have been the first defect inactivating the ancestral *CYP21P* gene. The presence of the splicing defect in intron 2 of one of the gorilla genes suggests that this mutation occurred before the gorilla lineage diverged from the ancestral human-chimpanzee lineage. If so, this defect should also be present in some of the chimpanzee genes, and the sequence of events in the ancestral *CYP21P* gene could have been this: splicing defect by substitution in intron 2 more than 6 Myr ago, 8-bp deletion in exon 3 approximately 6 Myr ago, and T insertion in exon 7, as well as stop codon–generating mutation in exon 8, less than 6 Myr ago. All these events presumably occurred in the branch of the tree marked by the *Hosa-CYP21P* gene (fig. 5).

In this context, one might ask how representative are the data generated by the study of a single cell line from each primate species. We have addressed this issue in detail elsewhere (Kawaguchi and Klein, in press); here it suffices to say that genomic DNA isolated from peripheral blood lymphocytes of multiple individuals of each of the three species displays the same or similar RFLP as that isolated from the cell lines. It is therefore likely that no gross rearrangements occurred in the *C4-CYP21* regions of the cell lines used and that the gene organization found in these lines is representative of each of the studied species.

At least some of the relationships among the studied

primate *CYP21* genes can be explained by postulating expansion and contraction of the *C4-CYP21* region. Thus, although the *Gogo-CYP21-1*01* and **02* genes occupy corresponding positions on the homologous chromosomes, the former is genetically more closely related to *Gogo-CYP21-2*01* (which occupies a different locus) than to *Gogo-CYP21-1*02*. We suggest, therefore, that *Gogo-CYP21-2*01* arose from *Gogo-CYP21-1*01* by recent duplication and that the *Gogo-CYP21-1*01* and *1*02* genes diverged before this event. Similarly, the fact that all the orangutan sequences group together on a single branch of the phylogenetic tree probably means that the genes expanded after contraction to a single locus and that the event occurred after the orangutan lineage separated from the common ancestor of humans, chimpanzee, and gorilla. As argued elsewhere (Kawaguchi et al. 1991), *de novo* duplication (in contrast to expansion of contracted genes) is an unlikely explanation for the outgrouping of the *Popy* sequences.

Frequent contraction and expansion of the *C4-CYP21* region is also indicated by the fact that the haplotypes with one or three pairs of *C4* and *CYP21* genes are quite common in human populations, especially in some ethnic groups (reviewed by Collier et al. 1989). Similarly, the fact that the two *CYP21* (*C4*) genes commonly found in the mouse (Chaplin et al. 1986) are genetically equidistant to the two human *CYP21* (*C4*) genes most likely means that in both species the number of *CYP21* (*C4*) genes in a single haplotype contracted, at some stage of evolution, to one and then expanded to two again. The consequence of the contraction and expansion is that the genes in each species are homogenized; that is to say, they resemble each other more closely than any one of them resembles any particular gene in another species (= concerted evolution). The most likely mechanism responsible for the contraction and expansion is unequal crossing-over.

Acknowledgments

We thank Lynne Yakes for editorial assistance. The work was supported in part by a grant from the Fonds der Chemischen Industrie, Frankfurt.

References

Amor M, Parker KL, Globerman H, New MI, White PC (1988) Mutation in the *CYP21B* gene (Ile-172→Asn) causes steroid 21-hydroxylase deficiency. Proc Natl Acad Sci USA 85:1600–1604

Carroll MC, Campbell RD, Porter RR (1985*a*) Mapping of steroid 21-hydroxylase genes adjacent to complement component C4 genes in HLA, the major histocompatibility complex in man. Proc Natl Acad Sci USA 82:521–525

Carroll MC, Palsdottir A, Belt KT, Porter RR (1985*b*) Deletion of complement C4 and steroid 21-hydroxylase genes in the HLA class III region. EMBO J 4:2547–2552

Chaplin DD, Galbraith LJ, Seidman JG, White PC, Parker KL (1986) Nucleotide sequence analysis of murine 21-hydroxylase genes: mutations affecting gene expression. Proc Natl Acad Sci USA 83:9601–9605

Chiou S-H, Hu M-C, Chung B-C (1990) A missense mutation at Ile^{172}→Asn or Arg^{356}→Trp causes steroid 21-hydroxylase deficiency. J Biol Chem 265:3549–3552

Collier S, Sinnott PJ, Dyer PA, Price DA, Harris R, Strachan T (1989) Pulsed field gel electrophoresis identifies a high degree of variability in the number of tandem 21-hydroxylase and complement *C4* gene repeats in the 21-hydroxylase deficiency haplotypes. EMBO J 8:1393–1402

Conner BJ, Reyes AA, Morin C, Itakura K, Teplitz RL, Wallace RB (1983) Detection of sickle cell β^S-globin allele by hybridization with synthetic oligonucleotides. Proc Natl Acad Sci USA 80:278–282

Davis LG, Dibner MD, Battey JF (1986) Basic methods in molecular biology. Elsevier, New York

Donohoue PA, Jospe N, Migeon CJ, Van Dorp C (1989) Two distinct areas of unequal crossing-over within the steroid 21-hydroxylase genes produce absence of *CYP21B*. Genomics 5:397–406

Donohoue PA, van Dop C, McLean RH, White PC, Jospe N, Migeon CJ (1986) Gene conversion in salt-losing congenital adrenal hyperplasia with absent complement C4B protein. J Clin Endocrinol Metab 62:995–1002

Dunham I, Sargent CA, Trowsdale J, Campbell RD (1987) Molecular mapping of the human major histocompatibility complex by pulsed-field gel electrophoresis. Proc Natl Acad Sci USA 84:7237–7241

Dupont B, Oberfield SE, Smithwick EM, Lee TD, Levine LS (1977) Close genetic linkage between HLA and congenital adrenal hyperplasia (21-hydroxylase deficiency). Lancet 2:1309–1312

Felsenstein J (1988) Phylogenies from molecular sequence: inferences and reliability. Annu Rev Genet 22:521–565

Globerman H, Amor M, Parker KL, New MI, White PC (1988) Nonsense mutation causing steroid 21-hydroxylase deficiency. J Clin Invest 82:139–144

Harada F, Kimura A, Iwanga T, Shimozawa K, Yata J, Sasazuki T (1987) Gene conversion-like events cause steroid 21-hydroxylase deficiency in congenital adrenal hyperplasia. Proc Natl Acad Sci USA 84:8091–8094

Higashi Y, Hiromasa T, Tanae A, Miki T, Nakura J, Kondo T, Ohura T, et al (1991) Effects of individual mutations in the *P-450(C21)* pseudogene on the P-450 (C21) activity and their distribution in the patient genomes of congenital steroid 21-hydroxylase deficiency. J Biochem 109:638–644

Higashi Y, Tanae A, Inoue H, Fujii-Kuriyama Y (1988*a*) Evidence for frequent gene conversion in the steroid 21-hydroxylase P-450 (C21) gene: implications for steroid 21-hydroxylase deficiency. Am J Hum Genet 42:17–25

Higashi Y, Tanae A, Inoue H, Hiromasa T, Fujii-Kuriyama Y (1988*b*) Aberrant splicing and missense mutations cause steroid-hydroxylase [P450(C21)] deficiency in humans: possible gene conversion products. Proc Natl Acad Sci USA 85:7486–7490

Higashi Y, Yoshioka H, Yanae M, Gotch O, Fujii-Kuriyama Y (1986) Complete nucleotide sequence of two steroid 21-hydroxylase genes tandemly arranged in human chromosome: a pseudogene and a genuine gene. Proc Natl Acad Sci USA 83:2841–2845

Holmes EC, Pesole G, Saccone C (1989) Stochastic models of molecular evolution and the estimation of phylogeny and rates of nucleotide substitution in the hominoid primates. J Hum Evol 18:775–794

Jospe N, Donohoue PA, van Dop C, McLean RH, Bias W, Migeon CJ (1987) Prevalence of polymorphic 21-hydroxylase gene (CA21HB) mutations in salt-losing congenital adrenal hyperplasia. Biochem Biophys Res Commun 142: 798–804

Kawaguchi H, Golubic M, Figueroa F, Klein J (1990) Organization of the chimpanzee *C4-CYP21* region: implications for the evolution of human genes. Eur J Immunol 20:739–745

Kawaguchi H, Klein J. Organization of *C4* and *CYP21* loci in gorilla and orangutan. Hum Immunol (in press)

Kawaguchi H, O'hUigin C, Klein J (1991) Evolution of primate *C4* and *CYP21* genes. In: Klein J, Klein D (eds) Molecular evolution of the major histocompatibility complex. Springer, Heidelberg, pp 357–382

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

Klein J, Bontrop RE, Dawkins RL, Erlich HA, Gyllensten UB, Heise ER, Jones PP, et al (1990) Nomenclature for the major histocompatibility complexes of different species: a proposal. Immunogenetics 31:217–219

Levine LS, Zachman M, New MI, Prader A, Pollack MS, O'Neill GJ, Yang SY, et al (1978) Genetic mapping of the 21-hydroxylase-deficiency gene within the HLA linkage group. N Engl J Med 299:911–915

Li W-H, Wu C-I, Luo C-C (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2:150–174

McAlpine PJ, Stranc LC, Boucheix C, Shows TB (1990) The 1990 catalog of mapped genes and report of the nomenclature committee. Cytogenet Cell Genet 55:5–76

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Press, Cold Spring Harbor, NY

Martin RD (1990) Primate origins and evolution: a phylogenetic reconstruction. Chapman & Hall, London

Matteson KJ, Phillips JA III, Miller Wl, Chung B-C, Orlando PJ, Frisch H, Ferrandez A, et al (1987) P450XXI (steroid 21-hydroxylase) gene deletions are not found in family studies of congenital adrenal hyperplasia. Proc Natl Acad Sci USA 84:5858–5862

Mayer WE, Jonker M, Klein D, Ivanyi P, van Seventer G, Klein J (1988) Nucleotide sequences of chimpanzee MHC class I alleles: evidence for trans-species mode of evolution. EMBO J 7:2765–2774

Miller WL (1988) Molecular biology of steroid hormone synthesis. Endocr Rev 9:295–318

Morel Y, Andre J, Uring-Lambert B, Hauptmann G, Bétuel H, Tosi M, Forest MG, et al (1989) Rearrangements and point mutations of P450c21 genes are distinguished by five restriction endonuclease haplotypes identified by a new probing strategy in 57 families with congenital adrenal hyperplasia. J Clin Invest 83:527–536

Mornet E, Crété P, Kuttenn F, Raux-Demay M-C, Boué J, White PC, Boué A (1991) Distribution of deletions and seven point mutations on CYP21B genes in three clinical forms of steroid 21-hydroxylase deficiency. Am J Hum Genet 48:79–88

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

New MI, Levine LS (1984) Recent advances in 21-hydroxylase deficiency. Annu Rev Med 35:649–663

Rasheed SA, Rongey RW, Bruszweski J, Nelson-Rees WA, Rabin H, Neubauer RH, Esra G, et al (1977) Establishment of a cell line with associated Epstein-Barr-like virus from a leukemic orang-utan. Science 198:407–409

Raum D, Awdeh Z, Anderson J, Strong L, Granados J, Teran L, Giblett E, et al (1984) Human C4 haplotypes with duplicated C4A or C4B. Am J Hum Genet 36:72–79

Rodrigues NR, Dunham I, Yu CY, Carroll MC, Porter RR, Campbell RD (1987) Molecular characterization of the HLA-linked steroid 21-hydroxylase B gene from an individual with congenital adrenal hyperplasia. EMBO J 6: 1653–1661

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

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Sinnott P, Collier S, Costigan C, Dyer PA, Harris R, Strachan T (1990) Genesis by meiotic unequal crossing-over of a *de novo* deletion that contributes to a steroid 21-hydroxylase deficiency. Proc Natl Acad Sci USA 87:2107–2111

Speiser PW, Dupont B, Rubinstein P, Piazza A, Kastelan A, New MI (1985) High frequency of nonclassical steroid 21-hydroxylase deficiency. Am J Hum Genet 37:650–667

Speiser PW, New MI (1985) Genetics of steroid 21-hydroxylase deficiency. Trends Genet 1:275–278

Speiser PW, New MI, White PC (1988) Molecular genetic analysis of nonclassical steroid 31-hydroxylase deficiency associated with HLA-B14,DR1. N Engl J Med 319:19–23

Steinmetz M, Stephan D, Dastoornikoo GR, Gibb E, Romaniuk R (1985) Methods in molecular immunology: chromosomal walking in the major histocompatibility complex. In: Lefkovits J, Pernis P (eds) Immunological methods. Academic Press, New York, pp 1–19

Tusie-Luna M-T, Speiser PW, Dumic M, New MI, White PC (1991) A mutation (Pro-30 to Leu) in CYP21 represents a potential nonclassic steroid 21-hydroxylase deficiency allele. Mol Endocrinol 5:685–692

Tusie-Luna M-T, Traktman P, White PC (1990) Determination of functional effects of mutations in the steroid 21-hydroxylase gene (CYP21) using recombinant vaccinia virus. J Biol Chem 265:20916–20922

Urabe K, Kimura A, Harada F, Iwanaga T, Sasazuki T (1990) Gene conversion in steroid 21-hydroxylase genes. Am J Hum Genet 46:1178–1186

Wallace AM, Beastal GH, Cook B, Currie AJ, Ross AM, Kennedy R, Girdwood RW (1986) Neonatal screening for congenital adrenal hyperplasia: a programme based on a novel direct radioimmunoassay for 17-hydroxyprogesterone in blood spots. J Endocrinol 108:299–308

White PC, Grossberger D, Onufer BJ, Chaplin DD, New MI, Dupont B, Strominger JL (1985) Two genes encoding steroid 21-hydroxylase are located near the fourth component of complement. Proc Natl Acad Sci USA 82:1089–1093

White PC, New MI, Dupont B (1986) Structure of human 21-hydroxylase genes. Proc Natl Acad Sci USA 83:5111–5115

——— (1987) Congenital adrenal hyperplasia. N Engl J Med 316:1519–1524

White PC, Vitek A, Dupont B, New MI (1988) Characterization of frequent deletions causing steroid 21-hydroxylase deficiency. Proc Natl Acad Sci USA 85:4436–4440