# Molecular Evolution of the *CYP2D* Subfamily in Primates: Purifying Selection on Substrate Recognition Sites without the Frequent or Long-Tract Gene Conversion

Yoshiki Yasukochi[1,*] and Yoko Satta[2]

[1]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-0033 Japan

[2]Department of Evolutionary Studies of Biosystems, The Graduate University for Advanced Studies, Shonan Village, Hayama, Kanagawa, 240-0193 Japan

*Corresponding author: E-mail: hyasukou@proof.ocn.ne.jp.

**Accepted:** March 18, 2015

## Abstract

The human cytochrome *P450* (*CYP*) *2D6* gene is a member of the *CYP2D* gene subfamily, along with the *CYP2D7P* and *CYP2D8P* pseudogenes. Although the CYP2D6 enzyme has been studied extensively because of its clinical importance, the evolution of the *CYP2D* subfamily has not yet been fully understood. Therefore, the goal of this study was to reveal the evolutionary process of the human drug metabolic system. Here, we investigate molecular evolution of the *CYP2D* subfamily in primates by comparing 14 *CYP2D* sequences from humans to New World monkey genomes. Window analysis and statistical tests revealed that entire genomic sequences of paralogous genes were extensively homogenized by gene conversion during molecular evolution of *CYP2D* genes in primates. A neighbor-joining tree based on genomic sequences at the nonsubstrate recognition sites showed that *CYP2D6* and *CYP2D8* genes were clustered together due to gene conversion. In contrast, a phylogenetic tree using amino acid sequences at substrate recognition sites did not cluster the *CYP2D6* and *CYP2D8* genes, suggesting that the functional constraint on substrate specificity is one of the causes for purifying selection at the substrate recognition sites. Our results suggest that the *CYP2D* gene subfamily in primates has evolved to maintain the regioselectivity for a substrate hydroxylation activity between individual enzymes, even though extensive gene conversion has occurred across *CYP2D* coding sequences.

**Key words:** CpG site methylation, *CYP2D6*, drug metabolism, gene conversion, molecular evolution, purifying selection.

## Introduction

Cytochrome P450 (CYP) enzymes catalyze the oxidation of exogenous and endogenous substrates, such as drugs, environmental chemicals, plant metabolites, fatty acids, and steroids. The enzymes are membrane-binding heme proteins, and they are widely expressed in prokaryotes and eukaryotes (Nelson et al. 1993). All the eukaryotic CYPs are considered to have been derived from a single common ancestor (Nelson 1999), although an ancestral enzyme has not yet been determined (Nelson et al. 2013; Sezutsu et al. 2013). A large variety of CYP proteins are necessary to interact with a large number of substrates, and the variety of substrates is considered to reflect the species-specificity for the habitat, environment, and diet (Gonzalez and Nebert 1990). In humans, *Homo sapiens*, the *CYP* gene family consists of 57 functional genes and 58 pseudogenes, whereas in mice, *Mus musculus*, there are 102 functional genes and 88 pseudogenes (Nelson et al. 2004). The putative number of *CYP* genes varies between organisms as shown in the Cytochrome P450 Homepage (Nelson 2009), including the Pacific transparent sea squirt, *Ciona savignyi*, and rice, *Oryza sativa*, with 97 and 323 functional genes, respectively, and as low as three expressed *CYP* genes in the budding yeast, *Saccharomyces cerevisiae*. Their genes are all members of the *CYP* superfamily and can be classified into two groups, with respect to their pattern of evolution (Thomas 2007; Gotoh 2012; Kawashima and Satta 2014). Products of the first group primarily function in metabolizing the endogenous substrates, and those genes undergo few gene duplications over long periods of time. On the other hand, products of the second group mainly metabolize exogenous substrates, and those genes have undergone frequent gene duplications or gene loss (birth-and-death evolution).

The hominin *CYP2D* subfamily, a member of the second group, consists of a functional *CYP2D6* and two paralogs, *CYP2D7* and *CYP2D8* (Nelson 2009), and the latter two genes often are not functional in some species (Yasukochi

and Satta 2011). The human CYP2D6 enzyme has a high affinity for alkaloids, and it detoxifies alkaloids (Fonne-Pfister and Meyer 1988). The enzyme is also important for metabolizing about 25% of commonly used drugs, such as antidepressants, β-blockers, and antiarrhythmics (Ingelman-Sundberg 2005). The three human *CYP2D* genes are located within a contiguous region of about 45 kb on chromosome 22 (Kimura et al. 1989). The nucleotide sequences of the two pseudogenes (*CYP2D7P* and *CYP2D8P*) are similar to that of *CYP2D6*. The *CYP2D7P* gene has a single frameshift mutation in the first exon, which causes a premature stop codon downstream. Alternative mRNAs from this pseudogene have been identified in human breast (Huang et al. 1997), lung (Huang et al. 1997), and brain (Pai et al. 2004; Gaedigk et al. 2005). Of them, a functional transcript was observed in some individuals in an Indian population (Pai et al. 2004). On the other hand, there is no observation of mRNA from *CYP2D8P* gene.

We previously identified the gene organization of the *CYP2D* subfamily in nonhuman primates, including the chimpanzee, *Pan troglodytes*, Sumatran orangutan, *Pongo pygmaeus abelii*, rhesus monkey, *Macaca mulatta*, and white-tufted-ear marmoset, *Callithrix jacchus* (Yasukochi and Satta 2011). Our findings reveal that the *CYP2D7* gene has been duplicated from *CYP2D6* in a stem lineage of humans and great apes, and the origin of *CYP2D6* and *CYP2D8P* genes in the human genome can be traced back to a stem lineage of the New World monkeys and Catarrhini at the latest. Two functional CYP2D isoforms have been observed in macaques and marmosets, although the copy number of the *CYP2D* genes is different among individuals in the case of macaques (Hichiya et al. 2004; Uno et al. 2010, 2011; Cooke et al. 2012). Furthermore, phylogenetic analyses for *CYP2D* genes suggest that the origin of the *CYP2D* subfamily can be traced back to before the divergence between amniotes and amphibians (Yasukochi and Satta 2011).

Substrate recognition sites (SRSs) are important for the catalysis of target substrates, and SRSs in the CYP2 family have been determined based on their homology with bacterial CYPs (Gotoh 1992). The 3D structure of the human CYP2D6 molecule revealed several functional important sites (Wolff et al. 1985; Modi et al. 1996; de Groot et al. 1999; Rowland et al. 2006; Unwalla et al. 2010). Rowland et al. (2006) obtained the crystal structure of CYP2D6 with 3.0-Å resolution and revealed that the basic structure of the CYP2D6 is similar to other CYP members. Unwalla et al. (2010) predicted the sites of the reaction center for known CYP2D6 substrates, based on a model of ligand-bound CYP2C5 complexes as a template. Wolff et al. (1985) and de Groot et al. (1999) have reported that most substrates of CYP2D6 have a basic nitrogen at a distance of about 5–7 Å (or 10 Å) from the site of oxidation and can interact with Asp-301or Glu-216 (de Groot et al. 1999).

The human CYP2D6 enzyme has been studied across various disciplines due to its clinical importance, including pharmacokinetics, pharmacogenetics, structural biology, and clinical medicine. Therefore, *CYP2D6* genetic variation in human populations has been also investigated extensively (Xie et al. 2001; Bradford 2002; Mizutani 2003; Raimundo et al. 2004; Sistonen et al. 2007; Fuselli et al. 2010; Contreras et al. 2011). More than 100 human *CYP2D6* alleles, including intron variations, in samples from around the world are described in the Human Cytochrome P450 Allele Nomenclature Database (http://www.cypalleles.ki.se/cyp2d6.htm). Additionally, unequal crossover events often cause copy number variation (CNV) of the gene (Heim and Meyer 1992; Panserat et al. 1995; Steen et al. 1995; Løvlie et al. 1996; Gaedigk et al. 2010). CNV and deleterious mutations of functional *CYP2D6* genes are categorized into four phenotypes based on their ability (rate) to metabolize drugs: A poor (PM), intermediate, efficient (EM), or ultrarapid metabolizer (UM) (Zanger et al. 2004). The PMs do not have any functional *CYP2D6*, and thus lack the ability to metabolize drugs. EMs correspond to a wild-type copy of *CYP2D6* with normal function. On the other hand, UMs can have up to 13 copies of *CYP2D6* gene and can rapidly metabolize medicines and thus are unaffected by drugs. The frequency of PMs is relatively higher in Europe than in other areas, whereas UMs are mainly found in North Africa (Sistonen et al. 2007).

Gene conversion among three *CYP2D* genes has occurred in human populations, and this probably also affects the genetic variation in drug metabolism (Hanioka et al. 1990; Heim and Meyer 1992). In human evolution, as people have only recently started to use drugs, drugs cannot be an evolutionary force for CYP2D evolution. Instead, the CYP2D enzyme must have been essential for the detoxification of foods containing plant toxins. Our previous study suggested that lineage-specific *CYP2D* gene expansion in vertebrates can reflect the requirement in an environment (Yasukochi and Satta 2011). This suggests that birth-and-death processes in *CYP2D* subfamily are associated with differences of feeding strategies among organisms. Nevertheless, there is little knowledge about the evolutionary pattern or mode of the *CYP2D* subfamily in primates. We here examine molecular evolution of the subfamily by comparing 14 *CYP2D* sequences from humans to New World monkey genomes.

## Materials and Methods

### Collection of Nucleotide Sequence Data of *CYP2D* Genes

We previously identified the gene organization of the human, chimpanzee, Sumatran orangutan, rhesus monkey, and white-tufted-ear marmoset *CYP2D* genes, based on nucleotide sequence data in the genome database and sequences determined by direct sequencing (Yasukochi and Satta 2011).

In addition, the *CYP2D6* sequence of the pygmy chimpanzee (Koch WH et al., unpublished data; GenBank accession number, DQ282163) was newly obtained from GenBank. They are human and orangutan *CYP2D6*, *CYP2D7P*, and *CYP2D8P*, pygmy chimpanzee (*Pan paniscus*) *CYP2D6*, common chimpanzee *CYP2D6*, *CYP2D7*, and *CYP2D8P*; and rhesus monkey and marmoset *CYP2D6* and *CYP2D8*. Sequences of the common chimpanzee (from now on we will refer to the common chimpanzee as the chimpanzee.) *CYP2D6* and *CYP2D7* genes were used according to the annotation of our previous study (Yasukochi and Satta 2011). In addition, three *CYP2D* homologs from each of the cynomolgus monkey (*Macaca fascicularis*) and marmoset were retrieved from the NCBI (National Center for Biotechnology Information) database (http://www.ncbi.nlm.nih.gov/).

As mentioned above, in humans there are more than 100 *CYP2D6* alleles that are categorized into four phenotypes. Thus, the elucidation of *CYP2D* molecular evolution is complicated by the presence of multiple C*YP2D6* alleles in a species. Here, we use the *CYP2D6*1* allele as the representative *CYP2D6* allele because this allele has the highest allele frequency in human populations (Sistonen et al. 2007). For putative pseudogenes, the character "*P*" was added after their gene names. Sequence data used in this study and analyses for each of the sequences (see below) are summarized in supplementary table S1, Supplementary Material online.

### Alignment of *CYP2D* Sequences and Test for Gene Conversion between Paralogous Genes

The alignment of nucleotide sequence data and translation into amino acids were performed using MEGA v5.10 (Tamura et al. 2011). The alignment was manually modified, and positions of deletions or insertions (indels) were determined. In the analysis, the pseudogene sequences were also translated according to the frame of *CYP2D6*, although they did not encode proteins, and transmembrane domains were predicted using SOSUI (Hirokawa et al. 1998).

The statistical test for detecting gene conversion was performed using the Two-Sample Runs Test (Takahata 1994) and GENECONV program version 1.81 (Sawyer 1989) which was run with default settings. We applied the Runs Test to *CYP2D6* and *CYP2D7*(*P*) genomic sequences of only human and chimpanzee origin because we were specifically focusing on the human *CYP2D6*. Indels were also treated as informative sites and were used because they were important for detecting gene conversion. The global test for recombination or gene conversion events was used with 10,000 permutations of the sequence alignment to assess the significance. Allelic pairs with Sim-*P* value $< 0.05$ were excluded from further analyses.

### Phylogenetic Analyses for *CYP2D* Sequences

The neighbor-joining (NJ) trees (Saitou and Nei 1987) for nucleotide sequences of intron 2 fragment in *CYP2D* genomic regions and amino acid sequences of the entire coding sequences (CDSs) were reconstructed based on the Kimura two-parameter (Kimura 1980) and *p*-distance models (Nei and Kumar 2000), respectively. The NJ tree was also constructed based on synonymous substitutions in the *CYP2D* CDSs. Bootstrap analyses for their NJ trees were performed using 1,000 replications. The maximum likelihood (ML) and maximum parsimony (MP) trees were implemented in the PHYLIP 3.69 package (Felsenstein 2009), and constructed on the basis of amino acid sequences at the SRSs. Bootstrap analyses were of 100 and 1,000 replicates for the ML and MP trees, respectively. For the ML tree, global rearrangement was allowed, and the input order of OTU was randomized with three jumbles during randomization. For the MP tree, the jumble option was set ten times. The ML and MP trees were visualized using the TreeView program version 1.6.6 (Page 1996). All positions containing any indels and missing data were eliminated from the data set (complete deletion option).

### Measure of Genetic Distance among *CYP2D* Sequences and Estimation of Natural Selection

The genetic distances of silent substitutions, which include both synonymous substitutions in the CDSs and changes in noncoding positions, were calculated by DnaSP 5.10.01 software (Librado and Rozas 2009). A sliding window analysis for the genetic distances of silent substitutions was also implemented in the DnaSP. The number of nonsynonymous substitutions per nonsynonymous site ($d_N$) and synonymous substitutions per synonymous site ($d_S$) was estimated with MEGA v5.10, using the distance of the modified Nei–Gojobori (Zhang et al. 1998) with the Jukes–Cantor correction (Jukes and Cantor 1969) for multiple substitutions. Standard errors were calculated using 1,000 bootstrapping replicates. Ratios of $d_N/d_S$ for nine functional *CYP2D6/7/8* and six *CYP2D6* CDSs were separately calculated at the SRSs and non-SRSs as well as over the entire coding region. In addition, we estimated $\omega$ ratios ($=d_N/d_S$) for 14 *CYP2D* CDSs including pseudogenes with the ML method by using CODEML in the PAML program version 4.8 (Yang 2007). Several models that estimate values of $\omega$ under different scenarios of selective pressure (M0, M1a, M2a, M7, and M8) were considered in this program. A Bayesian approach (Bayes empirical Bayes) in CODEML was used to predict positively selected codons. All indels and codons including nonsense and frameshift mutations were eliminated from the data set for the estimates of $\omega$ ratios.

### Amino Acid Sequences Variation for *CYP2D* Sequences in Primates

Amino acid variability among the 14 *CYP2D* CDSs was calculated using a Wu–Kabat plot (Wu and Kabat 1970).

The Wu–Kabat plot estimates the variability for each amino acid position, which is measured by the number of amino acids at a site divided by the maximum frequency of amino acid at the sites. Wu and Kabat (1970) originally evaluated the variability based on the sequence data for a single species, but here we used the sequences of orthologous genes from six different species, a method we refer to as "the modified Wu–Kabat plot." In this approach, if only a single amino acid appears at a site, the variability is the minimum value of 1, whereas if six different species have six different amino acid at a site, the value is 6/(1/6), or 36, at a maximum.

### Estimation of Nonfunctionalization Time of *CYP2D8* genes

To estimate nonfunctionalization times ($T_P$) of *CYP2D8P* genes in humans and orangutans, we calculated relative times of $T_P$ to species divergence time between the Hominidae and Old World monkeys ($T_S$). The relative times ($T_P/T_S$) were obtained from the number of nonsynonymous and synonymous substitutions between a functional gene and pseudogene according to the equation $T_P/T_S = 2\,(d_N/d_S - f)/(1 - f)$, by modifying the method of Sawai et al. (2008). The parameter $f$, which is the neutral fraction of substitutions, is determined by $f = d_{Nf}/d_{Sf}$, where $d_{Nf}$ and $d_{Sf}$ are nonsynonymous and synonymous divergence for the functional gene, respectively. We set the parameter $T_S$ to 35 Ma (Satta et al. 2004). The standard errors of $T_P/T_S$ were acquired by bootstrapping codons of *CYP2D8P* 10,000 times under a given $f$. In simulations, $T_P/T_S$ sometimes became negative, attributed to cases where $d_N/d_S < f$, due to insufficient time passed for accumulation of nonsynonymous substitutions since the loss of CYP2D8 function, suggesting that it is equivalent to the case of $T_P/T_S = 0$.

## Results

### Phylogenetic Relationships among *CYP2D* Genes in Primates

The NJ tree based on synonymous substitutions among 20 *CYP2D* CDSs of seven primates (fig. 1) showed that orthologous genes did not form a single clade. In the hominid, *CYP2D6* and *CYP2D7* were intermingled together, but *CYP2D8* did not cluster with them. In macaques, rhesus and cynomolgus monkey sequences grouped in a macaque-specific cluster. In marmosets, *CYP2D6* and *CYP2D8* clustered distinctly away from other primates. From this phylogeny, gene diversification was presumed to have occurred at least five times (diamonds 1–5 in fig. 1). Alternatively, the phylogeny shows that in orangutans, macaques, and marmosets, the conversion between *CYP2D6* and *CYP2D8* occurred in each species (diamonds 2, 4, and 5 in fig. 1). As the history of *CYP2D6* and *CYP2D7*(*P*) in humans and chimpanzees is not that simple, further phylogenetic analyses were performed to identify the time and location of ectopic gene exchanges.

### Gene Conversion between *CYP2D* Genes in Humans and Chimpanzees

For simplicity, *CYP2D6* and *CYP2D7*(*P*) genes in humans and chimpanzees were designated as H6, H7, C6, and C7, respectively. Informative sites of the four sequences were classified into three categories, supporting each of the three topologies. They were ((H6, C6), (H7, C7)), ((H6, H7), (C6, C7)), and ((H6, C7), (C6, H7)) (fig. 2*A*) and were named α, β, and γ, respectively. As *CYP2D6* and *CYP2D7*(*P*) duplication occurred before the divergence of humans and great apes, the topology α represents an original phylogenetic relationship. On the other hand, the topology β suggests gene conversion within each species, whereas the topology γ might be caused by recurrent substitutions. In fact, a site showing the topology γ was found only once (at the 2213 site) in the 6,924 bp compared, whereas α and β sites showed a block distribution (α1, α2, and β1–β3) (fig. 2*B* and supplementary fig. S1, Supplementary Material online).

We examined interlocus gene conversion (or recombination) using both Takahata's Two-Sample Runs Test and the GENECONV program. The Runs Test exhibited significant runs of both α and β informative sites (table 1) with $P$ values of $<0.001$. The analysis by GENECONV was performed using all the *CYP2D* genes of five primates, and it suggested significant gene conversion events in three β regions (supplementary table S2, Supplementary Material online).

To determine the timing of the conversion, we constructed a synonymous tree for each of the three β regions of H6, H7, C6, and C7, with orangutan *CYP2D6* and *CYP2D7P* (O6 and O7) as outgroups (fig. 3). In the trees for the three β regions, orthologous genes did not form a single cluster. For the β1 region, the distance between the orthologs, H6 and C6, was $d_{H6C6} = 0.013 \pm 0.004$ and comparable to the genome average (1.23%; Fujiyama et al. 2002). For the paralogous genes, $d_{C6C7}$ ($0.006 \pm 0.003$) was smaller than the genomic average. Both C6 and C7 were closer to H6 ($0.014 \pm 0.004$) than to H7 ($0.029 \pm 0.006$), suggesting that it is likely that C6 converts C7. These observations indicate that gene conversion between *CYP2D6* and *CYP2D7*(*P*) occurred at least once in the β1 region in an ancestral population of chimpanzees and humans. In the β2 and β3 regions, on the other hand, independent gene conversion seemed to have occurred in humans and chimpanzees, respectively. In addition, the average genetic divergences between H6/C6 and H7/C7 in the β2 and β3 regions were $0.031 \pm 0.007$ and $0.029 \pm 0.007$, respectively. The divergence also reflected the gene conversion between *CYP2D6* and *CYP2D7*(*P*) in an ancestor population of humans and chimpanzees.

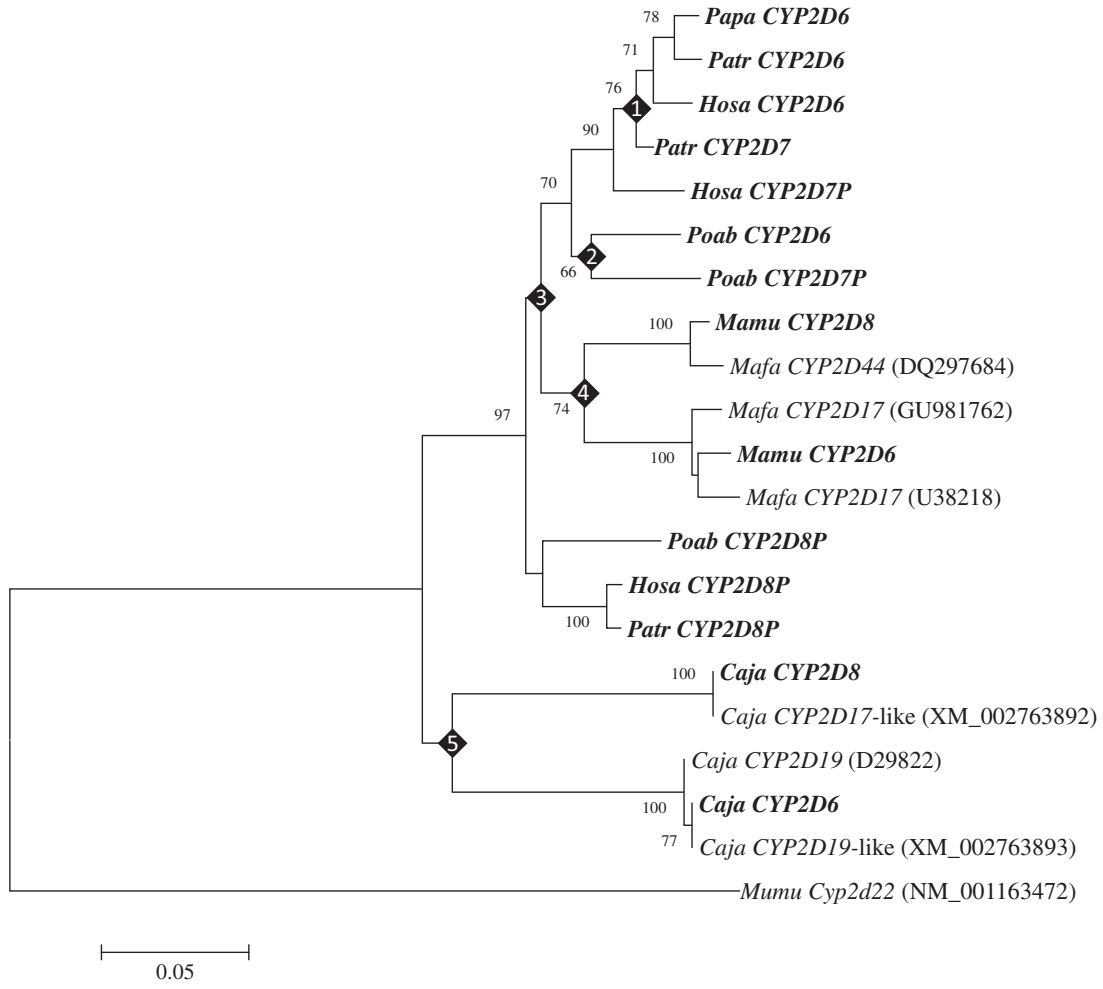


**Fig. 1.**—NJ tree of the *CYP2D* genes of primates, based on synonymous substitutions of the entire coding region. The modified Nei–Gojobori method was used as a distance measure with Jukes–Cantor correction. The internal node number represents bootstrap values (>50%). The diamond represents a gene divergence. The *CYP2D* sequences indicated in bold were obtained from the whole-genome sequence data, and nucleotide sequences of some fragments were determined in our previous study (Yasukochi and Satta 2011). The mouse *Cyp2d22* was used as an outgroup sequence. Hosa, human; papa, pygmy chimpanzee; patr, chimpanzee; poab, orangutan; mamu, rhesus monkey; mafa, cynomolgus monkey; caja, white-tufted-ear marmoset; mumu, house mouse.

## Large Genetic Distances in the *CYP2D* Region among Primates

We examined genetic divergence in the fragment of intron 2, which appeared to be scarcely affected by gene conversion across the *CYP2D* sequences of four Catarrhini species examined (supplementary fig. S2, Supplementary Material online). The recombination signal was not detected in intron 2 through both Runs test and GENECONV, but genetic distances ($d_{\_intron2}$) among sequences in the region were still large. For example, $d_{\_intron2}$ between the human and orangutan *CYP2D6* was $0.08 \pm 0.020$ (supplementary table S3*A*, Supplementary Material online). A sliding window analysis helped elucidate the large genetic distances across the *CYP2D* genomic region (supplementary fig. S3, Supplementary Material online). To examine whether the genetic divergence was enhanced uniquely in the *CYP2D* region, we calculated the $d_S$ of the *TCF20* (transcription factor 20) gene that was located adjacent to the *CYP2D8*(*P*) gene. The genetic distances among primate *TCF20* sequences were much lower than $d_{\_intron2}$ of *CYP2D* ($P < 0.001$). The value of $d_S$ between human and orangutan *TCF20* sequences was $0.03 \pm 0.002$ (supplementary table S3*B*, Supplementary Material online), in good agreement with the genome average (about 3%, Stewart and Disotell 1998; Satta et al. 2004; Peng et al. 2009). This result suggests that the large divergence is specific to the *CYP2D* nonconverted region.

As methylation of CpG sites often results in the large divergence between sequences, we compared GC content and the proportion of CpG sites in the *CYP2D* with those in *TCF20* CDSs. The GC contents of *CYP2D* CDSs and partial intron 2

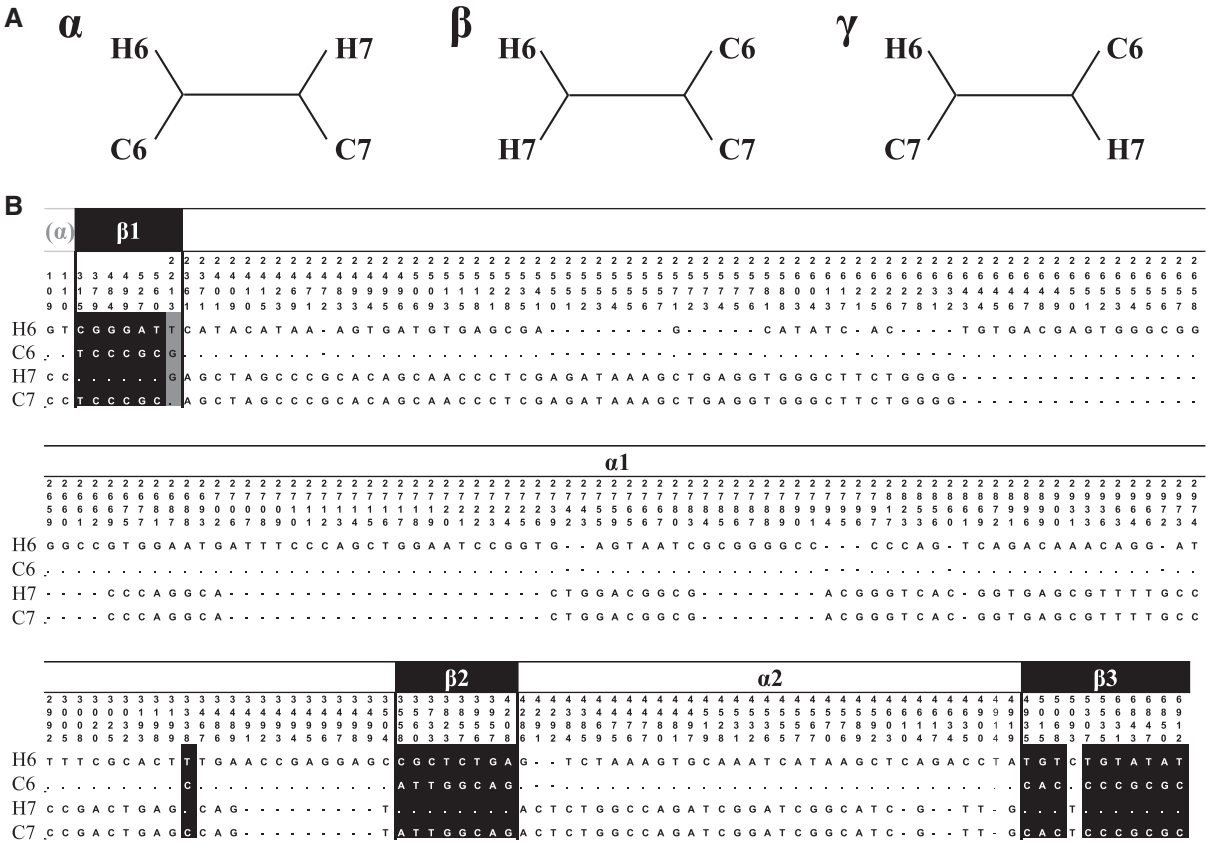


FIG. 2.—(*A*) The possible topologies at parsimony informative sites. In analysis, α topology was expected when a gene duplication event occurred before speciation, whereas β topology was expected when gene conversion or recombination occurred after its duplication. H6, human *CYP2D6* gene; H7, human *CYP2D7P*; C6, chimpanzee *CYP2D6*; C7, chimpanzee *CYP2D7*. (*B*) The alignment of nucleotide sequences based on parsimony informative sites in the *CYP2D6* and *CYP2D7*(*P*) of the human and chimpanzee. The white column indicates α topology. The filled black column indicates β topology. The filled gray column indicates γ topology.

sequences ranged from 63% to 69%, whereas those of the *TCF20* gene were approximately 52%. The average proportion of CpG sites in the entire *CYP2D* gene was much higher than that in the *TCF20* ($P < 0.001$; *CYP2D*, 6.4%; *TCF20*, 1.8%). In fact, when CpG methylation-related sites (CG, TG, and CA) were removed from *CYP2D* CDSs, their genetic distances became lower, for example, $d_S = 0.017 \pm 0.012$ between H6 and C6 intron 2 sequences (supplementary table S4, Supplementary Material online). These results suggest that genetic distances in the *CYP2D* region have been uniquely enhanced by the high mutation rate at the CpG sites.

## Amino Acid Sequences Variation for *CYP2D* Subfamily in Primates

The putative amino acid sequences of the *CYP2D* genes including pseudogenes were aligned. The aligned region is approximately 500 residues long (see supplementary fig. S4, Supplementary Material online). The amino acid sequences within some regions under the functional and structural constraints, such as heme-binding regions, were relatively conserved among the 14 *CYP2D* genes from the five primates. On the other hand, the sequence similarities of the transmembrane region were rather low.

The amino acid sequences at SRSs between *CYP2D7P* and *CYP2D8P* genes in Hominidae appeared to be more similar to those between *CYP2D6* and *CYP2D7P*. In contrast, at non-SRSs, amino acid sequences of *CYP2D6* and *CYP2D7P* were more similar to each other than those of *CYP2D7P* and *CYP2D8P* (supplementary fig. S4, Supplementary Material online). In the rhesus monkey and marmoset, amino acid sequences between *CYP2D6* and *CYP2D8* genes were more similar at non-SRSs than at SRSs. To confirm this observation, NJ trees of SRSs and non-SRSs for amino acid sequences and that of intron regions for nucleotide sequences were compared (fig. 4). In the SRS tree, all the *CYP2D6* genes from primates were monophyletic from the *CYP2D7*(*P*) and *CYP2D8*(*P*) (fig. 4*A*). The MP and ML trees also supported

**Table 1**
Two-Sample Runs Test for Detection of Gene Conversion

| Sister Group Status | A | B | *P* | *Z* | *k*[a] |
|---|---|---|---|---|---|
| (β, γ), α | 26 | 201 | $\approx 6.9 \times 10^{-25}$ | 10 | 5 |
| | | | $\approx 7.0 \times 10^{-26}$ | 9 | 4 |
| | | | $\approx 2.6 \times 10^{-27}$ | 8 | 4 |
| | | | $\approx 1.9 \times 10^{-28}$ | 7 | 3 |
| | | | $\approx 5.1 \times 10^{-30}$ | 6 | 3 |
| | | | $\approx 2.4 \times 10^{-31}$ | 5 | 2 |
| | | | $\approx 4.2 \times 10^{-33}$ | 4 | 2 |
| | | | $\approx 9.5 \times 10^{-35}$ | 3 | 1 |
| | | | $\approx 8.5 \times 10^{-37}$ | 2 | 1 |
| | | | $\ll 0.001$ | $\leq 10$ | |
| (α, γ), β | 202 | 25 | $\approx 5.9 \times 10^{-25}$ | 10 | 5 |
| | | | $\approx 6.3 \times 10^{-26}$ | 9 | 4 |
| | | | $\approx 2.3 \times 10^{-27}$ | 8 | 4 |
| | | | $\approx 1.7 \times 10^{-28}$ | 7 | 3 |
| | | | $\approx 4.7 \times 10^{-30}$ | 6 | 3 |
| | | | $\approx 2.3 \times 10^{-31}$ | 5 | 2 |
| | | | $\approx 4.1 \times 10^{-33}$ | 4 | 2 |
| | | | $\approx 9.5 \times 10^{-35}$ | 3 | 1 |
| | | | $\approx 8.5 \times 10^{-37}$ | 2 | 1 |
| | | | $\ll 0.001$ | $\leq 10$ | |

NOTE.—α, a sister group state, [(H6, C6),(H7, C7))]; β, a sister group state, [(H6, H7),(C6, C7))]; γ, a sister group state, [(H6, C7),(H7, C6))]. H6, human *CYP2D6*; H7, human *CYP2D7P*; C6, chimpanzee *CYP2D6*; C7, chimpanzee *CYP2D7*. A, site defined as one that does not support a certain sister group status; B, site defined as one that supports the specified sister group status; *Z*, continuum of the same sister group status (run).

[a]$Z = 2k$ when the number of *Z* is even, and $Z = 2k + 1$, when the number of *Z* is odd.

the same topology even though an outgroup sequence was changed (supplementary fig. S5, Supplementary Material online). The NJ trees of non-SRSs and intron regions showed that paralogous genes of each species formed an intermingled cluster for Hominidae (fig. 4*B* and *C*).

The variability within the CYP2D6 amino acid sequences was evaluated using a modified Wu–Kabat plot (See Materials and Methods) (fig. 5). The level of variability was not significantly different between SRSs and non-SRSs. High heterogeneities of amino acid variability were observed around αB and αJ′ helix tail ends and the β2-2 sheet, suggesting that a functional constraint was not so strong in these regions.

## Estimation of Nonfunctionalization Time of CYP2D8 Enzymes in Hominidae

A putative nonsense mutation in the *CYP2D8P* gene (site 344 between SRS-4 and SRS-5) was shared between human and chimpanzee lineages (fig. 6 and supplementary fig. S4, Supplementary Material online), but it was not shared with orangutans. Additional six independent nonsense and frameshift mutations were found in each of the human, chimpanzee, and orangutan *CYP2D8P* sequences. Based on these observations, pseudogenization occurred independently in a human/chimpanzee ancestor and in an orangutan lineage. No frameshift and nonsense mutations, however, were observed in the rhesus monkey or marmoset *CYP2D8* sequences, suggesting that they are functional. The single independent frameshift mutation of *CYP2D7P* was detected in human and orangutan, but not in the chimpanzee sequence.

To confirm the relaxation of functional constraint in the *CYP2D8P* pseudogenes, we estimated $\omega$ ratios under different selective scenarios for the specific lineages by the ML methods, based on the expected topology of *CYP2D8*(*P*) gene tree, (marmoset, (rhesus monkey, (orangutan, (human, chimpanzee)))). The analysis was performed in the similar manner of Zhao et al. (2010). The average $\omega$ value across all branches was 0.34 (table 2A), and the $\omega$ values on branches of human/chimpanzee/orangutan ($\omega = 1.00$, table 2B) *CYP2D8P* genes were higher than those across other branches ($\omega = 0.02$, table 2B and C).

It is interesting to know when the human and orangutan *CYP2D8P* genes have lost their function. The relative time ($T_P/T_S$) of pseudogenization of the gene was calculated using the method described in Materials and Methods (and see supplementary fig. S6, Supplementary Material online). We did not consider gene conversion in *CYP2D8P* from the *CYP2D6* and *CYP2D7*(*P*) genes, because *CYP2D8P* genes in Hominidae consistently formed clusters distinct from the other *CYP2D* genes in α and β regions (supplementary fig. S1, Supplementary Material online). The relative times ($T_P/T_S$) were 0.50 and 0.16 in human and orangutan *CYP2D8P*s, respectively. Bootstrap analyses showed that mean and standard errors of $T_P/T_S$ values were $0.54 \pm 0.29$ and $0.21 \pm 0.20$ in human and orangutan genes, respectively. The calibration time corresponding to $T_S$ was set to be the divergence time of 35 Ma between Hominidae and Old World monkeys (Satta et al. 2004). The pseudogenization times of the human and orangutan *CYP2D8P*s were estimated to be about 17.5 Ma ($19 \pm 10$ Ma) and 5.6 Ma ($7.4 \pm 7$ Ma), respectively. Although both estimates have large standard errors, the orangutan CYP2D8 enzyme has lost its function more recently than its counterpart in humans and chimpanzees.

## Selection Pressure Operating on *CYP2D* Genes

We calculated the mean $d_N$ and $d_S$ values for about 1,500-bp CDSs of functional *CYP2D6/D7/D8* genes for six primates (table 3). We then examined the ratio of $d_N/d_S$ in SRSs and non-SRSs separately. In the SRSs, the ratios were 0.79 among the three *CYP2D* genes and 0.48 among only *CYP2D6* genes, whereas those in non-SRSs were 0.32 and 0.24, respectively. The mean $d_N$ value of SRSs among the three *CYP2D* genes was the highest ($0.10 \pm 0.017$), and that of non-SRSs among *CYP2D6* genes was the lowest ($0.03 \pm 0.004$). A chi-square test showed no statistically significant differences in $d_N$ and $d_S$ values between SRSs and non-SRSs (supplementary table S5, Supplementary Material online). However, $d_N$ values at SRSs

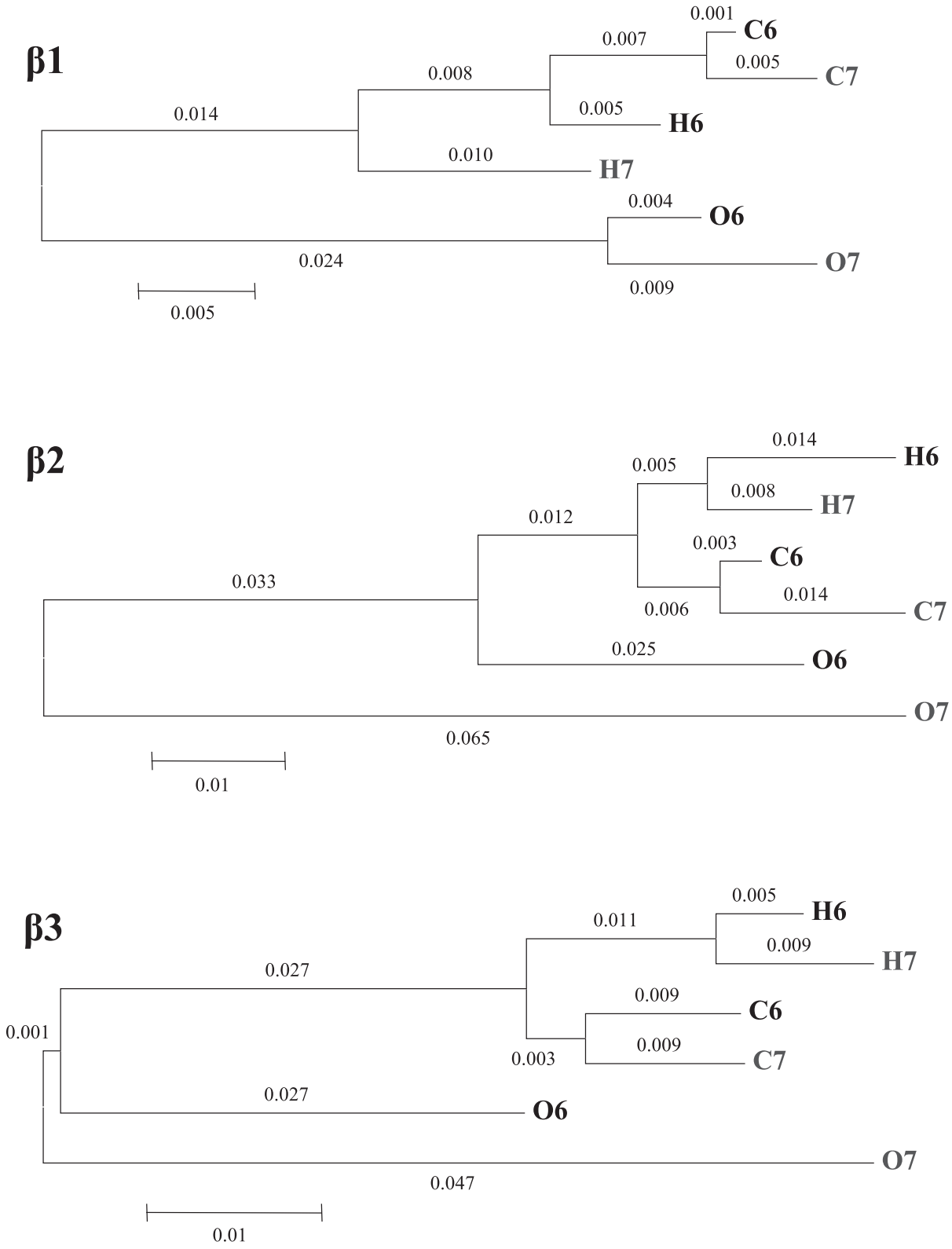


**Fig. 3.**—NJ trees based on the genetic distances of three β regions in *CYP2D6* and *CYP2D7*(*P*) genes of the human, chimpanzee, and orangutan. The genetic distances (three decimal places) were based on silent substitutions, which include synonymous substitutions in coding regions and changes in noncoding regions. H6, human *CYP2D6* gene; H7, human *CYP2D7P*; C6, chimpanzee *CYP2D6*; C7, chimpanzee *CYP2D7*; O6, orangutan *CYP2D6*; O7, orangutan *CYP2D7P*.

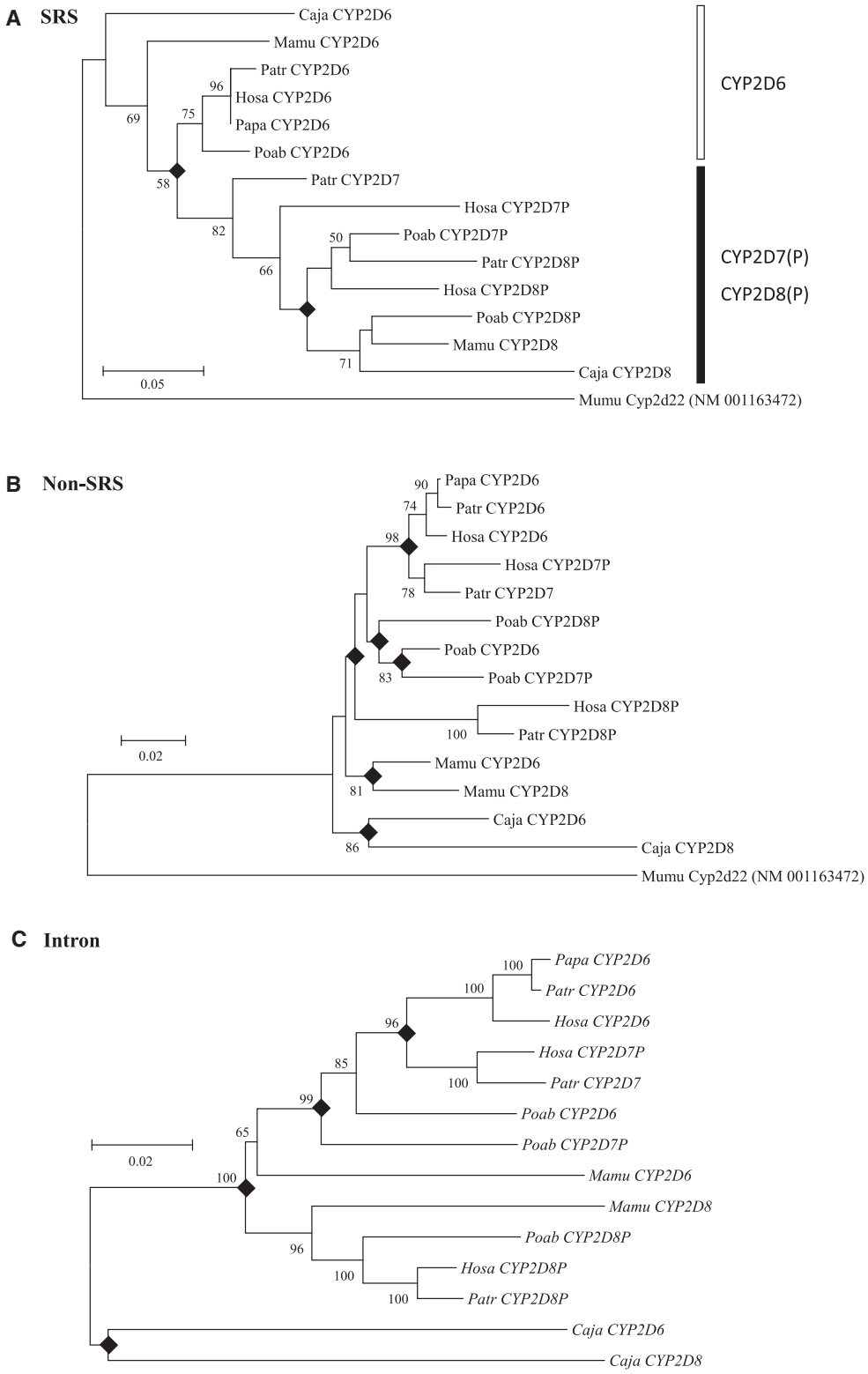


**Fig. 4.**—Three phylogenetic trees of primate *CYP2D* genes. (*A*) NJ tree based on amino acid sequences at SRSs. (*B*) NJ tree based on amino acid sequences at non-SRSs. (*C*) NJ tree based on nucleotide sequences in intron regions. A *p*-distance was used as a distance measure for amino acid sequences (*A* and *B*). A genetic distance of nucleotide sequences was calculated by using the Kimura two-parameter model (*C*). Diamond represents a gene divergence. Only bootstrap values over 50% are shown in the NJ tree. The mouse *Cyp2d22* was used as an outgroup sequence. Hosa, human; papa, pygmy chimpanzee; patr, chimpanzee; poab, orangutan; mamu, rhesus monkey; caja; white-tufted-ear marmoset; mumu, house mouse.

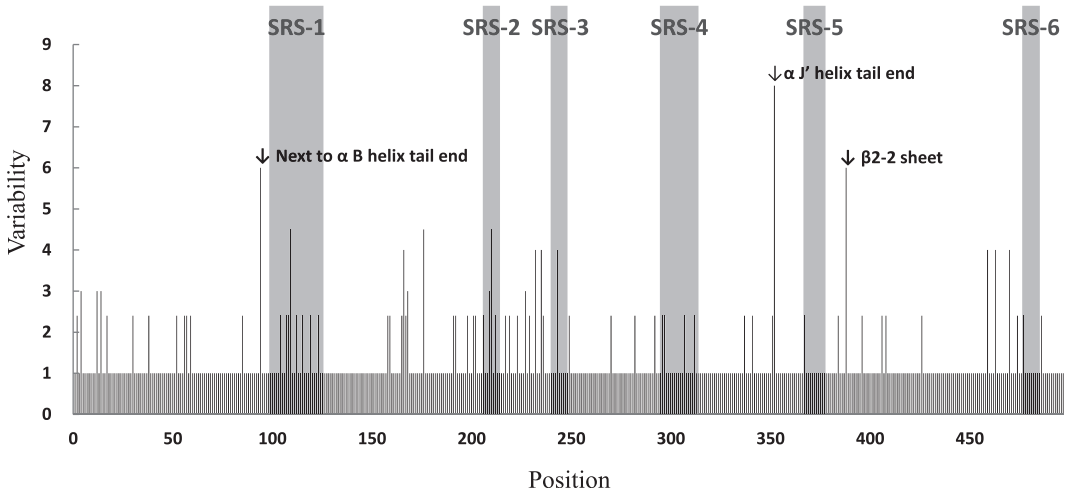


**Fig. 5.**—The variability level at amino acid residues among the *CYP2D6* genes in six primates by modified Wu–Kabat plot. Gray boxes indicate the SRSs identified by Gotoh (1992). The ordinate axis represents the variability of amino acid residues. The abscissa axis represents the position of the human CYP2D6 molecule.

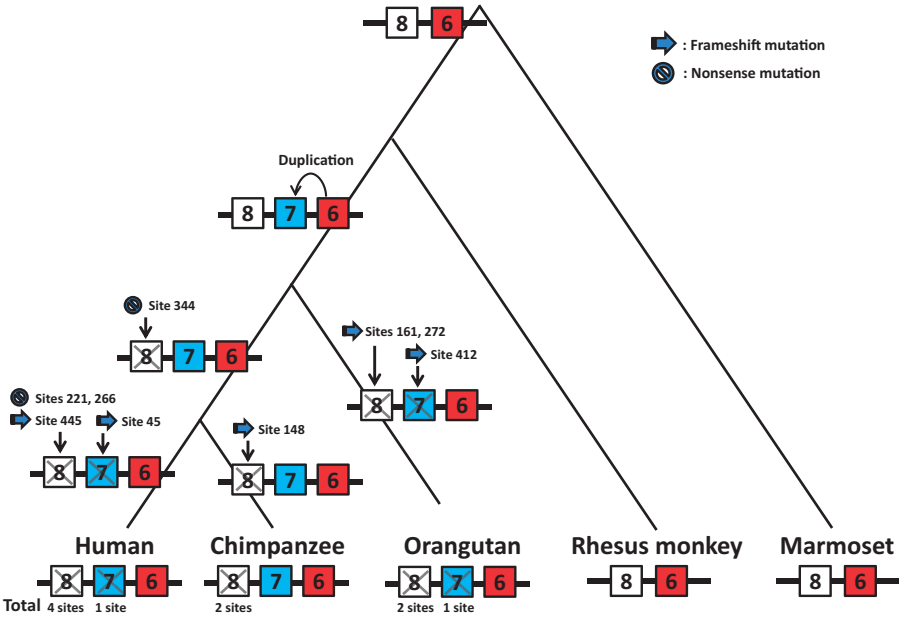


**Fig. 6.**—The chronological order of evolutionary events in the primate *CYP2D* gene cluster, based on the results of this study. The number in a box represents the CYP2D isoforms. A cross mark signifies a loss of function (pseudogene). The number of the site on a box represents a putative deleterious mutation site. A filled arrow and stop mark represent frameshift and nonsense mutations, respectively. The number of the site on the bottom represents the total number of their mutations.

among functional *CYP2D* genes were significantly higher than those at non-SRSs ($P = 5.8 \times 10^{-10}$ by Wilcoxon signed-rank test), although the length of SRSs was short.

By using the PAML program with different selective scenarios (models M0–M8), we estimated values of $\omega$ for functional *CYP2D* genes in five primates examined (supplementary table S6, Supplementary Material online). In the program, proportions $P_0$ of sites with $\omega < 1$ ($0 < \omega < 1$) were more than 75% whereas proportions $P_2$ of site with $\omega > 1$ were less than 2%.

## Discussion

### Gene Conversion in the *CYP2D* Cluster in Primates, but Not at the SRSs

Although some previous studies have reported that gene conversion affects evolution of *CYP2D* genes in humans (Kimura et al. 1989; Gonzalez and Nebert 1990; Masimirembwa et al. 1996), the phylogenetic tree based on amino acids at SRSs in this study indicates little effect of gene conversion between

**Table 2**

Estimates of $\omega$ Ratios for the Five Primate *CYP2D8(P)* Genes under Various Models

| | Model | np | $\omega$ | Likelihood | Model Comparison | *P* Value |
|---|---|---|---|---|---|---|
| A | All branches same $\omega$ | 10 | 0.34 | −3,220 | | |
| B | Hominidae branches $\omega_2$, other branches $\omega_1$ | 13 | $\omega_1 = 0.02$, $\omega_2 = 1.00$ | −3,205 | B versus A | $3.8 \times 10^{-8}$ |
| C | Hominidae branches $\omega_2 = 1$, other branches $\omega_1$ | 12 | $\omega_1 = 0.02$, $\omega_2 = 1.00$ | −3,205 | C versus B | 1.0 |

Note.—np, number of parameters.

**Table 3**

Mean $d_S$ and $d_N$ Distances among the Functional *CYP2D6/D7/D8* Genes and among *CYP2D6* Genes

| | No. of OTU | SRSs | | Non-SRSs | | Entire Coding Region | | |
|---|---|---|---|---|---|---|---|---|
| | | $\overline{d_N}$ | $\overline{d_N/d_S}$ | $\overline{d_N}$ | $\overline{d_N/d_S}$ | $\overline{d_N}$ | $\overline{d_S}$ | $\overline{d_N/d_S}$ |
| Primate functional *CYP2D6/D7/D8* genes[a] | 9 | $0.10 \pm 0.017$ | 0.79* | $0.04 \pm 0.004$ | 0.32* | $0.05 \pm 0.005$ | $0.12 \pm 0.010$ | 0.39* |
| Primate *CYP2D6* genes | 6 | $0.05 \pm 0.012$ | 0.48* | $0.03 \pm 0.004$ | 0.24* | $0.03 \pm 0.004$ | $0.11 \pm 0.010$ | 0.28* |

Note.—The $d_S$ and $d_N$ values are calculated by the modified Nei–Gojobori method with Jukes–Cantor ($R = 2.25$; Zhang et al. 1998).

[a]Excluding *CYP2D7P* genes of the human and orangutan and *CYP2D8P* genes of Hominidae.

*$P < 0.05$ ($d_N > d_S$; Wilcoxon signed-rank test).

SRS sequences of *CYP2D6* and *CYP2D8*. The SRSs identified by Gotoh (1992) were used in the present analysis, but de Graaf et al. (2007) have also proposed six SRSs. We reconstructed the NJ tree by using amino acids of the recognition sites identified by de Graaf et al. (2007), and confirmed the topological differences between SRSs and non-SRSs (data not shown). Although the definition of SRSs was slightly different between Gotoh (1992) and de Graaf et al. (2007), it is clear that the pattern of nucleotide substitution accumulation is different between the SRSs and non-SRSs based on both definitions.

In the comparison of amino acid sequences among *CYP2D* genes of five primates examined, the sequences of *CYP2D8* in the rhesus monkey and marmoset showed more similarities with those of human *CYP2D6* (H6) and *CYP2D7P* (H7) than the human *CYP2D8P* (H8), for example, 92% between H6 and rhesus monkey *CYP2D8* (R8), 91% between H7 and R8, and 89% between H8 and R8. However, *CYP2D8(P)* genes in all the five primates share two *Alu* elements, *AluSX* and *AluSg4*, in intron 1, whereas there is no *Alu* element in intron 1 of all *CYP2D6* and *CYP2D7P* genes (Yasukochi and Satta 2011). In addition, the phylogenetic tree of intron 2 fragment, which is scarcely affected by gene conversion, shows that *CYP2D8(P)* sequences form a distinct clade from *CYP2D6* and *CYP2D7(P)* sequences. Therefore, the *CYP2D7* gene is likely to have been duplicated from the *CYP2D6* after the divergence of hominid species and Old World monkeys at the latest.

## Difference in Metabolic Activity Level of CYP2D Enzymes among Primates

Our findings suggest that SRSs in CYP2D6 and CYP2D8 are scarcely affected by gene conversion to retain substrate specificity for each of their isoforms. CYP2D6 and CYP2D8 in the cynomolgus monkey metabolize the same compounds as humans, namely bufuralol and dextromethorphan, although their drug-metabolizing activity levels are different from those in humans (Uno et al. 2010). In addition, the CYP2D enzyme activity in macaques is more efficient than that in humans. In the marmoset, CYP2D19, a putative CYP2D6 ortholog, exhibits no activity for debrisoquine 4-hydroxylase but has high activity for debrisoquine 5-, 6-, 7-, and 8-hydroxylase, whereas CYP2D30, a putative CYP2D8 ortholog, exhibits high activity for the former and low activity for the latter (Cooke et al. 2012). These results suggest that the functional difference between CYP2D6 and CYP2D8 molecules is likely regioselectivity (and may be stereoselectivity) rather than the metabolism of different substrates. Amino acid sequences at SRSs have been likely maintained by purifying selection to retain this regioselectivity even though gene conversion affects the other regions of the gene.

Our results show that macaques and marmosets have at least two functional CYP2D isoforms, whereas Hominidae species appear to have one functional isoform only. Nishimuta et al. (2011) measured the relative activity of cynomolgus monkey CYP2D6 and CYP2D8 to human CYP2D6 by using the intrinsic clearance ($CL_{int}$) value, which is commonly used to measure hepatic drug metabolic activity, and revealed that the $CL_{int}$ in monkeys was higher than in the human. This result indicates more effective metabolism in monkeys. Although the actual activity in nonhuman Hominidae is unknown, human CYP2D enzyme had less activity than that of monkeys. It is interesting to know when this decrease in activity occurred.

## Enhanced Genetic Distances Caused by the Methylation of CpG Sites

Genetic distances at intron 2 that appeared unaffected by gene conversion between *CYP2D* orthologous genes in

primates were higher than expected, although all informative sites support the ordinary gene-cluster tree. On the other hand, the distances of *TCF20* gene located adjacent to the *CYP2D* region were in good agreement with the expectation of species divergence. In this study, we found that the substitution rate was specifically enhanced in *CYP2D* genes only (but not adjacent genes). This increased the mutation rate in the *CYP2D* genes could be due to CpG site methylation as GC content is higher in these genes relative to the *TCF20* gene. However, such CpG sites have been normally conserved by purifying selection if the gene is important (Uno and Osada 2011). The degeneration of CpG sites was possibly caused by a loss of functional constraint in the coding region of a gene. The one possible cause for relaxation of functional constraint on *CYP2D* in hominoids is discussed below.

### Pseudogenization of the *CYP2D* Gene in Hominidae

A nonsense mutation in *CYP2D8P*s was shared in the human and chimpanzee, but two frameshift mutations were independently found in the orangutan *CYP2D8P*, showing that *CYP2D8P*s have independently lost the function in common ancestors of human/chimpanzee and orangutan lineages. The pseudogenization times of human/chimpanzee and orangutan *CYP2D8P*s were estimated to be about 17.5 and 5.6 Ma, respectively. Previous studies reported that the divergence times of humans from orangutans were estimated at about 18 Ma (Satta et al. 2004; Steiper and Young 2006). Therefore, our estimation of the pseudogenization time suggests that the *CYP2D8* in human/chimpanzee may have immediately lost its function after the divergence of Hominidae species.

The average $\omega$ value across all branches in a tree of 14 *CYP2D* genes including both functional genes and pseudogenes estimated by CODEML was significantly lower than unity ($P < 1.4 \times 10^{-7}$ by likelihood-ratio test), indicating that purifying selection has generally acted on *CYP2D* genes. In addition, a two-ratio model exhibited that the $\omega$ value of Hominidae *CYP2D8P* clade was unity. Therefore, the evolutionary mode of the *CYP2D8P* genes may have switched to neutrality after the recent pseudogenization.

In general, animals are thought to perceive bitter or astringent taste as a method to avoid toxins in foods. The Old World monkeys tend to take in more antifeedants, such as condensed tannins and monoterpenes, compared with chimpanzees (Wrangham et al. 1998). Monkeys with smaller body size are thought to have an advantage for detoxifying plant metabolites, compared with humans and great apes, because the expression rates of microsomal enzyme activity negatively correlates with mammalian body size (Lambert 2002).

Ueno (2001) proposed the idea that primates have developed two different strategies to avoid unpredictable toxic risk. The first is the evolution of a detoxification system (physical adaptation), and the second is an enlargement of daily food diversity to dilute the harmful effect of toxin in foods (behavioral adaptation). Although both strategies are not mutually exclusive, this study supports that monkeys have utilized the former strategy primarily, whereas humans and great apes have utilized the latter. In fact, great apes probably ingest a variety of foods in a day to reduce the toxic effects of plants, although it has been reported that alkaloid-containing plants account for 14% of the diets of the chimpanzees examined (Simmen et al. 1999). Hominids are more likely to show insensitivity to the toxic effects than monkeys because of larger body size. It is possible that some *CYP2D* genes are pseudogenized in humans and great apes due to the relaxation of functional constraints in the detoxification system, whereas functional *CYP2D* genes have been retained in macaques and marmosets to efficiently detoxify substrates.

### Comparison of Amino Acid Residues at SRSs among CYP2D Molecules in Primates

To evaluate differences in *CYP2D* function, we compared amino acid residues at the SRSs of each *CYP2D* gene among primate species. Phe-120 at SRS-1 and Phe-483 at SRS-6 are important for CYP2D6 ligand binding (de Graaf et al. 2007), and in fact, Phe-120 was conserved in the primates examined, with the exception of Ile-120 in the chimpanzee CYP2D7 (Patr-CYP2D7) (supplementary fig. S7*A*, Supplementary Material online). As Phe-120 is responsible for direct $\pi$–$\pi$ interactions with aromatic rings of substrates (de Graaf et al. 2007), the Patr-CYP2D7 with Ile-120 may lack metabolic activity or have altered regioselectivity in metabolism. The comparison of amino acids at SRSs also shows that Phe-483 is conserved in all primate CYP2D molecules. The side chains of Phe-481 and Phe-483 are involved in interacting with an aromatic ring of a substrate although the importance for direct substrate recognition by these sites is ambiguous (de Groot et al. 1999; Rowland et al. 2006). Although all CYP2D6 and Patr-CYP2D7 molecules have Phe-481, all CYP2D8 molecules have Val-481 (supplementary fig. S7*B*, Supplementary Material online). In addition, Ala-482 is conserved in all CYP2D6s and Patr-CYP2D7, whereas Gly-482 is present in all CYP2D8s (supplementary fig. S7*C*, Supplementary Material online). As amino acids at 481 and 482 are unique for CYP2D6s/CYP2D7 and CYP2D8s, there is a possibility that those specificities confer differences in ligand regioselectivity in metabolism.

### Purifying Selection Has Acted on SRSs within the *CYP2D* Subfamily

All the values of the $d_N/d_S$ ratios for functional *CYP2D* genes in primates were significantly less than unity at SRSs ($P = 9.1 \times 10^{-4}$ by Wilcoxon signed-rank test), non-SRSs ($P = 2.9 \times 10^{-11}$), and over the entire coding region ($P = 2.9 \times 10^{-11}$). These results support purifying selection on the SRSs and non-SRSs within the *CYP2D* subfamily among different species. However, positive selection has been reported to act on

SRSs in four *CYP2* subfamilies (*CYP2A*, *2B*, *2C*, and *2D*) (Gotoh 1992), suggesting that different evolutionary forces have acted on SRSs between *CYP2D* orthologs and paralogs. Based on the higher $d_N/d_S$ ratios and $d_N$ values at SRSs among functional *CYP2D* genes relative to non-SRSs, we assume that the functional constraint on SRSs of the *CYP2D* subfamily is weaker than that on non-SRSs.

The inference of $\omega$ ratios under different scenarios of selective pressure in the PAML program showed a remarkably larger proportions of sites with $\omega < 1$ (>75%) than those with $\omega > 1$ (<2%). This result supports that amino acid sequences on the SRSs are totally conserved by purifying selection; however, some codons in the region may have been affected by diversifying selection. The Bayesian approach in the PAML analysis identified five positively selected sites (sites 167, 304, 373, 478, and 479; posterior probability, $P > 0.95$) under models M2a and M8 that allow for positive selection (in this case probably diversifying selection). Of the five codons, four are located within the SRSs, suggesting that the functional constraint of SRSs has been relaxed or diversifying selection has acted on their codons.

Although humans normally have only one functional *CYP2D* gene, mice have nine functional *Cyp2d* genes (Nelson et al. 2004), and there are 12 rodent *Cyp2d* genes that show evidence of positive selection (Thomas 2007). In contrast, we have found evidence of purifying selection in primate *CYP2D* genes. As a hypothesis, drug-metabolizing enzymes may have evolved by "plant–animal warfare" (Gonzalez and Nebert 1990; Heim and Meyer 1992), and the human CYP2D6 enzyme also has a very high affinity for plant toxins such as alkaloids (Fonne-Pfister and Meyer 1988). Ingelman-Sundberg (2005) has assumed that the difference in selective pressures for dietary detoxification potential resulted in the variance of the number of functional *CYP2D* genes between the human and mouse. The results of our study suggest that each *CYP2D* member in primates has evolved to retain different regioselectivity for a substrate in order to keep this function, and the moderate purifying selection may operate on each *CYP2D* gene, even though gene conversion widely affects the *CYP2D* genomic region.

In this study, we found that gene conversion and purifying selection have affected homogeneities between *CYP2D* paralogous genes in primates. In contrast, amino acids at SRSs have been moderately diversified between their paralogous genes. According to the number of functional *CYP2D* genes, macaques and marmosets appear to metabolize toxins in food more efficiently than Hominidae. Our findings, moreover, suggest that the methylation of CpG sites enhanced the divergence of *CYP2D* genes in the primates examined. In the near future, the development of a genome database for a variety of species will enable the analysis of intraspecies or interspecies genome-wide sequences.

## Supplementary Material

Supplementary tables S1–S6 and figures S1–S7 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

## Acknowledgments

This work was supported by Grant-in-Aid for Scientific Research (B) from Japan Society for the Promotion of Science (21370106). The authors are grateful to Max Planck Institute for Biology and Kyoto University for their kind contribution of sample collection. The authors also thank the Washington University School of Medicine, Genome Sequencing Center for the use of marmoset genome sequence data. We owe special thanks to Dr Naoyuki Takahata, Dr Jun Gojobori, and Dr Yukako Katsura for providing valuable comments. We also thank Dr Quintin Lau for the critical checking of the English language of this manuscript and two anonymous reviewers for their helpful comments.

## Literature Cited

Bradford LD. 2002. CYP2D6 allele frequency in European Caucasians, Asians, Africans and their descendants. Pharmacogenomics 3: 229–243.

Contreras AV, et al. 2011. Resequencing, haplotype construction and identification of novel variants of *CYP2D6* in Mexican Mestizos. Pharmacogenomics 12:745–756.

Cooke BR, Bligh SWA, Cybulski ZR, Ioannides C, Hall M. 2012. Debrisoquine metabolism and CYP2D expression in marmoset liver microsomes. Drug Metab Dispos. 40:70–75.

Felsenstein J. 2009. PHYLIP (Phylogeny Inference Package) ver.3.69. Distributed by the author, Department of Genome Sciences, University of Washington, Seattle, WA.

Fonne-Pfister R, Meyer UA. 1988. Xenobiotic and endobiotic inhibitors of cytochrome P-450dbl function, the target of the debrisoquine/sparteine type polymorphism. Biochem Pharmacol. 37:3829–3835.

Fujiyama A, et al. 2002. Construction and analysis of a human-chimpanzee comparative clone map. Science 295:131–134.

Fuselli S, et al. 2010. Evolution of detoxifying systems: the role of environment and population history in shaping genetic diversity at human *CYP2D6* locus. Pharmacogenet Genomics. 20:485–499.

Gaedigk A, et al. 2010. Identification of novel *CYP2D7-2D6* hybrids: non-functional and functional variants. Front Pharmacol. 1:121.

Gaedigk A, Gaedigk R, Leeder JS. 2005. *CYP2D7* splice variants in human liver and brain: does *CYP2D7* encode functional protein? Biochem Biophys Res Commun. 336:1241–1250.

Gonzalez FJ, Nebert DW. 1990. Evolution of the P450 gene superfamily. Trends Genet. 6:182–186.

Gotoh O. 1992. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. J Biol Chem. 267:83–90.

Gotoh O. 2012. The 50th anniversary and new horizons of cytochrome P450 research: expanding knowledge on the multiplicity and versatility of P450 and its industrial applications evolution of cytochrome P450 genes from the viewpoint of genome informatics. Biol Pharm Bull. 35: 812–817.

de Graaf C, et al. 2007. Molecular modeling-guided site-directed mutagenesis of cytochrome P450 2D6. Curr Drug Metab. 8:59–77.

de Groot MJ, Ackland MJ, Horne VA, Alex AA, Jones BC. 1999. Novel approach to predicting P450-mediated drug metabolism:

development of a combined protein and pharmacophore model for CYP2D6. J Med Chem. 42:1515–1524.
Hanioka N, Kimura S, Meyer UA, Gonzalez FJ. 1990. The human *CYP2D* locus associated with a common genetic defect in drug oxidation: a G1934→A base change in intron 3 of a mutant CYP2D6 allele results in an aberrant 3′ splice recognition site. Am J Hum Genet. 47: 994–1001.
Heim MH, Meyer UA. 1992. Evolution of a highly polymorphic human cytochrome P450 gene cluster: *CYP2D6*. Genomics 14:49–58.
Hichiya H, et al. 2004. Cloning and functional expression of a novel marmoset cytochrome P450 2D enzyme, CYP2D30: comparison with the known marmoset CYP2D19. Biochem Pharmacol. 68:165–175.
Hirokawa T, Boon-Chieng S, Mitaku S. 1998. SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics 14:378–379.
Huang Z, Fasco MJ, Kaminsky LS. 1997. Alternative splicing of CYP2D mRNA in human breast tissue. Arch Biochem Biophys. 343:101–108.
Huang Z, Fasco MJ, Spivack S, Kaminsky LS. 1997. Comparisons of CYP2D messenger RNA splice variant profiles in human lung tumors and normal tissues. Cancer Res. 57:2589–2592.
Ingelman-Sundberg M. 2005. Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): clinical consequences, evolutionary aspects and functional diversity. Pharmacogenomics J. 5:6–13.
Jukes T, Cantor C. 1969. Evolution of protein molecules. In: Munro H, editor. Mammalian protein metabolism. New York: Academic Press. pp. 21–132.
Kawashima A, Satta Y. 2014. Substrate-dependent evolution of cytochrome P450: rapid turnover of the detoxification-type and conservation of the biosynthesis-type. PLoS One 9:e100059.
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.
Kimura S, Umeno M, Skodaj RC, Meyert UA, Gonzalez FJ. 1989. The human debrisoquine 4-hydroxylase (*CYP2D*) locus: sequence and identification of the polymorphic *CYP2D6* gene, a related gene, and a pseudogene. Am J Hum Genet. 45:889–904.
Lambert JE. 2002. Digestive retention times in forest guenons (*Cercopithecus* spp.) with reference to chimpanzees (*Pan troglodytes*). Int J Primatol. 23:1169–1185.
Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452.
Løvlie R, Daly AK, Molven A, Idle JR, Steen VM. 1996. Ultrarapid metabolizers of debrisoquine: characterization and PCR-based detection of alleles with duplication of the *CYP2D6* gene. FEBS Lett. 392:30–34.
Masimirembwa C, Persson I, Bertilsson L, Hasler J, Ingelman-Sundberg M. 1996. A novel mutant variant of the *CYP2D6* gene (*CYP2D6*17*) common in a black African population: association with diminished debrisoquine hydroxylase activity. Br J Clin Pharmacol. 42:713–719.
Mizutani T. 2003. PM frequencies of major CYPs in Asians and Caucasians. Drug Metab Rev. 35:99–106.
Modi S, et al. 1996. A model for human cytochrome $P_{450}$ 2D6 based on homology modeling and NMR studies of substrate binding. Biochemistry 35:4540–4550.
Nei M, Kumar S. 2000. Molecular evolution and phylogenetics. Oxford: Oxford University Press.
Nelson DR. 1999. Cytochrome P450 and the individuality of species. Arch Biochem Biophys. 369:1–10.
Nelson DR. 2009. The cytochrome P450 homepage. Hum Genomics 4: 59–65.
Nelson DR, et al. 1993. The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. DNA Cell Biol. 12:1–51.
Nelson DR, et al. 2004. Comparison of cytochrome P450 (*CYP*) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. Pharmacogenetics 14:1–18.
Nelson DR, Goldstone JV, Stegeman JJ. 2013. The cytochrome P450 genesis locus: the origin and evolution of animal cytochrome P450s. Philos Trans R Soc Lond B Biol Sci. 368:20120474.
Nishimuta H, Sato K, Mizuki Y, Yabuki M, Komuro S. 2011. Species differences in intestinal metabolic activities of cytochrome P450 isoforms between cynomolgus monkeys and humans. Drug Metab Pharmacokinet. 26:300–306.
Page R. 1996. TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci. 12:357–358.
Pai HV, et al. 2004. A frameshift mutation and alternate splicing in human brain generate a functional form of the pseudogene cytochrome P4502D7 that demethylates codeine to morphine. J Biol Chem. 279: 27383–27389.
Panserat S, et al. 1995. An unequal cross-over event within the *CYP2D* gene cluster generates a chimeric *CYP2D7/CYP2D6* gene which is associated with the poor metabolizer phenotype. Br J Clin Pharmacol. 40:361–367.
Peng Z, Elango N, Wildman DE, Yi SV. 2009. Primate phylogenomics: developing numerous nuclear non-coding, non-repetitive markers for ecological and phylogenetic applications and analysis of evolutionary rate variation. BMC Genomics 10:247.
Raimundo S, et al. 2004. A novel intronic mutation, 2988G>A, with high predictivity for impaired function of cytochrome P450 2D6 in white subjects. Clin Pharmacol Ther. 76:128–138.
Rowland P, et al. 2006. Crystal structure of human cytochrome P450 2D6. J Biol Chem. 281:7614–7622.
Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 4:406–425.
Satta Y, Hickerson M, Watanabe H, O'hUigin C, Klein J. 2004. Ancestral population sizes and species divergence times in the primate lineage on the basis of intron and BAC end sequences. J Mol Evol. 59:478–487.
Sawai H, Go Y, Satta Y. 2008. Biological implication for loss of function at major histocompatibility complex loci. Immunogenetics 60: 295–302.
Sawyer S. 1989. Statistical tests for detecting gene conversion. Mol Biol Evol. 6:526–538.
Sezutsu H, Le Goff G, Feyereisen R. 2013. Origins of P450 diversity. Philos Trans R Soc Lond B Biol Sci. 368:20120428.
Simmen B, Hladik A, Ramasiarisoa PL, Iaconelli S, Hladik CM. 1999. Taste discrimination in lemurs and other primates, and the relationships to distribution of plant allelochemicals in different habitats of Madagascar. In: Rakotosamimmanana B, Rasamimanana H, Ganzhorn JU, Goodman SM, editors. New Directions in Lemur Studies. New York: Kluwer Academic/Plenum Press. pp. 201–219.
Sistonen J, et al. 2007. *CYP2D6* worldwide genetic variation shows high frequency of altered activity variants and no continental structure. Pharmacogenet Genomics. 17:93–101.
Steen VM, Molven A, Aarskog NK, Gulbrandsen AK. 1995. Homologous unequal cross-over involving a 2.8 kb direct repeat as a mechanism for the generation of allelic variants of human cytochrome P450 CYP2D6 gene. Hum Mol Genet. 4:2251–2257.
Steiper ME, Young NM. 2006. Primate molecular divergence dates. Mol Phylogenet Evol. 41:384–394.
Stewart CB, Disotell TR. 1998. Primate evolution—in and out of Africa. Curr Biol. 8:R582–R588.
Takahata N. 1994. Comments on the detection of reciprocal recombination or gene conversion. Immunogenetics 39:146–149.
Tamura K, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 28:2731–2739.
Thomas JH. 2007. Rapid birth-death evolution specific to xenobiotic cytochrome P450 genes in vertebrates. PLoS Genet. 3:e67.

Ueno Y. 2001. How do we eat? Hypothesis of foraging strategy from the viewpoint of gustation in primates. In: Matsuzawa T, editor. Primate origins of human cognition and behavior. Tokyo (Japan): Springer Japan. pp. 104–111.

Uno Y, Iwasaki K, Yamazaki H, Nelson DR. 2011. Macaque cytochromes P450: nomenclature, transcript, gene, genomic structure, and function. Drug Metab Rev. 43:346–361.

Uno Y, Osada N. 2011. CpG site degeneration triggered by the loss of functional constraint created a highly polymorphic macaque drug-metabolizing gene, *CYP1A2*. BMC Evol Biol. 11:283.

Uno Y, Uehara S, Kohara S, Murayama N, Yamazaki H. 2010. Cynomolgus monkey CYP2D44 newly identified in liver, metabolizes bufuralol, and dextromethorphan. Drug Metab Dispos. 38:1486–1492.

Unwalla RJ, et al. 2010. Using a homology model of cytochrome P450 2D6 to predict substrate site of metabolism. J Comput Aided Mol Des. 24: 237–256.

Wolff T, et al. 1985. Substrate specificity of human liver cytochrome P-450 debrisoquine 4-hydroxylase probed using immunochemical inhibition and chemical modeling. Cancer Res. 45:2116–2122.

Wrangham RW, Conklin-Brittain NL, Hunt KD. 1998. Dietary response of chimpanzees and cercopithecines to seasonal variation in fruit abundance. I. Antifeedants. Int J Primatol. 19:949–970.

Wu TT, Kabat EA. 1970. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. J Exp Med. 132:211–250.

Xie HG, Kim RB, Wood AJ, Stein CM. 2001. Molecular basis of ethnic differences in drug disposition and response. Annu Rev Pharmacol Toxicol. 41:815–850.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.

Yasukochi Y, Satta Y. 2011. Evolution of the *CYP2D* gene cluster in humans and four non-human primates. Genes Genet Syst. 86:109–116.

Zanger UM, Raimundo S, Eichelbaum M. 2004. Cytochrome P450 2D6: overview and update on pharmacology, genetics, biochemistry. Naunyn Schmiedebergs Arch Pharmacol. 369:23–37.

Zhang J, Rosenberg HF, Nei M. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc Natl Acad Sci U S A. 95:3708–3713.

Zhao H, Yang JR, Xu H, Zhang J. 2010. Pseudogenization of the umami taste receptor gene *Tas1r1* in the giant panda coincided with its dietary switch to bamboo. Mol Biol Evol. 27:2669–2673.

**Associate editor:** Yoshihito Niimura