A Novel Duplication Type of CYP2A6 Gene in African-American Population

Received October 26, 2006; accepted January 25, 2007

ABSTRACT:

Human CYP2A6 is responsible for the metabolism of nicotine and its genetic polymorphisms affect smoking behavior and risk of lung cancer. In the present study, we identified a novel type of CYP2A6 gene duplication that is created through an unequal crossover event with the CYP2A7 gene at 5.2 to 5.6 kilobases downstream from the stop codon. The novel duplication type of CYP2A6 was found in African Americans (n = 176) at an allele frequency of 1.7%, but was not found in European-American (n = 187), Korean (n = 209), or Japanese (n = 184) populations. The plasma cotinine/nicotine ratio in subjects possessing the novel CYP2A6 gene duplication with the CYP2A6*1 allele (10.8 ± 7.0, n = 4) was 1.4-fold higher than that in homozygotes of the wild type (8.0 ± 5.0, n = 87), although the difference was not statistically significant. The findings in the present study suggested that the novel duplicated CYP2A6 allele, which is specific for African Americans, would increase nicotine metabolism and may affect smoking behavior.

Cytochrome P450 is a superfamily of hemoproteins, many of which can metabolize xenobiotics such as clinically used drugs, procarcinogens, and environmental pollutants. The CYP2A gene subfamily comprises three genes, CYP2A6, CYP2A7, and CYP2A13, as well as a split pseudogene, CYP2A18. Among them, the genes coding functional enzymes are CYP2A6 and CYP2A13 (Yamano et al., 1990; Su et al., 2000). CYP2A6 is mainly expressed in liver and is responsible for nicotine metabolism (Nakajima et al., 1996). In addition, it can metabolize some pharmaceutical agents such as losigamone, letrozole, halothane, valproic acid, and disulfiram, and activates some procarcinogens such as 4-methyl nitrosourosoamino-1-(3-pyridyl)-1-butane and N-nitrosodiethylamine (Nakajima et al., 2002).

Genetic polymorphisms in the CYP2A6 gene are associated with large interindividual variability in nicotine metabolism, smoking behavior, and the risk of lung cancer (Fujieda et al., 2004). Several CYP2A6 alleles possess a single nucleotide polymorphism (SNP), which leads to an amino acid change, have decreased or no enzymatic activity. Other CYP2A6 alleles such as CYP2A6*1B or CYP2A6*12 are created by gene conversion or an unequal crossover event with the CYP2A7 gene (Ariyoshi et al., 2000; Oscarson et al., 2002). The CYP2A7 gene is located approximately 25 kb upstream of the CYP2A6 gene and shows an identity of 96.5% in the coding sequences (Yamano et al., 1990). Rao et al. (2000) reported a duplicated CYP2A6 allele, CYP2A6*1/2. This allele is created through an unequal crossover with the CYP2A7 gene at intron 8-exon 9 and the CYP2A6*4D allele is produced as the reciprocal product (Oscarson et al., 1999a). CYP2A6*4A and CYP2A6*4B alleles have also been reported as entire gene-deleted alleles (Oscarson et al., 1999b; Ariyoshi et al., 2004).

In a previous study (Nakajima et al., 2006), we found an African-American subject possessing the CYP2A6*1A, CYP2A6*1D (g.-1013A>G), and CYP2A6*1H (g.-745A>G) alleles, envisaging the existence of three copies of the CYP2A6 gene. However, genotyping analysis did not assign the CYP2A6*1/2 allele reported by Rao et al. (2000). In the present study, we investigated the CYP2A6 locus and identified the novel duplication type of CYP2A6 gene.

Materials and Methods

Chemicals and Reagents. Long and accurate (LA) TaqDNA polymerase and Blend TaqDNA polymerase were obtained from Takara (Shiga, Japan) and Toyobo (Osaka, Japan), respectively. Primers were commercially synthesized at Hokkaido System Sciences (Sapporo, Japan). The restriction enzymes were purchased from Takara, Toyobo, New England Biolabs (Ipswich, MA), and MBI Fermentas (Hanover, MD). Nicorette (nicotine gum containing 2 mg of nicotine) was obtained from Pfizer Japan (Tokyo, Japan). All other chemicals and solvents were of analytical or the highest grade commercially available.

Genomic DNA. This study was approved by the Human Studies Committee of Washington University School of Medicine (St. Louis, MO) and the Ethics Committees of Kanazawa University (Kanazawa, Japan) and Sapporo University (Sapporo, Japan). The restriction enzymes were purchased from Takara, Toyobo, New England Biolabs (Ipswich, MA), and MBI Fermentas (Hanover, MD). Nicorette (nicotine gum containing 2 mg of nicotine) was obtained from Pfizer Japan (Tokyo, Japan). All other chemicals and solvents were of analytical or the highest grade commercially available.

Genotyping of CYP2A6 Alleles. The genotyping of CYP2A6*1A (Nakajima et al., 2006), CYP2A6*1B1 (Nakajima et al., 2006), CYP2A6*1B2 (Nakajima et al., 2006), CYP2A6*1B13 (Nakajima et al., 2006), CYP2A6*1D (Nakajima et al., 2006), CYP2A6*1F (Nakajima et al., 2004), CYP2A6*1G (Nakajima et al., 2004), CYP2A6*1H (Nakajima et al., 2006), CYP2A6*1I (Nakajima et al., 2006), conventional CYP2A6*1/2 (Yoshida et al., 2002), CYP2A6*2 (Nakajima et al., 2000), CYP2A6*3 (Nakajima et al., 2000), CYP2A6*4A (Fukami et al., 2006), CYP2A6*4D (Fukami et al., 2006),
TABLE 1
Primers used in the present study

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>2A6FR-S</td>
<td>5′-AACTCCTCTGAGAGGACCT-3′</td>
</tr>
<tr>
<td>2A7FR-S</td>
<td>5′-AACTCCTCTGAGAGGACCT-3′</td>
</tr>
<tr>
<td>2A6FR4.2KS</td>
<td>5′-GATAAAGAACTTGAATGAAATG-3′</td>
</tr>
<tr>
<td>2A7FR5.4KS</td>
<td>5′-GATAAAGAACTTGAATGAAATG-3′</td>
</tr>
<tr>
<td>2A6FR6.8KR</td>
<td>5′-GACCTAGTGGTGGCAAGCC-3′</td>
</tr>
<tr>
<td>2A7FR7.8KR</td>
<td>5′-GACCTAGTGGTGGCAAGCC-3′</td>
</tr>
<tr>
<td>2A6FR13.8KR</td>
<td>5′-CCACAGGCGCTCTGAGATC-3′</td>
</tr>
<tr>
<td>2A7FR16.9KR</td>
<td>5′-CTCCCTTGAAGDGAATGAC-3′</td>
</tr>
</tbody>
</table>

CYP2A6*5 (Nakajima et al., 2001), CYP2A6*6 (Yoshida et al., 2002), CYP2A6*7 (Yoshida et al., 2002), CYP2A6*8 (Yoshida et al., 2002), CYP2A6*9 (Nakajima et al., 2006), CYP2A6*10 (Yoshida et al., 2002), CYP2A6*11 (Yoshida et al., 2002), CYP2A6*12 (Nakajima et al., 2004), CYP2A6*13 (Fukami et al., 2005b), CYP2A6*14 (Fukami et al., 2005b), CYP2A6*15 (Fukami et al., 2005b), CYP2A6*16 (Fukami et al., 2005b), CYP2A6*17 (Fukami et al., 2004), CYP2A6*18 (Fukami et al., 2005b), CYP2A6*19 (Fukami et al., 2005b), CYP2A6*20 (Fukami et al., 2005a), CYP2A6*21 (Nakajima et al., 2006), and CYP2A6*22 (Nakajima et al., 2006) was performed as described previously.

Determination of the Relative Gene Copy Ratio of CYP2A6/CYP2A7. To determine the copy number of the CYP2A6 gene, polymerase chain reaction (PCR) analyses were performed as described previously (Fukami et al., 2006). DNA samples from the subjects who were genotyped as CYP2A6*4A/CYP2A6*4A (n = 2, Japanese), CYP2A6*1/CYP2A6*4A (n = 4, Japanese), CYP2A6*1/CYP2A7*1 (n = 4, Japanese), and CYP2A6*1/conventional CYP2A6*1/2 (n = 1, Korean) were used for the standard curve.

PCR-RFLP to Identify the Crossover Region. To determine the crossover region of the novel duplicated CYP2A6 allele with the CYP2A7 gene, PCR restriction fragment length polymorphism (RFLP) was performed at the 3′-flanking region. The primer sets used were 2A6FR-S and 2A6FR13.8KR, 2A7FR-S and 2A7FR16.9KR, or 2A6FR-S and 2A7FR16.9KR (Table 1). The reaction mixture contained genomic DNA (200 ng), 1× LA-PCR buffer, 2.5 mM MgCl2, 0.4 mM deoxynucleoside-5′-triphosphates, 0.4 μM each primer, and 1 U of LA TaqDNA polymerase in a final volume of 25 μl. After an initial denaturation at 94°C for 1 min, the amplification was performed by denaturation at 98°C for 20 s, annealing and extension at 68°C for 12 min (with prolongation for 15 s per cycle during 13–26 cycles) for 26 cycles, followed by a final extension at 72°C for 10 min. The LA-PCR product was digested with Ael, EcoRI, EcoRV, or FspI at 37°C for 3 h. The digestion patterns were determined by electrophoresis in a 0.8% agarose gel.

Sequence Analysis of the 3′-Flanking Region of the CYP2A6 Gene. Sequence analysis was performed to examine the nucleotide sequences of the crossover region of the novel duplicated CYP2A6 allele. The PCR mixture contained genomic DNA (100 ng), 1× PCR buffer, 0.2 mM deoxynucleoside-5′-triphosphates, 0.4 μM 2A6FR4.2KS and 2A7FR7.8KR primers (Table 1), and 0.5 U of Blend TaqDNA polymerase in a final volume of 25 μl. After an initial denaturation at 94°C for 3 min, the amplification was performed by denaturation at 94°C for 25 s, annealing at 57°C for 25 s, and extension at 72°C for 3 min for 28 cycles, followed by a final extension at 72°C for 5 min. The PCR product was submitted to DNA sequencing using a Thermo Sequenase Dye Terminator Cycle Sequencing kit (GE Healthcare Bio-Science, Chalfont St. Giles, UK) with a Long-Read Tower DNA sequence (GE Healthcare Bio-Science).

Genotyping Method for the Novel Duplicated CYP2A6 Allele. Allele-specific-PCR was applied for the genotyping with the primer sets 2A6FR4.2KS and 2A6FR6.8KR or 2A7FR7.8KR (Table 1). The reaction mixture and PCR condition were the same as described above except for the primers. The CYP2A6*1 allele was amplified with the primer set 2A6FR4.2KS and 2A6FR6.8KR (2681 bp), and the novel duplicated CYP2A6 allele was amplified with the primer set 2A6FR4.2KS and 2A7FR7.8KR (2387 bp).

In Vivo Phenotyping of Nicotine Metabolism. Since 16 of 176 African Americans were judged as smokers based on the baseline presence of nicotine and cotinine, the phenotyping of in vivo nicotine metabolism was not performed in these subjects. Phenotyping analysis was performed in 160 African Americans in our previous study (Nakajima et al., 2006). In brief, the subjects chewed one piece of nicotine gum (Nicorette) for 30 min, chewing for 10 s per 30 s. Blood samples were collected from a cubital vein just before and 2 h after the start of chewing. The concentrations of nicotine and cotinine in the plasma samples were determined by high-performance liquid chromatography as described previously (Nakajima et al., 2000). The cotinine/nicotine ratio of the plasma concentration was calculated as an index of nicotine metabolism. Statistical analysis of the cotinine/nicotine ratios between the different genotypes was performed by Mann-Whitney U test. A value of P < 0.05 was considered statistically significant.

Results
A Subject Possessing Three CYP2A6 Alleles. In our recent study (Nakajima et al., 2006), we encountered an African-American subject who possessed the CYP2A6*1A, CYP2A6*1D, and CYP2A6*1H alleles. The subject was suspected to have three copies of the CYP2A6 gene. We determined the copy number of the CYP2A6 gene according to a PCR method described previously (Fukami et al., 2006). The ratios of the PCR products (exon 3 to intron 3) corrected by the fragment lengths of CYP2A6/CYP2A7 for the standard samples were as follows: 0.38 to 0.40 for CYP2A6*4A/CYP2A6*4A (n = 2), 1.67 to 1.80 for CYP2A6*1/CYP2A6*4A (n = 4), 2.55 to 2.95 for CYP2A6*1/CYP2A6*1 (n = 4), and 4.01 for CYP2A6*1/conventional CYP2A6*1 in the subject was 1.43. The ratios of the PCR products (exon 5 to intron 5) corrected by the fragment lengths of CYP2A6/CYP2A7 for the standard samples were as follows: 0.00 for CYP2A6*4A/CYP2A6*4A (n = 2), 0.75 to 0.83 for CYP2A6*1/CYP2A6*4A (n = 4), 1.47 to 1.61 for CYP2A6*1/CYP2A6*1 (n = 4), and 2.28 for CYP2A6*1/conventional CYP2A6*1 in the subject was 1.40. These results suggested that the subject had three copies of the CYP2A6 gene.

Identification of the Unequal Crossover Region of the Novel Duplicated CYP2A6 Allele. Genotyping analysis revealed that the subject did not have the conventional CYP2A6*1*2 allele reported by Rao et al. (2000). Thus, the crossover region was considered to be located downstream of exon 9 of the CYP2A6 gene. When the primer set 2A6FR-S and 2A6FR13.8KR was used, the 3′-flanking region of the CYP2A6 gene (13,507 bp, referring to the genomic sequences NG_00008.5) was amplified (Fig. 1). When the primer set 2A7FR-S and 2A7FR16.9KR was used, the 3′-flanking region of the CYP2A7 gene (16,471 bp) was amplified. When the primer set 2A6FR-S and 2A7FR16.9KR was used, the chimeric region of the novel duplicated CYP2A6 allele (15,300 bp by estimation) was amplified. PCR-RFLP analyses were performed to determine the crossover region. Based on the comparison of the RFLP patterns with four restriction enzymes, the crossover region was suspected to be approximately 6 kb downstream from the stop codon of the CYP2A6 gene. Sequence analysis demonstrated that the novel duplicated CYP2A6 allele has the CYP2A6 sequences up to 5.2 kb downstream of the stop codon, and the CYP2A7 sequences 5.6 kb downstream (Fig. 2). Thus, the crossover region was considered to be 5.2 to 5.6 kb downstream from the stop codon of the CYP2A6 gene.

Allele Frequency of the Novel Duplicated CYP2A6 Allele. We performed the genotyping analysis for the novel duplicated CYP2A6 allele in 187 European Americans, 176 African Americans, 209 Koreans, and 184 Japanese. Six of 176 African Americans were judged as carriers of the novel duplicated CYP2A6 allele, resulting in a prevalence of 3.4%. They were originally genotyped as CYP2A6*1/CYP2A6*1 (n = 4) and CYP2A6*1/CYP2A6*9 (n = 2). The geno-
typing method could not determine whether they have the duplicated CYP2A6 allele heterozygously or homozygously. However, it was suggested that they have three copies of the CYP2A6 gene, since the ratios of the copy numbers of CYP2A6/CYP2A7 in these six subjects were 1.41 to 1.49 (exon 3–intron 3) and 1.40 to 1.47 (exon 5–intron 5). Accordingly, the allele frequency of the novel duplicated CYP2A6 allele was determined as 1.7%. It was confirmed that the sequences around the crossover region of the novel duplicated allele in six subjects were identical. The novel duplicated CYP2A6 allele was not found in European Americans, Koreans, or Japanese.

**In Vivo Nicotine Metabolism in the Subjects Possessing the Novel Duplicated CYP2A6 Allele.** The cotinine/nicotine ratios in plasma were calculated as an index of nicotine metabolism in the African Americans. Since there was no association between the CYP2A6*1 suballeles (such as CYP2A6*1A, CYP2A6*1B, CYP2A6*1D, etc.) and the phenotype, the suballeles were combined as CYP2A6*1 alleles (Nakajima et al., 2006). Among six subjects possessing the novel duplicated CYP2A6 allele, the phenotype of one subject with the CYP2A6*9 allele could not be determined, because she was a smoker. Another subject possessing the novel duplicated CYP2A6 allele with the CYP2A6*9 allele showed a cotinine/nicotine ratio 6.5. However, the ratio was excluded from the data analysis because we could not determine whether his genotype was CYP2A6*/1/CYP2A6*/1/CYP2A6*9 or CYP2A6*/1/CYP2A6*9/CYP2A6*9. Interestingly, we found that the ratio in the subjects possessing the novel duplicated CYP2A6 allele (10.8 ± 7.0, n = 4) was 1.4-fold higher than that in homozygotes of the CYP2A6*1 allele (8.0 ± 5.0, n = 87), although the difference was not statistically significant. Further characterization with a much larger sample size may be required, but our data demonstrated that the novel duplicated CYP2A6 allele likely increased nicotine metabolism.

**Discussion**

Several years ago, Rao et al. (2000) found a duplication allele, CYP2A6*1×2. This allele is considered to be created through an unequal crossover event at intron 8 to exon 9, with the gene-deleted allele CYP2A6*4D as the reciprocal product. Although other deleted alleles such as CYP2A6*4A and CYP2A6*4B have been reported, the
**Fig. 2.** Sequence alignments of the 3' flanking region of the CYP2A6 gene, the CYP2A7 gene, and the novel duplicated CYP2A6 allele. Upper and lower nucleotide numbering refers to the next nucleotide of the stop codon as 1 for the CYP2A6 and CYP2A7 genes, respectively, with the reference sequence NG_000008.5. Deletions are denoted by dashes. Open boxes represent whether the sequences in the novel duplicated CYP2A6 allele are the same as the CYP2A6 gene or the CYP2A7 gene. Black and gray boxes represent the Alu elements and the AG-rich region, respectively.
reciprocal products have not been identified. In the present study, we found a novel duplicated CYP2A6 allele that is considered to be created by an unequal crossover with the CYP2A7 gene at 5.2 to 5.6 kb downstream from the stop codon. Interestingly, the Alu elements were found 4.7 to 6.4 kb downstream from the stop codon of the CYP2A6 gene (Fig. 2). Generally, homologous recombination between Alu elements might result in various genetic exchanges, including duplications, deletions, and translocations (Deininger and Batzer, 1999). For example, duplicated or multiplicaton alleles for CYP2A6 are well known, and it has been reported that repetitive sequences containing Alu elements would be involved in the generation of these alleles (Steen et al., 1995; Lundqvist et al., 1999).

Previously, Ariyoshi et al. (2004) reported that the entire CYP2A6 gene deletion allele CYP2A6*4B would be caused by unequal crossover at approximately 5.0 kb downstream from the stop codon of CYP2A6 gene. We could also confirm the crossover region in the CYP2A6*4B gene (data not shown). Since the crossover site is very close to that in the novel duplicated CYP2A6 allele found in this study, the possibility is suggested that the novel duplicated CYP2A6 allele may be the reciprocal product of the CYP2A6*4B allele (Fig. 3). So far, the CYP2A6*4B allele has been analyzed only in the Japanese population and found with an allele frequency of 0.24% (Ariyoshi et al., 2004). To investigate whether the CYP2A6*4B allele would also be found in African Americans, we performed a genotyping analysis according to the PCR-RFLP method by Ariyoshi et al. (2004). However, the genotyping was not accomplished owing to the presence of a SNP preventing the RFLP using BspHI. Therefore, a new genotyping method for CYP2A6*4B allele with allele-specific PCR was developed in the present study. The used primer sets were 2A6FR4.2kS or 2A7FR5.4kS and 2A6FR6.8kR (Table 1; Fig. 2). The CYP2A6*4 allele was amplified with the primer set 2A6FR4.2kS and 2A6FR6.8kR (2681 bp), and the CYP2A6*4B allele was amplified with the primer set 2A7FR5.4kS and 2A6FR6.8kR (2669 bp by estimation). Finally, the CYP2A6*4B allele was not found in African Americans as well as European Americans. The allele was found in a Korean subject and a Japanese subject, resulting in the allele frequency of 0.24% and 0.27%, respectively. Thus, the allele frequencies of the novel duplicated CYP2A6 allele and the CYP2A6*4B allele were not in concordance. Moreover, the duplicated CYP2A6 allele was not found in the Japanese population, although the frequency of the CYP2A6*4A allele was prominently high, as much as 20% (Schoedel et al., 2004; Nakajima et al., 2006). Although we have no plausible explanation, these results may suggest the possibility that reciprocal products caused by an unequal crossover event might not be equally inherited in the same populations.

It is interesting that all subjects with the novel duplicated CYP2A6 allele in the present study had the CYP2A6*1H allele, suggesting linkage of the duplication and the SNP g.-745A>G. Direct sequencing analyses demonstrated the presence of SNPs g.144G>A (exon 1), g.1620T>C (intron 2), g.3492C>T (exon 5), g.3570C>G (intron 5), g.5738C>T (exon 8), g.5843G>C (intron 8), g.6692C>G (3’-untranslated region) in the novel CYP2A6 allele in six subjects. The SNPs found in the exons are synonymous. An allele possessing these eight SNPs would be very similar to the haplotype 30 or 31 reported by Haberl et al. (2005).

Until now, we have evaluated in vivo nicotine metabolism using nicotine gum in nonsmokers (Nakajima et al., 2006). There is now much evidence that the genetic polymorphisms of the CYP2A6 gene are major determinants of the interindividual variability in nicotine metabolism. In the present study, we found that African-American subjects possessing the novel duplicated CYP2A6 allele showed increased (1.4-fold) nicotine metabolism compared with homozygotes of the wild type, consistent with the presence of three copies of the CYP2A6 gene. Our results are in agreement with a previous report by Rao et al. (2000) showing that individuals with the duplication allele CYP2A6*4*I*2 had a 1.4-fold higher plasma cotinine level than homozygotes of the wild type.

In conclusion, we found a novel duplicated CYP2A6 allele that is created by unequal crossover with the CYP2A7 gene at 5.2 to 5.6 kb downstream of the stop codon. The duplication allele was specific for African-American subjects at an allele frequency of 1.7%. Since the duplication allele likely increases nicotine metabolism, it might affect the smoking behavior or risk of lung cancer.

Acknowledgments. The enthusiasm and research support of Tracy Jones, Arnita Pitts, Phyllis Klein, and Ladonna Gaines (Washington University Center for Clinical Studies), and Margaret Ameayaw are greatly appreciated. We acknowledge Brent Bell for reviewing the manuscript.

Drug Metabolism and Toxicology, Division of Pharmaceutical Sciences, Graduate School of Medical Science, Kanazawa University, Kakuma-machi, Kanazawa, Japan (T.F., M.N., H.Y., Y.F., T.Y.); and Department of Medicine, Washington University School of Medicine, St. Louis, Missouri (H.L.M.)

References
Fujieda M, Yamazaki H, Saito T, Kiyotani K, Gyamfi MA, Sakurai M, Dosaka-Akita H,


**Address correspondence to:** Dr. Miki Nakajima, Drug Metabolism and Toxicology, Division of Pharmaceutical Sciences, Graduate School of Medical Sciences, Kanazawa University, Kakuma-machi, Kanazawa 920-1192, Japan. E-mail: nmiki@kenroku.kanazawa-u.ac.jp