Eur J Clin Pharmacol (2008) 64:465–475
DOI 10.1007/s00228-007-0445-7

PHARMACOGENETICS

The *CYP2D6* gene locus in South African Coloureds: unique allele distributions, novel alleles and gene arrangements

Andrea Gaedigk · Christa Coetsee

Received: 10 September 2007 / Accepted: 8 December 2007 / Published online: 17 January 2008

Abstract
Objective The highly polymorphic *CYP2D6* gene locus has been extensively scrutinised in the major ethnic populations, but little is known about the locus for many indigenous and unique admixed populations, including the Coloureds of South Africa. This study aimed at characterising the *CYP2D6* gene locus in Coloured subjects and predict their phenotype status.
Methods *CYP2D6* genotyping was performed on 99 Coloured subjects by long-range polymerase chain reaction (PCR) and PCR restriction fragment length polymorphism (RFLP). Testing included 25 allelic variants as well as gene duplications. Novel alleles were cloned and sequenced. A novel strategy for *CYP2D7/2D6* hybrid gene detection is described.
Results Thirteen alleles had a frequency of = 1%, three were infrequent (<1%) and 16 of the tested alleles were not detected. *CYP2D6*5* had a frequency of 17.2%, one of the highest ever observed in any population. Two novel alleles, *CYP2D6*64* and **65*, were identified. In addition, four samples carried *CYP2D7/2D6* hybrid genes, of which one matched the *CYP2D6*66* allele of a resequenced Caucasian control subject. *CYP2D6*66* is similar to *CYP2D6*16*, but its putative recombination point is further upstream. Genotyping identified three poor metabolisers (3%; predicted incidence 6.6%), and 12% of the population had an activity score of 0.5 indicative of intermediate metabolism.
Conclusions The Coloured population has a unique composition of alleles and a distinct frequency distribution. The preliminary data presented suggest decreased CYP2D6 activity compared with Caucasians. Follow-up studies including both genotyping and phenotyping will need to be conducted to further assess the relationship between genotype and phenotype in this population of complex ancestry.

Keywords CYP2D6 · Coloureds · Pharmacogenetics · Polymorphisms · Gene rearrangements

A. Gaedigk (✉)
Section of Developmental Pharmacology and Experimental Therapeutics, Children's Mercy Hospital & Clinics,
2401 Gillham Road,
Kansas City, MO 64108, USA
e-mail: agaedigk@cmh.edu

C. Coetsee
Faculty of Health Sciences, University of the Free State,
Bloemfontein, South Africa

Present address:
C. Coetsee
Biostatistics Department, Quintiles South Africa,
Bloemfontein, South Africa

Introduction

CYP2D6 is one of the best-investigated drug-metabolising enzymes and has been scrutinised extensively in all major ethnic populations [1]. In contrast, genotyping data are sparse or only emerging for many indigenous or geographically defined populations [2], and there is a void of information for unique admixed populations, including the Coloureds of South Africa. The population at Africa's southern tip is very diverse and consists of at least 20 ethnic groups and cultures, including the indigenous Khoi and San people. Four major racial groups are recognised (http://www.statsonline.gov.za/publications/P0302/P03022006.pdf). South Africa has a population of approximately 23.5 million and is composed of Africans (79.6%), whites (9.2%), Indian/Asians (2.4%) and Coloureds (8.8%). The African group encompasses an extraordinary variety of cultural and tribal identities, including the Nguni, Sotho and Khoi-Khoi groups. The white population group is com-

posed of descendants of European Caucasian immigrants who came to South Africa in previous centuries. The term Coloureds refers to a racially heterogeneous group of people who possess some degree of sub-Saharan ancestry in addition to substantial ancestry from Europe, Indonesia, South India, Ceylon, Madagascar, Mozambique, Mauritius, St. Helena, and western and southern Africa. People of this particular heritage are also referred to as Cape Coloureds, whereas Coloureds in KwaZulu-Natal are predominantly of British and Zulu heritage.

The impact of polymorphic *CYP2D6* expression on a subject's capability to metabolise numerous clinically used drugs is dramatic [1, 3, 4]. Depending on which alleles are present in an individual, a wide range of metabolic activity is observed among subjects: poor (PM), intermediate (IM), extensive (EM) and ultrarapid. Consequently, dose-related adverse events include toxicity in PMs due to higher than normal drug levels, to therapeutic failure due to extremely fast drug metabolism in subjects with a ultrarapid metabolism [5]. In Caucasians, four major nonfunctional alleles (*CYP2D6*3, *4, *5, *6*) with compiled frequencies of approximately 20–30% explain the majority of PM, whereas a number of reduced function alleles (*CYP2D6*9, *10, *41)* and alleles carrying functional gene duplications or multiplications (*CYP2D6*1xN, *2xN*) contribute to IM and ultrarapid metaboliser phenotypes, respectively. Other alleles appear to be restricted to certain ethnic groups or admixed populations. For example, the *CYP2D6*17* and **29* alleles have African origins and are found predominantly or exclusively in black Africans, African Americans and their descendants [2, 6, 7]. Yet other alleles are characterised by large differences in allele frequencies, such as *CYP2D6*10*, which is observed at a frequency of <5% in most populations but can reach frequencies over 50% in Asians [2].

Nonfunctional hybrid genes, such as *CYP2D6*13* and **16* [8, 9], have been known for a long time, but only limited information regarding their structure and frequency is available. Hybrids are composed of a 5′-*CYP2D7* portion and a 3′-*CYP2D6* portion (Fig. 1) and are believed to be products of large deletion events that fused respective gene regions [8, 9]. Because we detected an unusually high frequency of the *CYP2D6*5* deletion allele, the genotyping analysis was extended to include hybrid gene detection and characterisation.

Due to the unique and complex admixtue of South African Coloureds, it is impossible to predict their *CYP2D6* allele frequencies. The goal of this study was to characterise the *CYP2D6* gene locus in this unique population to generate basic knowledge about allele presence and genotype frequencies and use this information to predict their phenotype profile.

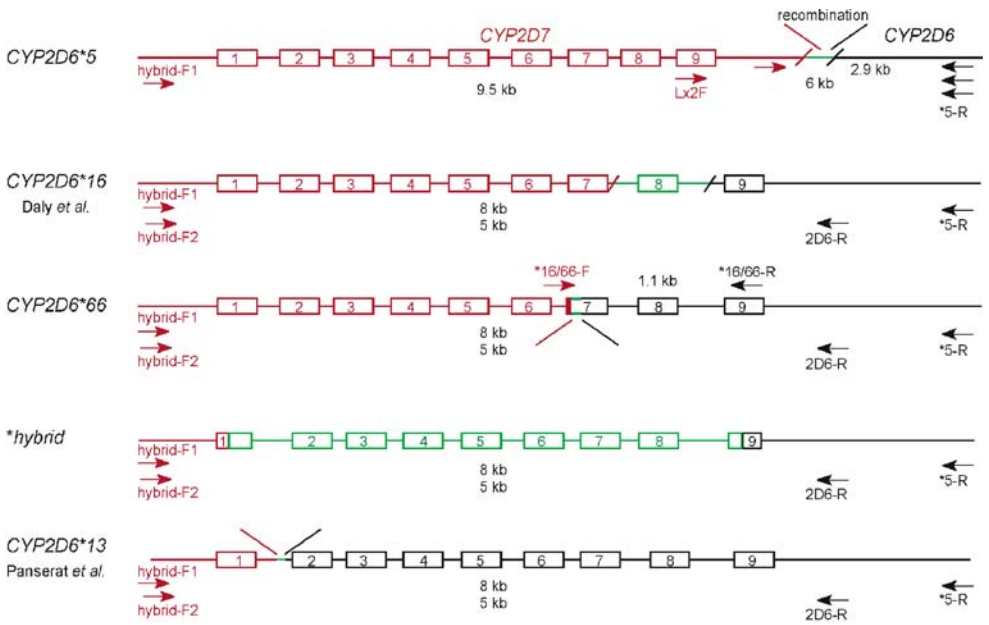

Fig. 1 Overview of the structure of the *CYP2D6*5* gene deletion and *CYP2D7/2D6* hybrid genes. The *CYP2D6*5* gene deletion and all *CYP2D7/2D6* hybrid structures are characterised by large deletions fusing proximal and distal regions within the locus. In *CYP2D6*5*, the entire *CYP2D6* gene is deleted, whereas the *CYP2D7* gene is completely retained. In the hybrid genes, 5′ portions of *CYP2D7* and 3′ portions of *CYP2D6* are fused together to create a number of hybrid genes containing different parts of each gene. Respective gene origins are given in *red* (*CYP2D7*) and *black* (*CYP2D6*). Sequences that could be derived from either gene or are not fully characterised are shown in *green*. *Arrows* denote approximate primer binding locations; primer names correspond to those given in Table 1. *Red and black arrows* denote *CYP2D7*- and *CYP2D6*-specific primers, respectively. Polymerase chain reaction product lengths are as indicated. The structures shown for *CYP2D6*13* and **16* are according to those described by Panserat et al. [9] and Daly et al. [8], respectively. *Brackets* indicate the region within recombination likely has occurred; *red and black lines* represent the most 5′ and 3′ position for which *CYP2D7* and *CYP2D6* sequences could unequivocally be determined; *green* indicates that the sequence could have been retained from either

Methods

Subjects and source of DNA

Blood samples from 100 subjects were obtained from the archives of the Department of Hematology and Cell Biology of the University of the Free State, Bloemfontain, South Africa. These samples were collected for genetic testing and were from individuals who had voluntarily disclosed their ethnicity. The samples were anonymised to conform to the study protocol that was approved by the Ethics Committee of the Faculty of Health Sciences of the Free State. Blood was collected in ethylenediaminetetraacetate (EDTA) tubes and stored at 4°C. DNA was isolated from 0.4 ml of blood using a chloroform/phenol extraction method adapted from the Promega Wizard Genomic DNA extraction kit (Promega, Madison, WI, USA). DNA quality was assured by agarose gel electrophoresis. DNA was obtained for 99 samples.

CYP2D6 genotyping

Genotype analysis was performed by long-range polymerase chain reaction (PCR) in combination with PCR restriction fragment length polymorphism (RFLP), allele-specific PCR and diagnostic long-range PCR reactions, as previously described in detail [10, 11] and references therein. Testing comprised *CYP2D6*2, *2A, *3, *4, *5, *6, *7, *8, *9, *10, *11, *17, *29, *36, *40, *41, *42, *45/46* and **56* allelic variants. Gene duplications were detected and characterised as described previously [10]. In addition, all subjects with homozygous genotypes were tested for the presence of hybrid genes such as *CYP2D6*13* and **16* [8, 9].

Detection and characterisation of *CYP2D7/2D6* hybrid genes

Identification of hybrid genes by a 8 kb amplification product Hybrid genes can be detected by amplifying the entire gene using a *CYP2D7*-specific primer located upstream of exon 1 (hybrid-F1) and a *CYP2D6*-specific primer (*5-R) binding downstream of the *CYP2D6* gene locus [8]. This assay (Table 1), adapted from Daly et al. [8] produces an 8-kb-long product from hybrid genes (Fig. 1). Of note, these primers also allow amplification of a 9-kb fragment from the *CYP2D6*5* deletion allele (Fig. 1). Even though the reaction conditions have extensively been optimised, assay results remain highly variable (i.e. amount of PCR product generated, presence of unspecific PCR products) and are sensitive to the quality of the genomic DNA. Alleles carrying hybrid genes do not produce the 6.6-kb-long product that is utilised for genotyping (i.e. the 6.6-kb genotyping product which was previously described in detail [11, 12]). Therefore, subjects carrying two hybrid alleles or a hybrid allele in combination with a *CYP2D6*5* are negative in the 6.6-kb PCR reaction, whereas subjects carrying any 6.6-kb-producing allele such as *CYP2D6*1, *2* in combination with a hybrid or *CYP2D6*5* allele will produce a genotype that reflects only one allele and consequently appears to be homozygous (e.g. a *CYP2D6 *1/*1* DNA may indeed be a **1/*5* or **1/hybrid*).

*Identification of CYP2D6*16 and *66 hybrid genes by a 1.1 kb amplification product* To specifically detect *CYP2D6*16, *66* or similar hybrid structures, an assay amplifying a 1.077-kb (1.1 kb)-long fragment was developed. PCR reactions were carried out with primers *16/66-F and *16/66-R in 8-μl volumes in the presence of 10–20 ng genomic DNA, and typically, 1–2 μl were analysed by agarose gel electrophoresis. Assay details are summarised in Table 1. The hybrid gene of a Caucasian individual previously identified to carry a hybrid gene was sequenced and deposited into GenBank (EU093102). This gene was initially called *CYP2D*16* [13] and was revised to *CYP2D6*66*.

Sequence analysis of hybrid genes For DNA sequencing, a nested PCR product of 5 kb encompassing the hybrid gene was generated from the initial 8-kb PCR amplicon with primers hybrid-F2 and 2D6-R, as detailed in Table 1. Alternatively, this product was generated directly from genomic DNA. The PCR products were treated with EXOSAP-IT (USB, Cleveland, OH, USA) and subsequently sequenced with primers that did not discriminate between *CYP2D6* and *CYP2D7* to determine the hybrid structure. Sequence gaps were filled in with *CYP2D6*- or *2D7*-specific primers to obtain consecutive full-length sequence. Sequencing was performed with DYEnamic ET dye terminator chemistry and a MegaBACE 500 capillary sequencer (GE, Piscataway, NJ, USA).

Improved strategy for hybrid gene detection and genotyping The size of the hybrid-specific amplicon was decreased from 8 kb to 5 kb. This was achieved by replacing the reverse primer (*5-R) with the primer also used to generate the *CYP2D6*-specific 6.6-kb genotyping fragment (2D6-R) and pairing it with a *CYP2D7*-specific forward primer (hybrid-F1). Assay details are given in Table 1. DNA samples with *CYP2D6*5/*5* and *CYP2D6*1/*1* genotypes served as negative controls and the previously characterized *CYP2D6*5/*66* DNA as positive control. To further characterise the 5-kb amplicon and demonstrate that it contained a hybrid gene arrangement, this fragment was used as a template to genotype for key sequence variations in exon 1 and exon 9. For the nested reamplification reactions, the 5-kb amplicon was diluted 1,000- to 2,000-fold, with 10 mM Tris ph 8 and 0.8 μl used for PCR. Specifically, the *CYP2D7*-specific T insertion in exon 1 (137insT) was detected with an

Table 1 Polymerase chain reaction (PCR) primers and reaction conditions

Primer/ orientation	Primer sequence 5′ to 3′	PCR amplicon length	Ann/ ext °C	Ext time	DMSO	Use of PCR/assay
*64-F	CTGGGCTGCACGCTAC*T* (100T, *10)	956 bp	65/72[c]	30 s	No	Allele-specific PCR to detect *CYP2D6*64*
*64-R	CGAAACCCAGGATCTGG*A* (1023T, *17)					
*64 IC-F	CTT CCG TGG AGT CTT GCA GG[b]	775 bp				
*64 IC-R	CCT GGG AGG TAG GTA GCC CTG[b]					
Hybrid-F1	CTC CGA CCA GGC CTT TCT ACC ACC	8 kb (*13, *16)	70/68[d]	10 min	5%	Long-range PCR to detect hybrid genes, adapted from [8]
*5-R	CAG GCA TGA GCT AAG GCA CCC AGA C	9.5 kb (*5)				
Hybrid-F2	TCC GAC CAG GCC TTT CTA CCA C	5018 kb (5 kb)	67/68[d]	6 min	5%	Long-range PCR to detect hybrid genes
2D6-R	ACT GAG CCC TGG GAG GTA GGT AG					
*16/66-F	CAG GAC GAA GGA GAG TGT CCC CT	1077 kb	56/68[d]	1 min	No	PCR to detect *CYP2D6*16* and *66 or similar
*16/66-R	GGA AAG CAA AGA CAC CAT G		58/72[c]			
Tins-F	CCC ACC AGG CCC CCT GCC ACT GCC- CGG GCT GGG CAA gCT (*Hind*III)	144 bp	72/72[c]	20 s	No	PCR followed by *Hind*III digest. 137insT cuts into 109+35 bp
Tins-R	GTT TGC TGG TGG TGG GGC AT					
L×2F[a]	GCC ACC ATG GTG TCT TTG CTT TCC TGG	603 bp	72/72[c]	30 s	No	Gene-specific PCR for *CYP2D6* exon 9
2D6-R	ACT GAG CCC TGG GAG GTA GGT AG					

PCR amplicon lengths are given in base pairs (bp) or kilobase pairs (kb)
Ann/ext °C annealing and extension temperatures in °C used for PCR cycling, *Ext time* reactions' extension time in seconds or minutes, *DMSO* dimethylsulfoxide added to a final concentration of 5% were indicated to enhance reaction performance
[a] L×2F was modified from that previously described by Johansson et al. [20]
[b] the ratio between the *64-F/*64-R and *64 IC-F/*64 IC-R primer sets was 2:1 (final primers concentrations in PCR reaction was 0.25 μM (*64-F and *64-R) and 0.125 μM (*64 IC-F and *64 IC-R), respectively
[c] JumpStart RED Taq polymerase or [d] JumpStart Accu Taq LA polymerase (Sigma, St Louis, MO, USA) were used for PCR reaction, and buffers provided.

assay modified from that described previously [14]. PCR was carried out with a forward primer that contained a partial *Hind*III restriction site (Tins-F) and a reverse primer (Tins-R) that bound to intron 1. PCR products containing 137insT were cut with *Hind*III [14]. The nature of exon 9 was determined by gene-specific PCR that utilised a *CYP2D6*-specific forward primer (L×2F) and the reverse primer also used to generate the 5-kb amplicon (2D6-R). A product was only formed from 5-kb templates containing *CYP2D6* exon 9 sequences. Specificity of this assay was demonstrated on a *CYP2D6*5/*5* DNA. Further details are given in Table 1.

Detection and characterisation of *CYP2D6*64* and *CYP2D6*65*

For *CYP2D6*64*, the 6.6-kb-long PCR product encompassing the *CYP2D6* gene and 1.5 kb of upstream region was cloned with the TOPO XL PCR cloning kit (Invitrogen, Carlsbad,

CA, USA). For *CYP2D6*65*, a 5-kb-long nested PCR product was generated due to poor yield of the 6.6-kb amplicon and cloned. Clones were genotyped for 100C>T and 1023C>T (*CYP2D6*64*) and 100C>T and 2850C>T (*CYP2D6*65*), respectively, to identify those carrying inserts of genes with the novel single nucleotide polymorphism (SNP) haplotypes. Subsequently, clones were entirely sequenced with appropriately spaced primers, DYEnamic ET dye terminator chemistry and a MegaBACE 500 capillary sequencer (GE). The sequences were compared with previously analysed allelic variants and allele designation obtained from the nomenclature committee (http://www.cypalleles.ki.se/).

To directly genotype genomic DNA for *CYP2D6*64* (100T, 1023T), allele-specific amplification was carried out with primers *64-F and *64-R. In order to control for PCR performance, a second primer pair was included in the reaction [internal control (IC) primers *64 IC-F and *64 IC-R]. *CYP2D6*64*-negative samples produced only the IC amplicon of 775 bp, whereas *CYP2D6*64*-positive samples produced two bands, the 775-bp IC and the 956-bp-long diagnostic *CYP2D6*64* amplicon. Assay details including primer sequences and PCR reaction conditions are given in Table 1. Of note, addition of the IC amplicon enhanced the specificity of the assay as it suppressed weak, unspecific amplification of the *CYP2D6*64* amplicon, which otherwise occurred in samples positive for either or both *CYP2D6*10* (100T, 1023C) and *CYP2D6*17* (100C, 1023T). Assay performance was validated on appropriate positive and negative samples.

Activity score assignment

The activity score (AS) is the sum of two values assigned to each allele reflecting their relative activity. The AS system has recently been described in detail elsewhere [11]. In this study, values for AS model A have been applied. Briefly, a value of 1 was given to the fully functional reference *CYP2D6*1* allele and 0 to nonfunctional alleles. Alleles carrying gene duplications or multiplications received double the value compared with that assigned to an allele with a single gene copy (e.g. *CYP2D6*2xN* received a value of 2). Reduced-activity alleles received a value of 0.5 to reflect reduced activity. For example, subjects with *CYP2D6*1/*29* and *CYP2D6*2xN/*5* genotypes received an AS of 1.5 and 2, respectively.

Results

CYP2D6 allele frequencies

DNA samples from 99 subjects who identified themselves as Coloureds were tested for over 30 allelic *CYP2D6* variants to provide a detailed genetic analysis and allow for the most accurate phenotype prediction possible with current knowledge. Table 2 summarises the allele frequencies determined in Coloureds and gives a comparison to those found in Caucasian and African American populations previously analysed [11]. Specifically, some alleles, such as *CYP2D6*4* and *41*, were lower compared with those in Caucasians but higher than those in African Americans. A number of alleles present in Caucasians and/or African Americans were absent (e.g. *CYP2D6*3, *6, *9*, etc.), while the *CYP2D6*5* gene deletion allele was detected at a relatively high frequency of 17.2%.

Detailed analysis of the *CYP2D6*5* gene deletion allele

To confirm the *CYP2D6*5* results obtained with a modified assay [11], all samples identified as either homozygous or heterozygous for a *CYP2D6*5* allele were reanalysed by amplifying a larger, approximately 6-kb-long, fragment, including the exon 9 portion of *CYP2D7* (Fig. 1). This assay was similar to that described by Fukuda et al. [15] which detects not only *CYP2D6*5* but also *CYP2D6* genes harbouring *CYP2D7*-derived downstream sequences. Such *CYP2D6* genes would appear as false-positives in *CYP2D6*5* assays targeting the intergenic region and producing amplicons ranging between 2.9 and 3.5 kb in length (i.e. assays used by many investigators for routine testing). However, no such *CYP2D6* gene structures were detected. All samples were 100% concordant between the *CYP2D6*5* assay utilised by Fukuda et al. [15] and our own modified assay. The observed *CYP2D6*5* genotype frequencies were in Hardy–Weinberg equilibrium.

Characterisation of *CYP2D6*66* and strategy for hybrid gene detection

Samples presenting with a homozygous genotype (e.g. *CYP2D6*1/*1* or *2/*2*) and negative for *CYP2D6*5* were subsequently scrutinised for the presence of hybrid genes such as *CYP2D6*13* or *16* (Fig. 1). The gold standard for their detection is long-range PCR that produces an 8-kb amplicon. In four samples, 8-kb PCR product was present, suggesting the presence of a hybrid gene. One of those samples produced a 1.1-kb product with a *CYP2D7*-specific primer binding to intron 6 and a *CYP2D6*-specific primer binding to exon 9 (Figs. 1 and 2). This assay detects hybrids with breakpoints upstream and downstream of respective primer binding sites (i.e. *CYP2D7* intron 6 and *CYP2D6* exon), including *CYP2D6*16*. The PCR product of the positive sample was partially sequenced and compared with a Caucasian reference allele, which has been designated *CYP2D6*66* by the nomenclature committee and deposited in GenBank under accession number

Table 2 Allele frequencies in Caucasians, African Americans and South African Coloureds

	Caucasians		African Americans		Coloureds	
Subjects	*n*=347		*n*=272		*n*=99	
Alleles	*n*=694	Freq %	*n*=544	Freq %	*n*=198	Freq %
**1*	253	36.46	179	32.90	52	26.77
**2*	3	0.43	45	8.27	20	10.10
**2A* (incl **35*)	156	22.48	31	5.70	10	5.05
**3*	7	1.01	1	0.18	0	0
**4*	137	19.74	21	3.86	14	7.07
**5*	23	3.31	35	6.43	31	17.17
**6*	7	1.01	3	0.55	0	0
**7*	1	0.14	0	0	0	0
**8, *11*	0	0	0	0	0/0	0/0
**9*	21	3.03	1	0.18	0	0
10* (incl **36*+10*)	15	2.16	16	2.94	4	2.53
**14*	0	0	0	0	1	0.51
**16* or **66*	0/1	0.14	0	0	0/1	0.51
**17*	3	0.43	104	19.12	25	12.63
**29*	1	0.14	41	7.54	9	4.55
**36*	0	0	3	0.55	0	0
**40*	0	0	3	0.55	0	0
**41*	56	8.07	10	1.84	7	3.54
**42*	0	0	3	0.55	0	0
**43*[2]	0	0	2	0.37	n/d	n/d
**45/46*	0	0	16	2.94	8	4.04
**56A/B*	0/0	0/0	0/1	0/0.18	0	0
**64*	0	0	0	0	2	1.01
**65*	0	0	0	0	1	0.51
**1×N*	2	0.29	2	0.37	3	1.52
**2×N*	1	0.14	3	0.55	2	1.01
**2A×N* (incl **35×N*)	4	0.58	6	1.10	0	0
**4×N*	3	0.43	15	2.76	0	0
**10×N*	0	0	1	0.18	0	0
**41×N*	1	0.14	0	0	0	0
**45×N*	0	0	1	0.18	0	0
Hybrid genes	n/d	n/d	n/d	n/d	3	1.52
Total	694	100	544	100	198	100
Nonfunctional alleles	179	25.8	85	15.6	50	26.8

The allele frequencies for the African American and Caucasian populations have previously been published [11]
Freq frequency

EU093102. Both sequences had a *CYP2D7* intron 6, which is clearly distinguishable from that of *CYP2D6*. Also, the first nucleotide of exon 7 (T) corresponded to *CYP2D7* (*CYP2D6* exhibits a G at the corresponding position 4776 in M33388), suggesting that the *CYP2D7* sequence extends into that exon. Beyond M33388 position 4873, however, the sequence was identified as *CYP2D6*. This narrows the region of hybridisation to 96 nucleotides between positions 4776 (*CYP2D7*) and 4873 (*CYP2D6*). The remaining three samples were negative in the *CYP2D6*16*-specific assay.

To improve hybrid gene detection and characterisation, a 5-kb-long PCR product was amplified from all samples that produced an 8-kb amplicon. This product was then further genotyped for the presence of the *CYP2D7* hallmark T insertion in exon 1 and a series of key nucleotides in exon 9 that identify *CYP2D6*. As shown in Fig. 3, the 5-kb template of all four samples with hybrid structures and the *CYP2D6*66* reference control had the *CYP2D7* 137insT and produced the *CYP2D6*-specific primers from exon 9.

Detection and characterisation of *CYP2D6*64* and **65*

Genotyping results also revealed deviation from known SNP haplotypes in three samples. These cases genotyped as following for three key SNPs. Cases 1 and 2: 100C/T, 1023C/T, 2850C/C; case 3: 100C/T, 1023C/C, 2850T/T.

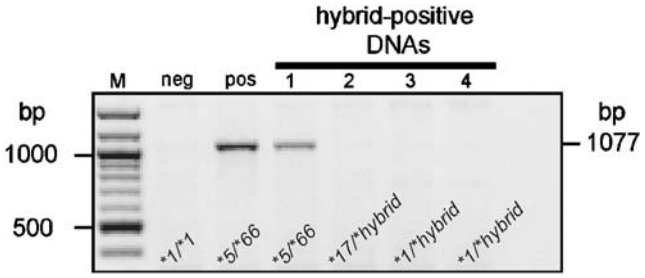

Fig. 2 Allele-specific amplification of *CYP2D6*16* and **66*. Primers binding to *CYP2D7* intron 6 and *CYP2D6* exon 9 were used to amplify a 1077-bp-long diagnostic fragment. As depicted in Fig. 1, this assay amplifies all hybrid structures that were created by recombination within the primer binding sites, including *CYP2D6*16* and **66*. No amplification was observed for *CYP2D6*1/*1* (*neg* negative control) and *CYP2D6*5/*5* (not shown) DNA samples. *Lane 1*: this sample produced hybrid-specific 8-kb- and 5-kb-long amplicons and produced the 1,077-bp polymerase chain reaction (PCR) product. *CYP2D6*66* was confirmed by partial sequence analysis and comparison with the fully sequenced reference DNA (*CYP2D6*5/*66*; *pos*, positive control). *Lanes 2–4*: these samples produced hybrid-specific 8-kb- and 5-kb-long amplicons, but not the 1,077-bp PCR product. The genotypes of all samples are as indicated. *M* 100-bp ladder

For the first two cases, 2850T was absent in the DNA carrying the *CYP2D6*17* SNP. In the third case, 2850T appeared to be located on a *CYP2D6*10* allele. To characterise these alleles, the *CYP2D6*-specific 6.6-kb-long range PCR product was cloned and respective alleles of interest identified by genotyping and entirely sequenced. Complete allele information and a comparison with other alleles is given in Fig. 4. The novel allele found in cases 1 and 2 was not a *CYP2D6*17* allele that lacked 2850T but had 100T and 1023T in a new haplotype. This allele was designated *CYP2D6*64* by the nomenclature committee. The novel allele in case 3 was confirmed to be a *CYP2D6*10* that had acquired 2850T and was designated *CYP2D6*65*.

Subsequently, an assay based on allele-specific amplification amplifying the 100T/1023T haplotype was developed. As shown in Fig. 5, a *CYP2D6*64*-specific PCR product and the internal control amplicon were generated only in the genomic DNA of the index case and clones derived thereof. DNAs carrying *CYP2D6*10* and/or **17* produced only the internal control amplicon. It should be noted that amplification specificity was achieved in the presence of the internal control amplicon and that some unspecific product is produced from *CYP2D6*10* and **17* in the absence of the internal control amplicon.

CYP2D6 genotypes and phenotype prediction

Of the 198 alleles scrutinised, 50 (26.8%) were genotyped as nonfunctional (*CYP2D6*4, *5, *14, *66* or other hybrids, Table 2), and three subjects were identified to be homozygous for two nonfunctional alleles. Even though the observed incidence of the PM phenotype was lower at 3.03% (n=3) than predicted by allele frequencies (6.7%, n=6.6) (Table 3), there was no statistically significant deviation from the Hardy–Weinberg equilibrium for genotypes with two, one or no nonfunctional alleles (p=0.17). Five alleles (2.53%) had duplications or multiplications of functional genes (*CYP2D6*1*×*N* and **2*×*N*), predicting ultrarapid metabolism in 2.3% subjects, i.e. those also having a fully functional second allele (*CYP2D6*1* or **2*) or having two *CYP2D6*1*×*N* or **2*×*N* alleles.

Phenotype prediction using the activity score

Figure 6 presents the distribution of AS groups in the study population in comparison with those observed in Caucasians and African Americans. A high proportion of subjects in the

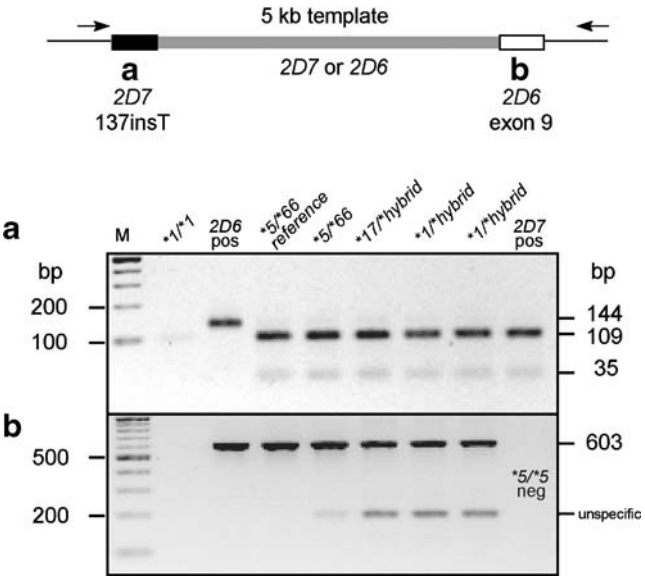

Fig. 3 Improved strategy to identify and characterise *CYP2D7/2D6* hybrid genes. The graph depicts the generic hybrid gene structure, i.e. *CYP2D7*- and *CYP2D6*-derived sequences in exons 1 and 9, respectively (indicated in *black and white boxes*) and *CYP2D7* or *CYP2D6* regions in between (*gray box*). Amplification of a 5-kb-long fragment with *CYP2D7*- and *CYP2D6*-specific primers encompasses the entire hybrid gene and serves as template to genotype key sequence variations in *CYP2D7* exon 1 [137 T insertion (insT)] and *CYP2D6* exon 9 (series of nucleotide variations). **a** Polymerase chain reaction restriction fragment length polymorphism (PCR-RFLP) assay detecting 137insT. *CYP2D7*-derived PCR fragments carrying the T insertion are cut by *Hind*III, whereas *CYP2D6*-derived sequences remain intact. **b** *CYP2D6* exon-9-specific PCR. PCR and PCR-RFLP fragment lengths are as indicated. *2D6 pos* 6.6 kb *CYP2D6* long-range amplicon of a *CYP2D6*1/*1* subject was used as positive control; *2D7 pos* plasmid containing the entire *CYP2D7* gene was used as positive control; **5/*5 neg* genomic DNA of a *CYP2D6*5/*5* subject was used as negative control; **5/*66 reference* the *CYP2D6*66* gene was entirely sequenced and subject's DNA used as positive control. *CYP2D7* exon 1 and *CYP2D6* exon 9 sequences were detected in all subjects that produced the 8-kb and 5-kb hybrid-specific amplicons. Amplification of a smaller, unspecific PCR product was only observed in the four study samples

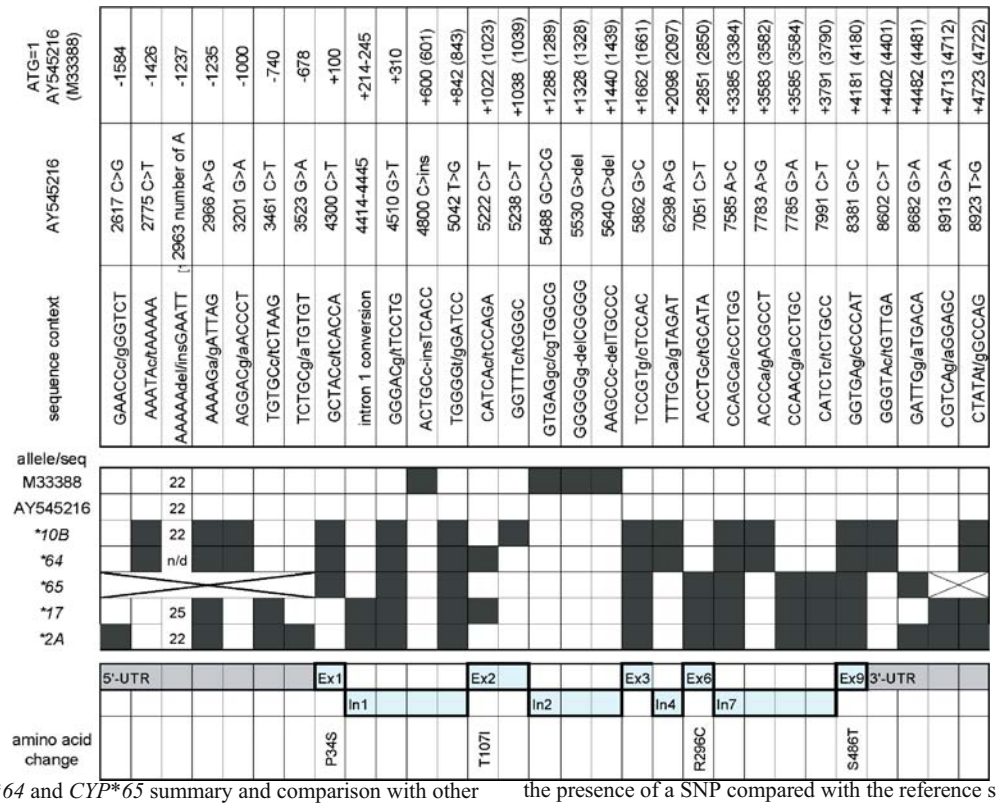

Fig. 4 *CYP2D6*64* and *CYP*65* summary and comparison with other alleles. A 6.6-kb-long polymerase chain reaction (PCR) product encompassing the entire gene region (*CYP2D6*64*) and a ~5-kb nested product thereof (*CYP2D6*65*) were cloned and sequenced. M33388 and AY545216 represent reference sequences. Sequences for *CYP2D6*2, *10B* and **17* were previously sequenced and are shown to demonstrate the presence of single nucleotide polymorphisms (SNPs) that are shared among respective alleles. *Black boxes* indicate the presence of a SNP compared with the reference sequences. The *top panels* provide the sequence context for each SNP and their respective positions in the reference sequences. Gene regions [5′-UTR, 3′-UTR, exons (*Ex*) and introns (*In*)] and amino acid changes for non-synonymous SNPs are as indicated in the *bottom panels*. *X* denotes the region for which no sequence was obtained for *CYP2D6*65*, and *n/d* indicates that the number of As could not be resolved

AS groups of 0.5 and 1 was noted for Coloureds. Whereas the number of subjects with an AS of 0.5 corresponded well with the number predicted, the number of subjects in the AS-1 group was higher than expected (41% vs. 33%).

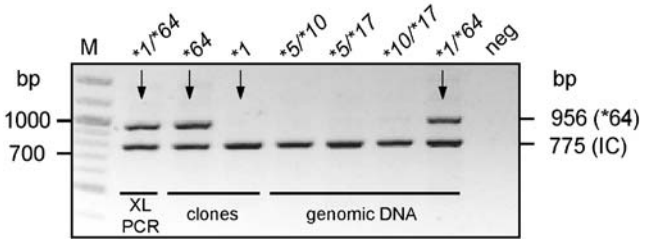

Fig. 5 Allele-specific polymerase chain reaction (PCR) assay to detect *CYP2D6*64*. Allele-specific PCR amplified from alleles carrying the *CYP2D6*64*-defining 100T/1023T haplotype. An internal control (*IC*) PCR product was coamplified to ensure assay performance, i.e. all samples produced the IC, whereas only *CYP2D6*64* carriers generated the diagnostic 956-bp-long fragment. The diagnostic *CYP2D6*64* amplification product was generated from the 6.6-kb-long genotyping product (*XL PCR*) as well as genomic DNA of the index case and the second carrier (not shown). Clones representing the two alleles of the index case amplified as expected, further demonstrating the linkage between 100T and 1023T

Discussion

This is the first investigation of the *CYP2D6* locus in Coloureds of South Africa. Considering the unique heritage of this population group, along with the high degree of diversity among Africans in general [2], finding a unique *CYP2D6* allele composition, including some rare as well as novel alleles, was not surprising. Notable is the relatively high frequency of the *CYP2D6*5* allele of 17.2%, one of the highest ever observed in any population [2]. Comparably high frequencies, albeit observed in small numbers of subjects, were recently also reported by Sistonen et al. [2] for the San (n=14, 14.3%), an indigenous people living in South Africa; the southeastern and southwestern Bantu (n= 16, 18.3%), but interestingly not the northeastern Bantu (n= 24, 4.2%). Frequencies of ≥10% for *CYP2D6*5* were also reported for the sub-Saharan Mbuti, but also other central/ south Asian populations. This high frequency of *CYP2D6*5* in Coloureds, however, does not lead to an exceptionally high frequency of PM, as the cumulative frequency of all other nonfunctional allelic variants is low, at 9.6%. In fact, the only other loss-of-function alleles observed were

Table 3 Observed and expected allele and genotype frequencies

Observed/predicted	Percent (*n*)	Expected by HWE	Percent (*n*)
Observed PM (*0/*0)	3.03 (3)	Expected PM	6.6 (6.6)
Observed heterozygous (*0/*other)	45.5 (45)	Expected heterozygous	38.3 (37.9)
Observed homozygous (*other/*other)	51.5 (51)	Expected homozygous (*other/*other)	55.2 (54.6)

PM poor metabolisers; *0, null or nonfunctional allele, *other, fully functional or reduced function allele

*CYP2D6*4* (7.1%), *CYP2D6*14* (0.5%) and hybrid genes (2.0%).

Given the unexpectedly high frequency of *CYP2D6*5* alleles, we systematically reviewed our procedures to rule out any experimental artifacts. First, the majority of *CYP2D6*5* carriers also produced a *CYP2D6*-derived 6.6-kb fragment indicating not only heterozygosity but also demonstrating that the DNA sample is of good quality and capable of supporting XL-PCR amplification. Secondly, all potential homozygous *CYP2D6*5/*5* subjects were tested for hybrid genes to discriminate between true homozygous *CYP2D6*5/*5* and *CYP2D6*5/*hybrid* genotype assignments. Third, all DNA samples that produced only faint XL-PCR products or failed to amplify *CYP2D6* products and/or internal control products were further evaluated by *CYP2D6*-specific assays, which do not rely on XL-PCR amplification. Those samples interpreted as *CYP2D6*5/*5* were consistently negative for any *CYP2D6*-derived amplification, whereas heterozygous *CYP2D6*5* samples produced consistent results for assays performed on both the XL-PCR template and directly from genomic DNA. Finally, all potential *CYP2D6*5* carriers had concordant results in two independent *CYP2D6*5* assays.

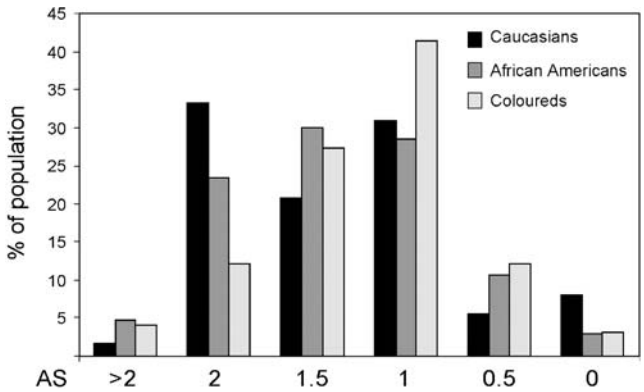

Fig. 6 Activity score (AS) distribution in Coloureds, Caucasians and African Americans. AS is the sum of the value assigned to each allele reflecting its activity [11]. Each AS group contains subjects with genotypes that confer similar CYP2D6 activity. Subjects with an AS of 0 (no functional alleles) are poor metabolisers, whereas an AS >2 on the other end of the distribution contains genotypes with one functional allele and one allele carrying a functional gene duplication. A subject with an AS ≥0.5 has a certain probability of presenting with an ultrarapid, extensive or intermediate metaboliser phenotype, as described in detail elsewhere [11]. Coloureds are predicted to have overall lower CYP2D6 activity (also referred to as a right shift) compared with Caucasians due to a higher frequency of reduced activity alleles. Coloureds, *n*=99 (*open bars*); Caucasians, *n*=347 (*gray bars*); African Americans, *n*=272 (*black bars*)

In the *CYP2D6*5* allele, the entire *CYP2D6* gene, along with flanking upstream and downstream regions, has been lost, whereas hybrid genes are characterised by a large deletion that removes partial gene sequences along intergenic sequences and fuses the remaining 5′-*CYP2D7* and 3′-*CYP2D6* portions [8]. Frequency data for hybrid genes, such as *CYP2D6*13* and **16*, are scarce, which is likely due to the cumbersome task of generating an approximately 8-kb-long PCR product [8]. Because the *CYP2D6*5* deletion was relatively abundant in Coloureds, we suspected that hybrid genes may also be present. Indeed, four potential carriers were identified, but a definitive allele call could not be made due to inconsistent amplification of the 8-kb product. However, the resequenced control DNA performed well under the assay conditions, suggesting that amplification difficulties may be due to other factors, e.g. residual impurities left behind during the phenol-based DNA extraction procedure. To overcome these problems and simplify hybrid gene characterisation, a shorter 5-kb PCR product was generated and subsequently genotyped for key sequence variations in exon 1 (*CYP2D7*) and exon 9 (*CYP2D6*). This approach confirmed all four subjects as hybrid carriers. One was subsequently assigned as *CYP2D6*66*, whereas the remaining three samples defied further characterisation by sequence analysis (i.e. sequence traces appeared to be composites of multiple sequence templates present in the reaction, which are suspected to be nonsensical/artificial amplicons formed during the PCR process). Nonetheless, the combined approach of amplifying a hybrid-specific 5-kb product and subsequent genotype analysis proved to be a valid approach, as demonstrated on the fully characterised *CYP2D6*66* DNA (Fig. 3).

The DNA now defined as *CYP2D6*5/*66* was initially consistent with a *CYP2D6*5/*16* genotype [8] (also see http://www.cypalleles.ki.se/). It produced the 8-kb-long hybrid-specific PCR product off the hybrid allele, a 9.5-kb-long *CYP2D6*5*-specific product and was also positive in an assay that detects hybrid structures with recombination points between *CYP2D7* exon 6 and *CYP2D6* exon 9 (Fig. 1). Resequencing of the entire hybrid gene of this phenotypic PM [13] revealed, however, that the recombination point is

located within the first half of exon 7 (the first nucleotide in exon 7 corresponds to *CYP2D7* exon 7, nucleotide 98 to *CYP2D6*, whereas the nucleotides in between match both genes) and not within intron 7, exon 8 or intron 8, as postulated by Daly et al. for *CYP2D6*16* [8]. One of our study subjects perfectly matched this hybrid sequence and was assigned *CYP2D6*66*. Additional hybrid genes previously genotyped as *CYP2D6*16* are being investigated to determine whether they also harbour *CYP2D6*66*-like sequence or indeed are *CYP2D6*16*. Unfortunately, no sequence was deposited for *CYP2D6*16* for more detailed comparisons.

Since both *CYP2D6*5/*5* and *CYP2D6*5/hybrid* genotypes cause poor metabolism, such subjects would not benefit from additional hybrid testing. However, any subjects presenting with a homozygous genotype for functional or partially functional alleles, such as *CYP*1*1, *10/*10* or *17*/17*, etc., could conceivably carry a nonfunctional hybrid and their phenotype prediction would change accordingly. Also, depending on the assay methodology, hybrid genes could cause unusual or weird genotyping results or allele calls. Consequently, knowledge and awareness about hybrid genes and their structures in general will aid in assay result interpretation and allow initiation of targeted additional testing if necessary.

The novel *CYP2D6*64* and *65* alleles may also be products of recombination, as they are hybrids between certain *CYP2D6* alleles or haplotypes. As shown in Fig. 4, the former is a hybrid between *CYP2D6*10* and *17* and the latter between *CYP2D6*10* and *2*. The impact of the novel haplotypes on function remains unknown. Since 2850C>T ($R^{296}C$) does not appear to diminish activity in the CYP2D6.2 isoenzyme when compared with CYP2D6.1 [16], we would not anticipate this SNP to diminish the activity of CYP2D6.65 but expect characteristics similar to that seen for CYP2D6.10.

We have not found any evidence of *CYP2D6*64* or *CYP2D6*65* (i.e. presence of 100T and 1023T in the absence of 2850T or with any other SNP) in any Caucasian or African American subjects previously studied. Regarding *CYP2D6*64*, a sample heterozygous for 100C>T, 1023C>T and 2850C>T would be interpreted as *CYP2D6*10/*17* (100T assigned to *CYP2D6*10*, 1023T and 2850T assigned to *CYP2D6*17*) according to current nomenclature. Heterozygosity for these SNPs is also compatible with a *CYP2D6*2/*64* genotype assignment, and only *CYP2D6*64*-specific testing would allow discrimination. The function of *CYP2D6*64* remains to be further investigated, but one may speculate that it may have properties of both *CYP2D6*10* (reduced activity due to unstable protein) and *CYP2D6*17* (level of reduction appears to be substrate dependent) [17]. Its impact on phenotype prediction may be limited, as we found no *CYP2D6*10/*17* candidate subjects among African Americans, suggesting that *CYP2D6*64* is rare in this population (>0.001). This allele may predominantly be found in Coloureds and possibly population(s) that contributed to their ethnic admixture. Therefore, routine testing for the *CYP2D6*64* allele does not appear to be warranted. Since we encountered only a single *CYP2D6*65* allele, the challenge of testing for linkage of SNPs that are distant, and lack of evidence that 2850T would alter protein function, we did not further pursue this allele in our population samples. Understanding *CYP2D6*64* and *CYP2D6*65* may, however, be helpful in interpreting genotyping results where the presence of respective SNPs cannot be reconciled with known haplotypes.

Based on the frequency of nonfunctional and reduced-function alleles and corresponding AS group distribution in Coloureds, one would expect a mean dextromethorphan/dextrorphan urinary metabolite ratio that is higher (i.e. lower CYP2D6 activity) compared with that observed in Caucasians and African Americans. However, whether this prediction, as suggested by the observed number of subjects in the AS groups, is accurate requires confirmation in an independent population sample that is both genotyped and phenotyped. Observed numbers of genotypes and consequently the number of subjects in AS groups 1 and 2 that deviated from those expected may be under- and over-estimated, respectively. One possible explanation may be the relatively small number of 99 subjects in the study, or the sampled population may comprise subjects of different lines of heritage (e.g. Cape Coloureds and Coloureds of KwaZulu-Natal). No information about the heritage was collected under the approved protocol, as all samples were anonymised to protect personal information, and collection of detailed demographic data would benefit future studies.

The incidence of HIV infection in South Africa is one among the highest in the world. In 2006, the national estimated HIV prevalence was 10.8% across all ethnicities and a staggering 29.1% among pregnant woman http://www.avert.org/safricastats.htm, accessed Aug 3, 2007). It appears that HIV infection can impact CYP2D6 activity on a population basis as well as have intraindividual variability. O'Neil et al. [18] reported an overall right shift towards lower activity in a mostly Caucasian HIV-positive cohort as well as phenotype switching events in a number of subjects. This phenomenon was also observed by Werner et al. (*BJCP*, manuscript submitted). Such a shift may have dramatic consequences in populations with high proportions of IMs, including Coloureds (HIV prevalence 1.9%) and other black South African populations. HIV patients with genotypes leading to reduced baseline CYP2D6 activity may have a higher risk of switching to a PM phenotype and encounter dose-dependent adverse drug reactions. However, the impact of HIV goes beyond CYP2D6, as this is only one of many drug-metabolising

enzymes whose activity may be altered via cytokines such as interleukin (IL)-1β, IL-6, tumor necrosis factor (TNF)-α and interferon (IFN)-α or γ [19].

To date little is known about phenotype and genotype and their relationship of phase I and II drug-metabolising enzymes in South Africans of any ethnicity. This study is a first step in that direction. The data presented not only demonstrate that Coloureds are unique in respect to the complement of *CYP2D6* alleles present and their frequencies, but also underpin the importance of characterising unique admixed populations. Clearly, personalised medicine in a country such as South Africa faces many challenges but also presents many opportunities to improve drug therapy.

Acknowledgements We thank Dr. JS Leeder for his support, in-depth discussions and reviewing the manuscript, and Mrs. L Ndjountché for technical assistance.

The study was supported in part by funds provided by the Division of Clinical Pharmacology & Medical Toxicology (AG) and the Department of Hematology and Cell Biology, University of the Free State and the National Research Foundation of South Africa (CC).

References

1. Zanger UM, Raimundo S, Eichelbaum M (2004) Cytochrome P450 2D6: overview and update on pharmacology, genetics, biochemistry. Naunyn-Schmiedeberg's Arch Pharmacol 369:23–37
2. Sistonen J, Sajantila A, Lao O, Corander J, Barbujani G, Fuselli S (2007) CYP2D6 worldwide genetic variation shows high frequency of altered activity variants and no continental structure. Pharmacogenet Genomics 17:93–101
3. Kirchheiner J, Nickchen K, Bauer M, Wong M-L, Licinio J, Roots I et al (2004) Pharmacogenetics of antidepressants and antipsychotics: the contribution of allelic variations to the phenotype of drug response. Mol Psychiatry 9:442–472
4. Spina E, de Leon J (2007) Metabolic drug interactions with newer antipsychotics: a comparative review. Basic Clin Pharmacol Toxicol 100:4–22
5. Ingelman-Sundberg M (2004) Genetic polymorphisms of cytochrome P4502D6 (CYP2D6): clinical consequences, evolutionary aspects and functional diversity. Pharmacogenomics J 5:6–13
6. Gaedigk A, Bradford LD, Marcucci KA, Leeder JS (2002) Unique CYP2D6 activity distribution and genotype-phenotype discordance in African Americans. Clin Pharmacol Ther 72:76–89
7. Wennerholm A, Johansson I, Hidestrand M, Bertilsson L, Gustafsson LL, Ingelman-Sundberg M (2001) Characterization of the CYP2D6*29 allele commonly present in a black Tanzanian population causing reduced catalytic activity. Pharmacogenetics 11:417–427
8. Daly AK, Fairbrother KS, Andreassen OA, London SJ, Idle JR, Steen VM (1996) Characterization and PCR-based detection of two different hybrid CYP2D7P/CYP2D6 alleles associated with the poor metabolizer phenotype. Pharmacogenetics 6:319–328
9. Panserat S, Mura C, Gerard N, Vincent-Viry M, Galteau MM, Jacq-Aigrain E 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
10. Gaedigk A, Ndjountché L, Divakaran K, Bradford LD, Zineh I, Oberlander TF et al (2007) Cytochrome P4502D6 (CYP2D6) gene locus heterogeneity: characterization of gene duplication events. Clin Pharmacol Ther 81:242–251
11. Gaedigk A, Simon SD, Pearce RE, Bradford LD, M.J. K, Leeder JS (2007) The CYP2D6 activity score: Translating genotype information into a qualitative measure of phenotype. Clin Pharmacol Ther DOI 10.1038/sj.clpt.6100406
12. Gaedigk A, Bradford LD, Alander SW, Leeder JS (2006) CYP2D6*36 gene arrangements within the CYP2D6 locus: Association of CYP2D6*36 with poor metabolizer status. Drug Metab Dispos 34:563–569
13. Gaedigk A, Gotschall RR, Forbes NS, Simon SD, Kearns GL, Leeder JS (1999) Optimization of cytochrome P4502D6 (CYP2D6) phenotype assignment using a genotyping algorithm based on allele frequency data. Pharmacogenetics 9:669–682
14. Gaedigk A, Gaedigk R, Leeder JS (2005) CYP2D7 splice variants in human liver and brain: Does CYP2D7 encode functional protein? Biochem Biophys Res Comm 336:1241–1250
15. Fukuda T, Maune H, Ikenaga Y, Naohara M, Fukuda K, Azuma J (2005) Novel structure of the CYP2D6 gene that confuses genotyping for the CYP2D6*5 allele. Drug Metab Pharmacokinet 20:345–350
16. Marcucci KA, Pearce RE, Crespi C, Leeder JS, Gaedigk A (2002) Characterization of cytochrome P450 2D6.1 (CYP2D6.1), CYP2D6.2 and CYP2D6.17 activities toward model CYP2D6 substrates dextromethorphan, bufuralol, and debrisoquine. Drug Metab Dispos 30:1–7
17. Shen H, He MM, Liu H, Wrighton SA, Wang L, Guo B et al (2007) Comparative metabolic capabilities and inhibitory profiles of CYP2D6.1, CYP2D6.10, and CYP2D6.17. Drug Metab Dispos 35:1292–1300
18. O'Neil WM, Gilfix BM, Markoglou N, Di Girolamo A, Tsoukas CM, Wainer IW (2000) Genotype and phenotype of cytochrome P450 2D6 in human immunodeficiency virus-positive patients and patients with acquired immunodeficiency syndrome. Eur J Clin Pharmacol 56:231–240
19. Aitken AE, Richardson TA, Morgan ET (2006) Regulation of drug-metabolizing enzymes and transporters in inflammation. Annu Rev Pharmacol Toxicol 46:123–149
20. Johansson I, Lundqvist E, Dahl M-L, Ingelman-Sundberg M (1996) PCR-based genotyping for duplicated and deleted CYP2D6 genes. Pharmacogenetics 6:351–355