ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

GENOMICS

Genomics 81 (2003) 166–174

www.elsevier.com/locate/ygeno

High-throughput analysis of informative *CYP2D6* compound haplotypes

Benjamin Fletcher,[a] David B. Goldstein,[b] Amanda L.R. Bradman,[a,c] Michael E. Weale,[a] Neil Bradman,[a] and Mark G. Thomas[a,*]

[a] *The Centre for Genetic Anthropology, Department of Biology and Department of Anthropology, University College London, The Darwin Building, Gower St., London WC1E 6BT, UK*

[b] *Galton Laboratory, Department of Biology, University College London, The Darwin Building, Gower St., London WC1E 6BT, UK*

[c] *School of Medicine, University of Leeds, Leeds LS2 9JT, UK*

Received 1 July 2002; accepted 19 November 2002

Abstract

We describe a high-throughput protocol for detecting key polymorphisms in the drug-metabolizing enzyme gene *CYP2D6* and a number of linked microsatellites that is both fast and relatively inexpensive to perform. This approach employs GeneScan technology to enable a researcher to determine rapidly the status of seven simple nucleotide polymorphisms in *CYP2D6* and also to assay repeat number variation at five closely linked dinucleotide microsatellite loci. The method requires only three PCRs and two GeneScan runs per sample. We anticipate that this will be of value to researchers in three different ways: (1) rapid discrimination of common *CYP2D6* alleles, (2) high-resolution haplotyping for association studies involving chromosome 22q13.1 using microsatellite variation, and (3) generation of compound haplotypes for investigating the evolution of *CYP2D6* variation. We also report compound haplotype frequencies for an Ashkenazi Jewish and a British sample.

Keywords: CYP2D6; Pharmacogenetics; Genotyping; Haplotype; Microsatellites; SNPs; Jews; British; Population; Variation

It is now clear that much of the observed variation in drug efficacy and safety has a hereditary basis, arising from polymorphisms in genes coding for drug-metabolizing enzymes (DMEs). The frequency of DME alleles can vary greatly between different human populations, and this has implications for both medical and evolutionary studies. Consequently, there has been growing academic and commercial interest in the extent, nature, and causes of DME variation. However, it is currently unknown to what extent interpopulation variability is due to selection, drift, or other processes. Studies seeking to elucidate this have been held back by the absence of efficient low-cost high-throughput procedures for characterizing variability at the molecular level in and around DME loci. Ultimately, the identification of the molecular basis of pharmacogenetically relevant variation combined with the technology to screen individuals for the presence of specific alleles allows for the prediction of drug response and individualized treatment.

Polymorphism in debrisoquine/sparteine oxidation is arguably the most highly studied pharmacogenetic trait. The DNA sequence encoding the enzyme has been localized to the 4.3-kb, nine-exon cytochrome P450 2D6 (*CYP2D6*) gene found at chromosome 22q13.1. To date more than 48 mutations and 53 alleles of *CYP2D6* have been characterized in European populations [1]. The poor metabolizer (PM) phenotype follows an autosomal recessive pattern of inheritance [2]. An allele duplication consisting of multiple functional copies of *CYP2D6* confers the ultrarapid metabolizer (URM) phenotype [3]. Individuals who demonstrate normal levels of *CYP2D6* activity are referred to as extensive metabolizers. Intermediate metabolizers (IMs) typically produce lower than normal levels of functional enzyme.

Medical interest in *CYP2D6* polymorphism arises because the *CYP2D6* enzyme has been found to process more than 30 prescribed and over-the-counter drugs, including antiarrhythmics, antihypertensives, β blockers, monoamine

* Corresponding author. Fax: +44-0-20-7679-7096.
E-mail address: m.thomas@ucl.ac.uk (M.G. Thomas).

0888-7543/03/$ – see front matter
doi:10.1016/S0888-7543(02)00031-9

oxidase inhibitors, morphine derivatives, antipsychotics, and tricyclic antidepressants [4]. The extreme variation in the activity of *CYP2D6* caused by these polymorphisms can produce an undesired response to drug therapies in a number of ways. In the case of individuals with the PM phenotype, prodrugs may not be converted to the active form in sufficient quantity to be therapeutic, e.g., absence of the analgesic effect of codeine when it is not converted to morphine [5]. Alternatively, failure to clear a drug may lead to its accumulation and induce an adverse drug reaction (hyperresponse). When individuals with the URM phenotype are administered normal clinical doses of drugs, they clear them too quickly. In the latter case, the drug is either absent from blood serum or present at subtherapeutic quantities with the consequence that subjects require correspondingly higher doses of drugs to receive therapeutic benefit, e.g., therapeutic failure of normal antidepressant doses [6]. In the case of the IM phenotype, status has been hypothesized to be potentially important for drugs with narrow therapeutic windows. Several studies have suggested that variation in *CYP2D6* can affect neurological disease susceptibility [7] as well as being a potential modulator of cancer risk [8]. Evolutionary interest in *CYP2D6* arises because of the primary role that DMEs have in the metabolism of dietary toxins such as alkaloids [9].

Microsatellites are short tandemly repeated nucleotide motifs of two to five bases that are found throughout the genomes of eukaryotes, e.g., $(CA)_n$. Many of these loci have been found to be highly polymorphic, with variation occurring in the number of repeat units in different alleles [10]. Consequently, microsatellites are useful in many types of modern genetic study, including forensics [11], phylogenetic reconstruction [12], dating mutational events [13], demographic history [14], linkage analysis [15], testing for selection [16], mapping complex genetic diseases [17], and identifying quantitative trait loci [18]. More recently, compound haplotypes made up of combinations of microsatellites and "simple nucleotide polymorphisms" (SNPs) have been used extensively to differentiate populations [19], study human migration in prehistory [20], date events in prehistory [21], and estimate the age of mutations [22]. We use the definition of SNPs as simple nucleotide polymorphisms, in preference to the more commonly used "single nucleotide polymorphisms," as it is inclusive of indel mutations. The study of compound haplotypes is likely to assume increasing importance given that recent studies suggest that linkage disequilibrium is highly structured into discrete "blocks" separated by recombination hot spots [23].

The construction of multiplex PCR systems has been widely used to increase the throughput and decrease the cost of typing large numbers of SNPs and microsatellite loci [24]. Here we describe a method for typing seven previously reported SNPs in *CYP2D6* and five closely linked microsatellite loci. Selection of polymorphisms for inclusion in the *CYP2D6* SNP protocol was based on: (1) identification of key mutations that result in altered *CYP2D6* activity and (2) mutual compatibility of markers within the multiplex system. The SNP protocol was designed to ensure that five distinct loci could be amplified by multiplex PCR, digested with a cocktail of restriction enzymes, and then typed by the presence or absence of DNA fragments of specific size and dye label on an ABI-377 automated sequencer or equivalent for GeneScan analysis. GeneScan is an integrated genotyping technology platform developed by Applied Biosystems (http://home.appliedbiosystems.com/) consisting of an electrophoretic gel tank (either slab or capillary based) combined with a laser detection system and analysis and laboratory information management software. Several primers were modified to incorporate mismatches at a single position to create a restriction enzyme recognition site when the naturally occurring polymorphism alone did not result in the presence/absence of a restriction site.

The SNPs detected are numbered as indicated in the standardized *CYP2D6* nomenclature (http://www.imm.ki.se/CYPalleles/CYP2D6.htm): C100T (Pro34Ser), C1023T (Thr107Ile), G1846A (splicing defect), 1863 +9-bp repeat (172-174FRP rep), A2549del (frameshift), 2613-2615AGAdel (Lys281del), and C2850T (Arg296Cys). These polymorphisms can be used to distinguish between *CYP2D6*2* [3], *CYP2D6*3* [25], *CYP2D6*4* [26], *CYP2D6*9* [27], *CYP2D6*10* [28], *CYP2D6*17* [29], and *CYP2D6*30* [1]. *CYP2D6*3* and **4* alleles either fail to produce enzymes or produce defective ones with no activity. *CYP2D6*9*, **10*, and **17* produce enzymes with reduced activity relative to the wild type, *CYP2D6*1*. The *CYP2D6*2* allele, and to a lesser extent the *CYP2D6*1* and *CYP2D6*4* alleles, are sometimes duplicated. In the cases of *CYP2D6*2* and *CYP2D6*1*, allele duplication confers the URM phenotype. The phenotypic effect of *CYP2D6*30* has not yet been fully characterized.

The five previously reported dinucleotide microsatellites, *D22S276*, *D22S279*, *D22S284*, *D22S423*, and *CYP2D8P*, are all less than 1 cM from the *CYP2D* gene cluster and have been used in association studies to explore the role of *CYP2D6* in neurological disease [30]. The typing of these markers in families should be useful in creating high-resolution haplotypes for linkage disequilibrium mapping studies to explore variation in *CYP2D6* or other genes localized to 22q13.1. This approach has already been successfully employed for numerous genetic disorders including Batten disease [31], Huntington's disease [32], and cystic fibrosis [33]. Furthermore, the *CYP2D6*4* allele has already been shown to be associated with long-type (repeat length $\geq$22) *CYP2D8P* alleles [34].

The multiplex PCR protocols required considerable optimization. Primers for all loci had to be designed to ensure that all products fell within a limited size range (75–400 bp) that was suitable for GeneScan analysis and that the different amplicons and their associated restriction enzyme digestion products could be discriminated using a combination of fragment size and fluorescent ABI dye label (HEX, TET, and FAM). The method requires only three PCRs and

Table 1
SNP protocol: expected and observed product sizes after multiplex enzyme digestion

Polymorphism in *CYP2D6*	Allele(s)	Discriminating enzyme	Cut size (bp)		Uncut size (bp)		Dye label
			Expected	Observed[a]	Expected	Observed[a]	
100 C→T	*4, *10	*Nhe*I	79 (C)	79.5	94 (T)	94.5	TET
1023 C→T	*17	*Xba*I	89 (T)	92.0	108 (C)	110.0	HEX
1846 G→A	*4	*Avr*II	292 (G)	291.5	307 (A)	306.5	FAM
2549 A→del	*3	*Tat*I	221 (A)	221.5	243 (del)	243.5	HEX
2850 C→T	*2, *17	*Fsp*I	79 (C)	78.5	211 (T)	211.5	FAM

[a] Using ABI-377/TAMRA-350.

two runs per sample on an ABI-377 automated sequencer or equivalent. The resulting amplicons in these multiplex protocols are analyzed in the presence of GS-350 or GS-500 molecular size standards labeled with the fluorescent dye TAMRA (PE–Applied Biosystems). Due to the high degree of sequence similarity between *CYP2D6* and its two related pseudogenes, *CYP2D7* and *CYP2D8P*, a specific preamplification PCR is required to provide a *CYP2D6*-only template for subsequent amplification reactions. We also report frequencies of compound haplotypes, consisting of all the markers described above, for an Ashkenazi Jewish and a British sample.

Results and discussion

Both protocols have been employed successfully to amplify DNA from samples taken as blood or buccal swabs and extracted with standard phenol/chloroform procedures, silica/guanidinium thiocyanate methods, and commercial kits (Qiagen, Inc., Valencia, CA, USA). To date we have determined over 2500 compound genotypes and haplotypes comprising all 12 polymorphic markers. The SNP protocol was recently used to assay *CYP2D6* variation as part of a study into how human population genetic structure affects the evaluation of drug safety, efficacy, and response [35].

The protocols were designed to ensure, as far as possible, similar signal intensities following GeneScan analysis of all observed products. Protocol performance was optimized by the following: (1) careful primer design and the modification of previously published primer sequences so that all primer pairs within a multiplex amplify at a similar optimal annealing temperature, (2) varying primer concentrations and using the lowest concentration that would give peak heights of around 2000 units in GeneScan analysis, (3) varying PCR annealing temperatures and using the highest temperature that would allow reliable amplification of all regions, and (4) using the lowest $MgCl_2$ concentration permissible for amplification of all regions. In our experience, the use of TaqStart MAb increases the specificity of the PCR. As specificity is an important factor in optimizing multiplex amplification, we used this reagent in all reactions. Some variation was observed in intensity of signal peaks between different DNA samples. However, absence of clear peaks for one or more loci was observed only in samples containing DNA that was either severely degraded or present at very low concentrations. Prudent design and dye labeling of the primers to detect the 1846 G → A (*CYP2D6*4*) and 2549 A → del (*CYP2D6*3*) mutations combined with the superior resolution of GeneScan analysis also allowed for the detection of two further indel mutations in their amplicons: the 1855–1863 +9-bp insertion (*CYP2D6*30*) and the 2615 3-bp deletion (*CYP2D6*9*). Subjects with these mutations produce digestion products with sizes slightly different from those observed in wild type: a 301-bp rather than 292-bp FAM-labeled product for *CYP2D6*30* and a 218-bp rather than 221-bp HEX-labeled product for *CYP2D6*9*.

The sizing of DNA fragments using GeneScan technology is not absolute [24]. The actual observed product sizes assigned to PCR products using GeneScan analysis software, although consistent across runs, can differ from the predicted product size by up to 6 nucleotides. Therefore, accurate scoring of SNPs and microsatellite repeat lengths using the protocols required calibration against DNA samples of known microsatellite repeat number and *CYP2D6* genotype. Observed product sizes for the SNPs were generally in good agreement with the predicted sizes (within ±0.5 bp), the exception being the 1023 C → T-containing product used to detect *CYP2D6*17* appearing 2 bp larger than expected (Tables 1 and 2). Observed microsatellite repeat lengths for the CEPH (http://www.cephb.fr/) sample 1347-02 are provided in Table 3 as a standard. Comparable results should be attainable using other slab gel- or capillary-based systems. DNA standards are also available from the authors.

Sample data generated by the protocols to produce compound haplotypes are presented in Table 4. Data were ob-

Table 2
SNP protocol: additional indels detected

Polymorphism in *CYP2D6*	Allele	Size (bp)		Dye label
		Expected	Observed[a]	
1855–1863 +9-bp insertion	*30	301	300.5	FAM
2615 AGA deletion	*9	218	218.5	HEX

[a] Using ABI-377/TAMRA-350.

Table 3
Microsatellite protocol: observed microsatellite repeat scores for CEPH standard sample 1347-02

Locus	Repeat length of allele
D22S276	220–226
D22S279	124–126
D22S284	116–116
D22S423	252–252
CYP2D	340–342

tained from Ashkenazi Jewish and British families consisting of both parents and one or more offspring, but only data on the unrelated parents are shown. Phase of parental haplotypes was assigned from their offspring's genotype data according to the method outlined in [36]. The Ashkenazi Jewish data on unrelated parents were used to test for linkage disequilibrium between each microsatellite locus and the alleles detected at the *CYP2D6* locus, using the exact test described in [37]. The results were *D22S279* ($p = 0.080$), *D22S423* ($p = 0.101$), *D22S284* ($p = 0.210$), *CYP2D8P* ($p < 0.0001$), and *D22S276* ($p = 0.0002$). We note that the order of increasing p values matches the order of loci according to the June 2002 Freeze of the human genome physical map (midpoints relative to the first base of *CYP2D6: D22S279* = −1.50 Mb, *D22S423* = −2.14 Mb, *D22S284* = −2.21 Mb, *CYP2D8P* = 14.7 kb, *D22S276* = −0.51 Mb), but not the order of loci according to the genetic map presented in [32] (relative to *CYP2D6: D22S279* = −0.9 cM, *D22S423* = −0.6 cM, *D22S284* = −0.3 cM, *CYP2D8P* = −0.0 cM, *D22S276* = +0.4 cM). When we compared the Ashkenazi Jewish and British unrelated parents using only *CYP2D6* allele frequencies, we found no significant difference between the two samples (using the exact test of [37], $p = 0.493$). However, when we added the haplotype information from the five microsatellite loci, we were able to distinguish the two samples ($p = 0.047$), illustrating the greater resolving power available with the compound haplotypes.

We compared our data with previously published data on *CYP2D6* allele frequencies in Western Europeans ($n = 3292$) [38] and Saudi Arabians ($n = 202$) [9]. Our British sample did not differ significantly from the Western European sample ($p = 0.487$ using exact test), despite the very high Western European sample size. *CYP2D6*4,* the primary cause of PM phenotype, was found at a frequency of 15.9% in the Ashkenazi. This is slightly below the levels found in the British (19.8%) and the Western Europeans (20.0%) but considerably in excess of levels in Saudi Arabians (3.0%). This difference is reflected in a lower F_{ST} value between Ashkenazi and British ($F_{ST} = -0.005$) and between Ashkenazi and Western Europeans ($F_{ST} = 0.002$) than between Ashkenazi and Saudi Arabians ($F_{ST} = 0.042$). The greater similarity of Ashkenazi Jews to Western Europeans (including the British) may be due to inward gene flow into the Ashkenazi Jewish community from surrounding populations [38].

The protocol described here has advantages over previous typing methods in (1) being more cost effective, being less labor intensive, and producing a higher throughput than simplex PCR/RFLP methods; (2) amplifying from low-yield DNA sources such as buccal swab samples as well as high-yield DNA sources such as blood and tissue; (3) requiring less investment of time and financial resources than other high-throughput typing technologies such as GeneChip CYP450 Assay Arrays (Affymetix, Inc., Santa Clara, CA, USA); and (4) that there are no licensing issues.

We anticipate several uses for these protocols. On its own the *CYP2D6* SNP protocol enables a researcher to quickly and cheaply characterize individuals for the most common CYP*2D6* alleles. When combined with previously published methods to detect *CYP2D6*5* alleles (that are not compatible with multiplexing), this approach can identify ~90% of individuals in Western European populations with PM phenotype [39]. *CYP2D6*4* is also found at relatively high frequencies in several non-European populations: 8.3% in Canadian Inuit [40], 8.5% in African Americans [41], and 11.3% in Turks [42]. Due to the low incidence of defective alleles in Africans and Oriental Asians, PM phenotypes are rare in these populations compared with Western Europeans [43]. However, the remaining two alleles that are detected by the protocol and confer reduced levels of metabolism are common in African and Oriental Asian populations. In published datasets for Asian populations *CYP2D6*10* ranges from 40% in Japanese [44] to 70% in Taiwanese Chinese [45]; *CYP2D6*17* ranges in published datasets for African populations from 9% in Ethiopians [46] to 34% in Zimbabweans [47]. The genotypic information that the protocol provides can be used in conjunction with known *CYP2D6* allele frequencies in different ethnic groups as a guide to deciding which other polymorphisms should be characterized, depending on whether a study is concerned with identifying URM or IM phenotypes. For example, *CYP2D6*2* alleles can be further characterized into three distinct haplotypes that confer different enzyme activities: the multiduplicated $CYP2D6*2_{(n)}$ [3], *CYP2D6*2* [−1496G], and *CYP2D6*2* [−1496C] [48]. The multiduplicated *CYP2D6*2* and variants cause the overexpression of CYP2D6 enzyme that results in the URM phenotype: *CYP2D6*2* [−1496G] has several amino acid differences from the wild type (*CYP2D6*1*), but comparable enzyme activity; the variant with the 5′ flanking sequence −1496C mutation in the presumed transcription initiation site causes the gene to be transcribed less efficiently and the *CYP2D6*2* [−1496C]/*0 (null) genotype is believed to be responsible for at least 50–60% of IM phenotypes in Europeans.

In stand-alone use the *CYP2D6* microsatellite protocol provides a cost-effective way of providing high-resolution haplotypes for association studies of the 22q13.1 region. Genes that have been found to be within a few centimorgans

Table 4
Compound haplotypes in Ashkenazi Jewish and British samples

CYP2D6 allele	Microsatellite haplotype[a]	Ashkenazi	British	*CYP2D6* allele	Microsatellite haplotype[a]	Ashkenazi	British
*CYP2D6*1*	15 25 22 27 18	1	—		19 21 21 26 17	1	—
	17 19 18 19 17	1	—		19 23 24 19 17	—	1
	17 19 22 21 17	—	1		19 24 21 19 17	1	—
	17 19 23 19 17	1	—		19 25 18 19 18	1	—
	17 19 25 19 21	—	1		19 25 22 19 17	1	1
	17 19 25 26 17	1	—		19 25 23 19 17	2	1
	17 20 18 19 18	1	1		19 25 23 19 18	1	2
	17 20 19 19 17	1	—		19 25 23 20 18	2	—
	17 20 22 19 17	—	1		19 25 23 25 17	1	—
	17 20 23 19 21	1	—		19 25 23 28 18	—	1
	17 20 23 20 18	1	—		19 26 23 19 18	1	—
	17 20 24 19 21	—	1		19 26 25 19 17	1	—
	17 20 25 19 21	—	1		20 17 18 19 18	—	1
	17 21 17 27 18	—	1		20 17 18 19 21	1	—
	17 21 26 19 18	1	—		20 19 24 19 17	—	1
	17 21 26 21 22	1	—		20 20 18 19 17	—	1
	17 22 20 19 17	1	1		20 21 21 19 17	—	1
	17 22 20 19 22	—	1		20 25 23 19 17	1	1
	17 23 19 24 18	—	1		20 25 23 19 21	1	1
	17 23 22 20 17	—	1		20 25 28 19 22	—	1
	17 23 23 19 17	2	—		21 17 23 19 17	1	—
	17 23 23 20 21	1	—		21 17 23 19 21	1	—
	17 24 22 19 17	—	1		21 23 18 19 21	1	—
	17 24 23 19 17	—	1		21 26 25 21 20	4	1
	17 25 23 19 17	—	2	*CYP2D6*2*	15 25 23 20 20	—	1
	17 25 23 19 18	1	—		17 17 23 28 18	—	1
	17 25 23 19 21	1	—		17 18 23 20 21	—	1
	17 25 23 20 18	—	1		17 18 24 20 17	2	1
	17 25 23 20 21	—	1		17 18 24 24 17	1	—
	17 25 23 25 17	1	—		17 19 18 19 17	2	—
	17 26 18 19 18	—	1		17 19 18 20 17	—	1
	17 26 23 20 21	1	—		17 19 22 20 18	—	1
	17 29 23 19 17	1	—		17 19 23 19 17	2	—
	18 19 18 21 22	—	1		17 19 23 20 17	2	—
	18 19 23 19 16	1	—		17 19 23 20 21	—	1
	18 19 23 19 21	1	—		17 19 25 20 18	—	1
	18 19 23 27 18	1	—		17 20 20 21 22	—	1
	18 19 24 27 18	—	2		17 20 21 20 18	3	—
	18 20 18 19 17	1	—		17 20 25 19 16	—	1
	18 20 20 20 18	1	—		17 21 17 27 18	—	1
	18 20 23 19 17	—	1		17 21 22 20 17	1	—
	18 20 23 19 21	1	—		17 21 23 20 17	—	1
	18 20 23 20 21	2	—		17 21 26 21 22	2	—
	18 20 25 19 17	—	1		17 24 28 22 17	1	—
	18 21 21 19 21	1	—		17 25 22 19 17	1	—
	18 21 23 19 21	1	—		17 25 23 18 24	1	—
	18 23 26 27 21	—	1		17 29 23 19 17	1	—
	18 24 22 19 18	—	1		18 16 22 20 19	—	1
	18 25 21 19 21	—	1		18 19 19 19 16	—	1
	18 25 24 19 17	1	—		18 19 19 20 18	—	1
	18 28 20 19 22	—	1		18 19 23 20 16	—	1
	19 19 18 19 17	—	1		18 19 23 20 21	1	—
	19 19 18 19 18	—	1		18 20 22 21 17	1	—
	19 19 22 19 17	—	1		18 20 23 20 21	1	—
	19 19 22 19 21	—	1		18 20 23 21 19	—	1
	19 19 22 20 18	1	—		18 21 20 19 18	—	1
	19 19 23 19 18	1	1		18 21 23 21 19	—	1
	19 20 18 25 17	1	—		18 25 20 20 17	—	1
	19 20 23 19 17	2	—		18 25 23 26 20	1	—
	19 20 23 19 21	1	—		19 17 18 21 21	3	—
	19 20 23 20 21	1	—		19 18 18 27 17	—	1
	19 20 24 19 18	1	—		19 18 23 20 17	—	1

(continued on next page)

Table 4 *(continued)*

CYP2D6 allele	Microsatellite haplotype[a]	Ashkenazi	British	*CYP2D6* allele	Microsatellite haplotype[a]	Ashkenazi	British
	19 19 18 19 17	—	1		17 20 24 26 19	1	—
	19 19 19 20 19	1	—		17 20 25 27 18	—	1
	19 20 23 19 16	1	—		17 21 17 27 18	—	1
	19 20 23 20 18	—	1		17 22 20 26 18	—	1
	19 20 23 21 17	1	—		17 24 23 27 18	—	1
	19 20 23 23 17	1	—		17 25 22 27 18	2	—
	19 21 21 20 17	—	1		17 28 20 26 18	—	1
	19 24 23 20 21	1	—		18 19 22 27 18	—	1
	19 24 23 20 25	1	—		18 19 23 21 18	1	—
	19 24 24 20 21	1	—		18 19 23 27 18	3	—
	19 25 23 19 17	—	1		18 19 24 27 18	—	2
	19 25 23 20 17	—	1		18 19 24 28 18	—	1
	19 25 23 20 18	1	—		18 20 18 27 18	1	—
	19 25 23 20 21	2	1		18 20 21 27 18	1	—
	19 25 23 24 17	1	—		18 20 22 19 18	1	—
	19 26 23 19 17	1	—		18 20 22 27 17	1	—
	19 26 23 20 21	1	—		18 20 22 27 18	2	1
	20 19 25 20 17	—	1		18 20 23 27 18	—	2
	20 21 21 19 17	—	1		18 24 23 27 18	—	1
	20 23 23 20 21	—	1		18 25 20 26 17	—	1
	20 23 25 19 17	—	1		18 25 22 27 18	—	1
	20 24 21 20 18	—	1		19 17 18 27 18	—	1
	20 24 24 20 16	—	1		19 19 23 27 18	1	—
	20 25 18 20 18	—	1		19 20 19 23 17	1	—
	21 17 23 19 21	1	—		19 20 23 27 17	1	—
	21 19 18 20 21	1	—		19 23 23 26 18	1	—
	21 26 25 21 20	1	—		19 25 23 25 17	1	—
*CYP2D6*3*	19 25 23 19 21	—	1		21 26 25 27 18	1	—
	19 27 25 19 17	—	1	*CYP2D6*5*[b]	17 25 23 27 18	1	—
*CYP2D6*4*	17 17 20 20 18	—	1		18 16 22 19 18	—	1
	17 19 20 27 18	1	—		18 19 23 27 18	1	—
	17 20 18 27 18	—	1	*CYP2D6*9*	19 26 23 19 17	—	1
	17 20 22 20 21	—	1	*CYP2D6*17*	17 22 19 24 18	1	—
	17 20 23 25 18	—	1		17 23 21 24 18	1	—
	17 20 23 27 17	—	1	Total		126	106

Note. The haplotypes presented are for unrelated parents only. The genotypes of offspring were used to resolve phase.

[a] Microsatellite haplotype comprises repeat sizes for microsatellite loci in the following order: D22S279, D22S423, D22S284, CYP2D8P, D22S276.

[b] *CYP2D6*5* alleles are not detected in the SNP protocol but were inferred from the pedigrees.

of this region include somatosatin receptor, interleukin-2 receptor b, platelet-derived growth factor-β polypeptide, adenylosuccinate lyase, *N*-acetylgalactosaminidase, thyroid autoantigen-ku, aconitase hydroxylase, diaphorase–cytochrome-b_5 reductase, and peripheral benzodiazepine receptor [49]. Haplotype phase can be either assigned via the use of family samples [36] or inferred by statistical techniques.

Using the two protocols in concert to generate compound haplotypes provides researchers with a new way to explore evolutionary and clinical questions arising from *CYP2D6* variation. It has been proposed that the high degree of variation in *CYP2D6* may arise from microenvironmental heterogeneity in diet causing diversifying selection that would favor different levels of *CYP2D6* activity in populations with different dietary histories [50]. The compound haplotype data that can be generated by the protocols permits researchers to address this question through estimating the age of mutations affecting enzyme function using coalescence theory [22] and testing for selection acting on the different alleles using an intraallelic variability-based approach [16].

In clinical studies, it should be possible to use the protocols described here to find new *CYP2D6* variants by identifying distinct genealogical clades of microsatellite haplotypes that are associated with poor metabolism, but not with known defective alleles. While in the case of coding region variants direct sequencing is likely to be more cost-effective, microsatellite haplotypes would be particularly useful in a search for noncoding region variants that cause poor metabolism. An association between specific alleles of the *CYP2D8P* microsatellite locus and the *CYP2D6*4* allele has been shown previously [34], and we would expect compound haplotypes consisting of two or more closely linked microsatellites to show an even greater association with *CYP2D6* alleles. For example, microsatellites *CYP2D8P* and *D22S276* are both located within approximately 500 kb of the *CYP2D6* gene, and we note that haplotypes composed of these two markers define different

Table 5
Primer sequences and concentrations for *CYP2D6* microsatellite protocol

Primer name	Sequence (5′–3′)	Dye label	Final conc. (μM)
D22S279-u	GAT CCA GCC TGT GTA TCA GAA T	—	0.43
D22S279-l	TTT TGG TGT TAG AGT GGT GTT ATC	FAM	0.43
D22S423-u	CAA GAG CAT CTG TGA GAC AAC TT	—	0.60
D22S423-l	GAG TGA GTG ACT GAG TAA ATG TAG TG	TET	0.60
D22S284-u	CCC TCC TGA AGT CAG ATG GA	—	0.35
D22S284-l	GAG CAA GAC CCT GTC TCA AGA	HEX	0.35
CYP2D-u	CCC TCC TGA AGT CAG ATG GA	FAM	0.65
CYP2D-l	GAG CAA GAC CCT GTC TCA AGA	—	0.65
D22S276-u	AAA TGG GCT TGT AAA G AA AAA T A	—	0.20
D22S276-l	AGT GTC CTT CAG TTC CTC CTC T	FAM	0.20

high-frequency modal clusters (the modal haplotype and its one mutation step neighbors) in the three major *CYP2D6* alleles identified in Ashkenazi Jews. All three modal cluster frequencies are well in excess of what would be expected assuming no linkage and taking into account the observed microsatellite allele frequencies (*CYP2D6*1* = 45% observed versus 26% expected, $n = 60$, $p < 0.0032$; *CYP2D6*2* = 29% observed versus 16% expected, $n = 42$, $p = 0.043$; *CYP2D6*4* = 75% observed versus 11% expected, $n = 20$, $p < 0.001$). Given the apparent block-like recombination recently identified in the autosomal genome [23] assemblages of well-characterized blocks comprising both SNPs and microsatellites may add considerable power to such studies in the future.

Materials and methods

Multiplex PCR amplification of microsatellite sequences using the CYP2D MS protocol

Amplification reactions were performed in a BioMetra (Uno II) thermal cycler using a 10-μl reaction volume. Reaction conditions were 200 μM dNTPs, 10 mM Tris–HCl (pH 9.0), 0.1% Triton X-100, 0.01% gelatin, 50 mM KCl, 2.2 mM MgCl, 0.13 units *Taq* polymerase (HT Biotech, Cambridge, UK), 9.3 nM TaqStart monoclonal antibody (Mab) (Clontech), and primers to the concentrations given in Table 5. Cycling parameters were 5 min at 95°C followed by 20 cycles of 1 min at 94°C, 1 min at 52°C, 1 min at 72°C; 15 cycles of 1 min at 94°C, 1 min at 51.5°C, 1 min at 72°C; 10 cycles of 1 min at 94°C, 1 min at 51°C, 1 min at 72°C; and then a final incubation step of 72°C for 10 min.

PCR preamplification of CYP2D6-specific fragment

Amplification reactions were performed in a BioMetra (Uno II) thermal cycler using a 10-μl reaction volume. Reaction conditions were 5% DMSO, 200 μM dNTPs, 10 mM Tris–HCl (pH 9.0), 0.1% Triton X-100, 0.01% gelatin, 50 mM KCl, 2.2 mM MgCl, 0.13 units *Taq* polymerase (HT Biotech), 9.3 nM TaqStart Mab (Clontech), and primers to the concentrations given in Table 2. Cycling parameters were 5 min at 95°C followed by 35 cycles of 10 s at 94°C, 1 min at 63°C, 1 min at 72°C, and then a final incubation step of 72°C for 10 min.

Multiplex PCR amplification from XL template with the CYP2D6 SNP protocol

Amplification reactions were performed in a BioMetra (Uno II) thermal cycler using a 10-μl reaction volume. Reaction conditions were 200 μM dNTPs, 10 mM Tris–HCl (pH 9.0), 0.1% Triton X-100, 0.01% gelatin, 50 mM KCl, 2.2 mM MgCl, 0.13 units *Taq* polymerase (HT Biotech), 9.3 nM TaqStart Mab (Clontech), primers to the concentrations given in Table 6, and 1 μl of 1/100 dilution of XL PCR product as template. Cycling parameters were 5 min at 95°C followed by 35 cycles of 10 s at 94°C, 1 min at 62°C, 1 min at 72°C, and then a final incubation step of 72°C for 10 min.

Multiplex restriction enzyme digestion of SNP PCR products using the CYP2D6 protocol

Digestions were carried out in 384-well microtiter plates in a final volume of 8 μl. Each reaction consisted of 2 μl of SNP PCR product, 1× concentration Tango/Y+ buffer (Fermentas, Inc., Hanover, NH, USA), 0.33 u *Nhe*I, 1.00 u *Xba*I, 0.22 u *Avr*II, 0.2 u *Tat*I, and 0.33 u *Fsp*I. Plates were incubated at 37°C overnight. Predicted sizes and associated polymorphism status for each dye-labeled PCR digestion product are given in Tables 1 and 2.

GeneScan analysis of microsatellite and SNPs

The microsatellite PCR products and SNP digestion products were run on an ABI-377 automated sequencer: 1.0-μl aliquots of the microsatellite PCR products or the SNP digestion products were mixed with 2.0 μl of loading buffer (formamide:dextran blue:TAMRA-labeled size standard, in the ratio 12:2:1). PCR products from both protocols required electrophoresis on a 36-cm, 5% gel, with a 2-h run

Table 6
Primer sequences and concentrations for XL PCR and *CYP2D6* SNP protocol

Primer name	Sequence (5′–3′)	Dye label	Final conc. (μM)
CYP2D6-XL-u	TGC TCC TGG TGG ACC TGA TGC	—	0.50
CYP2D6-XL-l	GTC CGG CCC TGA CAC TCC TT	—	0.50
*CYP2D6*2*-u	CCG TTC TGT CCC GAG TAT GCT C	FAM	0.13
*CYP2D6*2*-l	TCG GCC CCT GCA CTG TTT	—	0.13
*CYP2D6*3/9*-u	AGC TGG ATG AGC TGC TAA CTG AG**T**	—	0.12
*CYP2D6*3/9*-l	CCC CAA ATG ACC TCC AAT TCT G	HEX	0.12
*CYP2D6*4/30*-u	CCG CAT CTC CCA CCC C**T**	—	0.65
*CYP2D6*4/30*-l	TGG GGT CTC CTG GAA TGT CCT T	FAM	0.65
*CYP2D6*10*-u	ACG CTG GGC TGC AcG CTA **G**	—	0.18
*CYP2D6*10*-l	GAA GCA GTA TGG TGT GTT CTG GAA GT	TET	0.09
*CYP2D6*17*-u	GTG GTC GTG CTC AAT GGG CT	HEX	0.05
*CYP2D6*17*-l	CCC GAA ACC CAG GAT CT**A** G	—	0.08

Note. Boldface indicates mismatches designed into primer.

time for the SNP products (GS-350) and a 2-h 30-min run time for the microsatellite products (GS-500).

Notes on preparation of PCR and digestion mixes

All PCR reagents except the *Taq* polymerase and TaqStart Mab were premixed in batches sufficient for 96 reactions and stored at −20°C. The *Taq* and TaqStart Mab were mixed as 2 parts of 5 u/μl *Taq:*1 part 7 μM TaqStart Mab and stored at −20°C in 20-μl aliquots. Primers were also mixed and stored as a 10× stock to save time and reduce errors associated with pipetting small volumes. To minimize the time that the *Taq* enzyme was in contact with primers and other reagents, 1 μl of DNA template was first placed at the bottom of each sample's 0.2-ml PCR tube. Only then was the PCR premix, containing all primers and buffer reagents, thawed out. The *Taq*/TaqStart mix was added to the other PCR components just prior to amplification and the mixture vortexed thoroughly. Nine microliters of the PCR mix was pipetted into the lid of each PCR tube. Finally, the reaction tubes were placed in a thermal cycler.

For convenience and to minimize freeze–thawing, digestion buffer and water were premixed in batches sufficient for 96 digests and stored at −20°C. Separately, restriction enzymes were premixed and stored at −20°C. Two microliters of PCR product was aliquoted into the bottom of each microtiter well. The diluted buffer was then thawed out, and the restriction enzyme mix was added and then mixed thoroughly before use. Eight microliters of the digestion mix was pipetted into each microtiter well.

Acknowledgments

We thank Ann Daly (Department of Pharmacological Sciences, University of Newcastle-upon-Tyne, UK) and Ulrich Griese (Dr. Margarete Fischer-Bosch-Institut für Klinische Pharmakologie, Stuttgart, Germany) for supplying standards of known *CYP2D6* genotype. We thank Howard Cann (Fondation Jean Dausset–CEPH, France) for supplying standards of known allelic type for the microsatellites. B.F. was supported by a Biotechnology and Biological Sciences Research Council Cooperative Awards in Science and Engineering studentship awarded to the UCL Department of Biology (98/A2/G/04331).

References

[1] D. Marez, et al., Polymorphism of the cytochrome P450 CYP2D6 gene in a European population: characterization of 48 mutations and 53 alleles, their frequencies and evolution, Pharmacogenetics 7 (1997) 193–202.
[2] S. Kimura, M. Umeno, R.C. Skoda, U.A. Meyer, F.J. Gonzalez, 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 (1989) 889–904.
[3] I. Johansson, et al., Inherited amplification of an active gene in the cytochrome P450 CYP2D locus as a cause of ultrarapid metabolism of debrisoquine, Proc. Natl. Acad. Sci. USA 90 (1993) 11825–11829.
[4] D.W. Nebert, Polymorphisms in drug-metabolizing enzymes: what is their clinical relevance and why do they exist? Am. J. Hum. Genet. 60 (1997) 265–271.
[5] S.H. Sindrup, K. Brosen, The pharmacogenetics of codeine hypoalgesia, Pharmacogenetics 5 (1995) 335–346.
[6] L. Bertilsson, et al., Molecular basis for rational megaprescribing in ultrarapid hydroxylators of debrisoquine, Lancet 341 (1993) 63.
[7] A. Atkinson, et al., CYP2D6 is associated with Parkinson's disease but not with dementia with Lewy bodies or Alzheimer's disease, Pharmacogenetics 9 (1999) 31–35.
[8] C.R. Wolf, C.A. Smith, D. Forman, Metabolic polymorphisms in carcinogen metabolising enzymes and cancer susceptibility, Br. Med. Bull. 50 (1994) 718–731.
[9] R.A. McLellan, M. Oscarson, J. Seidegard, D.A. Evans, M. Ingelman-Sundberg, Frequent occurrence of CYP2D6 gene duplication in Saudi Arabians, Pharmacogenetics 7 (1997) 187–191.
[10] W. Messier, S.H. Li, C.B. Stewart, The birth of microsatellites, Nature 381 (1996) 483.
[11] A.K. Lindqvist, et al., Chromosome-specific panels of tri- and tetranucleotide microsatellite markers for multiplex fluorescent detection and automated genotyping: evaluation of their utility in pathology and forensics, Genome Res. 6 (1996) 1170–1176.
[12] A.M. Bowcock, et al., High resolution of human evolutionary trees with polymorphic microsatellites, Nature 368 (1994) 455–457.

[13] J.C. Stephens, et al., Dating the origin of the CCR5-Delta32 AIDS-resistance allele by the coalescence of haplotypes, Am. J. Hum. Genet. 62 (1998) 1507–1515.
[14] A. Di Rienzo, et al., Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories, Genetics 148 (1998) 1269–1284.
[15] M.R. Nelen, et al., Localization of the gene for Cowden disease to chromosome 10q22–23, Nat. Genet. 13 (1996) 114–116.
[16] M. Slatkin, G. Bertorelle, The use of intraallelic variability for testing neutrality and estimating population growth rate, Genetics 158 (2001) 865–874.
[17] R. Feakes, et al., Exploring the dense mapping of a region of potential linkage in complex disease: an example in multiple sclerosis, Genet. Epidemiol. 17 (1999) 51–63.
[18] P.J. Fisher, et al., DNA pooling identifies QTLs on chromosome 4 for general cognitive ability in children, Hum. Mol. Genet. 8 (1999) 915–922.
[19] S.A. Tishkoff, et al., Global patterns of linkage disequilibrium at the CD4 locus and modern human origins, Science 271 (1996) 1380–1387.
[20] T. Zerjal, et al., Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis, Am. J. Hum. Genet. 60 (1997) 1174–1183.
[21] M.G. Thomas, et al., Origins of Old Testament priests, Nature 394 (1998) 138–140.
[22] D.B. Goldstein, et al., Age estimates of two common mutations causing factor XI deficiency: recent genetic drift is not necessary for elevated disease incidence among Ashkenazi Jews, Am. J. Hum. Genet. 64 (1999) 1071–1075.
[23] D.B. Goldstein, Islands of linkage disequilibrium, Nat. Genet. 29 (2001) 109–111.
[24] M.G. Thomas, N. Bradman, H.M. Flinn, High throughput analysis of 10 microsatellite and 11 diallelic polymorphisms on the human Y-chromosome, Hum. Genet. 105 (1999) 577–581.
[25] M. Kagimoto, M. Heim, K. Kagimoto, T. Zeugin, U. Meyer, Multiple mutations of the human cytochrome P450IID6 gene (CYP2D6) in poor metabolizers of debrisoquine. Study of the functional significance of individual mutations by expression of chimeric genes, J. Biol. Chem. 265 (1990) 17209–17214.
[26] A.C. Gough, et al., Identification of the primary gene defect at the cytochrome P450 CYP2D locus, Nature 347 (1990) 773–776.
[27] R. Tyndale, et al., Identification of a new variant CYP2D6 allele lacking the codon encoding Lys-281: possible association with the poor metabolizer phenotype, Pharmacogenetics 1 (1991) 26–32.
[28] H. Yokota, et al., Evidence for a new variant CYP2D6 allele CYP2D6J in a Japanese population associated with lower in vivo rates of sparteine metabolism, Pharmacogenetics 3 (1993) 256–263.
[29] C. Masimirembwa, I. Persson, L. Bertilsson, J. Hasler, M. Ingelman-Sundberg, 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 (1996) 713–719.
[30] K. Wilhelmsen, et al., Is there a genetic susceptibility locus for Parkinson's disease on chromosome 22q13? Ann. Neurol. 41 (1997) 813–817.
[31] H.M. Mitchison, et al., Refined localization of the Batten disease gene (CLN3) by haplotype and linkage disequilibrium mapping to D16S288–D16S383 and exclusion from this region of a variant form of Batten disease with granular osmiophilic deposits, Am. J. Med. Genet. 57 (1995) 312–315.
[32] F.B. Atac, B. Elibol, F. Schaefer, The genetic analysis of Turkish patients with Huntington's disease, Acta Neurol. Scand. 100 (1999) 195–198.
[33] S. Raskin, et al., Cystic fibrosis in the Brazilian population: DF508 mutation and KM-19/XV-2C haplotype distribution, Hum. Biol. 69 (1997) 499–508.
[34] S. Tanaka, et al., Association of CYP2D microsatellite polymorphism with Lewy body variant of Alzheimer's disease, Neurology 50 (1998) 1556–1562.
[35] J.F. Wilson, et al., Population genetic structure of variable drug response, Nat. Genet. 29 (2001) 265–269.
[36] A. Nejati-Javaremi, C. Smith, Assigning linkage haplotypes from parent and progeny genotypes, Genetics 142 (1996) 1363–1367.
[37] M. Raymond, F. Rousset, An exact test for population differentiation, Evolution 49 (1995) 1280–1283.
[38] M.F. Hammer, et al., Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes, Proc. Natl. Acad. Sci. USA 97 (2000) 6769–6774.
[39] A. Gaedigk, et al., Optimization of cytochrome P4502D6 (CYP2D6) phenotype assignment using a genotyping algorithm based on allele frequency data, Pharmacogenetics 9 (1999) 669–682.
[40] M. Jurima-Romet, et al., CYP2D6-related oxidation polymorphism in a Canadian Inuit population, Can. J. Physiol. Pharmacol. 75 (1997) 165–172.
[41] W.E. Evans, et al., Genetic basis for a lower prevalence of deficient CYP2D6 oxidative drug metabolism phenotypes in black Americans, J. Clin. Invest. 91 (1993) 2150–2154.
[42] S. Aynacioglu, et al., Low frequency of defective alleles of cytochrome P450 enzymes 2C19 and 2D6 in the Turkish population, Clin. Pharmacol. Ther. 66 (1999) 185–192.
[43] C.M. Masimirembwa, J.A. Hasler, Genetic polymorphism of drug metabolising enzymes in African populations: implications for the use of neuroleptics and antidepressants, Brain Res. Bull. 44 (1997) 561–571.
[44] T. Tateishi, et al., Analysis of the CYP2D6 gene in relation to dextromethorphan O-demethylation capacity in a Japanese population, Clin. Pharmacol. Ther. 65 (1999) 570–575.
[45] M. Garcia-Barcelo, et al., Genetic analysis of the CYP2D6 locus in a Hong Kong Chinese population, Clin. Chem. 46 (2000) 18–23.
[46] E. Aklillu, et al., Frequent distribution of ultrarapid metabolizers of debrisoquine in an Ethiopian population carrying duplicated and multiduplicated functional CYP2D6 alleles, J. Pharmacol. Exp. Ther. 278 (1996) 441–446.
[47] C.M. Masimirembwa, I. Johansson, J.A. Hasler, M. Ingelman-Sundberg, Genetic polymorphism of cytochrome P450 CYP2D6 in Zimbabwean population, Pharmacogenetics 3 (1993) 275–280.
[48] S. Raimundo, et al., Elucidation of the genetic basis of the common 'intermediate metabolizer' phenotype for drug oxidation by CYP2D6, Pharmacogenetics 10 (2000) 577–581.
[49] U.J. Kim, et al., A bacterial artificial chromosome-based framework contig map of human chromosome 22q, Proc. Natl. Acad. Sci. USA 93 (1996) 6297–6301.
[50] L.F. Jorge, M. Eichelbaum, E.U. Griese, T. Inaba, T.D. Arias, Comparative evolutionary pharmacogenetics of CYP2D6 in Ngawbe and Embera Amerindians of Panama and Colombia: role of selection versus drift in world populations, Pharmacogenetics 9 (1999) 217–228.