Int J Mol Epidemiol Genet 2010;1(4):310-319
www.ijmeg.org /IJMEG1007002

Original Article

CYP2D6 and *CYP2C19* in Papua New Guinea: High frequency of previously uncharacterized *CYP2D6* alleles and heterozygote excess

Nicolas von Ahsen[*,1], Mladen Tzvetkov[*,2], Harin A Karunajeewa[3], Servina Gomorrai[4], Alice Ura[4], Jürgen Brockmöller[2], Timothy M E Davis[3], Ivo Mueller[4], Kenneth F Ilett[3], Michael Oellerich[1]

[1]Department of Clinical Chemistry, University Medicine Göttingen, Germany; [2]Department of Clinical Pharmacology, University Medicine Göttingen, Germany; [3]School of Medicine and Pharmacology, University of Western Australia, Australia; [4]Papua New Guinea Institute of Medical Research, Madang, Papua New Guinea. [*]Equal contributors.

Received July 7, 2010; accepted August 10, 2010; available online August 18, 2010

Abstract: Purpose: A high frequency of previously unknown *CYP2D6* alleles have been reported in Oceania populations. Genetic and functional properties of these alleles remain unknown. Methods: We performed analyses of the genetic variability of *CYP2D6* and *CYP2C19* genes using AmpliChip genotyping in cohorts from two distinct Papua New Guinea (PNG) populations (Kunjingini, n=88; Alexishafen, n=84) focussing on the genetic characterisation of PNG-specific alleles by re-sequencing. Results: Previously unknown *CYP2D6* alleles have population frequencies of 24% (Kunjingini) and 12% (Alexishafen). An allele similar to *CYP2D6**1, but carrying the 1661G>C substitution, was the second most frequent *CYP2D6* allele (20% Kunjingini and 10% Alexishafen population frequency). Sequencing suggests the *CYP2D6**1661G>C allele originated from a cross-over between *CYP2D6**1 and *2 and thus is predicted to confer fully active CYP2D6 enzyme. Two additional predicted full activity alleles [1661G>C;4180G>C] and 31G>A were found in the Kunjingini cohort (frequencies 3 c/c and 1%, respectively) and a novel predicted reduced activity allele [100C>T;1039C>T] was found in the Alexishafen cohort (frequency 2%). A high frequency of ultra-rapid (15%) and notably low frequencies of intermediate and poor CYP2D6 metabolizers (<5%) and a high frequency of poor CYP2C19 metabolizers were observed in PNG. Both *CYP2D6* and *CYP2C19* showed heterozygote excess that may be explained by exogamy and recent introduction of alleles by migration that are yet to reach HWE in relatively isolated populations. Conclusion: The *CYP2D6**1661 allele common in Oceania may be regarded as functionally equivalent to the full activity *CYP2D6*1* allele.

Keywords: *CYP2D6* gene polymorphism, *CYP2C19*, Papua New Guinea, novel alleles, heterozygote excess

Introduction

The Cytochrome P450 enzymes (CYP) are the major phase I enzymes involved in drug metabolism. Among CYPs, CYP2C9, CYP2D6 and CYP2C19 are highly polymorphic and together account for ~40% of hepatic phase I metabolism [1]. CYP2D6 contributes to the biotransformation of 20-25% of drugs in clinical use [2]. The aim of pharmacogenetics is to study the different responses to drugs based on inherited factors and to use this information to develop individualized therapy [3]. However, there are considerable regional and inter-ethnic differences in *CYP2D6* variant allele frequencies [4, 5]. In a group with European ancestry, the most common *CYP2D6* variant allele, *CYP2D6*4,* is associated with absent enzyme activity. In Asian populations, the *CYP2D6*10* variant and, in African populations, the *CYP2D6*17* variant are more relevant for reduced CYP2D6 activity. Ethiopian/Saudi Arabian populations have a particularly high frequency of *CYP2D6* gene duplications that lead to ultra-rapid metabolism of CYP2D6 substrates [2]. Sistonen et al. reported a ~30% frequency of a *CYP2D6**1661 allele in a small PNG cohort (n=17) but further characterisation of this allele has not been carried out [4]. *CYP2C19*2* and **3* deficiency alleles are known to be exceptionally frequent in Asian and

Pacific populations including Papua New Guinea (PNG) [6].

Since both CYP2D6 and CYP2C19 are involved in the metabolism of antimicrobial (including antiparasitic) agents that are in common use in developing countries such as PNG [7], we have assessed the genetic variation at these loci in two PNG populations. These findings should facilitate an understanding of *CYP2D6* genetic variation in populations from other parts of the world.

Patients and Methods

Study sites

The present study utilised baseline blood samples taken as part of a large clinical trial comparing the efficacy of different antimalarial treatments in children living in Kunjingini (East Sepik Province) and Alexishafen (Madang Province) (**Figure 1**) [8]. Both study sites are situated in the northern part of mainland PNG at latitude approximately 4° and have an equatorial climate with very high rainfall and humidity. However, they differ significantly with regard to migration history, lifestyle and accessibility.

Alexishafen is a coastal settlement situated 20km from the provincial capital Madang (population 30,000). The Madang north coast area was initially settled by people speaking both Austronesian and Papuan languages, with the major local language in the Alexishafen area being "Bel", a member of the Austronesian language family [8]. The development of Madang as a significant administrative centre and major port serving local industry (tuna fishing, copra and cocoa plantations) has resulted in significant transmigration of other Melanesian and Papuan ethnic groups into the Alexishafen area and subsequent inter-marriage. Most of the population around Alexishafen are subsistence farmers and fishermen.

Kunjingini is part of Wosera sub-district, East Sepik Province, situated inland on the alluvial plains between the Torricelli Range and the Sepik River [9] 75 km south-west of the provincial capital Wewak. The Wosera was settled by members of the Abelam ethnolinguistic group that speak Papuan Sepik languages. Most of the Wosera population is engaged in subsistence farming. Due to the high population density, the area has seen significant recent out-migration, while in-migration is rare. The Wosera is situated 340km NW of Alexishafen with two major rivers (Sepik and Ramu River) separating the two areas. Because there is no road link, migration between the two populations has been limited.

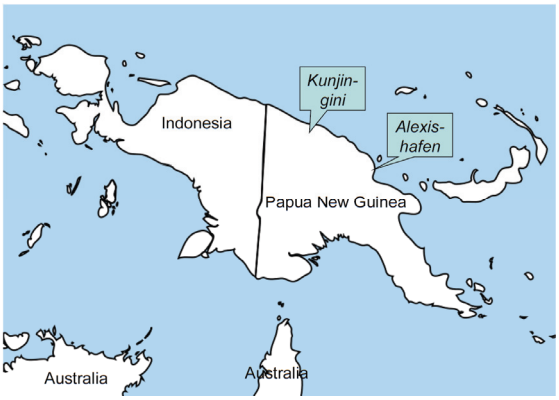

Figure 1. Location of studied populations. Map of Papua New Guinea with northern mainland sites Kunjingini (East Sepik Province) and Alexishafen (Madang Province) shown.

Subjects

The comparative treatment trial [8] was approved by the institutional review boards of the University of Western Australia (Approval no. 1060) and of the PNG Institute of Medical Research (MRAC 04.14). Parental consent was given. From the first patients enrolled at each site, a total of 84 blood samples from Alexishafen and 88 from Kunjingini were available and these were genotyped under supplementary ethical approval (MRAC 07/08). The median age of children from whom the samples were taken was 3.0 (range 0.5-5.3) years and 57% were males. There were no significant differences in age or sex between children from the two study sites.

Amplichip genotyping and allele scoring

The AmpliChip® (Roche, Mannheim, Germany) is the first FDA-approved pharmacogenetic gene chip for comprehensive *CYP2D6* and *CYP2C19* genotyping [10, 11]. The AmpliChip holds ~15,000 probe cells that allow comprehensive genotyping of more than 20 *CYP2D6* alleles

including gene duplication and deletion. In addition, the *CYP2C19* *2 and *3 alleles can be genotyped. Genotyping was performed according to the manufacturer's instructions. In brief, PCR products encompassing *CYP2D6* and *CYP2C19* gene loci were generated, fragmented and labelled. They were further processed (hybridization, staining) on an Affymetrix GeneChip Fluidics Station 450Dx and scanned on a Affymetrix GeneChip Scanner 3000Dx. Data were processed using the AmpliChip CYP450 Data Analysis Software V2.0.0 Beta to generate genotype calls. The following *CYP2D6* SNPs are tested as part of the AmpliChip P450 procedure: -1584C>G, 31G>A; 100C>T; 138insT; 883G>C; 1023C>T; 1039C>T; 1659G>A; 1661G>C; 1707delT; 1758G>T; 1758G>A; 1846G>A; 1976G>A; *20 cluster (1973insG, 1978C>T, and 1979T>C), 2539-2542delAACT; 2549delA; 2613-2615delAGA; 2850C>T; 2935A>C; 3183G>A; 3198C>G; 3277T>C; 4042G>A; *36 gene conversion; 4180G>C. The SNPs positions are numbered according the distance from the A of the ATG start codon in the reference sequence M33388. Based on these SNPs the alleles *CYP2D6**1 to *4, *6 to *11, *14, *15, *17, *19, *20, *25, *26, *29-*31, *35, *36, *40 and *41 are assigned. For *CYP2C19* the *2 and *3 alleles are genotyped. Putative phenotypes are assigned by the AmpliChip software e.g. *CYP2D6**1/*4 results in extensive metabolizer (EM) and *CYP2D6**4/*41 results in intermediate metabolizer (IM). In addition we used the Gaedigk et al. activity score system with its finer scale [12].

CYP2D6 sequencing

A 5092 bp fragment including the *CYP2D6* gene was amplified in a long-range PCR reaction using primers 5'-CCA GAA GGC TTT GCA GGC TTC A -3' and 5'-ACT GAG CCC TGG GAG GTA GGT A-3' that were previously shown to amplify specifically *CYP2D6*, but not *CYP2D7P* and *CYP2D8P* [13]. The unused primers and ddNTPs were removed by incubation with exonuclease I and shrimp alkaline phosphatase (both from USB, Staufen, Germany). The PCR products were sequenced using BigDyeTerminator v1.1 Cycle sequencing kit (Applied Biosystems, Darmstadt, Germany) with primers CYP2D6_seq1 (5'- AAA CGG CAC TCA GGA CTA AC), CYP2D6_seq2 (5'- CTG GAG GTG GAG CTG GAC TT), CYP2D6_seq3 (5'- GGG TCT TCC CTG AGT GCA AA), CYP2D6_seq4 (5' GGA GGG GAC ATC TCA GAC ATG), CYP2D6_seq5 (5'- GAG GGG AGG CTG GGC AAA AG) and CYP2D6_seq6 (5'- TCG GGC GCA GGA CTA GTT GA).

Statistical analysis

Genetic marker analyses were performed as implemented in the PowerMarker 3.25 (exact test for Hardy-Weinberg equilibrium (HWE), gene/allele summary statistics) [14], Genepop 4.0 (F_{IS} estimates [15], heterozygote deficiency and excess) [16] and Popgene 1.32 software (observed/expected hetero-/homozygosities). The PHASE 2.1 software [17] was used to reconstruct haplotypes for each individual. Ten independent runs with different seed numbers (used seeds 2, 1536, 2936, 3123, 4957, 5283, 6757, 7992, 8633, 9045) were performed in order to avoid seed-biased allele assignments. Frequencies were compared by Fisher's exact test (two-sided mid-p values) or the Mann-Whitney test as appropriate.

Results

CYP2D6

Genotype data are shown in **Table 1** as inferred phenotypes. The total no call rate for *CYDP2D6* in these patient samples was 33%, and significantly higher in the Kunjingini (45%) than in the Alexishafen (20%) samples. "No call" signifies that automated genotype assignment was impossible due to presence of unknown alleles. Both cohorts had a high rate of UM (~15%). The rate of extensive metabolizers (EM) was significantly higher in the Alexishafen (58%) than the Kunjingini (41%) sample. It was inappropriate to test for significant differences between Kunjingini and Alexishafen populations in CYP2D6 poor metabolizers (PM), intermediate metabolizers (IM) and ultra-rapid metabolizers (UM) because of sample numbers and presence of unknown SNPs. PM and IM were not detected at all in the Kunjingini cohort and were rare in the Alexishafen cohort.

To analyse the dataset including "no call" alleles further the AmpliChip raw data SNP report was used and alleles were determined by Bayesian forecasting with PHASE 2.1. **Table 2** gives a more detailed analysis of the observed *CYP2D6* alleles. The *CYP2D6**1 allele was most prevalent (~60% in both cohorts). The *CYP2D6**1XN gene duplication was also preva-

Table 1. Predicted phenotypes for *CYP2D6* and *CYP2C19*

CYP2D6	Total		Kunjingini		Alexishafen		p
no call	57	(33%)	40	(45%)	17	(20%)	<0.001
Classical phenotype acc. to AmpliChip software							
PM	1	(1%)	0	(0%)	1	(1%)	0.244
IM	3	(2%)	0	(0%)	3	(4%)	0.057
EM	85	(49%)	36	(41%)	49	(58%)	0.028
UM	26	(15%)	12	(14%)	14	(17%)	0.600
Activity score							
0.0	1	(1%)	0	(0%)	1	(1%)	0.244
0.5	2	(1%)	0	(0%)	2	(2%)	0.120
1.0	15	(9%)	4	(5%)	11	(13%)	0.044
1.5	6	(3%)	0	(0%)	6	(7%)	0.006
2.0	113	(66%)	66	(75%)	47	(56%)	0.008
2.5	3	(2%)	1	(1%)	2	(2%)	0.430
3.0	32	(19%)	17	(19%)	15	(18%)	0.771
Sum	172		88		84		
CYP2C19							
PM	68	(40%)	35	(40%)	33	(39%)	0.938
EM	104	(60%)	53	(60%)	51	(61%)	0.938
Sum	172		88		84		

CYP2D6 and CYP2C19 inferred phenotypes from two Papua New Guinean populations. p-values compare Kunjingini and Alexishafen cohorts. The activity score according to model A [12] is given for all observed *CYP2D6* genotypes. The p-value for AS score between group comparison is 0.036 (Mann-Whitney test). The assignment is putative for novel alleles (see results). p-values are given as two-sided Fisher's exact test mid-p values.

lent (~12% in both cohorts). Overall, the *CYP2D6* genotype distribution was highly significant different between the two cohorts (p<0.001). The frequencies of the gene deletion (*CYP2D6**5), and *CYP2D6**4 and *10 alleles were significantly higher in the Alexishafen compared to the Kunjingini cohort. The Kunjingini cohort has only one non-functional allele (*5 [frequency 3%]). The Alexishafen cohort carries more non- and reduced-activity alleles (*CYP2D6**4 [3%]; *5 [8%]; *10 [3%]) and the novel reduced-activity allele [100C>T;1039C>T] (2%; **Table 2**, see also below). The Kunjingini cohort had a higher mean AS (2.15) than the Alexishafen cohort (1.96, p=0.036).

CYP2D6 showed an excess of heterozygotes (**Table 3**). In accordance with this finding, a negative F_{IS} estimate (within-population inbreeding coefficient) was seen [15]. The Kunjingini cohort showed a tendency to violate HWE (see below) and this was significant in the analysis of the combined cohort (**Table 3**).

Characterization of novel CYP2D6 alleles

Several haplotypes were identified that could not be assigned to any previously known *CYP2D6* allele, either by the AmpliChip software or by direct comparison with the cytochrome P450 allelic database (www.cypalleles.ki.se). A previously undefined *CYP2D6* haplotype, referred to here as *CYP2D6*1661* allele, was observed with a frequency of 15% in the sample (**Table 2**). The *CYP2D6*1661* allele carried a 1661G>C substitution but lacked 2850C>T or 4180G>C substitutions. We re-sequenced 3961 bp of the *CYP2D6* gene in three heterozygous and one hemizygous carrier of the 1661G>C allele. This covered the whole *CYP2D6* coding region, 2064 bp of intron sequence and 251 bp

Table 2. *CYP2D6* and *CYP2C19* allele frequencies

CYP2D6 alleles	AS	Total		Kunjingini		Alexishafen		p
*1	1	210	(61%)	110	(63%)	100	(60%)	0.545
*2	1	1	(0%)	0	(0%)	1	(1%)	0.244
Duplication (*1XN)	2	40	(12%)	19	(11%)	21	(13%)	0.678
Deletion (*5)	0	18	(5%)	5	(3%)	13	(8%)	0.040
*4	0	5	(1%)	0	(0%)	5	(3%)	0.013
*10	0.5	5	(1%)	0	(0%)	5	(3%)	0.013
*41	0.5	5	(1%)	1	(1%)	4	(2%)	0.133
31G>A	1	1	(0%)	1	(1%)	0	(0%)	0.744
[100C>T;1039C>T]	0.5	3	(1%)	0	(0%)	3	(2%)	0.058
1661G>C	1	51	(15%)	35	(20%)	16	(10%)	0.008
[1661G>C;4180G>C]	1	5	(1%)	5	(3%)	0	(0%)	0.044
Sum		344		176		168		

Comparison of *CYP2D6* alleles/SNPs called by AmpliChip P450 analysis from two Papua New Guinea populations. Allele frequencies are given in brackets. If possible alleles are given in the star nomenclature. Novel alleles are indicated by their variant sites. Haplotypes were reconstructed using ten independent runs of the PHASE software. The haplotype assignment was consistent in all ten PHASE runs. p-values compare Kunjingini and Alexishafen cohorts. AS; activity score according to model A [12], putative assignment for novel alleles. The *CYP2D6**41 allele may also have occurred as duplicated allele (PHASE probability 13%).The 31G>A haplotype may also have occurred as [31G>A;1661G>C] (PHASE probability 50%). The [100C>T; 1039C>T] haplotype may also have occurred as [100C>T;1039C>T; 1661G>C] (PHASE probability 16%). The [1661G>C;4180G>C] haplotype may also have occurred as 1661G>C and 4180G>C in trans (PHASE probability 2%).
p-values are given as two-sided Fisher's exact test mid-p values.

Table 3. Population genetic descriptive statistics

	Observed heterozygosity	Expected heterozygosity	Observed homozygosity	Expected homozygosity	Fis	HWE, p
CYP2C19						
Kunjingini	0.602	0.580	0.398	0.420	-0.026	0.299
Alexishafen	0.750	0.667	0.250	0.333	-0.112	0.407
All samples	0.674	0.636	0.326	0.364	-0.070	0.739
CYP2D6						
Kunjingini	0.659	0.560	0.341	0.440	-0.133	0.076
Alexishafen	0.620	0.620	0.381	0.384	-0.005	0.117
All samples	0.640	0.590	0.361	0.410	-0.068	0.003

Observed and expected hetero- and homozygosities for *CYP2C19* and *CYP2D6* and resulting F_{IS} according to [15]. Exact test for departure from HWE given as p-value.

of the proximal promoter region. Genetic variants typical for the *CYP2D6**2 allele were observed in the 5' part but were missing in the 3' part of the 1661G>C allele (see **Figure 2**). The *CYP2D6*1661* allele may have originated from a crossing-over event between alleles *CYP2D6**1 and *2. There were no additional variants in the 1661C>G allele that differed from those known in the fully active *CYP2D6**1 allele. Therefore we assigned a putative activity

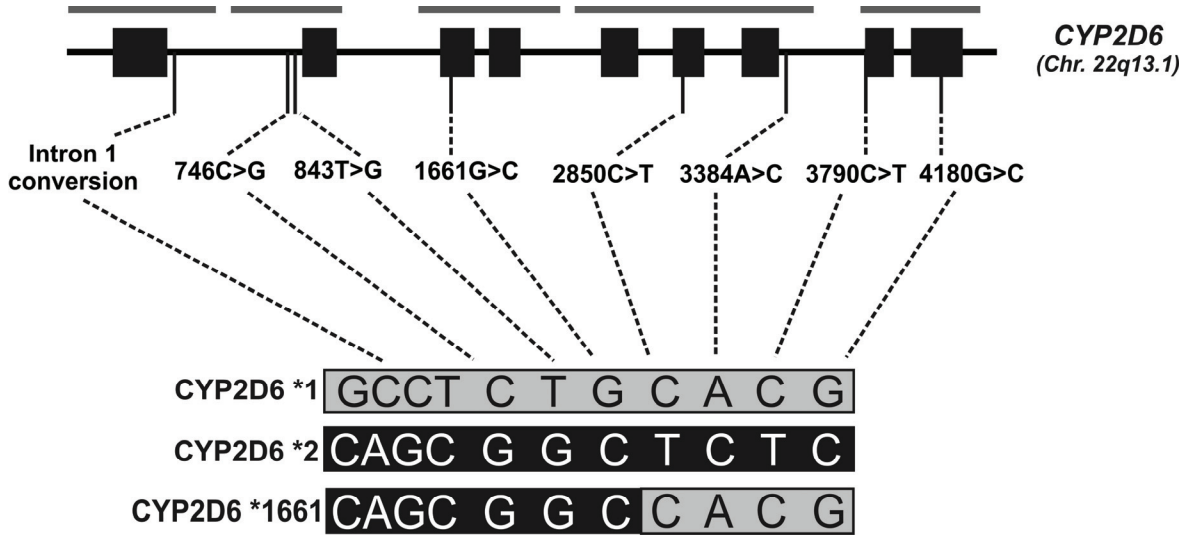

Figure 2. Illustration of the novel *CYP2D6*1661* allele (*CYP2D6**1661C>G) with high prevalence in the Papua New Guinean population. Exon-intron structure of the gene and the localisation of single polymorphic loci characteristic for *CYP2D6* *2 are shown. The regions of *CYP2D6* gene depicted in grey were analysed by re-sequencing in carriers of the 1661G>C allele. The combination of alleles of single polymorphic loci within the gene in the *CYP2D6* allele 1661G>C is compared with corresponding combinations of *CYP2D6* alleles 1 and 2 (obtained from Kimura et al. ([34]) and the genome reference sequence respectively). Suggested crossover between *CYP2D6**1 and *2 is illustrated as a combination of grey and black shading for the *CYP2D6*1661* allele.

score of 1 according to model A [12]. The *CYP2D6*1661* allele was not in HWE in the Kunjingini population ($p<0.001$) but in the Alexishafen population ($p=1$).

Re-sequencing of a heterozygous [100C>T; 1039C>T] allele carrier showed a combination of single variants 100C>T, 843T>G, 1039C>T, 2097G>A, 3384A>C, 4180G>C. This allele differed from the previously known allele *CYP2D6**10B by an absence of the G>C exchange at position 1661, a variant known not to cause functional changes in enzyme expression or activity. Therefore we assigned an activity score of 0.5 due to presence of the 100C>T as for the *CYP2D6**10 reduced-activity allele.

CYP2C19

All *CYP2C19* blood samples (n=172) had successful genotype calls. There was a high proportion (40%) of predicted CYP2C19 PMs that was equal in both patient cohorts. Despite the comparable frequencies of predicted CYP2C19 PMs (**Table 1**), we observed a different genetic basis for non-functional alleles (**Table 4**). The *CYP2C19**2 allele frequency was significantly higher in the Kunjingini cohort while the frequency of the *CYP2C19**3 allele was significantly higher in the Alexishafen cohort. The *CYP2C19**1 allele was of comparable frequency in both cohorts. The *CYP2C19* genotype distribution was not different between the two cohorts ($p=0.815$) and both were in HWE. *CYP2C19* also showed an excess of heterozygotes (**Table 3**) and, therefore, a negative F_{IS} estimate was seen.

Discussion

Pharmacogenetics has great potential to allow pharmacotherapy to be tailored to the individual based on drug metabolism and action. Ethnicity has generally proven to be non-specific in predicting pharmacogenetically relevant features [18, 19, 20]. It is accepted that the genetic variability between any two individuals within any one group is almost as high as the genetic variability between any two individuals anywhere in the world [21]. Therefore the knowledge of population specific variants in pharmacogenetically relevant genes such as *CYP2D6* will help to improve prediction overall.

This study undertook genotyping of *CYP2D6* and *CYP2C19* using the AmpliChip P450 gene chip

Table 4. *CYP2C19* genotype and allele data

CYP2C19	Total		Kunjingini		Alexishafen		p
*1/*1	19	(11%)	11	(13%)	8	(10%)	0.550
*1/*2	61	(35%)	38	(43%)	23	(27%)	0.032
*1/*3	24	(14%)	4	(5%)	20	(24%)	<0.001
*2/*2	31	(18%)	22	(25%)	9	(11%)	0.013
*3/*3	6	(3%)	2	(2%)	4	(5%)	0.323
*2/*3	31	(18%)	11	(13%)	20	(24%)	0.061
Sum	172		88		84		
*1	123	(36%)	64	(36%)	59	(35%)	0.779
*2	154	(45%)	93	(53%)	61	(36%)	0.002
*3	67	(19%)	19	(11%)	48	(29%)	<0.001
Sum	344		176		168		

Comparison of *CYP2C19* genotypes and alleles called by AmpliChip P450 analysis in two Papua New Guinean populations. p-values compare Kunjingini and Alexishafen cohorts.
p-values are given as two-sided Fisher's exact test mid-p values.

in cohorts from two Melanesian populations in PNG. *CYP2D6* genotyping revealed at least 3 novel alleles in these cohorts. The most frequent allele *CYP2D6*1661* (15% in the total sample) was previously assigned on a SNP basis by Sistonen et al. [4] who also found this allele prevalent in Papuan populations. The *CYP2D6**1661 allele was the second most frequent allele in both populations studied while the *CYP2D6**2 was present in only one child (0.2%; **Table 2**). This is in concordance with previous data [4] and indicates high relevance of *CYP2D6*1661* in the PNG populations. Of note the *CYP2D6*1661* alleles have also been observed in Eastern Europe, the Middle East and Subsaharan Africa [4]. Therefore we completely re-sequenced the *CYP2D6**1661 allele. Our data suggest that *CYP2D6**1661 originates from a crossing over event between of the *CYP2D6**1 and *CYP2D6**2 alleles within the regions between exon 3 and 6 (**Figure 2**). Both alleles confer full CYP2D6 activity and the *CYP2D6**1661 allele does not contain any additional polymorphisms. Therefore, full activity is likely for the *CYP2D6**1661 allele. Interestingly, the 1661G>C SNP violated HWE in the Kunjingini but not in the Alexishafen cohort; the observed absence of homozygote carriers and excess of heterozygote carriers may be explained by introduction of this allele by migration of a single heterozygote into the population. This results in a relative homozygote deficiency for several generations [22].

In contrast to Sistonen et al. [4], we found no isolated 4180G>C haplotype (PHASE probability only 0.02) but a [1661G>C;4180G>C] haplotype (PHASE probability 0.98). Another allele with putative full activity was characterized by 31G>A. This allele occurred only once and could also be [31G>A;1661G>C](PHASE probability 0.50). Presence of 1661G>C (PHASE probability 0.08) as well as *CYP2D6**41 (PHASE probability 0.13) in the duplicated gene could not be confirmed but is theoretically possible based on haplotype assignments. The duplicated allele *CYP2D6**1XN frequencies (12%) in both cohorts are high compared to other populations [2] but are comparable to or below those reported in Saudi Arabians [23] and Ethiopians [24].

The high frequency of duplicated active alleles in the PNG population is mirrored by a low frequency of low activity alleles. In the Kunjingini cohort, all cases had an activity score according to Gaedigk et al. [12] of ≥1 (**Table 1**). The Alexishafen cohort had more known (*CYP2D6**4, *5, *10) and novel putative reduced-activity alleles [100C>T;1039C>T] but still the frequency of cases with an activity score <1 was only 3%. We were not in a position to corroborate these findings by investigation of the phenotype but, based on the known genotype-phenotype correlations, we expect that medications or CYP2D6 substrates in general undergo extensive to even ultra-rapid metabolism in the studied populations. Toxicity associated with PM

should be almost absent.

The *CYP2D6* genetic diversity in the Alexishafen cohort was higher than in Kunjingini. The presence of both Austronesian and Papua speakers as well as increased transmigration into the Alexishafen area may have contributed to increased genetic diversity by admixture of alleles of different origins. By contrast, the Kunjingini area is geographically more isolated and linguistically more uniform. Minority (deficient) alleles may thus have been selected out or lost through genetic drift, consistent with observations in Pacific Islanders of a lower genetic variability in large island interiors versus more intermixed Austronesian speaking groups living along coastlines [25]. *CYP2D6* was not in HWE in the total sample and showed a similar tendency in the Kunjingini sub-group (**Table 3**). We are not aware of formal marriage restrictions in PNG but observed homozygosities were always smaller than expected. This favours outbreeding, including through arranged marriages between different clans of the same tribe, due to strong taboos against inbreeding causing exogamy [26]. The F_{IS} was negative in all cohorts and genes which indicates a departure from HWE towards a heterozygote excess (**Table 3**). Such negative F_{IS} values are to be expected in small populations [27]. In general, heterozygote excess has been much less discussed in the literature than heterozygote deficiency [28], especially in the multiple allele case as for *CYP2D6*.

For CYP2C19 we observed an exceptional high rate of 40% predicted PM. This was due to *CYP2C19**2 and *3 deficiency alleles. Our findings are in line with the previously-reported high frequencies in the Melanesian people from Vanuatu [29] and East Sepik [6]. Interestingly, while the Kunjingini cohort showed comparably high frequencies for *CYP2C19**2 and *CYP2C19**3 as reported for the other geographically (and linguistically) close East Sepik population [6], the *CYP2C19**3 allele was significantly more common in the Alexishafen cohort. Whether this is due to different migration histories, genetic drift or simply by chance in the studied samples is unclear. However, *CYP2C19**3 is significantly more common in the Austronesian speaking Madang population, while the Papuan speaking non-Austronesian populations in the Sepik and the also Austronesian speaking populations in Vanuatu have similar frequencies.

Deviations from HWE with global heterozygote excess and homozygote deficiency as observed for both *CYP2D6* and *CYP2C19* in this study can be caused by viability selection [30]. However, given the PNG population structure, it would be difficult or impossible to determine whether heterozygosity is maintained due to viability selection of certain alleles [31]. In addition to a heterozygote advantage, the population size, differential fertility (including confounding by genotype) and eventually differing allele frequencies in males and females may contribute to heterozygote excess [22, 32]. Analysis of genotypes classified by activity scores did not suggest violation of HWE (data not shown) which argues against a strong selection of certain phenotypes. Further factors include genetic drift, lack of panmixia, admixture and migration effects may play important roles in shaping *CYP2D6* allelic diversity in small populations such as those studied here. A bottleneck analysis was not feasible due to the small number of polymorphic loci studied [33].

In conclusion, we confirmed a high frequency of the *CYP2D6**1661 allele in PNG. The *CYP2D6**1661 allele was classified as fully active allele by sequence based characterization. We found a very low frequency of PM and a high frequency of UM in samples from the studied PNG populations compared to European populations. This study also confirms the high frequency of *CYP2C19* PM which is typical for Oceanian populations. The observed pattern of violation of HWE with heterozygote excess and homozygote deficiency could be explained by peculiarities in relatively isolated PNG populations with exogamy and recent introduction of alleles by migration. The characterization of these common PNG alleles has the potential to also improve the CYP2D6 phenotype prediction in other world populations.

Acknowledgements

The authors note with deep regret that Servina Gomorrai passed away after completion of the study.

The comparative clinical trial from which the samples were taken was sponsored by WHO Western Pacific Region, Rotary Against Malaria (PNG), and the National Health and Medical Research Council (NHMRC) of Australia (grant 353663). AmpliChip gene chips for this study

were kindly provided by Roche Diagnostics (Mannheim, Germany). TMED is supported by an NHMRC Practitioner Fellowship. We acknowledge the excellent technical assistance of Sandra Hartung, Reiner Andag and Karoline Jobst. We are most grateful to Sister Valsi Kurian and the staff of Alexishafen Health Centre for their kind co-operation during the performance of the field-work in this study. We also thank Sister Maria Goretti, Jovitha Lammey, Wesley Sikuma, Donald Paiva, Bernard ("Ben") Maamu, Olive Oa, Petronilla Wapon, Jane Simbrandu, Merilyn Uranoli, Peter Maku, John Taime, Irwin Lau and Michelle England for clinical, laboratory and/or logistic assistance.

Please address correspondence to: Nicolas von Ahsen, MD, Georg-August University, Department of Clinical Chemistry, Robert-Koch-Str. 40, 37099 Göttingen, Germany. Fax +49(551)39-12501, E-mail: nahsen@gwdg.de

References

[1] Ingelman-Sundberg M. Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): clinical consequences, evolutionary aspects and functional diversity. Pharmacogenomics J 2005; 5: 6-13.

[2] Eichelbaum M, Ingelman-Sundberg M, Evans WE. Pharmacogenomics and individualized drug therapy. Annu Rev Med 2006; 57: 119-137.

[3] Weinshilboum R. Inheritance and drug response. N Engl J Med 2003; 348: 529-537.

[4] Sistonen J, Sajantila A, Lao O, Corander J, Barbujani G, Fuselli S. CYP2D6 worldwide genetic variation shows high frequency of altered activity variants and no continental structure. Pharmacogenet Genomics 2007; 17: 93-101.

[5] Sistonen J, Fuselli S, Palo JU, Chauhan N, Padh H, Sajantila A. Pharmacogenetic variation at CYP2C9, CYP2C19, and CYP2D6 at global and microgeographic scales. Pharmacogenet Genomics 2009; 19: 170-179.

[6] Masta A, Lum JK, Tsukahara T, Hwaihwanje I, Kaneko A, Paniu MM, Sapuri M, Takahashi N, Ishizaki T, Kobayakawa T, Hombhanje FW. Analysis of Sepik populations of Papua New Guinea suggests an increase of CYP2C19 null allele frequencies during the colonization of Melanesia. Pharmacogenetics 2003; 13: 697-700.

[7] Desta Z, Zhao X, Shin JG, Flockhart DA. Clinical significance of the cytochrome P450 2C19 genetic polymorphism. Clin Pharmacokinet 2002; 41: 913-958.

[8] Karunajeewa HA, Mueller I, Senn M, Lin E, Law I, Gomorrai PS, Oa O, Griffin S, Kotab K, Suano P, Tarongka N, Ura A, Lautu D, Page-Sharp M, Wong R, Salman S, Siba P, Ilett KF, Davis TM. A trial of combination antimalarial therapies in children from Papua New Guinea. N Engl J Med 2008; 359: 2545-2557.

[9] Genton B, al-Yaman F, Beck HP, Hii J, Mellor S, Narara A, Gibson N, Smith T, Alpers MP. The epidemiology of malaria in the Wosera area, East Sepik Province, Papua New Guinea, in preparation for vaccine trials. I. Malariometric indices and immunity. Ann Trop Med Parasitol 1995; 89: 359-376.

[10] de Leon J, Susce MT, Murray-Carmichael E. The AmpliChip CYP450 genotyping test: Integrating a new clinical tool. Mol Diagn Ther 2006; 10: 135-151.

[11] Heller T, Kirchheiner J, Armstrong VW, Luthe H, Tzvetkov M, Brockmöller J, Oellerich M. AmpliChip CYP450 GeneChip: a new gene chip that allows rapid and accurate CYP2D6 genotyping. Ther Drug Monit 2006; 28: 673-677.

[12] Gaedigk A, Simon SD, Pearce RE, Bradford LD, Kennedy MJ, Leeder JS. The CYP2D6 activity score: translating genotype information into a qualitative measure of phenotype. Clin Pharmacol Ther 2008; 83: 234-242.

[13] Sachse C, Brockmöller J, Bauer S, Roots I. Cytochrome P450 2D6 variants in a Caucasian population: allele frequencies and phenotypic consequences. Am J Hum Genet 1997; 60: 284-295.

[14] Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 2005; 21: 2128-2129.

[15] Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution 1984; 38: 1358-1370.

[16] Rousset F. GENEPOP'007: a complete re-implementation of the GENEPOP. Mol Ecol Res 2008; 8: 103-106.

[17] Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 2003; 73: 1162-1169.

[18] Doyle JM. What race and ethnicity measure in pharmacologic research. J Clin Pharmacol 2006; 46: 401-404.

[19] Harty L, Johnson K, Power, A. Race and ethnicity in the era of emerging pharmacogenomics. J Clin Pharmacol 2006; 46: 405-407.

[20] Jorde LB, Wooding SP. Genetic variation, classification and 'race'. Nat Genet 2004; 36: S28-S33.

[21] Witherspoon DJ, Wooding S, Rogers AR, Marchani EE, Watkins WS, Batzer MA, Jorde LB. Genetic similarities within and between human populations. Genetics 2007; 176: 351-359.

[22] Neel JV, Ward RH. The genetic structure of a tribal population, the Yanomama Indians. VI. Analysis by F-statistics (including a comparison with the Makiritare and Xavante). Genetics 1972; 72: 639-666.

[23] McLellan RA, Oscarson M, Seidegard J, Evans DA, Ingelman-Sundberg M. Frequent occurrence of CYP2D6 gene duplication in Saudi Arabians. Pharmacogenetics 1997; 7: 187-191.

[24] Aklillu E, Herrlin K, Gustafsson LL, Bertilsson L, Ingelman-Sundberg M. Evidence for environ-

mental influence on CYP2D6-catalysed debrisoquine hydroxylation as demonstrated by phenotyping and genotyping of Ethiopians living in Ethiopia or in Sweden. Pharmacogenetics 2002; 12: 375-383.
[25] Friedlaender JS, Friedlaender FR, Reed FA, Kidd KK, Kidd JR, Chambers GK, Lea RA, Loo JH, Koki G, Hodgson JA, Merriwether DA, Weber JL. The genetic structure of Pacific Islanders. PLoS Genet 2008; 4: e19-.
[26] Long JC, Naidu JM, Mohrenweiser HW, Gershowitz H, Johnson PL, Wood JW, Smouse PE. Genetic characterization of Gainj- and Kalam-speaking peoples of Papua New Guinea. Am J Phys Anthropol 1986; 70: 75-96.
[27] Harpending H. Genetic structure of small populations. Ann Rev Anthropol 1974; 3: 229-243.
[28] Karlin S. Levels of multiallelic overdominance fitness, heterozygote excess and heterozygote deficiency. Theor Popul Biol 1990; 37: 129-149.
[29] Kaneko A, Bergqvist Y, Taleo G, Kobayakawa T, Ishizaki T, Bjorkman A. Proguanil disposition and toxicity in malaria patients from Vanuatu with high frequencies of CYP2C19 mutations. Pharmacogenetics 1999; 9: 317-326.
[30] Alvarez G. Deviations from Hardy-Weinberg proportions for multiple alleles under viability selection. Genet Res 2008; 90: 209-216.
[31] Wiesenfeld SL, Gajdusek DC. Genetic structure and heterozygosity in the Kuru region, eastern highlands of New Guinea. Am J Phys Anthropol 1976; 45: 177-189.
[32] Thompson EA, Roberts DF. Kinship structure and heterozygosity on Tristan da Cunha. Am J Hum Genet 1980; 32: 445-452.
[33] Cornuet JM, Luikart G. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 1996; 144: 2001-2014.
[34] Kimura S, Umeno M, Skoda RC, Meyer UA, Gonzalez FJ. 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 1989; 45: 889-904.