7-2023

Pharmacogenomic variation in the Malagasy population: implications for the antimalarial drug primaquine metabolism.

Estee Y. Cramer

Jacquelaine Bartlett

Ernest R. Chan

Andrea Gaedigk
Children's Mercy Kansas City

Arsene C. Ratsimbasoa

See next page for additional authors

Recommended Citation

Cramer EY, Bartlett J, Chan ER, et al. Pharmacogenomic variation in the Malagasy population: implications for the antimalarial drug primaquine metabolism. Pharmacogenomics. 2023;24(11):583-597. doi:10.2217/pgs-2023-0091

Research Article

For reprint orders, please contact: reprints@futuremedicine.com

Pharmacogenomics

Pharmacogenomic variation in the Malagasy population: implications for the antimalarial drug primaquine metabolism

Estee Y Cramer[1,2], Jacquelaine Bartlett[3], Ernest R Chan[†,3,4], Andrea Gaedigk[5], Arsene C Ratsimbasoa[6,7], Rajeev K Mehlotra*[,1], Scott M Williams**[,3] & Peter A Zimmerman***[,1]

[1]Center for Global Health & Diseases, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA
[2]Department of Biostatistics & Epidemiology, School of Public Health & Health Sciences, University of Massachusetts Amherst, Amherst, MA 01003, USA
[3]Population & Quantitative Health Sciences, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA
[4]Cleveland Institute for Computational Biology, Case Western Reserve University, Cleveland, OH 44106, USA
[5]Division of Clinical Pharmacology, Toxicology & Therapeutic Innovation, Children's Mercy Research Institute (CMRI), Kansas City, MO 64108, USA
[6]University of Fianarantsoa, Fianarantsoa, Madagascar
[7]Centre National d'Application de Recherche Pharmaceutique (CNARP), Antananarivo, Madagascar
*Author for correspondence: rkm@case.edu
**Author for correspondence: smw154@case.edu
***Author for correspondence: paz@case.edu
[†]These authors contributed equally to this work

Aim: Antimalarial primaquine (PQ) eliminates liver hypnozoites of *Plasmodium vivax. CYP2D6* gene variation contributes to PQ therapeutic failure. Additional gene variation may contribute to PQ efficacy. Information on pharmacogenomic variation in Madagascar, with *vivax* malaria and a unique population admixture, is scanty. **Methods:** The authors performed genome-wide genotyping of 55 Malagasy samples and analyzed data with a focus on a set of 28 pharmacogenes most relevant to PQ. **Results:** Mainly, the study identified 110 coding or splicing variants, including those that, based on previous studies in other populations, may be implicated in PQ response and copy number variation, specifically in chromosomal regions that contain pharmacogenes. **Conclusion:** With this pilot information, larger genome-wide association analyses with PQ metabolism and response are substantially more feasible.

First draft submitted: 23 May 2023; Accepted for publication: 11 July 2023; Published online: 8 August 2023

Keywords: copy number variation • genome-wide array • Madagascar • malaria treatment • *Plasmodium vivax* • primaquine

The treatment and control of *Plasmodium vivax* malaria are complicated by a latent, undetectable liver-stage form in the parasite life cycle, known as the hypnozoite [1,2]. Hypnozoites can cause clinical relapses weeks to months after clearance of the acute blood-stage infection. This can increase the disease burden and the potential for unexpected transmission, as well as impede efforts to eliminate *vivax* malaria [3,4]. Primaquine (PQ) and tafenoquine, both 8-aminoquinoline antimalarials, are the only known drug therapies for *vivax* latent liver-stage infection and relapse [5–8]. Hepatic CYP2D6 is the enzyme primarily responsible for the phase I metabolism of these drugs; there is a growing body of evidence indicating that the efficacy of PQ bioactivation is impacted by the highly polymorphic *CYP2D6* gene. Several recent studies have demonstrated that decreased or absent CYP2D6 activity (i.e., intermediate metabolizer or poor metabolizer phenotype) contributes to PQ therapeutic failure [9–14].

In addition to *CYP2D6*, variation in other genes may contribute to PQ metabolism, transport, response and to overall therapeutic efficacy [15–20]. An early *in vitro* study identified hepatic enzymes CYP1A2 and CYP2D6 as involved in the metabolism of PQ [17]. Recent reports have indicated that other CYP enzymes – CYP2C8 [18,20], CYP2C9 [20], CYP2C19 [15,16] and CYP3A5 [20] – as well as monoamine oxidase A (MAOA) [15], UDP-

Future Medicine

glucuronosyltransferase 2B7 (UGT2B7) [16], ATP-binding cassette transporters G2 (ABCG2) [16] and ABCB1 [19,21] and solute carrier organic anion transporters 1A2 (SLCO1A2), SLCO1B1 and SLCO2B1 [19] also contribute to PQ metabolism or transport, and variants in these genes may play a role in variable treatment outcomes for *vivax* malaria. It is therefore likely that variation in genes other than *CYP2D6* also affect the efficacy of PQ in killing hypnozoites [22].

Madagascar is an island nation in the Indian Ocean, approximately 400 km (250 miles) off the east African coast. Malagasy populations offer a unique, rich history of admixture between east African and southeast Asian ancestral populations [23–27]. Across this geographic landscape, contributions from Arabic, Indian, Papuan and/or Jewish populations have been suggested [28]. *Vivax* malaria is endemic in Madagascar [29], and PQ use can support efforts toward national malaria elimination [30]. Research on pharmacogenomic variation in Madagascar is very limited. As part of the authors' ongoing malaria epidemiological studies in the western highlands fringe region of Madagascar [29,31–33], they have recently investigated *CYP2D6* variation in Madagascar using commercially available TaqMan™ genotyping assays [11] and targeted next-generation sequencing [34]. To the best of the authors' knowledge, only one report has examined this population for variation in another gene, *CYP2C8*, that may affect *vivax* treatment [35].

With other reports indicating that variation in genes other than *CYP2D6* may contribute to differences in PQ metabolism as well as response [15–20], the paucity of pharmacogenomic information for Madagascar represents an important knowledge gap. In the present study, the authors performed genome-wide SNP typing of 55 Malagasy samples with the Infinium Multi-Ethnic Genotyping Array (MEGA) and analyzed data with a focus on a set of 28 pharmacogenes (Supplementary Table 1). Most of these genes have been shown or implicated to have a role in either PQ metabolism or transport, are specified in the Pharmacogenomics Knowledge Base as Very Important Pharmacogene tier one or two, or both. In view of a key role of MAO enzymes in the metabolism of 8-aminoquinolines [36,37], *MAOA* and *MAOB* genes were also included in the current survey. In addition, since PQ and certain metabolites undergo glucuronidation [38–40], *UGT* genes encoding the most clinically important hepatic UGT enzymes were also included [41]. The array data were also used to identify copy number variation (CNV). To visualize population relationships of the Malagasy samples, principal component analysis (PCA) was applied to genotype data from the whole genome and from those genes most relevant to PQ.

Materials & methods

Study sites, study participants, sample collection & processing

The samples (n = 59) used in the present study were part of malaria epidemiological studies conducted by the authors' group in the western highlands fringe region of Madagascar [31]. Samples were collected between March and August 2015 from unrelated individuals of all ages, most of whom were male. Detailed information about the study sites, participants, blood sample collection and DNA extraction has been described previously [11,31]. Briefly, the population sampled included rural communities in the Mandoto district (n = 55) and a migrant population in the rural Ampasimpotsy community of Tsiroanomandidy district (n = 4) relocated from the capital city, Antananarivo. Both communities self-identified as being predominantly of the Merina highland ethnic group, the largest ethnic group in Madagascar.

Infinium multi-ethnic genotyping array analysis

The MEGA array is a consortium-designed custom array with 1.7 M markers (MEGA_Consortium_v1-1_15071648_A1). The marker manifest is based on the genome build GRCh37. This array is the early version of other MEGA arrays (Infinium Expanded Multi-Ethnic Genotyping Array [MEGA^{EX}] with 2M markers, and the Multi-Ethnic Global BeadChip [MEG] with 1.7 M markers).

DNA samples were sent to the Genotyping Core at Hussman Institute for Human Genomics (HIHG), University of Miami Miller School of Medicine, where their quality was assessed and the arrays were run following the Illumina-recommended protocol. Post-DNA quality control (QC), 55 samples (Mandoto, 51; Ampasimpotsy, 4) with highest DNA quality scores were subjected to MEGA genotyping. One HIHG control sample (201207526-01, HapMap ID NA12760) was included to evaluate array performance.

Postgenotyping processing & quality control

Initial genotype calling was performed using Illumina's GenomeStudio software v2.0.0. No Call threshold of 0.15 was used as recommended for Infinium products. Since the MEGA is a custom array, Illumina does not provide

an off-the-shelf cluster file. Therefore, a cluster file generated through another project that used the same array to analyze 864 samples was utilized to improve genotype calling. Sample and SNP metrics, such as call rate, p10GC score, call frequency (Freq) and GenTrain score, are defined by Illumina in various resources [42,43]. QC of the raw genotype data was performed using the GenomeStudio software v2.0.5 [44]. QC steps were followed as recommended by Illumina [42,45] and have been described by various authors [46,47] (Supplementary Document 1).

Post-quality control reports & further analyses

A post-QC, final report (.CSV file) and custom PLINK report (.MAP and .PED files) [48] were generated. The final report was used to perform PCA, and the PLINK report was used for further analysis of the genotype data.

PLINK analysis

PLINK (1.90b6.11) was used to indicate duplicate SNPs and determine minor allele frequencies (MAFs) and missing genotype rates. After duplicates were removed and those SNPs that do not have a known chromosome (Chr) location were excluded, monomorphic and polymorphic variation was categorized for the whole genomic region.

To extract SNP information, the following databases were used: for location and SNP ID, Ensembl [49], dbSNP [50] and Genome Aggregation Database [51]; for genomic sequence position, dbSNP and the Pharmacogene Variation Consortium (PharmVar) [52,53]; for variant impact, dbSNP and PharmVar; for star (*) allele/suballele, PharmVar and UGT Alleles nomenclature [54] (the latter was limited, as it was not up to date [55]); and for SNP frequencies in global, European, African, south Asian and east Asian populations, the 1000 Genomes Project (phase III) [56].

Copy number variation analysis

CNV analysis was performed using cnvPartition CNV Analysis Plug-in for GenomeStudio [48]. This program identifies regions of CNV in a project's samples based on allele frequency (B allele frequency [BAF]) and signal intensity (log R ratio [LRR]) data derived from the project. The cnvPartition algorithm estimates copy number by comparing the observed BAF and LRR for each locus to predict BAFs and LRRs of different copy number scenarios [57,58]. For regions with copy numbers other than two, the algorithm also assigns a confidence score for the copy number that is called. The confidence threshold allows users to filter out CNV regions that have low confidence values. The default value of 35 was determined empirically using normal HapMap samples on the Illumina Human1M BeadChip [57]. The current study used a more stringent cutoff value of 100 for the confidence threshold and ten for the minimum probe count (default value three). In addition, the authors excluded sex chromosomes and did not include copy-neutral loss of heterozygosity (a phenomenon whereby one of two homologous chromosomal regions is lost, but various mechanisms have ensured the presence of two identical copies of such region in the genome) in the final results.

Principal component analysis

PCA was performed in R v4.1.2 [59] using the packages SNPRelate and gdsfmt [60]. PCA used the genome-wide data from five superpopulations in the 1000 Genomes Project (European, African, American, south Asian and east Asian; n = 2504) and the Malagasy samples (n = 55). SNPs that appeared in both sample datasets were used to perform two separate PCAs, one for the whole genome and the other for only the 28 pharmacogenes.

Results

GenomeStudio processing & quality control of the genotype data

The pre- and post-QC total number of SNPs (1,708,621 and 1,502,272, respectively), their chromosome-wise distribution and sample and SNP metrics such as call rate, p10GC score, call Freq and GenTrain score, are presented in Supplementary Document 1.

Post-quality control distribution of monomorphic & polymorphic SNPs

Among the 1,502,272 SNPs, 24,718 were duplicates and 8757 SNPs were not mapped to a single Chr location. After the removal of these, 1,468,797 SNPs were available for analysis. Of these, 40% were monomorphic and 60% were polymorphic. The MAF distribution of the polymorphic SNPs in four classes (rare, <1%; low, 1–5%;

Table 1. Number of monomorphic and polymorphic SNPs in pharmacogenes in the Malagasy†.

Gene	Total	Monomorphic	Polymorphic	Rare	Low	Moderate	High
CYP2D6	13	12	1	0	0	0	1
CYP1A2	46	40	6	0	2	2	2
CYP2C8	59	39	20	2	4	4	10
CYP2C9	48	26	22	9	2	2	9
CYP2C19	44	29	15	1	4	1	9
CYP3A5	46	35	11	0	2	2	7
MAOA	18	11	7	0	0	1	6
MAOB	34	27	7	1	0	1	5
UGT2B7	46	19	27	3	3	1	20
ABCG2	146	86	60	4	15	3	38
SLCO1A2	156	71	85	6	22	18	39
SLCO1B1	138	86	52	6	15	2	29
SLCO2B1	56	31	25	2	8	3	12
ABCB1	152	74	78	9	13	10	46
CYP2A6	18	14	4	2	2	0	0
CYP2A13	18	16	2	0	2	0	0
CYP2B6	71	38	33	8	4	2	19
CYP2E1	39	20	19	3	4	2	10
CYP3A4	50	37	13	4	0	2	7
UGT1A‡	188	124	64	18	3	2	41
UGT2B4	45	26	19	1	0	4	14
UGT2B10	21	11	10	2	4	0	4
UGT2B15	28	17	11	3	4	1	3
UGT2B17	13	11	2	0	1	1	0
Total	1493	900	593	84	114	64	331

Rare, <1%; low, 1–5%; moderate, >5–10%; high, >10%.
†Details provided in Supplementary Table 1.
‡Five *UGT1A* gene locus members (*1A1*, *1A3*, *1A4*, *1A6* and *1A9*) located on Chr 2: 234580544..234681946.

moderate, >5–10%; and high, >10%) is presented in Supplementary Table 2; 54% of the SNPs were in the high-frequency class.

There were a total of 1493 SNPs among the 28 interrogated pharmacogenes. The five *UGT1A* gene locus members, *1A1*, *1A3*, *1A4*, *1A6* and *1A9*, were analyzed collectively, as they represent alternative first exons combined with the common exons 2–4 and 5a by splicing out any intervening sequence [61]. 60% of the pharmacogene SNPs were monomorphic and 40% were polymorphic (Supplementary Table 2). The MAF distribution of the polymorphic SNPs in the four classes is also presented in Supplementary Table 2; 56% of the SNPs were in the high-frequency class.

Distribution of SNPs varied widely by the pharmacogene: from 13 SNPs (*CYP2D6*, *UGT2B17*) to 188 SNPs (*UGT1A* gene locus; Table 1). Only *ABCG2*, *SLCO1A2*, *SLCO1B1*, *ABCB1* and the *UGT1A* gene locus had >100 SNPs; the rest of the genes had 13–71 SNPs. The distribution of monomorphic (n = 900) and polymorphic (n = 593) SNPs and the MAF distribution of the polymorphic SNPs in the four classes are also presented in Table 1. Notably, out of a total of just 13 SNPs, *CYP2D6* had only one polymorphic SNP; the remaining 12 SNPs were monomorphic.

Polymorphic SNP types & frequencies in pharmacogenes

For the polymorphic SNPs, their locations, SNP IDs, genomic sequence positions and variant impacts, the star alleles/suballeles they represent and their MAFs, are presented in Supplementary Table 3 in chromosomal order. In addition, the corresponding frequencies of the Malagasy minor allele in global, European, African, south Asian and east Asian populations are also presented in Supplementary Table 3. The majority of these SNPs are intronic (n = 384), whereas 110 are in coding regions or impact splicing; the latter are presented separately in Supplementary Table 4. Of these, 70 are nonsynonymous, six impact splicing and 34 are synonymous SNPs.

Table 2. Defining variants for functionally important *CYP* alleles in the Malagasy.

Defining variant/s	*CYP* allele	Variant impact	Functional consequence†	Clinical Pharmacogenetics Implementation Consortium‡	Primaquine metabolism§	Malagasy¶	European#	African#	South Asian#	East Asian#
rs776746	*CYP3A5*3*	Splice defect	Inactive	No function	+	0.409	0.929	0.298	0.668	0.714
rs10264272	*CYP3A5*6*	Splice defect	Inactive	No function	+	0.109	<0.001	0.129	<0.001	0
rs41303343	*CYP3A5*7*	Frameshift	Decreased	No function	?	0.118	<0.001	0.102	<0.001	<0.001
rs4244285	*CYP2C19*2*	Splice defect	Inactive	No function	+	0.118	0.147	0.178	0.324	0.308
rs4986893	*CYP2C19*3*	Trp212X	Inactive	No function	?	0.027	<0.001	<0.001	0.004	0.063
rs1057910	*CYP2C9*3*	Ile359Leu	Decreased	No function	+	0.009	0.068	0.012	0.11	0.033
rs28371685	*CYP2C9*11*	Arg335Trp	Decreased	Decreased function	+	0.027	0.003	0.019	0.002	<0.001
rs11572103	*CYP2C8*2*	Ile269Phe	Decreased	–	+	0.091	0.003	0.152	0.019	<0.001
rs10509681, rs11572080	*CYP2C8*3*	Lys399Arg, Arg139Lys	Decreased	–	+	0.027	0.113	0.02	0.041	<0.001
rs28371759	*CYP3A4*18*	Leu293Pro	Decreased	–	?	0.155	0	0.001	<0.001	0.019
rs35599367	*CYP3A4*22*	Splice defect	Decreased	–	?	0.009	0.044	0.009	0.006	0
rs28399454	*CYP2A6*17*	Val365Met	Decreased	–	?	0.045	<0.001	0.109	<0.001	0
rs3745274	*CYP2B6*6*	Gln172His††	Decreased	Decreased function	?	0.255	0.196	0.32	0.281	0.215
rs28399499	*CYP2B6*18*	Ile328Thr	Decreased	No function	?	0.073	<0.001	0.07	<0.001	0

†Functionality assessment information is taken from [62,63], who used *in vitro* or *in vivo* data, and phenotype prediction methods.
‡Clinical Pharmacogenetics Implementation Consortium clinical function assignment guidelines (https://cpicpgx.org/guidelines/); no guidelines available.
§+: Role in the metabolism shown or suggested (Table 3); ?: Role in the metabolism not known.
¶Malagasy 'allele' frequencies represent SNP frequencies, which may be different from star (*) allele frequencies.
#The star (*) allele frequencies for European, African, south Asian and east Asian populations were taken from [63], who used whole exome and whole genome sequencing data from 141,614 individuals across 12 populations.
††Gln172His is also associated with other *CYP2B6* alleles. However, **6* is the most common and functionally well-characterized allele [65].

Functionally most relevant variants & star alleles

Among the 110 SNPs listed in Supplementary Table 4, 15 are SNPs that are used to define 14 functionally important star alleles in seven *CYP* genes (Table 2) [62,63]. These 14 star alleles were classified as impaired activity (intermediate metabolizer or poor metabolizer phenotype) alleles [63], and nine of those are classified by the Clinical Pharmacogenetics Implementation Consortium as having decreased or no function [64]. Of these 15 SNPs, seven SNPs in four *CYP* genes have been shown or suggested to affect PQ metabolism and/or treatment response (Table 3) [15,16,18,20]. In addition, there are eight SNPs in six other drug-metabolizing enzyme or drug transporter genes that have also been shown or suggested to affect PQ metabolism or transport and/or response (Table 3) [15,16,19].

A previous study has suggested that a 3-SNP haplotype in the *ABCB1* transporter gene (rs1045642, Ile1215=; rs2032582, Ser893Ala; rs1128503, Gly412=) may have a small effect on parasitemia clearance rate in response to treatment [19]. These SNPs were present at high frequencies in the Malagasy samples (Supplementary Table 4).

Copy number variable analysis

A total of 424 CNVs were observed in the Malagasy samples, and 116 CNVs were observed in the HIHG control sample. CNVs included deletions (estimated copy numbers = zero or one) and duplications (estimated copy numbers = three or four). Their distribution, CNV confidence scores and sizes are presented in Supplementary Table 5. Among the Malagasy samples, they ranged from estimated one CNV per sample to 25 CNVs per sample; 12 samples had ≥10 CNVs. CNVs were identified on all autosomal chromosomes except Chr 18; no CNV was identified on Chr 18 in the HIHG control sample as well. Of all the CNVs in the Malagasy samples, 387 had confidence scores between 100 and 499, 21 had confidence scores between 500 and 999 and 16 had confidence scores over >1000. Only two CNVs were <1000 base pairs (bp) in length, 31 were 1000–9999 bp, 210 were 10,000–99,999 bp, 164 were 100,000–1,000,000 bp and 17 were >1,000,000 bp in length. The largest CNV was 2,387,974 bp in length (estimated copy number = 3).

The same CNVs were identified in more than one Malagasy sample; examples and selected results are presented in Supplementary Table 6 & Figure 1, respectively. In two samples, a region with estimated three copies was

Table 3. Primaquine-related pharmacogene variants in the Malagasy.[†]

Gene	Variant ID	Variant impact	Allele[‡]	Minor allele	Major allele	Minor allele frequency	Ref.
UGT2B7	rs28365063	Arg124=	**3*	G	A	0.136	[16]
ABCG2	rs2231142	Gln141Lys	[§]	A	C	0.191	[16]
CYP3A5	rs776746	Splice defect	**3*	G	A	0.409	[20]
	rs10264272	Splice defect	**6*	A	G	0.109	[20]
CYP2C19	rs4244285	Splice defect	**2*	A	G	0.118	[15,16]
CYP2C9	rs1057910	Ile359Leu	**3*	C	A	0.009	[20]
	rs28371685	Arg335Trp	**11*	A	G	0.027	[20]
CYP2C8	rs11572103	Ile269Phe	**2*	A	T	0.091	[18,20]
	rs11572080	Arg139Lys	**3*	A	G	0.027	[18,20]
SLCO2B1	rs12422149	Arg312Gln	[§]	A	G	0.136	[19]
SLCO1B1	rs2306283	Asn130Asp	Many, including **14*, **15*	A	G	0.264	[19]
	rs11045819	Pro155Thr	Many, including **14*	A	C	0.027	[19]
	rs4149056	Val174Ala	Many, including **5*, **15*	G	A	0.082	[19]
SLCO1A2	rs11568563	Glu172Asp	**3*	C	A	0.018	[19]
MAOA	rs6323	Arg297=	[§]	C	A	0.400	[15]

[†]Implicated by others in primaquine metabolism or transport and/or treatment response in Malaysian [15], Thai [16] and Brazilian [18–20] populations.
[‡]Listed in Pharmacogene Variation Consortium, other nomenclature databases or literature.
[§]Could not be found/has not been determined.

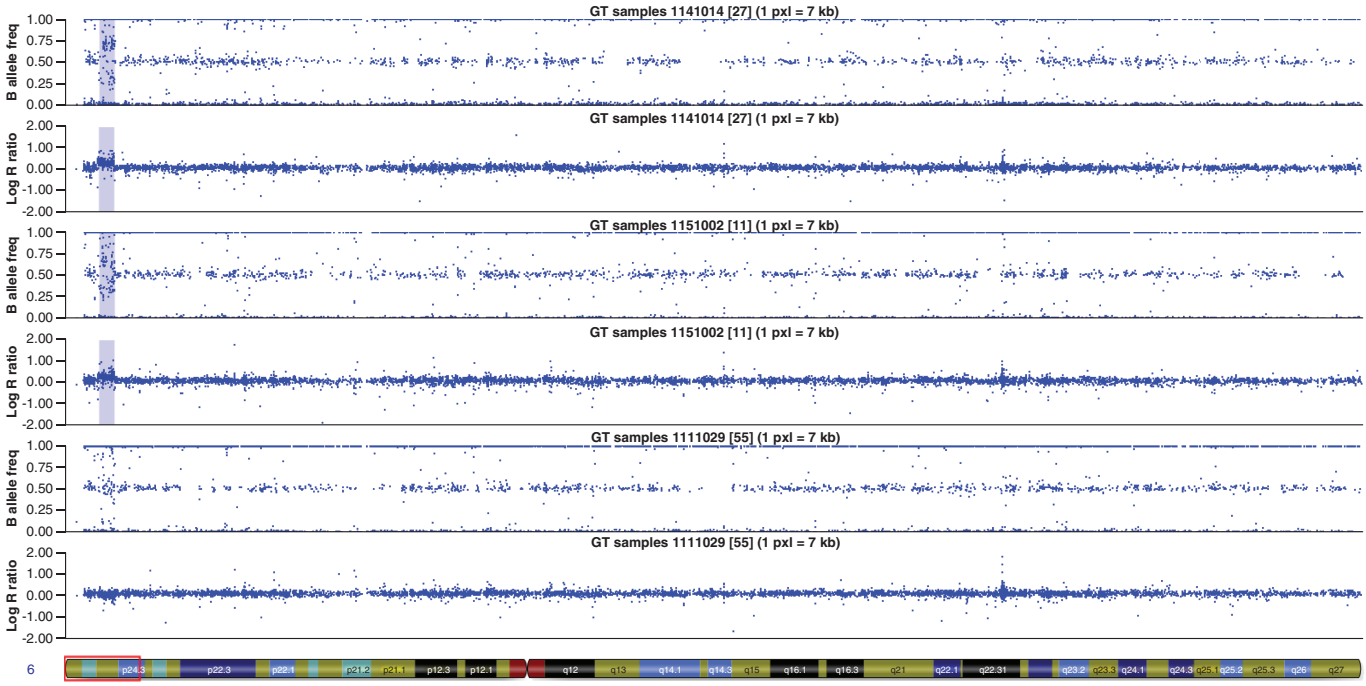

Figure 1. Copy number variation (CNV) 3 on chromosome 6 (bookmarked or shaded, coordinates: 257,999–381,513) identified in samples 1141014 (CNV confidence score 510) and 1151002 (CNV confidence score 332). Note a 'split' in B allele frequency (>0.50 and <0.50) and a 'bump' in the log R ratio (>0.00) plots of those samples, not seen in another sample, 1111029, in which this variation was not identified.

identified on Chr 6. Their BAF and LRR plots in the duplication-containing region (bookmarked or shaded) are noticeably different from those of another sample in which this CNV was not identified (Figure 1).

In some samples, a CNV was identified in a chromosomal region that contains a pharmacogene of interest (Table 4). In one sample, a region with estimated three copies was identified on Chr 4 where *UGT2B10* is located (coordinates: 69,681,712–69,697,741 [Supplementary Table 1]). In addition, in four samples a region with estimated three copies, and in two samples, a region with estimated one copy were identified on Chr 10 where *CYP2E1* is located (coordinates: 135,340,867–135,352,627 [Supplementary Table 1]).

Table 4. Copy number variation in a chromosomal region containing a pharmacogene of interest.[†]

Sample ID	Copy number variation value	Copy number variation confidence score	Chromosome	Start[‡]	End[‡]	Size[§]
1111051	3	112	4	69,679,570	69,715,945	36,375
1134163	3	689	10	135,233,541	135,386,432	152,891
1134154	3	665	10	135,233,541	135,387,399	153,858
1041501	3	651	10	135,241,711	135,380,121	138,410
1020505	3	237	10	135,080,873	135,377,847	296,974
1111051	1	343	10	135,250,209	135,377,847	127,638
1141004	1	241	10	135,242,873	135,377,847	134,974

[†] *UGT2B10* (Chr 4) and *CYP2E1* (Chr 10), Supplementary Table 1.
[‡] Genome build GRCh37.
[§] Value in base pairs.

Principal component analysis

For the genome-wide PCA, 636,381 SNPs were utilized that were in both the 1000 Genomes Project and Malagasy datasets. The results of this genome-wide analysis of 2559 samples are shown in Figure 2A. The east Asian superpopulation (green) and the south Asian superpopulation (yellow) had the tightest clusters of points, demonstrating the least amount of variation across their genomes. The European superpopulation (blue) did not have much variance; however, there appeared to be two distinct European clusters. The African superpopulation (black), as the oldest and most diverse superpopulation, showed more variation than the other nonadmixed continental populations. The American superpopulation (pink) is the newest superpopulation and contains samples that are admixed (people with Mexican ancestry in Los Angeles, USA, and Puerto Ricans in Puerto Rico). There was a large overlap between the American superpopulation and the south Asian and European superpopulations. Also, some American samples had a genetic makeup similar to that of the African superpopulation. The Malagasy population formed a distinct cluster (orange) between the African and east Asian superpopulations.

For the pharmacogene PCA, only 438 out of 1493 SNPs were in both the 1000 Genomes Project and the Madagascar sample datasets. The analysis included 2559 samples, and the results are shown in Figure 2B. Although only a relatively small number of SNPs were available for this analysis, the PCA plot still shows evidence of the continental clusters. However, the clusters overlap more than those in the genome-wide PCA plot and have substantial variability within them. Consistent with the genome-wide PCA results (Figure 2A), the Malagasy samples (orange) were distributed between the African (black) and east Asian (green) superpopulations but were also mixed with the other superpopulations.

Discussion

By performing genome-wide SNP analysis of 55 samples from the western highlands fringe region of Madagascar, the authors asked, *"What knowledge can be gained regarding their pharmacogenomic variation, CNV and population relationships?"* This exploratory study is significant for the following reasons: First, the knowledge of pharmacogenomic variation in Madagascar is sparse. One previous study on 153 individuals from southern Madagascar has reported an allele frequency of 15% ± 2.9% for *CYP2C8*2* (rs11572103, Ile269Phe) [35]. However, the frequency of this allele was 9.1% among the current study's 55 participants from a different region. The authors' recent study reported *CYP2D6* genetic variation for 211 individuals [11] in which eight star alleles (*CYP2D6*2*, **4*, **10*, **17*, **29*, **40*, **41* and **45*), the *CYP2D6*5* gene deletion, gene duplications (*CYP2D6*1x2*, **2x2*, **4x2* and **35x2*) and alleles with tandem arrangements (*CYP2D6*36+*10* and **36x2+*10*) were detected. Also, their recent next-generation sequencing-based *CYP2D6* gene resequencing of 96 of the 211 Malagasy samples revealed new variants and haplotypes [34]. To the best of the authors' knowledge, there are no other reports on the pharmacogenomic variation in Malagasy populations. The present study takes the first step in exploring a set of 28 pharmacogenes selected based on their putative role in PQ metabolism, transport and response, as shown by others, and/or their overall significance for individualized drug therapy. It provides valuable information in identifying 110 coding or splicing SNPs in the Malagasy population and the first account of 13 star alleles of seven *CYP* genes, which cause impaired activity and contribute to intermediate metabolizer or poor metabolizer phenotype (Table 2) [62,63]. Six of these alleles (*CYP3A5*3*, **6* and **7*; *CYP2C19*2*; *CYP3A4*18*; and *CYP2B6*6*) seem to be present at frequencies of ≥10%. Hence, this study provides important information to better understand the role of genetic variation

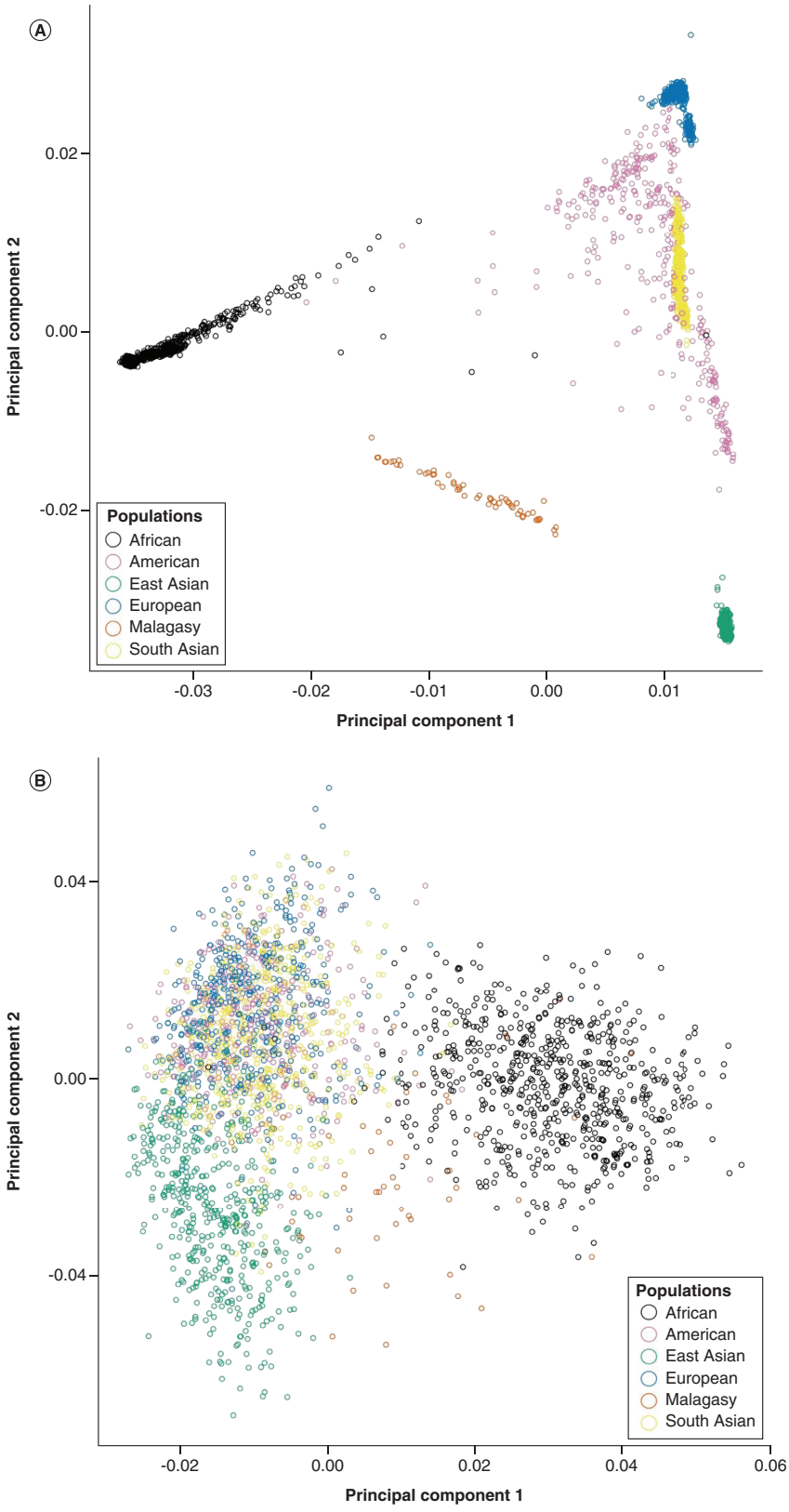

Figure 2. Principal component analysis. (A) Analysis of 636,381 SNPs across the genome in five 1000 Genomes Project superpopulations and the Malagasy population. **(B)** Analysis of 438 pharmacogene-specific SNPs in five 1000 Genomes Project superpopulations and the Malagasy population.

in drug response in Malagasy populations. It is to be noted that the Malagasy 'allele' frequencies represent SNP frequencies, which in some cases may differ from star allele frequencies. Because an SNP may be a part of more than one star allele (or haplotype), the SNP frequency represents the cumulative frequency of those star alleles (e.g., rs3745274 [Gln172His] in *CYP2B6* tags 15 star alleles in addition to the common *CYP2B6*6*) [66].

Second, the study provides valuable information on 15 SNPs (ten nonsynonymous, three impact splicing and two synonymous) in ten genes that, based on previous studies in Malaysian [15], Thai [16] and Brazilian [18–20] populations, may be implicated in PQ metabolism, transport and response (Table 3). Primaquine has been shown to be metabolized principally via CYP2D6 in animal models, *in vitro* enzymatic assays and in human studies [11]. Clinical and laboratory evidence suggests that the efficacy of PQ may depend on genetic variation in *CYP2D6*. *Vivax* malaria relapses following PQ treatment have predominantly been observed in patients carrying *CYP2D6* decreased function and/or nonfunctional alleles in various combinations [9–14]. However, several studies have shown that variation in drug-metabolizing enzymes other than CYP2D6 and drug transporters may also contribute to PQ response and to overall therapeutic efficacy [15,16,18–20]. In fact, large, inter-individual variability in CYP2D6 activity and in the pharmacokinetics of probe drug substrates are observed within genotype groups as well as activity score groups, which remains unexplained [22,67]. While a small portion of this variability may be explained by rare or novel *CYP2D6* variants, additional genetic modifiers together with nongenetic factors are therefore likely to also contribute to phenotypic variability. Understanding these modifiers will allow for more accurate prediction of a patient's metabolizer status for PQ and response to the treatment [22].

Third, the study provides the first account of CNVs, specifically in chromosomal regions that contain pharmacogenes, for Madagascar. In recent years, various algorithms have been developed to identify CNV regions using data from a growing number of different genome-wide SNP arrays [68–70]. Because array-based technologies can only infer CNVs, CNV calling algorithms are prone to producing vastly different findings. Each array and algorithm has different strengths and weaknesses, and therefore they are suited to different types of data. The parameters chosen and their settings may also impact CNV detection and CNV size [68,70]. Using cnvPartition algorithm, CNVs were identified in all 55 samples, ranging from one to 25 CNVs per sample, with 99.5% of the CNVs 1100–2,387, 974 bp in length. Certain CNVs were identified in more than one sample (Supplementary Table 6 & Figure 1). Certain CNVs were identified in chromosomal regions containing the *UGT2B10* and *CYP2E1* pharmacogenes (Table 4 & Supplementary Table 1). However, this algorithm-based identification of CNVs in chromosomal regions that contain pharmacogenes is not necessarily indicative of *UGT2B10* and *CYP2E1* gene copy numbers; additional testing is required to characterize these CNVs precisely. Nevertheless, it is interesting to note that CNVs for *UGT2B10* are rare to none [71,72], whereas those for *CYP2E1* have been well described and functionally characterized [69]. In worldwide populations, *CYP2E1* duplications are much more common (Europeans, 1–3%; Africans, 4–10%; Asians, ≤1%) than deletions (0%–<1%) [69]. A study using liver tissue samples of European descent donors and HumanHap300v2 SNP array data in conjunction with PennCNV algorithm identified *CYP2E1* duplications (CNV = 3 or 4) of approximately 95–145 kb (frequency 1.7%) and deletions (CNV = 1) of approximately 95 kb (frequency <1%), located at Chr 10 coordinates 135,225–135,402 Mb; these findings were confirmed by quantitative PCR using TaqMan copy number assays [69]. Collectively, these findings are similar to those of the present study (Table 4) and thus point to the importance of more in-depth characterization of CNVs in Malagasy populations.

Finally, consistent with previous, more robust genetic and linguistic analyses showing that Malagasy populations have emerged from an admixture between African and east Asian populations, genome-wide PCA clearly shows that the Malagasy population is distributed between the African and east Asian superpopulations (Figure 2A). A recent study has confirmed this admixture and its timeline [73]. The PCA results provide validity to the post-QC genotype data and genomic analyses of this study's Malagasy samples. On the other hand, the pharmacogene PCA did not show the same population relationships for the Malagasy samples (Figure 2B). This indicates that the pharmacogene patterns of differentiation deviated from those of the genome-wide background.

PCAs of pharmacogene-specific SNPs have been performed for various populations [74,75], including ethnically diverse individuals from Canada [75]. Compared with the current study, these studies included a larger number of samples (n = 109 and 112, two different populations [74]; n = 524 [75]), pharmacogenes (n = 65 [74]; n = 220 [75]) and variants (n = 163,722 [74]; n = 2,094 [75]). In the current study's pharmacogene PCA plot, the tightness of the clusters may have been impacted by the small number of Malagasy samples as well as the limited number of pharmacogenes. However, the clustering was not affected by the limited number of SNPs used; the study on the Canadian cohort has shown that PCA can be successfully applied to discriminate clearly between population

clusters using a limited set of markers (n = 250–500) [75]. It is known that genetic and environmental factors have influenced pharmacogenomic characteristics in populations across the globe [76]. Therefore, it is possible that due to unique population admixture, diet and exposure to novel xenobiotics, the pharmacogenomic landscape of Malagasy populations has been shaped differently.

Limitations

There are a number of limitations: first, the small number of individuals, all male, analyzed in this study [31]. Due to the relatively small number of samples, the authors did not perform a statistical comparison of SNP frequencies with those of African and Asian populations that are considered ancestral to Malagasy populations [26–28]. The small sample size may also have partly affected the distribution of monomorphic and polymorphic SNPs overall (40% and 60%, respectively) and in the 28 pharmacogenes (60% and 40%, respectively); some rare or low-frequency SNPs may have turned up as monomorphic.

Second, the inherent limitations of the Infinium MEGA: Illumina and collaborators from several consortia developed this array in October 2014 for genome-wide association studies in populations of diverse ancestries and backgrounds. The two groups most involved with the creation of the MEGA were the Consortium on Asthma among African-ancestry Populations in the Americas and the Population Architecture using Genomics and Epidemiology. Subsequently, the MEGA prerelease version of the array was replaced by the MEGA^{EX} precommercial version and then by an improved commercial version MEG, containing resynthesized, better-quality probes. In this study's initial genotype data, the authors observed that 1.8% of SNPs (31,395 out of 1,708,621) did not pass the No Call threshold of 0.15 recommended by Illumina. Then, during QC, 10.43% of SNPs (174,954 out of 1,677,226) were identified as incorrectly clustered and had to be manually removed, as they could not be rectified by manual reclustering. Thus, a total of 12.08% of SNPs (206,349 out of 1,708,621) had to be removed from the data. Although there are studies in which all three arrays, the MEGA, MEGA^{EX} and MEG, were used for genotyping [77,78], a detailed QC analysis, including percentage of the SNPs removed, is not available in those studies to compare with the present findings.

Post-QC, the MEGA provides limited data regarding variants in the pharmacogenes of interest; most of the genes had <100 SNPs (Table 1). Many functionally relevant and common SNPs are not on the array (e.g., rs12248560 defining *CYP2C19*17*, associated with increased expression) [62,63] or did not pass QC and were removed (e.g., rs2279343 [Lys262Arg] that is part of numerous *CYP2B6* star alleles [65], rs7439366 [His268Tyr] defining *UGT2B7*2* [79,80] and rs1065852 [Pro34Ser], which is found in many *CYP2D6* star allele, including *CYP2D6*10*, an allele that is frequent in Malagasies and east Asians [62,63]). In fact, the highly polymorphic *CYP2D6* had the lowest number of variants (12 monomorphic and one polymorphic; Table 1). At the time this study was conceived in September 2017, the MEGA was the best option for studying a population of mixed ancestry.

Finally, although the MEGA allows CNV identification, it appears to provide rather limited information. The limited SNP coverage for many pharmacogenes, and the percentage of SNPs that did not pass QC, may have affected the detection of CNVs in the chromosomal regions containing pharmacogenes (e.g., *CYP2D6*). In addition, based on the same starting and ending positions of the array markers, the authors observed 'the same' CNVs in more than one Malagasy sample (Supplementary Table 6). However, it is unclear whether the observed duplications/deletions at a particular region were due to a single, frequently occurring CNV (recurring instances of a CNV with matching breakpoints) or were due to similar CNVs with distinct breakpoints that overlap.

Conclusion & future perspective

Although the CYP2D6-dependent pathway appears crucial for the bioactivation of PQ, researchers are also starting to interrogate other pharmacogenes that are potentially involved in PQ metabolism and response to *vivax* infection [15,16,18–20]. Untangling not only the complex routes of PQ metabolism but also the contribution of genomic variation to metabolite formation, especially those that are active against *P. vivax* stages, is essential to eventually developing optimized PQ dosing strategies. In order to address *P. vivax* elimination challenges for the global population at risk, optimizing PQ treatment across human genetic diversity may be needed. Therefore, it is crucial to learn how the efficacy of PQ therapy is affected by genomic variation at interindividual as well as interpopulation levels. Although preliminary in nature, the present study significantly contributes to narrowing the knowledge gap in pharmacogenomic research in Madagascar and identifies variants that, based on previous studies in other populations, may be implicated in PQ response.

The authors have recently launched a study collecting a much larger number of samples. The goals of this study are achieving a better understanding of *CYP2D6* as well as pharmacogenomic variation in Madagascar, and performing genetic association analyses with PQ metabolism and response to *vivax* infection. In this new study, adequate representation of males and females will be ensured, as gender differences are known to impact drug metabolism, efficacy and safety [81–83]. In addition to *CYP2D6*-focused genetic analyses [11,34], authors will analyze the pharmacogenomic landscape using advanced genome-wide arrays [84,85]; one such array that might be suitable for this purpose is the Infinium Global Diversity Array with Enhanced Pharmacogenomics Content [86]. Although CNV calling may be vastly improved due to having more such variants, future array data may still need to be complemented by confirmatory CNV assays to validate findings. Such a comprehensive analysis approach will help identify relevant genomic variation that impacts PQ metabolism and treatment outcomes in uniquely admixed populations of Madagascar, where malaria remains a major public health problem.

Summary points

- CYP2D6 primarily mediates the metabolism of an 8-aminoquinoline antimalarial, primaquine (PQ), a drug of choice against latent malaria caused by *Plasmodium vivax*.
- Clinical and laboratory evidence has suggested that the efficacy of PQ may depend on *CYP2D6* variation. However, recent studies have indicated that other drug-metabolizing enzyme and drug transporter genes may also be involved.
- The authors have been investigating *CYP2D6* genetic variation in Madagascar, where *vivax* malaria is endemic. In addition, the Malagasy populations offer a unique, rich admixture history between east African and southeast Asian ancestral populations. Information on pharmacogenomic variation in Madagascar is scanty.
- In the present study, the authors applied whole genome genotyping to study pharmacogenomic variation most relevant to PQ, copy number variation, and population relationships of 55 Malagasy samples.
- This study reports for the first time, to the authors' knowledge, a large number of coding or splicing variants, including those that, based on previous studies in other populations, may be implicated in PQ response. In addition, there are copy number variations, specifically in chromosomal regions that contain pharmacogenes.
- The present study significantly contributes to narrowing the knowledge gap in pharmacogenomic research in Madagascar, which will help identify relevant genomic variation that impacts PQ metabolism and treatment outcomes.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/suppl/10.2217/pgs-2023-0091

Author contributions

SM Williams and PA Zimmerman conceived the study and secured the funding. RK Mehlotra prepared the DNA samples for the genotyping, performed GenomeStudio analyses and wrote the manuscript. EY Cramer and J Bartlett analyzed the post-quality control genotype data. ER Chan downloaded and processed the 1000 Genomes Project data and assisted in the principal component analysis. A Gaedigk contributed to data interpretation and manuscript preparation. AC Ratsimbasoa was the principal collaborator and overall supervisor of the field studies. All authors contributed to the manuscript editing and approved the final version.

Acknowledgments

The authors thank all study participants, local health officials, field doctors and project technicians for their participation and support. The authors appreciate the help of Ioanna Konidari (director, Genotyping Core, HIHG, University of Miami) for performing Infinium Multi-Ethnic Genotyping Array analysis. RK Mehlotra is indebted to Kim Brustoski (Population and Quantitative Health Sciences, CWRU), Simone Edelheit (Genomics Core, CWRU) and Jose Gonzales (Technical Applications Scientist II, Illumina) for their invaluable help with GenomeStudio processing and quality control of the genotype data. In addition, RK Mehlotra thanks Kim Brustoski for generating Supplementary Figure 1 (Supplementary Document 1) and critically reading the manuscript and Jasmine Olvany and Quentin Watson for their comments on the manuscript. The authors' sincere thanks are also due to Alexandra Stewart and James McCarthy for critically reading the manuscript.

Financial & competing interests disclosure

This study was supported by R01 grants from the NIH (AI097366 to PA Zimmerman and AI148469 to SM Williams and PA Zimmerman). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Institutional review board statement/informed consent statement

The study was approved by the University Hospitals of Cleveland Institutional Review Board (#09-13-01), CWRU IRB (STUDY20191615), the Division of Microbiology and Infectious Diseases/NIAID/National Institutes of Health (#13-0067) and the Madagascar Ministry of Health Ethics Committee (#099 and #52 MSANP/SG/AMM/CERBM). Written informed consent was obtained from all subjects, or subject guardians, prior to enrollment.

Data availability statement

Data from this study are available through one supplementary document, four tables, two figures and five (Supplementary Tables 2–6) in the manuscript.

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Baird JK. Basic research of *Plasmodium vivax* biology enabling its management as a clinical and public health problem. *Front. Cell. Infect. Microbiol.* 11, 696598 (2021).
2. Flannery EL, Kangwanrangsan N, Chuenchob V *et al. Plasmodium vivax* latent liver infection is characterized by persistent hypnozoites, hypnozoite-derived schizonts, and time-dependent efficacy of primaquine. *Mol. Ther. Methods Clin. Dev.* 26, 427–440 (2022).
3. White NJ. Determinants of relapse periodicity in *Plasmodium vivax* malaria. *Malar. J.* 10, 297 (2011).
4. White NJ, Imwong M. Relapse. *Adv. Parasitol.* 80, 113–150 (2012).
5. Lacerda MVG, Llanos-Cuentas A, Krudsood S *et al.* Single-dose tafenoquine to prevent relapse of *Plasmodium vivax* malaria. *N. Engl. J. Med.* 380(3), 215–228 (2019).
6. Llanos-Cuentas A, Lacerda MVG, Hien TT *et al.* Tafenoquine versus primaquine to prevent relapse of *Plasmodium vivax* malaria. *N. Engl. J. Med.* 380(3), 229–241 (2019).
7. Poespoprodjo JR, Burdam FH, Candrawati F *et al.* Supervised versus unsupervised primaquine radical cure for the treatment of falciparum and vivax malaria in Papua, Indonesia: a cluster-randomised, controlled, open-label superiority trial. *Lancet Infect. Dis.* 22(3), 367–376 (2022).
8. Rodrigo C, Rajapakse S, Fernando D. Tafenoquine for preventing relapse in people with *Plasmodium vivax* malaria. *Cochrane Database Syst. Rev.* 9, CD010458 (2020).
9. Bennett JW, Pybus BS, Yadava A *et al.* Primaquine failure and cytochrome P-450 2D6 in *Plasmodium vivax* malaria. *N. Engl. J. Med.* 369(14), 1381–1382 (2013).

•• **This is the first report of *CYP2D6* variation associated with primaquine treatment failure.**

10. Ingram RJ, Crenna-Darusallam C, Soebianto S, Noviyanti R, Baird JK. The clinical and public health problem of relapse despite primaquine therapy: case review of repeated relapses of *Plasmodium vivax* acquired in Papua New Guinea. *Malar. J.* 13, 488 (2014).

• **This is another, earlier case report of *CYP2D6* variation associated with primaquine treatment failure.**

11. Mehlotra RK, Gaedigk A, Howes RE, Rakotomanga TA, Ratsimbasoa AC, Zimmerman PA. *CYP2D6* genetic variation and its implication for vivax malaria treatment in Madagascar. *Front. Pharmacol.* 12, 654054 (2021).

•• **This article is the first report of *CYP2D6* variation in Madagascar.**

12. Stewart AGA, Zimmerman PA, McCarthy JS. Genetic variation of *G6PD* and *CYP2D6*: clinical implications on the use of primaquine for elimination of *Plasmodium vivax*. *Front. Pharmacol.* 12, 784909 (2021).
13. Suarez-Kurtz G. Impact of *CYP2D6* genetic variation on radical cure of *Plasmodium vivax* malaria. *Clin. Pharmacol. Ther.* 110(3), 595–598 (2021).
14. Choi S, Choi H, Park SY *et al.* Association between CYP2D6 phenotype and recurrence of *Plasmodium vivax* infection in South Korean patients. *Malar. J.* 21(1), 289 (2022).
15. Ariffin NM, Islahudin F, Kumolosasi E, Makmor-Bakry M. Effects of MAO-A and CYP450 on primaquine metabolism in healthy volunteers. *Parasitol. Res.* 118(3), 1011–1018 (2019).

- **This study describes *MAOA* and *CYP2C19* variation and primaquine response in Malaysians.**

16. Chamnanphon M, Gaedigk A, Puangpetch A *et al.* Pharmacogene variation in Thai *Plasmodium vivax* relapse patients treated with a combination of primaquine and chloroquine. *Pharmgenomics Pers. Med.* 13, 1–12 (2020).

- **This study describes *CYP2C19, UGT2B7* and *ABCG2* variation and primaquine response in Thais.**

17. Li XQ, Bjorkman A, Andersson TB, Gustafsson LL, Masimirembwa CM. Identification of human cytochrome P(450)s that metabolise anti-parasitic drugs and predictions of *in vivo* drug hepatic clearance from *in vitro* data. *Eur. J. Clin. Pharmacol.* 59(5-6), 429–442 (2003).

- **This *in vitro* study identified CYP2D6 and CYP1A2 involved in primaquine metabolism.**

18. Silvino AC, Costa GL, Araujo FC *et al.* Variation in human cytochrome P-450 drug-metabolism genes: a gateway to the understanding of *Plasmodium vivax* relapses. *PLOS ONE* 11(7), e0160172 (2016).

- **This study describes *CYP2C8* variation and primaquine response in Brazilians.**

19. Sortica VA, Lindenau JD, Cunha MG *et al. SLCO1A2*, *SLCO1B1* and *SLCO2B1* polymorphisms influences chloroquine and primaquine treatment in *Plasmodium vivax* malaria. *Pharmacogenomics* 18(15), 1393–1400 (2017).

- **This study describes *SLCO1A2, SLCO1B1* and *SLCO2B1* variation and primaquine response in Brazilians.**

20. Sortica VA, Lindenau JD, Cunha MG *et al.* The effect of SNPs in CYP450 in chloroquine/primaquine *Plasmodium vivax* malaria treatment. *Pharmacogenomics* 17(17), 1903–1911 (2016).

- **This study describes *CYP2C8, CYP2C9* and *CYP3A5* variation and primaquine response in Brazilians.**

21. Hayeshi R, Masimirembwa C, Mukanganyama S, Ungell AL. The potential inhibitory effect of antiparasitic drugs and natural products on P-glycoprotein mediated efflux. *Eur. J. Pharm. Sci.* 29(1), 70–81 (2006).

22. Gaedigk A, Dinh JC, Jeong H, Prasad B, Leeder JS. Ten years' experience with the CYP2D6 activity score: a perspective on future investigations to improve clinical predictions for precision therapeutics. *J. Pers. Med.* 8(2), 15 (2018).

23. Brucato N, Kusuma P, Cox MP *et al.* Malagasy genetic ancestry comes from an historical Malay trading post in southeast Borneo. *Mol. Biol. Evol.* 33(9), 2396–2400 (2016).

24. Kusuma P, Brucato N, Cox MP *et al.* Contrasting linguistic and genetic origins of the Asian source populations of Malagasy. *Sci. Rep.* 6, 26066 (2016).

25. Kusuma P, Cox MP, Pierron D *et al.* Mitochondrial DNA and the Y chromosome suggest the settlement of Madagascar by Indonesian sea nomad populations. *BMC Genomics* 16, 191 (2015).

26. Pierron D, Heiske M, Razafindrazaka H *et al.* Strong selection during the last millennium for African ancestry in the admixed population of Madagascar. *Nat. Commun.* 9(1), 932 (2018).

27. Pierron D, Heiske M, Razafindrazaka H *et al.* Genomic landscape of human diversity across Madagascar. *Proc. Natl Acad. Sci. USA* 114(32), e6498–e6506 (2017).

28. Heiske M, Alva O, Pereda-Loth V *et al.* Genetic evidence and historical theories of the Asian and African origins of the present Malagasy population. *Hum. Mol. Genet.* 30(R1), R72–R78 (2021).

29. Howes RE, Franchard T, Rakotomanga TA *et al.* Risk factors for malaria infection in central Madagascar: insights from a cross-sectional population survey. *Am. J. Trop. Med. Hyg.* 99(4), 995–1002 (2018).

30. National Malaria Control Programme of Madagascar. National strategic plan for malaria control in Madagascar 2018–2022. Progressive malaria elimination from Madagascar (2017).

31. Howes RE, Chan ER, Rakotomanga TA *et al.* Prevalence and genetic variants of G6PD deficiency among two Malagasy populations living in *Plasmodium vivax*-endemic areas. *Malar. J.* 16(1), 139 (2017).

32. Menard D, Barnadas C, Bouchier C *et al. Plasmodium vivax* clinical malaria is commonly observed in Duffy-negative Malagasy people. *Proc. Natl Acad. Sci. USA* 107(13), 5967–5971 (2010).

33. Willie N, Mehlotra RK, Howes RE *et al.* Insights into the performance of SD Bioline Malaria Ag P.f/Pan rapid diagnostic test and *Plasmodium falciparum* histidine-rich protein 2 gene variation in Madagascar. *Am. J. Trop. Med. Hyg.* 98(6), 1683–1691 (2018).

34. Chan ER, Mehlotra RK, Pirani KA *et al. CYP2D6* gene resequencing in the Malagasy, a population at the crossroads between Asia and Africa: a pilot study. *Pharmacogenomics* doi: 10.2217/pgs-2021-0146 (2022) (Epub ahead of print).

- •• **This recent article describes new *CYP2D6* variants and haplotypes in Madagascar.**

35. Paganotti GM, Gramolelli S, Tabacchi F *et al.* Distribution of human *CYP2C8*2* allele in three different African populations. *Malar. J.* 11, 125 (2012).

- **This study describes *CYP2C8* variation in Madagascar.**

36. Chaurasiya ND, Ganesan S, Nanayakkara NP, Dias LR, Walker LA, Tekwani BL. Inhibition of human monoamine oxidase A and B by 5-phenoxy 8-aminoquinoline analogs. *Bioorg. Med. Chem. Lett.* 22(4), 1701–1704 (2012).

- **This study describes a connection between MAOA and MAOB enzymes and primaquine metabolism.**

37. Chaurasiya ND, Liu H, Doerksen RJ, Nanayakkara NPD, Walker LA, Tekwani BL. Enantioselective interactions of anti-infective 8-aminoquinoline therapeutics with human monoamine oxidases A and B. *Pharmaceuticals (Basel)* 14(5), 398 (2021).

- **This study describes a connection between MAOA and MAOB enzymes and primaquine metabolism.**

38. Avula B, Tekwani BL, Chaurasiya ND *et al.* Metabolism of primaquine in normal human volunteers: investigation of phase I and phase II metabolites from plasma and urine using ultra-high performance liquid chromatography-quadrupole time-of-flight mass spectrometry. *Malar. J.* 17(1), 294 (2018).

•• **This study provides a comprehensive view of pathways for primaquine metabolism in humans.**

39. Avula B, Tekwani BL, Chaurasiya ND *et al.* Profiling primaquine metabolites in primary human hepatocytes using UHPLC-QTOF-MS with 13 C stable isotope labeling. *J. Mass Spectrom.* 48(2), 276–285 (2013).

40. Fasinu PS, Avula B, Tekwani BL *et al.* Differential kinetic profiles and metabolism of primaquine enantiomers by human hepatocytes. *Malar. J.* 15, 224 (2016).

41. Rowland A, Miners JO, Mackenzie PI. The UDP-glucuronosyltransferases: their role in drug metabolism and detoxification. *Int. J. Biochem. Cell Biol.* 45(6), 1121–1132 (2013).

42. Illumina. Infinium® genotyping data analysis. www.illumina.com/Documents/products/technotes/technote_infinium_genotyping_data_analysis.pdf

43. Illumina. Illumina GenCall data analysis software. www.illumina.com/Documents/products/technotes/technote_gencall_data_analysis_software.pdf

44. Illumina. GenomeStudio v2.0.5. https://support.illumina.com/array/array_software/genomestudio/downloads.html

45. Illumina. GenomeStudio® genotyping module v2.0 software guide. https://support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/genomestudio/genomestudio-2-0/genomestudio-genotyping-module-v2-user-guide-11319113-01.pdf

46. Patel H, Lee SH, Breen G, Menzel S, Ojewunmi O, Dobson RJB. The COPILOT raw Illumina genotyping QC protocol. *Curr. Protoc.* 2(4), e373 (2022).

47. Zhao S, Jing W, Samuels DC, Sheng Q, Shyr Y, Guo Y. Strategies for processing and quality control of Illumina genotyping arrays. *Brief Bioinform.* 19(5), 765–775 (2018).

48. GenomeStudio 2.0 plug-ins. Illumina. https://support.illumina.com/downloads/genomestudio-2-0-plug-ins.html

49. Cunningham F, Allen JE, Allen J *et al.* Nucleic acids research 50(D1), D988–D995 (2022).

50. Sherry ST, Ward M, Sirotkin K. dbSNP—database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome research.* 9(8), 677–679 (1999).

51. Chen S, Francioli LC, Goodrich JK. A genome-wide mutational constraint map quantified from variation in 76,156 human genomes. *bioRxiv.* doi: 10.1101/2022.03.20.485034 (2022) (Epub ahead of print).

52. Gaedigk A, Casey ST, Whirl-Carrillo M, Miller NA, Klein TE. Pharmacogene Variation Consortium: a global resource and repository for pharmacogene variation. *Clin. Pharmacol. Ther.* 110(3), 542–545 (2021).

53. Gaedigk A, Casey ST, Whirl-Carrillo M, Miller NA, Klein TE. Pharmacogene Variation Consortium: A Global Resource and Repository for Pharmacogene Variation. *Clinical pharmacology and therapeutics* 111(3), 542–545 (2021).

54. UGT Nomenclature Committee. www.pharmacogenomics.pha.ulaval.ca/ugt-alleles-nomenclature/

55. Mackenzie PI, Bock KW, Burchell B *et al.* Nomenclature update for the mammalian UDP glycosyltransferase (UGT) gene superfamily. *Pharmacogenet. Genomics* 15(10), 677–685 (2005).

56. Fairley S . 48(D1) D941-7 (2020)., Lowy-Gallego E, Perry E, Flicek P. The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic acids research* 48(D1), D941–7 (2020).

57. Illumina. DNA copy number and loss of heterozygosity analysis algorithms. www.illumina.com/documents/products/technotes/technote_cnv_algorithms.pdf

58. Illumina. Interpreting infinium assay data for whole-genome structural variation. www.illumina.com/Documents/products/technotes/technote_cytoanalysis.pdf

59. R Core Team. R: a language and environment for statistical computing. *R Foundation for Statistical Computing* Vienna, Austria (2021). www.R-project.org/

60. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28(24), 3326–3328 (2012).

61. Barbarino JM, Haidar CE, Klein TE, Altman RB. PharmGKB summary: very important pharmacogene information for *UGT1A1*. *Pharmacogenet. Genomics* 24(3), 177–183 (2014).

62. Zhou Y, Ingelman-Sundberg M, Lauschke VM. Worldwide distribution of cytochrome P450 alleles: a meta-analysis of population-scale sequencing projects. *Clin. Pharmacol. Ther.* 102(4), 688–700 (2017).

- **This study provides a comprehensive account of cytochrome P450 gene variation in worldwide populations.**

63. Zhou Y, Lauschke VM. The genetic landscape of major drug metabolizing cytochrome P450 genes – an updated analysis of population-scale sequencing data. *Pharmacogenomics J.* doi: 10.1038/s41397-022-00288-2 (2022) (Epub ahead of print).

- **This study provides an updated, comprehensive account of cytochrome P450 gene variation in worldwide populations.**

64. CPIC (Clinical Pharmacogenetics Implementation Consortium). The Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines. https://cpicpgx.org/guidelines/
65. Zanger UM, Klein K. Pharmacogenetics of cytochrome P450 2B6 (*CYP2B6*): advances on polymorphisms, mechanisms, and clinical relevance. *Front. Genet.* 4, 24 (2013).
66. PharmVar Pharmacogene Variation Consortium. PharmVar gene *CYP2B6*. www.pharmvar.org/gene/CYP2B6
67. Rudesheim S, Selzer D, Fuhr U, Schwab M, Lehr T. Physiologically-based pharmacokinetic modeling of dextromethorphan to investigate interindividual variability within CYP2D6 activity score groups. *CPT Pharmacometrics Syst. Pharmacol.* doi: 10.1002/psp4.12776 (2022) (Epub ahead of print).
68. Dellinger AE, Saw SM, Goh LK, Seielstad M, Young TL, Li YJ. Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays. *Nucleic Acids Res.* 38(9), e105 (2010).
69. Tremmel R, Klein K, Winter S, Schaeffeler E, Zanger UM. Gene copy number variation analysis reveals dosage-insensitive expression of CYP2E1. *Pharmacogenomics J.* 16(6), 551–558 (2016).
70. Tsuang DW, Millard SP, Ely B *et al.* The effect of algorithms on copy number variant detection. *PLOS ONE* 5(12), e14456 (2010).
71. Santos M, Niemi M, Hiratsuka M *et al.* Novel copy-number variations in pharmacogenes contribute to interindividual differences in drug pharmacokinetics. *Genet. Med.* 20(6), 622–629 (2018).
72. Coe BP, Witherspoon K, Rosenfeld JA *et al.* Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat. Genet.* 46(10), 1063–1071 (2014).
73. Alva O, Leroy A, Heiske M *et al.* The loss of biodiversity in Madagascar is contemporaneous with major demographic events. *Curr. Biol.* 32(23), 4997–5007; e4995 (2022).
74. Bharti N, Banerjee R, Achalere A, Kasibhatla SM, Joshi R. Genetic diversity of 'Very Important Pharmacogenes' in two South-Asian populations. *Peer J.* 9, e12294 (2021).
75. Visscher H, Ross CJ, Dube MP *et al.* Application of principal component analysis to pharmacogenomic studies in Canada. *Pharmacogenomics J.* 9(6), 362–372 (2009).
76. Li J, Zhang L, Zhou H, Stoneking M, Tang K. Global patterns of genetic diversity and signals of natural selection for human ADME genes. *Hum. Mol. Genet.* 20(3), 528–540 (2011).

- **This study provides comprehensive genetic characteristics of absorption, distribution, metabolism and excretion (ADME) genes in worldwide populations.**

77. Blout Zawatsky CL, Shah N, Machini K *et al.* Returning actionable genomic results in a research biobank: analytic validity, clinical implementation, and resource utilization. *Am. J. Hum. Genet.* 108(12), 2224–2237 (2021).
78. McCcoy TH, Castro VM, Snapper LA, Hart KL, Perlis RH. Efficient genome-wide association in biobanks using topic modeling identifies multiple novel disease loci. *Mol. Med.* 23, 285–294 (2017).
79. Bhasker CR, McKinnon W, Stone A *et al.* Genetic polymorphism of UDP-glucuronosyltransferase 2B7 (*UGT2B7*) at amino acid 268: ethnic diversity of alleles and potential clinical significance. *Pharmacogenetics* 10(8), 679–685 (2000).
80. Duguay Y, Baar C, Skorpen F, Guillemette C. A novel functional polymorphism in the uridine diphosphate-glucuronosyltransferase 2B7 promoter with significant impact on promoter activity. *Clin. Pharmacol. Ther.* 75(3), 223–233 (2004).
81. Mezzalira S, Toffoli G. The effects of sex on pharmacogenetically guided drug treatment. *Pharmacogenomics* 22(15), 959–962 (2021).
82. Yang L, Li Y, Hong H *et al.* Sex differences in the expression of drug-metabolizing and transporter genes in human liver. *J. Drug Metab. Toxicol.* 3(3), 1000119 (2012).
83. Zanger UM, Schwab M. Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol. Ther.* 138(1), 103–141 (2013).
84. Arbitrio M, Scionti F, Di Martino MT *et al.* Pharmacogenomics biomarker discovery and validation for translation in clinical practice. *Clin. Transl Sci.* 14(1), 113–119 (2021).
85. Van Der Lee M, Kriek M, Guchelaar HJ, Swen JJ. Technologies for pharmacogenomics: a review. *Genes (Basel)* 11(12), 1456 (2020).
86. Illumina. Infinium™ Global Diversity Array with Enhanced PGx-8 v1.0. www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/infinium-global-diversity-array-with-enhanced-pgx-data-sheet-m-gl-00031/infinium-global-diversity-array-with-enhanced-pgx-data-sheet-M-GL-00031.pdf