ACS Infectious Diseases

Article

pubs.acs.org/journal/aidcbc

Next-Generation Sequencing of *Plasmodium vivax* Patient Samples Shows Evidence of Direct Evolution in Drug-Resistance Genes

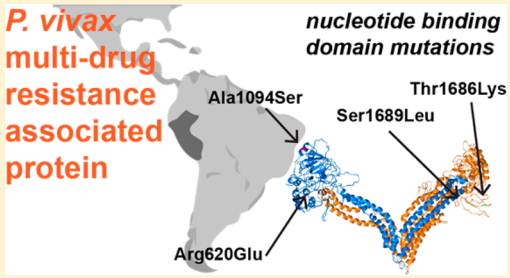
Erika L. Flannery,[†] Tina Wang,[†] Ali Akbari,[‡] Victoria C. Corey,[†] Felicia Gunawan,[†] A. Taylor Bright,[†] Matthew Abraham,[†] Juan F. Sanchez,[§] Meddly L. Santolalla,[§] G. Christian Baldeviano,[§] Kimberly A. Edgel,[§] Luis A. Rosales,[§] Andrés G. Lescano,[§] Vineet Bafna,[‡] Joseph M. Vinetz,[†] and Elizabeth A. Winzeler*[,†]

[†]School of Medicine and [‡]Jacobs School of Engineering, University of California—San Diego, 9500 Gilman Drive, La Jolla, California 92093, United States

[§]U.S. Naval Medical Research Unit No. 6 (NAMRU-6), Avenida Venezuela Cuadra 36 S/N, Centro Médico Naval, Lima Callao 02, Peru

Supporting Information

ABSTRACT: Understanding the mechanisms of drug resistance in *Plasmodium vivax*, the parasite that causes the most widespread form of human malaria, is complicated by the lack of a suitable long-term cell culture system for this parasite. In contrast to *P. falciparum*, which can be more readily manipulated in the laboratory, insights about parasite biology need to be inferred from human studies. Here we analyze the genomes of parasites within 10 human *P. vivax* infections from the Peruvian Amazon. Using next-generation sequencing we show that some *P. vivax* infections analyzed from the region are likely polyclonal. Despite their polyclonality we observe limited parasite genetic diversity by showing that three or fewer haplotypes comprise 94% of the examined genomes, suggesting the recent introduction of parasites into this geographic region. In contrast we find more than three haplotypes in putative drug-resistance genes, including the gene encoding dihydrofolate reductase-thymidylate synthase and the *P. vivax* multidrug resistance associated transporter, suggesting that resistance mutations have arisen independently. Additionally, several drug-resistance genes are located in genomic regions with evidence of increased copy number. Our data suggest that whole genome sequencing of malaria parasites from patients may provide more insight about the evolution of drug resistance than genetic linkage or association studies, especially in geographical regions with limited parasite genetic diversity.



KEYWORDS: *malaria, Plasmodium vivax, genomics, haplotype, clone, recombination, mutation*

BACKGROUND

At present there is still no long-term sustainable culture system for *Plasmodium vivax*, the parasite responsible for 47% of malaria outside of sub-Saharan Africa.[1] *P. vivax* malaria, which can be fatal and is endemic in 57 countries,[1] is primarily treated with chemotherapy. Chloroquine plus primaquine is used in many areas of the world, such as South America, where resistance is not widespread,[2] while artemisinin-based combination therapies (ACTs) are used in high-transmission areas, such as Papua New Guinea, where chloroquine-resistant *P. vivax* has been observed.[3] Antifolates, such as the sulfadoxine-pyrimethamine combination (Fansidar), are not recommended for use because of widespread parasite resistance but have been used extensively in the past. Pyrimethamine was added to table salt in some countries, including Brazil in the 1950s.[4,5] Although *P. vivax* alleles conferring resistance to pyrimethamine are well known[6−9] and are found in the Amazon basin, alleles conferring resistance to chloroquine, artemisinin, and primaquine (the only therapy that prevents vivax malaria relapse) in *P. vivax* remain obscure.

To improve surveillance for drug resistance and to prevent complications from treatment with inadequate therapies it would be desirable to identify molecular markers of drug resistance. However, the lack of long-term culture methods for *P. vivax* complicates the identification of resistance genes using *in vitro* evolution or genetic crosses. On the other hand, given that there is no shortage of patients with *P. vivax* malaria, population genetic studies may provide an attractive method for mapping genes involved in drug resistance. In fact genome-wide association studies were recently used to identify loci involved in *P. falciparum* artemisinin resistance.[10−13] Although population-based approaches are theoretically possible with *P. vivax*, to utilize these techniques it would be desirable to understand the frequency of polyclonal infections, levels of population structure, and rates of parasite inbreeding in different locations as this can affect the design of population-based studies.

Received: April 20, 2015
Published: July 13, 2015

DOI: 10.1021/acsinfecdis.5b00049
ACS Infect. Dis. 2015, 1, 367−379

Although researchers have sought to address such questions about *P. vivax* populations using microsatellite markers and *P. vivax* DNA derived from patient infections, microsatellite markers are genetically less stable than single nucleotide variants (SNVs) and may provide overestimates of genetic diversity. Conversely, reliance on too few microsatellite markers may underestimate diversity, especially in closely related parasites.[14] Here we use next-generation sequencing (NGS) to analyze the genomes of *P. vivax* parasites collected from patients in an isolated mining camp in Peru, a country that has experienced a resurgence in malaria cases over the last half a century following near elimination.[15,16] Even with a small sample size we find evidence of intrainfection polyclonality, interinfection clonality, and inbreeding, all factors that could affect sample size calculations in genetic mapping studies. Despite a relatively homogeneous population, we nevertheless find evidence of recent evolution in known drug-resistance genes, indicating that the direct examination of genome sequences and candidate drug-resistance genes may provide insight into the genetic basis of drug resistance in this parasite.

RESULTS AND DISCUSSION

Whole Genome Sequencing Identifies Polyclonality in the Madre de Dios Region. To investigate parasite genetic diversity and clonality in the Peruvian Amazon, 10 human whole blood samples (Mdio01, Mdio03−11) were collected from *P. vivax*-infected patients presenting to a local health post in the Madre de Dios region of Southeastern Peru (Table S1). Blood was drawn from patients and filtered with CF11 cellulose powder (Whatman) to remove human white cells before DNA isolation. Microsatellite analysis was first performed at five markers to establish a baseline understanding of the clonality for each sample (Table 1), and then whole genome sequencing was performed to further investigate the composition of the samples. Illumina paired-end sequencing yielded a mean coverage of 10.7× per sample with an average of 29.2% of the reads aligning to the *P. vivax* Sal1 reference genome. Because of the poor coverage in four of the samples, they were not included in further whole genome analysis. The six samples (Mdio03, 05, 06, 08, 10, and 11) used for subsequent analysis had a mean coverage of 16.6× (Table 2).

Table 1. Microsatellite Repeat Size for Five Markers (bp)[a]

		microsatellite marker				
sample	collection date	14.185	4.2771	6.34	12.335	7.67
Mdio01	May 09, 2012	283	85	160	170	101
Mdio03	April 21, 2012	286	80	160	164	103
Mdio04	April 26, 2012	283	85/96	160	162	101
Mdio05	April 21, 2012	270	85	160	170	101
Mdio06	May 11, 2012	283	85/98	158	160	101
Mdio07	April 23, 2012	286	98	162	N/A	101
Mdio08	May 02, 2012	283	85/98	160	160	103
Mdio09	May 14, 2012	283	80/95	134	170	99
Mdio10	May 07, 2012	270	85	158	162	101
Mdio11	April 19 2012	283	85	134	170	103

[a]Alleles for each microsatellite loci are reported as PCR product size. Three to four unique genotypes are identified at each marker. All haplotypes observed are unique and nonoverlapping. bp, base pairs; N/A, not available.

All 10 samples were used for microsatellite analysis, and 4 of these samples (Mdio04, 6, 8, and 9) had multiple microsatellite repeat sizes suggesting preliminary evidence of polyclonality in these infections (Table 1). Polyclonal infections could arise from mosquito injection of sibling parasites derived from a single meiosis into a person, from "superinfection" from multiple infectious mosquito bites, with each delivering subsequent genetically unrelated parasites, or from a combination of a relapse and a new infection. To gain further insight into the cause of polyclonality in these samples we genotyped each strain at 14 651 polymorphic sites that were confidently called in all samples (Methods, Table S2). Of these, a mean of 11 665 sites were polymorphic per sample (Table 3).

Table 2. Whole-Genome Resequencing Statistics for 10 Patient Whole Blood Samples Analyzed[a]

sample	total reads (millions)	mean coverage (*x*)	% reads aligned to *P. vivax* *SalI*	% bases covered by 5 or more reads	% bases covered by 10 or more reads
Mdio01	22.1	6.35	25.7	6.8	4.7
Mdio03	18.5	8.23	19.9	50	24
Mdio04	13.8	4.24	20.4	9.3	1.5
Mdio05	14.2	22.3	49.0	76	65
Mdio06	14.7	28.1	55.4	78	70
Mdio07	13.3	4.23	20.7	9.4	1.3
Mdio08	13.5	10.6	30.3	62	36
Mdio09	6.72	1.74	18.3	1.0	0.5
Mdio10	11.5	16.1	43.5	73	57
Mdio11	38.0	14.4	21.7	62	35
average	15.8	10.7	29.2	39	27

[a]Four samples (Mdio01, Mdio04, Mdio07, and Mdio09) were excluded from genotyping studies because the *P. vivax* read coverage was low. This is most likely due to high levels of contaminating human DNA or in the case of Mdio01, contamination with *P. falciparum* DNA.

Within a given sample, many polymorphic sites had mapped reads bearing two different alleles (mixed reads, designated the major and minor alleles based on read frequency), suggesting that the infection consisted of more than one parasite clone. The NGS data showed that all samples carried mixed-allele polymorphisms, ranging from 6% of genotyped sites in Mdio3 to 28% of the sites in Mdio10, compared to the microsatellite data, which showed that only two of the six high-coverage samples were polyclonal. We additionally calculated the within-host diversity metric, F_{ws},[17,18] an indicator of the multiplicity of infection (MOI) that considers the relative proportions of clones within a sample as well as the genetic divergence of the sample from the local parasite population. This metric is more robust to locus-specific artifacts that may occur using microsatellite analysis to calculate MOI. The F_{ws} for the Madre de Dios isolates was between 0.46 and 0.83 (Table 3), where values $\leq$0.7 are considered highly diverse. We previously observed F_{ws} metrics between 0.97 and 0.94 for monoclonal *P. vivax* relapse infections.[14] Thus, these data suggest that the Madre de Dios infections may be polyclonal, particularly Mdio10, which has the lowest F_{ws} and the greatest number of polymorphic sites in the genotyped set.

Madre de Dios *P. vivax* Genomes Are Composed of DNA from Three Hypothetical Haplotypes. Previous population studies of *P. vivax*, specifically those in the Peruvian Amazon, have revealed substantial genetic diversity among parasite populations with the vast majority of every infection examined appearing to be composed of a different haplotype using tandem repeat polymorphic markers.[19−21] Indeed our

DOI: 10.1021/acsinfecdis.5b00049
ACS Infect. Dis. 2015, 1, 367−379

Table 3. Genotyping Summary[a]

strain	Mdio03	Mdio05	Mdio06	Mdio08	Mdio10	Mdio11
genotyped	32 199	32 199	32 199	32 199	32 199	32 199
SNVs covered >5 reads for major allele	21 139	30 915	31 424	25 487	29 887	26 041
SNVs passing filtering criteria with unmapped contigs	15 380	15 380	15 380	15 380	15 380	15 380
genotyped sites mapping to 14 chromosomes	14 651	14 651	14 651	14 651	14 651	14 651
polymorphic sites						
single allele SNVs	8494	10 989	11 240	11 023	8868	9360
mixed allele SNVs	1016	1119	1141	1319	4132	1286
F_{ws}	0.770	0.827	0.818	0.829	0.464	0.781
SNVs in coding regions[b]						
nonsynonymous	2465	3272	3360	3274	2617	2371
synonymous	1718	1952	2298	2264	1830	1718
unique SNVs	7	2	3	1	1	0

[a]Single-nucleotide variants (SNVs) were called in the six samples with the highest mean read coverage. There were 14 651 sites in all 6 samples where a definitive base call could be made and mapped to one of the 14 chromosomal contigs. [b]Mixed-allele sites are excluded from the analysis.

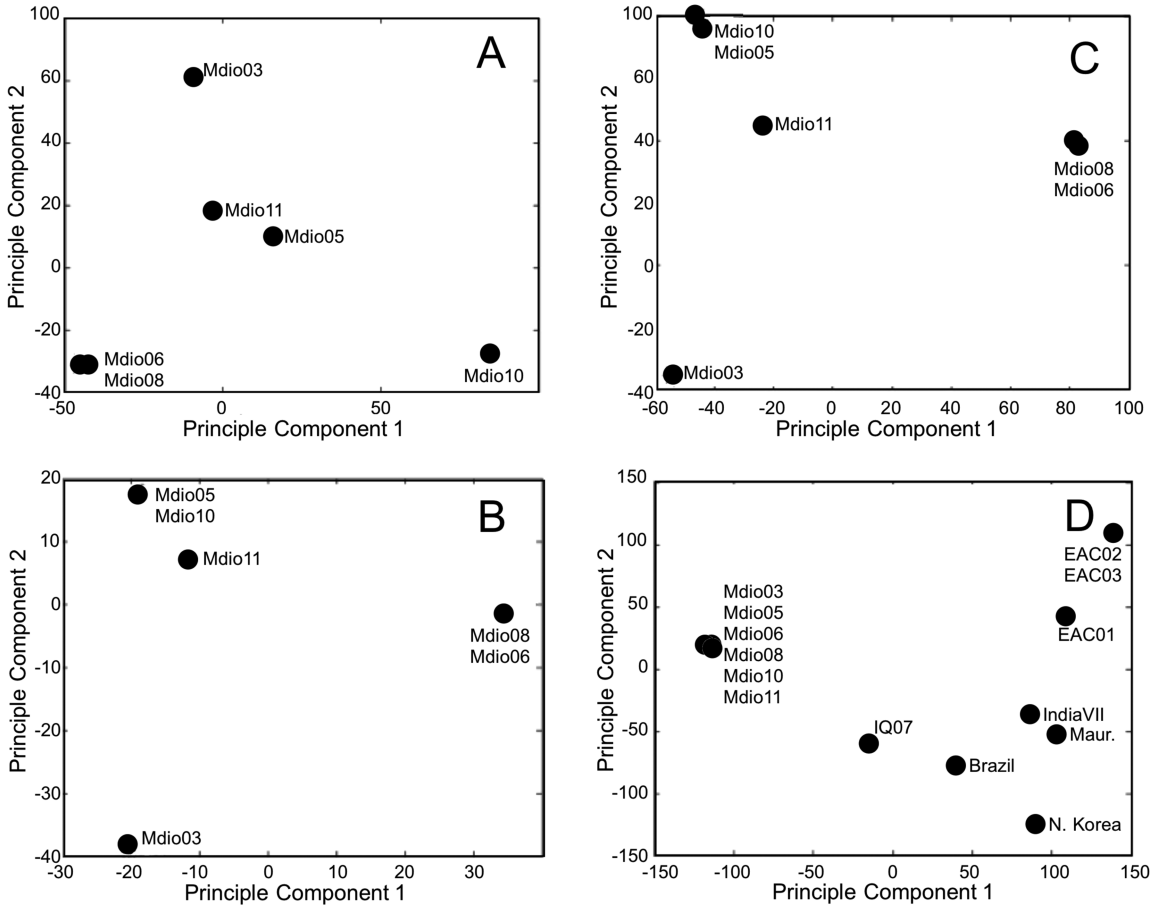


Figure 1. Principal component analysis of *P. vivax* genome sequences. (A) PCA of Madre de Dios genome sequences with all alleles in polyclonal sites included (14 651 sites). (B) PCA of genotypes excluding polyclonal sites (9499 sites). (C) PCA of major alleles with polyclonal sites included (14 651 sites). (D) PCA of Madre de Dios samples combined with worldwide isolates.[14,22,23] EAC01, 02, 03 are from East Africa. Data is from the short read sequence archive, EAC01/03 SRA057904; North Korea I, SRP000316; Mauritania I, SRP000493; Brazil I, SRP007883; India VII, SRP007923; and IQ07, SRP003406.

microsatellite analysis using just four markers (excluding 4.2771 which was polyclonal) showed a unique haplotype for each of the samples tested in our small population. Because the transmission of malaria in Madre de Dios is not expected to be high, we would expect the parasites genotyped here to be more closely related. Therefore, we assessed the relatedness of the parasites comprising each infection. Principle component analysis (PCA) using the 14 651 genotyped sites showed that the Mdio06 and Mdio08 infections were nearly identical (Figure 1A). Furthermore, although Mdio05 and Mdio10 are somewhat unrelated when considering all alleles, they are nearly indistinguishable if polyclonal sites are excluded (Figure 1B) or if only the major allele is considered for the PCA (Figure 1C). The data suggest the clones present at the highest frequency in the Mdio05 and 10 infections are nearly identical but that the minor allele clones are different. When comparing the Madre

DOI: 10.1021/acsinfecdis.5b00049
ACS Infect. Dis. 2015, 1, 367−379

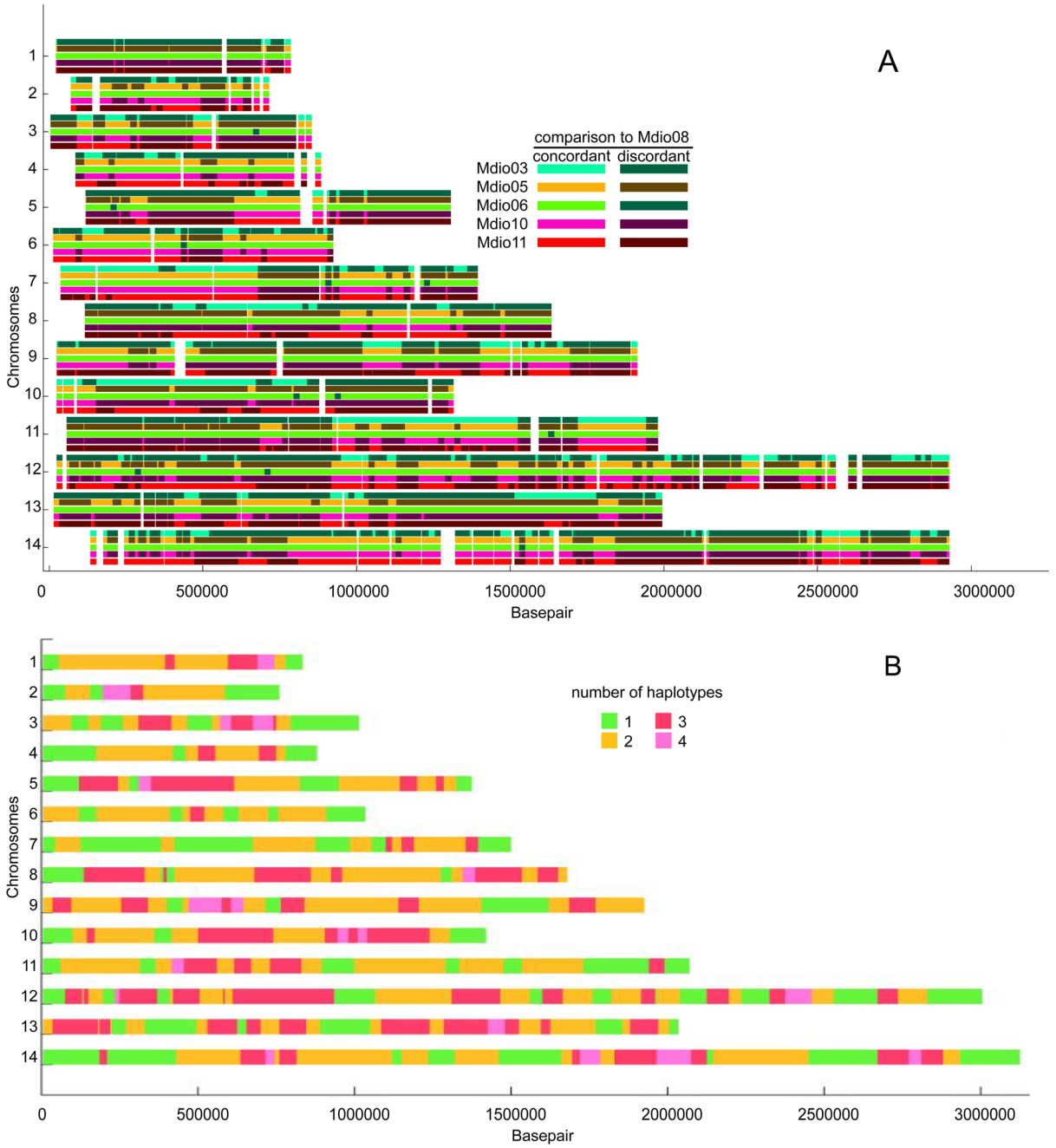


Figure 2. Similarity of the Madre de Dios isolates over the entire genome. (A) Pairwise comparisons showing regions of identity and the difference between isolates Mdio03, 05, 06, 10, and 11 compared to Mdio08 were conducted using the 9499 sites where no second allele was observed. Light colors represent the conservation of sequence; dark shading represents mismatches. (B) Number of haplotypes calculated for each region of the chromosome assessed by calculating a pairwise distance matrix for 50 000 kb overlapping windows.

de Dios population with other global *P. vivax* isolates (from Brazil, India, North Korea, Mauritania,[22] Peru,[23] and East Africa[14]), PCA showed that all six Madre de Dios samples were very closely related, even when considering Brazil I and IQ07, both *P. vivax* isolates from the Amazon basin (Figure 1D).

Given the similarity of the population as a whole, we further sought to determine if the parasites would share chromosomal regions containing identical alleles, interspersed with unrelated regions, suggestive of sexual recombination. We performed pairwise comparisons of allele composition along each chromosome using the set of 9499 SNVs with no minor allele (Figure 2A). These data supported the PCA clustering data showing that Mdio06 and Mdio08 and Mdio05 and Mdio10 were closely related. Additionally, of the 5071 sites that were identical within the set of 6 sequences (and the nonancestral allele), most were spatially clustered along the chromosomes, indicating linkage. For example, Mdio03, 05, 10, and 11 are virtually identical to one another on chromosome 1: Mdio10 and Mdio11 differ at only 10 of 449 genotyped sites in contrast to Mdio08 and Mdio11, which differ at 231 of the 449 sites. In contrast, Mdio08 shares the same haplotype with Mdio06, Mdio05, and Mdio11 in parts of chromosomes 4 and 6.

The data also show evidence of meiotic recombination among the haplotypes as was previously shown.[14] To determine

DOI: 10.1021/acsinfecdis.5b00049
ACS Infect. Dis. 2015, 1, 367−379

Table 4. Coding Changes Relative to the *Sal*I Reference Sequence for Known Drug-Resistance Genes[a]

Chr	Pos.	AA	ref	03	05	06	08	10	11	Br	IQ07	IndVII	Mau	NK
P. vivax Bifunctional Dihydrofolate Reductase-Thymidylate Synthase (*pvdhfr*, PVX_089950)														
5	964763	S58R	C	A	A	G	G	A	A	A	G	G	C	C
5	964796	Y69	T	C	C	C	C	C	C	C	C	T	T	T
5	964856	G89	T	T/G	T/G	T/G	T/G	T	T	G	G	G	G	G
5	964939	S117N	G	A	A	A	A	A	A	A	A	A	G	G
5	965106	I173L	A	A	C	A	A	C/A	A	C	A	A	A	A
5	966035	E482D	A	A/T	A	A	A	A	A	A	A	A	A	A
5	966144	L519	C	C	C	C	C	C	C	A	A	A	A	A
P. vivax Multidrug Resistance Gene 1 (*pvmdr1*, PVX_080100)														
10	363223	T958M	G	A	A	A	A	A	A	A	A	A	A	A
P. vivax Multidrug Resistance Associated Protein 1 (*pvmrp1*, PVX_124085)														
14	2043012	S1689L	G	A/G	G/A	G	G	G	G	G	G	G	G	G
14	2043021	T1686K	G	T	G/T	G/T	G	G	G/T	G	G	G	G	G
14	2043859	Q1407E	G	C	C	C	C	C	G	C	G	C	C	C
14	2044798	A1094S	C	C	A	A	A	A	C	A	C	C	C	C
14	2045050	V1010M	C	T	T	T	T	T	T	T	T	T	T	T
14	2046126	V651G	A	A	A	A/C	A	C/A	A	A	A	A	A	A
14	2046219	R620Q	C	C	T	C	C	T	C	C	C	C	C	C
14	2047233	R282M	C	A	A	A	A	A	A	A	A	A	A	A
14	2047816	E88Q	C	G	G	C	C	G	C	C	G	C	G	C
Chloroquine Resistance Transporter Ortholog (*pvcrt-o*, PVX_087980)														
1	330262	5′ UTR	T	C	C	C	C	C	C	T	T	Y	C	C
1	331151	intron	T	C	C	C	C	C	C	C	C	C	C	C
1	331819	intron	G	G	G	A	A	G	G	G	G	G	G	G
1	332453	intron	T	C	C	C	C	C	C	C	C	C	C	C
1	332874	intron	A	C	C	C	C	C	C	C	C	C	C	C
1	333469	intron	G	G/A	G	G/A	G/A	G	G	G	G	G	G	G
1	333473	intron	G	G	G	G/T	G/T	T	T	G	G	G	G	G
1	333518	intron	A	G	G	A	A	G	G	A	G	G	G	A
1	333544	intron	T	C	C	C	C	C	C	T	C	C	C	T
1	334324	intron	A	A/T	A	T	A/T	A/T	A/T	A	A	T	A	A

[a]Mutations in Madre de Dios isolates identified in the 14 651 confidently called set and mutations identified in Brazil I (Br), India VII (IndVII), Mauritania I (Mau), and North Korea (NK) isolates identified from plasmodb.com were included. No single-nucleotide variants leading to coding changes were observed for *cytochrome bc1* (NC_007243), dihydropteroate synthetase (PVX_123230), gtp cyclohydrolase (PVX_123830), chloroquine resistance transporter ortholog (PVX_087980), or kelch 13 (PVX_083080) orthologue. Y, pyrimidine.

the number of founding parasites that may have contributed to the Madre de Dios breeding populations, we examined blocks of co-inherited markers (Methods). We found that 8.2 million bases (26% of the genome) were in regions that were identical across all isolates (a single haplotype). On the other hand, 43 and 25% of the genome (Figure 2B) were composed of two and three haplotypes, respectively. The remaining 6% of the genome contained four or five haplotypes. Six haplotypes (the maximum number that can be observed) were never observed. The isolated regions with more than three haplotypes typically spanned likely recombination junctions, were in regions containing known polymorphic genes, or were in genomic regions bearing signatures of a possible genome amplification event (see below). For instance, of the 10 allelic differences between Mdio10 and Mdio11 on chromosome 1, 3 are in a polymorphic gene (PVX_093645, bp 717 749 to 728 479) that contains 12 coding and noncoding substitutions. Mdio03 had the most SNVs that were present in only one strain (unique SNVs, Table 3), and it had only seven. These data suggest that two to three parasites contributed to a recombining parasite population in Madre de Dios and that differences observed within these parental genotypes are due to intrinsic mutation rates.

Exploration of Orthologs of Known Drug-Resistance Genes and Identification of Potential Novel Resistance-Associated Genes. We also examined orthologs of known drug-resistance genes in *P. falciparum* to determine if any of these genes in *P. vivax* showed evidence of selection. All *P. vivax* malaria patients are currently treated with primaquine and chloroquine in Peru (CQ 25 mg/kg, PQ 0.5 mg/kg for 7 days),[24] although chloroquine resistance is found in the Amazon region and contributes to treatment failure in some countries.[25−27] The genes involved in chloroquine resistance in *P. vivax* are unknown, but good candidates include the orthologs of the *P. falciparum* genes that are involved in resistance, the chloroquine resistance transporter (*pvcrt-o*, PVX_087980), and the multidrug resistance transporter (*pvmdr1*, PVX_080100) (Table 4). *Pvmdr1* (T958M) bore a single variant relative to the *Sal*I reference sequence; this same allele change has arisen, presumably independently in Thailand[28] and in Madagascar.[29] We found no coding changes in *pvcrt-o*, although no exonic coding changes in *pvcrt-o* have been observed even in parasites resistant to chloroquine.[29] Nevertheless, we did find six intronic changes and one mutation in 5′ UTR, placing it in the top 1% of genes with intronic changes (Table 4). Interestingly, increased expression of *pvcrt-o* in

DOI: 10.1021/acsinfecdis.5b00049
ACS Infect. Dis. 2015, 1, 367−379

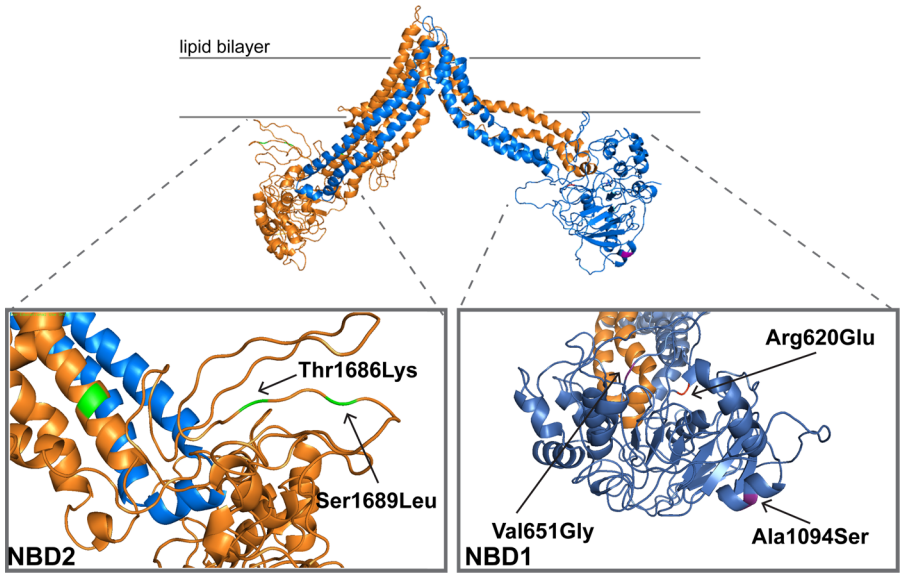


Figure 3. Homology model of PvMRP1 using *C. elegans* PGP-1 as a template. This ABC transporter protein is located in the parasite plasma membrane and has two nucleotide binding domains (NBDs) conserved between the two proteins. The locations of SNVs identified in NBD1 and NBD2 are shown in the insets. SNVs identified but not shown are in transmembrane domains.

chloroquine-resistant parasites from the Amazon has been observed.[30]

Furthermore, we found no amino acid changes in the mitochondrially encoded gene, *plviomp3*, which encodes the cytochrome bc1 complex, the target of atovaquone in *P. falciparum*.[31] We also found no nonsynonymous mutations relative to *Sal*I in *pvkelch13* (PVX_083080), the *P. vivax* ortholog of *pfkelch13*,[10] a gene conferring reduced sensitivity to artemisinin in *P. falciparum* (reviewed in ref 32) presumably because ACTs are not currently used in the Amazon region.

A number of coding changes were observed in the *P. vivax* multidrug resistance associated protein (*pvmrp1*, PVX_124085), which bore nine nonsynonymous SNVs and two synonymous SNVs (Table 4). Position 1466K variants are associated with antifolate resistance in *P. falciparum*,[33] and the disruption of this gene in *P. falciparum* leads to increased primaquine sensitivity.[34] *Pvmrp1* was previously noted because of an unexpectedly high number of coding variants observed in another South American *P. vivax* patient isolate.[23] A homology model using the *Caenorhabditis elegans* multidrug resistance protein, PGP-1,[35] shows several of the mutations observed in the Madre de Dios population occurring in structurally conserved regions of the protein including the nucleic acid binding domain in the intracellular space (Figure 3).

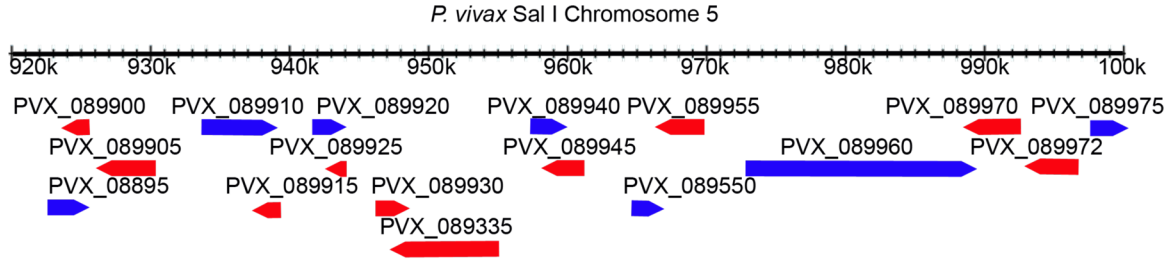
Although antifolates are no longer recommended in Peru we found that coding mutations in the antifolate pathway were common, with all isolates carrying the S117N and S58R dihydrofolate reductase (*dhfr*) mutations, known in combination to confer high levels of resistance to pyrimethamine in *P. vivax* (Table 4).[8,36] Notably, we found two independent S58R nucleotide transversions (C → A, C → G). Triallelic mutations are a feature shared by only 19 other genes, including the apical merozoite antigen (*ama-1*; PVX_092275) (Table S3), suggesting selection pressure on this nucleotide position. A subset (Mdio05 and Mdio10) also carried the I173L allele associated with treatment failure.[36] Given how different these strains are from those in other countries (Figure 1D), it is expected that these mutations arose independently and were not imported. We found there were no novel variants in *dihydropteroate synthetase* (*dhps;* PVX_123230) or *GTP cyclohydrolase* (*gtpch;* PVX_123830), other genes in the dihydrofolate biosynthesis pathway.

It has been shown, on the basis of mapping study experiences with chloroquine- and pyrimethamine-resistant *P. falciparum* field isolates, that patterns of linkage for disequilibrium could be used to detect genes conferring drug resistance.[37,38] These studies have indicated that genes involved in antifolate and chloroquine resistance have arisen infrequently and have likely been transmitted through populations by selective sweeps.[37,38] In contrast to what was observed for *P. falciparum*, the region around *dhfr* on chromosome 5 (*pvdhfr*, bp 964 590 to 966 464) did not show strong evidence of linkage disequilibrium within the Madre de Dios parasite genomes (an 80 kb surrounding region showing two haplotypes is displayed in Figure 4), suggesting that mutant *pvdhfr* has arisen more than once.

Evolution of Other Genes. To further identify genes under selection and potential genes not known to be involved in drug resistance we identified genes with a high nonsynonymous substitution to synonymous substitution ratio (K_a/K_s). Of the top 20 genes with high K_a/K_s, 18 are hypothetical, reflecting the lack of information about the interaction of *P. vivax* with the human immune system. Those genes with functional annotations typically played possible roles in immune evasion, including members of the msp3 multigene family (Table S4), as well as PVX_092625 which encodes an uncharacterized protein, the ortholog of a possible *P. falciparum* vaccine candidate, predicted to be surface expressed.[39] No statistically significant K_a/K_s ratios were observed for orthologs of known *P. falciparum* drug-resistance genes (Table S5). The majority of genes (3773 out of 5507 queried), on the other hand, had no nonsynonymous variants in any of the isolates, reflecting the overall lack of diversity in the population.

Genome Regions around *pvmrp1* and Genes of the Folate Biosynthetic Pathway Show Evidence of Gene Amplification. Changes in the gene copy number have been shown to or are predicted to confer drug resistance in laboratory-evolved *P. falciparum* clones[40] and *P. falciparum* field isolates[41] and are relatively common. We thus looked for

DOI: 10.1021/acsinfecdis.5b00049
ACS Infect. Dis. 2015, 1, 367−379



Position	921057	921121	921829	921886	922670	927165	933556	937847	943806	956274	958396	958505	961498	962548	**964763**	**964796**	**964939**	972559	976603	980374	980484	980542	980552	986288	989702	993180	994083	994491	994694
AA affect						S1038		Intron	F75L		V2461I	G282V			S58R	Y69	S117N		S1224	L2481	F25118Y	G25537	Q25541E	V445 3I			V6446	E5110D	T4443P
Gene						PVX_089905		PVX_089910	PVX_089925		PVX_089940	PVX_089955			PVX_089963	PVX_089969	PVX_089939		PVX_089960	PVX_089974	PVX_089984	PVX_089992	PVX_089996	PVX_089996			PVX_089972	PVX_089972	PVX_089972
Sal I	C	G	C	G	G	A	G	T	A	G	G	G	A	A	C	T	G	C	C	G	T	G	C	G	G	G	T	T	T
Mdio03	T	A	C	A	C	C	T	C	C	A	A	T	G	A	A	C	A	A	T	G	A	T	C	G	A	G	T	T	G
Mdio05	T	A	C	A	C	C	T	C	C	A	A	T	G	A	A	C	A	A	T	G	A	T	C	G	A	G	T	T	G
Mdio06	T	A	A	A	C	C	T	C	C	A	A	T	G	G	G	C	A	A	C	T	T	G	G	A	A	T	C	G	G
Mdio08	T	A	A	A	C	C	T	C	C	A	A	T	G	G	G	C	A	A	C	T	T	G	G	A	A	T	C	G	G
Mdio10	T	A	C	A	C	C	T	C	C	A	A	T	G	A	A	C	A	A	T	G	A	T	C	G	A	G	T	T	G
Mdio11	T	A	C	A	C	C	T	C	C	A	A	T	G	A	A	C	A	A	T	G	A	T	C	G	A	G	T	T	G
Brazil I	T	A	C	A	G	C	T	T	C	A	A	T	G	A	C	C	A	A	T	T	T	G	G	G	A	G	T	T	T
IQ07	T	A	C	A	C	C	T	T	C	A	G	T	G	A	C	C	A	A	T	T	A	T	C	G	A	T	C	T	G
India VII	T	A	C	A	C	C	T	T	C	A	G	T	G	A	C	T	A	A	C	T	T	T	C	G	A	G	C	T	G
Mauritania	C	A	C	G	G	C	G	T	C	A	G	T	G	A	C	T	G	A	C	T	T	G	C	G	A	T	C	T	G
No Korea	T	A	C	A	C	C	G	T	C	A	G	T	G	A	C	T	G	A	C	G	T	G	C	G	A	G	T	T	G
EAC01	C	A	C	A	G	C	G	T	C	A	G	T	G	A	C	T	A	A	C	T	T	T	C	G	A	G	T	T	G
EAC02	C	A	C	A	G	C	G	T	C	A	G	T	G	A	C	T	A	A	C	T	T	T	C	G	A	G	T	T	G
EAC03	C	A	C	A	G	C	G	T	C	A	G	T	G	A	C	T	A	A	C	T	T	T	C	G	A	G	T	T	G

Figure 4. Chromosome 5 alleles surrounding *P. vivax dhfr*. Genotyped sites with no minor alleles in the Madre de Dios set were used to avoid possible phasing errors. Positions in bold are *dhfr*. Red and green alternate shading denote open reading frames. Blue, orange, and purple shading indicate identical haplotypes in this 80 kb region on chromosome 5.

potential increased read coverage across the genome in the Madre de Dios *P. vivax* sequences using an automated CNV-calling algorithm.[42] On average, 2.8% of the genome was potentially amplified and 1.5% showed evidence of a deletion event (Figure 5A). Deleted regions typically encoded members of multigene families. For example, a 100 kb region on chromosome 12 encodes members of the PfamD family, early transcribed membrane proteins, and exported proteins. An ∼200 kb region on chromosome 3 was deleted in Mdio5, Mdio6, and Mdio10 and encoded ∼15 VIR proteins, 7 plasmodium exported proteins, a PfamD family protein member, and an early transcribed member protein as well as "hypothetical" proteins.

In contrast to what is found in *P. falciparum*[43,44] and *P. vivax* in Thailand[45] we found little evidence of increased read coverage near *pvmdr1* on chromosome 10 (361 701−366 095) nor could we find evidence of an amplification event near *pvcrt-o* (chr1:330 260 to 334 540). On the other hand, the region around *pvmrp1* (chr14:2 042 042−2 048 077) in Mdio06, Mdio08, and Mdio05 has increased read coverage suggestive of a copy number variant in that region (Figure 5B). Although amplification of this gene in *P. vivax* or *P. falciparum* patient isolates have not been reported, *P. falciparum* parasites exposed to atovaquone in the laboratory also acquired a CNV that encompassed this gene,[46] providing some evidence that extra copies provide an adaptive advantage.

Our data showed that *pvdhfr* is contained within a region with significantly increased read coverage in Mdio08 (Figure 5C), and manual inspection also showed increased read coverage in Mdio06. In addition, we also found that four isolates had significantly increased read coverage around the *P. vivax* dihydrofolate synthase gene (*pvdhfs*; PVX_116635), which encodes the protein upstream of dihydrofolate reductase in the folate biosynthesis pathway and converts 7,8-dihydropteroate, and L-glutamate to 7,8-dihydrofolate through the hydrolysis of ATP. Recombination events surrounding smaller amplified regions (e.g., 2 kb) were conclusively observed upstream and downstream of *pvdhfr*, suggesting that the genome may have harbored a large 30−40 kb amplification at one time, but the amplified region may have been partially

DOI: 10.1021/acsinfecdis.5b00049
ACS Infect. Dis. 2015, 1, 367−379

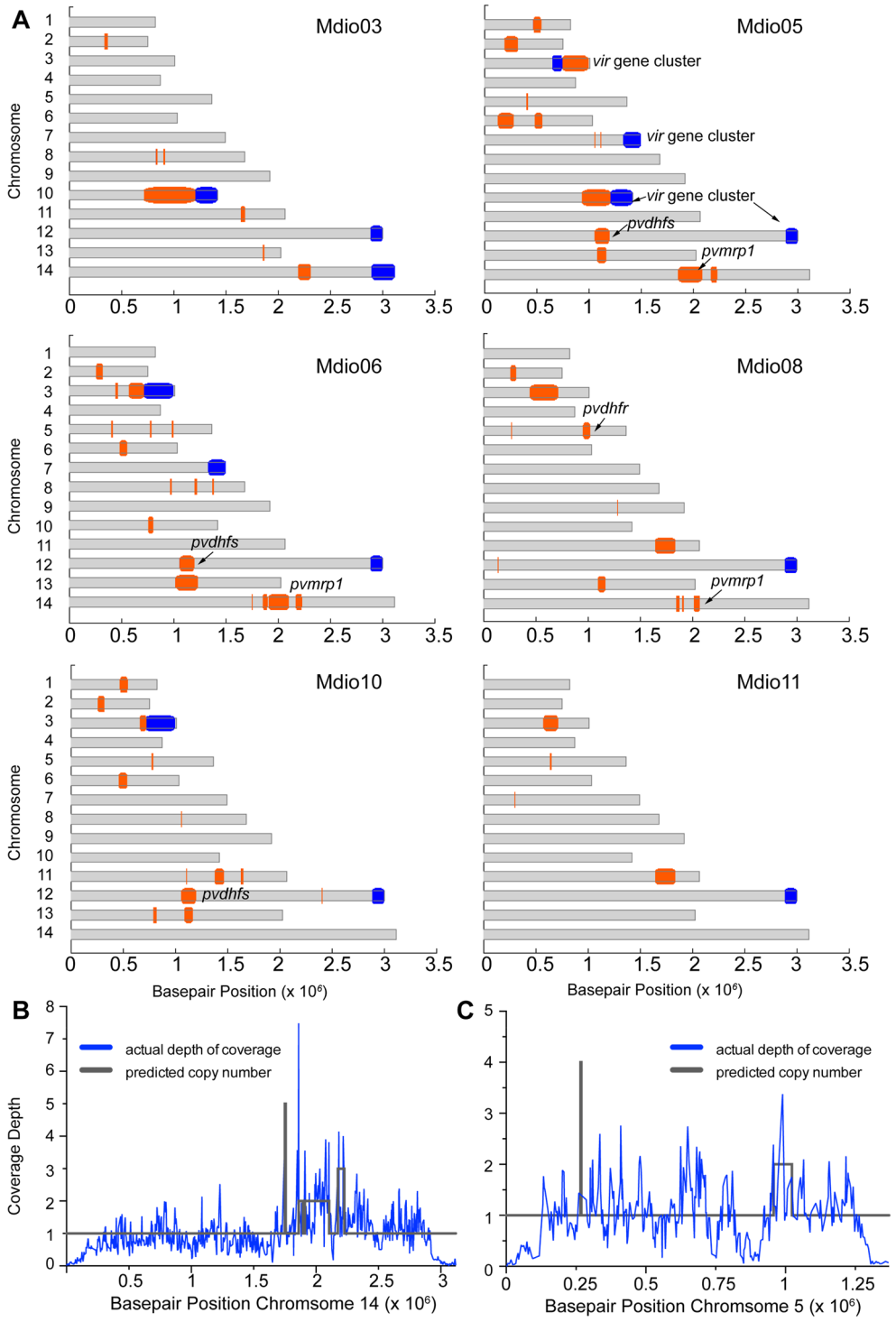


Figure 5. Copy number variants identified in Madre de Dios isolates using a systematic calling algorithm. (A) Distribution of increased (red) and decreased (blue) read coverage in all isolates. Sensitivity is based on overall read coverage for an isolate, and regions may have been missed in isolates with overall lower sequencing depth (e.g., Mdio03). (B) Normalized coverage depth and predicted copy number for Mdio06 chromosome 14 showing the amplification of *pvmrp1*. (C) Normalized coverage depth and predicted copy number for Mdio08 chromosome 5, highlighting amplification in *pvdhfr*.

lost as antifolate drug pressure was removed. Sequencing to a higher depth of coverage or the use of long read sequencing technology such as PacBio could provide more resolution. Further evidence for *pvdhfs* amplification as well as *dhfr* SNVs is suggestive of selective pressure from the past indiscriminate use of antifolates in the 1950s and 1960s.[4,5]

Amplified regions with no clearly known drug-resistance genes were also observed (Table S6) and could represent novel resistance markers. Examples of such amplified regions include chromosome 1:473 919−541 139 bp and a number of regions on chromosome 3 that were observed in all but Mdio03. Although sequencing coverage is too low to establish exact boundaries of these events, one possibly interesting gene that

DOI: 10.1021/acsinfecdis.5b00049
ACS Infect. Dis. 2015, 1, 367−379

was potentially amplified in this region is PVX_000585, which is predicted to encode a homologue of the *E. coli* multidrug transporter, *EmrE*. This membrane protein exchanges small aromatic drugs such as ethidium bromide and methyl viologen (paraquat) across the membrane in *E. coli*[47] and belongs to the same transporter class as *pfcrt*.[48] Orthologs are missing from most eukaryotes but are found in other apicomplexan parasites such as *Toxoplasma gondii* and *Babesia bovis* (OrthoMCL DB v.5.0). Further work will be needed to establish whether these changes are significant.

CONCLUSIONS

Our data highlight potential challenges with using genetic mapping methods to identify genes involved in antimalarial drug resistance. These include inbred populations, which can skew the true diversity of the population if diversity is measured by only a few markers; polyclonal infections, which can affect SNV calling and haplotype inference; and the appearance of de novo mutations which may or may not be associated with resistance in an otherwise homogeneous genetic background. In addition, genome amplifications, which may or may not be correctly mapped to their correct genome location, are likely to be common. All of these features may impact the study design.

On the other hand, our studies demonstrate that insight into the genetic basis of drug resistance can be gained simply by examining the genomes of parasites using high-resolution methods. This method was previously used to identify alleles associated with resistance to the antibiotic clindamycin in *P. falciparum* isolates from Peru[49] and GTP cyclohydrolase gene amplification events in many laboratory and field isolates. The GTP cyclohydrolase event was hypothesized to confer resistance to antifolate drugs because of its presence in the folate biosynthetic pathway[50,51] and was eventually shown to help parasites survive exposure to antifolate drugs by increasing drug flux through the pathway in the presence of mutations in *pfdhfr* that cause a loss of fitness.[52] In our study we identify known and novel mutations potentially involved in drug resistance. Relatively uncharacterized genes that could be further tested for association with resistance to currently used therapies (primaquine and chloroquine) include *pvmrp1* as well as PVX_000585, the *P. vivax* homologue of the *Escherichia coli* multidrug resistance gene, *EmrE*.

Genes involved in resistance to small molecules are now routinely found in *P. falciparum* using laboratory evolution and whole genome sequencing.[53] Identification of these genes allows for monitoring of the emergence of drug resistance in parasite populations. Additionally, knowledge of resistance genes can aide in the design of novel molecules that are effective at targeting the known target in the presence of these known population variants. The identification of resistance genes using laboratory evolution relies on *in vitro* culture systems and would be challenging for *P. vivax*. On the other hand, because parasites in a region such as Madre de Dios could effectively be clonal, such isolated parasite populations might offer opportunities to directly observe the effects of drug exposure on the population. Alleles that appeared in the population only after a new therapy was introduced might be readily identified in an inbred parasite population such as this. Thus, archiving parasite material before and after the introduction of a new therapy could eventually give valuable insight into the evolution of drug resistance in the field.

Our studies also have implications for the design of clinical trials aiming to test novel radical cure drugs that might replace primaquine, the only licensed drug that can eliminate dormant *P. vivax* and *P. ovale* hypnozoites. Although initial clinical trials might be performed with travelers or soldiers returning from endemic regions, groups that would not be expected to be reinfected, larger patient studies in endemic regions could be complex—if a patient returned with an infection that was genotypically very similar to his original infection, then one might assume that the drug candidate was not effective at preventing relapse when in fact the patient had actually been reinfected. Thus, estimates of the number of parasite haplotypes circulating in a region would be needed before trials could be effectively designed. This method would be particularly suited for an area where transmission is low and thus less parasite diversity is expected. In regions where diversity is high and patients are infected with multiple strains, it may prove more difficult to observe mutations related to drug pressure.

METHODS

Sample Collection. This study was approved by the Human Subjects Protection Program of the University of California San Diego, the Institutional Review Board of NAMRU-6, Lima, Peru and the Directorate of Health, Madre de Dios, Peru. Peripheral blood samples for next-generation sequencing were collected from patients with self-reported fever and microscopy-confirmed *P. vivax* malaria. Patients presented at the Peruvian Ministry of Health facility of Bajo Puquiri, locally called Delta-1, in the Madre de Dios department, Peru. This location is approximately 8 h from the capital city of Puerto Maldonado, over difficult terrain. *P. vivax* monoinfection was later confirmed by PCR as previously described.[54] Blood was collected using EDTA-containing Vacutainer tubes. Tubes were centrifuged at 1000*g* to remove plasma, and the packed red blood cells were resuspended in a 2× volume of sterile phosphate-buffered saline (1.8 mM KH_2PO_4, 10 mM Na_2HPO_4, 137 mM NaCl, 2.7 nM KCl). Within 1 h of collection diluted blood was filtered through CF11 cellulose powder (Whatman) to remove white blood cells.[55] Eluted blood was pelleted by centrifugation at 1000*g* for 10 min and stored at −20 °C for a maximum period of 2 to 3 weeks before being shipped in batches to the NAMRU-6 laboratory in a liquid-nitrogen dry shipper.

Microsatellite Analysis. *P. vivax* microsatellite characterization was conducted using the following previously published loci: 14.185, 12.335, 7.67, 6.34, 4.2771, and 3.35.[56,57] Microsatellites were amplified with primers previously described,[56] and fluorescently labeled PCR products were analyzed on an Applied Biosystems Prism 3130xl Avant Genetic Analyzer (Applied Biosystems). A minimum peak height of 200 fluorescent units was used to define amplification products.

Whole Genome Sequencing. Genomic DNA was isolated from frozen whole blood samples using a DNeasy blood and tissue kit (Qiagen), and whole genome paired end-sequencing libraries were created using the NEBNext DNA library prep kit (New England Biolabs). DNA was sheared using Covaris adaptive focused acoustics and size selected on an agarose gel. DNA libraries were clustered and run on an Illumina HiSeq2000 (Illumina, Inc., San Diego) according to the manufacturer's instructions. Base calls were made using Consensus Assessment of Sequence and Variation v. 1.8.2 (Illumina). Sequencing reads in fastq format are available on the National Center for Biotechnology Information's short read archive (PRJNA272364). Reads were aligned to the *P. vivax*

DOI: 10.1021/acsinfecdis.5b00049
ACS Infect. Dis. 2015, 1, 367−379

*Sal*I reference (v. 10.0) using the Burrows-Wheeler aligner with the default options.[58,59] Picard was used to remove PCR duplicates and evaluate the sequencing quality.

Single nucleotide variants (SNVs) were called simultaneously in the Madre de Dios set of 10 using the Genome Analysis Toolkit Unified Genotyper tool. Variant sites within SNV clusters (cluster window = 10, SNP cluster = 3) and with low mapping quality (MQ < 50) were filtered out. Additionally, variant sites with a combined depth of coverage of less than 20 reads (DP < 20) were excluded. A total of 32 199 high-quality SNVs were identified and genotyped in each of the 10 samples. Because we were interested in high-quality SNVs that would allow us to determine relatedness of the strains, we further filtered the variants. Overall read coverage was low in Mdio01, Mdio04, Mdio07, and Mdio09; therefore, these samples were excluded from further analysis. For the samples included, Mdio03, Mdio05, Mdio06, Mdio08, Mdio10, and Mdio11, sites were filtered if they did not have calleable alleles in all of these samples or a read coverage of <5. Additional custom filtering was designed to retain sites where we were confident of the zygosity. Homozygous SNVs were called if >80% of the reads were the major allele, and heterozygous SNVs were called if 50−80% of the reads represented the major allele. SNVs with more than one alternate allele were also excluded. This resulted in 15 380 confidently calleable sites for determining clonality located on both contigs and chromosomes. SNVs indentified in contigs were further removed to yield a high-quality set of 14 651 genotyped sites (Table 4).

Diversity and Relatedness Analyses. To characterize within-host diversity we used the F_{ws} statistic[18,55] ($F_{ws} = 1 - (H_w/H_s)$, where H_w is the within-host heterozygosity and H_s is the population-level heterozygosity), using the 14 651 genotyped sites. The principle component analysis was performed using the Matlab (R2012b, The Mathworks) princomp command. Calculation of the ratio of the non-synonymous substation rate to the synonymous substitution rate (K_a/K_s) was done using the modified Yang−Neilsen algorithm.[60] Haplotype blocks were calculated using a set of 4430 SNVs with no minor allele and which were variant in one Madre de Dios isolate relative to the others. The number of haplotypes present was calculated for 50 kb windows (moving over each chromosome in 5 kb steps) by calculating the pairwise distance matrix for all sample pairs within the window. The pairwise distance estimate, d, was calculated as $d = n_d/n_c$, where n_d is the number of nonidentical sites and n_c is the number of sites compared. Values for the distance matrix that were less than 5×10^{-5} were considered to be similar haplotype blocks between samples.

Pairwise Comparisons. Pairwise comparisons were completed using either all 15 380 genotyped sites or sites where no minor allele was shared (9571 sites left). Chromosome maps were made to represent whether alternate alleles were shared between samples.

Homology Modeling. Amino acid sequences of PvMRP1 were retrieved from PlasmoDB (search ID: PVX_124085) and were cross referenced with Uniprot. A template search was conducted against the SWISS-MODEL template library, with the highest scoring template being chosen for model building. A web-based homology model was produced using the SWISS-MODEL pipeline which operates via Promod-II and Modeler platforms. The PyMOL Molecular Graphics System (version 1.3, 2010 Schrodinger, LLC.) was then used to display and label the protein structure.

CNV Analysis. Plasmodium Type Uncovering Software (Platypus)[61] was used to identify regions of the genome where a significant increase or decrease in read coverage was observed. This software sensitively detects the largest changes in read signal after GC bias normalization. Changes were also observed for the various unassembled contigs (data not shown).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsinfecdis.5b00049.

Demographic characteristics of the sample population. Tri-allelic SNVs observed in the Madre de Dios population. Top 20 genes with the largest K_a/K_s ratios. K_a/K_s ratios for *P. vivax* orthologs of known *P. falciparum* drug-resistance genes. (PDF)

Genomic regions with significantly increased or decreased read coverage identified using paired-end NGS data. (XLSX)

Single nucleotide variants identified in each of the Madre de Dios whole genome sequencing samples. (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: ewinzeler@ucsd.edu.

Author Contributions

E.A.W. and E.L.F. wrote the manuscript. E.A.W, E.L.F., A.T.B., F.G., V.C., M.A., A.A., V.B., and T.W. analyzed the data. A.T.B. prepared samples for genomic sequencing. J.F.S., M.L.S., G.C.B., K.A.E., and L.A.R. performed sample collection and microsatellite analysis. A.G.L. collected the samples and analyzed the clinical data. J.M.V. analyzed the clinical data and genetic phenotypes. All authors edited the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

E.A.W. is supported by grants R01AI103058 and R01AI090141 from NIAID, grants from the Bill and Melinda Gates Foundation (OPP1040406), and the Medicines for Malaria Venture. E.L.F. is supported by an F32 NRSA fellowship (5F32AI102567). T.W. receives support from the UCSD Genetics Training Program through an institutional training grant from the National Institute of General Medical Sciences (T32 GM008666). The work in Peru was supported by U19AI089681 from NIAID (JMV). A.A. and V.B. were supported in part by a grant from the NSF (IIS-1318386). A.G.L. is sponsored by training grant 2D43 TW007393 awarded to NAMRU-6 by the Fogarty International Center of the U.S. National Institutes of Health. The views expressed in this article are those of the authors only and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, or the U.S. Government. Several authors of this manuscript are employees of the U.S. Government. This work was prepared as part of their duties. Title 17 U.S.C. § 105 provides that 'Copyright protection under this title is not available for any work of the United States Government.' Title 17 U.S.C. § 101 defines a U.S. Government work as a work prepared by a military service member or employee of the U.S. Government as part of that person's official duties.

DOI: 10.1021/acsinfecdis.5b00049
ACS Infect. Dis. 2015, 1, 367−379

REFERENCES

(1) WHO *World Malaria Report 2014*; World Health Organization: Geneva, 2014; pp 1−276.

(2) WHO *Guidelines for the Treatment of Malaria*; World Health Organization: Geneva, 2010; pp 1−194.

(3) Ratcliff, A., Siswantoro, H., Kenangalem, E., Wuwung, M., Brockman, A., Edstein, M. D., Laihad, F., Ebsworth, E. P., Anstey, N. M., Tjitra, E., and Price, R. N. (2007) Therapeutic response of multidrug-resistant Plasmodium falciparum and P. vivax to chloroquine and sulfadoxine-pyrimethamine in southern Papua, Indonesia. *Trans. R. Soc. Trop. Med. Hyg. 101* (4), 351−9.

(4) Newton, P., and White, N. (1999) Malaria: new developments in treatment and prevention. *Annu. Rev. Med. 50*, 179−192.

(5) Val, F. F., Sampaio, V. S., Cassera, M. B., Andrade, R. T., Tauil, P. L., Monteiro, W. M., and Lacerda, M. V. (2014) Plasmodium vivax malaria elimination: should innovative ideas from the past be revisited? *Memorias do Instituto Oswaldo Cruz 109* (5), 522−4.

(6) Ganguly, S., Saha, P., Chatterjee, M., and Maji, A. K. (2014) Prevalence of polymorphisms in antifolate drug resistance molecular marker genes *pvdhfr* and *pvdhps* in clinical isolates of *Plasmodium vivax* from Kolkata, India. *Antimicrob. Agents Chemother. 58* (1), 196−200.

(7) Auliff, A. M., Balu, B., Chen, N., O'Neil, M. T., Cheng, Q., and Adams, J. H. (2012) Functional analysis of *Plasmodium vivax* dihydrofolate reductase-thymidylate synthase genes through stable transformation of *Plasmodium falciparum*. *PLoS One 7* (7), e40416.

(8) Hastings, M. D., Maguire, J. D., Bangs, M. J., Zimmerman, P. A., Reeder, J. C., Baird, J. K., and Sibley, C. H. (2005) Novel Plasmodium vivax dhfr alleles from the Indonesian Archipelago and Papua New Guinea: association with pyrimethamine resistance determined by a Saccharomyces cerevisiae expression system. *Antimicrob. Agents Chemother. 49* (2), 733−40.

(9) Ding, S., Ye, R., Zhang, D., Sun, X., Zhou, H., McCutchan, T. F., and Pan, W. (2013) Anti-folate combination therapies and their effect on the development of drug resistance in Plasmodium vivax. *Sci. Rep. 3*, 1008.

(10) Ariey, F., Witkowski, B., Amaratunga, C., Beghain, J., Langlois, A. C., Khim, N., Kim, S., Duru, V., Bouchier, C., Ma, L., Lim, P., Leang, R., Duong, S., Sreng, S., Suon, S., Chuor, C. M., Bout, D. M., Menard, S., Rogers, W. O., Genton, B., Fandeur, T., Miotto, O., Ringwald, P., Le Bras, J., Berry, A., Barale, J. C., Fairhurst, R. M., Benoit-Vical, F., Mercereau-Puijalon, O., and Menard, D. (2014) A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria. *Nature 505* (7481), 50−55.

(11) Dondorp, A. M., Nosten, F., Yi, P., Das, D., Phyo, A. P., Tarning, J., Lwin, K. M., Ariey, F., Hanpithakpong, W., Lee, S. J., Ringwald, P., Silamut, K., Imwong, M., Chotivanich, K., Lim, P., Herdman, T., An, S. S., Yeung, S., Singhasivanon, P., Day, N. P., Lindegardh, N., Socheat, D., and White, N. J. (2009) Artemisinin resistance in Plasmodium falciparum malaria. *N. Engl. J. Med. 361* (5), 455−67.

(12) Miotto, O., Almagro-Garcia, J., Manske, M., Macinnis, B., Campino, S., Rockett, K. A., Amaratunga, C., Lim, P., Suon, S., Sreng, S., Anderson, J. M., Duong, S., Nguon, C., Chuor, C. M., Saunders, D., Se, Y., Lon, C., Fukuda, M. M., Amenga-Etego, L., Hodgson, A. V., Asoala, V., Imwong, M., Takala-Harrison, S., Nosten, F., Su, X. Z., Ringwald, P., Ariey, F., Dolecek, C., Hien, T. T., Boni, M. F., Thai, C. Q., Amambua-Ngwa, A., Conway, D. J., Djimde, A. A., Doumbo, O. K., Zongo, I., Ouedraogo, J. B., Alcock, D., Drury, E., Auburn, S., Koch, O., Sanders, M., Hubbart, C., Maslen, G., Ruano-Rubio, V., Jyothi, D., Miles, A., O'Brien, J., Gamble, C., Oyola, S. O., Rayner, J. C., Newbold, C. I., Berriman, M., Spencer, C. C., McVean, G., Day, N. P., White, N. J., Bethell, D., Dondorp, A. M., Plowe, C. V., Fairhurst, R. M., and Kwiatkowski, D. P. (2013) Multiple populations of artemisinin-resistant *Plasmodium falciparum* in Cambodia. *Nat. Genet. 45*, 648.

(13) Takala-Harrison, S., Clark, T. G., Jacob, C. G., Cummings, M. P., Miotto, O., Dondorp, A. M., Fukuda, M. M., Nosten, F., Noedl, H., Imwong, M., Bethell, D., Se, Y., Lon, C., Tyner, S. D., Saunders, D. L., Socheat, D., Ariey, F., Phyo, A. P., Starzengruber, P., Fuehrer, H. P., Swoboda, P., Stepniewska, K., Flegg, J., Arze, C., Cerqueira, G. C., Silva, J. C., Ricklefs, S. M., Porcella, S. F., Stephens, R. M., Adams, M., Kenefic, L. J., Campino, S., Auburn, S., MacInnis, B., Kwiatkowski, D. P., Su, X. Z., White, N. J., Ringwald, P., and Plowe, C. V. (2013) Genetic loci associated with delayed clearance of *Plasmodium falciparum* following artemisinin treatment in Southeast Asia. *Proc. Natl. Acad. Sci. U. S. A. 110* (1), 240−5.

(14) Bright, A. T., Manary, M. J., Tewhey, R., Arango, E. M., Wang, T., Schork, N. J., Yanow, S. K., and Winzeler, E. A. (2014) A high resolution case study of a patient with recurrent *Plasmodium vivax* infections shows that relapses were caused by meiotic siblings. *PLoS Neglected Trop. Dis. 8* (6), e2882.

(15) Roberts, D. R., Laughlin, L. L., Hsheih, P., and Legters, L. J. (1997) DDT, global strategies, and a malaria control crisis in South America. *Emerging Infect. Dis. 3* (3), 295−302.

(16) Aramburu Guarda, J., Ramal Asayag, C., and Witzig, R. (1999) Malaria reemergence in the Peruvian Amazon region. *Emerging Infect. Dis. 5* (2), 209−15.

(17) Auburn, S., Campino, S., Miotto, O., Djimde, A. A., Zongo, I., Manske, M., Maslen, G., Mangano, V., Alcock, D., MacInnis, B., Rockett, K. A., Clark, T. G., Doumbo, O. K., Ouedraogo, J. B., and Kwiatkowski, D. P. (2012) Characterization of within-host *Plasmodium falciparum* diversity using next-generation sequence data. *PLoS One 7* (2), e32891.

(18) Manske, M., Miotto, O., Campino, S., Auburn, S., Almagro-Garcia, J., Maslen, G., O'Brien, J., Djimde, A., Doumbo, O., Zongo, I., Ouedraogo, J.-B., Michon, P., Mueller, I., Siba, P., Nzila, A., Borrmann, S., Kiara, S. M., Marsh, K., Jiang, H., Su, X.-Z., Amaratunga, C., Fairhurst, R., Socheat, D., Nosten, F., Imwong, M., White, N. J., Sanders, M., Anastasi, E., Alcock, D., Drury, E., Oyola, S., Quail, M. A., Turner, D. J., Ruano-Rubio, V., Jyothi, D., Amenga-Etego, L., Hubbart, C., Jeffreys, A., Rowlands, K., Sutherland, C., Roper, C., Mangano, V., Modiano, D., Tan, J. C., Ferdig, M. T., Amambua-Ngwa, A., Conway, D. J., Takala-Harrison, S., Plowe, C. V., Rayner, J. C., Rockett, K. A., Clark, T. G., Newbold, C. I., Berriman, M., MacInnis, B., and Kwiatkowski, D. P. (2012) Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. *Nature 487*, 375.

(19) Kosek, M., Yori, P. P., Gilman, R. H., Calderon, M., Zimic, M., Chuquiyauri, R., Jeri, C., Pinedo-Cancino, V., Matthias, M. A., Llanos-Cuentas, A., and Vinetz, J. M. (2012) High degree of *Plasmodium vivax* diversity in the Peruvian Amazon demonstrated by tandem repeat polymorphism analysis. *Am. J. Trop. Med. Hyg. 86* (4), 580−6.

(20) Van den Eede, P., Erhart, A., Van der Auwera, G., Van Overmeir, C., Thang, N. D., Hung le, X., Anne, J., and D'Alessandro, U. (2010) High complexity of Plasmodium vivax infections in symptomatic patients from a rural community in central Vietnam detected by microsatellite genotyping. *Am. J. Trop. Med. Hyg. 82* (2), 223−7.

(21) Van den Eede, P., Van der Auwera, G., Delgado, C., Huyse, T., Soto-Calle, V. E., Gamboa, D., Grande, T., Rodriguez, H., Llanos, A., Anne, J., Erhart, A., and D'Alessandro, U. (2010) Multilocus genotyping reveals high heterogeneity and strong local population structure of the *Plasmodium vivax* population in the Peruvian Amazon. *Malar. J. 9*, 151.

(22) Neafsey, D. E., Galinsky, K., Jiang, R. H. Y., Young, L., Sykes, S. M., Saif, S., Gujja, S., Goldberg, J. M., Young, S., Zeng, Q., Chapman, S. B., Dash, A. P., Anvikar, A. R., Sutton, P. L., Birren, B. W., Escalante, A. A., Barnwell, J. W., and Carlton, J. M. (2012) The malaria parasite Plasmodium vivax exhibits greater genetic diversity than Plasmodium falciparum. *Nat. Genet. 44* (9), 1046−1050.

(23) Dharia, N. V., Bright, A. T., Westenberger, S. J., Barnes, S. W., Batalov, S., Kuhen, K., Borboa, R., Federe, G. C., McClean, C. M., Vinetz, J. M., Neyra, V., Llanos-Cuentas, A., Barnwell, J. W., Walker, J. R., and Winzeler, E. A. (2010) Whole-genome sequencing and microarray analysis of *ex vivo Plasmodium vivax* reveal selective pressure on putative drug resistance genes. *Proc. Natl. Acad. Sci. U. S. A. 107* (46), 20045−20050.

(24) Durand, S., Cabezas, C., Lescano, A. G., Galvez, M., Gutierrez, S., Arrospide, N., Alvarez, C., Santolalla, M. L., Bacon, D. J., and Graf, P. C. (2014) Efficacy of three different regimens of primaquine for the

DOI: 10.1021/acsinfecdis.5b00049
ACS Infect. Dis. 2015, 1, 367−379

prevention of relapses of Plasmodium vivax malaria in the Amazon Basin of Peru. *Am. J. Trop. Med. Hyg. 91* (1), 18−26.
(25) Marques, M. M., Costa, M. R., Santana Filho, F. S., Vieira, J. L., Nascimento, M. T., Brasil, L. W., Nogueira, F., Silveira, H., Reyes-Lecca, R. C., Monteiro, W. M., Lacerda, M. V., and Alecrim, M. G. (2014) *Plasmodium vivax* chloroquine resistance and anemia in the western Brazilian Amazon. *Antimicrob. Agents Chemother. 58* (1), 342−7.
(26) de Santana Filho, F. S., Arcanjo, A. R., Chehuan, Y. M., Costa, M. R., Martinez-Espinosa, F. E., Vieira, J. L., Barbosa, M. G., Alecrim, W. D., and Alecrim, M. G. (2007) Chloroquine-resistant *Plasmodium vivax*, Brazilian Amazon. *Emerging Infect. Dis. 13* (7), 1125−6.
(27) Chehuan, Y. F., Costa, M. R., Costa, J. S., Alecrim, M. G., Nogueira, F., Silveira, H., Brasil, L. W., Melo, G. C., Monteiro, W. M., and Lacerda, M. V. (2013) In vitro chloroquine resistance for *Plasmodium vivax* isolates from the Western Brazilian Amazon. *Malar. J. 12*, 226.
(28) Rijken, M. J., Boel, M. E., Russell, B., Imwong, M., Leimanis, M. L., Phyo, A. P., Muehlenbachs, A., Lindegardh, N., McGready, R., and Rénia, L. (2011) Chloroquine resistant vivax malaria in a pregnant woman on the western border of Thailand. *Malar. J. 10*, 113.
(29) Barnadas, C., Ratsimbasoa, A., Tichit, M., Bouchier, C., Jahevitra, M., Picot, S., and Ménard, D. (2008) Plasmodium vivax resistance to chloroquine in Madagascar: clinical efficacy and polymorphisms in pvmdr1 and pvcrt-o genes. *Antimicrob. Agents Chemother. 52* (12), 4233−4240.
(30) Melo, G. C., Monteiro, W. M., Siqueira, A. M., Silva, S. R., Magalhaes, B. M., Alencar, A. C., Kuehn, A., del Portillo, H. A., Fernandez-Becerra, C., and Lacerda, M. V. (2014) Expression levels of pvcrt-o and pvmdr-1 are associated with chloroquine resistance and severe Plasmodium vivax malaria in patients of the Brazilian Amazon. *PLoS One 9* (8), e105922.
(31) Fry, M., and Pudney, M. (1992) Site of action of the antimalarial hydroxynaphthoquinone, 2-[trans-4-(4′-chlorophenyl) cyclohexyl]-3-hydroxy-1,4-naphthoquinone (566C80). *Biochem. Pharmacol. 43* (7), 1545−53.
(32) Winzeler, E. A., and Manary, M. J. (2014) Drug resistance genomics of the antimalarial drug artemisinin. *Genome Biology 15* (11), 544.
(33) Dahlstrom, S., Veiga, M. I., Martensson, A., Bjorkman, A., and Gil, J. P. (2009) Polymorphism in PfMRP1 (*Plasmodium falciparum* multidrug resistance protein 1) amino acid 1466 associated with resistance to sulfadoxine-pyrimethamine treatment. *Antimicrob. Agents Chemother. 53* (6), 2553−6.
(34) Raj, D. K., Mu, J., Jiang, H., Kabat, J., Singh, S., Sullivan, M., Fay, M. P., McCutchan, T. F., and Su, X.-z. (2009) Disruption of a Plasmodium falciparum multidrug resistance-associated protein (PfMRP) alters its fitness and transport of antimalarial drugs and glutathione. *J. Biol. Chem. 284* (12), 7687−7696.
(35) Jin, M. S., Oldham, M. L., Zhang, Q., and Chen, J. (2012) Crystal structure of the multidrug transporter P-glycoprotein from Caenorhabditis elegans. *Nature 490* (7421), 566−569.
(36) Leartsakulpanich, U., Imwong, M., Pukrittayakamee, S., White, N. J., Snounou, G., Sirawaraporn, W., and Yuthavong, Y. (2002) Molecular characterization of dihydrofolate reductase in relation to antifolate resistance in Plasmodium vivax. *Mol. Biochem. Parasitol. 119* (1), 63−73.
(37) Nair, S., Williams, J. T., Brockman, A., Paiphun, L., Mayxay, M., Newton, P. N., Guthmann, J. P., Smithuis, F. M., Hien, T. T., White, N. J., Nosten, F., and Anderson, T. J. (2003) A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. *Molecular biology and evolution 20* (9), 1526−36.
(38) Wootton, J. C., Feng, X., Ferdig, M. T., Cooper, R. A., Mu, J., Baruch, D. I., Magill, A. J., and Su, X. Z. (2002) Genetic diversity and chloroquine selective sweeps in Plasmodium falciparum. *Nature 418* (6895), 320−3.
(39) Villard, V., Agak, G. W., Frank, G., Jafarshad, A., Servis, C., Nebie, I., Sirima, S. B., Felger, I., Arevalo-Herrera, M., Herrera, S., Heitz, F., Backer, V., Druilhe, P., Kajava, A. V., and Corradin, G. (2007) Rapid identification of malaria vaccine candidates based on alpha-helical coiled coil protein motif. *PLoS One 2* (7), e645.
(40) Hoepfner, D., McNamara, C. W., Lim, C. S., Studer, C., Riedl, R., Aust, T., McCormack, S. L., Plouffe, D. M., Meister, S., Schuierer, S., Plikat, U., Hartmann, N., Staedtler, F., Cotesta, S., Schmitt, E. K., Petersen, F., Supek, F., Glynne, R. J., Tallarico, J. A., Porter, J. A., Fishman, M. C., Bodenreider, C., Diagana, T. T., Movva, N. R., and Winzeler, E. A. (2012) Selective and specific inhibition of the *Plasmodium falciparum* lysyl-tRNA synthetase by the fungal secondary metabolite cladosporin. *Cell Host Microbe 11* (6), 654−63.
(41) Nair, S., Nash, D., Sudimack, D., Jaidee, A., Barends, M., Uhlemann, A. C., Krishna, S., Nosten, F., and Anderson, T. J. (2006) Recurrent gene amplification and soft selective sweeps during evolution of multidrug resistance in malaria parasites. *Mol. Biol. Evol. 24* (2), 562−573.
(42) Manary, M. J., Singhakul, S. S., Flannery, E. L., Bopp, S. E., Corey, V. C., Bright, A. T., McNamara, C. W., Walker, J. R., and Winzeler, E. A. (2014) Identification of pathogen genomic variants through an integrated pipeline. *BMC Bioinf. 15* (1), 63.
(43) Wilson, C. M., Volkman, S. K., Thaithong, S., Martin, R. K., Kyle, D. E., Milhous, W. K., and Wirth, D. F. (1993) Amplification of pfmdr 1 associated with mefloquine and halofantrine resistance in Plasmodium falciparum from Thailand. *Mol. Biochem. Parasitol. 57* (1), 151−60.
(44) Price, R. N., Uhlemann, A. C., Brockman, A., McGready, R., Ashley, E., Phaipun, L., Patel, R., Laing, K., Looareesuwan, S., White, N. J., Nosten, F., and Krishna, S. (2004) Mefloquine resistance in Plasmodium falciparum and increased pfmdr1 gene copy number. *Lancet 364* (9432), 438−47.
(45) Lin, J. T., Patel, J. C., Kharabora, O., Sattabongkot, J., Muth, S., Ubalee, R., Schuster, A. L., Rogers, W. O., Wongsrichanalai, C., and Juliano, J. J. (2013) Plasmodium vivax isolates from Cambodia and Thailand show high genetic complexity and distinct patterns of P. vivax multidrug resistance gene 1 (pvmdr1) polymorphisms. *Am. J. Trop. Med. Hyg. 88* (6), 1116−23.
(46) Bopp, S. E. R., Manary, M. J., Bright, A. T., Johnston, G. L., Dharia, N. V., Luna, F. L., McCormack, S., Plouffe, D., McNamara, C. W., Walker, J. R., Fidock, D. A., Denchi, E. L., and Winzeler, E. A. (2013) Mitotic Evolution of *Plasmodium falciparum* Shows a Stable Core Genome but Recombination in Antigen Families. *PLoS Genet. 9* (2), e1003293.
(47) Yerushalmi, H., Lebendiker, M., and Schuldiner, S. (1995) EmrE, an Escherichia coli 12-kDa multidrug transporter, exchanges toxic cations and H+ and is soluble in organic solvents. *J. Biol. Chem. 270* (12), 6856−63.
(48) Chinappi, M., Via, A., Marcatili, P., and Tramontano, A. (2010) On the mechanism of chloroquine resistance in Plasmodium falciparum. *PLoS One 5* (11), e14064.
(49) Dharia, N. V., Plouffe, D., Bopp, S. E., Gonzalez-Paez, G. E., Lucas, C., Salas, C., Soberon, V., Bursulaya, B., Kochel, T. J., Bacon, D. J., and Winzeler, E. A. (2010) Genome-scanning of Amazonian Plasmodium falciparum shows subtelomeric instability and clindamycin-resistant parasites. *Genome Res. 20* (11), 1534−1544.
(50) Kidgell, C., Volkman, S. K., Daily, J., Borevitz, J. O., Plouffe, D., Zhou, Y., Johnson, J. R., Le Roch, K., Sarr, O., Ndir, O., Mboup, S., Batalov, S., Wirth, D. F., and Winzeler, E. A. (2006) A systematic map of genetic variation in Plasmodium falciparum. *PLoS Pathog. 2* (6), e57 Epub 2006 Jun 23. .
(51) Nair, S., Miller, B., Barends, M., Jaidee, A., Patel, J., Mayxay, M., Newton, P., Nosten, F., Ferdig, M. T., and Anderson, T. J. (2008) Adaptive copy number evolution in malaria parasites. *PLoS Genet. 4* (10), e1000243.
(52) Heinberg, A., Siu, E., Stern, C., Lawrence, E. A., Ferdig, M. T., Deitsch, K. W., and Kirkman, L. A. (2013) Direct evidence for the adaptive role of copy number variation on antifolate susceptibility in Plasmodium falciparum. *Mol. Microbiol. 88* (4), 702−712.
(53) Flannery, E. L., Fidock, D. A., and Winzeler, E. A. (2013) Using Genetic Methods To Define the Targets of Compounds with Antimalarial Activity. *J. Med. Chem. 56*, 7761.

DOI: 10.1021/acsinfecdis.5b00049
ACS Infect. Dis. 2015, 1, 367−379

(54) Snounou, G., Viriyakosol, S., Zhu, X. P., Jarra, W., Pinheiro, L., do Rosario, V. E., Thaithong, S., and Brown, K. N. (1993) High sensitivity of detection of human malaria parasites by the use of nested polymerase chain reaction. *Mol. Biochem. Parasitol. 61* (2), 315−20.
(55) Venkatesan, M., Amaratunga, C., Campino, S., Auburn, S., Koch, O., Lim, P., Uk, S., Socheat, D., Kwiatkowski, D. P., Fairhurst, R. M., and Plowe, C. V. (2012) Using CF11 cellulose columns to inexpensively and effectively remove human DNA from Plasmodium falciparum-infected whole blood samples. *Malar. J. 11*, 41.
(56) Imwong, M., Sudimack, D., Pukrittayakamee, S., Osorio, L., Carlton, J. M., Day, N. P., White, N. J., and Anderson, T. J. (2006) Microsatellite variation, repeat array length, and population history of *Plasmodium vivax*. *Mol. Biol. Evol. 23* (5), 1016−8.
(57) McCollum, A. M., Soberon, V., Salas, C. J., Santolalla, M. L., Udhayakumar, V., Escalante, A. A., Graf, P. C., Durand, S., Cabezas, C., and Bacon, D. J. (2014) Genetic variation and recurrent parasitaemia in Peruvian *Plasmodium vivax* populations. *Malar. J. 13*, 67.
(58) Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics 25* (14), 1754−60.
(59) Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics 25* (16), 2078−9.
(60) Zhang, Z., Li, J., and Yu, J. (2006) Computing Ka and Ks with a consideration of unequal transitional substitutions. *BMC Evol. Biol. 6*, 44.
(61) Manary, M. J., Singhakul, S. S., Flannery, E. L., Bopp, S. E., Corey, V. C., Bright, A. T., McNamara, C. W., Walker, J. R., and Winzeler, E. A. (2014) Identification of pathogen genomic variants through an integrated pipeline. *BMC Bioinf. 15*, 63.

DOI: 10.1021/acsinfecdis.5b00049
ACS Infect. Dis. 2015, 1, 367−379