DISSECTING THE EFFECTS OF SELECTIVE PRESSURES ON THE GENOMES OF CO-ENDEMIC
PLASMODIUM VIVAX AND *PLASMODIUM FALCIPARUM* IN CAMBODIA

Christian Michael Parobek

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the
Curriculum of Genetics and Molecular Biology in the School of Medicine.

Chapel Hill
2016

Approved by:

Steven R. Meshnick

Jeffrey A. Bailey

Kristina De Paris

Corbin D. Jones

Jonathan J. Juliano

Praveen Sethupathy

ABSTRACT

Christian Michael Parobek: Dissecting the Effects of Selective Pressures on the Genomes of Co-endemic *Plasmodium vivax* and *Plasmodium falciparum* in Cambodia
(Under the direction of Jonathan J. Juliano)

Each year, the malaria parasites *Plasmodium vivax* and *Plasmodium falciparum* together infect hundreds of millions of people and cause hundreds of thousands of deaths. In Cambodia, these two parasites are co-endemic, transmitted by the same vectors, and cause disease in the same populations. Recently, pointed efforts have been made in western Cambodia to eliminate malaria. This selective pressure provides a natural experiment that enables direct comparison of *P. vivax* and *P. falciparum* populations. Here, we explore the genetic similarities and differences between co-endemic *P. vivax* and *P. falciparum* from Cambodia through three questions to be answered. First, we conducted a pilot study in which we deep sequenced two important antigens - *pvmsp1* and *pvcsp* - from 48 *P. vivax* clinical isolates and explored the contrasting population genetics of these two antigens. We discovered that, in population genetic studies, marker choice has a profound effect on study outcomes, and that *P. vivax* and *P. falciparum* orthologs can host very different signatures of selection. Next, through whole-genome sequencing of 80 *P. falciparum* isolates, we found that population substructuring associates with artemisinin-combination therapy partner-drug *in vitro* IC_{50} and with clinical outcomes. This finding suggests that the *P. falciparum* population has in part responded to selective pressures with complex demographic changes. Finally, we whole-genome sequenced 70 sympatric *P. vivax* isolates. We found that, in spite of control efforts and in contrast to *P. falciparum*, the genetic diversity within the *P. vivax* population remains high and has been more rapidly expanding. We identified a minority of orthologous loci that have opposite signatures of selection in *P. vivax* and *P. falciparum*. Surprisingly, we found

evidence of several strong and recent selective sweeps within the *P. vivax* population at transcriptional regulatory loci. This finding suggests that *P. vivax* may rely on a nuanced response to selective pressure, modulating transcript levels as a means to maintain population resilience.

ACKNOWLEDGEMENTS

interpretation. Drs. Meshnick and Bailey have been close at hand from the inception of this work, and have offered guidance at critical junctures. Dr. Jones has been a wealth of knowledge pertaining to sequencing options and population genetics concepts. Drs. De Paris and Sethupathy taught excellent infectious-disease and population-genetics courses, and have brought their expertise to bear on this project.

Many others deserve recognition. I would like to thank my past mentors, Kenneth Poss and Viravuth Yin, who gave me foundational skills that I have been able to build on at UNC. Throughout graduate school, Andrew Morgan has provided stimulating conversation and thoughtful commentary pertaining to these projects. And, of course, none of this work would have been possible without the administration and administrative support provided by the MD/PhD program and the Genetics and Molecular Biology curriculum.

Finally, I want to thank my entire family. Each one of you has been so supportive, and loving. Thank you. My parents and grandma have been positive, encouraging, and a joy to be around. For my entire life, you have been there, and I am beginning to realize just how blessed I am for that. Last, but not least, I want to thank my beautiful wife, Bisset. You are so wise, even-keeled, and fun. I know this process would not have been near as rewarding without you at my side. You are my hero, and I love you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

3D7	*Plasmodium falciparum* reference-genome strain
ACT	artemisinin-combination therapy
AFS	allele-frequency spectrum
AIC	Akaike information criterion
AL	artemether-lumefantrine, an ACT
AMA	apical merozoite antigen
AMOVA	analysis of molecular variation
AS-AQ	artesunate-amodiaquine, an ACT
AS-MQ	artesunate-mefloquine, an ACT
BIC	Bayesian information criterion
CNV	copy-number variation
CP	population identifier prefix
CQ	chloroquine
CR	central-repeat region of the circumsporozoite surface protein
CSP	circumsporozoite protein
DALY	disability-adjusted life year
DHA-PPQ	dihydroartemisinin-piperaquine, an ACT
dN/dS	the ratio of nonsynonymous to synonymous mutations
EHH	extended-haplotype homozygosity
F_{ST}	Wright's fixation index
F_{WS}	measure of within-isolate clonality
G6PD	glucose-6-phosphate dehydrogenase
GATK	genome-analysis toolkit
GDP	gross domestic product
GO	gene ontology

HARB	high-affinity reticulocyte binding cluster
HRP2	histidine-rich protein 2, the target antigen for a malaria rapid diagnostic test
HTA	heteroduplex tracking assay
IC_{50}	half-maximal inhibitory drug concentrations
iHS	integrated haplotype score, a haplotype-based test for detecting sweeps
IQR	interquartile range
IRS	indoor residual spraying
ITN	insecticide-treated bednets
LD	linkage disequilibrium
MDR	multidrug-resistance transporter
MK	McDonald-Kreitman test
MOI	multiplicity of infection
MQ	mefloquine
MSP	major surface protein / merozoite-stage protein
N_{eff}	effective population size
nS_L	a haplotype-based test for detecting sweeps
P.	*Plasmodium*
PCA	principal components analysis
PCT	parasite clearance time
PPQ	piperaquine
RBC	red blood cell
RDT	rapid diagnostic test
Sal1	*Plasmodium vivax* reference-genome strain
S_{NN}	Hudson's nearest-neighbor statistic
TACT	triple artemisinin-combination therapy, artemisinin plus two partner drugs
WHO	World Health Organization

η_G/η_D factor of population size expansion or contraction

θ population mutation rate

π nucleotide diversity

CHAPTER 1: THE CLINICAL BURDEN, CAUSES, AND PATHOGENESIS OF MALARIA

1.1   Epidemiology

Malaria is a mosquito-borne, parasitic illness that affects hundreds of millions of people each year. Although global efforts to eliminate malaria have reduced its burden, in 2015 there were still 214 million malaria cases (range: 149 - 303 million) and 438,000 deaths (range: 236,000 - 635,000) [1]. Thus, despite recent advances in the fight against the disease, malaria is still a leading cause of morbidity and mortality throughout much of the world.

Malaria is a disease of poverty [2]. Of the 106 countries with reported active malaria transmission in 2000 [1], nearly all were tropical or subtropical low-income developing countries. Within these countries, the most affected individuals are children under five [3] (who are at increased risk prior to developing partial immunity), the immunocompromised and pregnant women [4] (as malaria parasites find a home in the placenta), and people who spend time outdoors (for example farming or in forested areas) (eg. [5,6]). In 2010, malaria cost nearly 70,000,000 disability adjusted life years (DALYs) globally [3], and it is estimated that persistent malaria infection during childhood is associated with up to a 50% decrease in adult income [7]. Globally, countries with nationwide malaria transmission have 1.6% decreased growth in gross domestic product (GDP) than countries with less of a malaria burden [2]. These trends likely hold true in Southeast Asia, and specifically in Cambodia, where the present studies take place. In 2013, the last year for which there is published data, the Cambodian Ministry of Health and the National Malaria Programme reported there were roughly 2000 confirmed malaria cases per month [8], with some evidence of recent case-incidence increases [9].

1.2   Malaria Biology, Lifecycle, and Host Responses

Malaria is caused by parasites of the genus *Plasmodium*. Malaria parasites infect a variety of hosts, ranging from reptiles and birds to mammals, primates, and humans. *Plasmodium* parasites are

transmitted by mosquitos of the genus *Anopheles*, which are ubiquitous throughout the world. Five of these *Plasmodium* species can cause disease in humans: *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae*, and *P. knowlesi*. Together, *P. falciparum* and *P. vivax* cause the majority of cases globally, with *P. falciparum* being more lethal.

All *Plasmodium* parasites share a similar, and complex, lifecycle. The meiotic and sole diploid stage occurs in the *Anopheles* midgut. After meiosis, haploid sporozoites are produced, which migrate to the mosquito salivary glands. While the mosquito draws a bloodmeal from a potential host, sporozoites deposit in the host interstitial space, from which they migrate into the bloodstream and infect hepatocytes [10]. After several days of intra-hepatic development (variation is species specific), infected hepatocytes release schizonts into the bloodstream. Schizonts rupture, each releasing numerous merozoites, which infect circulating red blood cells (RBCs). Infected RBCs undergo multiple cycles of parasite replication, rupture, and re-infection, causing both a recurring fever and a rapid increase in parasitemia. As a byproduct of bloodstage infection, sexual forms, called gametocytes, are produced, which are ingested by mosquitoes, repeating the lifecycle.

With repeated *Plasmodium* infections, hosts develop partial immunity that may be strain transcending; however, absent continued exposure this immunity wanes, though it does not completely disappear [11]. Consistent with this, most malaria cases occurring in the United States are among travellers returning from visiting friends and relatives who live in malaria endemic settings. Many of these travellers formerly lived in malaria endemic settings and presumably had partial immunity to malaria, which was lost after immigrating to the United States [12]. Importantly, the correlates of immunity to malaria (both *P. falciparum*-type and *P. vivax*-type) remain unknown, hindering vaccine development.

1.3   Clinical Characteristics and Treatment

The clinical manifestations of malaria are caused by blood-stage parasitemia, and are most pronounced in children and adults without immunity [13]. Typically, clinical symptoms commence within days to weeks of infection, but can manifest more than two months after infection [14]. Malaria can

present as uncomplicated or severe. In general, malaria lacks a pathognomonic feature, but rather presents as an undifferentiated fever. Patients presenting with uncomplicated disease have a fever that often abates and recurs with some periodicity, and can also present with headache, nausea, hepatomegaly, and splenomegaly. Severe cases can present with evidence of organ dysfunction including a dramatic anemia due to erythrocyte lysis, acute respiratory distress, renal compromise, shock, or neurologic symptoms due to sequestration of infected erythrocytes in the cerebral vasculature resulting in severe inflammation.

Treatment guidelines are species-specific and are determined on a country-by-country basis (dictated by each nation's health ministry or malaria control program). Global and national guidelines are frequently updated in response to updated drug resistance profiling. Since 2001, artemisinin-combination therapy (ACT) has been the WHO-recommended first-line therapy for *P. falciparum*. ACT is now also a WHO-recommended therapy for *P. vivax*, and is the preferred treatment for *P. vivax* in areas of high *P. vivax* chloroquine resistance [15]. The importance of artemisinin for global public health was recently recognized when the 2015 Nobel Prize for Medicine was shared by the scientist who isolated artemisinin, Tu Youyou.

1.4 Malaria Prevention

Beyond treating active malaria cases, prevention strategies are effective at reducing transmission and cases [16]. The primary preventative strategies are methods of vector controls, including indoor residual spraying (IRS), distribution and use of insecticide-treated bednets (ITNs), mass-screen-and-treat or mass-drug-administration campaigns, prophylactic pharmacotherapy in pregnant women, the use of transmission-blocking pharmaceutical agents in malaria patients, deployment of malaria village workers, and continued efforts to develop an effective vaccine.

Although the correlates of immunity against *P. falciparum* and *P. vivax* are poorly understood, pharmaceutical companies, global malaria control efforts, and academic research groups have expended vast resources pursuing a malaria vaccine. Malaria vaccines in the general (i.e. non-pregnant) population can be grouped into three broad classes: pre-erythrocytic, erythrocytic (also known as "blood-stage"), and

transmission-blocking. Two vaccines against the pregnancy-specific *P. falciparum* antigen *var2csa* are also undergoing active development.

Anti-*P. falciparum* vaccine candidates have progressed further than vaccines against *P. vivax*, with a pre-erythrocytic (CSP-based) [17,18] and an erythrocytic candidate (AMA1-based) [19] having completed advanced clinical trials in Africa. Both of these vaccines, however, showed disappointing efficacy, due in large part to standing antigenic diversity leading to vaccine escape [19,20]. Novel antigens with more constrained genetic diversity have been identified and are now being pursued for vaccine development [21,22]. Going forward, longitudinal studies of antibody titers against the entire array of potential malaria antigens as immunity develops in children hold great potential for identifying and quantifying the antigenic correlates of protection against malaria. Protein microarray and antigen phage displays are enabling this work [21,23,24].

1.5 Emerging Drug Resistance and Elimination Efforts in Southeast Asia

In 2009, a significant number of *P. falciparum* clinical cases with a delayed parasite clearance time (PCT) in response to artemisinin treatment were recognized in western Cambodia [25]. The assertion that these cases were "artemisinin resistant" (despite the lack of outright treatment failure), spurred a debate over the inappropriate use of the term "resistance" [26,27]. Unfortunately, there is now evidence of true artemisinin and ACT treatment failures in *P. falciparum* cases [28]. In some parts of Cambodia, ACT failure rates now approach 40% [29]. For simplicity, I will use "artemisinin resistant" and "artemisinin resistance" to refer both to cases of delayed clearance and outright failure, unless otherwise specified.

To neutralize the threat of potential artemisinin resistance, the WHO, with funding support from the Gates Foundation, developed the Global Plan for Artemisinin Resistance Containment [30]. The stated goal of this ambitious plan was to eliminate artemisinin-resistant drug alleles from western Cambodia. In focal areas of artemisinin resistance, it called for increased use of vector-control tools like insecticide-treated bednets. It also provided for increased availability of authentic antimalarial combination therapies, as well as increased monitoring, including active case detection, focal screen and treat programs, mass screen and treat programs, and mass-drug-administration programs.

Although this effort reduced the number of *P. falciparum* cases in western Cambodia, the frequency of artemisinin-resistant infections increased [31]. The resistance phenotype has now been reported in five countries within the Mekong region [1,32]. Unfortunately, elimination of artemisinin and partner-drug resistance is no longer a realistic goal of malaria-control. Rather, control efforts are focused on treating with enhanced cocktails of antimalarial drugs, and new efforts are underway to test out new treatment regimens in Cambodia. Moreover, there is much discussion regarding the development of new pharmacotherapies, and there are a number of new compounds in development [33–35].

1.6   The Genetics of Artemisinin and Partner-Drug Resistance

Despite the early fears of artemisinin-resistant malaria, years passed before a "resistance marker" was identified. Early association studies identified several genomic regions that were associated with resistance phenotypes, including a portion of chromosome 13 [36,37]. In 2013, a large population-based sequencing study identified multiple populations of genetically distinct artemisinin-resistant parasites circulating in Cambodia [38], but did not identify a genetic resistance marker. Then, in 2014 Ariey *et al.* found that parasites subjected to continuous artemisinin exposure developed non-synonymous mutations in the *P. falciparum kelch* K13 propeller domain (located on chromosome 13). The authors especially noted the presence of a C580Y mutation, and found that this mutation increased in frequency from 2001-2012 [39]. Follow-up work demonstrated unequivocally that C580Y and other K13-domain mutations do confer resistance to artemisinin *in vitro* [40]. Several mechanisms for *P. falciparum kelch* K13-mediated artemisinin resistance have been proposed. First, it was recognized that mutations in the human homolog KEAP1 are implicated as driver mutations in non-small-cell lung cancer [41]. Malaria-specific studies have shown associations between *P. falciparum* kelch mutations and increased phosphatidylinositol-3-kinase levels [42] and delayed parasite maturation through the intraerythrocytic stage [43].

A great concern is that artemisinin resistance in *P. falciparum* will spread to or develop in Africa. The fact that there have been multiple independent origins of artemisinin resistance in Southeast Asia [44] heightens the urgency of monitoring African *P. falciparum* clinical isolates for evidence of novel artemisinin-resistance markers in *kelch* or other genes. Collectively, we and others have found that

5

although there is deep genetic diversity at the *kelch* locus in African isolates, there is no evidence of a selective sweep around this locus among African *P. falciparum* populations [45–47]. This surprising finding could be due to (1) the lack of requisite haplotype background that enhances the selective advantage conferred by *kelch* mutations [48] from African isolates, or (2) disadvantageous allele pairings that could hinder the transmission of resistance alleles from Southeast Asian isolates to African isolates, similar to a mechanism uncovered previously in South American *P. falciparum* parasites [49].

Even as the threat of artemisinin resistance grows, its combination-therapy partner drugs are also losing effectiveness due to growing resistance. In Cambodia, increasing failure of artemisinin plus mefloquine (MQ) led to the adoption of artemisinin plus piperaquine (PPQ) as front-line therapy in 2012 [28]. Now, artemisinin/piperaquine combination therapy is also losing its effectiveness [28,50]. Though not a closed book, the molecular basis for MQ resistance in *P. falciparum* parasites is better understood than artemisinin. Numerous studies have found that MQ resistance is associated with copy-number variations or mutations at drug-resistance transporters [51–53]. Studies are ongoing to identify validated PPQ molecular resistance markers. However, *in vitro* and *ex vivo* resistance testing of parasites in Cambodia has shown dramatic increases in IC_{50} to piperaquine among parasite isolates.

1.7   *Plasmodium vivax* in an Era of Artemisinin Resistance

Globally, *Plasmodium vivax* causes at least 80 million cases of malaria each year [54], and, in Cambodia, *P. vivax* accounts for >50% of the total malaria burden [8]. In global guidelines, ACTs are now the recommended treatment for *P. vivax* in areas of chloroquine resistance [15]. In Cambodia, both CQ and ART are used for treatment of bloodstage *P. vivax*, and ACT is becoming the preferred treatment as it is now frontline therapy for *P. falciparum* [29]. Although *P. vivax* - like *P. falciparum* - develops resistance to antimalarial therapy, there have been no reports of artemisinin-resistant *P. vivax* clinical cases. Additionally, there is no evidence of potentially drug-resistance-causing mutations in the *P. vivax kelch* 13 propeller domain (our unpublished observations and [55]). In addition, although genetic markers of chloroquine resistance are present [56], Cambodian *P. vivax* remains sensitive to that former first-line therapy [29]. Despite its sustained and broad sensitivity to pharmacotherapy, *P. vivax* case incidence has

increased in Cambodia, even as *P. falciparum* cases have fallen dramatically [8]. There is growing

concern - both in Cambodia and around the world - that the burden of *P. vivax* malaria is increasing and

that it will be more difficult to eliminate than *P. falciparum* malaria.

1.8   Summary and Questions to be Answered

Malaria is a widespread and potentially deadly infectious disease caused by *Plasmodia spp.*

parasites. Though control efforts in the last decade have dramatically reduced the burden of malaria, it

still infects hundreds of millions of people a year, and kills hundreds of thousands. Despite these

advances, there are some concerning trends:

1. Resistance to first-line combination therapies is becoming more prevalent among *P. falciparum*

   from Southeast Asia, raising fears that resistance alleles may spread to other parts of the world,

   increasing global malaria-attributable morbidity and mortality.

2. Even as the incidence of *P. falciparum* malaria has decreased by >80% in parts of Cambodia,

   cases of *P. vivax* malaria have increased, raising concerns *P. vivax* will be more difficult to

   eliminate.

These trends raise several questions that must be answered to build on the successes of malaria-

elimination programs:

1. To what extent does next-generation sequencing marker choice affect the conclusions drawn from

   these studies?

2. What are the selective pressures that shape global genetic diversity at *P. vivax* antigens, and how

   does this compare with the pressures on *P. falciparum* orthologs?

3. Why is *P. vivax* more resilient to control measures than *P. falciparum*, and what are the genetic

   mechanisms of this enhanced resilience?

4. What are the genetic determinants of ACT partner-drug resistance, and how do partner-drug and

   artemisinin resistance interact?

5. Why are there multiple co-localized but genetically distinct populations of ACT-resistant *P.*

   falciparum circulating in Cambodia?

CHAPTER 2: DIFFERING PATTERNS OF SELECTION AND GEOSPATIAL GENETIC DIVERSITY WITHIN TWO LEADING *PLASMODIUM VIVAX* CANDIDATE VACCINE ANTIGENS[1]

2.1   Introduction

Plasmodium vivax causes 80 to 300 million infections per year and over 2.5 billion people remain at risk of infection despite malaria elimination efforts [57]. Now, concern over *P. vivax* is growing due to reports of increasingly severe disease [58], emerging chloroquine resistance [59], and multidrug resistance [60]. Ultimately, an effective vaccine will be important for controlling *P. vivax* malaria [61]. The fact that humans naturally develop partial immunity to *P. vivax* and *P. falciparum* lends hope for effective vaccines against these parasites; however, because the majority of global malaria research funding targets *P. falciparum* [54,62], only a handful of *P. vivax* antigens are currently being considered for vaccine development [63]. Among these are *P. vivax* merozoite surface protein 1 (*pvmsp1*) and circumsporozoite protein (PvCSP).

PvMSP1, an erythrocytic vaccine candidate, plays an important role in reticulocyte invasion [64]. Its C-terminus contains a 42 kDa region, which is processed into 33 and 19 kDa fragments (**Figure 2.1A**). The 33 kDa fragment contains two high-affinity reticulocyte binding clusters (HARBs) (20 kDa and 14 kDa), and antibodies against the HARBs confer protection in monkeys [65]. In humans, antibodies to the 42 kDa region have also been associated with clinical protection, making this region an attractive vaccine candidate [66–69]. Another vivax protein, PvCSP, is a pre-erythrocytic vaccine candidate and is critical in sporozoite motility and hepatocyte invasion [70]. *P. vivax* circumsporozoite protein has an immunogenic central repeat, consisting of two major types of nonapeptide repeats (VK210 and VK247 – there is also a

---

rarer repeat type termed vivax-like) flanked by highly conserved 59 and 39 regions (**Figure 2.1B**). The *P. falciparum* ortholog of *pvcsp*, as formulated in RTS,S, is the most advanced *P. falciparum* vaccine candidate to date, showing modest efficacy at one year interim analysis in a Phase III trial [17].

Despite this knowledge of *pvmsp1* and PvCSP, little is known about the geospatial genetic diversity of these antigens. Variation in these antigens may become a mechanism of vaccine resistance if strain-specific immunity is important for protection, as has been seen in some *P. falciparum* vaccine candidates [71]. Vaccine trials of *P. falciparum* AMA1 and MSP2 as well as genetic crosses using *P. chabaudi* underscore the importance of strain-specific immunity as a determinant of outcome [19,72–74]. Additionally, despite initial evidence that strain-specific immunity may not impact RTS,S efficacy [75–77], the incomplete protection afforded by the RTS,S vaccine in Phase II and III trials [17,78,79] has prompted a careful examination of strain-specific responses to this vaccine. Thus, as momentum grows for field trials of *P. vivax* vaccine antigens, carefully designed population genetic studies of *P. vivax* vaccine candidates will be key to assess the need for multivalent vaccine formulations.

To better understand the selective forces on, and geospatial genetic diversity associated with *pvmsp1* and *pvcsp*, we used the Illumina sequencing platform to determine haplotypes for 42 kDa region of *pvmsp1* (n=44) and we used the PacBio and Illumina platforms to sequence the complete *pvcsp* gene (n=47) from Cambodian isolates [80]. To dissect the immune selection acting on these regions, we studied these sequences using population genetic tests of selection and models of tandem repeat evolution. To evaluate the global genetic diversity of *pvmsp1* and *pvcsp*, we extracted worldwide *pvmsp1* and *pvcsp* sequence data available in GenBank (n=238 for *pvmsp1* and n=412 for *pvcsp*) (**Figure 2.2**), and studied our sequence data alongside the sequences from GenBank. Finally, we compare the performance of Illumina and PacBio sequencing to traditional Sanger sequencing, and discuss the potential and challenges of next-generation sequencing for population genetic studies of malaria parasite antigens.

2.2   Materials and Methods

2.2.1 Parasite isolates

Clinical samples from a previous study were used for this study [81]. Written informed consent was acquired from each individual and the study was approved by the IRB at University of North Carolina, the IRB of the Naval Medical Research Unit #2, Jakarta, Indonesia, and the Cambodian National Ethical Committee for Health Research. Briefly, blood spots were collected from 109 patients with uncomplicated *P. vivax* malaria, presenting to a clinic in Chumkiri, Cambodia during 2006–07. We selected 48 subjects with a multiplicity of infection (MOI) of one (n=20) or two (n=28) for sequencing. MOI was determined by heteroduplex tracking assay (HTA) [80,82]. Briefly, in an HTA, radiolabeled DNA probes are annealed to genomic DNA and drawn through a non-denaturing gel matrix. The number of bands observed represents the number of conformational differences present among heteroduplexes, and is a proxy for the number of infection clones (MOI). Details of the method have been published elsewhere [83].

2.2.2 Amplification of pvmsp1 and pvcsp

The *pvmsp1* 42 kDa region (nucleotides 3973–5239 of Sal1 PVX_099980, www.PlasmoDB.org) was amplified using primers F: 5'-CAG GAC TAC GCC GAG GAC TA-3' and R: 5'-GGA GGA AAA GCA ACA TGA GC-3' and an Eppendorf Mastercycler (Eppendorf, Hauppauge, NY) in 50 mL reactions containing 5 μL 10x Qiagen Hotstar Master Mix (Qiagen, Valencia, CA), 0.25 μL Qiagen Hotstar Taq, 300 nM forward primer, 300 nM reverse primer, 1 μL 10 mM dNTPs, and 5 μL 5–10 mM template. Cycling conditions were: 95℃ x 15m; 35 cycles of 95℃ x 45s, 55℃ x 45s, 72℃ x 3m; and 72℃ x 10m. The *pvcsp* gene (PVX_119355) was performed by nested PCR. The outer step used primers F: 5'-GGC AAA CTC ACA AAC ATC CA-3' and R: 5'-TGC GTA AGC GCA TAA TGT GT-3'. Reactions were as above except for 600 nM forward primer, 600 nM reverse primer, 1 μL 10 mM dNTPs, 5 μL 5–10 mM template, 6 μL of 25 mM MgCl$_2$, and 28.75 μL H$_2$O. Cycling conditions were: 95℃ x 15m; 25 cycles of 95℃ x 45s, 45℃ x 45s, 72℃ x 3m; and 72℃ x 10m. The inner step used 600 nM of each of the primers F: 5'-AAA CAG CCA AGG CCT ACA A-3' and R: 5'-GAC GCC GAA AAT ATT GGA TG-3' using 5–

10 μL of the initial amplification. The cycling conditions were: 95℃ x 15m; 25 cycles of 95℃ x 45s, 54℃ x 45s, 72℃ x 3m; and 72℃ x 10m.

2.2.3 Amplicon sequencing and sequence determination

  pvmsp1 and *pvcsp* amplicons were fragmented by acoustic shearing (Covaris, Woburn, MA) using the following settings: 10% duty cycle, 5.0 intensity, 200 cycles per burst, and frequency sweeping mode. Forty-eight barcoded libraries were prepared using the NEXTflex multiplex library kit (Bioo Scientific, Austin, Texas), each containing the pooled *pvmsp1* and *pvcsp* amplicons from one patient. Libraries were sequenced on the Illumina HiSeq2000, using the paired-end 100 base pair chemistry (Illumina, San Diego, CA). We used Lasergene SeqMan NGen v.3.1.1 (DNASTAR, Madison, WI) to assemble *pvmsp1* short reads *de novo* and to determine SNP frequency within each assembly. For purposes of comparison and confirmation, we re-sequenced 9 *pvmsp1* amplicons with differing MAFs: 3 samples with all MAFs>90%; 3 samples with all MAFs between 60% and 90%; 3 samples with MAF<60% for at least one SNP. Sanger-sequence haplotypes were compared to predicted Illumina haplotypes. Based on these comparisons, only predicted *pvmsp1* haplotypes with MAF>60% at all polymorphic sites were used in our analysis. In addition to Illumina sequencing, *pvcsp* amplicons were sequenced using PacBio Circular Consensus Sequencing (Pacific Biosciences, Menlo Park, CA). One PacBio SMRT cell produced a total of 12,103 reads with a minimum of 36 circular consensus coverage, which were used for this study. These were further filtered, removing truncated reads or reads with errors in the barcode. This left 8430 reads (3979 forward and 4451 reverse). Clustering attempted to minimize false positive haplotypes due to erroneous base calls and PCR slippage within the tandem repeat region. For each sample, haplotypes were created by clustering reads, allowing reads differing only by indels of 1 and 2 bases and low quality mismatches to collapse. Low quality was defined as either a mismatching base Q<30 or any Q<25 within an 11 basepair region centered on the mismatch, as has been applied previously to rigorous SNP discovery from shotgun data [84]. To overcome artifacts of PCR infidelity due to slippage events leading to shortened repeats and false haplotypes, we set a high threshold requiring that co-occurring haplotypes of the same repeat type beat high frequency in order to exclude the low

frequency variation/stuttering in the repeat region. Haplotype repeat type was then determined by translation and the most frequent haplotype of each major repeat type (VK210 and VK247) present was kept >0.5%. Additional haplotypes of major repeat types were kept if they were common (>20%) and thus unlikely to be due simply to low frequency slippage events. In total, across all samples, 4081 of the 8430 reads clustered contributed to utilized haplotypes.

The long-read haplotypes determined through consensus clustering were used as templates for short-read alignment using *Bowtie2* v2.1.0 [85], with very-sensitive alignment parameters and stringent filtering for Mapping Quality Score and Alignment Score. Final sequence predictions were used for the analyses in this paper and were deposited in GenBank under accession numbers JX461243-JX461285 and KJ173797- KJ173802 for *pvcsp*, and JX461286-JX461333 for *pvmsp1*.

Rarefaction curves of haplotypes were calculated using *EstimateS* v9.0. Individual-based curves using sampling without replacement were estimated [86] and extrapolated to 2x the actual sample number [87]. Rarefaction plots were visualized in the *R* base package (http://cran.us.r-project.org/).

2.2.4 Acquisition of published sequences for inter-population comparisons

GenBank was queried for population sets published prior to August 1, 2013, which included sequence data for the 42 kDa region of *pvmsp1* and the whole-gene of *pvcsp*. Sequences from a recent publication [88] were excluded because the isolates were collected over the course of a 12 year period. The authors provide evidence that the haplotype distribution of this population changed substantially over time, making this population inappropriate for our analysis of selection.

2.2.5 Assessing selection on pvmsp1 and pvcsp

Population datasets with >25 sequences that were collected over a span of ≤4 years were included for analysis of selection. We used *DnaSP* v5.1 to perform tests of selection [89]. We calculated polymorphism and Tajima's *D* across *pvmsp1* and the *pvcsp* constant regions using a 50 bp sliding window with a 25 bp step size. We also performed 1000 coalescent simulations with recombination to determine a 95% confidence interval and centile for each Tajima's *D* estimate [90]. To test for long-term selection, we used the McDonald-Kreitman (MK) test [91]. Skew was calculated using Fisher's exact test

(two tailed). For the *pvmsp1* 42 kDa region amplicons reported here and by others, 15 *Plasmodium knowlesi pkmsp1* isolates from Thailand [92] (accession numbers JF837339-JF837353) were used as the interspecies outgroup. Three insertions and deletions occurred in the 42 kDa region of *pvmsp1* relative to *pkmsp1*, and were not considered. We could not obtain MK estimates for *pvcsp* sequences due to numerous insertions and deletions relative to *pkcsp*. For analysis of *pvcsp* repeats, we performed pairwise comparisons of untranslated repeat units within individual *pvcsp* sequences [93]. We calculated skewness and mean nucleotide differences between repeat units, as previously reported [94]. Similar to the methods of Dias et al., 2013, we also calculated dN/dS on the first 1–459 bases of all 32 VK210 repeat regions and the first 1–540 bases of all 15 VK247 repeat regions. This analysis was performed in *MEGA5*, using the Nei-Gojobori method [95].

2.2.6 Phylogenetics and statistics to determine population structure

Interpopulation heterogeneity was first assessed using Wright's fixation index (F_{ST}). Pairwise fixation values between *pvmsp1* populations were calculated in *DnaSP*. Site-specific fixation values for pairwise comparisons among Cambodia, NW Thailand, S Thailand, India, and Turkey were generated using the analysis of molecular variance (AMOVA) function within *Arlequin* v3.11 [96].

Neighbor-joining trees for *pvmsp1*, *pvcsp* VK210, and *pvcsp* VK247 were drawn using the *ape* package for *R* [97]. To generate trees based off *pvmsp1*, distance calculations between haplotypes were performed in *MEGA5* using the maximum composite likelihood method to construct a neighbor-joining tree file. For the *pvcsp* CR, we used *MS_Align* (v.2.0) [98,99] to create genetic distance matrices separately comparing both the VK210 and VK247 repeat arrays. *MS_Align* generates an event-based genetic distance using a model of tandem repeat evolution (expansion, deletion, substitution). Cost parameters for *MS_Align* were set to 0.1 for amplification or contraction and 5 for repeat insertion or deletion. A pairwise cost table of repeat-to-repeat mutations was created in *MEGA5* using the maximum composite likelihood method and used as input for *MS_Align* [93,100]. *MS_Align* output matrices were used by *FastME* [101,102] to construct neighbor-joining trees with balanced branch-length estimation. To cluster geographic groups, we calculated Hudson's nearest-neighbor statistic (S_{NN}) [103]. Input was in the

13

form of a pairwise distance matrix between all haplotypes for each phylogeny. For this statistic, highly distant populations have values approaching 1 while panmictic populations have values near 0.5. To test the reproducibility of the geographic clustering predicted by S_{NN}, 1000 jackknife samplings were constructed for both *pvmsp1* and *pvcsp* VK210 and VK247 populations using *Fast UniFrac* [104]. For each jackknife replicate, 5 individuals, based on the size of the smallest population, were randomly selected from each population and used to redraw trees. Observed splits between geographic populations were quantified and used to assign confidence to predicted geographic clusters. To evaluate potential mutational paths connecting all *pvmsp1* haplotypes, we constructed a median-joining network using *NETWORK* v4.6 (Fluxus Engineering, Suffolk, England) [105]. This method expresses multiple plausible evolutionary paths in the form of cycles. A similar analysis was not completed for *pvcsp* due to the variable length of CR haplotypes.

2.3 Results

2.3.1 pvmsp1 sequences

We Illumina sequenced *pvmsp1* 42 kDa-fragments (**Figure 2.1A**) from 48 patients, and compared these to Sanger sequencing data for selected samples. Illumina haplotypes with a major allele frequency of >60% agreed with Sanger haplotypes in every case tested (n=6). Illumina haplotypes with a major allele frequency of <60% did not consistently agree with Sanger haplotypes (n=3). Thus, we were able to build 44 complete *pvmsp1* 42 kDa haplotypes (26 unique haplotypes) with a major allele frequency of >60% at all polymorphic sites (**Table 2.1**). The average coverage depth for all isolates was >800 reads per base, with all bases having ≥100 reads of coverage. Haplotype accumulation (rarefaction) curves were estimated, and then further extrapolated to show that our sample captured fewer than half the total *pvmsp1* haplotypes in this region of Cambodia (**Figure 2.3**). In addition to these isolates, we identified 238 submissions in GenBank [106–110] (**Table 2.2**) containing either the whole-gene or 42 kDa-region sequence information.

2.3.2 Detecting signatures of selection within pvmsp1

The interaction between human host and the parasite has had a profound impact on the parasite genome, leaving behind characteristic ''signatures'' of natural selection [111], which are detectable using population genetics approaches to examine sequence diversity. We first assessed nucleotide diversity (**Figure 2.4A**), and observed a spike of polymorphism in the region between the two HARBs (positions 4348–4731 in the Sal1 reference). We termed this the ''intervening region''. To test whether the diversity in the intervening region is due to long-term selection, we used the McDonald-Kreitman (MK) test [91] to compare the ratio of non-synonymous to synonymous nucleotide polymorphisms between the Cambodian *P. vivax* population and a Thai *P. knowlesi* population [92]. We observed a highly elevated MK ratio (*p*=0.00427) in the intervening region but not in the HARBs (data not shown) or the entire 42 kDa region (*p*=0.681), suggesting that the intervening region is under long-term selective pressure (**Table 2.3**). To determine whether the long-term selective pressure shaping the intervening region is potentially due to human immunity, we assessed balancing selection in this region, as balancing selection within a malaria antigen suggests that the antigen is a target of the human immune system [111]. We applied Tajima's *D* test of neutrality [112] to five geographically distinct *P. vivax* populations (all populations with n>25, accounting for 190 of 238 available sequences) (**Table 2.1**, **Figure 2.4B**). In panmictic populations with an uncomplicated demographic history [111], the Tajima's *D* statistic can indicate whether a nucleotide sequence is under directional (*D*<0) or balancing selection (*D*>0). Populations not subjected to recent bottlenecks (i.e. Cambodia, India, and NW Thailand, [106,110]) demonstrated a significant signature of balancing selection in the *pvmsp1* 42 kDa region (**Table 2.1**). This signature occurred specifically in the intervening region (**Figure 2.3B**), and is consistent with the conclusion that human immunity targets the intervening region. The three regions of the *pvmsp1* fragment that are considered vaccine candidates were each assessed for diversity in the Cambodian population [64,113]. In contrast to the intervening region, the 20 kDa HARB (Sal1 positions 4021–4347) and 14 kDa HARB (Sal1 positions 4732–4941) showed no coding polymorphisms and no evidence of balancing selection, similar to recent reports [113]. The 19

15

kDa fragment (Sal1 nucleotide positions 4918–5239) also showed limited diversity, with only a K1709E

substitution, and no evidence of balancing selection.

2.3.3 Geospatial genetic diversity at the pvmsp1 42 kDa region

Although the *pvmsp1* 42 kDa region contains potential vaccine candidates [64,113], the 42 kD

region's global genetic diversity has not been carefully evaluated. To study *pvmsp1* 42 kDa diversity, we

calculated Wright's Fixation index (F_{ST}) [114] for each pairwise comparison between five diverse

populations (**Table 2.4**). F_{ST} values between naturally evolving parasite populations (Cambodia, NW

Thailand, and India) approached zero, showing a high degree of genetic similarity, while comparisons

with populations that have undergone a recent bottleneck (S Thailand and Turkey) showed a high degree

of genetic distance due to their limited number of haplotypes. Similarly, F_{ST} values calculated for each

variable site demonstrate a high degree of homogeneity in pairwise comparisons between the Cambodia,

NW Thailand, and India populations across all sites, and substantial heterogeneity between S Thailand

and Turkey across all sites (**Figure 2.5**). This is evidence that balancing selection maintains a similar

range of alleles in the *pvmsp1* 42 kDa region of multiple geographically diverse naturally evolving *P.

vivax* populations. To visualize whether 42 kDa sequences cluster according to geography, we compared

all unique haplotypes in a single neighbor-joining tree, which revealed little clustering according to

geographic origin (**Figure 2.6**). We quantified the extent of this clustering using Hudson's nearest-

neighbor statistic (S_{NN}), which assesses how frequently a variant's nearest neighbor is from the same

population [103]. In both global and pairwise comparisons, *pvmsp1* 42 kDa sequences from naturally

evolving populations in Cambodia, India, and NW Thailand showed no evidence of strong geographic

clustering (**Table 2.5**). To further confirm this finding, a neighbor-joining consensus tree was created and

underwent 1000 jackknifed replicates (**Figure 2.7A**). Results showed that the predicted splits between

most populations occurred only less than 50% of the time, providing strong evidence that there is minimal

geographic clustering of *pvmsp1* 42 kDa sequences. To better understand the evolutionary relationships

between *pvmsp1* haplotypes from around the world, we employed a median-joining network to describe

the set of potential mutational paths between all available global *pvmsp1* 42 kDa sequences [105]. The

network shows extensive admixture of parasite populations from diverse locales, with numerous mutational paths connecting haplotypes (**Figure 2.8**). With the exception of populations from S Thailand and Turkey, which have undergone recent bottlenecks, these data provide further evidence that there is no clustering by geography.

2.3.4 pvcsp sequences

We sequenced the complete *pvcsp* gene from 43 isolates using the PacBio and Illumina platforms. *De novo* assembly of the Illumina paired-end short reads was not possible, due to over-collapse in the central repeat (CR) region, resulting in inappropriately short CRs. In contrast, PacBio long reads allowed the gene to be sequenced in its entirety and, after clustering, predicted 47 *pvcsp* haplotypes within the 43 samples. Reported error rates for PacBio sequencing have been high, especially for indels [115]; however, the use of Circular Consensus Sequencing allows single DNA fragments to be read multiple times, decreasing the error rate of the final predicted sequence. To check the accuracy of PacBio haplotypes, individual haplotypes were used as a template for alignment of Illumina reads from the same clinical isolate. The addition of Illumina reads corrected only a single 1-bp deletion in a single haplotype. Therefore, after clustering, PacBio-predicted haplotypes have an error rate of 1/(~1200 basepairs/sequence x 47 sequences), or approximately 0.002%.

Considering the entire gene, there were 24 unique haplotypes at the nucleotide level, and most genetic diversity was within the CR (**Figure 2.1**). Both nonapeptide repeat array types - VK210 (total n=32, range 17-21 repeat units) and VK247 (total n=15, range 20-21 repeat units) – were represented in our Cambodian population, with no VK210–VK247 hybrids (reviewed in [116]). The average Illumina short-read depth for each isolate was >1000, with all bases having ~5 reads of coverage. In addition to our isolates, we identified one cohort of nearly complete *pvcsp* sequences (n=27), and 12 cohorts of CR sequences (n=385) [117–122] (**Table 2.1**). An extrapolated rarefaction curve showed that we sampled more than two thirds of the *pvcsp* CR haplotypes in this part of Cambodia, and that there are significantly fewer *pvcsp* CR variants in this region of Cambodia than *pvmsp1* 42 kDa variants (**Figure 2.3**).

2.3.5 Detecting signatures of selection within pvcsp

In contrast to *pvmsp1*, the 5' and 3' non-repeat regions of *pvcsp* had no significant signatures of selection either by the MK test (data not shown) or Tajima's *D* test (**Table 2.1**). The 5' non-repeat region in the Cambodian cohort showed a non-significant signature of balancing selection (**Table 2.1** and **Figure 2.3D**), which was due to a G38N amino acid polymorphism. This polymorphism also was observed in 6/16 parasites from the Latin Pacific region (JQ511263-JQ511276, JQ511279, JQ511286) and 2/27 parasites from Colombia (GU339072 and GU339085). The 39 non-repeat region had little evidence of balancing selection, with Tajima's *D* values <0 (**Table 2.1** and **Figure 2.4D**). Within *pvcsp*, an 18 amino-acid C-terminal motif known as Region II (amino acid residues 311–328 in Sal1) is important for parasite invasion of hepatocytes [123] and purportedly contains both B and T-cell epitopes [124,125]. Among all Cambodia and Colombia parasite isolates, this motif is completely conserved at the nucleotide and protein level, with an amino-acid sequence of EWTPCS VTCGVGVRVRRR, similar to previous reports [113].

To better understand the selective forces acting upon the *pvcsp* CR, we assessed the dN/dS ratio for Cambodian VK210 and VK247 [118]. Strikingly, synonymous substitutions were strongly favored in both VK210 (dN/dS = 0.267; Z test *p*<0.001) and VK247 (dN/dS = 0.166; Z test *p*<0.001) repeats. This is consistent with the finding that VK210 and VK247 isolates from around the world consistently demonstrate a depressed dN/dS ratio, suggesting that the VK210 and VK247 repeat regions are both under strong purifying selection [118]. The CR of *P. falciparum csp* is thought to evolve by slipped-strand mispairing [94]. To understand if a similar mechanism works in the *pvcsp* repeats, we studied the mismatch distribution of pairwise genetic distances between untranslated repeat units within each VK210 and VK247 repeat array type in Cambodia. Consistent with another study [120], we observed a strong right skew in the proportion of genetic differences between pairwise VK210 repeat comparisons, and between pairwise VK247 repeat comparisons, evidence that *pvcsp* repeats have a high proportion of identical or nearly identical repeats (data not shown). This finding is consistent with a continuous and rapid expansion and contraction of repeats by slipped-strand mispairing, which may be a mechanism to evade host immunity [94].

2.3.6 Geospatial genetic diversity at the pvcsp central repeat

A recent study assessed global genetic diversity in the *pvcsp* CR, but did not define the correlates of differentiation between populations [118]. Moreover, this report investigated CR diversity by using a subset of the repeat region that was invariant in length. This approach may not reflect true population structure as it only assesses repeats early in the CR. Indeed, we have found that certain repeat types do cluster in locations within the repeat arrays (data not shown). To more rigorously study the global diversity of the *pvcsp* CR, we modeled CR repeat expansion, contraction, and substitution using *MS_Align*, which calculates an event-based genetic distance between CR haplotypes [98]. From these data, we constructed neighbor-joining trees for global VK210 and VK247 repeat arrays isolates (**Figure 2.9** and **Figure 2.10**). In contrast to *pvmsp1*, the VK210 and VK247 trees revealed striking geographic clustering by country and continent. We quantified clustering using Hudson's S_{NN}, and observed strong genetic differentiation between most geographically diverse parasite populations, in contrast to *pvmsp1* (**Table 2.5**). To confirm this finding, neighbor-joining consensus trees for both VK210 and VK247 were subjected to 1000 jackknife replicates and the reproducibility of predicted splits between populations was tested demonstrating a strong correlation between genetic distance and geography (**Figure 2.7B–C**).

We were able to define the peptide sequence basis of the clustering observed among *pvcsp* CR repeats. For VK210 repeats, almost all (81/84) Latin American repeat arrays contained either a 59 (GDRADGQPA) 4 or an internal (GDRADGQPA) 3–4, while very few (11/278) of the Asian sequences contained one or both of these features. Similarly, for VK247 repeat arrays, all (34/34) Latin American sequences began with a single EDGAGDQPG repeat, while only one (1/44) Asian sequence began with this repeat. These sequence features may represent a reliable method to assign sequences to a geographic region.

2.4   Discussion

This study (1) presents the first population set of *pvmsp1* and *pvcsp* sequences from Cambodia, (2) identifies a signature of putative immune-mediated, frequency-dependent selection in the *pvmsp1* 42 kDa region and the *pvcsp* CR, and (3) provides the most comprehensive evaluation to date of geospatial

genetic diversity for these genes. We also demonstrate the feasibility of using a next-generation

sequencing approach to study the genetic diversity of malaria antigens.

A distinguishing feature of this study is the use of next-generation sequencing methods to

generate *P. vivax* amplicon sequence data from clinical isolates. This work represents a first step into this

largely unexplored territory. As a relatively new technology, next-generation sequencing methods must be

validated before use in molecular epidemiological studies. We provide evidence that the dominant

Illumina-predicted *pvmsp1* haplotypes are consistent with Sanger sequencing, and are fit for comparison

with population sets generated by traditional sequencing methods. Methods for predicting multiple

haplotypes from short-read sequencing are under development and will need further validation. We also

demonstrate the ability of combined PacBio-Illumina haplotypes to predict *pvcsp* VK210 and VK247

haplotypes out of individual mixed infections. As next-generation sequencing methods are utilized more

frequently for population genetic studies of infectious diseases, the methods introduced here will be

further improved and will help to provide greater insight into *Plasmodia* population genetics.

2.4.1 Evidence of selection in both pvmsp1 and pvcsp

We found compelling genetic evidence that the *pvmsp1* 42 kDa intervening region is under strong

immune pressure in multiple panmictic populations. Results from the MK test suggested that this region

is under sustained selective pressure (Table 3); however, because a positive MK test can signify balancing

selection or weak negative selection [126,127], we tested the hypothesis that this region is under

balancing selection using Tajima's *D* test of neutrality. Since multiple populations showed strong

evidence of balancing selection by Tajima's *D* (**Table 2.1**, **Figure 2.4B**), we conclude that the intervening

region is undergoing continual diversifying, balancing selection. An alternative hypothesis is that the

positive Tajima's *D* values are an artifact of recent population contractions. Because (1) a positive

Tajima's *D* was observed in multiple populations, and (2) other regions of *pvmsp1* contained negative

Tajima's *D* values, we conclude that the 42 kDa intervening region of *pvmsp1* undergoes frequency-

dependent (and likely immune-mediated) balancing selection.

Because *pvmsp1* is a merozoite surface antigen, it is highly accessible to antibodies and complement. The predicted structure of the 42 kDa region shows that the 33 kDa fragment covers the 19 kDa fragment [66,128], limiting its exposure to the human immune system relative to the 33 kDa fragment. This observation could explain the extensive balancing selection present in the 33 kDa fragment (specifically, the intervening region) but not in the 19 kDa fragment. Additionally, this finding suggests that the sliding window approach for evaluating polymorphism and balancing selection may help generate hypotheses about functionally important (19 kDa fragment, for example) or immunologically dominant (the intervening region, for example) regions of *P. vivax* proteins.

For *pvmsp1*, Tajima's D and F_{ST} were inversely correlated. Populations with strong evidence of high Tajima's D in the *pvmsp1* intervening region showed a low genetic differentiation by F_{ST}. This suggests that in naturally evolving populations, diversification of this region is extensive and maintains a similar range of genetic diversity despite geographic distance. Populations that have undergone a recent bottleneck show a low Tajima's D with relatively few variants and strong genetic differentiation from more diverse populations. This suggests that if strain-specific immune responses are important in vaccine efficacy, vaccines may work more effectively if other interventions can be used to bottleneck the population, thus decreasing its genetic diversity [106].

The central repeat region (CR) is a primary immunodominant region of PvCSP. Though alignment-based methods to assess for selection (Tajima's D, for example) cannot be employed in a tandem repeat region, there is wide-ranging evidence that selective pressures shape the genetic composition of the *pvcsp* CR [129–133], including new evidence hinting that hosts develop strain-specific immunity to *P. falciparum* NANP repeats of varying lengths [134]. Indeed, the presence of two distinct repeat types (VK210 and VK247) may itself be evidence of selection as suggested in a study of the *P. cynomolgi csp* CR [132].

Our analysis of the two CR array types, VK210 and VK247, also suggests that selection is occurring in this region. In pairwise comparisons of nucleotide and amino acid differences, we observed a positive skew showing decreased differences among repeat units. This finding is consistent with Patil et

al.'s study of *pvcsp* isolates from Brazil [120,135], and provides further evidence that both VK210 and VK247 repeat arrays may continuously evolve via slipped-strand mispairing [94]. Furthermore, consistent with a recent study of selection in worldwide *pvcsp* isolates [118], we found that Cambodian *pvcsp* VK210 and VK247 isolates have a strong bias toward synonymous substitutions. This signature of purifying selection is consistent with reports from *pfcsp* [136–138] and suggests that there are a limited number of amino acid polymorphisms allowable within this repeat region. Taken together, these findings suggest that expansion, contraction, and rearrangement of repeat units, rather than generation of novel repeat units through mutation, maintain genetic diversity at the *pvcsp* locus in both VK210 and VK247 variants. This phenomenon may be responsible for immune evasion [120,135].

Although these two *P. vivax* genes are orthologs of well-characterized vaccine candidate antigens from *P. falciparum* malaria, substantial differences are seen in the effects of immune selection between these genes and their orthologs. Previous reports have shown that the functionally similar *pfmsp1* 42 kDa fragment has relatively low nucleotide diversity and lacks evidence of balancing selection by Tajima's *D* [139]. *P. falciparum csp*, on the other hand, shows a high level of nucleotide diversity [140–142] and modest Tajima's *D* elevations in the C-terminal T cell epitopes [140,143]. These patterns are in stark contrast to our observations in *P. vivax*, and this highlights the need for *P. vivax*-specific studies to determine appropriate candidate vaccine antigens.

Finally, our analysis of the *pvmsp1* 42 kDa region underscores the importance of selecting an appropriate parasite population for population-genetic studies. We did not observe signatures of balancing selection in *pvmsp1* populations from S Thailand or Turkey. This is likely due to bottlenecks secondary to robust malaria control measures employed in S Thailand [106] and limited human migration in Turkey [110]. Thus, appropriate selection of panmictic populations for these studies is critical.

2.4.2 Differing patterns of geospatial genetic diversity at pvmsp1 and pvcsp

Using both tree-based and statistical methods [144], we found that *pvcsp*, but not *pvmsp1*, showed strong clustering by geography (**Tables 2.4-2.5**, and **Figures 2.6–2.10**). For *pvmsp1*, we observed little geographic clustering among naturally evolving parasite populations, suggesting that immune selection

22

maintains similar *pvmsp1* alleles around the globe. Notably similar findings have been described in Duffy Binding Protein and Thrombospondin-related anonymous protein in *P. vivax* malaria [113], while a recent global survey of diversity in the Apical Membrane Antigen 1 found evidence of geographically restricted haplotypes [145]. In contrast to *pvmsp1*, we found that pvcsp variants demonstrate strong evidence of geographic clustering. This juxtaposition between *pvmsp1* and *pvcsp* sequences is similar to what has previously been described for merozoite and sporozoite antigens in *P. falciparum* [146]. The population sets included in this survey were collected in different years. While it is known that novel *P. vivax* surface antigen types can appear in the course of a decade [109], it is difficult to

assess the magnitude of this effect on our analyses. As more *pvmsp1* and *pvcsp* population sets are collected, this will become clearer.

It is interesting that the CR of *pvcsp* shows evidence of multiple forms of selection: (1) the depressed number of non-synonymous mutations suggests purifying selection, (2) the differences in CR genotypes between geographic locations suggests directional selection, and (3) the genetic composition of the repeats suggests rapid expansion and contraction, possibly due to immune selection. It is unclear what drives the first two signatures of selection. We hypothesize a model in which purifying selection within a population limits the amino acid composition of repeats due to functional constraints of the protein, while directional selection between populations is driven by environmental factors.

One environmental factor that may explain both the purifying and directional selection of parasite *pvcsp* CR sequences is the mosquito vector. The circumsporozoite protein is expressed in the mosquito during oocyst development [147] and in the salivary glands [148,149]. It is also critical in sporozoite motility [70]. We found no overlap in the distribution of Anopheline species between the countries from Asia and Latin America included in this study (data not shown) [150–152]. Furthermore, there is substantial evidence that different Anopheline species and strains show differential ability to be infected by malaria [153–156].

Regardless of the cause of the differing patterns of geospatial genetic diversity we observed in *pvmsp1* and *pvcsp*, the observation itself has significance for vaccine design. The malaria vaccine field is

just beginning to unravel how antigenic diversity within a single parasite population can reduce vaccine efficacy [157]. Our findings highlight an additional level of complexity that will hinder the implementation of a *P. vivax* vaccine – antigenic variability. While the effects of immune cross-reactivity against different antigenic variants aren't fully known, the extensive intrapopulation variability seen in *pvmsp1* may necessitate a highly multivalent *pvmsp1* vaccine, while the dramatic interpopulation variability seen in *pvcsp* suggests that a PvCSP-based vaccine that is effective in one part of the globe may not be effective in other regions. Thus, a thorough understanding of the geospatial genetic diversity of candidate vaccine antigens must inform antigen selection for vaccine design.

2.4.3 Data Availability

The *pvcsp* sequences generated through this work are deposited at GenBank under accession numbers JX461243-JX461285 and KJ173797-KJ173802. The *pvmsp1* sequences generated through this work are deposited at GenBank under accession numbers JX461286-JX461333.

2.4.4 Author Contributions

Conceived and designed the experiments: JJJ CMP JAB. Performed the experiments: CMP. Analyzed the data: CMP JJJ JAB NJH. Contributed reagents/materials/analysis tools: DS WOR NJH. Wrote the paper: CMP JJJ JAB.

2.4.5 Declaration of Interests

The authors declare no conflicts of interest.

2.4.6 Funding

views expressed in this paper are those of the authors and do not represent the official position of the U.S. Department of Defense, NIH, or UNC Chapel Hill. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

pvmsp1: 42 kDa region							
Cambodia	44	62	24.8	0.020	26	0.950	2.08[1]
India	28	64	24.9	0.021	27	0.997	1.32[1]
NW Thailand	65	62	24.9	0.020	34	0.968	2.42[1]
S Thailand	67	42	6.46	0.005	5	0.336	-0.986
Turkey	30	33	8.33	0.007	3	0.536	-0.001
pvcsp: N- and C-terminal domains							
Cambodia	47	-	-	-	24	-	-
N-terminal non-repeat		3	0.971	0.003	3	0.500	0.901
C-terminal non-repeat		2	0.318	0.001	2	0.159	-0.538
Columbia	27	-	-	-	27	-	-
N-terminal non-repeat		2	0.285	0.001	2	0.143	-0.954
C-terminal non-repeat		0	-	-	1	0.000	-

Table 2.1 Summary population genetic data for *Plasmodium vivax* antigens.

This table includes all population sequence sets which contained sufficient numbers to perform allele-based tests of neutrality. Population sets which included sequence data only for *pvcsp* repeat regions alone are not summarized here. * *p*<0.05; [1] number of haplotypes; [2] within-population variant sites; [3] average number of nucleotide differences; [4] nucleotide diversity; [5] number of haplotypes; [6] haplotype diversity.

Gene	Region	Country	Year	Loci Included	N	Accession No.	Reference
pvmsp1	SE. Asia	Cambodia	2006-2007	42 kDa region	44	JX461286-JX461333	Present Study
	SE. Asia	NW Thailand	2006-2007	Whole gene	65	GQ890872-GQ890916; AF435595-AF435615	Jongwutiwes 2010
	SE. Asia	S Thailand	2006-2007	Whole gene	67	GQ890975 - GQ891041	Jongwutiwes 2010
	SE. Asia	N Thailand	1997-1998	Whole gene	19	AF435595-AF435614	Putaporntip 2002
	SE. Asia	Bangkok	1997	Whole gene	1	AF435615	Putaporntip 2002
	Asia Pacific	Vanuatu	1996,1998	Whole gene	2	AF435632, AF435634	Putaporntip 2002
	S. Asia	India	2004-2007	42 kDa	28	EU430452-EU430479	Thakur 2008
	S. Asia	India	1999	Whole gene	1	AF435639	Putaporntip 2002
	S. Asia	Bangladesh	1994	Whole gene	5	AF435616-AF435620	Putaporntip 2002
	E. Asia	S. Korea	1996-2009	Whole gene	8	HQ171934-HQ171941	Han 2011
	E. Asia	S. Korea	1998	Whole gene	4	AF435635-AF435638	Putaporntip 2002
	W. Asia	Turkey	2007-2008	Whole gene	30	AB564559-AB564588	Zeyrek 2010
	S. Amer	NW. Brazil	1995,1997	Whole gene	8	AF435622-AF435631	Putaporntip 2002
pvcsp	Asia Pacific	Cambodia	2006-2007	Whole gene	47	JX461243-JX461285, KJ173797-KJ173802	Present Study
	Asia Pacific	Thailand	2009*	CR	45	HQ011279-HQ011323	N/A
	S. Asia	India	2008*	CR	79	FJ491063-FJ491141	N/A
	E. Asia	Korea	2006*	CR	20	DQ859736-DQ859767	N/A
	E. Asia	Tibet	2008*	CR	37	FJ601724-FJ601761	N/A
	Oceania	Papua New Guinea	1999, 2001-2003	CR	18	EU031819-EU031836	Henry-Halldin 2011
	S. Asia	Sri Lanka	1998-2000	CR	60	JQ362595-JQ362654	Dias 2013
	S. Amer	Columbia	2009	Whole gene	27	GU339059-GU339086	Hernandez-Martinez 2011
	S. Amer	Brazil	2004-2005	CR	45	FJ845383-FJ845391	Patil 2010
	S. Amer	Guatemala	2012*	CR	7	KC154040-KC154046	Mendizabal-Cabrera[†]
	S. Amer	Latin Pacific	2012*	CR	24	JQ511263-JQ511286	Gonzalez-Ceron[†]
	S. Amer	Brazil	1997-1999	CR	41	DQ978649-DQ978689	Santos-Ciminera 2007
	S. Amer	Honduras	2010-2011	CR	9	JQ903593 - JQ903601	Lopez 2012

Table 2.2 *pvmsp1* and *pvcsp* sequences included in this study.

The PlasmoDB gene identifier is PVX_099980 for *pvmsp1* and PVX_119355 for *pvcsp*.Indicates the year the sequences were made available in GenBank. [†]Indicates study unpublished but sequences available in GenBank.

Inter-species McDonald-Kreitman comparisons				
	42 kDa region		42 kDa intervening region	
	Synonymous	Non-synonymous	Synonymous	Non-synonymous
Fixed	96	93	32	28
Polymorphic	34	39	10	31
	p=0.681		p=0.00427	

Table 2.3 McDonald-Kreitman test for selection in *pvmsp1*.

Evidence for long-term selective pressure on the *pvmsp1* 42 kDa region and the 42 kDa intervening region was assessed with the McDonald-Kreitman test, using *P. knowlesi* msp1 as the outgroup comparator. A Fisher's exact test (two tailed) was used to determine significance.

pvmsp1 Global F_{ST} 0.340				
pvmsp1 Pairwise	Cambodia	India	NW Thailand	S Thailand
India	0.031			
NW Thailand	0.000	0.043		
S. Thailand	0.449	0.433	0.366	
Turkey	0.361	0.329	0.403	0.796

Table 2.4 Interpopulation *F*-statistics for *pvmsp1*.

F_{ST} values compare the relatedness of a gene among different populations of the same species. Reported values compare the relatedness of *pvmsp1* 42 kDa alleles for pairwise comparisons between Cambodia, India, NW Thailand, S Thailand, and Turkey. F_{ST} values approaching 0 indicate greater relatedness, while values approaching 1 indicate substantial interpopulation variability. Global F_{ST} statistic calculated between all *pvmsp1* populations with n>25 indicates that relatively little genetic distance exists between the sampled populations. However, pairwise comparisons demonstrate that some populations exhibit a high degree of genetic similarity (Cambodia and India, for example) while other populations are more dissimilar (S Thailand and Turkey, for example).

pvmsp1 Global S_{nn} 0.410*	Cambodia	NW Thailand	S Thailand	India
Cambodia				
NW Thailand	0.318			
S Thailand	0.780	0.865		
India	0.673	0.779*	0.865	
Turkey	0.897	0.946	0.750	0.917

VK210 Global S_{nn} 0.517*	Cambodia	Thailand	India	Korea	Tibet	PNG	Sri Lanka	Brazil	Honduras
Cambodia									
Thailand	0.511								
India	0.824	0.778							
Korea	0.950*	0.92	0.977*						
Tibet	0.968*	1.000*	0.686	0.864					
PNG	0.7621	0.633	0.920*	0.923*	0.913				
Sri Lanka	1.000*	1.000*	0.955*	0.958*	0.926*	0.987*			
Brazil	0.844*	0.958*	0.943*	0.955*	0.969*	0.855	1.000*		
Honduras	0.824	1.000	0.976*	1.000	1.000*	1.000*	1.000*	0.881	
Guatemala	1.000*	0.933	0.970*	1.000	0.957*	0.929	0.960*	0.957*	0.708

VK247 Global S_{nn} 0.872*	Cambodia	Thailand	PNG
Cambodia			
Thailand	1.000*		
PNG	0.904*	0.939*	
Columbia	1.000*	0.981*	0.935*

Table 2.5 S_{NN} statistics for the *pvmsp1* 42 kDa region and the *pvcsp* central repeat region.

S_{NN} values approaching 1 indicate genetic isolation while values near 0.5 indicate that two geographically disparate population may approximate panmixia global and pairwise S_{NN} values show stronger geographic clustering among *pvcsp* VK210 and VK247 repeats than among *pvmsp1* 42 kDa regions.
* indicates significance to ($p<0.05$) after Bonferroni correction for multiple comparisons.
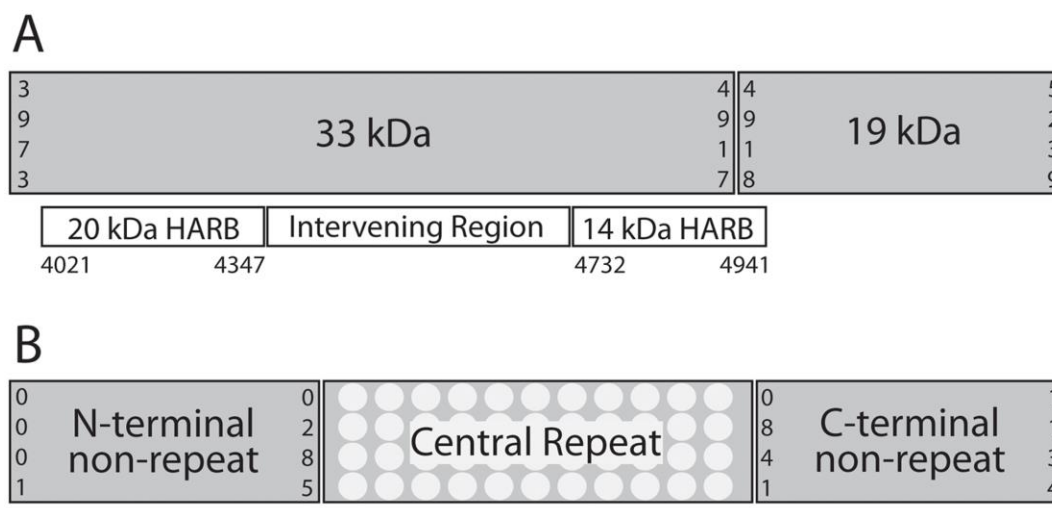
**Figure 2.1 Protein domains and immunologically-relevant regions
of *pvmsp1* 42 kDa region and *pvcsp*.**

For both genes, numbers indicate coordinates according to the Sal1 reference genes. Sequences for *pvmsp1* (PVX_099980) and *pvcsp* (PVX_119355) were accessed August 14, 2012 from PlasmoDB.org. (**A**) The *pvmsp1* 42 kDa region is composed of two primary subunits – a 33 kDa and a 19 kDa subunit. Other sub-regions, including the 20 kDa and 14 kDa HARBs have been previously defined and studied. Here, we define the region between the HARBs as the "intervening region." (**B**) The *pvcsp* gene is composed of three regions – an N-terminal non-repeat region, a central repeat region, and a C-terminal non-repeat region. The central repeat region consists of two major nonapeptide repeat types, termed VK210 and VK247. Approximate locations of *pvcsp* regions I and II are noted with horizontal lines in the N- and C-terminal non-repeat regions, respectively.
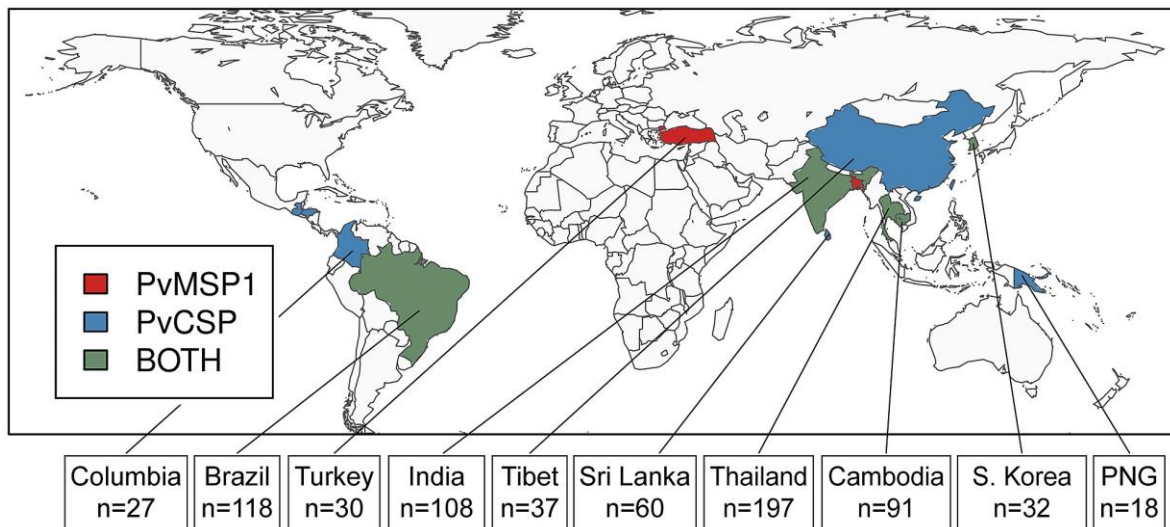
Figure 2.2 Geographic distribution of *P. vivax* populations contributing to this study.

In total, we identified 13 populations with *pvmsp1* 42 kDa fragment sequences and 13 populations with *pvcsp* central repeat or whole-gene sequences. These populations were collected from 14 countries, pictured above. For countries with n≥10 isolates, the total number of *pvmsp1* and *pvcsp* isolates is marked.
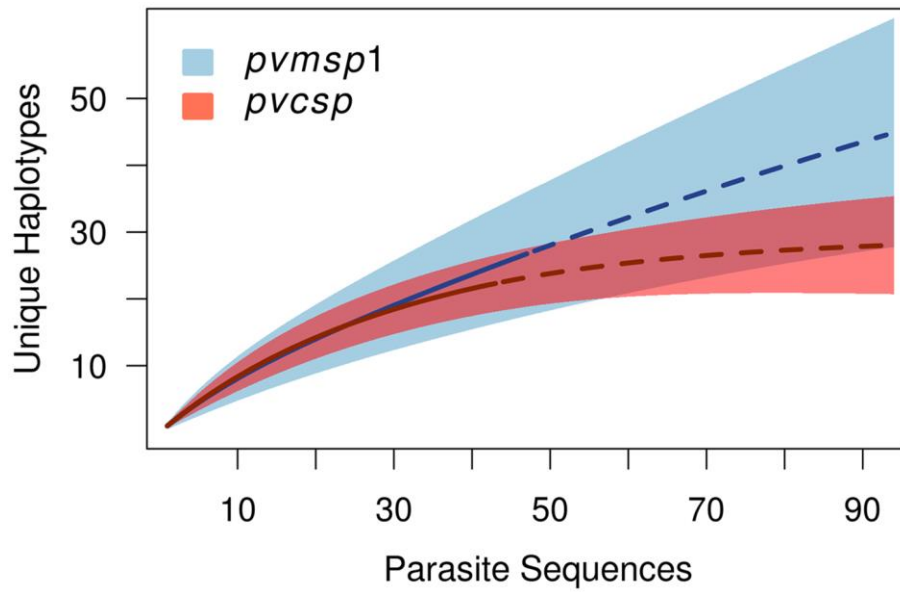
Figure 2.3 Haplotype rarefaction curves for the Cambodian cohort.

Calculated rarefaction curves are depicted by solid blue (*pvmsp1*) and red (*pvcsp*) lines. Dotted lines represent rarefaction values extrapolated according to the methods of Cowell, et al. The 95% CIs of rarefaction estimates for *pvmsp1* and *pvcsp* are demarked by light blue and light red shaded areas, respectively.
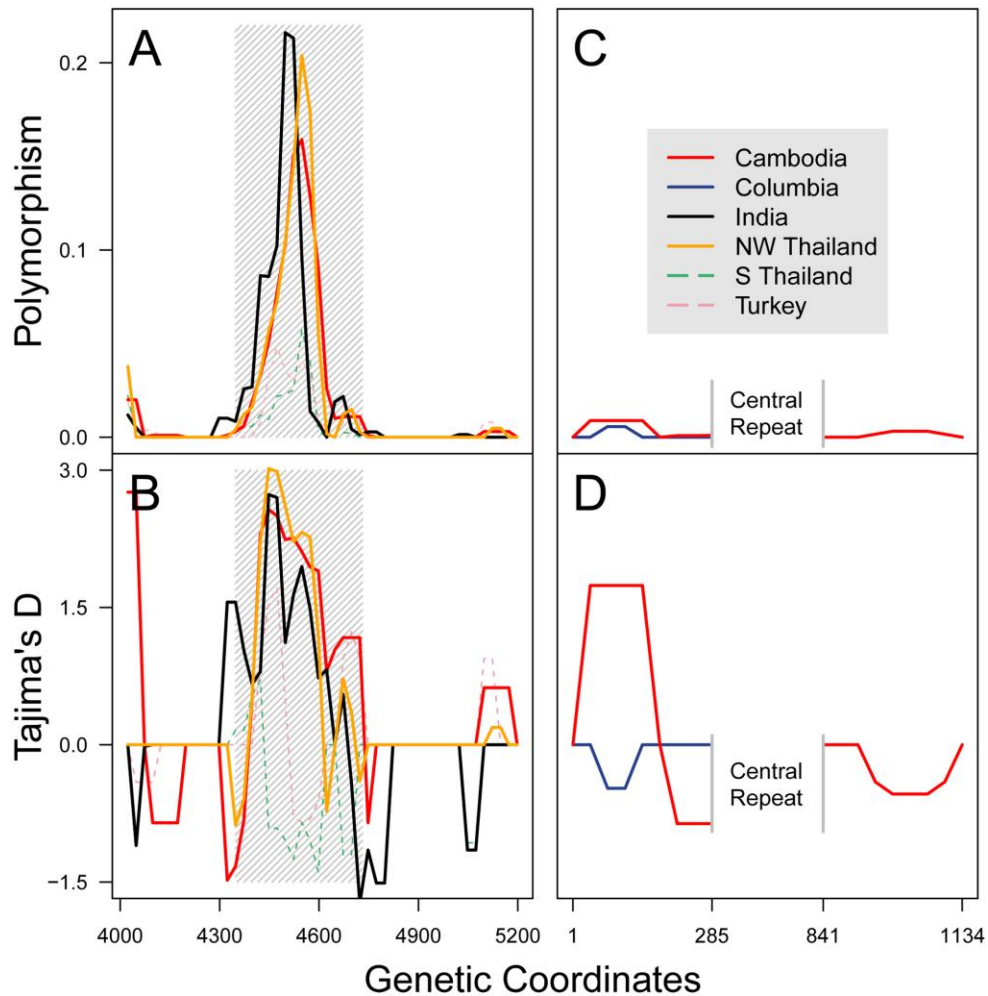
Figure 2.4 Nucleotide diversity and Tajima's *D* across the *pvmsp1* 42 kDa region and the whole *pvcsp* gene.

Polymorphism (nucleotide diversity, π) (**A**) and Tajima's *D* (**B**) were calculated across the *pvmsp1* amplicon for five diverse populations. A sliding window (50 bp window and 25 bp step size) was used to achieve a high resolution analysis. Grey hatches demark the intervening region (nucleotides 4348–4731). For *pvcsp*, N-terminal and C-terminal non-repeat regions were analyzed for nucleotide polymorphism (**C**) and evidence of balancing selection (**D**) using a sliding window. Putatively panmictic populations are marked with a solid line, while populations known to be subject to strong selective forces are marked with dotted lines. All coordinates are based on Sal1 *pvmsp1* and *pvcsp* reference sequences. Results of sliding-window analyses were fairly stable to varying window size (results not shown).
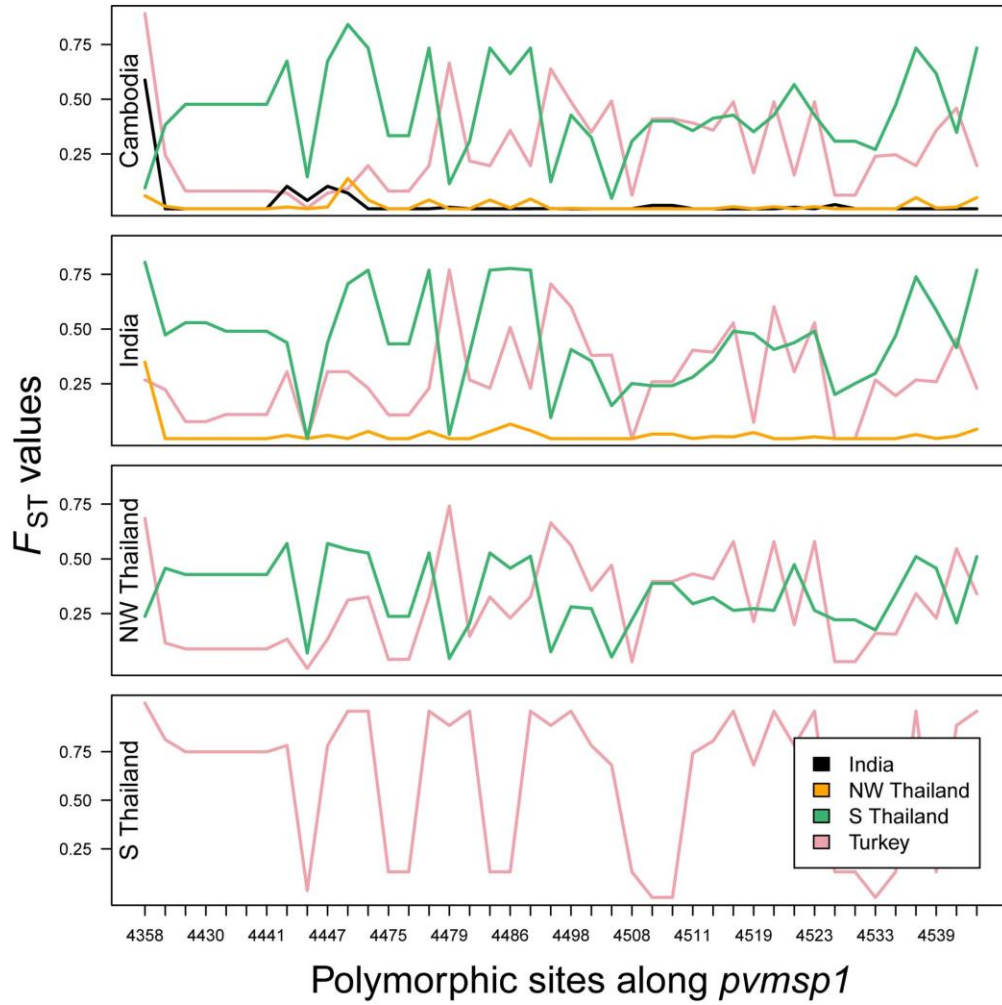
Figure 2.5 F_{ST} values at polymorphic sites within the *pvmsp1* 42 kDa intervening region.

Available parasite populations with n>25 individuals (Cambodia, India, NW Thailand, S Thailand, and Turkey) share 42 variable sites within the 42 kDa intervening region of *pvmsp1*. F_{ST} values for each variable site were calculated in a pairwise manner between all five populations. F_{ST} values approaching 0 indicate limited inter-population variability at that site, while values approaching 1 indicate substantial inter-population variability. Coordinates are reported for every third polymorphic site.
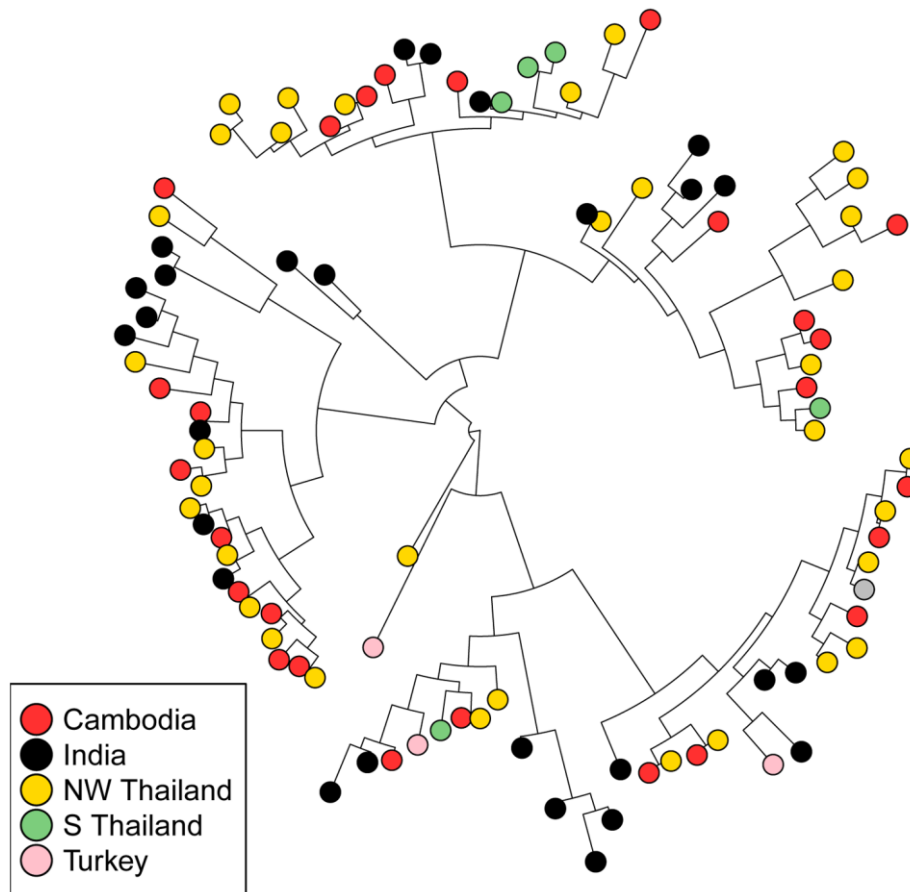
Figure 2.6 Neighbor-joining tree of 42 kDa regions from *pvmsp1* isolates shows significant admixture across geography.

All unique 42 *pvmsp1* population set with n>25 were plotted on a single unrooted, neighbor-joining phylogenetic tree. The Sal1 reference sequence is marked in grey.
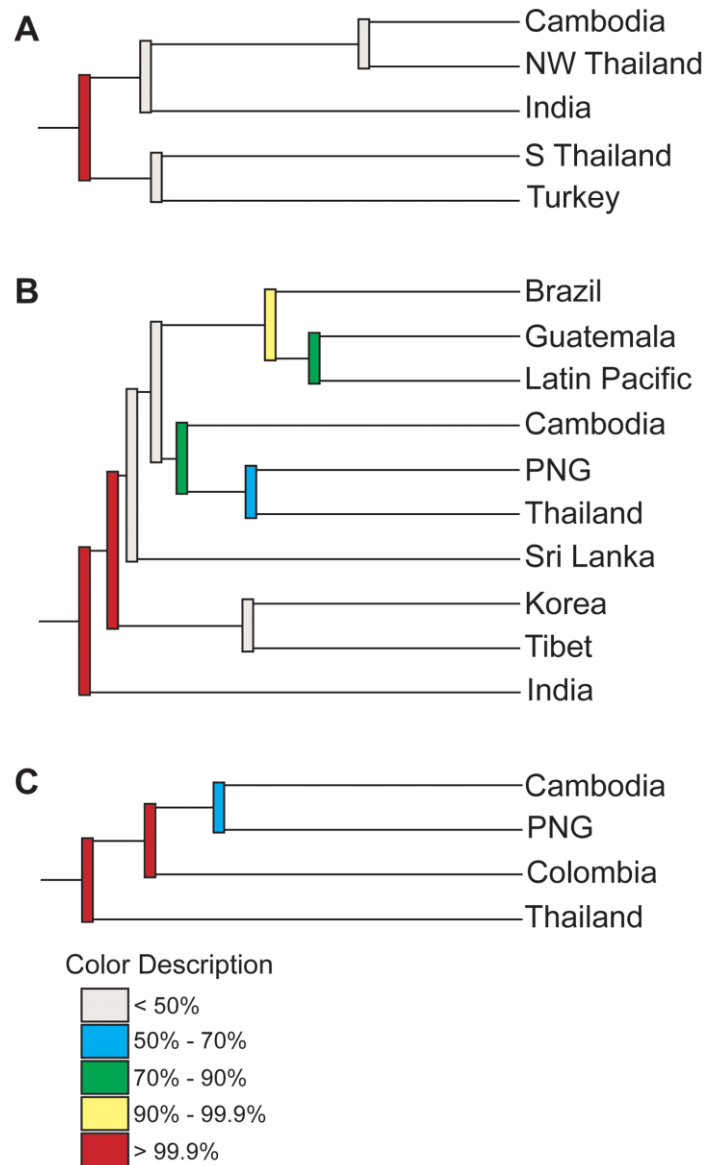
Figure 2.7 Jackknifed consensus trees demonstrate reproducible geographic clustering in *pvcsp* VK210 and VK247 isolates, but not *pvmsp1*.

The reproducibility of population clustering was assessed using 1000 jackknifed phylogenies. Individual populations clustered together or apart in each of the 1000 jackknifed phylogenies, and the frequency of a split between any two populations was quantified. Populations with grey bars (<50% splits) were genetically similar, while populations with red bars (>99.9% splits) were highly genetically distinct. Phylogenies were built from the *pvmsp1* 42 kDa region (A), the *pvcsp* VK210 central repeat (B), and *pvcsp* VK247 central repeat (C).
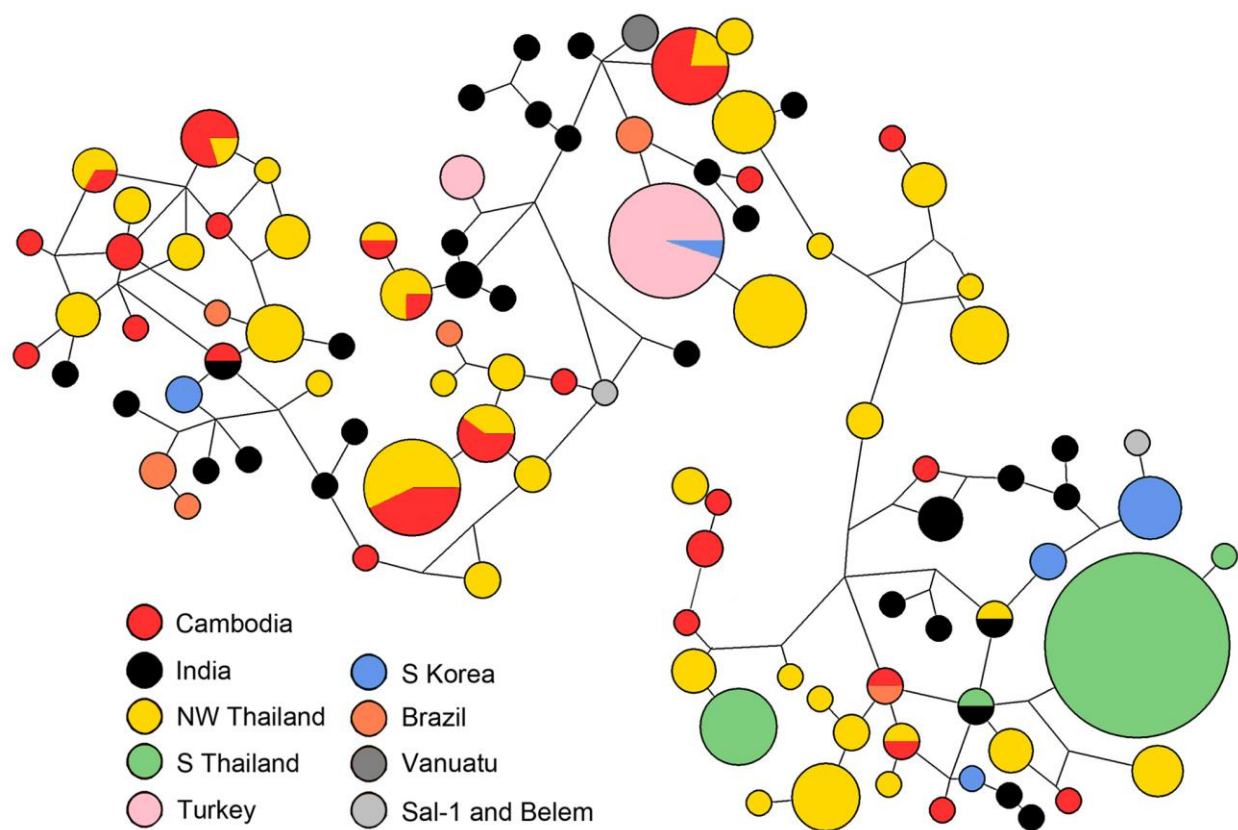
Figure 2.8 Median-joining network of diverse *pvmsp1* populations proposes multiple mutational paths between geographically diverse populations.

286 *pvmsp1* 42 kDa sequences from diverse geographical regions were used as input to create an unrooted median-joining network. This network is a visual representation of the mutational paths that may explain the observed sequence diversity. Each node represents an allele, node size represents the frequency of that allele (range n = 1 to n = 54), and node color corresponds to country of origin. Cycles within the diagram represent alternative evolutionary pathways. Corners represent obligate intermediate sequences that were not observed among the sampled alleles. Line length is not proportional to genetic distance.
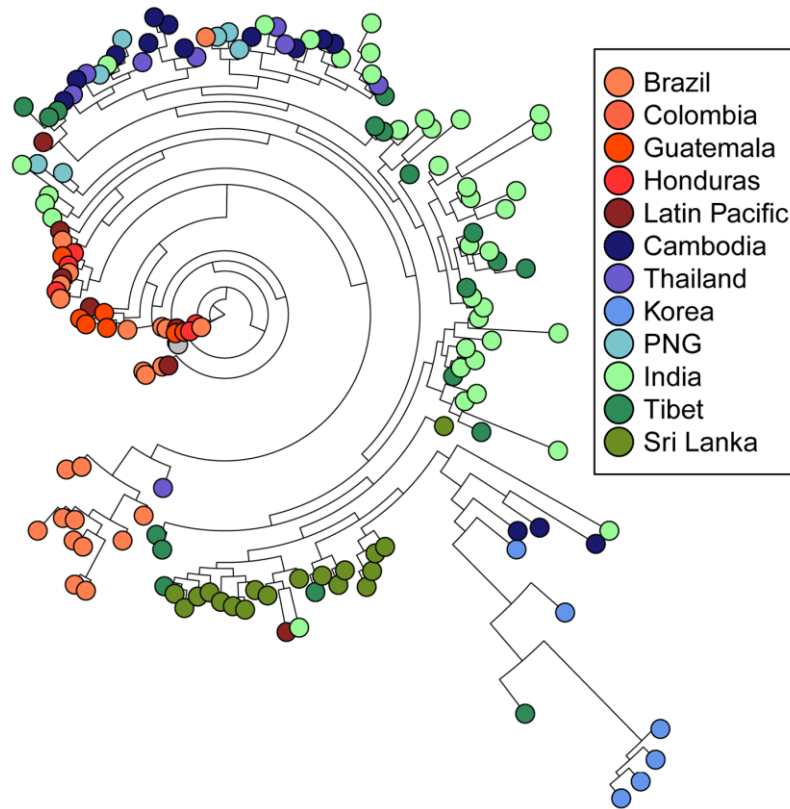
Figure 2.9 Neighbor-joining tree of *pvcsp* VK210 repeat arrays shows genetic clustering according to geography.

All unique repeat array haplotypes from each *pvcsp* VK210 population set were plotted on a single unrooted, neighbor-joining phylogenetic tree. Visual inspection reveals strong geographic clustering by region and country. Latin American sequences are in shades of red, South East Asian sequences are in shades of blue, and South and Central Asian sequences are in shades of green.
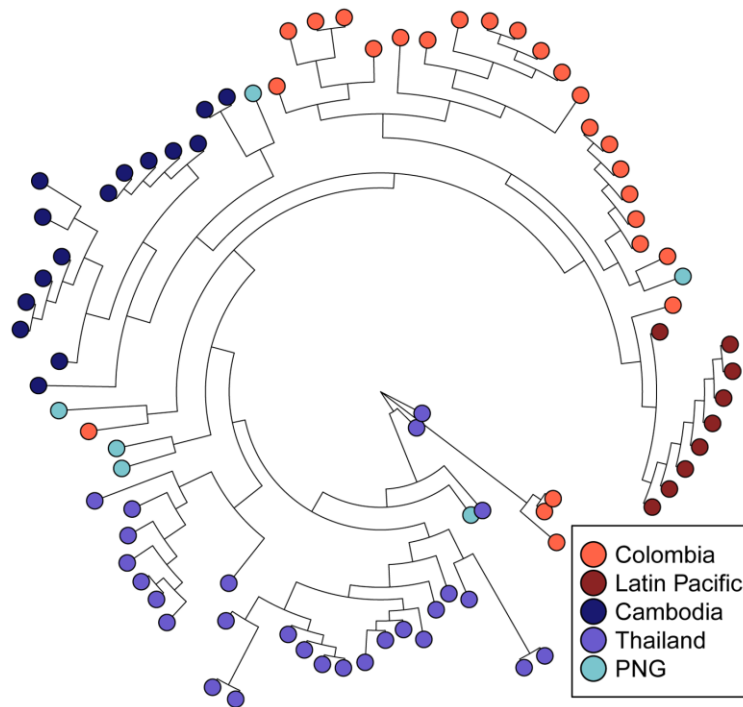
Figure 2.10 Neighbor-joining tree of *pvcsp* VK247 repeat arrays shows genetic clustering according to geography.

All repeat array haplotypes from each *pvcsp* VK247 population set were plotted on a single unrooted, neighbor-joining phylogenetic tree. Latin American sequences are in shades of red, South East Asian sequences are in shades of blue.

CHAPTER 3: GENETIC ARCHITECTURE OF ARTEMISININ-RESISTANT MALARIA IN CAMBODIA ASSOCIATES WITH PARTNER-DRUG RESISTANCE[2]

3.1 Introduction

Mortality due to malaria nearly halved over the last fifteen years, due in part to the adoption of artemisinin-based combination therapy (ACT) as first-line therapy. Resistance to ACT represents a major threat to malaria control efforts worldwide and has now been reported in five countries in Southeast Asia [158]. Efforts are underway to stem the rising tide of resistance in the region, and new antimalarial regimens are desperately needed.

One strategy that has been proposed is the use of triple ACT (TACT) regimens [159]. ACT is traditionally comprised of a potent, short-acting artemisinin derivative paired with a longer-acting partner drug, such as piperaquine, mefloquine, or lumefantrine. This allows ACT to be administered over just three days, but leads to a terminal phase of partner drug monotherapy [160,161]. It has long been recognized that, as artemisinin resistance increases and parasites take longer to clear, the probability of developing resistance to the partner drug also increases. TACTs would combine two partner drugs with artemisinin based on the epidemiologic observation that resistant parasites seem to display sensitivity to piperaquine when mefloquine resistant and visa versa [162–164].

In western Cambodia, increasing rates of artesunate-mefloquine (AS-MQ) treatment failure led Cambodia to adopt dihydroartemisinin-piperaquine (DHA-PPQ) as its first-line ACT in 2012 [163,165]. DHA-PPQ failure in Cambodia appears to be due to emerging piperaquine resistance in the setting of already established artemisinin resistance [162,163,165,166]. Research into the genetics of artemisinin

---

[2] This chapter is currently in submission. Full citation: Parobek CM*, Parr JB*, Balasubramanian S, Chaorattanakawee S, Gosi P, Barnett E, Lanteri CA, Lon C, Saunders DL, Lin JT*, Juliano JJ*. Asterisks indicate co-first and co-senior authors.

resistance in Southeast Asia has borne fruit with the discovery of specific resistance markers and their association with genetically distinct, clonally expanded parasite populations [166–168]. However, the role of ACT partner drugs in shaping these parasite populations has not been explored.

Using 78 samples collected from three provinces of Cambodia collected between 2011-2014, we investigated the genetic architecture of the *P. falciparum* population using whole genome sequencing. We confirmed the genotype of several loci using Sanger sequencing, in particular candidate indels, and assessed copy number of resistance loci using real-time PCR and digital droplet PCR. Finding stark genetic structuring among the samples, we used population modeling to deduce the origins and recent histories of parasite clusters. We then correlated the genetic findings to available *ex vivo* susceptibility data to piperaquine and mefloquine and *in vivo* response to dihydroartemisinin-piperaquine (DHA-PPQ) therapy. We found that piperaquine and mefloquine resistance developed in genetically distinct parasite populations that have expanded clonally from a central ancestral population. Our genome-wide data confirm that - at present - malaria strains that are highly resistant to piperaquine remain sensitive to mefloquine, and vice versa. These findings suggest that ongoing trials to test triple combination therapies containing both piperaquine and mefloquine may, in the short term, be effective. However, the emergence of triple-therapy resistant *P. falciparum* remains a distinct threat.

3.2   Materials and Methods

3.2.1 Study population

We studied parasites collected from 78 adults with uncomplicated *P. falciparum* infection from three provinces of Western Cambodia: Oddar Meanchey in the North (n = 61), Battambang (n = 8) in the West and Kampot in the South (n = 9). These samples were collected between June 2011 and December 2013. Sixty of the samples in Oddar Meanchey came from an *in vivo* efficacy study of DHA-PPQ in which 54% of patients developed recrudescent parasitemia requiring rescue therapy with AS-MQ (clinical trial protocol WR1877 [NCT01849640]) [169]. An additional 18 *P. falciparum* clinical isolates from inside and outside of Oddar Meanchey Province were selected at random from an *in vitro* surveillance study (protocol WR1576), for which clinical follow-up data was not collected [170]. Both

42

studies were approved by the Cambodian National Ethics Committee for Health Research and the Walter Reed Army Institute of Research Institutional Review Board. The molecular studies described herein were approved by the University of North Carolina Institutional Review Board (14-0419). All subjects provided informed consent.

3.2.2 Drug susceptibility testing

Immediate *ex vivo* drug susceptibilities to piperaquine and mefloquine were measured using a histidine-rich protein 2 (HRP-2) enzyme-linked immunosorbent assay [171–173].

3.2.3 Whole-genome sequencing

Leukodepleted DNA from 91 *P. falciparum* infections was used for whole-genome sequencing. The ratio of parasite DNA to host DNA was determined using a quantitative PCR assay, and isolates with ≥20% *P. falciparum* DNA were sequenced on the HiSeq2000 Sequencing System (Illumina, San Diego, CA) [174]. Sequence reads were aligned to the *P. falciparum* 3D7 (v3) genome using the *bwa mem* algorithm. Isolates with 5-fold coverage at ≥60% of the genome were considered for variant calling, resulting in calls for 78 isolates. Variant calling was performed using the GATK *UnifiedGenotyper* utility.

3.2.4 Candidate markers for piperaquine, mefloquine and artemisinin resistance.

We confirmed specific candidate gene polymorphisms using alternate methods to assess the accuracy of whole-genome calls. We performed PCR and Sanger sequencing to detect microindel IIg in the *P. falciparum* multidrug resistance protein 2 gene (*pfmrp2*) and to genotype *pfmdr1* D1264Y [175–178]. We conducted evaluation of copy-number variation of candidate loci. This included digital droplet PCR to detect X5r domain amplification of chromosome five, previously associated with *in vitro* piperaquine resistance, and reported real time PCR data to assess *pfmdr1* copy number amplification, associated with mefloquine resistance (with CN≥1.5 considered to represent increased copy number) [170,179,180]. We then interrogated the WGS for additional polymorphisms associated with ACT resistance, including two other candidate markers of piperaquine resistance (*pfcrt* C350R and a repeat polymorphism in ABC transporter B family member 6), six recently described genetic background

markers common to artemisinin-resistant "founders" and the propeller domain of the *P. falciparum kelch* protein *K13* (a subset of which had been reported previously) [169,181–183].

3.2.5 Assessment of population structure and demography

Whole-genome variant data were used to determine the principal components of genetic variation. Cluster assignments were determined using a nonparametric *k*-means approach [184], and clusters were assigned a unique "Cambodian Parasite" (CP) identifier. Observed minor allele frequencies were calculated and compared to the frequencies expected from an idealized population, modeled using *ms* [185]. Pairwise F_{ST} for single-nucleotide polymorphisms between sub-populations was calculated and rendered in R (*R* CoreTeam, Vienna, Austria). The folded site-frequency spectrum was used to infer demographic parameters and likelihoods under a number of two-population models (split with isolation and split with migration) [186]. The clonality of each population cluster was rigorously tested by comparing point estimates of linkage disequilibrium decay with bootstrap replicates drawn randomly from all sequenced isolates.

3.2.6 Statistical Analysis

Median piperaquine and mefloquine IC_{50} values were compared between PCA clusters using the Kruskall-Wallis test. The Wilcoxon-Mann-Whitney test was used to compare the median piperaquine IC_{50} values of isolates with and without select candidate resistance markers. Fisher's exact test was used to compare the frequency of increased *pfmdr1* CN amongst recrudescent parasites and those that did not recrudesce. We analyzed 42-day cumulative risk of recrudescent malaria via survival-analysis with Kaplan-Meier curves and compared the risk by cluster using Cox proportional hazards models. Statistical analyses were completed using *SAS* 9.4 (SAS Institute, Cary, NC) and figures generated using *R*.

3.3   Results

We evaluated the IC_{50} values for piperaquine and mefloquine among our isolates and noted a trend of alternate resistance profiles among parasites (**Figure 3.1**). Evaluation of previously reported candidate piperaquine resistance loci showed no correlation with *ex vivo* resistance (data not shown). Mefloquine resistance was correlated to *pfmdr1* copy number increases, as previously reported [187].

44

Whole-genome sequencing produced, on average, greater than ten-fold coverage in 94.7% of coding regions and greater than five-fold coverage in 99.5% of coding regions. Variant calling with stringent filtering produced 7,228 high-quality SNPs genome-wide within this population. The allele-frequency spectrum showed overrepresentation of alleles at both minor and intermediate frequencies, suggesting underlying population structure and an expanding population (**Figure 3.2**). Using a principal components analysis (PCA) of the genetic variation amongst these samples, we summarized the underlying population structure. The first two principal components explained 52% of the variance and revealed four distinct population subclusters (**Figure 3.3 A-B**). Components beyond the third PC were less informative (**Figure 3.4 A-D**). *K*-means clustering using the top three principal components assigned individuals to four genetically distinct clusters (**Figure 3.5**). We found that dissimilarity was not explained by geography, and is therefore driven by a factor that supersedes genetic drift. All clusters except for one (named CP4) contained parasites from more than one geographic location (data not shown).

The pattern of opposing drug resistance profiles for piperaquine and mefloquine correlates with underlying genetic structure found by whole genome sequencing, a trend that is observed in geographically distant sites. Parasite isolates demonstrating higher *ex vivo* resistance to piperaquine were enriched in two clusters - CP3 and CP4 - while mefloquine-resistant parasites were enriched in one cluster - CP1. This is depicted in **Figure 3.3 A** and **Figure 3.3 B**, where nodes in the PCA represent individual clinical isolates and are sized according to their respective piperaquine and mefloquine IC_{50} values. Further comparative analysis of median piperaquine and mefloquine IC_{50} values by cluster shows different drug-resistance profiles among the four groups (**Figure 3.3 C-D**). The median piperaquine IC_{50} values were: 19.3nM (IQR 11.8 to 27.5) for CP1, 36.5nM (17.4 to 45.3) for CP2, 35.5nM (13.4 to 61.3) for CP3, and 57.7nM (29.3 to 80.2) for CP4 (Kruskal-Wallis, *p*=0.002). The median mefloquine IC_{50} values were: 143.0nM (IQR 53.9 to 173.9) for CP1, 9.1nM (1.2 to 63.7) for CP2, 47.8nM (23.0 to 62.7) for CP3, and 33.1nM (4.8 to 42.7) for CP4 (Kruskal-Wallis, *p*=0.004).

Pairing *in vivo* data with the PCA results, we found that patients failing DHA-PPQ therapy carried parasites that were disproportionately represented in certain clusters and would likely respond to AS-MQ salvage treatment. Among the 58 subjects followed for seven weeks after DHA-PPQ treatment, there were no recrudescence events in CP1 (n=11, 0%), two in CP2 (n=8, 25%), four in CP3 (n=16, 25%), and 18 in CP4 (n=23, 78%). Survival analysis demonstrated that subjects infected with parasites from cluster CP4 were more likely to suffer recrudescence after treatment with DHA-PPQ than those from other clusters (hazard ratio 7.9, 95% CI 3.0-20, p<0.001, **Figure 3.6**). All subjects with recurrent malaria were treated with AS-MQ per national guidelines [187,188]. After this rescue treatment for the initial recurrence, one subject had a second recurrent malaria infection during a limited follow-up period (median 15 days, range four to 33 days). Genotyping demonstrated the initial recurrence was not a DHA-PPQ failure. However, the initial recurrence was the same genotype as the second recurrence after AS-MQ rescue therapy, suggesting the patient was likely reinfected with a mefloquine-resistant strain. These data show no early treatment failures when AS-MQ was given for DHA-PPQ failure, but follow-up times were insufficient to rule out late treatment failures. Additionally, data from the original study showed only one of 35 (2.9%) recrudescence isolates had increased *pfmdr1* CN, whereas ten of 59 (16.9%) initial non-recurring isolates had increased CN (p=0.05) [187,188].

Parasites in CP2, the central cluster, displayed the least overall partner drug resistance (**Figure 3.3 C-D**) and displayed the highest genetic diversity. To explore the hypothesis that CP2 is a potential ancestral source population for the other groups, we fit population models based on demographic parameters inferred from the allele-frequency spectra within each cluster. Fitting two-population models to the CP2-CP1/CP4 cluster pairs demonstrated that CP2 closely resembles an ancestral population that gave rise to each of the other subpopulations (**Table 3.1**). CP3 was unsuitable for model fitting because of its extreme clonality. Analysis of linkage decay in each of the four populations revealed a high degree of linkage disequilibrium in these populations, indicating that CP1, CP3, and CP4 are clonal or near-clonal, suggesting epidemic expansion due to fitness advantages associated with drug resistance to commonly used antimalarials (**Figure 3.7**). The clonal nature and overall low genetic diversity within CP3 and CP4

is supported by the distribution of pairwise SNP differences between individuals within the subpopulations (**Figure 3.8**). In order to support the demographic models, we evaluated pairwise F_{ST} between population clusters. This showed that CP2 is more genetically similar to the outer groups (CP1, CP3 and CP4) than the outer groups are amongst themselves (mean F_{ST} of 0.41 vs. 0.87, respectively). Pairwise F_{ST} calculated for the outer groups revealed most polymorphic positions in the genomes to be either highly divergent or nearly identical (**Figure 3.9**). Overall, these findings support the hypothesis that CP1, CP3 and CP4 are clonal expansions derived from CP2, an admixed source population. However, the extremely low genetic diversity within the CP3 population does not allow us to directly assess the alternate hypothesis, that CP4 is not a direct expansion from CP2, but rather a second step in a CP2-CP3-CP4 demographic history.

To understand these findings in context of the recent and intensive work in Cambodian *P. falciparum* population genetics, we compared the genetic diversity in our clusters with the genetic architecture [38,189]. Miotto et al. identified five distinct compartments of *P. falciparum* genetic diversity in Cambodia [38]. Based on 7,214 SNPs common to both datasets, we determined that our populations fall within the scope of diversity described in 2013 (**Figure 3.10 A-C**), and that CP2 has a genetic signature most similar to the admixed KHA population (**Figure 3.10 D**).

Analysis of the distribution of candidate piperaquine, mefloquine, and artemisinin resistance loci across the four CP clusters highlights the clonal nature of CP1, CP3, and CP4 versus the heterogeneity found within CP2, while also showing that the clusters differentiate somewhat by kelch subtype, *pfmdr1* CN profile, and the presence or absence of a full suite of artemisinin-resistance background markers (**Figure 3.11**). CP2, the diverse source population, generally contains parasites with fewer artemisinin-resistance background markers; 15 of 21 (83%) possess at least one wild type allele, and five (28%) possess multiple wild type alleles. As expected, four of the five parasites with multiple wild type alleles remain wild type at the *kelch13* locus. In contrast, CP1, CP3, and CP4 parasites carry nearly the full set of artemisinin resistance background markers accompanied by kelch mutations. CP3 and CP4, the piperaquine-resistant clusters, contained parasites that generally harbored the *kelch13* C580Y mutation

and, by and large, did not display increased copy number at *pfmdr1* or contain the *pfmrp2* microindel.

Conversely, the mefloquine-resistant CP1 parasites were primarily *kelch13* R539T mutants, many

displayed increased *pfmdr1* CN, and all tested isolates carried the *pfmrp2* microindel II-g. None of the

samples differed in their genotypes at several candidate piperaquine resistance loci (*pfcrt* C350R, X5r

copy number, and *mdr6* repeat).

3.4   Discussion

3.4.1 Associations between population structure and partner-drug IC$_{50}$ in Cambodia

Our findings show that resistance to piperaquine and mefloquine has developed on distinct

genetic backgrounds that are not explained by geography. These findings have important clinical and

public health implications. They help explain the phenomenon that AS-MQ rescue therapy is successful

in individuals who fail DHA-PPQ treatment, and they suggest that, in the short term, AS-MQ will remain

effective in the setting of DHA-PPQ failure.

Currently, trials of triple artemisinin combination therapies (TACTs) are underway in Cambodia

[190]. These are predicated upon evidence that the mechanisms of MQ and PPQ resistance may be

antagonistic. Mefloquine resistance has been shown to sensitize parasites to piperaquine *in vitro*, which

mirrors historical observations of a reciprocal relationship between chloroquine (a relation of piperaquine)

and mefloquine resistance [191–193]. Similarly, artesunate-amodiaquine (ASAQ) and artemether-

lumefantrine (AL), regimens commonly used in Africa, have been found to exert opposing selective

effects on parasite polymorphisms in the *pfmdr1* and *pfcrt* genes [194,195]. Our findings are largely

consistent with these prior observations. Specifically, the most mefloquine-resistant subgroup (CP1) was

the most piperaquine sensitive, the most piperaquine-resistant subgroup (CP4) was the most mefloquine

sensitive, and no parasites were strongly resistant to both mefloquine and piperaquine. Moreover, recent

studies in western Cambodia have shown declining *pfmdr1* copy number amplification as piperaquine

resistance emerged [196–198]. Our study suggests that the loss of *pfmdr1* duplications is not the result of

copy-number deamplification, acquisition of piperaquine resistance, and then expansion of the same

parasite clones, but rather expansion of very different genetic clones associated with piperaquine

resistance. These lines of evidence suggest that resistant parasites in Cambodia tend to represent multidrug resistant lines (based on *ex vivo* resistance data to other antimalarials not reported here), resistant to either piperaquine of mefloquine, rather than extensively drug resistant lines resistant to both partner drugs. This would mean that TACT would target non-overlapping drug-resistant populations, likely increasing the overall efficacy of initial therapy for malaria.

Despite this evidence in support of TACT trials, our data present some unsettling findings. First, several isolates remain sensitive to both drugs, raising the concern that resistance to both partner drugs may emerge within these isolates by a novel mechanism if TACTs are implemented. Secondly, rare parasites among CP3 and CP4 groups had increased *pfmdr1* copy number - a genetic marker of MQ resistance. These findings could be of no import to public health, or they could be a harbinger of triple-therapy resistant *P. falciparum*.

Whole-genome sequencing allowed us to appreciate clinically relevant substructure and population histories. Our findings suggest that the partner-drug resistant subpopulations (CP1, 3, and 4) have undergone, or are currently undergoing an epidemic expansion. These subgroups harbored nearly the full set of artemisinin-resistance background markers, were uniformly *kelch13* mutants, and demonstrated increased IC_{50} to either mefloquine or piperaquine. While the CP3 and CP4 populations overlap previously described subpopulations in Cambodia (**Figure 3.10**), a higher percentage of our isolates appear in these CP3 and CP4 populations compared to the previous studies, suggesting a recent expansion. Within-subpopulation pairwise genetic diversity (defined as the number of SNPs between the isolates) is much lower for CP1, 3, and 4 as compared to the CP2 subpopulation (**Figure 3.8**), causing elevated linkage disequilibrium within these subpopulations (**Figure 3.7**). Demographic modeling brings further clarity to the history of these subpopulations. Our models suggest that CP2 acted as a source from which CP1 and CP4 split and underwent clonal expansion (the low diversity in CP3 precluded these models). Consistent with epidemiological data that MQ-resistance arose in Cambodia prior to PPQ-resistance [199,200], we found that the MQ-resistant CP1 subpopulation separated from CP2 prior to the separation of the PPQ-resistant CP4 subpopulation (**Table 3.1**).

A concerning finding was that the expanding CP4 subpopulation had a strikingly high recrudescence rate after DHA-PPQ therapy. It is unknown what genetic drivers may cause this phenotypic difference between CP4 and CP3, which also had overall elevated DHA-PPQ IC_{50} values. F_{ST} comparisons between these two populations showed large genomic regions of near perfect homology contrasting with regions of high divergence. These findings support - but cannot pinpoint - yet-undiscovered genetic variation that confers increased parasite fitness to CP4 parasites. Additional studies involving larger numbers of well characterized samples with elevated piperaquine IC_{50} will help elucidate the mechanistic differences between CP3 and CP4 parasites, and have the potential to uncover elusive genetic markers of piperaquine resistance and clinical failure. As our samples were collected during the early emergence of piperaquine resistance in Cambodia with modest increases in IC_{50}, ongoing sample collection and genomic surveillance will be crucial for understanding how these subpopulations continue to evolve.

It is clear that artemisinin resistance has had profound effect on the population structure in Cambodia [201,202] (for example, the high prevalence of R539T in CP1 and of C580Y in CP3 and 4). Unfortunately, *ex vivo* artemisinin phenotype data by the ring-stage survival assay [203] are unavailable for these samples, precluding *ex vivo* evidence regarding the extent of artemisinin resistance in these subpopulations. Overall, only five of the 78 samples that underwent whole-genome sequencing in our study did not harbor *kelch13* mutations, but these five samples still harbored multiple mutant artemisinin resistance background markers believed to be conducive to the acquisition of a *kelch13* mutation (**Figure 3.11**) [201]. This is consistent with previous population genetic studies showing that *P. falciparum* parasites in the region have undergone an extreme, recent genetic bottleneck due to artemisinin selection pressure that has resulted in groups of highly differentiated clonal subpopulations [201,204] . The fact that these background markers are largely fixed in our cohort does not bode well for the therapeutic lifespan of future partner drugs used in ACT in the region. As many have remarked, ACT in its current form (two drugs for three days) may no longer be a viable strategy in Southeast Asia [199,205].

50

In conclusion, we used a population genetics approach to provide insight into the differing genetic backgrounds upon which piperaquine and mefloquine resistance occur in Cambodia. Although our study was limited to a single country, the findings and methods are likely applicable to other countries in Asia where ACT resistance is actively spreading [204]. As reports of ACT failure become more commonplace, further study of partner-drug resistance throughout the region and evaluation of novel ACT regimens is imperative.

3.4.2 Data Availability

Sequencing reads were deposited at the Sequence Read Archive [SRP063498]. Sequences for *pfmdr1* [accessions KT833623-740] and *pfmrp2* [KT876268-381] amplicons were deposited at GenBank.

3.4.3 Author Contributions

CMP, JBP, JTL, CL, DLS and JJJ designed the study. CMP, JBP, SB, JTL, JJJ, CAL, CL, DLS, SC, and PG collected the data. CMP, JBP, SB, SC, PG, EB, CAL, CL, DLS, JTL, and JJJ did experiments and analyzed the data. CMP, JBP, JTL, DLS and JJJ interpreted the data and prepared the report. JTL, JJJ, CL and DLS oversaw the project.

3.4.4 Funding

Data	Model	θ	s	η_{CP2}	η_X	$m_{CP2 \rightarrow X}$	$m_{X \rightarrow CP2}$	T_{SPLIT}	Log Likelihood	AIC[†]
CP2-CP1	Static	382.98	-	-	-	-	-	-	-3202.43	6406.86
CP2-CP1	Iso	388.13	0.866	18.13	0.05	-	-	0.074	-340.90	691.80
CP2-CP1	IM	367.55	0.999	5.45	0.051	0.165	0.129	8.531	-251.80	517.60
CP2-CP3	Static	*	*	*	*	*	*	*	*	*
CP2-CP3	Iso	*	*	*	*	*	*	*	*	*
CP2-CP3	IM	*	*	*	*	*	*	*	*	*
CP2-CP4	Static	363.92	-	-	-	-	-	-	-3717.13	7436.26
CP2-CP4	Iso	365.19	0.947	18.11	0.053	-	-	0.092	-231.71	473.42
CP2-CP4	IM	374.68	0.185	4.47	0.052	0.339	2.61	0.023	-218.08	450.16

Table 3.1 Inferred parameters for demographic models.

Best-fit inferred parameter estimates for alternative two-population models fit between CP2 and CP1/CP3/CP4. Static models no split and no growth (positive or negative). Iso models a split with subsequent fluctuating population sizes but without migration. IM models a split with subsequent fluctuating population sizes and migration. The parameter s is the proportion of an ancestral combined population that gave rise to CP2, while $1 - s$ is the proportion of the ancestral population that gave rise to either CP1/CP3/CP4. η_1 and η_2 are the effective population size of CP2 and CP1/CP3/CP4, respectively. $m_{1 \rightarrow 2}$ is the migration rate from CP2 to CP1/CP3/CP4. $m_{2 \rightarrow 1}$ is the migration rate from CP1/CP3/CP4 to CP2. T_{SPLIT} is the time since the two populations split, measured in θ generations. θ is the population mutation rate. [†] AIC: Akaike Information Criterion; * Population CP3 was not amenable to model fitting because of its high clonality.

**Figure 3.1 Piperaquine and mefloquine IC$_{50}$ values associate
with parasite subpopulations in Cambodia.**

Subjects who had both piperaquine and mefloquine IC$_{50}$ values are displayed. Those who later had an episode of recrudescence are displayed using solid orange triangles, while those who did not have an episode of recrudescence or did not have follow-up data available are displayed using solid navy circles. Two isolates that had recrudescence with very high piperaquine IC$_{50}$ values (307.0 and 397.1nM) are represented with hollow symbols.

Figure 3.2 Allele frequency spectrum demonstrates an excess of low-frequency and intermediate-frequency alleles.

A folded allele frequency spectrum for synonymous (red), nonsynonymous (orange), genic (light blue) and intragenic variants (dark blue) regions are shown for the 78 *Plasmodium falciparum* isolates. The population has an excess of low frequency alleles, similar to other *P. falciparum* populations, and suggestive of a population expansion. However, an excess of intermediate frequency alleles is also observed when compared to expected values under a neutral evolution model, suggesting substructuring.
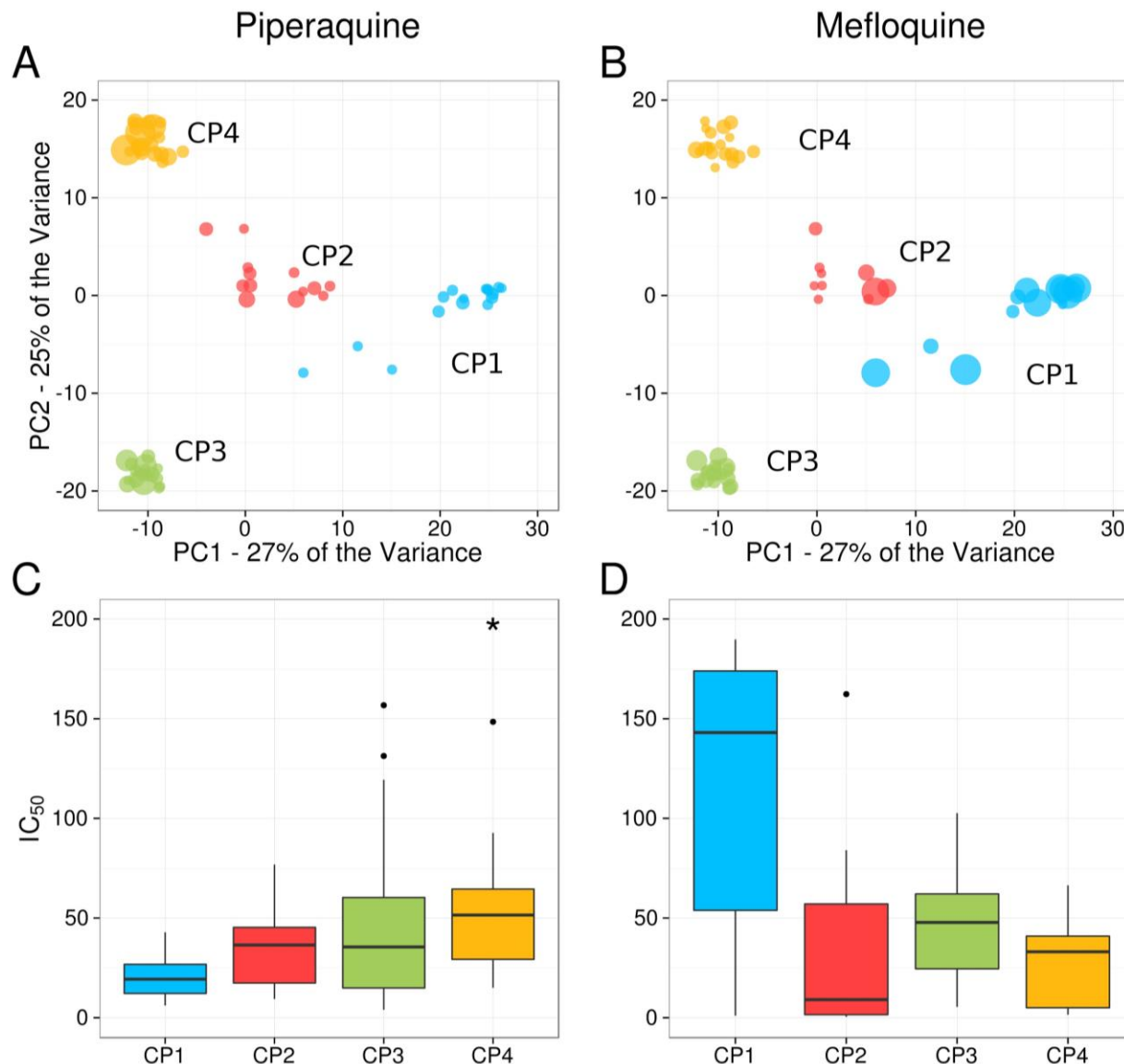
Figure 3.3 Principal components analysis demonstrates distinct clusters with non-overlapping resistance profiles.

In the upper panels, nodes are scaled according to IC_{50} values, with the largest nodes representing IC_{50} values of 200nM or greater. Panel **(A)** displays piperaquine IC_{50} values, and panel **(B)** displays mefloquine IC_{50} values. Blue samples represent with cluster CP1, red with CP2, green with CP3 and yellow with CP4. In the lower panels, IC_{50} values by cluster are depicted for piperaquine in panel **(C)** and mefloquine in panel **(D)**. Boxplots display the median (horizontal line), interquartile range (box), and the lowest and highest data point within 1.5 times the IQR (whiskers) of IC_{50} values [values outside this range are plotted as individual points]. Two outliers with piperaquine IC_{50} values of 307.0 and 397.1nM were clipped from cluster CP4 in panel **(C)** and labeled with an asterisk (*). There was a statistically significant difference among the four groups for piperaquine (*p*=0.002) and for mefloquine (*p*=0.004) IC_{50} values. Noise was added to reduce overplotting. Abbreviations: PC, principal component; IC_{50}, half-maximal inhibitory concentrations.
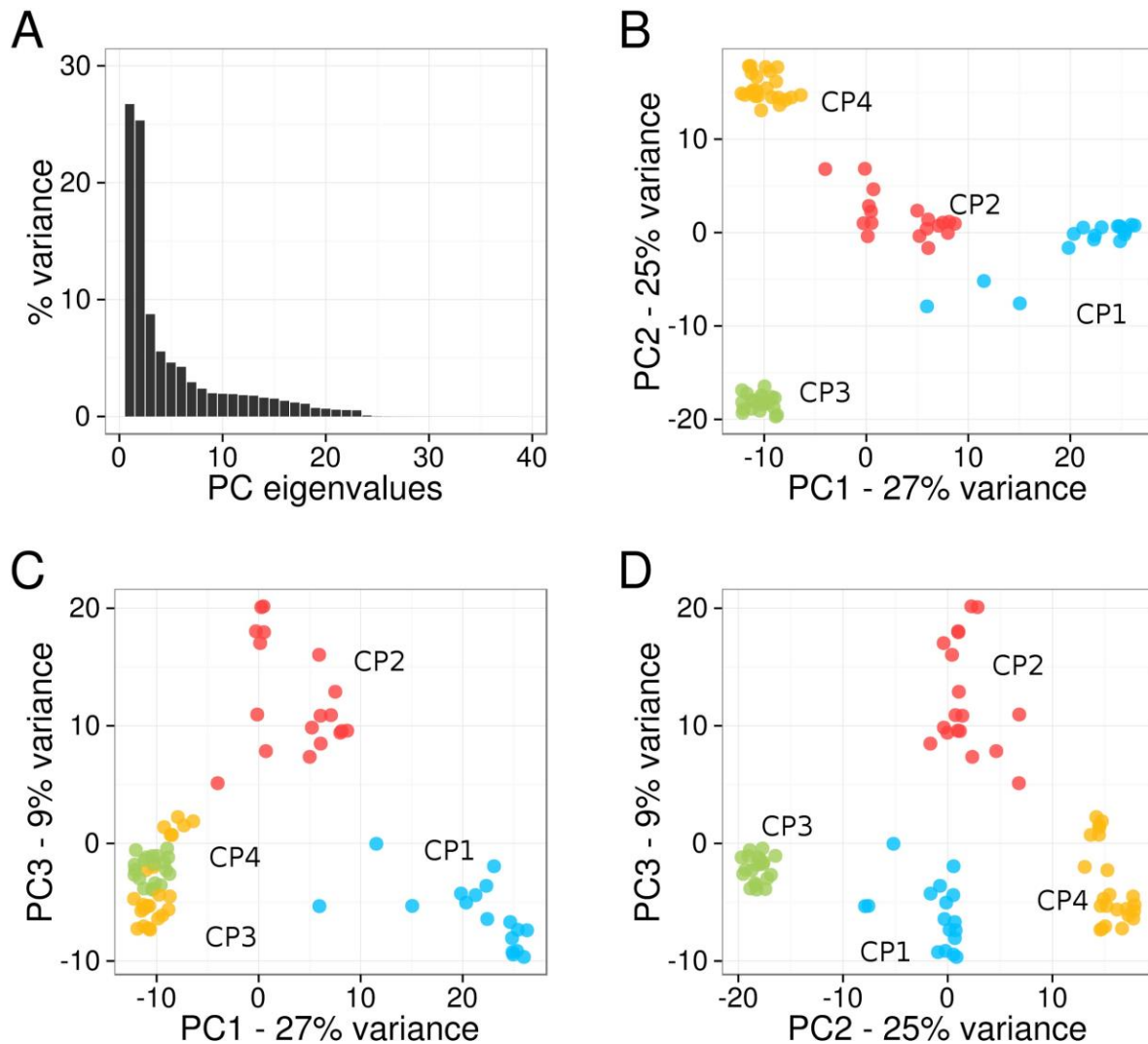
Figure 3.4 Eigenanalysis demonstrates that the majority of genetic diversity is described by the top three principal components.

Eigenvalues are presented in **(A)**, demonstrating that the top three components describe over 60% of genetic diversity in this population. **(B, C, D)** show genetic differentiation among sequenced isolates, as described by the top principal components. In each view, clear separation is seen between subpopulations.
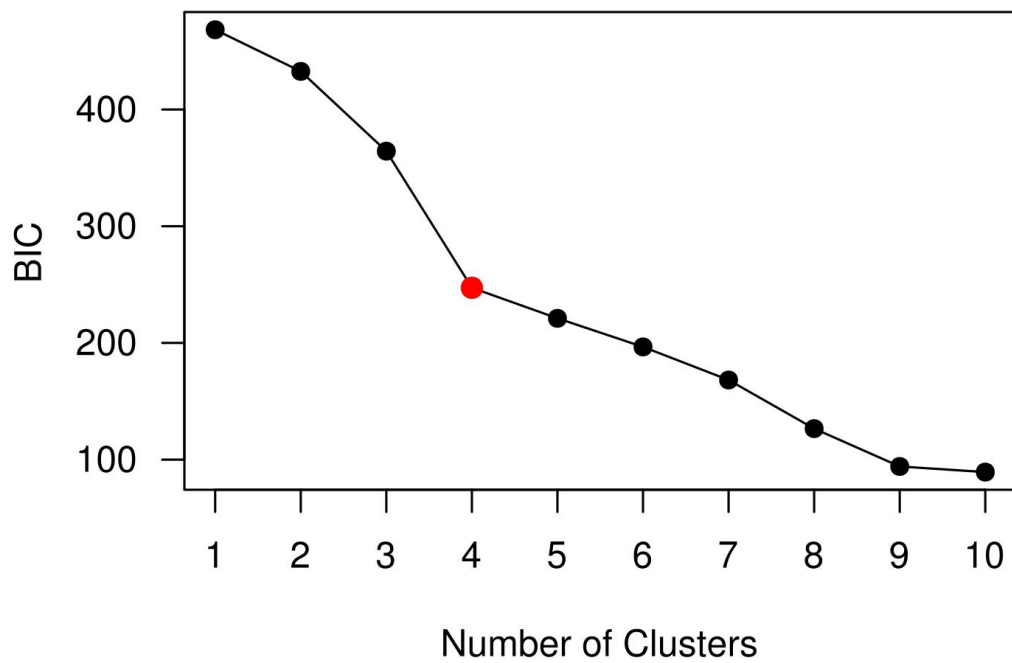
Figure 3.5 *K*-means clustering identifies four genetic clusters.

Increasing the number of clusters from n=3 to n=4 yields a substantial improvement in Bayesian information criterion (BIC). While subsequent increases do yield BIC improvements, the improvement in goodness of fit is not commensurate with the cost associated with increased degrees of freedom.
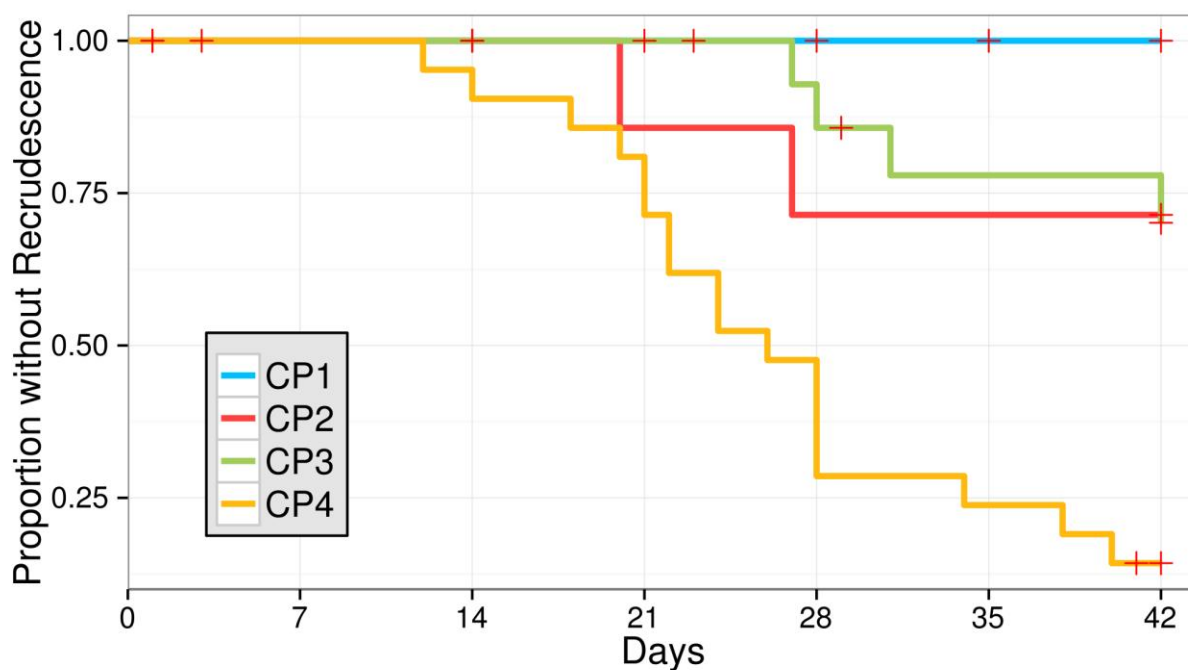
Figure 3.6 Recrudescence after DHA-PPQ therapy is associated with the CP4 cluster.

Kaplan-Meier curves demonstrate the risk of recrudescence by cluster (Wald $p<0.001$ for CP4 compared to other clusters). Red crosses represent patients censored for withdrawal or development of new infection.
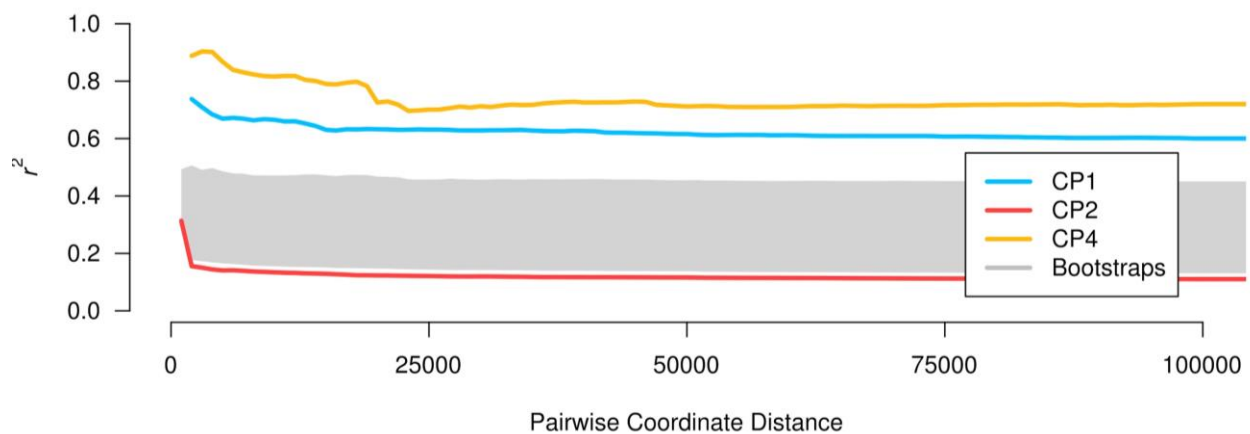
Figure 3.7 Parasites in drug-resistant subpopulations have extended linkage disequilibrium (LD), supporting clonal expansion.

The extent of clonal expansion among the *P. falciparum* clusters was determined by assessing linkage decay (r^2). Point estimates of linkage decay for each of the clusters were calculated pairwise between variants up to 200,000 bases apart, and are shown by colored lines. CP3 is absent because its extreme clonality prevented calculation of r^2. To determine the significance of these results, a null distribution of linkage decay was calculated from 100 bootstrapped subpopulations, drawn randomly with replacement from the 78 *P. falciparum* samples.
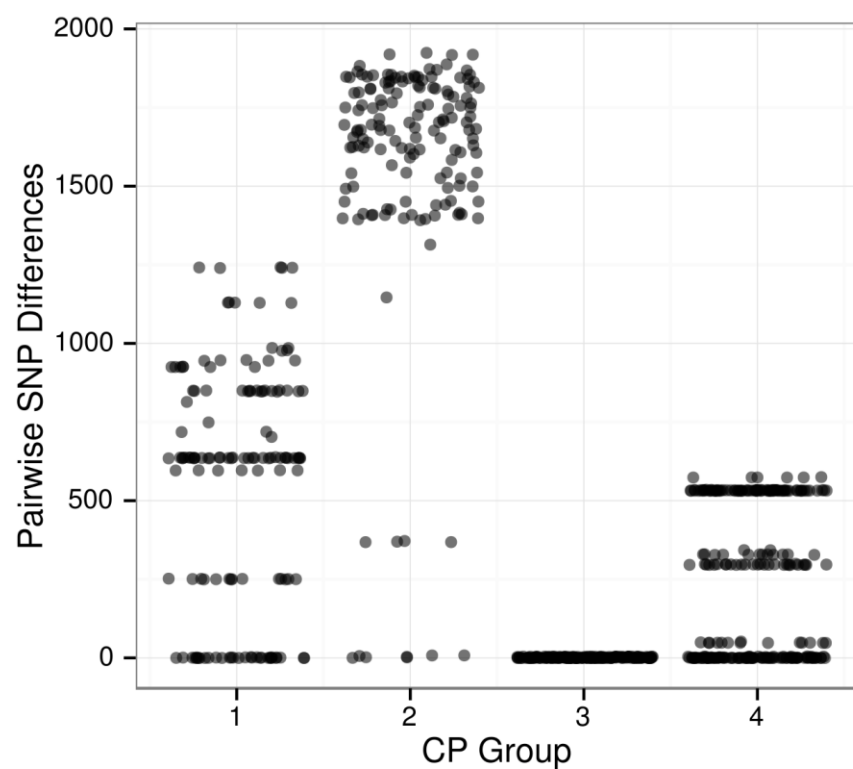
Figure 3.8 Parasites in drug-resistant subpopulations have decreased pairwise SNP differences, consistent with clonal expansion.

For each CP group, the number of pairwise SNP differences was calculated and plotted. CP3 and CP4 have substantially lower numbers of pairwise SNP differences between isolates, consistent with the hypothesis that these are clonally expanding subpopulations.
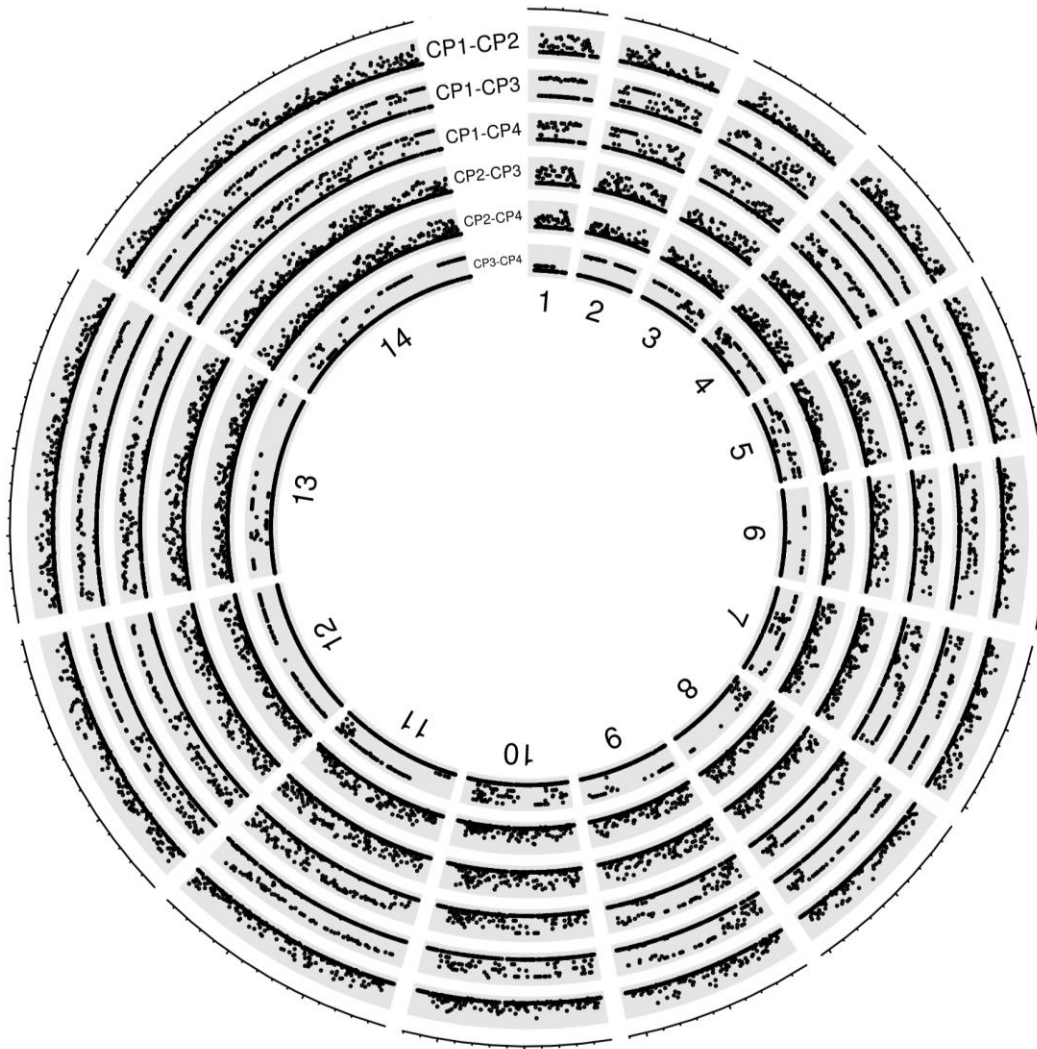
Figure 3.9 Pairwise F_{ST} comparison of four *P. falciparum* subpopulations pinpoints identical and divergent genomic regions between subpopulations.

Each circular track represents the pairwise F_{ST} comparison for all genes (dots) between the reported CP populations. Each segment of the track represents a chromosome, which are numbered in the center of the figure. Within each track, the lower and upper y-axis limits are zero and one, respectively. F_{ST} values at or approaching zero indicate that a locus is genetically similar between the two populations, and F_{ST} values at or approaching one indicate that a locus is genetically dissimilar between the two populations. In particular, the comparison between CP3 and CP4 (the inner-most track) highlights identical and divergent genomic regions, providing clues for the genetic determinants of piperaquine resistance and clinical outcomes.
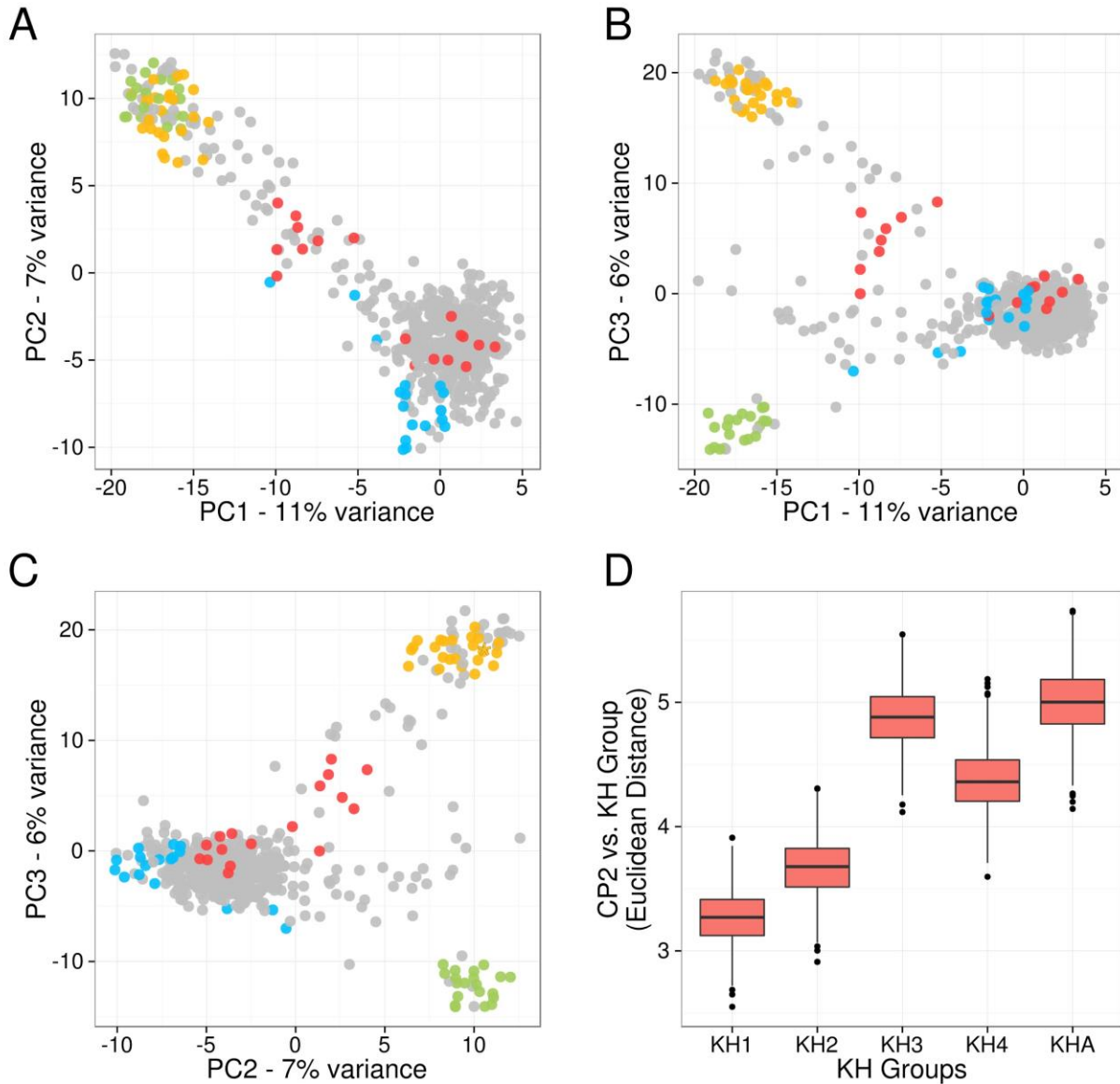
Figure 3.10 Comparisons of *P. falciparum* clusters with previously described Cambodian subpopulations.

Panel **(A-C)** displays projections for the top three principal components for analysis of 78 *P. falciparum* isolates sequenced in this study (CP1 as blue, CP2 as red, CP3 as green, CP4 as yellow), compared to 293 isolates previously described Cambodian isolates (gray) from Pailin, Pursat, Ratanakiri, and Tasanh (Miotto *et al.*, 2013). The scope of genetic diversity described in the present study falls within the range previously described. Panel **(D)** displays comparisons of allele-frequencies between CP2 (the founder population described in the present study) with previously described Cambodian subpopulations (KH#). Of the 1886 subpopulation-defining variants described by Miotto et al. in 2013, 619 of these were also used in this study as markers of genetic diversity, and the allele frequency at these markers were used for matching. Comparisons were bootstrapped 1000 times, and the boxplots display the median (horizontal line), IQR (box), and 1.5 times the IQR (whiskers). CP2 is most genetically similar to the KHA subgroup. Noise has been added to the PCA to reduce overplotting.
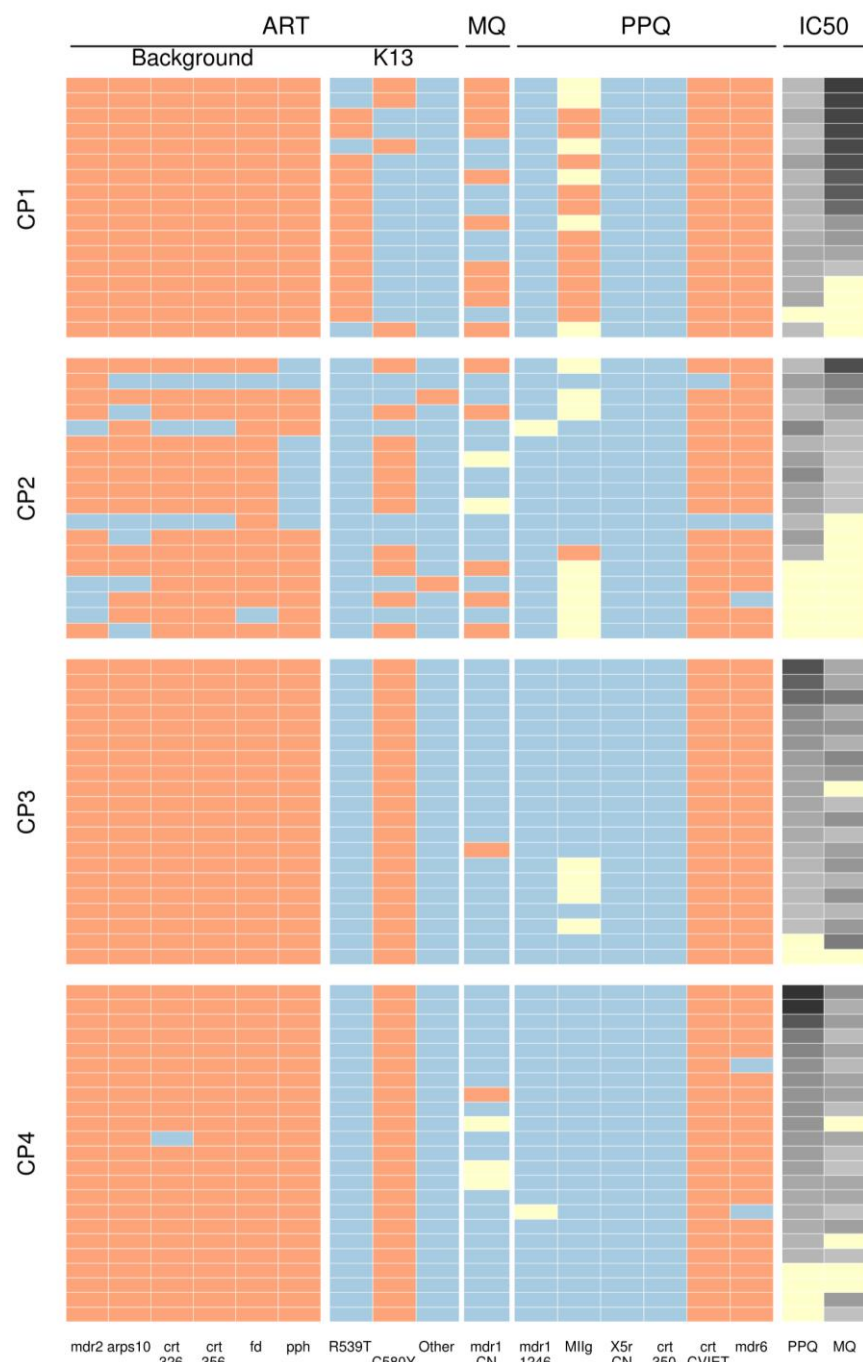
Figure 3.11 Haplotype map describing associations between piperaquine and mefloquine IC$_{50}$ values and genotypes of known and candidate markers of resistance.

IC$_{50}$ values are displayed in gray scale, with lighter shades indicating lower IC$_{50}$ values and black indicating samples with values ≥200nM. Additional columns code data, with blue blocks indicating wild-type (3D7) genotype, red blocks indicating the presence of the mutation of interest, and yellow blocks indicating missing data. We also included both known and candidate markers for artemisinin (including recently described candidate background resistance markers), mefloquine, and piperaquine.

CHAPTER 4: DIFFERING MECHANISMS OF POPULATION-LEVEL RESILIENCE FOR CO-ENDEMIC *PLASMODIUM VIVAX* AND *PLASMODIUM FALCIPARUM*[3]

4.1   Introduction

Western Cambodia, which is the site of emerging artemisinin-resistant *Plasmodium falciparum* [206,207], has been the focus of intense multimodal malaria-control interventions. These interventions, which were designed to contain artemisinin resistance in *P. falciparum*, have included increased vector control, increased surveillance, and improved access to quality artemisinin-combination therapy (ACT) [208]. Although these interventions were intended to curtail *P. falciparum* malaria, enhanced diagnosis and treatment and use of insecticide treated bednets are also effective against *P. vivax* malaria. *P. vivax* and *P. falciparum* affect the same populations and are transmitted by the same mosquito species in this region [209]. As a result of containment efforts, between 2009 to 2013 there was an 81% decrease in total malaria infections. While there has been a marked decline in *P. falciparum* cases, *P. vivax* cases have increased and are now more common than *P. falciparum* [210]. *P. vivax* has proven its resilience throughout the Mekong region as it has become the predominant species [211–213]. Despite the failure of these interventions to contain artemisinin resistant *P. falciparum*, they have created a natural experiment in which the genomic responses of sympatric *P. falciparum* and *P. vivax* populations can be studied. This contrast in outcomes for the Cambodian *P. vivax* and *P. falciparum* populations raises a question - what are the genetic signals that correlate with population-level resilience for these two species, and how do they differ.

---

[3] This chapter is currently in preparation for journal submission. Current citation: Parobek CM, Saunders DL, Barnett EJ, Lon C, Lanteri CA, Balasubramanian S, Brazeau NF, DeConti DK, Garba DL, Spring MD, Chour CM, Bailey JA, Lin JT*, Juliano JJ. Differing mechanisms of population-level resilience for co-endemic *Plasmodium vivax* and *Plasmodium falciparum*.

Because both *P. falciparum* and co-endemic *P. vivax* are affected by many control interventions, studies inclusive of *P. vivax* are needed. Sympatric populations provide a vehicle for understanding differences in transmission dynamics, population structure, and life history. Importantly, they provide a means for understanding how these species respond differently to control interventions. They may also provide a proving ground for policies that address the elimination of both *P. vivax* and *P. falciparum*. The few published population genetics studies of sympatric *P. vivax* and *P. falciparum* populations have revealed key insights into the differences between these two species. Microsatellite studies conducted in the Solomon Islands, Papua New Guinea, Indonesia, and Cambodia have painted a consistent picture of higher genetic diversity in *P. vivax* populations than in sympatric *P. falciparum* populations [214–217]. Additionally, these reports detail that *P. falciparum* populations tend to be clonal with episodic transmission and structure-by-geography, while *P. vivax* populations enjoy a higher effective population size, more stable transmission, and increased gene flow between geographic islands. Studies of *ama1* in sympatric populations uncovered similar trends [218,219].

These studies of sympatric parasite populations spawn as many questions as they resolve. Specifically, it is unclear why this pattern of higher *P. vivax* diversity and more consistent transmission are repeated across continents and time. On a basic level, we need to know whether these results are driven by species-specific differences in microsatellite and antigen mutation rates, and if they hold up at a whole-genome level. If these differences are real, are they a result of control programs doing a poorer job at attenuating *P. vivax* transmission (for example because transmission can occur at lower parasitemias and even before clinical disease is recognized, or because there is a reservoir of genetic diversity in the dormant hypnozoites among the population), or do *P. vivax* parasites simply respond differently to similar assaults. If mechanistic differences do account for these differences, what additional information is needed before these insights can be translated into intervention? The answers to these questions may aid in designing elimination programs that address both *P. vivax* and *P. falciparum* effectively.

Genome-by-population studies of sympatric *P. vivax* and *P. falciparum* isolates can help answer these questions. Importantly, whole-genome studies can peer into demographic history and also

comprehensively search for loci subject to selective pressures. The population genetics of the *P. falciparum* parasites in artemisinin-resistance-era Cambodia have been carefully studied [220–222]; however, an in-depth genomic study of *P. vivax* in Cambodia during the resistance-containment campaign is lacking.

To understand the genome-wide species-specific patterns of selection in sympatric *P. vivax* and *P. falciparum* populations in Cambodia, we conducted whole-genome sequencing of co-endemic parasites sampled from a primary site and two nearby sites. These sites were designated as "Zone 2" during the containment campaign, and as such were subject to comprehensive and multimodal malaria control efforts [223]. Using these data, we compared *P. vivax* and *P. falciparum* population structure, polyclonality, and selection signatures to uncover differences in how these two species respond to malaria control and elimination efforts.

4.2   Materials and Methods

4.2.1 Sample Collection

Clinical isolates were collected from 2009-2013 by the Armed Forces Research Institute of Medical Sciences in three Cambodian provinces - Oddar Meanchey (northwest), Battambang (west), or Kampot (southwest). Uncomplicated *P. vivax* or *P. falciparum* malaria patients presenting to study-site clinics gave written informed consent for their participation in this study. Study staff collected and leukodepleted venous blood, and administered treatment in accordance with Cambodian National Malaria Control Program guidelines. During this period, both *P. falciparum* malaria monoinfections and mixed *P. falciparum* and *P. vivax* malaria infections were treated with dihydroartemisnin-piperaquine (DHA-PPQ). While chloroquine (CQ) was the primary antimalarial used for *P. vivax* monoinfection during this period, the concerns for chloroquine resistance in Cambodia led to a change to DHA-PPQ as the treatment of choice for vivax malaria in 2012. Molecular studies were approved by the IRB at the University of North Carolina, some other IRB, and the Cambodian National Ethical Committee for Health Research.

P. vivax and *P. falciparum* sample sequencing. For *P. vivax*, to remove contaminating host DNA, whole blood was leukodepleted using Plasmodipur™ filters (Euro-Diagnostica®, Malmö, Sweden). The

66

ratio of parasite-DNA to host-DNA was determined using a qPCR assay [224], and isolates with ≥20% *P. vivax* DNA were considered for sequencing. Clinical isolates with high plasmodium:human DNA content were sequenced on the HiSeq2000™ or HiSeq2500™ using 100 or 125 base-pair paired-end chemistry (Illumina, San Diego, CA). Data are made available at the Sequence Read Archive, and accessions are listed in **Table 4.6**. For *P. falciparum*, data from previously reported isolated were used and reanalyzed in this study to allow for comparable analysis methods between species. More stringent analysis filters altered the number of variants called compared to previous reports (Parobek *et al.*, *in submission*).

4.2.2 Sequence Analysis

Sequence reads were aligned to the *P. falciparum* 3D7 (v3) and *P. vivax* Sal1 (v3) genomes using *bwa mem* [225], which allows for a hybrid end-to-end and local alignment approach. To increase sensitivity through hypervariable regions, we increased the clip penalty for local alignment to L=15, which was the minimum value that allowed for with default sensitivity. PCR and optical duplicates were removed from alignments using the Picard Tools *MarkDuplicates* utility [226] and local realignment of highly entropic regions was performed using the GATK *IndelRealigner* utility. Isolates with five-fold coverage at ≥80% of the genome in the case of *P. vivax*, and ≥60% of the genome in the case of *P. falciparum*, were considered for variant calling and further analyses.

Variant calling for *P. falciparum* was performed as described previously (Parobek *et al.*, *in submission*). Variants were called for *P. vivax* samples jointly using the GATK *UnifiedGenotyper* [227,228]. Variants were filtered stringently using cutoffs responsive to the underlying distribution of quality scores: Quality-by-Depth (QD) ≥ 25, Mapping Quality (MQ) ≥ 55, Fisher Score (FS) ≤ 10, Map-Quality Rank Sum ≥ -5.0, and Read Position Rank Sum ≥ -5.0, extreme Filtered Depth (DP ≤ 10th and ≥ 90th percentile), and an overall quality (QUAL) below the log-scaled inflection point. In addition, only biallelic variant records with at least 5-fold coverage in 100% of isolates were considered. Variants in low-complexity regions were identified using *Tandem Repeat Finder* [229] and excluded. Members of highly paralogous gene families [230] were excluded. Furthermore, we excluded 60000 bases of sequence space from the beginning and end of each chromosome.

4.2.3 Tests for Within-Host Diversity

To test whether infections were multiclonal or polyclonal, we used the F_{WS} statistic [231,232]. To enable direct F_{WS} comparisons between *P. falciparum* and *P. vivax* isolates, which had different numbers of loci, we bootstrapped each calculation 1000 times, selecting 5000 variable sites for each isolate and each replicate. The maximum and minimum F_{WS} bootstrap values were identified to provide a generous upper and lower confidence interval for each F_{WS} whole-data point estimate. Additionally, *P. vivax* isolates were screened for multiplicity of infection (MOI) using ultra-deep sequencing of the highly polymorphic *pvmsp1* locus [233], which is more sensitive to complex infections than microsatellite fragment analysis [234]. Deep-sequencing of the *pvmsp1* locus was performed using the Ion Torrent platform, and the reads for each individual were clustered using *SeekDeep*, an iterative *de novo* clustering algorithm (Hathaway, *in submission*; github.com/bailey-lab/SeekDeep). Experimental procedures were performed as previously described [234]. For the purposes of the present study, a sample was deemed multiclonal if a minor clone existed at ≥ 10% read frequency.

4.2.4 Tests for Population Structure

We determined population substructuring using principal component analysis (PCA) and cluster assignments were determined using a nonparametric *k*-means approach. PCA was calculated using *adegenet* [184].

4.2.5 Demographic Inference

Five one-population demographic scenarios were fit to the observed site-frequency spectra at synonymous sites in both the *P. vivax* and *P. falciparum* populations [235]. First, a model of constant population size was fit in which there is no increase or decline in the effective population size over time. Second, a model of population decline was fit, in which the effective population size began to decrease at some time, *T*, in the past. Third and fourth, exponential and two-epoch growth models were fit, in which the effective population size began either exponential growth or a sudden size change at *T*. Finally, a "bottleneck" scenario of sudden decline followed by exponential growth was fit. For each model, the following parameter space was explored as applicable: time (*T*, 0.01 - 5.0), the factor of population

contraction (η_D, 0.001 - 1.0) and expansion (η_G, 1 - 100), and the ancestral θ. One hundred independent runs were performed for each model and each dataset. Inter-run parameter values were compared to assess model convergence, and the run with the highest log-likelihood was selected.

4.2.6 Detecting Patterns of Selection

Inferred parameters for the best-fit one-population demographic scenario for both *P. falciparum* and *P. vivax* were used to parameterize coalescent simulations. Each gene in the genome (excepting genes from highly paralogous families and in chromosomal telomeres) was simulated 1000 times using the *R* package *coala* [236], a wrapper for *ms* [237]. The resulting dataset represented a model population with the same demographic history as our Cambodian *P. vivax* and *P. falciparum* populations, but provided a null distribution for population genetic tests. From this resource, we determined a null distribution for Tajima's *D* values for both *P. vivax* and *P. falciparum* to aid in identifying genes under unexpectedly strong balancing or directional selection.

4.2.7 Copy-Number Analysis

Due to the high AT content of the *P. vivax* and *P. falciparum* genomes, we identified copy-number variants using a tailored approach for identifying segmental duplications in genomes with high AT content (DeConti, *in submission*). Predicted copy-number variants were confirmed using a probabilistic framework for structural variant discovery [238], and a subset of *P. falciparum mdr1* copy number variants were confirmed using qPCR as described previously (Parobek *et al.*, *in submission*).

4.2.8 Scans for Positive Selection

Because there is as yet no fine-scale map of recombination for the *P. vivax* genome, we initially sought a map-independent haplotype-based approach. nS_L, a modification of iHS, obviates the need for a genetic map, reduces its dependence on recombination and demographic events [239], and may afford increased sensitivity to detect soft selective sweeps. This statistic has proved sensitive in identifying selection in other non-model organisms [240]. As a secondary test, we constructed a genetic recombination map for both *P. falciparum* and *P. vivax* using *LDhat interval* [241] and performed the iHS haplotype-based test for directional selection [242]. iHS was calculated using iHH_0/iHH_1, where

69

subscripts 0 and 1 arbitrarily denote alleles, agnostic to ancestral or derived status. For both nS_L and iHS, because of the lack of outgroup sequences, we ignored the sign. We plotted iHS and nSL scores that were normalized according to allele frequency bins. Extended haplotype homozygosity (EHH) was calculated for the selected and unselected allele. These three haplotype-based tests for selection were performed using the program *selscan* [243].

4.3   Results

4.3.1 Sequencing sympatric P. vivax and P. falciparum populations

We generated whole-genome short sequence reads for 78 Cambodian *P. vivax* field isolates and 93 *P. falciparum* field isolates, and aligned these sequence data to the Sal1 and 3D7 reference genomes, respectively. Among *P. vivax* isolates, 70 (90%) had five-fold or greater coverage across at least 80% of the genome, while among *P. falciparum* isolates, 80 (86%) had five-fold or greater coverage across at least 60% of the genome. Beyond tandem repeats, we excluded from analysis roughly 4 Mb of sequence space from both the *P. vivax* and *P. falciparum* genomes, containing approximately 300 and 400 genes, respectively. These were primarily subtelomeric regions containing members of highly paralogous gene-families. In total we identified 61,448 high-quality *P. vivax* SNPs and 6,734 *P. falciparum* SNPs.

4.3.2 P. vivax infections are more complex than P. falciparum

Because *Plasmodia* infections are frequently multiclonal, we investigated the extent of multiclonality among our sequenced field isolates. *P. vivax* infections were more polyclonal (defined as $F_{WS} < 0.95$) than *P. falciparum* infections ($p<0.0001$, Fisher's exact), a finding which remained unchanged after bootstrapping to account for the difference in the number of variants identified (**Figure 4.1**). To confirm the extreme polyclonality predicted for *P. vivax* by F_{WS}, we deep sequenced the hypervariable region of the 42kda domain of the merozoite surface protein 1 (*pvmsp1*) (identified in [244] and used in [245]). Amplicon deep sequencing confirmed the highly multiclonal nature of *P. vivax* infections: of the 52 *P. vivax* isolates from Oddar Meanchey province, 47 isolates were able to be deep sequenced, of which 24 (51%) were multiclonal. Furthermore, these results were in good agreement with F_{WS} results: 17/21 isolates with an $F_{WS} \geq 0.95$ were monoclonal by *pvmsp1* sequencing, while 19/26

isolates with an $F_{WS} < 0.95$ were multiclonal (Fisher's exact test $p=0.0004$). Subsequent to this, most analyses were performed both for the 28 *P. vivax* infections that were monoclonal by the F_{WS} metric and for all *P. vivax* samples (i.e. multiclonal and monoclonal infections) to assess the effect of multiclonality of our results.

4.3.3 Determining within-population genetic diversities

Principal component analysis (PCA) revealed no population substructure among *P. vivax* isolates. In contrast, we previously showed that *P. falciparum* parasites were partitioned into subpopulations, and we confirmed these findings with a more stringent variant calling analysis here (**Figure 4.2**) (Parobek *et al.*, *in submission*). These partitions do not correspond to collection site, date of collection, or multiplicity of infection. To confirm visual inspection of partitioning, we used *k*-means clustering to identify the number of *P. vivax* and *P. falciparum* clusters present. *K*-means clustering confirmed that all *P. vivax* isolates were part of a single cluster, and that the *P. falciparum* was subdivided into four clusters (**Figure 4.3**). As described previously (Parobek *et al., in submission*), the central *P. falciparum* population (here referred to as CP2) is most closely related to the admixed *P. falciparum* described by Miotto *et al.* [246] and likely represents the founder population for the other populations. Henceforth most analyses were performed both using all *P. falciparum* samples and using the 18 parasites of the central admixed population (known as CP2).

4.3.4 Type and timescale of demographic events

To better understand the differences in the population structures, we first examined the allele frequency spectra (AFS) of *P. vivax* and *P. falciparum* populations. Spectra were calculated by variant type (synonymous, non-synonymous, genic, and intergenic) and compared to the spectrum expected in a simulated coalescent population with no natural selection, constant population size, and complete random mating (**Figure 4.4**). We also determined the AFS among *P. vivax* monoclonal infections and among the parasites in the *P. falciparum* complete population and CP2 population alone. We observed an excess of intermediate-frequency derived alleles in the collective Cambodian *P. falciparum* AFS, with a local maximum at 19 - 21. This intermediate peak is likely driven by the presence of multiple parasite *P.*

falciparum subpopulations, and when the core CP2 subpopulation alone was analyzed, these excess intermediate frequency alleles disappeared (**Figure 4.5**). In contrast to the *P. falciparum* case, we observed an excess of low-frequency derived alleles in the Cambodian *P. vivax* AFS (both entire population and monoclonals only) compared to the neutral simulations, which suggests a population expansion. We used these qualitative comparisons as priors to inform demographic model selection.

To choose a best-fit demographic scenario and identify optimal population parameters, we adopted an efficient diffusion approach, which approximates a population of *N* individuals varying through discrete generations. Using this approach, the following demographic scenarios were fit to observed *P. vivax* and *P. falciparum* allele frequency spectra - constant population size, decline, exponential increase, two-epoch increase, and bottleneck with subsequent exponential growth. Toy scenarios are provided in **Figure 4.6**. We used the Akaike Information Criterion (AIC) to inform model selection. For both *P. vivax* (all samples and monoclonals only) and the *P. falciparum* CP2 (core population only) datasets, the positive exponential growth model was the best-fit model (**Table 4.1-1.3**). Interestingly, when considered as a whole, the best-fit model for the entire *P. falciparum* population, which includes satellite populations that are likely clonal expansions of multidrug-resistant parasites (Parobek *et al.*, *in submission*), was one of bottleneck and extreme growth. The best-fit models predicted that the *P. vivax* population has been expanding resulting in a larger effective population size (θ=850 for *P. vivax* monoclonals with best fit exponential growth model) than the sympatric *P. falciparum* population (θ=240 for *P. falciparum* CP2 with best fit exponential growth model).

4.3.5 Selection and sweeps on orthologous genes

The population structure of *P. falciparum* in Cambodia has been shaped by the intensive use of artemisinins and their partner drugs [246,247] (Parobek *et al.*, *in submission*). Previously, we showed that this structure likely includes and admixed founder population (CP2) with clonal expansion of three other populations (Parobek *et al.*, *in submission*). However, little is known about the selective forces that shape the *P. vivax* population. To understand the selective landscape shaping the *P. vivax* and the admixed *P. falciparum* population (CP2), we used LD-based tests to identify genomic regions that have undergone

72

selective sweeps. For this haplotype-based test for selection, we focused on the subset of monoclonal *P. vivax* isolates (n=28) and the CP2 *P. falciparum* group (n=18), which are predominantly monoclonal (15/18).

For *P. vivax*, we performed the nS_L test for positive selection, which represented 45,701 SNPs (**Figure 4.7**). Strikingly, among the strongest 15 strongest signals of positive selection, five were in close proximity to potential regulators of gene expression (three AP2-domain containing transcription factors and two SET-domain containing proteins, which are histone modulators). Among these 15 regions, four were neighboring transporters (MDR1, MDR2, MRP1, and an ABC-transporter) (**Table 4.4**). Notably, a previous comparison of a 1990s-era Peruvian *P. vivax* isolate to the reference Sal1 strain found an increased dN/dS signal at AP2-containing transcription factors and at *pvmrp1* [248]. In fact, the strongest selective sweep among the *P. vivax* isolates was on chromosome 14, adjacent to an AP2-domain containing transcription factor (PVX_122680). To determine the extent of haplotype homozygosity (i.e. how big is the sweep in terms of genomic coordinates), we turned to the extended haplotype homozygosity test (EHH). The focal SNP in this region of increased nS_L score was used for EHH analysis. We identified a 100 kb region of strong linkage disequilibrium around the principle chromosome 14 locus (**Figure 4.8**) in isolates with the selected allele. In addition to the AP2-domain transcription factor, this region contained 25 genes (**Table 4.5**). Analysis of the entire *Plasmodium vivax* population using nS_L yielded qualitatively similar results (**Figure 4.9**). Furthermore, the iHS statistic - which has been used extensively in malaria studies - was calculated for these same loci, yielding similar results (**Figure 4.9**).

Another prominent sweep encompassed the *pvmdr1* locus. We explored whether this locus has been responsible for a recent sweep. We compared key drug-resistance SNP frequencies in *pvmdr1* to frequencies in Cambodian isolates collected several years earlier from Kampot province, from 2006-2007 [249]. Two key mutations (Y976F and F1076L) existed at roughly the same frequency (89% previously vs 77% more recently; 87% vs 90%, respectively). This high similarity in allele frequency from samples collected before and during the global plan to contain artemisinin resistance provides evidence that the

sweep encompassing the *pvmdr1* locus was likely not driven by ACT-resistance containment efforts, and may be the result of CQ pressure (similar to *pfmdr1* mutations in *P. falciparum*). Similar to another study, we did not find any evidence of copy number variations encompassing the *pvmdr1* locus [249].

A similar analysis of selective sweeps was performed for the 5,158 SNPs of the *P. falciparum* CP2 founder population. Both *nS*L and iHS statistics revealed a single locus under strong directional selection. Surprisingly, this region was near the *pfama1* gene (< 20 kb) (which also was the 15th highest signal in *P. vivax*). Previous genome scans have identified this locus as one under a strong selective sweep [250,251]. This finding could be due to an incomplete sweep at this locus. To more closely investigate the selective forces acting on *pfama1*, we looked at Tajima's *D* for this locus. Gene-wide estimate of Tajima's *D* was approximately 0.3, indicating that *pfama1* is not under extreme balancing selection in genome-wide comparisons of Tajima's *D*.

4.3.6 Copy Number analysis

Segmental duplications (also known as copy-number variants, or CNVs) can drive the evolution of parasite populations [252]. To identify CNVs in these genome data, we used a probabilistic framework for detecting variants that incorporates multiple types of evidence, as well as a custom search algorithm that accommodates the high AT content of *Plasmodium* species. First, to validate our CNV-calling approach, we performed qPCR on 54 of the whole-genome sequenced *P. falciparum* isolates to determine copy-number at *pfmdr1* - a known CNV locus in Cambodia. We identified seven isolates with a segmental duplication encompassing this locus. In comparison with experimental results, we found that both genome-wide CNV-detection methods identified five of these seven isolates, and detected no false positives, demonstrating reasonable sensitivity and good specificity.

When all *P. falciparum* samples were analyzed, 14 isolates had detectable *pfmdr1* duplications, including five with 3 copies and two with 4 copies. Beyond *pfmdr1*, we identified three additional loci with CNVs in *P. falciparum*: the plasmepsin II locus (chromosome 14), *rh2b* & *rh2a* (chromosome 13), and *rab2* (chromosome 8). For plasmepsin II, 40 samples contained a duplication event, including 16 samples with three copies. We identified two unique segmental duplications that encompassed the *rh2b*

74

and *rh2a* loci. One event spanned both genes (n=8) while some samples contained only an *rh2b* duplication. Finally, a large duplication event spanned the *rab2* locus (ras-related protein Rab-2) in a single isolate. This duplication also encompassed PF3D7_1230800 (unknown function), PF3D7_1230900 (serine/threonine protein kinase RIO1), PF3D7_1231000 (unknown function), PF3D7_1231200 (unknown function), PF3D7_1231300 (unknown function), and PF3D7_1231400 (amino acid transporter). Similar to the most comprehensive *P. falciparum* CNV analysis to date [252], we found no evidence of selective sweeps around CNVs.

We identified a limited number of copy-number variants in the *P. vivax* genomes. The following genes were found to have several copies. Sixteen Cambodian *P. vivax* had a 2x duplication encompassing the Duffy binding protein (*dbp*, PVX_110810), which was previously described [253] in *P. vivax* isolates worldwide. We also discovered one *P. vivax* isolate which contained 4x *pvdbp* duplication event. Additionally, we identified 13 *P. vivax* isolates with a deletion that includes the two 5'-most exons of the cytoadherence-linked asexual protein (*clag*, PVX_094265) deletion, as well as an exported protein of unknown function (PVX_094270). Deletions of *clag* are common in *P. falciparum* cultured strains [254]. Similar to *P. falciparum*, there was no evidence of selective sweeps that encompassed segmental duplications in the *P. vivax* genome. In addition, similar to another study, we did not find any evidence of copy number variations encompassing the *pvmdr1* locus [249].

4.3.7 Genome scans for balancing selection

Genome-wide scans for balancing selection have been instrumental in identifying factors that shape populations. For *P. falciparum* populations, scans for balancing selection have contributed greatly to our understanding of the breadth of host-parasite antigen-antibody interactions. This dataset affords us the unique opportunity to investigate the signatures of balancing selection in co-endemic *P. vivax* and *P. falciparum* populations. We calculated Tajima's *D* for every non-excluded gene in the *P. vivax* and *P. falciparum* genome (approximately 5,300 and 5,200 genes, respectively), and observed, in both species, mean value *D* < 0, with a strong right skew (**Figure 4.10**). Because Tajima's *D* is sensitive to demographic history, we used the best-fit demographic models - as determined above - to parameterize

coalescent simulations on a gene-by-gene basis. The results of these simulations suggested that the Tajima's *D* distribution we observed largely reflects the population expansion that both *P. vivax* and *P. falciparum* have undergone (data not shown).

4.3.8 Ortholog analyses

Studying whole-genomes from sympatric *P. vivax* and *P. falciparum* populations allows us to investigate the congruity of selective forces acting on orthologs. A realistic explanation for the mechanisms of population resilience underlying the dramatic differences we observe in *P. vivax* and *P. falciparum* could be asymmetric treatment and responses of orthologs. For example, one likely scenario for explaining the differences between *P. falciparum* and *P. vivax* is that there have been particular genetic loci in *P. falciparum* that are subjected to strong selective pressures, when orthologous genes from *P. vivax* have not been exposed to such pressures.

There are 1,852 genes that have a single *P. vivax* / *P. falciparum* ortholog [255]. For each orthologous pair, we studied selection dynamics to identify both concordant and discordant ortholog pairs. As above, we calculated Tajima's *D* for each gene with an ortholog from *P. vivax* and *P. falciparum* and also performed a per-exon analysis (data not shown). We also investigated a new concept, the idea of **Δ**Tajima's D, to study the genes with discordant signatures of selection. Overall, the signature of selection was concordant between orthologs (values near 0), though there were a subset ortholog pairs with extreme negative or extreme positive **Δ**Tajima's *D* (**Figure 4.11**). We investigated the top 1% of orthologous gene pairs with opposing signals of selection by GO-term analysis, but found no enrichment. It is possible that the tails of the **Δ**Tajima's *D* distribution hold meaningful biological signals, but that any signal is obscured by the high proportion of genes with unknown function in the tails of this distribution.

4.4   Discussion

4.4.1 Species-specific mechanisms of population-level resilience

In this study, we whole-genome sequenced 70 *P. vivax* and 80 co-endemic *P. falciparum* clinical isolates from three sites in western Cambodia. This dataset represents a unique resource for comparing the effects of selective forces have had on the genomes of these *Plasmodium* species. Because these isolates

were collected during a concerted malaria-elimination campaign, they can shed light on the different responses of these two species to control efforts. We applied population-genetic methods to investigate these responses.

On the basis of epidemiological data, it has been observed that *P. vivax* may be more difficult to eliminate than *P. falciparum* [8]. However, it is as-yet unclear why this may be the case. Understanding the long-term and recent demographic histories, as well as comparing the respective population structures, of Cambodian *P. vivax* and *P. falciparum* may be of great utility to malaria control programs by shedding light on how the population has responded to interventions and documenting how the genome has changed due the imposed selection. Sympatric whole-genome data for *P. vivax* and *P. falciparum* are particularly suited for such an investigation. Similar to several recent studies of co-endemic *P. falciparum* and *P. vivax* isolates [256–261], we observed more genetic structuring among *P. falciparum* isolates, which were split among four highly separated populations, than among *P. vivax*. Our results add the first whole-genome evidence to the increasingly common observation that *P. vivax* is typically has a simpler, less fractured population structure more admixed than sympatric *P. falciparum,* even in the face of large bed-net distributions and intensive diagnosis and treatment. The demographic models support the population structure analysis as well. Given our sample sizes, we have power for robust detection of ancient demographic events, and potentially power for detection of recent events [262]. Best-fit demographic models indicated that *P. vivax* has undergone steady exponential growth in the region with genetic diversity within the *P. vivax* population has expanding more rapidly compared the admixed founder population CP2. Consistent with this, a demography-adjusted estimate of the population mutation rate (θ) - a proxy for effective population size (N_{eff}) - found that the *P. vivax* N_{eff} is substantially larger than *P. falciparum*. The overall population demographic history encompassing all isolates is best described by an exponential growth model, consistent with population splits and expansion of what appear to be clonal lines of drug resistant parasites (Parobek *et al.*, *in submission*).

These demographic scenarios match epidemiological observations in the region [8]. Specifically, even as *P. falciparum* cases have dropped rapidly, *P. vivax* cases have increased. This could be due to (1)

77

the ability of *P. vivax* infections to go dormant or undetected during hypnozygony, allowing the

population to maintain occult genetic diversity, (2) earlier transmission of *P. vivax* due to earlier

gametocytemia, allowing transmission events to occur prior to clinical symptoms and presentation for

treatment, or (3) a release of *P. vivax* competitive inhibition due to falling *P. falciparum* cases [263,264].

It could also be because *P. vivax* has unique genetic mechanisms to maintain its resilience in spite of

control measures. If there are different mechanisms between *P. vivax* and *P. falciparum*, these differences

might be seen in scans for selection.

These data enabled direct comparison of selective signatures in *P. vivax* and *P. falciparum*

orthologs. While it is known that orthologs operate under different selective regimes [244], the ability to

simultaneously investigate thousands of orthologs presents a rich resource for generating hypotheses

about the biological differences between *P. vivax* and *P. falciparum*. We used best-fit demographic

scenarios to construct null distributions of Tajima's *D* values for Cambodian *P. vivax* and *P. falciparum*.

Using these distributions, we identified genes in both *P. vivax* and *P. falciparum* under extreme

directional or balancing selection. On a gene-wise basis, both species showed a negative mean Tajima's *D*

value with strong right skew (**Figure 4.10**), similar to previous reports from Africa in *P. falciparum*

[265]. Among the minority of genes with signatures of balancing selection, candidate erythrocytic stage

vaccine antigens in both species were over-represented. Analysis of **Δ**Tajima's *D* between orthologous

gene pairs revealed a small minority of genes that have highly discordant marks of selection. As loci

within the *P. falciparum* and *P. vivax* genomes receive better annotation, this signal in particular may

point the way to important biological differences between these two species.

We also employed haplotypic measures, nS_L and iHS, to assess for genomic regions in the

populations under recent directional selection and to identify incomplete sweeps. To date, *P. falciparum*

studies have demonstrated that the strongest signatures of selection are nearby multidrug transporters,

antimalarial targets, and antigens [250,251,266–268]. Similarly, in *P. vivax*, we identified some selective

sweeps with the focal variant occurring in close proximity to known and potential drug-resistance genes

(*pvmdr1*, *pvmdr2*, *pvmrp1*, and an ABC-family transporter; **Table 4.4**).

However, surprisingly, in our *P. vivax* population sample, the strongest selective sweeps occurred in close proximity to multiple AP2- and SET-domain containing proteins. Both of these classes of proteins are likely transcriptional regulators, raising the possibility that *P. vivax* is responding to selective pressures by altering its transcriptional profile. Fully one-third of the strongest signatures of directional selection in *P. vivax* were likely driven by transcriptional regulators. These results suggest that while *P. falciparum* populations experience sweeps of drug-resistance loci among most studies [250,266,267,269,270], *P. vivax* populations may rely more heavily on complex transcriptional mechanisms to evade selective pressures. Of note, the statistical tests for directional selection used here can not determine the presence of a sweep if a mutation is at fixation. Within our samples, it is clear that several drug resistance mutations are nearly fixed within the population, including the CVIET haplotype *pfcrt* associated with chloroquine resistance (**Figure 3.11**), thus masking these selected areas in our analysis. These data suggest that molecular surveillance for mutations in drug-resistance genes may be of comparatively limited utility in *P. vivax*, if indeed transcriptional regulation plays an important role in promoting parasite fitness.

This finding is not, however, not an absolute difference between species. A closer look at previously published *P. falciparum* scans for selection in multiple locations in Africa reveals a striking finding. Park et al. found a region on chromosome 6 strongly with strong signal of positive selection in pyrimethamine-resistant parasites - this region contained, among other things, a SET-domain containing protein (PF3D7_0629700), which was nearly at the center of the selective sweep [270], and this is the ortholog of the SET-domain protein on chromosome 11 (PVX_114585). Nwakanma et al. also found this region to be under "exceptionally strong" directional selection in the Gambia [271], as have other studies [250,266]. Though it is known that transcriptional timing can be an important component of *P. falciparum* drug-resistance responses [43], these results suggest an underappreciation for the role of transcriptional regulation modification in *P. falciparum* fitness.

Among the CP2 *P. falciparum* population, we identified only a single locus with a convincing signature of directional selection. This relative lack of loci under directional selection could be due to

unrecognized multiclonal infections in the CP2 group, leading to haplotype breakdown during sequence analysis. It could also be due to undetected fixed sweeps, which nS_L and iHS are insensitive to detect. Surprisingly, this locus was in close proximity to the *pfama1* locus. While past scans for directional selection have identified a strong signature at this locus [250,251], it remains unclear whether the *pfama1* locus or a nearby gene is driving this sweep. Because the nS_L and iHS statistics are sensitive for incomplete sweeps, a strong signature of balancing selection in a subdomain of *pfama1* could be causing this signature. This finding bears further investigation.

We also sought to identify and characterize segmental duplications in the *P. vivax* and *P. falciparum* genomes, and, as our data are uncaptured, they are particularly useful for studies of structural variation. Segmental duplications in *Plasmodia* species can be associated with increased fitness in certain settings (eg. the *P. falciparum* multidrug resistance transporter), and studies of copy-number changes have proved instrumental in uncovering unique *P. vivax* biology [253]. A striking finding was the highly duplicated nature of the *P. vivax* Duffy binding protein (*pvdbp*) in this Cambodian population. A conserved segmental duplication encompassing the *pvdbp* locus was previously described in a handful of globally sourced *P. vivax* isolates, suggesting a recent duplication event and rapid sweep, perhaps because it conferred a RBC-invasion advantage. That the *pvdbp* duplication is so common in our sample (17/70 isolates, including a single 4x duplication) suggests an important mechanism that may transcend invasion of Duffy-negative human hosts. Similar to a recent study of segmental duplications in the *P. falciparum* genome [43,252], we found no evidence that copy-number events are driving selective sweeps in either *P. falciparum* or *P. vivax*.

Our study is subject to some limitations. First, although we have curated our data carefully - performing extensive tests to determine the best alignment, filtering, and variant-calling approaches - unappreciated error at these steps could spawn spurious variants and skew the allele frequency spectrum. The risk of this is heightened in *P. vivax*, for which extensive recognized and unrecognized paralogous families present significant mapping and variant-calling challenges [272].

Second, our demographic inferences are constructed upon Wright-Fisher assumptions. Malaria parasites have a complex lifecycle, including stages in human and mosquito, with multiple clonal generations occurring within the host bloodstream and frequent bottlenecks during transmission [273]. Such realities violate the assumptions of standard population genetic models, and may complicate the process of population-genetic inference from genetic data that relies on these models. These peculiarities of the malaria lifecycle may skew the allele frequency spectrum toward increased singletons - even at neutral sites - leading to quantitatively or even qualitatively inappropriate demographic conclusions [268,274]. Even if the models do approximate the genetic features of *Plasmodia*, our power to accurately infer recent population events from the site-frequency spectrum, may be limited for very recent events [262]. Selection bias could further limit our ability to infer population dynamics from our sample, as we sequenced parasites (1) causing patent disease, and (2) with a high enough density for whole-genome sequencing.

Third, while *P. vivax* and *P. falciparum* populations in western Cambodia have been exposed to similar selective pressures over time, these pressures are certainly not identical, especially with respect pharmacotherapy. For roughly ten years, from the early 2000s until 2012, national guidelines in Cambodia mandated artemisinin-based therapy for *P. falciparum* infections and chloroquine-based therapy for *P. vivax* [28]. Despite this important difference, *P. vivax* and *P. falciparum* populations almost certainly experienced cross-treatment, particularly in cases of mixed infections or empiric diagnosis and treatment.

Despite these shortcomings, our study has fundamental strengths. We have a reasonable sample size, including the largest *P. vivax* population sequenced from any single country. Importantly, none of these samples were hybrid captured, giving us greater confidence in the quantitative accuracy of calls in mixed infections and structural variants. This co-endemic *P. vivax* and *P. falciparum* genome-wide sequencing dataset is a first, and will continue to provide useful grounds for mining.

Our findings raise important questions. First, while the total number of *P. falciparum* cases in Cambodia decreased, our demographic models show a slowly increasing N_{eff} - at least for the CP2

founder population. Although it is possible that our demographic models are detecting ancient rather than recent trends, another study of isolates in this area identified the same trend [275], lending credence to our findings and suggesting that recent control efforts have not significantly decreased *P. falciparum* genetic diversity in the region.

That *P. vivax* - and to some extent *P. falciparum* - may modulate transcriptional regulation in their response to control efforts is a finding that bears closer investigation. A multi-step approach to investigating these findings should be adopted. First, we need to see whether the findings that strong selective sweeps near potential transcriptional modulators (in our case, AP2-domain transcription factors and SET-domain chromatin modifiers) are replicable in other populations. For example, it would be interesting to see whether these signals hold in *P. vivax* populations from places with prolonged zero-transmission seasons (eg. North Korea). A strong signal on *P. falciparum* chromosome 6 centered near a SET-domain chromatin modifier has already been identified in multiple studies of African isolates [250,266,270,271].

Once confirmed, the selected alleles in these sweeps should be studied for their specific role in transcriptional regulation. In *P. falciparum*, *in vitro* studies could uncover the role of selected alleles in transcriptional changes throughout the parasite lifecycle. While culture conditions certainly do not mirror field conditions in an elimination setting, such experiments would provide well-controlled and evidence. Determining the role of selected *P. vivax* variants will be more challenging. Modifications to *Plasmodium knowlesi* homologs could be tested in culture [276]; humanized mice could be infected with transgenic *P. vivax* allowing for transcriptional analysis through the liver stage [277,278]; or monkeys could be infected with transgenic *P. vivax*, allowing for transcriptional analysis through the blood stage [279]. If such experiments demonstrate that selected variants do indeed modulate transcription, this would provide key evidence of an advanced parasitic response to selective pressure. It would also suggest that tracking highly fit *P. vivax* parasites could be complicated, and would provide further evidence in support of the idea that *P. vivax* will be the more challenging species to eliminate.

In summary, we have presented evidence that *P. vivax* and *P. falciparum* populations in Cambodia have responded in substantively different ways to the selective pressure of the artemisinin-resistance containment campaign. While *P. falciparum* has seen population splitting with clonal and sporadic transmission, the *P. vivax* population remains admixed, with strong growth genetic diversity perhaps enabled by a high frequency of multiple infections. There are clear differences in the genes under selection in *P. vivax* and *P. falciparum*, with many orthologs under strongly opposite signals of selection. Finally, while both *P. vivax* and *P. falciparum* experience selective sweeps, the strongest sweeps in *P. vivax* occur nearby transcriptional regulators while, in several previously published *P. falciparum* datasets, the strongest sweeps occur near drug-resistance genes. These findings highlight potentially important differences between *P. vivax* and *P. falciparum*, suggest that *P. vivax* will be more difficult to eliminate, and indicate that a true understanding of *P. vivax* biology will require additional studies of this parasite.

4.4.2 Data Availability

Unaligned sequence data are available at the Sequence Read Archive. Accessions are in **Table 4.6**.

4.4.3 Author Contributions

DLS, CL, CL and MDS designed and executed the original studies (WR1576 and WR1877). CMC coordinated the execution of this study with the Cambodian National Malaria Control program. CMP, DLG, SB, JTL and JJJ designed and executed the experiments. CMP, JAB, DKD, EJB and NB contributed to the bioinformatic analyses. CMP, DLS, JTL, JAB and JJJ wrote the manuscript.

4.4.3 Declaration of Interests

The authors declare no competing financial interests.

Model	θ	η_D* (0.001 - 1.0)	η_G* (1.0 - 100)	T* (0.01 - 5.0)	Log Likelihood	AIC
Static	1868.499	-	-	-	-911.145	1824.29
NegG	5260.287	0.328	-	4.583	-1007.829	2021.65
PosG-Epoch	898.556	-	9.819	0.672	-28.569	63.138
PosG-Exp	850.687	-	19.999	1.032	-28.077	62.154
BG	4053.125	0.0307	4.0639	0.367	-28.020	64.04

Table 4.1 Demographic history of *P. vivax* monoclonals.

Highest log-likelihood parameter sets were selected from 100 model-fit replicates. * Ranges explored for optimization are provided for optimizable parameters.

Model	θ	η_D* (0.001 - 1.0)	η_G* (1.0 - 100)	T* (0.01 - 5.0)	Log Likelihood	AIC
Static	1568.496	-	-	-	-601.569	1205.14
NegG	1406454	0.00111	-	4.755	-602.773	1211.546
PosG-Epoch	848.495	-	7.798	0.547	-31.500	79.00
PosG-Exp	816.378	-	14.342	0.830	-27.594	61.188
BG	816.373	0.972	14.332	0.839	-27.593	63.186

Table 4.2 Demographic history of *P. vivax* entire population

Highest log-likelihood parameter sets were selected from 100 model-fit replicates. * Ranges explored for optimization are provided for optimizable parameters.

Model	θ	η_D* (0.001 - 1.0)	η_G* (1 - 100)	T* (0.01 - 5.0)	Log Likelihood	AIC
Static	396.184	-	-	-	-27.336	56.672
NegG	396.228	0.9997	-	0.820	-27.340	68.68
PosG-Epoch	266.604	-	1.691	1.729	-24.789	55.578
PosG	240.603	-	1.942	4.994	-24.736	55.472
BG	304.952	0.998	1.538	2.282	-24.738	57.476

Table 4.3 Demographic history for the core *P. falciparum* population (CP2).

Highest log-likelihood parameter sets were selected from 100 model-fit replicates. * Ranges explored for optimization are provided for optimizable parameters.

Chr	Focal SNP Stats		Sweep Region Stats		
	Location	nS$_L$	Closest Plausible Genetic Driver	Gene ID	Distance*
14	797870	6.175	transcription factor with AP2 domain(s), putative (ApiAP2)	PVX_122680	-4462
14	1650897	4.386	heterochromatin protein 1 (HP1); H3 lysine methyltransferase (SET10)	PVX_123682; PVX_123685	999 9918
10	351691	4.207	multidrug resistance protein 1, putative (MDR1)	PVX_080100	10010
04	577543	4.201	serine-repeat antigen 4 (SERA) serine-repeat antigen 5 (SERA)	PVX_003825; PVX_003830	789 -1691
02	85446	-3.948	numerous *vir* and *fam* members	-	-
03	368737	-3.914	hypothetical genes	-	-
12	2421187	3.591	multidrug resistance protein 2 (MDR2)	PVX_118100	-4361
13	396427	3.589	ABC transporter B family member 7, putative (ABCB7)	PVX_084521	-7326
11	749030	-3.567	SET domain containing protein	PVX_114585	46877
11	1900445	-3.487	transcription factor with AP2 domain(s), putative (ApiAP2)	PVX_113370	-31184
06	574362	-3.438	merozoite TRAP-like protein, putative (MTRAP)	PVX_111290	6929
09	1693066	-3.349	transcription factor with AP2 domain(s), putative (AP2-O)	PVX_092760	7201
02	145835	3.318	multidrug resistance-associated protein 1, putative (MRP1)	PVX_097025	7807
07	934541	3.268	hypothetical genes	-	-
09	1252731	3.219	heat shock protein 70, putative; apical membrane antigen 1 (AMA1)	PVX_092310 PVX_092275	5'UTR -21528

Table 4.4 Top fifteen *P. vivax* genomic regions under recent directional selection.

nS$_L$ analysis was performed in *P. vivax* monoclonal infections to avoid false haplotype breakdown in multiclonal clinical isolates. * Distance from focal SNP to potential driver gene; a negative sign indicates upstream.

| Gene ID	Gene Description	Position
PVX_122645	pre-mRNA-processing factor 40, putative (PRP40)	Pv_Sal1_chr14: 745,554 - 749,056 (+)
PVX_122650	choline-phosphate cytidylyltransferase, putative	Pv_Sal1_chr14: 752,305 - 755,482 (+)
PVX_122655	hypothetical protein, conserved	Pv_Sal1_chr14: 756,864 - 759,127 (+)
PVX_122660	protein transport protein SEC20, putative	Pv_Sal1_chr14: 759,434 - 761,371 (-)
PVX_122665	hypothetical protein, conserved	Pv_Sal1_chr14: 762,592 - 765,741 (-)
PVX_122670	U4/U6.U5 tri-snRNP-associated protein 2, putative	Pv_Sal1_chr14: 768,843 - 770,900 (+)
PVX_122675	DNA replication licensing factor MCM4, putative	Pv_Sal1_chr14: 772,315 - 775,182 (+)
PVX_122680	transcription factor with AP2 domain(s), putative (ApiAP2)	Pv_Sal1_chr14: 785,708 - 793,408 (+)
PVX_122685	hypothetical protein, conserved	Pv_Sal1_chr14: 794,062 - 796,623 (-)
PVX_122690	hypothetical protein, conserved	Pv_Sal1_chr14: 797,092 - 798,234 (+)
PVX_122695	hypothetical protein, conserved	Pv_Sal1_chr14: 799,176 - 799,918 (-)
PVX_122700	hypothetical protein, conserved	Pv_Sal1_chr14: 800,245 - 802,852 (+)
PVX_122705	hypothetical protein, conserved	Pv_Sal1_chr14: 803,430 - 803,912 (-)
PVX_122710	40S ribosomal protein S19, putative (RPS19)	Pv_Sal1_chr14: 805,253 - 806,554 (+)
PVX_122715	hypothetical protein, conserved	Pv_Sal1_chr14: 807,324 - 809,996 (-)
PVX_122720	hypothetical protein, conserved	Pv_Sal1_chr14: 811,474 - 812,043 (-)
PVX_122725	ferredoxin, putative	Pv_Sal1_chr14: 814,528 - 815,735 (-)
PVX_122730	glycerol-3-phosphate 1-O-acyltransferase, putative (G3PAT)	Pv_Sal1_chr14: 816,300 - 817,556 (+)
PVX_122735	hypothetical protein, conserved	Pv_Sal1_chr14: 819,451 - 824,001 (+)
PVX_122740	structural maintenance of chromosome 2, putative	Pv_Sal1_chr14: 824,711 - 829,052 (-)
PVX_122742	conserved Plasmodium protein, unknown function	Pv_Sal1_chr14: 830,213 - 832,084 (+)
PVX_122745	hypothetical protein, conserved	Pv_Sal1_chr14: 832,095 - 834,081 (+)
PVX_122750	hypothetical protein, conserved	Pv_Sal1_chr14: 834,505 - 836,979 (-)
PVX_122755	translocation protein SEC63, putative (SEC63)	Pv_Sal1_chr14: 839,917 - 842,636 (-)
PVX_122760	hypothetical protein, conserved	Pv_Sal1_chr14: 844,891 - 846,828 (+)

Table 4.5 Loci falling within the principal *P. vivax* selective sweep, occurring on chromosome 14.

Identifier, description, and location of genes that occur within the sweep are provided.
While the sweep haplotype extends approximately 50 Kb in each direction,
genes occurring within 5 Kb of the focal SNP are highlighted.

Species	Province	SRA Accession Numbers
P. vivax	Battambang	SRR2315729, SRR2315849, SRR2316038, SRR2316105, SRR2316478, SRR2316872, SRR2316895, SRR2317109, SRR2317489
P. vivax	Kampot	SRR2317560, SRR2315958, SRR2315988, SRR2316010, SRR2316011, SRR2316013, SRR2316015, SRR2316017, SRR2316032
P. vivax	Oddar Meanchey	SRR2316034, SRR2316036, SRR2316041, SRR2316044, SRR2316048, SRR2316051, SRR2316083, SRR2316087, SRR2316091, SRR2316095, SRR2316098, SRR2316101, SRR2316108, SRR2316111, SRR2316115, SRR2316117, SRR2316172, SRR2316252, SRR2316299, SRR2316344, SRR2316398, SRR2316422, SRR2316531, SRR2316564, SRR2316625, SRR2316658, SRR2316724, SRR2316755, SRR2316807, SRR2316839, SRR2316865, SRR2316867, SRR2316875, SRR2316879, SRR2316883, SRR2316887, SRR2316889, SRR2316890, SRR2316891, SRR2316892, SRR2316893, SRR2316894, SRR2316920, SRR2316921, SRR2316922, SRR2316923, SRR2316924, SRR2316925, SRR2316926, SRR2316970, SRR2317000, SRR2317001, SRR2317142, SRR2317171, SRR2317223, SRR2317268, SRR2317321, SRR2317322, SRR2317410, SRR2317444
P. falciparum	Battambang	SRR2317584, SRR2317585, SRR2317700, SRR2317711, SRR2317722, SRR2318021, SRR2318061, SRR2318484, SRR2318682, SRR2318694
P. falciparum	Kampot	SRR2317586, SRR2317587, SRR2317692, SRR2317693, SRR2317694, SRR2317695, SRR2317696, SRR2317697, SRR2317698
P. falciparum	Oddar Meanchey	SRR2317699, SRR2317701, SRR2317702, SRR2317703, SRR2317704, SRR2317705, SRR2317706, SRR2317707, SRR2317708, SRR2317709, SRR2317710, SRR2317712, SRR2317713, SRR2317714, SRR2317715, SRR2317716, SRR2317717, SRR2317718, SRR2317719, SRR2317720, SRR2317721, SRR2317723, SRR2317724, SRR2317725, SRR2317726, SRR2317727, SRR2317728, SRR2317729, SRR2317730, SRR2318019, SRR2318020, SRR2318023, SRR2318024, SRR2318031, SRR2318033, SRR2318034, SRR2318035, SRR2318039, SRR2318040, SRR2318041, SRR2318056, SRR2318126, SRR2318130, SRR2318177, SRR2318219, SRR2318297, SRR2318319, SRR2318368, SRR2318415, SRR2318446, SRR2318479, SRR2318576, SRR2318622, SRR2318670, SRR2318675, SRR2318676, SRR2318677, SRR2318678, SRR2318679, SRR2318680, SRR2318681, SRR2318683, SRR2318684, SRR2318685, SRR2318686, SRR2318688, SRR2318689, SRR2318690, SRR2318691, SRR2318692, SRR2318693, SRR2318702, SRR2318703, SRR2318704

Table 4.6 Sequence Read Archive accession numbers for *P. vivax* and *P. falciparum* isolates whole-genome sequenced in this study.

Accessions are classified according to species (*P. vivax* or *P. falciparum*) and province of collection (Battambang, Kampot, or Oddar Meanchey).
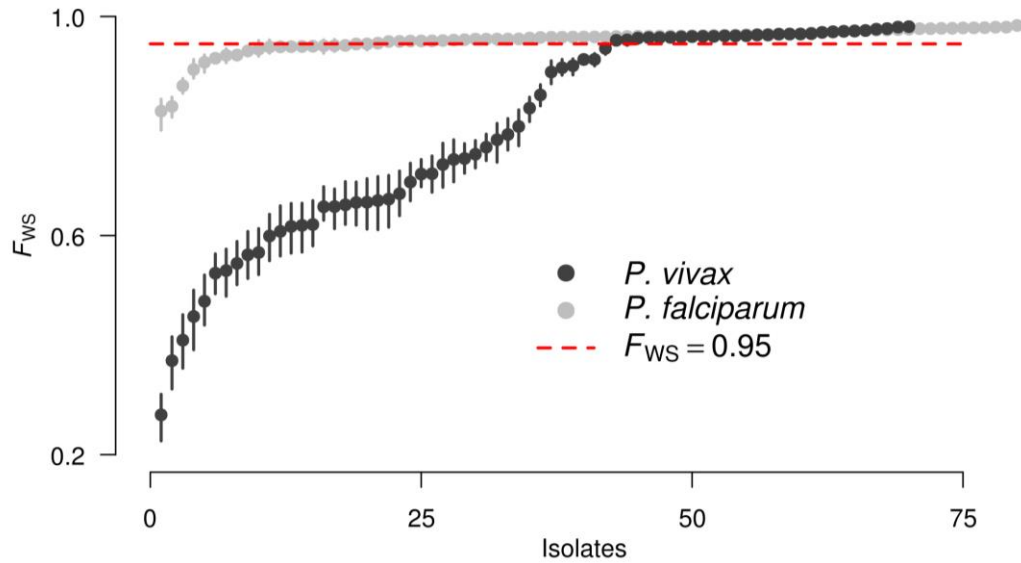
Figure 4.1 Multiplicity of infection for *P. vivax* and *P. falciparum*.

(A) F_{WS} calculated for *P. vivax* and *P. falciparum*. Points represent point estimate of F_{WS} for each sample in the respective population. Vertical bars represent the maximum and minimum value in 1000 bootstrap replicates, which downsampled the number of variants to be equal for *P. vivax* and *P. falciparum*, to correct for the increased number of *P. vivax* variants. **(B)** Summary bargraph representing the number of *P. vivax* and *P. falciparum* clinical isolates considered monoclonal or polyclonal, based on a cutoff of $F_{WS} < 0.95$ being considered polyclonal.
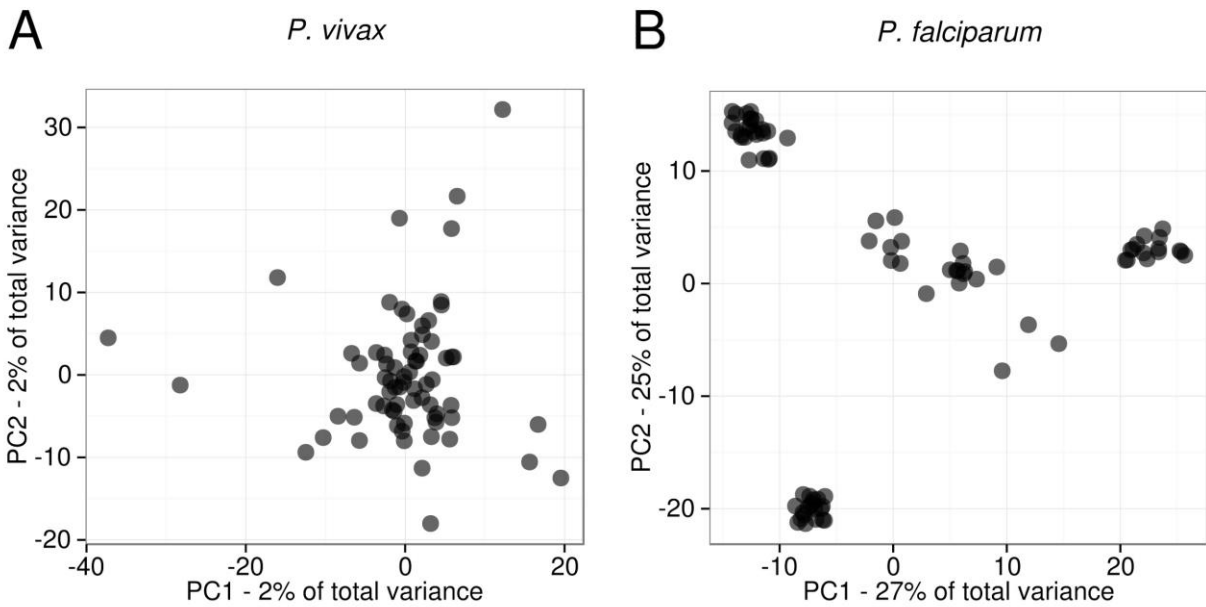
Figure 4.2 Principal components analysis of *P. vivax* and *P. falciparum* populations from Cambodia.

Calculated from genetic distance matrices which were calculated from single-nucleotide variant data from whole-genome sequencing data for each of 70 *P. vivax* clinical isolates **(A)** and 80 *P. falciparum* clinical isolates **(B)**. Noise was added to allow visualization of isolates.
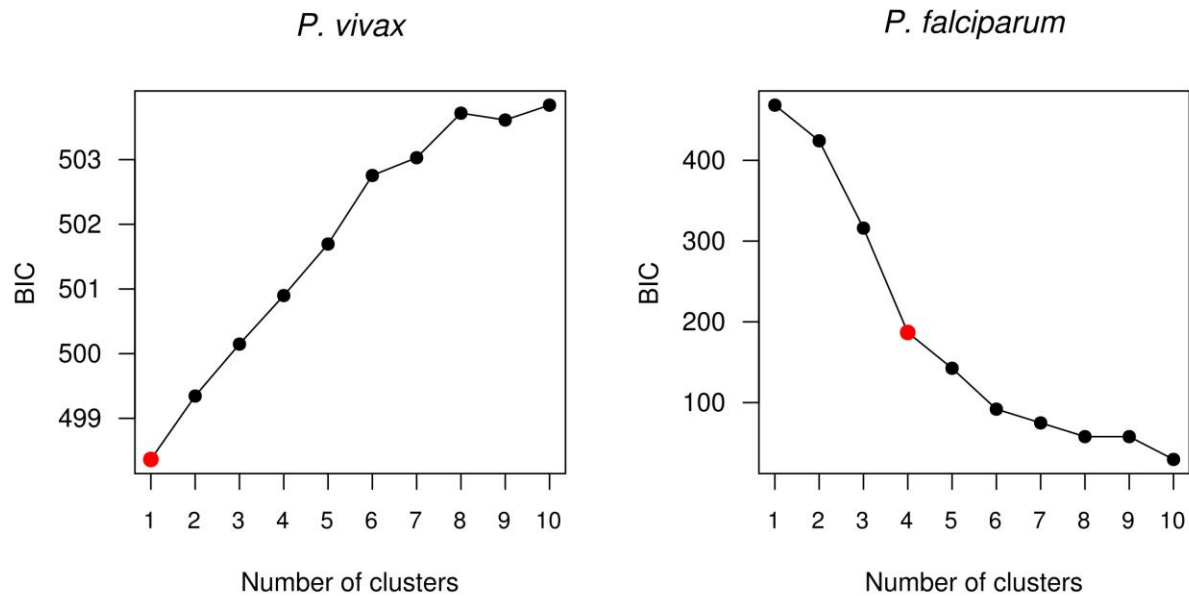
Figure 4.3 *K*-means clustering provides evidence for the number of subpopulations in
P. vivax and *P. falciparum* populations.

Nonparametric *k*-means clustering of total genetic diversity in *P. vivax* and *P. falciparum* isolates. For *P. vivax*, clustering suggests that all parasites are drawn from the same population. For *P. falciparum*, Increasing the number of clusters from n=3 to n=4 yields a substantial improvement in Bayesian information criterion (BIC). While subsequent increases do yield BIC improvements, the improvement in goodness of fit is not commensurate with the cost associated with increased degrees of freedom.
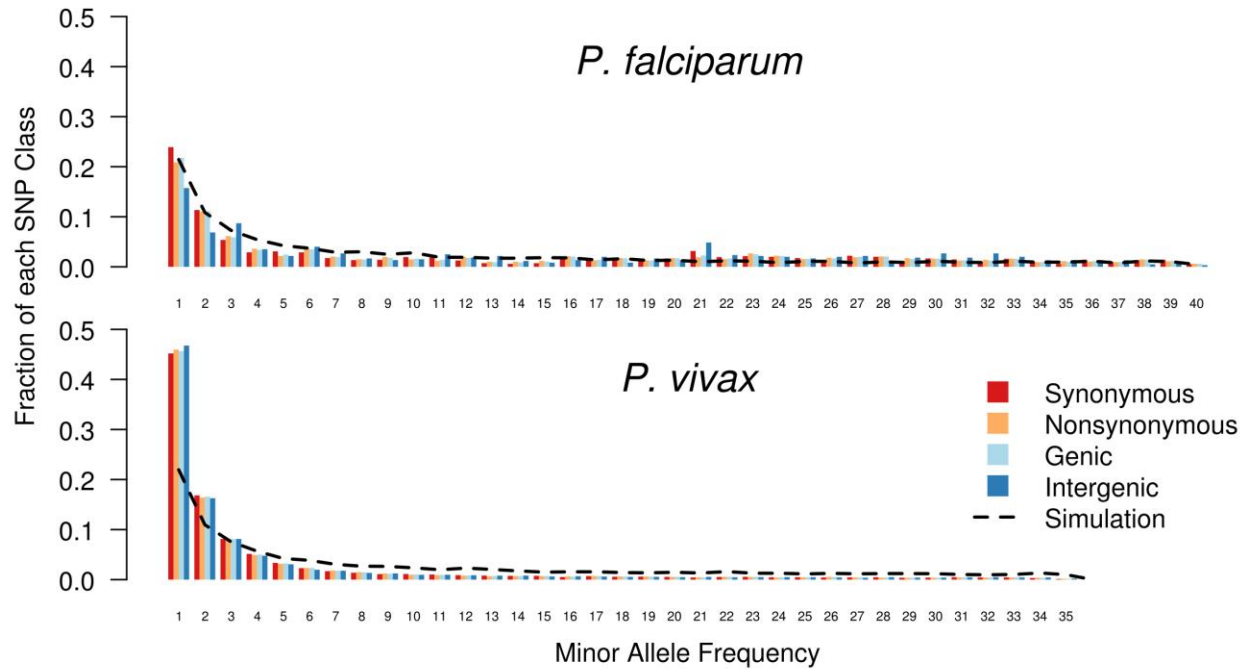
Figure 4.4 Allele-frequency spectra suggest strong, sustained growth for the *P. vivax* population, but muted population expansion for *P. falciparum*.

We observed a preponderance of low-frequency alleles in the *P. vivax* population. Under a Wright-Fisher model of genetic evolution, this suggests that the *P. vivax* population has undergone a recent expansion. In contrast, for *P. falciparum*, we observed an excess in intermediate-frequency minor alleles, reflecting the subdivided population structure.
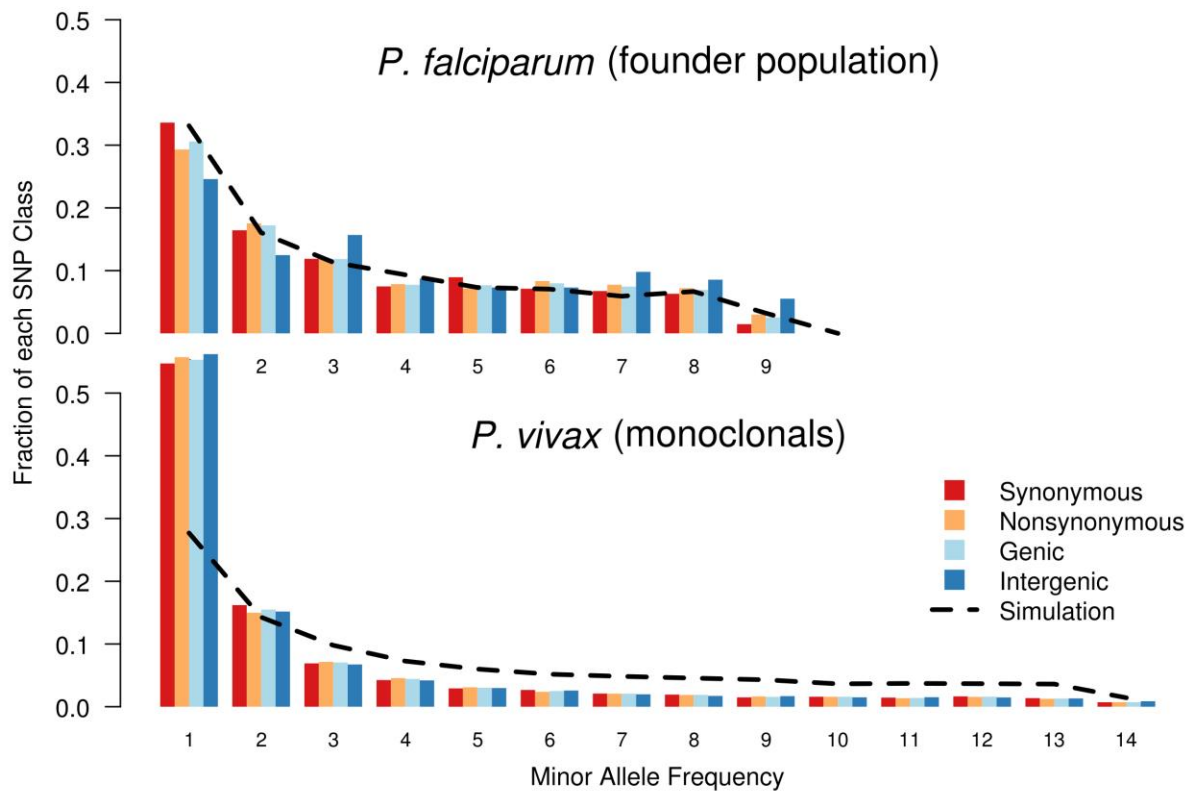
**Figure 4.5 Allele-frequency spectra for *P. falciparum* CP2 founder population and
P. vivax monoclonal isolates.**

When considered in isolation, there is no increase in intermediate-frequency alleles within the *P. falciparum* CP2 founder population. Similar to the entire *P. vivax* population, the *P. vivax* monoclonal isolates have a strong overrepresentation of low-frequency alleles, suggesting an expanding population size.
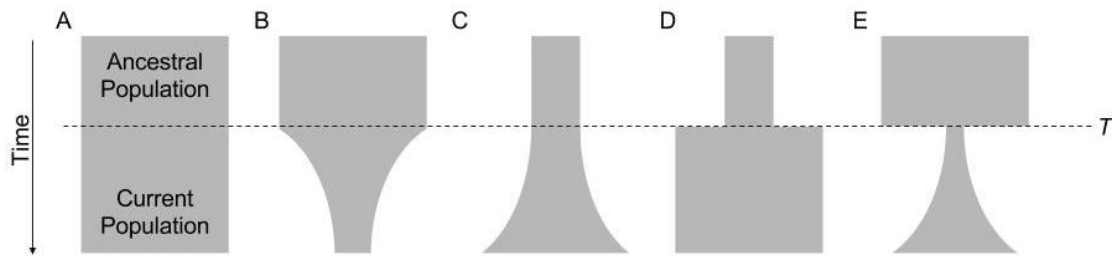
Figure 4.6 Five one-population models were used for demographic inference.

The following models were fitted to observed *P. vivax* and *P. falciparum* allele frequency spectra under using a diffusion approximation to population evolution: **(A)** a model of no change in effective population size through time; **(B)** a model of population decline beginning at time *T*; **(C)** A model of exponential population expansion, beginning at time *T*; **(D)** a two-epoch model of sudden population expansion at time *T*; (E) a model of rapid population decline followed by exponential growth, beginning at time *T*.
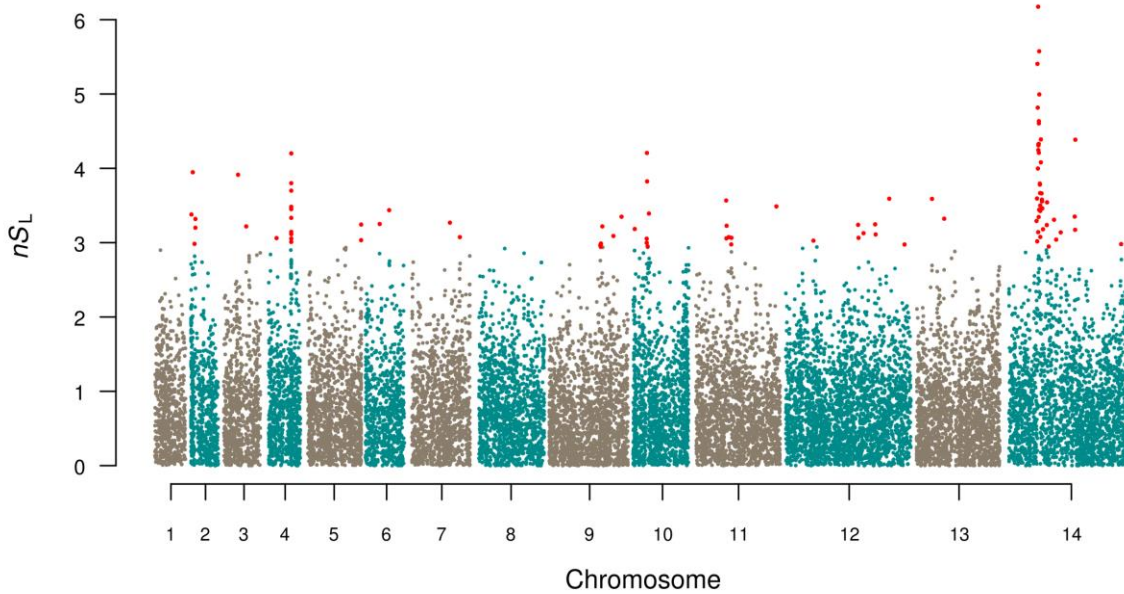
Figure 4.7 Haplotype sweep provides evidence of positive selection in Cambodian *P. vivax*.

Manhattan plot of normalized nS_L values is shown. Each point corresponds to a SNP, and the top 0.5% of hits are colored in red. This view suggests that there are several genomic regions under positive selection, including areas near transcription factors (AP2-containing domain), chromatin regulators (SET10, HP1), antigens under known positive selection (SERA4 & 5), and drug resistance genes (MDR1).
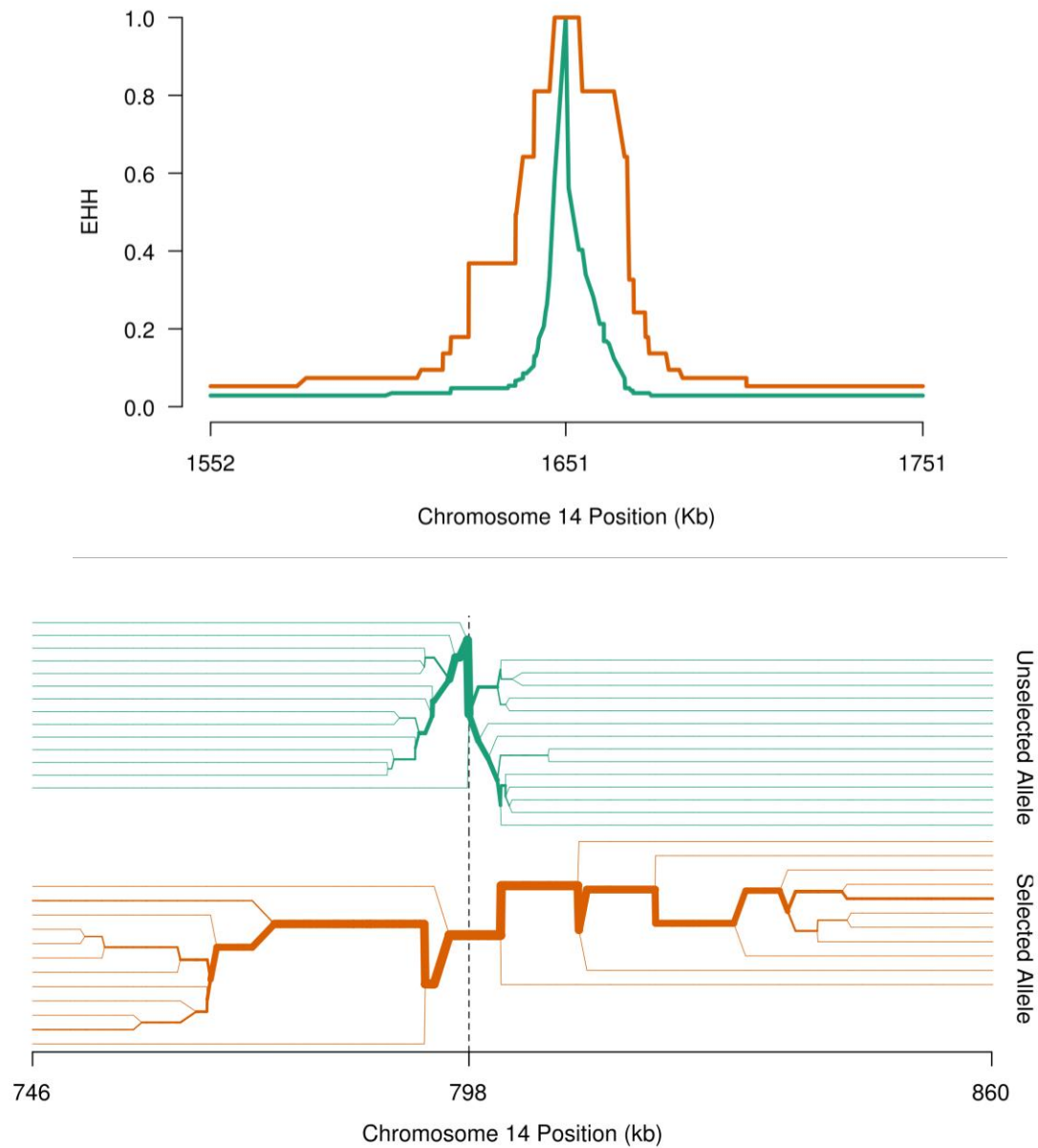
Figure 4.8 Evidence of recent positive selection centered around an AP2-domain containing transcription factor.

The upper panel shows the decay of extended haplotype homozygosity (EHH) for the selected (orange) and unselected (blue) haplotypes. It is evident that a large region surrounding the selected focal variant is in strong linkage disequilibrium. The lower panel shows a haplotype bifurcation diagram centered on the focal SNP. Haplotypes extend approximately 50 Kb in either direction from the center SNP. Haplotype frequency correlates with line thickness.
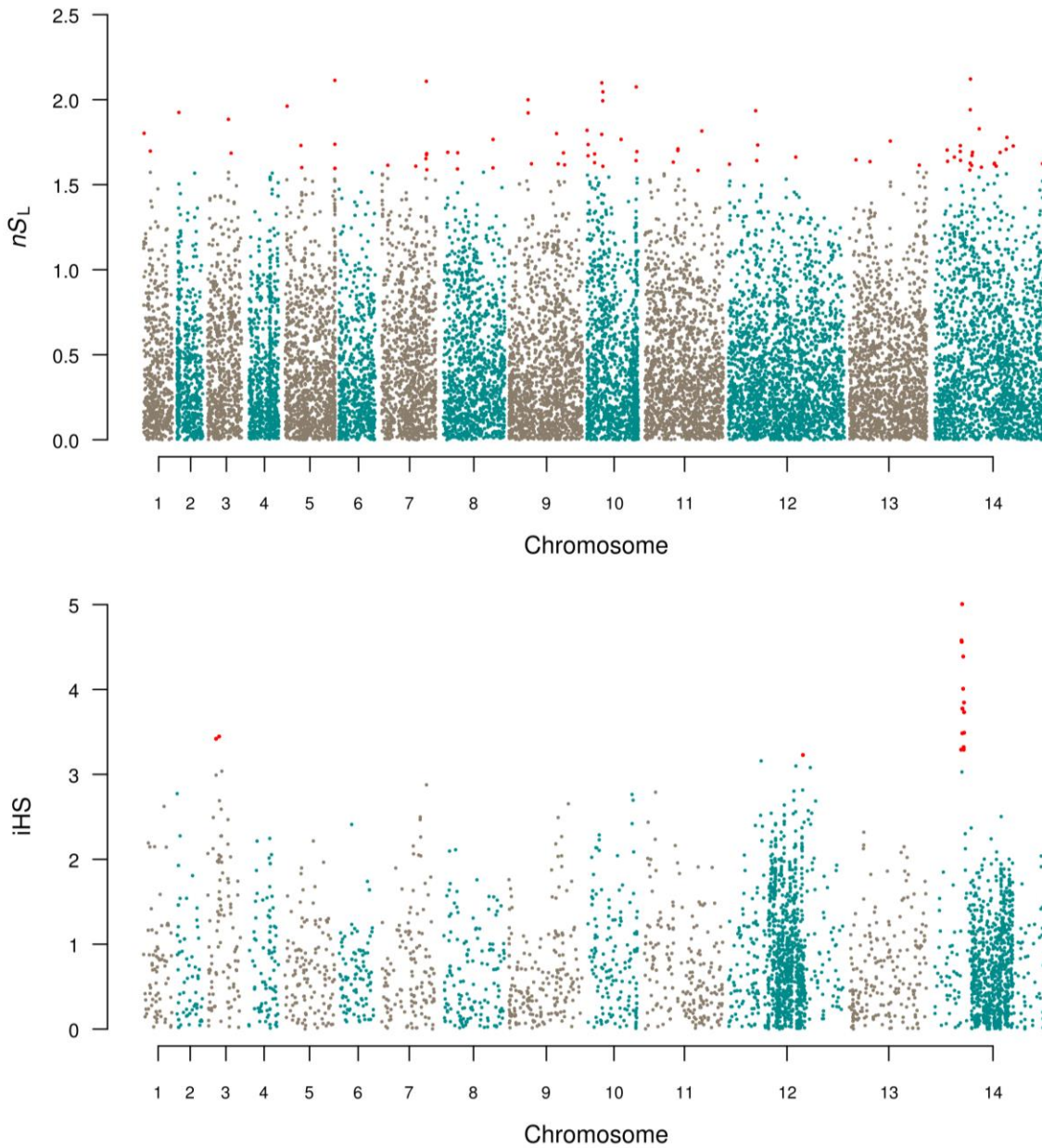
Figure 4.9 Additional haplotype-based scans for directional selection in *P. vivax*.

The upper panel depicts results of nS_L analysis for the entire *P. vivax* sample. As many *P. vivax* infections have a high multiplicity of infection, haplotype-based test results are muted. Nevertheless, many of the same loci as in the monoclonal analysis are identified. The lower panel depicts results of iHS analysis applied to the *P. vivax* monoclonal samples. The strongest signal matches that in the nS_L monoclonal analysis, with the focal variant occurring in close proximity to an AP2-domain transcription factor.

Genome-wide Tajima's D
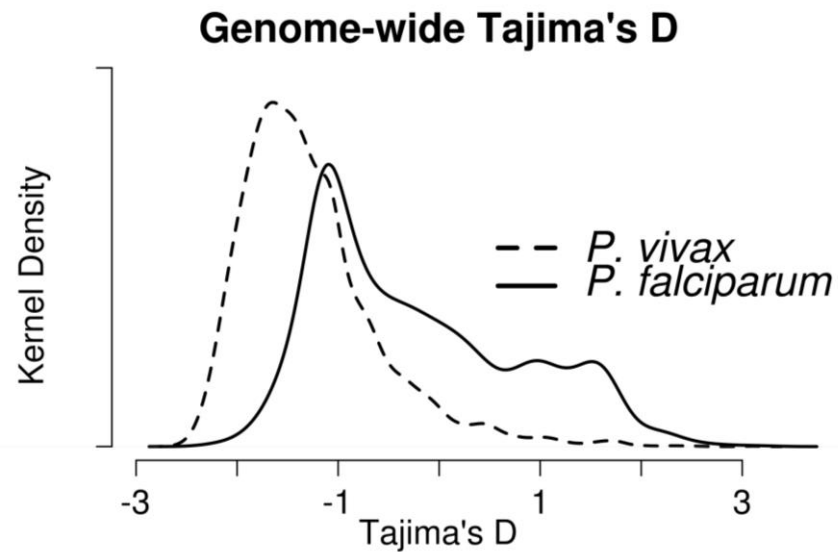
Figure 4.10 Distributions of *P. vivax* and *P. falciparum* gene-wise Tajima's *D* values.

Tajima's *D* was calculated for each gene in the entire *P. vivax* and *P. falciparum* population set. Both distributions have a median value $D < 0$, and a strong right skew. The negative median values may reflect the history of population expansion. The strong right skew is likely driven by a small subset of genes under strong balancing selection. In the distribution of *P. falciparum* values, there is a strong overrepresentation of *D* values in the range from 0.0 to 2.0. This is due to "balanced" allele frequencies at many loci caused multiple discrete subpopulations.
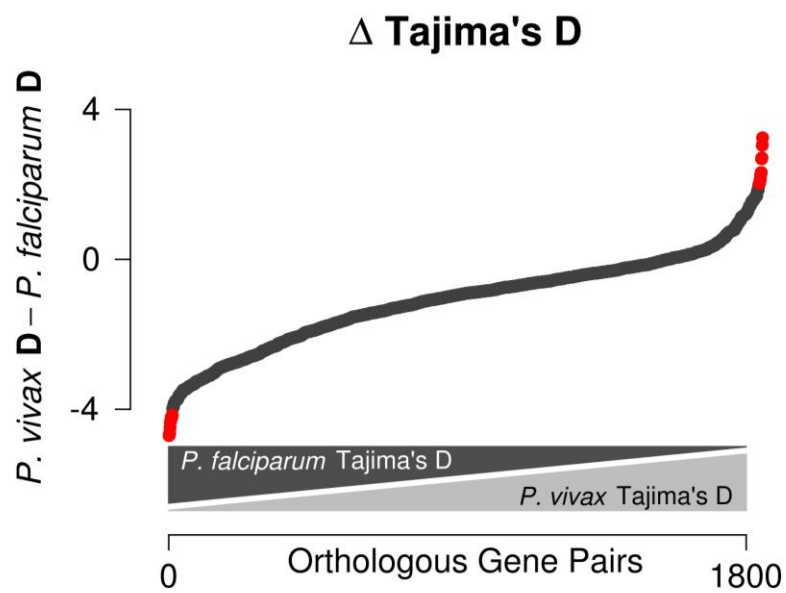
Figure 4.11 Comparison of Tajima's *D* values between *P. vivax* and *P. falciparum* orthologs.

For the 1852 *P. falciparum* and *P. vivax* genes with a single cross-species ortholog, the difference in Tajima's *D* value was calculated to identify ortholog pairs with discordant signatures of selection. The top 1% of orthologous gene pairs with opposing signals are highlighted.

CHAPTER 5: DISCUSSION AND FUTURE DIRECTIONS

5.1   Principal Findings

P. vivax and *P. falciparum* are pathogens with many genetic mysteries yet to solve. Through these studies, we have provided compelling evidence regarding the population-level genetic similarities and differences between these two *Plasmodia* species. In our first study, we deep sequenced the (major surface antigen 1) *pvmsp1* and (circumsporozoite protein) *pvcsp* genes from Cambodian clinical isolates [280]. We compared the genetic diversity in Cambodian *pvmsp1* and *pvcsp* sequence sets with previously published *pvmsp1* and *pvcsp* sequence sets from around the world, finding markedly different patterns of geospatial genetic diversity at these loci. Notably, *pvcsp* sequences clustered continent - or even country - while *pvmsp1* sequences from around the world were intermixed, similar to a previous study of another blood-stage antigen, *pvama1* [145]. Not only do these patterns reflect the differing biology and selective pressures that shape these loci, but they also highlight the need for thoughtful marker selection and for caution when drawing conclusions about population structure and differentiation from genetic markers. In this study, we identified a region of dramatically elevated genetic diversity in *pvmsp1*, and have used this region as a tool in several studies ([56] and unpublished). Finally, orthologous CSP domains are under opposite patterns of selection in *P. vivax* and *P. falciparum* [140], demonstrating that a true understanding of *P. vivax* biology cannot be inferred from *P. falciparum*, but rather will require *P. vivax*-specific investigation.

In our second study, we whole-genome sequenced nearly 80 *P. falciparum* clinical isolates from Cambodia and compared patterns of genetic diversity with patterns of measured ACT partner-drug resistance. We found that the Cambodian *P. falciparum* population is fractured into at least four subpopulations, which have different artemisinin and partner-drug resistance profiles. We found that ACT partner-drug resistance to piperaquine and mefloquine have arisen on different genetic backgrounds. We

also found that infections with parasites of one rapidly expanding near-clonal subpopulations of *P. falciparum* (CP4) have a much higher treatment-failure rate than infections with parasites from other subpopulations. These observations may have implications for malaria treatment policy in Cambodia. We have provided a genetic explanation for the observation that DHA-PPQ treatment failures can be treated successfully with AS-MQ [28]. In the short-term, triple-artemisinin therapy (TACT) that incorporates an artemisinin derivative, PPQ, and MQ may be advisable, and such studies are now underway [281]. In the long-term however, TACTs may hasten the onset of even more multidrug-resistant malaria. Ultimately, short of malaria elimination or an effective vaccine, novel antimalarials that are introduced as carefully controlled combination therapies will be the best solution for malaria drug-resistance.

In our third study, we compared these Cambodian *P. falciparum* clinical isolates to sympatric *P. vivax* clinical isolates. It was previously known that *P. falciparum* cases are declining in Cambodia, even as *P. vivax* cases are increasing [8]. We found genetic evidence in support of these epidemiological observations: *P. vivax* had a higher effective population size and was more admixed compared to *P. falciparum*, suggesting stronger, more sustained transmission. We also found that *P. falciparum* and *P. vivax* seem to have different response mechanisms to selective pressures. While both *P. falciparum* and *P. vivax* has responded with copy-number duplications or mutations in drug-resistance genes, the *P. vivax* population has seen several selective sweeps near transcriptional regulators - transcription factors (with an AP2 domain) and chromatin modifiers (with a SET domain). Thus, while both species rely on simple "drug-resistance variations," *P. vivax* may alter its gene-expression profile, though it is currently unclear which transcripts are affected.

In conclusion, these three studies reveal dramatic differences in *P. vivax* and *P. falciparum* at multiple levels (transcriptional, population structure, response to interventions, selection at antigens). They support the conclusions that *P. vivax* may be more difficult to eliminate than *P. falciparum* and that effective *P. vivax* control will require tailored interventions.

5.2   Primary Limitations and Strengths

A primary limitation of this work is that it has been observational. While these data are raw material for many hypotheses, drawing causal links between observed genetic signatures and biological processes is tenuous at best. *In vitro* experiments to test our hypotheses and further explore our conclusions are needed. An elegant example of how population genetics and molecular epidemiology can direct laboratory experiments is seen in the recent discovery of *P. falciparum* kelch K13-domain mutations. Ariey *et al.* demonstrated - primarily observationally - that specific mutations in this domain are associated with increased artemisinin resistance [39]. However, to confirm these conclusions, classical transgenic *in vitro* experiments were needed. The following year, a separate group provided such evidence by engineering K13-domain mutations into and out of *P. falciparum* laboratory strains, and observing the expected changes in artemisinin-resistance profiles [40]. Our work thus far has generated many hypotheses regarding organismal differences between *P. vivax* and *P. falciparum* biology, but rigorous experimental confirmation is necessary.

The entirety of our work has been performed using Cambodian clinical malaria cases. This has afforded us a detailed view into the population genetics of *Plasmodia* in Cambodia. Of the 106 nations with active malaria transmission [1], Cambodia may host the most fascinating malaria genetics, due to both the *in situ* emergence of artemisinin resistance (and resistance to most other antimalarials) and a recent history of intense malaria control interventions. It seems likely that the recently colorful genetic history of Cambodian *Plasmodia* may limit the generalizability of our findings. However, time after time we see reported in the literature the same broad population-level trends we have observed in Cambodia - specifically that, suggesting that our findings may have a degree of generalizability [282–287]. These reports describe how, global co-endemic *P. falciparum* populations tend to be clonal with episodic transmission and structure-by-geography, while *P. vivax* populations enjoy a higher effective population size, more stable transmission, and increased gene flow between geographic islands.

A limitation for our work - and for the broader field of malaria genetics - is the lack of appropriate models for *Plasmodium* genetic variation [288]. While most tools for demographic inference

or modeling make Wright-Fisher assumptions, free-living *Plasmodium* populations clearly violate these assumptions, especially with respect to non-overlapping generations and a finite, constant population size. For example, it is unclear how to take a *Plasmodium spp.* census - does one estimate population size when parasite numbers are increasing in log-linear fashion during the blood stage? Or during the bottleneck that occurs through host-to-vector (or vector-to-host) transmission? Or during the within-vector meiotic stage? While more appropriate models are being developed [289], the field is at present left without a satisfactory tool. Genetic models have been important for our work - from estimating the significance of observed Tajima's *D* values using coalescent simulations [280] to drawing inferences regarding the demographic history of *P. falciparum* and *P. vivax*. The fact that the models we have used may not appropriately describe *Plasmodia* biology could compromise the quantitative or qualitative integrity of our conclusions.

Despite these limitations, our work represents a significant advance for the field of malaria genetics. In collaboration with co-authors from our first publication (see [280]), we have led the field in developing a framework for generating, analyzing, and interpreting malaria antigen deep-sequencing data. We have also contributed crucial data regarding how *P. falciparum* population structure interfaces with ACT partner-drug resistance. Finally, in undertaking the first genome-by-population study of co-endemic *P. vivax* and *P. falciparum*, we now have solid hypotheses for how these species differ in their responses to malaria control and elimination efforts, and these results will provide a compass for future studies.

5.3 Unanswered Questions and Future Directions

Our studies beg additional investigation. Regarding ACT partner-drug resistance among Cambodian *P. falciparum*, there are still many unknowns. It is unclear whether mefloquine (MQ) and piperaquine (PPQ) resistance are antagonistic to one another, or whether they have - by chance - occurred on different genetic backgrounds. The roughly inverse correlation between MQ and PPQ IC50s that we observed raises the possibility that antagonism exists; however, occasional parasites with elevated MQ and PPQ IC50s suggest otherwise. Previous studies have demonstrated antagonism between PPQ and MQ *in vitro* [290], but it is unclear whether this translates into resistance antagonism. Additional population

studies could shed light. If PPQ and MQ resistance are antagonistic, one would expect the maintenance of these distinct population subgroups, with contrasting ACT partner-drug resistance profiles. *In vitro* and clinical surveillance studies could also increase our understanding. If an antagonistic relationship does exist, *in vitro* studies will be necessary to define the molecular nature of this relationship.

We have also presented the first genome-by-population study of co-endemic *P. falciparum* and *P. vivax*. A surprising discovery was the stark differences in the types of genes driving selective sweeps in *P. falciparum* and *P. vivax* malaria. To date, *P. falciparum* studies have demonstrated that the strongest signatures of selection are nearby multidrug transporters, antimalarial targets, and antigens [291–295]. In contrast, in our *P. vivax* population sample, we observed the strongest selective sweeps in close proximity to multiple AP2- and SET-domain containing proteins. Both of these classes of proteins are likely transcriptional regulators, raising the possibility that *P. vivax* is responding to selective pressures by altering its transcriptional profile.

Transcriptional regulation by pathogens in response to sustained drug pressure has been observed in fungal pathogens [296]. In our *P. vivax* population, it is crucial to first identify the selected allele at these AP2-domain and SET-domain loci, then to determine their transcriptional effects. RNA-Seq and ChIP-Seq experiments may demonstrate that these alleles affect transcription of drug-resistance transporters, or they may hint at an undescribed mechanism. Absent a reliable *P. vivax* culture system, these studies may be pursued in *P. knowlesi* culture [297]. We already have some familiarity with these studies. Unpublished work with a collaborator has shown that mutations in the AP2 DNA-binding domain of a *P. falciparum* transcription factor are associated with escape from antibody pressure. Work is ongoing to determine whether these mutations affect transcriptional patterns. Short of performing these difficult experiments in *P. vivax* or *P. knowlesi* culture, we could search for cis-acting regulatory elements that may be targets of AP2-binding domain transcription factors, to at least determine which genes might be influenced by these alleles.

In fungal eukaryotic pathogens, sustained drug pressure can also cause chromosomal rearrangements, which in turn drive gene expression patterns [296]. It would be interesting to explore *P.*

vivax structural rearrangements as a compensatory mechanism for drug pressure. With our whole-genome sequencing data for both *P. falciparum* and *P. vivax*, we have an opportunity to perform the first genome-wide investigation of structural variants in these parasites. Our data may be especially useful for such studies because - in contrast to many population-level genome sequencing projects - our isolates were sequenced directly, rather than being subject to capture. A preliminary look has revealed some large and unexpected structural rearrangements in *P. vivax* isolates. As more genome-wide data become available for both *P. vivax* and *P. falciparum*, these questions will be answered in a systematic way.

REFERENCES

1. World Health Organization. World Malaria Report 2015. World Health Organization; 2015.

2. Gallup JL, Sachs JD. The economic burden of malaria. Am J Trop Med Hyg. 2001;64: 85–96.

3. Murray CJL, Rosenfeld LC, Lim SS, Andrews KG, Foreman KJ, Haring D, et al. Global malaria mortality between 1980 and 2010: a systematic analysis. Lancet. 2012;379: 413–431.

4. Schantz-Dunn J, Nour NM. Malaria and pregnancy: a global health perspective. Rev Obstet Gynecol. 2009;2: 186–192.

5. Erhart A, Ngo DT, Phan VK, Ta TT, Van Overmeir C, Speybroeck N, et al. Epidemiology of forest malaria in central Vietnam: a large scale cross-sectional survey. Malar J. 2005;4: 58.

6. Thanh P, Pham T, Van Hong N, Van Van N, Van Malderen C, Valérie O, et al. Epidemiology of forest malaria in Central Vietnam: the hidden parasite reservoir. Malar J. 2015;14: 86.

7. Bleakley H. Health, Human Capital, and Development. Annu Rev Econom. 2010;2: 283–310.

8. Maude RJ, Chea N, Po L, Tol B, Pengby N, de la Torre SEC, et al. Spatial and temporal epidemiology of clinical malaria in Cambodia 2004–2013. Malar J. 2014;13: 385.

9. Global Fund Statement on Cambodia's Programs against Malaria. In: www.theglobalfund.org [Internet]. 20 Oct 2015. Available: http://www.theglobalfund.org/en/news/2015-10-20_Global_Fund_Statement_on_Cambodia_s_Programs_against_Malaria/

10. Prudêncio M, Rodriguez A, Mota MM. The silent path to thousands of merozoites: the Plasmodium liver stage. Nat Rev Microbiol. 2006;4: 849–856.

11. Doolan DL, Dobaño C, Baird JK. Acquired immunity to malaria. Clin Microbiol Rev. 2009;22: 13–36, Table of Contents.

12. Cullen KA, Mace KE, Arguin PM. Malaria Surveillance — United States, 2013. MMWR Surveill Summ. 2016;65: 1–22.

13. Taylor SM, Molyneux ME, Simel DL, Meshnick SR, Juliano JJ. Does This Patient Have Malaria? JAMA. 2010;304. doi:10.1001/jama.2010.1578

14. Schwartz E, Parise M, Kozarsky P, Cetron M. Delayed Onset of Malaria — Implications for Chemoprophylaxis in Travelers. N Engl J Med. 2003;349: 1510–1516.

15. Guidelines for the Treatment of Malaria. Geneva: World Health Organization; 2015.

16. Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U, et al. The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. Nature. 2015;526: 207–211.

17. Agnandji ST, Lell B, Soulanoudjingar SS, Fernandes JF, Abossolo BP, Conzelmann C, et al. First results of phase 3 trial of RTS,S/AS01 malaria vaccine in African children. N Engl J Med. 2011;365: 1863–1875.

18. The R, S Clinical Trials Partnership. A Phase 3 Trial of RTS,S/AS01 Malaria Vaccine in African Infants. N Engl J Med. 2012;367: 2284–2295.

19. Thera MA, Doumbo OK, Coulibaly D, Laurens MB, Ouattara A, Kone AK, et al. A field trial to assess a blood-stage malaria vaccine. N Engl J Med. 2011;365: 1004–1013.

20. Neafsey DE, Juraska M, Bedford T, Benkeser D, Valim C, Griggs A, et al. Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine. N Engl J Med. 2015;373: 2025–2037.

21. Raj DK, Nixon CP, Nixon CE, Dvorin JD, DiPetrillo CG, Pond-Tor S, et al. Antibodies to PfSEA-1 block parasite egress from RBCs and protect against malaria infection. Science. 2014;344: 871–877.

22. Douglas AD, Baldeviano GC, Lucas CM, Lugo-Roman LA, Crosnier C, Bartholdson SJ, et al. A PfRH5-based vaccine is efficacious against heterologous strain blood-stage Plasmodium falciparum infection in aotus monkeys. Cell Host Microbe. 2015;17: 130–139.

23. Doolan DL, Yunxiang M, Berkay U, Suman S, Siddiqua H, Conrad V, et al. Profiling humoral immune responses to P. falciparum infection with protein microarrays. Proteomics. 2008;8: 4680–4694.

24. Crompton PD, Kayala MA, Traore B, Kayentao K, Ongoiba A, Weiss GE, et al. A prospective analysis of the Ab response to Plasmodium falciparum before and after a malaria season by protein microarray. Proc Natl Acad Sci U S A. 2010;107: 6958–6963.

25. Dondorp AM, François N, Poravuth Y, Debashish D, Phyo AP, Joel T, et al. Artemisinin Resistance in Plasmodium falciparum Malaria. N Engl J Med. 2009;361: 455–467.

26. Meshnick S. Perspective: artemisinin-resistant malaria and the wolf. Am J Trop Med Hyg. 2012;87: 783–784.

27. White NJ. Counter perspective: artemisinin resistance: facts, fears, and fables. Am J Trop Med Hyg. 2012;87: 785.

28. Spring MD, Lin JT, Manning JE, Vanachayangkul P, Somethy S, Bun R, et al. Dihydroartemisinin-piperaquine failure associated with a triple mutant including kelch13 C580Y in Cambodia: an observational cohort study. Lancet Infect Dis. 2015;15: 683–691.

29. Amaratunga C, Chanaki A, Pharath L, Seila S, Sokunthea S, Sivanna M, et al. Dihydroartemisinin–piperaquine resistance in Plasmodium falciparum malaria in Cambodia: a multisite prospective cohort study. Lancet Infect Dis. 2016;16: 357–365.

30. World Health Organization. Global Plan for Artemisinin Resistance Containment. World Health Organization; 2011.

31. World Health Organization. Update on Artemisinin Resistance - April 2012. World Health Organization; 2012.

32. Ashley EA, Mehul D, Fairhurst RM, Chanaki A, Parath L, Seila S, et al. Spread of Artemisinin Resistance in Plasmodium falciparum Malaria. N Engl J Med. 2014;371: 411–423.

33. White NJ, Sasithon P, Phyo AP, Ronnatrai R, François N, Podjanee J, et al. Spiroindolone KAE609 for Falciparum and Vivax Malaria. N Engl J Med. 2014;371: 403–410.

34. Phillips MA, Lotharius J, Marsh K, White J, Dayan A, White KL, et al. A long-duration dihydroorotate dehydrogenase inhibitor (DSM265) for prevention and treatment of malaria. Sci Transl Med. 2015;7: 296ra111.

35. Phyo AP, Jittamala P, Nosten FH, Pukrittayakamee S, Imwong M, White NJ, et al. Antimalarial activity of artefenomel (OZ439), a novel synthetic antimalarial endoperoxide, in patients with Plasmodium falciparum and Plasmodium vivax malaria: an open-label phase 2 trial. Lancet Infect Dis. 2016;16: 61–69.

36. Cheeseman IH, Miller BA, Nair S, Nkhoma S, Tan A, Tan JC, et al. A Major Genome Region Underlying Artemisinin Resistance in Malaria. Science. 2012;336: 79–82.

37. Takala-Harrison S, Clark TG, Jacob CG, Cummings MP, Miotto O, Dondorp AM, et al. Genetic loci associated with delayed clearance of Plasmodium falciparum following artemisinin treatment in Southeast Asia. Proc Natl Acad Sci U S A. 2013;110: 240–245.

38. Miotto O, Almagro-Garcia J, Manske M, Macinnis B, Campino S, Rockett KA, et al. Multiple populations of artemisinin-resistant Plasmodium falciparum in Cambodia. Nat Genet. 2013;45: 648–655.

39. Ariey F, Witkowski B, Amaratunga C, Beghain J, Langlois A-C, Khim N, et al. A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. Nature. 2014;505: 50–55.

40. Straimer J, Gnädig NF, Witkowski B, Amaratunga C, Duru V, Ramadani AP, et al. Drug resistance. K13-propeller mutations confer artemisinin resistance in Plasmodium falciparum clinical isolates. Science. 2015;347: 428–431.

41. Singh A, Misra V, Thimmulappa RK, Lee H, Ames S, Hoque MO, et al. Dysfunctional KEAP1-NRF2 interaction in non-small-cell lung cancer. PLoS Med. 2006;3: e420.

42. Mbengue A, Alassane M, Souvik B, Trupti P, Haining L, Guillermina E, et al. A molecular mechanism of artemisinin resistance in Plasmodium falciparum malaria. Nature. 2015;520: 683–687.

43. Mok S, Ashley EA, Ferreira PE, Zhu L, Lin Z, Yeo T, et al. Drug resistance. Population transcriptomics of human malaria parasites reveals the mechanism of artemisinin resistance. Science. 2015;347: 431–435.

44. Takala-Harrison S, Jacob CG, Arze C, Cummings MP, Silva JC, Dondorp AM, et al. Independent emergence of artemisinin resistance mutations among Plasmodium falciparum in Southeast Asia. J Infect Dis. 2015;211: 670–679.

45. Taylor SM, Parobek CM, DeConti DK, Kayentao K, Coulibaly SO, Greenwood BM, et al. Absence of putative artemisinin resistance mutations among Plasmodium falciparum in Sub-Saharan Africa: a molecular epidemiologic study. J Infect Dis. 2015;211: 680–688.

46. Kamau E, Campino S, Amenga-Etego L, Drury E, Ishengoma D, Johnson K, et al. K13-propeller polymorphisms in Plasmodium falciparum parasites from sub-Saharan Africa. J Infect Dis. 2015;211: 1352–1355.

47. MalariaGEN Plasmodium falciparum Community Project. Genomic epidemiology of artemisinin resistant malaria. Elife. 2016;5. doi:10.7554/eLife.08714

48. Miotto O, Amato R, Ashley EA, MacInnis B, Almagro-Garcia J, Amaratunga C, et al. Genetic architecture of artemisinin-resistant Plasmodium falciparum. Nat Genet. 2015;47: 226–234.

49. Molina-Cruz A, Garver LS, Alabaster A, Bangiolo L, Haile A, Winikor J, et al. The human malaria parasite Pfs47 gene mediates evasion of the mosquito immune system. Science. 2013;340: 984–987.

50. Chaorattanakawee S, Saunders DL, Sea D, Chanarat N, Yingyuen K, Sundrakes S, et al. Ex Vivo Drug Susceptibility Testing and Molecular Profiling of Clinical Plasmodium falciparum Isolates from Cambodia from 2008 to 2013 Suggest Emerging Piperaquine Resistance. Antimicrob Agents Chemother. 2015;59: 4631–4643.

51. Reed MB, Saliba KJ, Caruana SR, Kirk K, Cowman AF. Pgh1 modulates sensitivity and resistance to multiple antimalarials in Plasmodium falciparum. Nature. 2000;403: 906–909.

52. Wongsrichanalai C, Chansuda W, Meshnick SR. Declining Artesunate-Mefloquine Efficacy against Falciparum Malaria on the Cambodia–Thailand Border. Emerg Infect Dis. 2008;14: 716–719.

53. Eastman RT, Dharia NV, Winzeler EA, Fidock DA. Piperaquine resistance is associated with a copy number variation on chromosome 5 in drug-pressured Plasmodium falciparum parasites. Antimicrob Agents Chemother. 2011;55: 3908–3916.

54. Baird JK. Evidence and implications of mortality associated with acute Plasmodium vivax malaria. Clin Microbiol Rev. 2013;26: 36–57.

55. Popovici J, Kao S, Eal L, Bin S, Kim S, Ménard D. Reduced polymorphism in the Kelch propeller domain in Plasmodium vivax isolates from Cambodia. Antimicrob Agents Chemother. 2015;59: 730–733.

56. Lin JT, Patel JC, Kharabora O, Sattabongkot J, Muth S, Ubalee R, et al. 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. 2013;88: 1116–1123.

57. Mueller I, Galinski MR, Baird JK, Carlton JM, Kochar DK, Alonso PL, et al. Key gaps in the knowledge of Plasmodium vivax, a neglected human malaria parasite. Lancet Infect Dis. 2009;9: 555–566.

58. Tjitra E, Anstey NM, Sugiarto P, Warikar N, Kenangalem E, Karyana M, et al. Multidrug-resistant Plasmodium vivax associated with severe and fatal malaria: a prospective study in Papua, Indonesia. PLoS Med. 2008;5: e128.

59. Price RN, Douglas NM, Anstey NM. New developments in Plasmodium vivax malaria: severe disease and the rise of chloroquine resistance. Curr Opin Infect Dis. 2009;22: 430–435.

60. Marfurt J, Jutta M, de Monbrison F, Sara B, Laetitia B, Ivo M, et al. Molecular Markers of In Vivo Plasmodium vivax Resistance to Amodiaquine Plus Sulfadoxine-Pyrimethamine: Mutations in pvdhfr and pvmdr1. J Infect Dis. 2008;198: 409–417.

61. Greenwood B, Targett G. Do we still need a malaria vaccine? Parasite Immunol. 2009;31: 582–586.

62. PATH Malaria Vaccine Initiative. Staying the course? Malaria research and development in a time of economic uncertainty. 2011.

63. Arévalo-Herrera M, Chitnis C, Herrera S. Current status of Plasmodium vivax vaccine. Hum Vaccin. 2010;6: 124–132.

64. Espinosa AM, Sierra AY, Barrero CA, Cepeda LA, Cantor EM, Lombo TB, et al. Expression, polymorphism analysis, reticulocyte binding and serological reactivity of two Plasmodium vivax MSP-1 protein recombinant fragments. Vaccine. 2003;21: 1033–1043.

65. Collins WE, Kaslow DC, Sullivan JS, Morris CL, Galland GG, Yang C, et al. Testing the efficacy of a recombinant merozoite surface protein (MSP-1(19) of Plasmodium vivax in Saimiri boliviensis monkeys. Am J Trop Med Hyg. 1999;60: 350–356.

66. Blackman MJ, Heidrich HG, Donachie S, McBride JS, Holder AA. A single fragment of a malaria merozoite surface protein remains on the parasite during red cell invasion and is the target of invasion-inhibiting antibodies. J Exp Med. 1990;172: 379–382.

67. Guevara Patiño JA, Holder AA, McBride JS, Blackman MJ. Antibodies that inhibit malaria merozoite surface protein-1 processing and erythrocyte invasion are blocked by naturally acquired human antibodies. J Exp Med. 1997;186: 1689–1699.

68. Udhayakumar V, Anyona D, Kariuki S, Shi YP, Bloland PB, Branch OH, et al. Identification of T and B cell epitopes recognized by humans in the C-terminal 42-kDa domain of the Plasmodium falciparum merozoite surface protein (MSP)-1. J Immunol. 1995;154: 6022–6030.

69. Nwuba RI, Sodeinde O, Anumudu CI, Omosun YO, Odaibo AB, Holder AA, et al. The human immune response to Plasmodium falciparum includes both antibodies that inhibit merozoite surface protein 1 secondary processing and blocking antibodies. Infect Immun. 2002;70: 5328–5331.

70. Sultan AA. Molecular mechanisms of malaria sporozoite motility and invasion of host cells. Int Microbiol. 1999;2: 155–160.

71. Takala SL, Plowe CV. Genetic diversity and malaria vaccine design, testing and efficacy: preventing and overcoming "vaccine resistant malaria." Parasite Immunol. 2009;31: 560–573.

72. Genton B, Betuela I, Felger I, Al-Yaman F, Anders RF, Saul A, et al. A recombinant blood-stage malaria vaccine reduces Plasmodium falciparum density and exerts selective pressure on parasite populations in a phase 1-2b trial in Papua New Guinea. J Infect Dis. 2002;185: 820–827.

73. Pattaradilokrat S, Cheesman SJ, Carter R. Linkage group selection: towards identifying genes controlling strain specific protective immunity in malaria. PLoS One. 2007;2: e857.

74. Martinelli A, Cheesman S, Hunt P, Culleton R, Raza A, Mackinnon M, et al. A genetic approach to the de novo identification of targets of strain-specific immunity in malaria parasites. Proc Natl Acad Sci U S A. 2005;102: 814–819.

75. Enosse S, Dobaño C, Quelhas D, Aponte JJ, Lievens M, Leach A, et al. RTS,S/AS02A malaria vaccine does not induce parasite CSP T cell epitope selection and reduces multiplicity of infection. PLoS Clin Trials. 2006;1: e5.

76. Kumkhaek C, Phra-Ek K, Rénia L, Singhasivanon P, Looareesuwan S, Hirunpetcharat C, et al. Are extensive T cell epitope polymorphisms in the Plasmodium falciparum circumsporozoite antigen, a leading sporozoite vaccine candidate, selected by immune pressure? J Immunol. 2005;175: 3935–3939.

77. Alloueche A, Milligan P, Conway DJ, Pinder M, Bojang K, Doherty T, et al. Protective efficacy of the RTS,S/AS02 Plasmodium falciparum malaria vaccine is not strain specific. Am J Trop Med Hyg. 2003;68: 97–101.

78. Abdulla S, Salim N, Machera F, Kamata R, Juma O, Shomari M, et al. Randomized, controlled trial of the long term safety, immunogenicity and efficacy of RTS,S/AS02D malaria vaccine in infants

living in a malaria-endemic region. Malar J. 2013;12: 11.

79. Olotu A, Fegan G, Wambua J, Nyangweso G, Awuondo KO, Leach A, et al. Four-year efficacy of RTS,S/AS01E and its interaction with malaria exposure. N Engl J Med. 2013;368: 1111–1120.

80. Lin JT, Juliano JJ, Kharabora O, Sem R, Lin F-C, Muth S, et al. Individual Plasmodium vivax msp1 variants within polyclonal P. vivax infections display different propensities for relapse. J Clin Microbiol. 2012;50: 1449–1451.

81. Rogers WO, Sem R, Tero T, Chim P, Lim P, Muth S, et al. Failure of artesunate-mefloquine combination therapy for uncomplicated Plasmodium falciparum malaria in southern Cambodia. Malar J. 2009;8: 10.

82. Givens MB, Lin JT, Lon C, Gosi P, Char MC, Lanteri CA, et al. Development of a capillary electrophoresis-based heteroduplex tracking assay to measure in-host genetic diversity of initial and recurrent Plasmodium vivax infections in Cambodia. J Clin Microbiol. 2014;52: 298–301.

83. Ngrenngarmlert W, Kwiek JJ, Kamwendo DD, Ritola K, Swanstrom R, Wongsrichanalai C, et al. Measuring allelic heterogeneity in Plasmodium falciparum by a heteroduplex tracking assay. Am J Trop Med Hyg. 2005;72: 694–701.

84. Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, et al. An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature. 2000;407: 513–516.

85. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9: 357–359.

86. Colwell RK, Mao CX, Jing C. INTERPOLATING, EXTRAPOLATING, AND COMPARING INCIDENCE-BASED SPECIES ACCUMULATION CURVES. Ecology. 2004;85: 2717–2727.

87. Colwell RK, Chao A, Gotelli NJ, S.-Y. L, Mao CX, Chazdon RL, et al. Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. Journal of Plant Ecology. 2012;5: 3–21.

88. Kang J-M, Ju H-L, Kang Y-M, Lee D-H, Moon S-U, Sohn W-M, et al. Genetic polymorphism and natural selection in the C-terminal 42 kDa region of merozoite surface protein-1 among Plasmodium vivax Korean isolates. Malar J. 2012;11: 206.

89. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25: 1451–1452.

90. Hudson RR. Estimating the recombination parameter of a finite population model without selection. Genet Res. 1987;89: 427–432.

91. McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in Drosophila. Nature. 1991;351: 652–654.

92. Jongwutiwes S, Buppan P, Kosuvin R, Seethamchai S, Pattanawong U, Sirichaisinthop J, et al. Plasmodium knowlesi Malaria in humans and macaques, Thailand. Emerg Infect Dis. 2011;17: 1799–1806.

93. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary

genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28: 2731–2739.

94. Hughes AL. The evolution of amino acid repeat arrays in Plasmodium and other organisms. J Mol Evol. 2004;59: 528–535.

95. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol. 1986;3: 418–426.

96. Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online. 2005;1: 47–50.

97. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 2004;20: 289–290.

98. Bérard S, Nicolas F, Buard J, Gascuel O, Rivals E. A fast and specific alignment method for minisatellite maps. Evol Bioinform Online. 2006;2: 303–320.

99. Bérard S, Rivals E. Comparison of minisatellites. J Comput Biol. 2003;10: 357–372.

100. Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci U S A. 2004;101: 11030–11035.

101. Desper R, Gascuel O. Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. J Comput Biol. 2002;9: 687–705.

102. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4: 406–425.

103. Hudson RR. A new statistic for detecting genetic differentiation. Genetics. 2000;155: 2011–2014.

104. Hamady M, Lozupone C, Knight R. Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. ISME J. 2010;4: 17–27.

105. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999;16: 37–48.

106. Jongwutiwes S, Putaporntip C, Hughes AL. Bottleneck effects on vaccine-candidate antigen diversity of malaria parasites in Thailand. Vaccine. 2010;28: 3112–3117.

107. Putaporntip C, Jongwutiwes S, Sakihama N, Ferreira MU, Kho W-G, Kaneko A, et al. Mosaic organization and heterogeneity in frequency of allelic recombination of the Plasmodium vivax merozoite surface protein-1 locus. Proc Natl Acad Sci U S A. 2002;99: 16348–16353.

108. Thakur A, Alam MT, Sharma YD. Genetic diversity in the C-terminal 42 kDa region of merozoite surface protein-1 of Plasmodium vivax (PvMSP-1(42)) among Indian isolates. Acta Trop. 2008;108: 58–63.

109. Han E-T, Wang Y, Lim CS, Cho JH, Chai J-Y. Genetic diversity of the malaria vaccine candidate merozoite surface protein 1 gene of Plasmodium vivax field isolates in Republic of Korea. Parasitol Res. 2011;109: 1571–1576.

110.   Zeyrek FY, S.-I. T, Yuksel F, Doni N, Palacpac N, Arisue N, et al. Limited Polymorphism of the Plasmodium vivax Merozoite Surface Protein 1 Gene in Isolates from Turkey. Am J Trop Med Hyg. 2010;83: 1230–1237.

111.   Weedall GD, Conway DJ. Detecting signatures of balancing selection to identify targets of anti-parasite immunity. Trends Parasitol. 2010;26: 363–369.

112.   Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;123: 585–595.

113.   Chenet SM, Tapia LL, Escalante AA, Durand S, Lucas C, Bacon DJ. Genetic diversity and population structure of genes encoding vaccine candidate antigens of Plasmodium vivax. Malar J. 2012;11: 68.

114.   Hudson RR, Slatkin M, Maddison WP. Estimation of levels of gene flow from DNA sequence data. Genetics. 1992;132: 583–589.

115.   Carneiro MO, Carsten R, Ross MG, Gabriel SB, Nusbaum C, DePristo MA. Pacific biosciences sequencing technology for genotyping and variation discovery in human data. BMC Genomics. 2012;13: 375.

116.   Lim CS, Tazi L, Ayala FJ. Plasmodium vivax: recent world expansion and genetic identity to Plasmodium simium. Proc Natl Acad Sci U S A. 2005;102: 15523–15528.

117.   Henry-Halldin CN, Sepe D, Susapu M, McNamara DT, Bockarie M, King CL, et al. High-throughput molecular diagnosis of circumsporozoite variants VK210 and VK247 detects complex Plasmodium vivax infections in malaria endemic populations in Papua New Guinea. Infect Genet Evol. 2011;11: 391–398.

118.   Dias S, Wickramarachchi T, Sahabandu I, Escalante AA, Udagama PV. Population genetic structure of the Plasmodium vivax circumsporozoite protein (Pvcsp) in Sri Lanka. Gene. 2013;518: 381–387.

119.   Hernandez-Martinez MA, Escalante AA, Arevalo-Herrera M, Herrera S. Antigenic Diversity of the Plasmodium vivax Circumsporozoite Protein in Parasite Isolates of Western Colombia. Am J Trop Med Hyg. 2011;84: 51–57.

120.   Patil A, Orjuela-Sánchez P, da Silva-Nunes M, Ferreira MU. Evolutionary dynamics of the immunodominant repeats of the Plasmodium vivax malaria-vaccine candidate circumsporozoite protein (CSP). Infect Genet Evol. 2010;10: 298–303.

121.   Santos-Ciminera PD, Alecrim M das GC, Roberts DR, Quinnan GV Jr. Molecular epidemiology of Plasmodium vivax in the State of Amazonas, Brazil. Acta Trop. 2007;102: 38–46.

122.   Ana L, Ortiz A, Coello J, Sosa-Ochoa W, Torres RE, Banegas EI, et al. Genetic diversity of Plasmodium vivax and Plasmodium falciparum in Honduras. Malar J. 2012;11: 391.

123.   Cerami C, Kwakye-Berko F, Nussenzweig V. Binding of malarial circumsporozoite protein to sulfatides [Gal(3-SO4)beta 1-Cer] and cholesterol-3-sulfate and its dependence on disulfide bond formation between cysteines in region II. Mol Biochem Parasitol. 1992;54: 1–12.

124.   Sinnis P, Clavijo P, Fenyö D, Chait BT, Cerami C, Nussenzweig V. Structural and functional properties of region II-plus of the malaria circumsporozoite protein. J Exp Med. 1994;180: 297–306.

125.   Seth RK, Bhat AA, Rao DN, Biswas S. Acquired immune response to defined Plasmodium vivax antigens in individuals residing in northern India. Microbes Infect. 2010;12: 199–206.

126.   Tetteh KKA, Stewart LB, Ochola LI, Amambua-Ngwa A, Thomas AW, Marsh K, et al. Prospective identification of malaria parasite genes under balancing selection. PLoS One. 2009;4: e5568.

127.   Charlesworth J, Eyre-Walker A. The McDonald-Kreitman test and slightly deleterious mutations. Mol Biol Evol. 2008;25: 1007–1015.

128.   Blackman MJ, Whittle H, Holder AA. Processing of the Plasmodium falciparum major merozoite surface protein-1: identification of a 33-kilodalton secondary processing product which is shed prior to erythrocyte invasion. Mol Biochem Parasitol. 1991;49: 35–44.

129.   Arevalo-Herrera M, Roggero MA, Gonzalez JM, Vergara J, Corradin G. Mapping and comparison of the B cell epitopes recognized on the Plasmodium vivax circumsporozoite protein by immune Colombians and immunized Aotus monkeys. Ann Trop Med Parasitol. 1998;92: 539–551.

130.   Arnot DE, Barnwell JW, Tam JP, Nussenzweig V, Nussenzweig RS, Enea V. Circumsporozoite protein of Plasmodium vivax: gene cloning and characterization of the immunodominant epitope. Science. 1985;230: 815–818.

131.   Herrera S, Escobar P, de Plata C, Avila GI, Corradin G, Herrera MA. Human recognition of T cell epitopes on the Plasmodium vivax circumsporozoite protein. J Immunol. 1992;148: 3986–3990.

132.   Hughes AL. Circumsporozoite protein genes of malaria parasites (Plasmodium spp.): evidence for positive selection on immunogenic regions. Genetics. 1991;127: 345–353.

133.   Nardin E, Clavijo P, Mons B, van Belkum A, Ponnudurai T, Nussenzweig RS. T cell epitopes of the circumsporozoite protein of Plasmodium vivax. Recognition by lymphocytes of a sporozoite-immunized chimpanzee. J Immunol. 1991;146: 1674–1678.

134.   Bowman NM, Congdon S, Mvalo T, Patel JC, Escamilla V, Emch M, et al. Comparative population structure of Plasmodium falciparum circumsporozoite protein NANP repeat lengths in Lilongwe, Malawi. Sci Rep. 2013;3: 1990.

135.   Ferreira MU, Hartl DL. Plasmodium falciparum: worldwide sequence diversity and evolution of the malaria vaccine candidate merozoite surface protein-2 (MSP-2). Exp Parasitol. 2007;115: 32–40.

136.   Escalante AA, Lal AA, Ayala FJ. Genetic polymorphism and natural selection in the malaria parasite Plasmodium falciparum. Genetics. 1998;149: 189–202.

137.   Rich SM, Hudson RR, Ayala FJ. Plasmodium falciparum antigenic diversity: evidence of clonal population structure. Proc Natl Acad Sci U S A. 1997;94: 13040–13045.

138.   Hartl DL. The origin of malaria: mixed messages from genetic diversity. Nat Rev Microbiol. 2004;2: 15–22.

139.   Pacheco MA, Poe AC, Collins WE, Lal AA, Tanabe K, Kariuki SK, et al. A comparative study of the genetic diversity of the 42kDa fragment of the merozoite surface protein 1 in Plasmodium falciparum and P. vivax. Infect Genet Evol. 2007;7: 180–187.

140.   Bailey JA, Mvalo T, Aragam N, Weiser M, Congdon S, Kamwendo D, et al. Use of massively

parallel pyrosequencing to evaluate the diversity of and selection on Plasmodium falciparum csp T-cell epitopes in Lilongwe, Malawi. J Infect Dis. 2012;206: 580–587.

141.   Jalloh A, Jalloh M, Matsuoka H. T-cell epitope polymorphisms of the Plasmodium falciparum circumsporozoite protein among field isolates from Sierra Leone: age-dependent haplotype distribution? Malar J. 2009;8: 120.

142.   Gandhi K, Thera MA, Coulibaly D, Traore K, Guindo AB, Doumbo OK, et al. Next Generation Sequencing to Detect Variation in the Plasmodium falciparum Circumsporozoite Protein. Am J Trop Med Hyg. 2012;86: 775–781.

143.   Weedall GD, Preston BMJ, Thomas AW, Sutherland CJ, Conway DJ. Differential evidence of natural selection on two leading sporozoite stage malaria vaccine candidate antigens. Int J Parasitol. 2007;37: 77–85.

144.   Zarate S, Pond SLK, Shapshak P, Frost SDW. Comparative Study of Methods for Detecting Sequence Compartmentalization in Human Immunodeficiency Virus Type 1. J Virol. 2007;81: 6643–6651.

145.   Arnott A, Mueller I, Ramsland PA, Siba PM, Reeder JC, Barry AE. Global Population Structure of the Genes Encoding the Malaria Vaccine Candidate, Plasmodium vivax Apical Membrane Antigen 1 (PvAMA1). PLoS Negl Trop Dis. 2013;7: e2506.

146.   Barry AE, Lee S, Buckee CO, Reeder JC. Contrasting Population Structures of the Genes Encoding Ten Leading Vaccine-Candidate Antigens of the Human Malaria Parasite, Plasmodium falciparum. PLoS One. 2009;4: e8497.

147.   Boulanger N, Charoenvit Y, Krettli A, Betschart B. Developmental changes in the circumsporozoite proteins of Plasmodium berghei and P. gallinaceum in their mosquito vectors. Parasitol Res. 1995;81: 58–65.

148.   Posthuma G, Meis JF, Verhave JP, Hollingdale MR, Ponnudurai T, Meuwissen JH, et al. Immunogold localization of circumsporozoite protein of the malaria parasite Plasmodium falciparum during sporogony in Anopheles stephensi midguts. Eur J Cell Biol. 1988;46: 18–24.

149.   Golenda CF, Starkweather WH, Wirtz RA. The distribution of circumsporozoite protein (CS) in Anopheles stephensi mosquitoes infected with Plasmodium falciparum malaria. J Histochem Cytochem. 1990;38: 475–481.

150.   World Health Organization. World Malaria Report 2011. 2011.

151.   Foley DH, Klein TA, Lee I-Y, Kim M-S, Wilkerson RC, Genelle H, et al. Mosquito Species Composition and Plasmodium vivax Infection Rates on Baengnyeong-do (Island), Republic of Korea. Korean J Parasitol. 2011;49: 313.

152.   Yoo D-H, Shin E-H, Park M-Y, Kim H-C, Lee D-K, Lee H-H, et al. Mosquito Species Composition and Plasmodium vivax infection Rates for Korean Army Bases near the Demilitarized Zone in the Republic of Korea, 2011. Am J Trop Med Hyg. 2012;88: 24–28.

153.   Adak T, Singh OP, Das MK, Wattal S, Nanda N. Comparative susceptibility of three important malaria vectors Anopheles stephensi, Anopheles fluviatilis, and Anopheles sundaicus to Plasmodium vivax. J Parasitol. 2005;91: 79–82.

154.   Marrelli MT, Honório NA, Flores-Mendoza C, Lourenço-de-Oliveira R, Marinotti O, Kloetzel JK. Comparative susceptibility of two members of the Anopheles oswaldoi complex, An. oswaldoi and An. konderi, to infection by Plasmodium vivax. Trans R Soc Trop Med Hyg. 1999;93: 381–384.

155.   Joshi D, Choochote W, Park M-H, Kim J-Y, Kim T-S, Suwonkerd W, et al. The susceptibility of Anopheles lesteri to infection with Korean strain of Plasmodium vivax. Malar J. 2009;8: 42.

156.   Silva ANM da, da Silva ANM, Santos CCB, Lacerda RN, Machado RLD, Póvoa MM. Susceptibility of Anopheles aquasalis and An. darlingi to Plasmodium vivax VK210 and VK247. Mem Inst Oswaldo Cruz. 2006;101: 547–550.

157.   Takala SL, Coulibaly D, Thera MA, Batchelor AH, Cummings MP, Escalante AA, et al. Extreme polymorphism in a vaccine antigen and risk of clinical malaria: implications for vaccine development. Sci Transl Med. 2009;1: 2ra5.

158.   Ashley EA, Dhorda M, Fairhurst RM, Amaratunga C, Lim P, Suon S, et al. Spread of artemisinin resistance in Plasmodium falciparum malaria. N Engl J Med. 2014;371: 411–423.

159.   Shanks GD, Edstein MD, Jacobus D. Evolution from double to triple-antimalarial drug combinations. Trans R Soc Trop Med Hyg. 2015;109: 182–188.

160.   Fairhurst RM. Understanding artemisinin-resistant malaria: what a difference a year makes. Curr Opin Infect Dis. 2015;28: 417–425.

161.   Taylor SM, Juliano JJ. Artemisinin combination therapies and malaria parasite drug resistance: the game is afoot. J Infect Dis. 2014;210: 335–337.

162.   Chaorattanakawee S, Saunders DL, Sea D, Chanarat N, Yingyuen K, Sundrakes S, et al. Ex Vivo Drug Susceptibility Testing and Molecular Profiling of Clinical Plasmodium falciparum Isolates from Cambodia from 2008 to 2013 Suggest Emerging Piperaquine Resistance. Antimicrob Agents Chemother. 2015;59: 4631–4643.

163.   Leang R, Taylor WRJ, Bouth DM, Song L, Tarning J, Char MC, et al. Evidence of Plasmodium falciparum Malaria Multidrug Resistance to Artemisinin and Piperaquine in Western Cambodia: Dihydroartemisinin-Piperaquine Open-Label Multicenter Clinical Assessment. Antimicrob Agents Chemother. 2015;59: 4719–4726.

164.   Lim P, Dek D, Try V, Sreng S, Suon S, Fairhurst RM. Decreasing pfmdr1 copy number suggests that Plasmodium falciparum in Western Cambodia is regaining in vitro susceptibility to mefloquine. Antimicrob Agents Chemother. 2015;59: 2934–2937.

165.   Spring MD, Lin JT, Manning JE, Vanachayangkul P, Somethy S, Bun R, et al. Dihydroartemisinin-piperaquine failure associated with a triple mutant including kelch13 C580Y in Cambodia: an observational cohort study. Lancet Infect Dis. 2015;15: 683–691.

166.   Ariey F, Witkowski B, Amaratunga C, Beghain J, Langlois A-C, Khim N, et al. A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. Nature. 2014;505: 50–55.

167.   Miotto O, Amato R, Ashley EA, MacInnis B, Almagro-Garcia J, Amaratunga C, et al. Genetic architecture of artemisinin-resistant Plasmodium falciparum. Nat Genet. 2015;47: 226–234.

168.   Takala-Harrison S, Clark TG, Jacob CG, Cummings MP, Miotto O, Dondorp AM, et al. Genetic loci associated with delayed clearance of Plasmodium falciparum following artemisinin treatment in

Southeast Asia. Proceedings of the National Academy of Sciences. 2012;110: 240–245.

169.   Spring MD, Lin JT, Manning JE, Vanachayangkul P, Somethy S, Bun R, et al. Dihydroartemisinin-piperaquine failure associated with a triple mutant including kelch13 C580Y in Cambodia: an observational cohort study. Lancet Infect Dis. 2015;15: 683–691.

170.   Chaorattanakawee S, Saunders DL, Sea D, Chanarat N, Yingyuen K, Sundrakes S, et al. Ex Vivo Drug Susceptibility Testing and Molecular Profiling of Clinical Plasmodium falciparum Isolates from Cambodia from 2008 to 2013 Suggest Emerging Piperaquine Resistance. Antimicrob Agents Chemother. 2015;59: 4631–4643.

171.   Chaorattanakawee S, Saunders DL, Sea D, Chanarat N, Yingyuen K, Sundrakes S, et al. Ex Vivo Drug Susceptibility Testing and Molecular Profiling of Clinical Plasmodium falciparum Isolates from Cambodia from 2008 to 2013 Suggest Emerging Piperaquine Resistance. Antimicrob Agents Chemother. 2015;59: 4631–4643.

172.   Rutvisuttinunt W, Chaorattanakawee S, Tyner SD, Teja-Isavadharm P, Se Y, Yingyuen K, et al. Optimizing the HRP-2 in vitro malaria drug susceptibility assay using a reference clone to improve comparisons of Plasmodium falciparum field isolates. Malar J. 2012;11: 325.

173.   Saunders DL, Suwanna C, Panita G, Charlotte L, Sok S, Worachet K, et al. Atovaquone-Proguanil Remains a Potential Stopgap Therapy for Multidrug-Resistant Plasmodium falciparum in Areas along the Thai-Cambodian Border. Antimicrob Agents Chemother. 2015;60: 1896–1898.

174.   Beshir KB, Hallett RL, Eziefula AC, Bailey R, Watson J, Wright SG, et al. Measuring the efficacy of anti-malarial drugs in vivo: quantitative PCR measurement of parasite clearance. Malar J. 2010;9: 312.

175.   Bahl A, Brunk B, Crabtree J, Fraunholz MJ, Gajria B, Grant GR, et al. PlasmoDB: the Plasmodium genome resource. A database integrating experimental and computational data. Nucleic Acids Res. 2003;31: 212–215.

176.   Conrad MD, LeClair N, Arinaitwe E, Wanzira H, Kakuru A, Bigira V, et al. Comparative impacts over 5 years of artemisinin-based combination therapies on Plasmodium falciparum polymorphisms that modulate drug sensitivity in Ugandan children. J Infect Dis. 2014;210: 344–353.

177.   Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013;30: 2725–2729.

178.   Veiga MI, Osorio NS, Ferreira PE, Franzen O, Dahlstrom S, Lum JK, et al. Complex Polymorphisms in the Plasmodium falciparum Multidrug Resistance Protein 2 Gene and Its Contribution to Antimalarial Response. Antimicrob Agents Chemother. 2014;58: 7390–7397.

179.   Lim P, Dek D, Try V, Sreng S, Suon S, Fairhurst RM. Decreasing pfmdr1 copy number suggests that Plasmodium falciparum in Western Cambodia is regaining in vitro susceptibility to mefloquine. Antimicrob Agents Chemother. 2015;59: 2934–2937.

180.   Eastman RT, Dharia NV, Winzeler EA, Fidock DA. Piperaquine resistance is associated with a copy number variation on chromosome 5 in drug-pressured Plasmodium falciparum parasites. Antimicrob Agents Chemother. 2011;55: 3908–3916.

181.   Miotto O, Amato R, Ashley EA, MacInnis B, Almagro-Garcia J, Amaratunga C, et al. Genetic

architecture of artemisinin-resistant Plasmodium falciparum. Nat Genet. 2015;47: 226–234.

182.   Okombo J, Abdi AI, Kiara SM, Mwai L, Pole L, Sutherland CJ, et al. Repeat polymorphisms in the low-complexity regions of Plasmodium falciparum ABC transporters and associations with in vitro antimalarial responses. Antimicrob Agents Chemother. 2013;57: 6196–6204.

183.   Pelleau S, Moss EL, Dhingra SK, Volney B, Casteras J, Gabryszewski SJ, et al. Adaptive evolution of malaria parasites in French Guiana: Reversal of chloroquine resistance by acquisition of a mutation in pfcrt. Proc Natl Acad Sci U S A. 2015;112: 11672–11677.

184.   Jombart T, Ahmed I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics. 2011;27: 3070–3071.

185.   Hudson RR. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics. 2002;18: 337–338.

186.   Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 2009;5: e1000695.

187.   Chaorattanakawee S, Saunders DL, Sea D, Chanarat N, Yingyuen K, Sundrakes S, et al. Ex Vivo Drug Susceptibility Testing and Molecular Profiling of Clinical Plasmodium falciparum Isolates from Cambodia from 2008 to 2013 Suggest Emerging Piperaquine Resistance. Antimicrob Agents Chemother. 2015;59: 4631–4643.

188.   Spring MD, Lin JT, Manning JE, Vanachayangkul P, Somethy S, Bun R, et al. Dihydroartemisinin-piperaquine failure associated with a triple mutant including kelch13 C580Y in Cambodia: an observational cohort study. Lancet Infect Dis. 2015;15: 683–691.

189.   Miotto O, Amato R, Ashley EA, MacInnis B, Almagro-Garcia J, Amaratunga C, et al. Genetic architecture of artemisinin-resistant Plasmodium falciparum. Nat Genet. 2015;47: 226–234.

190.   Maxmen A. Back on TRAC: New trial launched in bid to outpace multidrug-resistant malaria. Nat Med. 2016;22: 220–221.

191.   Veiga MI, Osorio NS, Ferreira PE, Franzen O, Dahlstrom S, Lum JK, et al. Complex Polymorphisms in the Plasmodium falciparum Multidrug Resistance Protein 2 Gene and Its Contribution to Antimalarial Response. Antimicrob Agents Chemother. 2014;58: 7390–7397.

192.   Barnes DA, Foote SJ, Galatis D, Kemp DJ, Cowman AF. Selection for high-level chloroquine resistance results in deamplification of the pfmdr1 gene and increased sensitivity to mefloquine in Plasmodium falciparum. EMBO J. 1992;11: 3067–3075.

193.   van Es HH, Karcz S, Chu F, Cowman AF, Vidal S, Gros P, et al. Expression of the plasmodial pfmdr1 gene in mammalian cells is associated with increased susceptibility to chloroquine. Mol Cell Biol. 1994;14: 2419–2428.

194.   Venkatesan M, Gadalla NB, Stepniewska K, Dahal P, Nsanzabana C, Moriera C, et al. Polymorphisms in Plasmodium falciparum chloroquine resistance transporter and multidrug resistance 1 genes: parasite risk factors that affect treatment outcomes for P. falciparum malaria after artemether-lumefantrine and artesunate-amodiaquine. Am J Trop Med Hyg. 2014;91: 833–843.

195.   Conrad MD, LeClair N, Arinaitwe E, Wanzira H, Kakuru A, Bigira V, et al. Comparative impacts

over 5 years of artemisinin-based combination therapies on Plasmodium falciparum polymorphisms that modulate drug sensitivity in Ugandan children. J Infect Dis. 2014;210: 344–353.

196.   Chaorattanakawee S, Saunders DL, Sea D, Chanarat N, Yingyuen K, Sundrakes S, et al. Ex Vivo Drug Susceptibility Testing and Molecular Profiling of Clinical Plasmodium falciparum Isolates from Cambodia from 2008 to 2013 Suggest Emerging Piperaquine Resistance. Antimicrob Agents Chemother. 2015;59: 4631–4643.

197.   Lim P, Dek D, Try V, Sreng S, Suon S, Fairhurst RM. Decreasing pfmdr1 copy number suggests that Plasmodium falciparum in Western Cambodia is regaining in vitro susceptibility to mefloquine. Antimicrob Agents Chemother. 2015;59: 2934–2937.

198.   Amaratunga C, Lim P, Suon S, Sreng S, Mao S, Sopha C, et al. Dihydroartemisinin-piperaquine resistance in Plasmodium falciparum malaria in Cambodia: a multisite prospective cohort study. Lancet Infect Dis. 2016;16: 357–365.

199.   Spring MD, Lin JT, Manning JE, Vanachayangkul P, Somethy S, Bun R, et al. Dihydroartemisinin-piperaquine failure associated with a triple mutant including kelch13 C580Y in Cambodia: an observational cohort study. Lancet Infect Dis. 2015;15: 683–691.

200.   Leang R, Taylor WRJ, Bouth DM, Song L, Tarning J, Char MC, et al. Evidence of Plasmodium falciparum Malaria Multidrug Resistance to Artemisinin and Piperaquine in Western Cambodia: Dihydroartemisinin-Piperaquine Open-Label Multicenter Clinical Assessment. Antimicrob Agents Chemother. 2015;59: 4719–4726.

201.   Miotto O, Amato R, Ashley EA, MacInnis B, Almagro-Garcia J, Amaratunga C, et al. Genetic architecture of artemisinin-resistant Plasmodium falciparum. Nat Genet. 2015;47: 226–234.

202.   Miotto O, Almagro-Garcia J, Manske M, Macinnis B, Campino S, Rockett KA, et al. Multiple populations of artemisinin-resistant Plasmodium falciparum in Cambodia. Nat Genet. 2013;45: 648–655.

203.   Witkowski B, Menard D, Amaratunga C, Fairhurst RM. Ring - stage Survival Assays (RSA) to evaluate the in - vitro and ex - vivo susceptibility of Plasmodium falciparum to artemisinins. 2013.

204.   Fairhurst RM. Understanding artemisinin-resistant malaria: what a difference a year makes. Curr Opin Infect Dis. 2015;28: 417–425.

205.   Shanks GD, Edstein MD, Jacobus D. Evolution from double to triple-antimalarial drug combinations. Trans R Soc Trop Med Hyg. 2015;109: 182–188.

206.   Spring MD, Lin JT, Manning JE, Vanachayangkul P, Somethy S, Bun R, et al. Dihydroartemisinin-piperaquine failure associated with a triple mutant including kelch13 C580Y in Cambodia: an observational cohort study. Lancet Infect Dis. 2015;15: 683–691.

207.   Dondorp AM, Nosten F, Yi P, Das D, Phyo AP, Tarning J, et al. Artemisinin resistance in Plasmodium falciparum malaria. N Engl J Med. 2009;361: 455–467.

208.   World Health Organization. Global Plan for Artemisinin Resistance Containment. World Health Organization, editor. World Health Organization; 2011.

209.   Durnez L, Mao S, Denis L, Roelants P, Sochantha T, Coosemans M. Outdoor malaria transmission in forested villages of Cambodia. Malar J. 2013;12: 329.

210. Maude RJ, Nguon C, Ly P, Bunkea T, Ngor P, Canavati de la Torre SE, et al. Spatial and temporal epidemiology of clinical malaria in Cambodia 2004-2013. Malar J. 2014;13: 385.

211. Zhou G, Sirichaisinthop J, Sattabongkot J, Jones J, Bjørnstad ON, Yan G, et al. Spatio-temporal distribution of Plasmodium falciparum and p. Vivax malaria in Thailand. Am J Trop Med Hyg. 2005;72: 256–262.

212. Cui L, Yan G, Sattabongkot J, Cao Y, Chen B, Chen X, et al. Malaria in the Greater Mekong Subregion: heterogeneity and complexity. Acta Trop. 2012;121: 227–239.

213. Wangroongsarb P, Sudathip P, Satimai W. Characteristics and malaria prevalence of migrant populations in malaria-endemic areas along the Thai-Cambodian border. Southeast Asian Journal of Tropical Medicine and Public Health. 2012; Available: http://www.ncbi.nlm.nih.gov/pubmed?term=characteristics%20and%20malaria%20prevalence%20o f%20migrant%20populations%20in%20malaria%20endemic%20areas%20along%20the%20thai-cambodian%20border.&cmd=correctspelling

214. Gray K-A, Dowd S, Bain L, Bobogare A, Wini L, Shanks GD, et al. Population genetics of Plasmodium falciparum and Plasmodium vivax and asymptomatic malaria in Temotu Province, Solomon Islands. Malar J. 2013;12: 429.

215. Jennison C, Arnott A, Tessier N, Tavul L, Koepfli C, Felger I, et al. Plasmodium vivax populations are more genetically diverse and less structured than sympatric Plasmodium falciparum populations. PLoS Negl Trop Dis. 2015;9: e0003634.

216. Noviyanti R, Coutrier F, Utami RAS, Trimarsanto H, Tirta YK, Trianty L, et al. Contrasting Transmission Dynamics of Co-endemic Plasmodium vivax and P. falciparum: Implications for Malaria Control and Elimination. PLoS Negl Trop Dis. 2015;9: e0003739.

217. Orjuela-Sánchez P, Sá JM, Brandi MCC, Rodrigues PT, Bastos MS, Amaratunga C, et al. Higher microsatellite diversity in Plasmodium vivax than in sympatric Plasmodium falciparum populations in Pursat, Western Cambodia. Exp Parasitol. 2013;134: 318–326.

218. Arnott A, Wapling J, Mueller I, Ramsland PA, Siba PM, Reeder JC, et al. Distinct patterns of diversity, population structure and evolution in the AMA1 genes of sympatric Plasmodium falciparum and Plasmodium vivax populations of Papua New Guinea from an area of similarly high transmission. Malar J. 2014;13: 233.

219. Ord RL, Tami A, Sutherland CJ. ama1 genes of sympatric Plasmodium vivax and P. falciparum from Venezuela differ significantly in genetic diversity and recombination frequency. PLoS One. 2008;3: e3366.

220. Miotto O, Almagro-Garcia J, Manske M, Macinnis B, Campino S, Rockett KA, et al. Multiple populations of artemisinin-resistant Plasmodium falciparum in Cambodia. Nat Genet. 2013;45: 648–655.

221. Takala-Harrison S, Jacob CG, Arze C, Cummings MP, Silva JC, Dondorp AM, et al. Independent emergence of artemisinin resistance mutations among Plasmodium falciparum in Southeast Asia. J Infect Dis. 2015;211: 670–679.

222. Miotto O, Amato R, Ashley EA, MacInnis B, Almagro-Garcia J, Amaratunga C, et al. Genetic architecture of artemisinin-resistant Plasmodium falciparum. Nat Genet. 2015;47: 226–234.

223. World Health Organization. Global plan for artemisinin resistance containment (GPARC). World Health Organization; 2011.

224. Beshir KB, Hallett RL, Eziefula AC, Bailey R, Watson J, Wright SG, et al. Measuring the efficacy of anti-malarial drugs in vivo: quantitative PCR measurement of parasite clearance. Malar J. 2010;9: 312.

225. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM [Internet]. arXiv.org. 2013. Available: http://arxiv.org/abs/1303.3997

226. http://broadinstitute.github.io/picard/ [Internet]. Available: http://broadinstitute.github.io/picard/

227. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43: 491–498.

228. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20: 1297–1303.

229. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27: 573–580.

230. Neafsey DE, Kevin G, Jiang RHY, Lauren Y, Sykes SM, Sakina S, et al. The malaria parasite Plasmodium vivax exhibits greater genetic diversity than Plasmodium falciparum. Nat Genet. 2012;44: 1046–1050.

231. Auburn S, Campino S, Miotto O, Djimde AA, Zongo I, Manske M, et al. Characterization of within-host Plasmodium falciparum diversity using next-generation sequence data. PLoS One. 2012;7: e32891.

232. Manske M, Miotto O, Campino S, Auburn S, Almagro-Garcia J, Maslen G, et al. Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. Nature. 2012;487: 375–379.

233. Parobek CM, Bailey JA, Hathaway NJ, Socheat D, Rogers WO, Juliano JJ. Differing patterns of selection and geospatial genetic diversity within two leading Plasmodium vivax candidate vaccine antigens. PLoS Negl Trop Dis. 2014;8: e2796.

234. Lin JT, Hathaway NJ, Saunders DL, Lon C, Balasubramanian S, Kharabora O, et al. Using Amplicon Deep Sequencing to Detect Genetic Signatures of Plasmodium vivax Relapse. J Infect Dis. 2015;212: 999–1008.

235. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 2009;5: e1000695.

236. Staab PR, Dirk M. Coala: an R framework for coalescent simulation. Bioinformatics. 2016; btw098.

237. Hudson RR. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics. 2002;18: 337–338.

238.   Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. Genome Biol. 2014;15: R84.

239.   Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. On detecting incomplete soft or hard selective sweeps using haplotype structure. Mol Biol Evol. 2014;31: 1275–1291.

240.   Schlamp F, van der Made J, Stambler R, Chesebrough L, Boyko AR, Messer PW. Evaluating the performance of selection scans to detect selective sweeps in domestic dogs. Mol Ecol. 2016;25: 342–356.

241.   McVean G, Awadalla P, Fearnhead P. A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics. 2002;160: 1231–1241.

242.   Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. PLoS Biol. 2006;4: e72.

243.   Szpiech ZA, Hernandez RD. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. Mol Biol Evol. 2014;31: 2824–2827.

244.   Parobek CM, Bailey JA, Hathaway NJ, Socheat D, Rogers WO, Juliano JJ. Differing patterns of selection and geospatial genetic diversity within two leading Plasmodium vivax candidate vaccine antigens. PLoS Negl Trop Dis. 2014;8: e2796.

245.   Lin JT, Hathaway NJ, Saunders DL, Lon C, Balasubramanian S, Kharabora O, et al. Using Amplicon Deep Sequencing to Detect Genetic Signatures of Plasmodium vivax Relapse. J Infect Dis. 2015;212: 999–1008.

246.   Miotto O, Almagro-Garcia J, Manske M, Macinnis B, Campino S, Rockett KA, et al. Multiple populations of artemisinin-resistant Plasmodium falciparum in Cambodia. Nat Genet. 2013;45: 648–655.

247.   Miotto O, Amato R, Ashley EA, MacInnis B, Almagro-Garcia J, Amaratunga C, et al. Genetic architecture of artemisinin-resistant Plasmodium falciparum. Nat Genet. 2015;47: 226–234.

248.   Dharia NV, Bright AT, Westenberger SJ, Barnes SW, Batalov S, Kuhen K, et al. 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. 2010;107: 20045–20050.

249.   Lin JT, Patel JC, Kharabora O, Sattabongkot J, Muth S, Ubalee R, et al. 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. 2013;88: 1116–1123.

250.   Amambua-Ngwa A, Park DJ, Volkman SK, Barnes KG, Bei AK, Lukens AK, et al. SNP Genotyping Identifies New Signatures of Selection in a Deep Sample of West African Plasmodium falciparum Malaria Parasites. Mol Biol Evol. 2012;29: 3249–3253.

251.   Mu J, Myers RA, Jiang H, Liu S, Ricklefs S, Waisberg M, et al. Plasmodium falciparum genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. Nat Genet. 2010;42: 268–271.

252.   Cheeseman IH, Miller B, Tan JC, Tan A, Nair S, Nkhoma SC, et al. Population Structure Shapes Copy Number Variation in Malaria Parasites. Mol Biol Evol. 2016;33: 603–620.

253.   Menard D, Chan ER, Benedet C, Ratsimbasoa A, Kim S, Chim P, et al. Whole genome sequencing of field isolates reveals a common duplication of the Duffy binding protein gene in Malagasy Plasmodium vivax strains. PLoS Negl Trop Dis. 2013;7: e2489.

254.   Trenholme KR, Gardiner DL, Holt DC, Thomas EA, Cowman AF, Kemp DJ. clag9: A cytoadherence gene in Plasmodium falciparum essential for binding of parasitized erythrocytes to CD36. Proc Natl Acad Sci U S A. 2000;97: 4029–4033.

255.   Chen F, Mackey AJ, Stoeckert CJ Jr, Roos DS. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res. 2006;34: D363–8.

256.   Noviyanti R, Coutrier F, Utami RAS, Trimarsanto H, Tirta YK, Trianty L, et al. Contrasting Transmission Dynamics of Co-endemic Plasmodium vivax and P. falciparum: Implications for Malaria Control and Elimination. PLoS Negl Trop Dis. 2015;9: e0003739.

257.   Jennison C, Arnott A, Tessier N, Tavul L, Koepfli C, Felger I, et al. Plasmodium vivax populations are more genetically diverse and less structured than sympatric Plasmodium falciparum populations. PLoS Negl Trop Dis. 2015;9: e0003634.

258.   Ord RL, Tami A, Sutherland CJ. ama1 genes of sympatric Plasmodium vivax and P. falciparum from Venezuela differ significantly in genetic diversity and recombination frequency. PLoS One. 2008;3: e3366.

259.   Gray K-A, Dowd S, Bain L, Bobogare A, Wini L, Shanks GD, et al. Population genetics of Plasmodium falciparum and Plasmodium vivax and asymptomatic malaria in Temotu Province, Solomon Islands. Malar J. 2013;12: 429.

260.   Arnott A, Wapling J, Mueller I, Ramsland PA, Siba PM, Reeder JC, et al. Distinct patterns of diversity, population structure and evolution in the AMA1 genes of sympatric Plasmodium falciparum and Plasmodium vivax populations of Papua New Guinea from an area of similarly high transmission. Malar J. 2014;13: 233.

261.   Orjuela-Sánchez P, Sá JM, Brandi MCC, Rodrigues PT, Bastos MS, Amaratunga C, et al. Higher microsatellite diversity in Plasmodium vivax than in sympatric Plasmodium falciparum populations in Pursat, Western Cambodia. Exp Parasitol. 2013;134: 318–326.

262.   Robinson JD, Coffman AJ, Hickerson MJ, Gutenkunst RN. Sampling strategies for frequency spectrum-based population genomic inference. BMC Evol Biol. 2014;14: 254.

263.   Boyd MF, Kitchen SF. Simultaneous inoculation with Plasmodium vivax and Plasmodium falciparum. Am J Trop Med Hyg. 1937;17: 855–861.

264.   Phimpraphi W, Waraphon P, Paul RE, Surapon Y, Supalarp P-A, Nipon T, et al. Longitudinal study of Plasmodium falciparum and Plasmodium vivax in a Karen population in Thailand. Malar J. 2008;7: 99.

265.   Amambua-Ngwa A, Tetteh KKA, Manske M, Gomez-Escobar N, Stewart LB, Deerhake ME, et al. Population genomic scan for candidate signatures of balancing selection to guide antigen characterization in malaria parasites. PLoS Genet. 2012;8: e1002992.

266.   Duffy CW, Assefa SA, Abugri J, Amoako N, Owusu-Agyei S, Anyorigiya T, et al. Comparison of genomic signatures of selection on Plasmodium falciparum between different regions of a country

with high malaria endemicity. BMC Genomics. 2015;16: 527.

267. Mobegi VA, Duffy CW, Amambua-Ngwa A, Loua KM, Laman E, Nwakanma DC, et al. Genome-wide analysis of selection on the malaria parasite Plasmodium falciparum in West African populations of differing infection endemicity. Mol Biol Evol. 2014;31: 1490–1499.

268. Chang H-H, Park DJ, Galinsky KJ, Schaffner SF, Ndiaye D, Ndir O, et al. Genomic sequencing of Plasmodium falciparum malaria parasites from Senegal reveals the demographic history of the population. Mol Biol Evol. 2012;29: 3427–3439.

269. Ocholla H, Preston MD, Mipando M, Jensen ATR, Campino S, MacInnis B, et al. Whole-Genome Scans Provide Evidence of Adaptive Evolution in Malawian Plasmodium falciparum Isolates. J Infect Dis. 2014;210: 1991–2000.

270. Park DJ, Lukens AK, Neafsey DE, Schaffner SF, Chang H-H, Valim C, et al. Sequence-based association and selection scans identify drug resistance loci in the Plasmodium falciparum malaria parasite. Proc Natl Acad Sci U S A. 2012;109: 13052–13057.

271. Nwakanma DC, Duffy CW, Amambua-Ngwa A, Oriero EC, Bojang KA, Pinder M, et al. Changes in Malaria Parasite Drug Resistance in an Endemic Population Over a 25-Year Period With Resulting Genomic Evidence of Selection. J Infect Dis. 2013;209: 1126–1135.

272. Hester J, Chan ER, Menard D, Mercereau-Puijalon O, Barnwell J, Zimmerman PA, et al. De novo assembly of a field isolate genome reveals novel Plasmodium vivax erythrocyte invasion genes. PLoS Negl Trop Dis. 2013;7: e2569.

273. Smith RC, Joel V-R, Marcelo J-L. The Plasmodium bottleneck: malaria parasite losses in the mosquito vector. Mem Inst Oswaldo Cruz. 2014;109: 644–661.

274. Chang H-H, Hartl DL. Recurrent bottlenecks in the malaria life cycle obscure signals of positive selection. Parasitology. 2015;142 Suppl 1: S98–S107.

275. Nkhoma SC, Shalini N, Salma A-S, Elizabeth A, Rose M, Phyo AP, et al. Population genetic correlates of declining transmission in a human pathogen. Mol Ecol. 2012;22: 273–285.

276. Zeeman A-M, der Wel AV, Kocken CHM. Ex vivo culture of Plasmodium vivax and Plasmodium cynomolgi and in vitro culture of Plasmodium knowlesi blood stages. Methods Mol Biol. 2013;923: 35–49.

277. Vaughan AM, Kappe SHI, Ploss A, Mikolajczak SA. Development of humanized mouse models to study human malaria parasite infection. Future Microbiol. NIH Public Access; 2012;7. doi:10.2217/fmb.12.27

278. Mikolajczak SA, Vaughan AM, Kangwanrangsan N, Roobsoong W, Fishbaugher M, Yimamnuaychok N, et al. Plasmodium vivax liver stage development and hypnozoite persistence in human liver-chimeric mice. - PubMed - NCBI [Internet]. [cited 29 Mar 2016]. Available: http://www.ncbi.nlm.nih.gov/pubmed/25800544

279. Joyner C, Barnwell JW, Galinski MR. No more monkeying around: primate malaria model systems are key to understanding Plasmodium vivax liver-stage biology, hypnozoites, and relapses. Front Microbiol. 2015;6: 145.

280. Parobek CM, Bailey JA, Hathaway NJ, Socheat D, Rogers WO, Juliano JJ. Differing patterns of

selection and geospatial genetic diversity within two leading Plasmodium vivax candidate vaccine antigens. PLoS Negl Trop Dis. 2014;8: e2796.

281.   Maxmen A. Back on TRAC: New trial launched in bid to outpace multidrug-resistant malaria. Nat Med. 2016;22: 220–221.

282.   Orjuela-Sánchez P, Sá JM, Brandi MCC, Rodrigues PT, Bastos MS, Amaratunga C, et al. Higher microsatellite diversity in Plasmodium vivax than in sympatric Plasmodium falciparum populations in Pursat, Western Cambodia. Exp Parasitol. 2013;134: 318–326.

283.   Noviyanti R, Coutrier F, Utami RAS, Trimarsanto H, Tirta YK, Trianty L, et al. Contrasting Transmission Dynamics of Co-endemic Plasmodium vivax and P. falciparum: Implications for Malaria Control and Elimination. PLoS Negl Trop Dis. 2015;9: e0003739.

284.   Jennison C, Arnott A, Tessier N, Tavul L, Koepfli C, Felger I, et al. Plasmodium vivax populations are more genetically diverse and less structured than sympatric Plasmodium falciparum populations. PLoS Negl Trop Dis. 2015;9: e0003634.

285.   Gray K-A, Dowd S, Bain L, Bobogare A, Wini L, Shanks GD, et al. Population genetics of Plasmodium falciparum and Plasmodium vivax and asymptomatic malaria in Temotu Province, Solomon Islands. Malar J. 2013;12: 429.

286.   Ord RL, Tami A, Sutherland CJ. ama1 genes of sympatric Plasmodium vivax and P. falciparum from Venezuela differ significantly in genetic diversity and recombination frequency. PLoS One. 2008;3: e3366.

287.   Arnott A, Wapling J, Mueller I, Ramsland PA, Siba PM, Reeder JC, et al. Distinct patterns of diversity, population structure and evolution in the AMA1 genes of sympatric Plasmodium falciparum and Plasmodium vivax populations of Papua New Guinea from an area of similarly high transmission. Malar J. 2014;13: 233.

288.   Chang H-H, Moss EL, Park DJ, Ndiaye D, Mboup S, Volkman SK, et al. Malaria life cycle intensifies both natural selection and random genetic drift. Proc Natl Acad Sci U S A. 2013;110: 20129–20134.

289.   Chang H-H, Hartl DL. Recurrent bottlenecks in the malaria life cycle obscure signals of positive selection. Parasitology. 2015;142 Suppl 1: S98–S107.

290.   Davis TME, Hamzah J, Ilett KF, Karunajeewa HA, Reeder JC, Batty KT, et al. In vitro interactions between piperaquine, dihydroartemisinin, and other conventional and novel antimalarial drugs. Antimicrob Agents Chemother. 2006;50: 2883–2885.

291.   Duffy CW, Assefa SA, Abugri J, Amoako N, Owusu-Agyei S, Anyorigiya T, et al. Comparison of genomic signatures of selection on Plasmodium falciparum between different regions of a country with high malaria endemicity. BMC Genomics. 2015;16: 527.

292.   Amambua-Ngwa A, Park DJ, Volkman SK, Barnes KG, Bei AK, Lukens AK, et al. SNP Genotyping Identifies New Signatures of Selection in a Deep Sample of West African Plasmodium falciparum Malaria Parasites. Mol Biol Evol. 2012;29: 3249–3253.

293.   Mobegi VA, Duffy CW, Amambua-Ngwa A, Loua KM, Laman E, Nwakanma DC, et al. Genome-wide analysis of selection on the malaria parasite Plasmodium falciparum in West African

populations of differing infection endemicity. Mol Biol Evol. 2014;31: 1490–1499.

294.   Mu J, Myers RA, Jiang H, Liu S, Ricklefs S, Waisberg M, et al. Plasmodium falciparum genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. Nat Genet. 2010;42: 268–271.

295.   Chang H-H, Park DJ, Galinsky KJ, Schaffner SF, Ndiaye D, Ndir O, et al. Genomic sequencing of Plasmodium falciparum malaria parasites from Senegal reveals the demographic history of the population. Mol Biol Evol. 2012;29: 3427–3439.

296.   Sanglard D, Coste A, Ferrari S. Antifungal drug resistance mechanisms in fungal pathogens from the perspective of transcriptional gene regulation. FEMS Yeast Res. 2009;9: 1029–1050.

297.   Kocken CHM, Ozwara H, van der Wel A, Beetsma AL, Mwenda JM, Thomas AW. Plasmodium knowlesi provides a rapid in vitro and in vivo transfection system that enables double-crossover gene knockout studies. Infect Immun. 2002;70: 655–660.