
DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU


HARVARD LIBRARY
Office for Scholarly Communication

Harnessing Advances in Genomics and Molecular Genetics to Inform Understanding of P. vivax Epidemiology, Evolution, and Drug Resistance

Citation

Buyon, Lucas Evan. 2022. Harnessing Advances in Genomics and Molecular Genetics to Inform Understanding of P. vivax Epidemiology, Evolution, and Drug Resistance. Doctoral dissertation, Harvard University Graduate School of Arts and Sciences.

Permanent link

https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37371925

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Committee on Higher Degrees in Biological Sciences in Public Health

have examined a dissertation entitled

Harnessing Advances in Genomics and Molecular Genetics to Inform Understanding of *P. vivax* Epidemiology, Evolution, and Drug Resistance

presented by Lucas Evan Buyon

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature _____

Typed name: Prof. Dyann Wirth

Signature _____

Typed name: Prof. Jeffrey Dvorin

Signature _____

Typed name: Prof. Daniel Hartl

Signature David Serre

Typed name: Prof. David Serre

Date: April 1, 2022

Harnessing Advances in Genomics and Molecular Genetics to Inform Understanding of *P. vivax* Epidemiology, Evolution, and Drug Resistance

A dissertation presented by

Lucas Evan Buyon

to

The Committee on Higher Degrees in Biological Sciences in Public Health

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biological Sciences in Public Health

Harvard University

Cambridge, Massachusetts

April 1st, 2022

Copyright

© 2022 –Lucas Evan Buyon
All rights reserved.

Dissertation Advisors: Daniel E. Neafsey and Manoj T. Duraisingh Lucas Evan Buyon

Harnessing Advances in Genomics and Molecular Genetics to Inform Understanding of *P. vivax* Epidemiology, Evolution, and Drug Resistance

Abstract

Plasmodium vivax and Plasmodium falciparum are two parasites that cause most malaria cases worldwide. *P. vivax* is chronically understudied compared to *P. falciparum*, and significant aspects of its biology remain a mystery. This dearth of understanding stems from the lack of an *in vitro* culture system for *P. vivax* and the difficulty in generating high-quality sequencing data from patient samples because of low parasitemia. However, novel techniques, including selective whole genome amplification, which allows the sequencing of *P. vivax* samples directly from the blood, and the development of culture systems of closely related parasites *Plasmodium knowlesi* and *Plasmodium cynomolgi*, have enabled large scale population genomic studies and molecular genetic experiments to study *P. vivax* epidemiology and biology. In this dissertation, I use these new scientific tools to explore *P. vivax* transmission dynamics in and assess risk case importation in Panama, use comparative population genomics to identify candidate *P. vivax* drug resistance loci, and validate a *P. vivax* gene, *pvmdr1*, for a functional role regarding drug resistance.

In **Chapter One**, I summarize advances in *P. vivax* genomics, the state of molecular surveillance for *P. vivax*, and what is known about the molecular basis of *P. vivax* drug resistance. In **Chapter Two,** I showcase work using selective whole genome amplification and sequencing to understand *P. vivax* population structure in a low transmission setting and the risk of *P. vivax* case importation in Panama. I find that there is a single highly related lineage of *P. vivax* parasites in Panama that has persisted for over a decade. I also uncover several likely imported cases and discuss possible future uses of molecular surveillance for identifying case

importation. In **Chapter Three,** I discuss the use of comparative selection scans to identify and prioritize candidate *P. vivax* drug resistance loci for experimental characterization. *P. knowlesi* has only recently been understood as a human infection but is not thought to have human to human transmission. Therefore, *P. knowlesi* has only recently been exposed to anti-malarial drugs, and its genome should not exhibit signs of selection due to pressure from these drugs. I hypothesized that genes with a signal of selection only in *P. vivax* populations (and not in *P. knowlesi* populations) and from regions with co-infection in *P. falciparum* and known drug resistance are likely candidate drug resistance causing loci. I compared signals of evolution in both *P. vivax* and *P. knowlesi* to identify a set of candidate drug resistance loci in *P. vivax* to prioritize for *in vitro* characterization with regards to drug resistance. In **Chapter Four,** I discuss my work interrogating a set of globally representative *pvmdr1* alleles for their role in mediating *P. vivax* drug resistance. I conducted a population genomic analysis of variation in this gene and identified 10 SNPs with minor allele frequencies greater than 5%, which exist as 23 unique haplotypes in natural populations. I took advantage of the recent development of a continuous culture system for the closely related parasite *P. knowlesi* and used CRISPR/Cas9 to construct transgenic *P. knowlesi* lines expressing these 23 *pvmdr1* haplotypes. I also mapped on known *pfmdr1* drug resistance polymorphisms on *pvmdr1* and constructed lines that contained these mutations. Finally, I constructed an overexpression plasmid containing *pvmdr1* to test the effect of copy number variation to mediate drug resistance. I then assayed all transgenic *pvmdr1* lines and the *P. knowlesi* YH1 line, to identify changes in their susceptibility, if any, to an array of antimalarial compounds. I found that *pvmdr1* mutations confer reduced susceptibility to mefloquine, lumefantrine, dihydroartemisinin, and halofantrine. I also discuss implications of these findings for *P. vivax* treatment, control, and molecular surveillance. In **Chapter Five,** I

explore the implications of this thesis on future molecular surveillance of *P. vivax* and on exploring its biology in the future.

Table of Contents

List of Figures and Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Appendix:

Acknowledgements

I want to start by thanking my advisors Dan and Manoj.

Dan- thank you for a taking a chance on me and helping mold me into the scientist I am today. I know that I have not been the easiest first student you could have had but thank you for being patient with me. I have a learned a lot from you; how to slow down and asses, rather than react, how to use the "Strong Inference" framework to tackle problems (thank you for making me read that paper), and how to be a better writer (apologies in advance for this run on sentence). It has been a fun journey growing with you and the lab, back from when it was just 4-5 people at the Broad, to the much larger group with global impact that it is now. I hope your future students give you a much easier time than your first one! Thank you for everything!

Manoj- From the very first time we met during interviews, I knew we would get along I have always been grateful for you sticking up for me, particularly through the trials and tribulations of my G1 year. Thank you for taking someone who had not touched a pipette in 6 years and making me into a confident bench scientist. I sincerely appreciate your mentorship and support during my PhD. I also am thankful for the supportive lab environment you have built- it is an amazing place to learn and grow as a scientist. I look forward to beating you in the lab fantasy football league for years to come. Thank you!

Duraisingh and Neafsey Labs – It is definitely a unique experience being co-mentored and being a part of two different research groups. There are so many people to thank, but I particularly want to thank **Angela Early** for all her patient mentorship of me from my first rotation in the Neafsey group and throughout the PhD. I also want to thank fellow team

"Dan-oj"" member, **Jacob Tennessen,** for his mentorship as well. I also want to thank Annie, Zack, Meg, Thais, Emily, Adi, Kristen, Sheena, and other Neafsey and Duraisingh lab members past and present for their help, support, and friendship. I would also like to thank Selina, Rob, Rebecca, Lola, and Aabha from the Wirth lab for their support and friendship as well.

I want to particularly thank **Brendan Elsworth**. Brendan, thank you for being incredibly patient with me as I learned how to be bench scientist. I could not have completed this thesis without you. I look forward to continuing our scientific chats in DC in the coming years!

I also want to thank my fellow PhD friends: **Kevin, James, Bobby (LRH4), Bolu, Noah, Priya, Meghann, Megan, and Georgia.** You all have been so supportive during these years, and I could not have done it without you. To my BPH friends **Caroline, Sam, Francesca, and Maji,** thank you for all your support and mentorship as well. I also want to thank my friends **Harlan, Tom, Dr. Andrew Navia, Glenn, Hace, Stan, Nico, and Keya** for all their love and support.

I am incredibly lucky to have learned from some of the best mentors in the public health field these past few years. I owe a great deal **Kelly Callahan,** and **Dyann Wirth**. I cannot thank you both enough for always taking the time give me advice and support. I am so lucky to have learned from two global health leaders. I also want to thank my high school global health teacher, **Eileen Dieck** for first igniting my passion for public health. Thank you all for your mentorship!

Thank you to my girlfriend **Lauren Monz**. Thank you for all the love and support the past three years. I am not sure why you stuck with me after starting to date me during my high-stress qualifying exam prep, but at least it was a dry-run for thesis-writing stress. Thank you

for caring for me after I tore my ACL, and for putting up with my neuroses. I love you and owe you a long vacation.

Lastly, I want to thank my family. I have been very fortunate to have been born into a family of scientists and am incredibly grateful for having scientific role models from a young age. To my aunt **Dr. Jill Buyon** and my uncle **Dr. Robert Clancy**, thank you for taking me into your lab at NYU when I was in high school and fostering my love of science. To my uncle, **Dr. Eric Lander**, thank you for showing me the highest peaks of science, and the incredible impact scientists can achieve for the greater good. I am very humbled to follow, in a small way, in your footsteps. I want to thank my aunt **Lori Lander** for basically being my second mother in Boston. I cannot begin to say thank you enough for always being there more through the trials and tribulations of this PhD. Thank you for being my rock. Thank you for my brothers, **Noah,** and **Nate** for listening my kvetching. Finally thank you to my parents **Julie and Ethan Buyon**. I am very lucky to have parents who actively encouraged my love of science and helped keep me on this path whenever I was doubting myself. I have always appreciated your listening to my frustrations and struggles and thank you for making the path as easy as it could be. Thank you for your unconditional love and support, this thesis is for you!

Chapter One: Introduction

1.1 *P. vivax* Epidemiology, Biology and Drug Resistance

Plasmodium vivax, a cause of malaria, is a major global health threat. *P. vivax* and *Plasmodium falciparum* are responsible for most human malaria cases, with *P. vivax* causing most malaria cases outside the African continent[1,2] **(Figure 1.1).** The significant progress towards reducing malaria transmission is threatened by the emergence of drug resistance in *P. falciparum* to all clinically used antimalarials[3]. *P. vivax* has documented *in vitro* and *ex vivo* resistance to antimalarial drugs, including chloroquine (CQ), mefloquine (MQ), sulfadoxine, and pyrimethamine (SP)[1,6–9], Drug-resistant *P. vivax* has been reported in many regions of the world[151,6–9]. High-grade CQ-resistance (CQR) in *P. vivax* has emerged in Indonesia, Malaysia, and Papua New Guinea, and has been associated with CQ treatment failure[10,11]. CQR has spread relatively slowly in *P. vivax* populations, compared to its emergence, and spread in *P. falciparum* populations. It has been hypothesized that there is reduced selection pressure in *P. vivax* compared to *P. falciparum* due to the ability of *P. vivax* to produce gametocytes early in infection, which allows the parasite to transmit before drug treatment[12]. *P. vivax* is also subjected to less selection pressure from drugs because of is its ability to relapse from liver hypnozoites, allowing for transmission after drug concentration has waned[12]. Reduced parasite biomass during *P. vivax* infection and different host cell preferences (restriction to reticulocytes) also possibly affects *P. vivax* exposure to selection pressure from antimalarials[12]. The blood-stage activity of Primaquine(PQ)/Tafenoquine (TQ)could also reduce the transmission of CQR parasites. The extent of *P. vivax* drug resistance and subsequent public health consequences remains poorly defined in most regions of the world[1].

Figure 1.1: Map of Estimated *Plasmodium vivax* case counts in 2020. This map is modified from the Malaria Atlas Project.

Despite more than two decades of research into drug-resistant *P. vivax*, molecular markers of resistance in this species remain elusive, even as molecular markers of drug resistance in *P. falciparum* have been identified[1,2]. This stems from the lack of *in vitro* culture and transgenic systems for *P. vivax,* which has been essential for studying *P. falciparum* drug resistance. Several polymorphisms in candidate *P. vivax* resistance genes are suggested as possible molecular markers of drug resistance[13,14]. However, the current evidence for the role of many of these mutations in drug resistance is relatively weak, potentially leading to false conclusions that could impact drug policy. Validated drug resistance mutations could be used to rapidly and cost-effectively survey drug resistance in an area over time. This information is essential for identifying the most effective treatment policy in a region and determining when it is appropriate to switch between first line antimalarials. Understanding the molecular basis of drug resistance and development of heterologous model systems to study candidate resistance gene will be useful for the development of future antimalarials and to develop effective treatment policy.

P. vivax orthologs of known *P. falciparum* drug resistance genes are the focus of much research into *P. vivax* drug resistance. However, only one gene, *pvdhfr*, has been validated as a drug resistance gene.[8,13–15] This could be due to biological differences between each species, which could lead to different genes mediating resistance, or due to different selection pressure from drugs. Despite the lack of molecular evidence for mutations in these genes to mediate *P. vivax* drug resistance, several studies have still cited them as markers of drug resistance and recommended changes in treatment policy[16–19]. The use of SNPs in candidate resistance genes as molecular markers of drug resistance without genetic validation emphasizes the urgent need to identify and validate *P. vivax* drug resistance markers to inform malaria control policy. *ex vivo* assays and clinical treatment failure in the presence of adequate drug concentrations in the blood are the only

reliable methods currently available for determining the level of drug resistance in a population. Verified genetic markers of drug resistance could be used to replace these laborious and costly approaches, allowing high-throughput and frequent sampling of parasite populations that can be used to inform drug policy in that region. Molecular surveillance for resistance markers would allow treatment regimens to be changed upon the emergence of significant drug resistance in a population[20,21]. Using population genomics and evolutionary biology approaches can help identify *P. vivax* candidate resistance genes beyond those based on orthology to *P. falciparum*. Identifying new candidate *P. vivax* drug resistance genes may reveal novel *P. vivax* specific mechanisms of drug resistance that have remained largely unexplored. Candidate resistance markers will need to be validated in a heterologous system, such as using transgenic systems in *P. knowlesi* or *P. cynomolgi,* two parasite species closely related biologically and evolutionarily to *P. vivax*[22,23].

Understanding the evolutionary and population dynamics of drug resistance will be critical for molecular surveillance, to both identify when these alleles arise and understand how they move throughout and between populations. Investigation of candidate *P. vivax* drug resistance genes using advances in *in vitro* culture, *ex vivo* assays, and sequencing will help make this possible and improve *P. vivax* control and elimination efforts.

1.2 *P. vivax* drug resistance

While there is strong evidence of *P. vivax* resistance to CQ in Papua New Guinea, and reports of CQR in the Brazilian Amazon and Ethiopia, little is known about the molecular basis of resistance in *P. vivax.* Below I summarize what is known about the molecular basis of *P. vivax* drug resistance and discuss several *P. vivax* drug resistance candidates.

1.2.1 *P. vivax* Multidrug Resistance Gene 1 (*pvmdr1*)

pvmdr1 is the ortholog of the *P. falciparum pfmdr1* gene, which mediates drug sensitivity to MQ, lumefantrine (LUM), and CQ in this species.[24] In both parasites, *mdr1* encodes a transmembrane protein localized to the digestive vacuole (DV), where the parasite digests host cell proteins and converts heme from hemoglobin into non-toxic hemozoin[25–27]. *pvmdr1*, due to its strong sequence conservation with *pfmdr1*, has subsequently become one of the primary candidate genes investigated in *P. vivax* drug sensitivity studies[28–30]. PvMDR1 localizes to the digestive vacuole when over-expressed in *P. knowlesi*, suggesting it plays a similar role to PfMDR1 in *P. vivax*[13].

PfMDR1 is thought to transport drugs, including CQ and MQ, into the DV[25,26,31]. Molecular epidemiological and functional genetic studies have associated amino acid changes at positions 86, 184, 1034, 1042, and 1246 in PfMDR1 with drug resistance[31,32]. CQ inhibits hemozoin formation in the DV, which leads to a toxic buildup of heme. The reduced amount of CQ in the DV leads to chloroquine resistance[31,33]. Mutations in *pfmdr1* are associated with CQR and with increased sensitivity to MQ and LUM[34,35]. MQ and LUM are thought to primarily act on cytoplasmic proteins, although their mechanisms of action are less clear than CQ. Increased transport of MQ and LUM by mutant PfMDR1 (and possibly by *pvmdr1*) alleles into the DV would sequester these compounds from their targets, thereby leading to reduced susceptibility[25–27].

Sequencing of *pvmdr1* across several regions of the world has revealed more than fifty polymorphisms in this gene and increases in *pvmdr1*copy number. These single nucleotide polymorphisms (SNPs) correspond to amino acid changes throughout the protein sequence, including in the ATP binding domains and multiple transmembrane regions.[36] No single SNP, or set of SNPs, have emerged as definitive drug resistance markers. However, six SNPs have been reported at high frequency in multiple studies in regions with reported drug resistance; (relative to

the Sal-1 reference) S513R, G698S, M908L, T958M, Y976F, and F1076L[29,30,37–39]. Amino acid changes at Y976F and F1076L, in particular, have been cited as possible markers of drug resistance[29,40,41]. However, these SNPs are also found in regions without reported CQR, making their association with drug resistance uncertain[40]. The T958M, Y976F, F1076L variants (and possibly others) in *Pvmdr1,* have been shown to arise independently, even within the same population[42]. This finding suggests that *pvmdr1* mutations can arise on different genetic backgrounds, and that they have not been subject to a selective sweep, possibly due to a lack of direct pressure from drugs[42]. If, and how, these SNPs play a role in mediating *P. vivax* drug resistance remains to be directly demonstrated in functional studies.

Studies exploring the relationship between these SNPs and drug sensitivity do not firmly establish or deny a role for these mutations to cause drug resistance. A survey of isolates from China that associated sequencing of *pvmdr1* with *ex vivo* measurements of drug susceptibility found an association between M908L and reduced susceptibility to CQ, MQ, pyronaridine (PYN), piperaquine (PIP), quinine (QN), artemisinin (ART), and dihydroartemisinin (DHA)[43]. A study in Cambodia found a correlation between Y976F and F1076L mutations and resistance to MQ and PIP, but not CQ[44]. Another study found an association between Y976F and reduced susceptibility to CQ[41]. Yet multiple other studies have found no relationship between polymorphisms in *pvmdr1* and drug resistance[30,43,45,46]. These discrepancies could stem from methodological differences in *ex vivo* phenotyping, and because several studies are conducted in regions with low rates of CQR. Differences between studies may also simply reflect real geographical differences. The differences between studies highlight the critical need to evaluate these polymorphisms in an *in vitro* culture system to assess their impact on drug resistance. Two *pvmdr1* haplotypes were episomally over-expressed in a *P. knowlesi* model system, both of which encoded the 698S, 908L, and 958T

mutations, but did not alter sensitivity to either CQ or MQ[13]. suggesting that at least in this genetic background and expression system, these mutations did not confer resistance. Another limitation is that many studies only sequence part of the gene, thus constraining and biasing our understanding of mutations to these regions of the gene.

Copy number variation (CNV) of *pvmdr1* is distributed globally at variable frequencies (7-31.6%), including Thailand, Cambodia, Laos, India, Brazil, French Guinea, and other countries [38,39,41,45,47–49]. CNVs in *pvmdr1* have been more clearly implicated as a potential cause of *P. vivax* drug resistance[41,47,48]. Sequencing of isolates in Thailand and Indonesia found a correlation between increases in IC_{50} of amodiaquine (AQ), ART, and MQ, measured by *ex vivo* drug susceptibility assays, and *pvmdr1* amplification[41,50]. Notably, this mirrors the similar effect of *pfmdr1* amplification on MQ resistance in *P. falciparum*[51].

1.2.2 *P. vivax* Chloroquine Resistance Transporter (*pvcrt*)

The *pvcrt* gene (also referred to as *pvcrt-o*) emerged as a candidate drug resistance gene due to its orthology with the *pfcrt* gene that mediates CQR in *P. falciparum* (**Fig 2A)**. The *pfcrt* gene was first identified as a determinant of CQR in a genetic cross between a CQ-sensitive parasites line and CQR parasites[52]. Numerous *pfcrt* polymorphisms have been implicated in causing CQ drug resistance[53–57]. Additionally, polymorphisms in *pfcrt* are associated with PIP resistance, and conversely, increased sensitivity to MQ and ART[55,56,58,59].

Unlike *pfcrt*, and in contrast to *pvmdr1*, very few SNPs (~10) in *pvcrt* have been reported, but most occur at very low frequency[28,38,60]. The most common PvCRT polymorphism is a lysine insertion at position 10 (K10)[28,38,60,61]. The K10 insertion has been observed in both Southeast Asian and South American parasites, but no association between the K10 insertion and *in vitro P. vivax* drug resistance has been found[17,60,61]. It is currently unknown if *pvcrt* variants associated

with CQR have a deleterious impact on parasite fitness, which is the case for *P. falciparum,* and could explain the low frequency of *pvcrt* mutations[62,63].

There is evidence that increased expression of *pvcrt* potentially mediates CQR[64–66]. Sá and colleagues recently performed a *P. vivax* genetic cross between the CQR *P. vivax* NIH-1993-R line and the CQS NIH-1993-S line. Bulk segregant analysis of blood-stage progeny implicated a 76 KB region on chromosome one, which includes *pvcrt*, as having a role in CQR. The study found no SNPs in *pvcrt but* did identify a TGAAGH motif with an increased number of repeats both upstream of the 5' UTR of *pvcrt,* as well as a deletion within intron nine of the gene, and that CQR progeny had increased expression of *pvcrt*.

Over-expression of *pvcrt* has been linked to drug resistance, as reported in several studies studying CQ treatment failure of *P. vivax* patients in Brazil[47,61,64]. However, a study of *pvcrt* expression in parasites from Indonesia, where there is high-grade drug resistance, found no relationship between *pvcrt* expression levels and *ex vivo* susceptibility to CQ, PIP, MQ, and Artesunate (AS)[67]. As with *pvmdr1*, whether these discrepancies are due to differences in the genetic background, such as *pvmdr1* polymorphisms and CNVs, or whether they are due to technical differences will require further investigation.

1.2.3 *P. vivax* Dihydrofolate Reductase-Thymidylate Synthase (*pvdhfr-ts*) and Dihydropteroate Synthase (*pvdhps*)

The essential enzymes dihydrofolate reductase-thymidylate synthase (PvDHFR-TS*)* and dihydropteroate synthase (PvDHPS) are involved in folate synthesis and are the targets of pyrimethamine and sulfadoxine (SP), respectively[8,9,68–71]. SP has long been used to treat *P. falciparum*, but is now primarily only used for intermittent preventive treatment in pregnancy due to widespread resistance[72]. PfDHPS and PfDHFR-TS are well understood in *P. falciparum* and highly conserved in *P. vivax*[73]. While SP has not been intentionally used for *P. vivax* treatment*,*

mutations in PvDHPS and PvDHFR-TS, conserved mutations known to confer resistance in *P. falciparum* are widespread in *P. vivax* populations around the world[74–76]. These conserved mutations are associated with *P. vivax* SP treatment failure [74–76]. Genomic surveillance of *pvdhfr-ts* and *pvdhps* could help understand which regions these drugs could be used to cheaply and safely treat *P. vivax* infection, although SP has not been recommended for *P. vivax*[73]. *pvdhfr-ts* and *pvdhps* also represent a key example of how molecular studies could be performed to characterize other resistance alleles and monitor the spread of resistance.

Amino acid changes in PvDHFR-TS at positions N50I, S58R, S117N, and I173L exist as double, triple, and quadruple mutants at frequencies ranging from as 20-90% in Malaysia, Thailand, Papua New Guinea, and Indonesia[78–80]. These mutations align with mutations N51I, C59R, S108N, and I164L in PfDHFR-TS, which are associated with pyrimethamine resistance [9,81]. An additional F57L mutation in PvDHFR-TS, which has no *P. falciparum* equivalent, has also been reported at high frequency as a double mutant with 58R or 117N, and a quadruple mutant with 58R/61M/117N[8,9,68,69].

Expression of mutant PvDHFR-TS in a yeast system demonstrated that these mutations confer high levels of pyrimethamine resistance. Single mutations at positions 57L and 117N, and double or triple mutants, 58R/117N and 117N/173L, 58R/117N/173L resulted in a 50-, 87-, 460-, 700- and 500-fold increase in resistance to pyrimethamine, respectively[9]. Auliff et al episomally expressed PvDHFR-TS in *P. falciparum* to characterize a single mutation, 117N. PvDHFR-TS alleles with 177N had 46-, 6-, 2- and 6-fold increases in resistance, relative to wild-type, to pyrimethamine, clociguanil, WR99210, and cycloguanil, respectively. PvDHFR-TS double mutant haplotypes of 57L/117T and 58R/117T, resulted in 67- and 114-fold increases in pyrimethamine resistance, 6- and a 4-fold increase in cycloguanil resistance, 10- and a 26-fold

increase in clociguanil resistance, and 22- and 10-fold increases in WR99210 resistance for each haplotype, respectively. They also found a PvDHFR-TS quadruple mutant (57L/58R/61M/117T) resulted in high resistance to pyrimethamine, cycloguanil, chlorcycloguan, and WR99210, with resistance increases of 8497-, 746-, 2565-, and 44-fold, respectively. Interestingly, expression of a triple PvDHFR-TS mutant (58R/61M/117T) resulted in susceptibilities similar to wild-type for all drugs except for pyrimethamine, which had a 58-fold increase in resistance. The authors hypothesized that 61M may be a compensatory mutation to offset possible fitness costs with carrying other resistance mutations. Studies assessing PvDHFR-TS mutants and SP treatment outcomes have confirmed that these results correlate with *in vivo* efficacy. SP treatment failure is 23 times more likely to occur in patients infected with *P. vivax* parasites containing the 57L/58R/61M/117T haplotype[69,82]. Structure analysis of PvDHFR-TS showed that a mutation at position 117 leads to steric conflict with pyrimethamine binding, resulting in resistance, similar to the equivalent mutation at position 108 in PfDHFR-TS[83]. These data suggest that *P. vivax* resistance to antifolates arises from molecular changes and is highly conserved with *P. falciparum*.

Mutations at amino acids 383 and 553 in PvDHPS, correspond with known drug resistance mutations 437 and 581 In PfDHPS[69,84]. Similar to SP resistance in *P. falciparum*, mutations in PvDHPS alone are thought to not be sufficient to provide resistance to SP[84]. Furthermore, the wild-type PvDHPS allele at position 585 is a Valine, which aligns with position 613 in PfDHPS, and is thought to reduce the binding to sulfadoxine[85]. This V585 allele is found commonly in isolates, suggesting possible innate resistance to sulfadoxine in this species.[69,85] Structure analysis of PvDHPS mutants supports the role of these mutations in providing resistance to sulfadoxine by reducing binding affinity to the mutant enzyme.[86] Particularly, parasites with both PvDHPS 383G

and 553G mutations and mutant PvDHFR-TS alleles, have been implicated in SP treatment failure[84].

Studies looking at the distribution of *pvdhfr-ts* alleles in natural populations have found a high prevalence (80-90%) of double, triple, and quadruple mutations in Malaysia, Thailand, India, Indonesia, Madagascar, and China, which all have high SP treatment failure rates[8,14,37,78,79,87]. Global and regional population genetic studies of *P. vivax* have found *pvdhfr-ts* and *pvdhps* in regions of the genome, with strong evidence of selection by linkage disequilibrium and integrated haplotype score metrics[88,89]. Strong signals of genetic differentiation were found between Ethiopian populations, which have a low prevalence of *pvdhfr-ts* mutants, compared to Indonesian or Thai populations, which have a high prevalence of *pvdhfr-ts* mutants, providing evidence of region-specific drug pressure[78]. Furthermore, a study comparing *P. vivax* isolates in Indonesia, Malaysia, and Thailand found signals of genetic differentiation of *pvdhfr-ts* and *pvdhps* between regions, which all have a high prevalence of *pvdhfr-ts* mutants, suggesting that mutant haplotypes can arise on different genetic backgrounds[88].

Two studies in India found a higher prevalence of *pvdhfr-ts* mutations in regions where *P. falciparum* and *P. vivax* are co-endemic than regions where *P. vivax* is the dominant parasite[87,90]. These results suggest that pressure of SP treatment of *P. falciparum* can lead to co-selection of *pvdhfr-ts* mutant alleles, which could also occur for other drugs, including ACT partner drugs.

1.2.4 Other Candidate *P. vivax* Resistance Genes

Population genetic studies looking for evidence of positive selection have identified several candidate *P. vivax* drug resistance genes[78,88,91–94]. These include the candidate genes described above (*pvmdr1, pvcrt, pvdhfr-ts, pvdhps*), as well as several additional genes of interest.

Several studies of parasites in the Greater Mekong Subregion (GMS) found evidence of selective sweeps via extended haplotype homozygosity (XP-EHH) and iHS tests around the *pvmdr2* and *pvmrp2* genes in Cambodia[88,94,95]. Another study in the Peruvian Amazon found a higher number of SNPs in *pvmrp2*, though this study did not evaluate any measures of selection on this gene[96]. The *P. falciparum* ortholog of *pvmrp2* has been implicated in conferring resistance to CQ, MQ, and PIP[97–99]. Similarly, *pfmdr2*, the *P. falciparum* ortholog of *pvmdr2,* has been implicated in pyrimethamine resistance[81].

Several studies have monitored the *P. vivax* ortholog of *pfkelch13*, *pvkelch12*, for polymorphisms that may lead to ART resistance[91,100,101]. While several polymorphisms have been found in *pvkelch12*, they do not align to the respective mutations in *pfkelch13* that mediate artemisinin resistance[91,100,101]. These results suggest a lack or reduced selection pressure from ART on *pvkelch12* to date. Notably, there is a lack of persistence of *pvkech12* SNPs over time in the GMS, and little polymorphism in *pvkelch12* in Papua New Guinea, where one might expect selection pressure from ART indirectly from *P. falciparum* treatment, or directly as the first-line therapy in this region[91,101]. These results suggest a lack of strong selection pressure from ART on *pvkelch12* at this time[91,101].

A study of *P. vivax* parasites in the Peruvian Amazon uncovered possible evidence of increased copy number of PVP01_0312700. PVP01_0312700 encodes *pvdmt2*, a homolog of the *E. coli emrE* gene. *emrE is* a multidrug transporter, but its *Plasmodium* homologs have not been evaluated for a role in drug resistance in any *Plasmodium* spp[93]. Additionally, a study of *P. vivax* found evidence of selection and significant genetic differentiation of the I165V variant in the *P. vivax plasmepsin IV* gene*,* an ortholog of the *P. falciparum plasmepsin II*, between Malaysian parasite populations, where there is a higher reported prevalence of CQR, and Thailand, where

there is reportedly a lower prevalence of CQR[88]. *P. falciparum plasmepsin II* has been implicated in mediating PIP resistance in this species [102]. The study also found significant genetic differentiation of variants in the candidate drug/metabolite transporters *pvdmt1*, and a CG2-related protein (PVP01_1450700) between Thailand and Malaysia[88]. While these genes have not previously been the subject of intense research focus, the evidence of recent selection or increased copy number variation at loci encompassing them suggests further research effort is merited.

P. vivax multidrug resistance protein 1 (*pvmrp1*) has emerged as a candidate drug resistance gene due to signals of selection in population genetic studies[92,93]. *pvmrp1* is an ortholog of the *pfmrp1* gene, which transports glutathione adducts out of the parasite, a process thought to help the parasite regulate oxidative stress[103,104]. *pfmrp1* knockouts in *P. falciparum* have increased susceptibility to quinine, CQ, ART, and PQ, as well as reduced parasite growth rate *in vitro*[104]. SNPs in *pfmrp1* have also been associated with AL resistance, particularly the amino acid change, I876V, which was found in recurrent infections after treatment with AL[105].

Genetic sequencing of *P. vivax* samples in South America has found strong signals of molecular evolution in this gene, with a high ratio of non-synonymous substitutions per synonymous site relative to the number of synonymous substitutions per synonymous site (d_N/d_S) [92,93]. Sequencing of a Peruvian isolate also demonstrated evidence of *pvmrp1* gene amplification[93]. Studies in Sudan, Thailand, Indonesia, and Cambodia found that *pvmrp1* is situated in a long region of homozygosity, which indicates recent selection[78,88,94,106].

Sequencing of a *P. vivax* isolate from a patient, with no mutations in CYP genes, that experienced PQ failure, found several SNPs in *pvmrp1*[107]. PQ and TQ are the only approved drugs that can eliminate hypnozoites, which is required to achieve radical cure of *P. vivax* [108–111]. CYPD26 oxidizes PQ into hydroxyl-metabolites, whose oxidation generates quinoneimine, and

subsequently generates hydrogen peroxide (H_2O_2) [112]. Qinoneimine is also a substrate for CPR, CYPD26's redox partner, which leads to the accumulation of H_2O_2 and subsequent antimalarial activity through oxidative stress in the parasite. PQ tolerance, leading to relapse of infection, is rare, but has been reported in cases following the correct treatment regimen[108,113]. Knockouts of *pfmrp1* have increased sensitivity to PQ in *P. falciparum*, further supporting that this protein plays a role in mediating tolerance to PQ, although the active metabolites were not tested[104]. The experimental evidence of *pfmrp1* mediating PQ tolerance and evidence of selection pressure on *pvmrp1* have led to this gene as a candidate for mediating PQ tolerance in *P. vivax*[92,93,106,114]. *pvmrp1* does not have a syntenic ortholog in *P. knowlesi.* Dharia and colleagues proposed a model of *pvmrp1* mediating PQ tolerance, where *pvmrp1* mutations improve the transport of glutathione adducts resulting from PQ treatment, and thus allow for PQ tolerance by mitigating oxidative damage. [92]

1.3. Assessing drug resistance in *P. vivax*

Linking *P. vivax* mutations to a drug sensitivity phenotype, *in viv*o and *ex vivo*, is essential to validate SNPs resistance markers. However, identification of resistance mutations in *P. vivax* remains difficult for some drugs, including CQ, in large part because defining drug-resistant *P. vivax* phenotypes is itself is difficult. I discuss the strengths and limitations of currently available methods to measure drug sensitivity in *P. vivax* to the different antimalarials used clinically, and the current evidence of drug resistance globally.

1.3.1 Measuring *in vivo* drug resistance

Identifying *in vivo* resistance to CQ and ACT is difficult in *P. vivax*, in part due to varying definitions of resistance. The clearest definition of CQR in *P. vivax* is the ability to grow in CQ concentrations that would normally kill (CQ concentration >100 ng/ml), although this cutoff for

resistance is subjective, and can miss identifying low-grade resistance[1,115]. A high rate of treatment failure by day 28 is often used as a measure of resistance[116–125]. However, most recurrences that occur after day 21, and as early as day 14, are associated with low blood concentrations of CQ (<100 ng/ml) at the time of treatment failure[116–125]. Different definitions of clinical resistance can lead to contrasting conclusions about the degree, or even the presence, of CQR[126]. Clinical studies that measure blood CQ levels at the day of treatment failure, or detect a high frequency of early treatment failures (≤14 days), can paint a clearer picture of resistance[127–132]. Such studies have demonstrated clear CQ resistance in Myanmar, Thailand, Ethiopia, Bolivia, and Brazil, with the prevalence of CQ treatment failure ranging from ~0.5%-10%[127–132]. Regions of Indonesia and Papua New Guinea have significantly higher rates of CQR in *P. vivax* populations, where treatment failure occurs in 20-97% of patients, with blood CQ levels >100 ng/ml, often including a high frequency of early treatment failures[7,10,118,133–136]. Treatment failures in such regions have made it necessary to switch to ACTs as the first-line treatment.

Several other factors can complicate the association of clinical failure with drug resistance. *P. vivax* treatment with CQ is usually combined with PQ, or more recently TQ, to achieve a "radical cure" to clear hypnozoites. Clinical trials that delay the addition of PQ/TQ until day 28 will mask CQR due to the blood-stage activity of these drugs[118,131,137,138]. High levels of immunity can also suppress parasitemia following treatment, reducing the rate of recrudescence[139–142]. Recrudesce can mistakenly be classified as having lower rates of treatment failure and drug resistance, compared to regions with lower immunity[139–142]. Genotyping *P. vivax* at the time of treatment and recrudescence has been used to identify relapse, where parasite clones are observed at recrudescence that were not present in the initial population[108,143]. However, the presence of a

homologous clone before and after treatment cannot differentiate between recrudescence and relapse, limiting its use to ruling out recrudescence but not relapse.

ACTs remain clinically active against CQR *P. vivax*, including artemisinin (ART) in combination with MQ, artemether in combination with LUM (combination denoted as AL) or dihydroartemisinin (DHA) in combination with piperaquine (PIP)[130,132,134,144–148]. A recent meta-analysis of AL compared to DHA-PIP showed a higher recrudescence rate in patients treated with AL in the first 42 days[146]. This is likely due to the longer half-life of PIP compared to LUM, which provides a longer-lasting prophylactic effect against relapse from hypnozoite. [146] This observation is supported by the more similar rates of recrudescence by day 63 between drugs, as well as the significant effect of PQ in both treatments[146]. An alternative explanation for this observation is that the use of AL in Ethiopia against *P. falciparum* has led to the selection of LUM resistant variants in the *P. vivax* population.

SP has not been used to treat *P. vivax* due to the concept that it was naturally resistant in earlier literature, which was likely an artifact of hypnozoite relapse or use of an already resistant [77]. SP has historically been used to treat *P. falciparum* and is still used for intermittent preventive treatment in pregnancy (IPTp), which is indiscriminate of species[3]. Notably, *P. vivax* SP resistance is relatively common, even though it was never intentionally treated with SP, demonstrating strong pressure from the treatment of *P. falciparum* [87,90].

1.3.2 Culturable heterologous model systems and reverse genetics

P. vivax ex vivo drug susceptibility assays can help to identify drug resistance *P. vivax*[150]. However, *ex vivo* assays are limited to short single cycle assays with relatively poor growth compared to *P. falciparum*[150]. *Ex vivo* assays also require patient isolates, which can limit their accuracy because they cannot be well controlled for parasite stage and parasitemia[150]. Improved

growth conditions and the use of multiple-cycle assays would reduce the differences observed from starting parasite age and parasitemia[150]. Improved *P. vivax ex vivo* assays will allow for the study of lower-acting drugs, including TQ and assays could help determine drug susceptibility in natural isolates and link drug phenotypes to specific resistance loci[151]. However, validation of candidate resistance alleles will remain unobtainable in *P. vivax* until robust *ex vivo* short-term culture, or a continuous culture system, are developed. In place of a stable culture system, heterologous genetic systems will be invaluable for validation studies in *P. vivax*.

P. falciparum has been used to episomally express *P. vivax dhfr* alleles, which demonstrated the role of polymorphisms observed in the field in mediating pyrimethamine resistance[70]. The role and protein sequence of DHFR-TS, and its resistance mutations, are relatively well conserved between *Plasmodium* spp., as well as to higher eukaryotes[9]. In contrast, the sequence of other candidate resistance genes, as well as mutations found within them, are more divergent between *Plasmodium* spp., which may impact their accuracy as model systems. As such, the field has moved towards using *Plasmodium* species that are more closely related to *P. vivax*. *P. knowlesi* primarily infects primates but can also cause outbreaks in humans[152,153]. The parasite is more closely related to *P. vivax* than *P. falciparum* is[152,153]. *P. knowlesi* has recently been adapted to grow in human RBCs, which opens up the possibility of conducting molecular genetic experiments[152,153]. *P. knowlesi* is genetically tractable and a CRISPR/Cas9 system has been developed for use in this species. CRISPR/Cas9 has successfully been used to replace *pkdhps* with its *P. vivax* ortholog for vaccine studies[154]. *P. knowlesi* has also been used to overexpress two *pvmdr1* variants, both of which localized to the digestive vacuole[13]. Expression of these variants did not alter sensitivity to CQ or MQ[13]. Whether the lack of an effect is due to the *pvmdr1* variants tested, *pvmdr1* not being involved in drug resistance, or due to technical reasons, such as poor

expression levels or the presence of the wild-type PkMDR1, will require further study. Allelic replacement using CRISPR/Cas9 could replace the native *pkmdr1* with *pvmdr1* variants, which would ensure it is expressed from the native promoter in the absence of *pkmdr1*. The same methods could also be used to study other candidate resistance genes.

More recently, *P. cynomolgi*, another primate malaria that is even more closely related to *P. vivax,* has been adapted to *in vitro* culture in rhesus macaque RBCs[155]. The closer evolutionary relationship between *P. cynomolgi* and *P. vivax* may make it a better model for drug resistance mechanisms. Transgenic methods have been developed for *P. cynomolgi in vivo* using a primate model, which was used to produce fluorescent or luminescent reporter lines for high-throughput screening of compounds active against hypnozoites[156,157]. *P. cynomolgi* can also develop hypnozoites, providing it a significant advantage of using over *P. knowlesi* as a *P. vivax* model system. As such, *P. cynomolgi* could be used to investigate candidate resistance markers of PQ or TQ. Further development of transfection for *in vitro* cultured *P. cynomolgi* and a CRISPR/Cas9 system will allow the use of *P. cynomolgi* as a model for understanding *P. vivax* biology. Advances in these model systems will enable molecular genetic characterization of polymorphisms in candidate *P. vivax* genes on drug susceptibility.

1.3.3 Genomics and Transcriptomics

Advances in genomics and transcriptomics over the past decade have opened a range of possibilities for conducting population genetics and molecular epidemiological studies of *P. vivax*. Whole Genome Sequencing (WGS) of *P. vivax* directly from patient samples can be difficult because of the typically low parasitemia, and the subsequent difficulty of separating the *P. vivax* DNA from the background human genome[158,159]. Leukocyte depletion, which removes the background human DNA, and hybrid selection to enrich parasite DNA, has been used to

successfully sequence *P. vivax* genomes from clinical samples[89,159,160]. Similarly, selective whole genome amplification (SWGA) approaches, which use a highly processive polymerase to amplify the parasite genome from the background human genome, for WGS is a low-cost and easily scalable method for conducting WGS on *P. vivax* clinical samples[158].

These technologies have opened the possibility of conducting large-scale population molecular epidemiological studies in *P. vivax* and can help conduct genome-wide association studies (GWAS). GWAS requires sequencing of clinical *P. vivax* samples and identifying loci associated with a phenotype. Such approaches have successfully been used to identify drug resistance loci in *P. falciparum*. Previous methods of sequencing *P. vivax* samples were expensive and difficult to generate quality data at scale due to the low parasitemia during *P. vivax* infection. Recent advances like SWGA can circumvent these challenges and increase the ease of conducting *P. vivax* GWAS. GWAS could also be used to highlight selected loci under different selective pressure (i.e., from drugs) between regions.

Genomic epidemiology is the use of genomic information to inform epidemiological investigation. In the infectious disease field, genomic epidemiology is concerned with using genomics to understand how diseases spread and evolve as they move into and through populations. Applications of genomics to malaria epidemiology include tracking the spread of drug resistance alleles, identifying novel drug resistance alleles, understanding transmission intensity and disease burden, understanding patterns of connectivity and relatedness to track transmission, and studying the co-evolution of the parasites and their mosquito vectors, among other use cases[89,91,160]. Genomic epidemiology can be used to identify pockets of high malaria transmission and distinguish between locally transmitted and imported cases, and sources and sinks of malaria transmission. Particularly as countries approach malaria elimination, distinguishing between local

and imported cases is critical to stamp out the last vestiges of malaria transmission and successfully achieve and maintain malaria elimination. Other uses of molecular epidemiology include surveillance of *P. falciparum histidine-rich protein 2* and *histidine-rich protein 3* deletions, which can lead to false negatives on malaria rapid diagnostic tests, and molecular markers of drug resistance to inform malaria treatment policy[161–164].

A limitation of whole-genome approaches is that they can miss rare variants due to lower sequence coverage, which limits the ability to uncover rare variants that could cause drug resistance[165,166]. Targeted sequencing approaches, such as amplicon sequencing where genes are amplified in a PCR reaction in parallel and then sequenced, allows for deep sequencing of select loci to uncover low-frequency variants and uncover the full range of genetic variation in a population[165,167,168]. *P. vivax* amplicon sequencing has been used to differentiate between recrudescence, reinfection, or relapse as a cause of recurrent infection, which could help determine if a patient has drug-resistant *P. vivax*[167]. Amplicon sequencing in *P. falciparum* has been used to identify low-frequency SNPs in the drug resistance genes *pfcrt, pfkelch13, pfmdr1, pfmrp1,* and *pfdhfr-ts*[166,169]. Development of *P. vivax* amplicon panels for candidate resistance genes could help identify SNPs associated with drug resistance[166,167,169].

The low parasitemia and lack of a culture system in *P. vivax* have also traditionally hindered the use of transcriptomics to study gene expression. Recent advances in the ability to conduct transcriptomics from clinical samples and single-cell transcriptomics from monkey-adapted lines demonstrate the feasibility to conduct *P. vivax* transcriptomic studies[170–172]. *P. vivax* transcriptomes from Cambodian patients treated with or without CQ, showed no differential expression of the candidate CQR genes *pvmdr1* and *pvcrt* between the two treatment groups[170]. The lack of an association could be due to a lack of drug resistance alleles in this sample set, or a

lack of change in gene expression in response to drug pressure[170]. Comparative transcriptomics between CQR and CQS patient isolates could identify candidate resistance loci that have elevated expression in CQR parasites since increased expression mediates *P. falciparum* drug resistance, and there is evidence that the same is true in *P. vivax*[65,170,17265,170,172].

1.4 Looking Forward

The *Plasmodium vivax* field is in a new era of epidemiology and molecular biology research focused on this parasite of public health importance. Advances in genomics and molecular genetics allow for the use of novel techniques to answer critical questions about parasite transmission dynamics, epidemiology, and drug resistance. As *Plasmodium vivax* elimination efforts accelerate, advances in knowledge of transmission dynamics and epidemiology can help support these efforts to highlight epidemiological factors that will impact the success of parasite elimination. Additionally, these new tools can help answer critical questions about the molecular basis of *P. vivax* drug resistance, a major threat to the success of *P. vivax* control and elimination. Finally, Genomic epidemiological techniques are promising for the malaria field to inform policy. The development of tools and frameworks that can synthesize and translate genomic information to inform malaria control policy will be critical as genomic epidemiological techniques enter common practice.

In this thesis, I present several studies using new tools in genomics and Plasmodium molecular biology to answer questions about the epidemiology and drug resistance biology of *P. vivax*. First, I present a use-case of genomic epidemiolocal data to understand *P. vivax* transmission in a low elimination setting. I also demonstrate work using population genomics data to identify signals of the evolution of candidate *P. vivax* drug resistance genes and highlight some possible other candidate transporters genes that exhibit strong signals of species-specific selection due to

drug pressure. Finally, I used a *P. vivax* population genomics dataset to identify a globally representative sample of circulating diversity in the *pvmdr1* gene, one of the strongest candidate *P. vivax* drug resistance genes. I then evaluate each haplotype in a *P. knowlesi* model system for its effects on mediating drug resistance and find that several mutations do lead to resistance for various antimalarial compounds. These results will provide the field with a set of validating resistance markers in *pvmdr1* to use in molecular surveillance and to inform *P. vivax* treatment policy.

1.5 References

1. Price, R. N. *et al.* Global extent of chloroquine-resistant Plasmodium vivax: a systematic review and meta-analysis. *Lancet Infect. Dis.* **14**, 982–991 (2014).
2. WHO. *World Malaria Report 2019*. vol. 1 (WHO, 2019).
3. WHO. *World Malaria Report 2019*. vol. 1 (WHO, 2019).
4. Haldar, K., Bhattacharjee, S. & Safeukui, I. Drug resistance in Plasmodium. *Nat. Rev. Microbiol.* **16**, 156–170 (2018).
5. Wicht, K. J., Mok, S. & Fidock, D. A. Molecular Mechanisms of Drug Resistance in Plasmodium falciparum Malaria. *Annu. Rev. Microbiol.* **74**, 431–454 (2020).
6. Baird, J. K. *et al.* In vivo resistance to chloroquine by Plasmodium vivax and Plasmodium falciparum at Nabire, Irian Jaya, Indonesia. *Am. J. Trop. Med. Hyg.* **56**, 627–631 (1997).
7. Fryauff, D. J. *et al.* Chloroquine-resistant Plasmodium vivax in transmigration settlements of West Kalimantan, Indonesia. *Am. J. Trop. Med. Hyg.* **59**, 513–518 (1998).
8. Hastings, M. D. *et al.* Novel Plasmodium vivax dhfr alleles from the Indonesian Archipelago and Papua New Guinea: association with pyrimethamine resistance determined by a Saccharomyces cerevisiae expression system. *Antimicrob. Agents Chemother.* **49**, 733–740 (2005).
9. Hastings, M. D. & Sibley, C. H. Pyrimethamine and WR99210 exert opposing selection on dihydrofolate reductase from Plasmodium vivax. *Proc. Natl. Acad. Sci.* **99**, 13137–13141 (2002).
10. Asih, P. B. S. *et al.* Phenotyping clinical resistance to chloroquine in Plasmodium vivax in northeastern Papua, Indonesia. *Int. J. Parasitol. Drugs Drug Resist.* **1**, 28–32 (2011).
11. Grigg, M. J. *et al.* Efficacy of Artesunate-mefloquine for Chloroquine-resistant Plasmodium vivax Malaria in Malaysia: An Open-label, Randomized, Controlled Trial. *Clin. Infect. Dis.* **62**, 1403–1411 (2016).
12. Schneider, K. A. & Escalante, A. A. Fitness components and natural selection: why are there different patterns on the emergence of drug resistance in Plasmodium falciparum and Plasmodium vivax? *Malar. J.* **12**, 15 (2013).
13. Verzier, L. H., Coyle, R., Singh, S., Sanderson, T. & Rayner, J. C. Plasmodium knowlesi as a model system for characterising Plasmodium vivax drug resistance candidate genes. *PLoS Negl. Trop. Dis.* **13**, e0007470 (2019).
14. Barnadas *et al.* A new high-throughput method for simultaneous detection of drug resistance associated mutations in Plasmodium vivax dhfr, dhps and mdr1 genes. *Malar. J.* **10**, 282 (2011).
15. Dharia, N. V. *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 A* **107**, 20045–20050 (2010).
16. Cheong, F.-W., Dzul, S., Fong, M.-Y., Lau, Y.-L. & Ponnampalavanar, S. Plasmodium vivax drug resistance markers: Genetic polymorphisms and mutation patterns in isolates from Malaysia. *Acta Trop.* **206**, 105454 (2020).
17. Zhao, Y. *et al.* Molecular surveillance for drug resistance markers in Plasmodium vivax isolates from symptomatic and asymptomatic infections at the China–Myanmar border. *Malar. J.* **19**, 281 (2020).
18. Tantiamornkul, K., Pumpaibool, T., Piriyapongsa, J., Culleton, R. & Lek-Uthai, U. The prevalence of molecular markers of drug resistance in Plasmodium vivax from the border regions of Thailand in 2008 and 2014. *Int J Parasitol Drugs Drug Resist* **8**, 229–237 (2018).
19. Wang, X. *et al.* Prevalence of molecular markers associated with drug resistance of Plasmodium vivax isolates in Western Yunnan Province, China. *BMC Infect Dis* **20**, 307 (2020).

20. Ehrlich, H. Y., Jones, J. & Parikh, S. Molecular surveillance of antimalarial partner drug resistance in sub-Saharan Africa: a spatial-temporal evidence mapping study. *Lancet Microbe* **1**, e209–e217 (2020).
21. Nkhoma, S., Molyneux, M. & Ward, S. Molecular surveillance for drug-resistant Plasmodiumfalciparum malaria in Malawi. *Acta Trop.* **102**, 138–142 (2007).
22. Loy, D. E. *et al.* Evolutionary history of human Plasmodium vivax revealed by genome-wide analyses of related ape parasites. *PNAS* **115**, E8450–E8459 (2018).
23. Loy, D. E. *et al.* Out of Africa: origins and evolution of the human malaria parasites Plasmodium falciparum and Plasmodium vivax. *Int. J. Parasitol.* **47**, 87–97 (2017).
24. Price, R. N. *et al.* The pfmdr1 Gene Is Associated with a Multidrug-Resistant Phenotype in Plasmodium falciparum from the Western Border of Thailand. *Antimicrob. Agents Chemother.* **43**, 2943–2949 (1999).
25. Rohrbach, P. *et al.* Genetic linkage of pfmdr1 with food vacuolar solute import in Plasmodium falciparum. *EMBO J.* **25**, 3000–3011 (2006).
26. Reiling, S. J. & Rohrbach, P. Monitoring PfMDR1 transport in Plasmodium falciparum. *Malar. J.* **14**, 270 (2015).
27. Sanchez, C. P., Rotmann, A., Stein, W. D. & Lanzer, M. Polymorphisms within PfMDR1 alter the substrate specificity for anti-malarial drugs in Plasmodium falciparum. *Mol. Microbiol.* **70**, 786–798 (2008).
28. Barnadas *et al.* Plasmodium vivax Resistance to Chloroquine in Madagascar: Clinical Efficacy and Polymorphisms in *pvmdr1* and pvcrt-o Genes. *Antimicrob. Agents Chemother.* **52**, 4233–4240 (2008).
29. Brega, S. *et al.* Identification of the Plasmodium vivax mdr-Like Gene (*pvmdr1*) and Analysis of Single-Nucleotide Polymorphisms among Isolates from Different Areas of Endemicity. *J. Infect. Dis.* **191**, 272–277 (2005).
30. Orjuela-Sánchez, P. *et al.* Analysis of single-nucleotide polymorphisms in the crt-o and mdr1 genes of Plasmodium vivax among chloroquine-resistant isolates from the Brazilian Amazon region. *Antimicrob. Agents Chemother.* **53**, 3561–3564 (2009).
31. Veiga, M. I. *et al.* Globally prevalent PfMDR1 mutations modulate Plasmodium falciparum susceptibility to artemisinin-based combination therapies. *Nat. Commun.* **7**, 11553 (2016).
32. Wurtz, N. *et al.* Role of Pfmdr1 in In Vitro Plasmodium falciparum Susceptibility to Chloroquine, Quinine, Monodesethylamodiaquine, Mefloquine, Lumefantrine, and Dihydroartemisinin. *Antimicrob. Agents Chemother.* **58**, 7032–7040 (2014).
33. Olafson, K. N., Ketchum, M. A., Rimer, J. D. & Vekilov, P. G. Mechanisms of hematin crystallization and inhibition by the antimalarial drug chloroquine. *Proc. Natl. Acad. Sci.* **112**, 4946–4951 (2015).
34. Reed, M. B., Saliba, K. J., Caruana, S. R., Kirk, K. & Cowman, A. F. Pgh1 modulates sensitivity and resistance to multiple antimalarials in Plasmodium falciparum. *Nature* **403**, 906–909 (2000).
35. Sidhu, A. B. S., Valderramos, S. G. & Fidock, D. A. pfmdr1 mutations contribute to quinine resistance and enhance mefloquine and artemisinin sensitivity in Plasmodium falciparum. *Mol. Microbiol.* **57**, 913–926 (2005).
36. Sá *et al.* Plasmodium vivax: allele variants of the mdr1 gene do not associate with chloroquine resistance among isolates from Brazil, Papua, and monkey-adapted strains. *Exp. Parasitol.* **109**, 256–259 (2005).
37. Huang, B. *et al.* Molecular surveillance of pvdhfr, pvdhps, and pvmdr-1 mutations in Plasmodium vivax isolates from Yunnan and Anhui provinces of China. *Malar. J.* **13**, 346 (2014).
38. Joy, S. *et al.* Drug resistance genes: pvcrt-o and pvmdr-1 polymorphism in patients from malaria endemic South Western Coastal Region of India. *Malar. J.* **17**, 40 (2018).

39. Vargas-Rodríguez, R. del C. M., da Silva Bastos, M., Menezes, M. J., Orjuela-Sánchez, P. & Ferreira, M. U. Single-Nucleotide Polymorphism and Copy Number Variation of the Multidrug Resistance-1 Locus of Plasmodium vivax: Local and Global Patterns. *Am. J. Trop. Med. Hyg.* **87**, 813–821 (2012).
40. Spotin, A. *et al.* Global assessment of genetic paradigms of *Pvmdr1* mutations in chloroquine-resistant Plasmodium vivax isolates. *Trans. R. Soc. Trop. Med. Hyg.* **114**, 339–345 (2020).
41. Suwanarusk *et al.* Amplification of *pvmdr1* associated with multidrug-resistant Plasmodium vivax. *J. Infect. Dis.* **198**, 1558–1564 (2008).
42. Schousboe, M. L. *et al.* Multiple Origins of Mutations in the mdr1 Gene—A Putative Marker of Chloroquine Resistance in *P. vivax*. *PLoS Negl. Trop. Dis.* **9**, e0004196 (2015).
43. Li, J. *et al.* Ex vivo susceptibilities of Plasmodium vivax isolates from the China-Myanmar border to antimalarial drugs and association with polymorphisms in *Pvmdr1* and Pvcrt-o genes. *PLoS Negl. Trop. Dis.* **14**, e0008255 (2020).
44. Chaorattanakawee, S. *et al.* Measuring ex vivo drug susceptibility in Plasmodium vivax isolates from Cambodia. *Malar. J.* **16**, 392 (2017).
45. Faway, E. *et al.* Plasmodium vivax multidrug resistance-1 gene polymorphism in French Guiana. *Malar. J.* **15**, 540 (2016).
46. Nyunt, M. H. *et al.* Clinical and molecular surveillance of drug resistant vivax malaria in Myanmar (2009-2016). *Malar. J.* **16**, 117 (2017).
47. Costa, G. L. *et al.* Assessment of copy number variation in genes related to drug resistance in Plasmodium vivax and Plasmodium falciparum isolates from the Brazilian Amazon and a systematic review of the literature. *Malar. J.* **16**, 152 (2017).
48. Imwong *et al.* Gene Amplification of the Multidrug Resistance 1 Gene of Plasmodium vivax Isolates from Thailand, Laos, and Myanmar. *Antimicrob. Agents Chemother.* **52**, 2657–2659 (2008).
49. Khim, N. *et al.* Effects of Mefloquine Use on Plasmodium vivax Multidrug Resistance. *Emerg. Infect. Dis.* **20**, 1637–1644 (2014).
50. Auburn, S. *et al.* Genomic Analysis Reveals a Common Breakpoint in Amplifications of the Plasmodium vivax Multidrug Resistance 1 Locus in Thailand. *J. Infect. Dis.* **214**, 1235–1242 (2016).
51. Preechapornkul, P. *et al.* Plasmodium falciparum pfmdr1 Amplification, Mefloquine Resistance, and Parasite Fitness. *Antimicrob. Agents Chemother.* **53**, 1509–1515 (2009).
52. Su, X., Kirkman, L. A., Fujioka, H. & Wellems, T. E. Complex polymorphisms in an approximately 330 kDa protein are linked to chloroquine-resistant *P. falciparum* in Southeast Asia and Africa. *Cell* **91**, 593–603 (1997).
53. Ecker, A., Lehane, A. M., Clain, J. & Fidock, D. A. PfCRT and its role in antimalarial drug resistance. *Trends Parasitol.* **28**, 504–514 (2012).
54. Fidock, D. A. *et al.* Mutations in the *P. falciparum* Digestive Vacuole Transmembrane Protein PfCRT and Evidence for Their Role in Chloroquine Resistance. *Mol. Cell* **6**, 861–871 (2000).
55. Johnson, D. J. *et al.* Evidence for a Central Role for PfCRT in Conferring Plasmodium falciparum Resistance to Diverse Antimalarial Agents. *Mol. Cell* **15**, 867–877 (2004).
56. Pulcini, S. *et al.* Mutations in the Plasmodium falciparum chloroquine resistance transporter, PfCRT, enlarge the parasite's food vacuole and alter drug sensitivities. *Sci. Rep.* **5**, 14552 (2015).
57. Summers, R. L. *et al.* Diverse mutational pathways converge on saturable chloroquine transport via the malaria parasite's chloroquine resistance transporter. *Proc. Natl. Acad. Sci.* **111**, E1759–E1767 (2014).

58. Phyo, A. P. *et al.* Dihydroartemisinin-piperaquine versus chloroquine in the treatment of Plasmodium vivax malaria in Thailand: a randomized controlled trial. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **53**, 977–984 (2011).
59. van der Pluijm, R. W. *et al.* Determinants of dihydroartemisinin-piperaquine treatment failure in Plasmodium falciparum malaria in Cambodia, Thailand, and Vietnam: a prospective clinical, pharmacological, and genetic study. *Lancet Infect. Dis.* **19**, 952–961 (2019).
60. Noisang, C., Meyer, W., Sawangjaroen, N., Ellis, J. & Lee, R. Molecular Detection of Antimalarial Drug Resistance in Plasmodium vivax from Returned Travellers to NSW, Australia during 2008–2018. *Pathogens* **9**, 101 (2020).
61. Silva, S. R. *et al.* Chloroquine resistance is associated to multi-copy pvcrt-o gene in Plasmodium vivax malaria in the Brazilian Amazon. *Malar. J.* **17**, 267 (2018).
62. Ord, R. *et al.* Seasonal carriage of pfcrt and pfmdr1 alleles in Gambian Plasmodium falciparum imply reduced fitness of chloroquine-resistant parasites. *J. Infect. Dis.* **196**, 1613–1619 (2007).
63. Petersen, I. *et al.* Balancing drug resistance and growth rates via compensatory mutations in the Plasmodium falciparum chloroquine resistance transporter. *Mol. Microbiol.* **97**, 381–395 (2015).
64. Melo, G. C. *et al.* Expression levels of pvcrt-o and pvmdr-1 are associated with chloroquine resistance and severe Plasmodium vivax malaria in patients of the Brazilian Amazon. *PloS One* **9**, e105922 (2014).
65. Sá, J. M. *et al.* Plasmodium vivax chloroquine resistance links to pvcrt transcription in a genetic cross. *Nat. Commun.* **10**, 1–10 (2019).
66. Silva, S. R. *et al.* Chloroquine resistance is associated to multi-copy pvcrt-o gene in Plasmodium vivax malaria in the Brazilian Amazon. *Malar. J.* **17**, 267 (2018).
67. Pava, Z. *et al.* Expression of Plasmodium vivax crt-o Is Related to Parasite Stage but Not Ex Vivo Chloroquine Susceptibility. *Antimicrob. Agents Chemother.* **60**, 361–367 (2015).
68. Asih, P. B. S. *et al.* Distribution of Plasmodium vivax pvdhfr and pvdhps alleles and their association with sulfadoxine–pyrimethamine treatment outcomes in Indonesia. *Malar. J.* **14**, 365 (2015).
69. Auliff *et al.* Amino acid mutations in Plasmodium vivax DHFR and DHPS from several geographical regions and susceptibility to antifolate drugs. *Am. J. Trop. Med. Hyg.* **75**, 617–621 (2006).
70. Auliff, Adams, J. H., O'Neil, M. T. & Cheng, Q. Defining the Role of Mutations in Plasmodium vivax Dihydrofolate Reductase-Thymidylate Synthase Gene Using an Episomal Plasmodium falciparum Transfection System. *Antimicrob. Agents Chemother.* **54**, 3927–3932 (2010).
71. Barnadas *et al.* Plasmodium vivax dhfr and dhps mutations in isolates from Madagascar and therapeutic response to sulphadoxine-pyrimethamine. *Malar. J.* **7**, 35 (2008).
72. WHO. Technical Expert Group meeting on intermittent preventive treatment in pregnancy (IPTp). (2007).
73. Hawkins, V. N., Joshi, H., Rungsihirunrat, K., Na-Bangchang, K. & Sibley, C. H. Antifolates can have a role in the treatment of Plasmodium vivax. *Trends Parasitol* **23**, 213–222 (2007).
74. Maguire, J. D. *et al.* Mefloquine is highly efficacious against chloroquine-resistant Plasmodium vivax malaria and Plasmodium falciparum malaria in Papua, Indonesia. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **42**, 1067–1072 (2006).
75. Pukrittayakamee, S., Vanijanonta, S., Chantra, A., Clemens, R. & White, N. J. Blood stage antimalarial efficacy of primaquine in Plasmodium vivax malaria. *J. Infect. Dis.* **169**, 932–935 (1994).

76. Tjitra, E., Baker, J., Suprianto, S., Cheng, Q. & Anstey, N. M. Therapeutic efficacies of artesunate-sulfadoxine-pyrimethamine and chloroquine-sulfadoxine-pyrimethamine in vivax malaria pilot studies: relationship to Plasmodium vivax dhfr mutations. *Antimicrob. Agents Chemother.* **46**, 3947–3953 (2002).
77. Hawkins, V. N., Joshi, H., Rungsihirunrat, K., Na-Bangchang, K. & Sibley, C. H. Antifolates can have a role in the treatment of Plasmodium vivax. *Trends Parasitol.* **23**, 213–222 (2007).
78. Auburn, S. *et al.* Genomic Analysis of Plasmodium vivax in Southern Ethiopia Reveals Selective Pressures in Multiple Parasite Mechanisms. *J. Infect. Dis.* **220**, 1738–1749 (2019).
79. Imwong *et al.* Association of Genetic Mutations in Plasmodium vivax dhfr with Resistance to Sulfadoxine-Pyrimethamine: Geographical and Clinical Correlates. *ANTIMICROB AGENTS CHEMOTHER* **11**, 7 (2001).
80. Marfurt, J. *et al.* Molecular markers of in vivo Plasmodium vivax resistance to amodiaquine plus sulfadoxine-pyrimethamine: mutations in pvdhfr and *pvmdr1*. *J. Infect. Dis.* **198**, 409–417 (2008).
81. Briolant, S. *et al.* The F423Y Mutation in the pfmdr2 Gene and Mutations N51I, C59R, and S108N in the pfdhfr Gene Are Independently Associated with Pyrimethamine Resistance in Plasmodium falciparum Isolates. *Antimicrob. Agents Chemother.* **56**, 2750–2752 (2012).
82. Hastings, M. D. *et al.* Dihydrofolate reductase mutations in Plasmodium vivax from Indonesia and therapeutic response to sulfadoxine plus pyrimethamine. *J. Infect. Dis.* **189**, 744–750 (2004).
83. Kongsaeree, P. *et al.* Crystal structure of dihydrofolate reductase from Plasmodium vivax: pyrimethamine displacement linked with mutation-induced resistance. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 13046–13051 (2005).
84. Imwong *et al.* Limited Polymorphism in the Dihydropteroate Synthetase Gene (dhps) of Plasmodium vivax Isolates from Thailand. *Antimicrob. Agents Chemother.* **49**, 4393–4395 (2005).
85. Korsinczky, M. *et al.* Sulfadoxine Resistance in Plasmodium vivax Is Associated with a Specific Amino Acid in Dihydropteroate Synthase at the Putative Sulfadoxine-Binding Site. *Antimicrob. Agents Chemother.* **48**, 2214–2222 (2004).
86. Yogavel, M. *et al.* Structure of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase–dihydropteroate synthase from Plasmodium vivax sheds light on drug resistance. *J. Biol. Chem.* **293**, 14962–14972 (2018).
87. Alam, M. T. *et al.* Similar trends of pyrimethamine resistance-associated mutations in Plasmodium vivax and P. falciparum. *Antimicrob. Agents Chemother.* **51**, 857–863 (2007).
88. Auburn, S. *et al.* Genomic analysis of a pre-elimination Malaysian Plasmodium vivax population reveals selective pressures and changing transmission dynamics. *Nat. Commun.* **9**, 2585 (2018).
89. Hupalo, D. N. *et al.* Population genomics studies identify signatures of global dispersal and drug resistance in Plasmodium vivax. *Nat. Genet.* **48**, 953–958 (2016).
90. Kaur, S. *et al.* Plasmodium vivax dihydrofolate reductase point mutations from the Indian subcontinent. *Acta Trop.* **97**, 174–180 (2006).
91. Brazeau, N. F. *et al.* Longitudinal Pooled Deep Sequencing of the Plasmodium vivax K12 Kelch Gene in Cambodia Reveals a Lack of Selection by Artemisinin. *Am. J. Trop. Med. Hyg.* **95**, 1409–1412 (2016).
92. Dharia, N. V. *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.* **107**, 20045–20050 (2010).

93. Flannery, E. L. *et al.* Next-Generation Sequencing of Plasmodium vivax Patient Samples Shows Evidence of Direct Evolution in Drug-Resistance Genes. *ACS Infect. Dis.* **1**, 367–379 (2015).
94. Parobek, C. M. *et al.* Selective sweep suggests transcriptional regulation may underlie Plasmodium vivax resilience to malaria control measures in Cambodia. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E8096–E8105 (2016).
95. Brashear, A. M. *et al.* Population genomics identifies a distinct Plasmodium vivax population on the China-Myanmar border of Southeast Asia. *PLoS Negl. Trop. Dis.* **14**, e0008506 (2020).
96. Cowell, A. N., Valdivia, H. O., Bishop, D. K. & Winzeler, E. A. Exploration of Plasmodium vivax transmission dynamics and recurrent infections in the Peruvian Amazon using whole genome sequencing. *Genome Med.* **10**, 52 (2018).
97. Mok, S. *et al.* Structural polymorphism in the promoter of pfmrp2 confers Plasmodium falciparum tolerance to quinoline drugs. *Mol. Microbiol.* **91**, 918–934 (2014).
98. Nogueira, F., Lopes, D., Alves, A. C. & Rosario, V. E. do. Plasmodium falciparum multidrug resistance protein (MRP) gene expression under chloroquine and mefloquine challenge. *J. Cell Anim. Biol.* **2**, 010–020 (2008).
99. Veiga, M. I. *et al.* Complex Polymorphisms in the Plasmodium falciparum Multidrug Resistance Protein 2 Gene and Its Contribution to Antimalarial Response. *Antimicrob. Agents Chemother.* **58**, 7390–7397 (2014).
100. Deng, S. *et al.* Genetic diversity of the Pvk12 gene in Plasmodium vivax from the China-Myanmar border area. *Malar. J.* **15**, 528 (2016).
101. Gresty, K., Anderson, K., Pasay, C., Waters, N. C. & Cheng, Q. Polymorphisms in Plasmodium falciparum Kelch 13 and *P. vivax* Kelch 12 Genes in Parasites Collected from Three South Pacific Countries Prior to Extensive Exposure to Artemisinin Combination Therapies. *Antimicrob. Agents Chemother.* **63**, (2019).
102. Bopp, S. *et al.* Plasmepsin II–III copy number accounts for bimodal piperaquine resistance among Cambodian Plasmodium falciparum. *Nat. Commun.* **9**, 1769 (2018).
103. Müller, S. Role and Regulation of Glutathione Metabolism in Plasmodium falciparum. *Molecules* **20**, 10511–10534 (2015).
104. Raj, D. K. *et al.* Disruption of a Plasmodium falciparum multidrug resistance-associated protein (PfMRP) alters its fitness and transport of antimalarial drugs and glutathione. *J. Biol. Chem.* **284**, 7687–7696 (2009).
105. Dahlström, S. *et al.* Plasmodium falciparum multidrug resistance protein 1 and artemisinin-based combination therapy in Africa. *J. Infect. Dis.* **200**, 1456–1464 (2009).
106. Bright *et al.* Genetic Analysis of Primaquine Tolerance in a Patient with Relapsing Vivax Malaria. *Emerg. Infect. Dis.* **19**, 802–805 (2013).
107. Bright *et al.* Genetic Analysis of Primaquine Tolerance in a Patient with Relapsing Vivax Malaria. *Emerg Infect Dis* **19**, 802–805 (2013).
108. Chiang, T.-Y., Lin, W.-C., Kuo, M.-C., Ji, D.-D. & Fang, C.-T. Relapse of imported vivax malaria despite standard-dose primaquine therapy: an investigation with molecular genotyping analyses. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* **18**, E232-234 (2012).
109. Hill, D. R. *et al.* Primaquine: report from CDC expert meeting on malaria chemoprophylaxis I. *Am. J. Trop. Med. Hyg.* **75**, 402–415 (2006).
110. Llanos-Cuentas, A. *et al.* Tafenoquine versus Primaquine to Prevent Relapse of Plasmodium vivax Malaria. *N. Engl. J. Med.* **380**, 229–241 (2019).
111. Taylor, W. R. J. *et al.* Short-course primaquine for the radical cure of Plasmodium vivax malaria: a multicentre, randomised, placebo-controlled non-inferiority trial. *Lancet Lond. Engl.* **394**, 929–938 (2019).

112. Camarda, G. *et al.* Antimalarial activity of primaquine operates via a two-step biochemical relay. *Nat. Commun.* **10**, 3226 (2019).
113. Townell, N., Looke, D., McDougall, D. & McCarthy, J. S. Relapse of imported Plasmodium vivax malaria is related to primaquine dose: a retrospective study. *Malar. J.* **11**, 214 (2012).
114. Bright *et al.* A High Resolution Case Study of a Patient with Recurrent Plasmodium vivax Infections Shows That Relapses Were Caused by Meiotic Siblings. *PLoS Negl. Trop. Dis.* **8**, (2014).
115. Baird, J. K. *et al.* In vivo resistance to chloroquine by Plasmodium vivax and Plasmodium falciparum at Nabire, Irian Jaya, Indonesia. *Am J Trop Med Hyg* **56**, 627–631 (1997).
116. Añez, A. *et al.* [Therapeutic response of Plasmodium vivax to chloroquine in Bolivia]. *Biomedica* **32**, 527–535 (2012).
117. Añez, A. *et al.* Resistance of infection by Plasmodium vivax to chloroquine in Bolivia. *Malar J* **14**, 261 (2015).
118. Baird *et al.* Treatment of chloroquine-resistant Plasmodium vivax with chloroquine and primaquine or halofantrine. *J. Infect. Dis.* **171**, 1678–1682 (1995).
119. Congpuon, K. *et al.* In vivo sensitivity monitoring of chloroquine for the treatment of uncomplicated vivax malaria in four bordered provinces of Thailand during 2009-2010. *J Vector Borne Dis* **48**, 190–196 (2011).
120. Getachew, S. *et al.* Chloroquine efficacy for Plasmodium vivax malaria treatment in southern Ethiopia. *Malar J* **14**, 525 (2015).
121. Guthmann, J.-P. *et al.* Plasmodium vivax resistance to chloroquine in Dawei, southern Myanmar. *Trop Med Int Health* **13**, 91–98 (2008).
122. Hwang, J. *et al.* In vivo efficacy of artemether-lumefantrine and chloroquine against Plasmodium vivax: a randomized open label trial in central Ethiopia. *PLoS One* **8**, e63433 (2013).
123. Phong, N. C. *et al.* Susceptibility of Plasmodium falciparum to artemisinins and Plasmodium vivax to chloroquine in Phuoc Chien Commune, Ninh Thuan Province, south-central Vietnam. *Malar J* **18**, 10 (2019).
124. Phyo, A. P. *et al.* Dihydroartemisinin-piperaquine versus chloroquine in the treatment of Plasmodium vivax malaria in Thailand: a randomized controlled trial. *Clin Infect Dis* **53**, 977–984 (2011).
125. Yohannes, A. M., Teklehaimanot, A., Bergqvist, Y. & Ringwald, P. Confirmed vivax resistance to chloroquine and effectiveness of artemether-lumefantrine for the treatment of vivax malaria in Ethiopia. *Am J Trop Med Hyg* **84**, 137–140 (2011).
126. Ferreira, M. U. *et al.* Monitoring Plasmodium vivax resistance to antimalarials: Persisting challenges and future directions. *Int. J. Parasitol. Drugs Drug Resist.* **15**, 9–24 (2021).
127. Añez, A. *et al.* [Therapeutic response of Plasmodium vivax to chloroquine in Bolivia]. *Biomed. Rev. Inst. Nac. Salud* **32**, 527–535 (2012).
128. Congpuon, K. *et al.* In vivo sensitivity monitoring of chloroquine for the treatment of uncomplicated vivax malaria in four bordered provinces of Thailand during 2009-2010. *J. Vector Borne Dis.* **48**, 190–196 (2011).
129. Guthmann, J.-P. *et al.* Plasmodium vivax resistance to chloroquine in Dawei, southern Myanmar. *Trop. Med. Int. Health TM IH* **13**, 91–98 (2008).
130. Hwang, J. *et al.* In vivo efficacy of artemether-lumefantrine and chloroquine against Plasmodium vivax: a randomized open label trial in central Ethiopia. *PloS One* **8**, e63433 (2013).
131. Ladeia-Andrade, S. *et al.* Monitoring the Efficacy of Chloroquine-Primaquine Therapy for Uncomplicated Plasmodium vivax Malaria in the Main Transmission Hot Spot of Brazil. *Antimicrob. Agents Chemother.* **63**, (2019).

132. Yohannes, A. M., Teklehaimanot, A., Bergqvist, Y. & Ringwald, P. Confirmed vivax resistance to chloroquine and effectiveness of artemether-lumefantrine for the treatment of vivax malaria in Ethiopia. *Am. J. Trop. Med. Hyg.* **84**, 137–140 (2011).
133. Fryauff, D. J. *et al.* The drug sensitivity and transmission dynamics of human malaria on Nias Island, North Sumatra, Indonesia. *Ann. Trop. Med. Parasitol.* **96**, 447–462 (2002).
134. Ratcliff, A. *et al.* Two fixed-dose artemisinin combinations for drug-resistant falciparum and vivax malaria in Papua, Indonesia: an open-label randomised comparison. *The Lancet* **369**, 757–765 (2007).
135. Sumawinata, I. W. *et al.* Very high risk of therapeutic failure with chloroquine for uncomplicated Plasmodium falciparum and *P. vivax* malaria in Indonesian Papua. *Am. J. Trop. Med. Hyg.* **68**, 416–420 (2003).
136. Sutanto, I. *et al.* Evaluation of chloroquine therapy for vivax and falciparum malaria in southern Sumatra, western Indonesia. *Malar. J.* **9**, 52 (2010).
137. Fukuda, M. M. *et al.* A randomized, double-blind, active-control trial to evaluate the efficacy and safety of a three day course of tafenoquine monotherapy for the treatment of Plasmodium vivax malaria. *PloS One* **12**, e0187376 (2017).
138. Wilairatana, P. *et al.* Efficacy of primaquine regimens for primaquine-resistant Plasmodium vivax malaria in Thailand. *Am. J. Trop. Med. Hyg.* **61**, 973–977 (1999).
139. McIntosh, H. M. & Olliaro, P. Artemisinin derivatives for treating uncomplicated malaria. *Cochrane Database Syst. Rev.* CD000256 (2000) doi:10.1002/14651858.CD000256.
140. Stepniewska, K. *et al.* In vivo parasitological measures of artemisinin susceptibility. *J. Infect. Dis.* **201**, 570–579 (2010).
141. WWARN. Artemether-lumefantrine treatment of uncomplicated Plasmodium falciparum malaria: a systematic review and meta-analysis of day 7 lumefantrine concentrations and therapeutic response using individual patient data. *BMC Med.* **13**, 227 (2015).
142. WWARN *et al.* Clinical determinants of early parasitological response to ACTs in African patients with uncomplicated falciparum malaria: a literature review and meta-analysis of individual patient data. *BMC Med.* **13**, 212 (2015).
143. Imwong *et al.* Relapses of Plasmodium vivax infection usually result from activation of heterologous hypnozoites. *J. Infect. Dis.* **195**, 927–933 (2007).
144. Abdallah, T. M. *et al.* Efficacy of artemether-lumefantrine as a treatment for uncomplicated Plasmodium vivax malaria in eastern Sudan. *Malar. J.* **11**, 404 (2012).
145. Abreha, T. *et al.* Comparison of artemether-lumefantrine and chloroquine with and without primaquine for the treatment of Plasmodium vivax infection in Ethiopia: A randomized controlled trial. *PLoS Med.* **14**, e1002299 (2017).
146. Commons, R. J. *et al.* The efficacy of dihydroartemisinin-piperaquine and artemether-lumefantrine with and without primaquine on Plasmodium vivax recurrence: A systematic review and individual patient data meta-analysis. *PLOS Med.* **16**, e1002928 (2019).
147. Karunajeewa, H. A. *et al.* A trial of combination antimalarial therapies in children from Papua New Guinea. *N. Engl. J. Med.* **359**, 2545–2557 (2008).
148. Senn, N. *et al.* Effectiveness of artemether/lumefantrine for the treatment of uncomplicated Plasmodium vivax and *P. falciparum* malaria in young children in Papua New Guinea. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **56**, 1413–1420 (2013).
149. WHO. WHO | Intermittent preventive treatment in pregnancy (IPTp). *WHO* http://www.who.int/malaria/areas/preventive_therapies/pregnancy/en/ (2019).
150. Rangel, G. W. *et al.* Enhanced Ex Vivo Plasmodium vivax Intraerythrocytic Enrichment and Maturation for Rapid and Sensitive Parasite Growth Assays. *Antimicrob Agents Chemother* **62**, (2018).
151. Russell, B. M. *et al.* Simple in vitro assay for determining the sensitivity of Plasmodium vivax isolates from fresh human blood to antimalarials in areas where *P. vivax* is endemic. *Antimicrob. Agents Chemother.* **47**, 170–173 (2003).

152. Lim, C. *et al.* Expansion of host cellular niche can drive adaptation of a zoonotic malaria parasite to humans. *Nat. Commun.* **4**, 1638 (2013).
153. Moon, R. W. *et al.* Adaptation of the genetically tractable malaria pathogen Plasmodium knowlesi to continuous culture in human erythrocytes. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 531–536 (2013).
154. Mohring, F. *et al.* Rapid and iterative genome editing in the malaria parasite Plasmodium knowlesi provides new tools for *P. vivax* research. *eLife* **8**, (2019).
155. Chua, A. C. Y. *et al.* Robust continuous in vitro culture of the Plasmodium cynomolgi erythrocytic stages. *Nat. Commun.* **10**, 3635 (2019).
156. Voorberg-van der Wel, A. M. *et al.* Dual-Luciferase-Based Fast and Sensitive Detection of Malaria Hypnozoites for the Discovery of Antirelapse Compounds. *Anal. Chem.* **92**, 6667–6675 (2020).
157. Voorberg-van der Wel, A. *et al.* Transgenic fluorescent Plasmodium cynomolgi liver stages enable live imaging and purification of Malaria hypnozoite-forms. *PloS One* **8**, e54888 (2013).
158. Cowell, A. N. *et al.* Selective Whole-Genome Amplification Is a Robust Method That Enables Scalable Whole-Genome Sequencing of Plasmodium vivax from Unprocessed Clinical Samples. *mBio* **8**, (2017).
159. Melnikov, A. *et al.* Hybrid selection for sequencing pathogen genomes from clinical samples. *Genome Biol.* **12**, R73 (2011).
160. Pearson, R. D. *et al.* Genomic analysis of local variation and recent evolution in Plasmodium vivax. *Nat. Genet.* **48**, 959–964 (2016).
161. Neafsey, D. E., Taylor, A. R. & MacInnis, B. L. Advances and opportunities in malaria population genomics. *Nat. Rev. Genet.* **22**, 502–517 (2021).
162. Bosco, A. B. *et al.* Molecular surveillance reveals the presence of pfhrp2 and pfhrp3 gene deletions in Plasmodium falciparum parasite populations in Uganda, 2017–2019. *Malar. J.* **19**, 300 (2020).
163. Chen, C. Development of antimalarial drugs and their application in China: a historical review. *Infect. Dis. Poverty* **3**, 9 (2014).
164. Kümpornsin, K., Kochakarn, T. & Chookajorn, T. The resistome and genomic reconnaissance in the age of malaria elimination. *Dis. Model. Mech.* **12**, dmm040717 (2019).
165. Gruenberg, M., Lerch, A., Beck, H.-P. & Felger, I. Amplicon deep sequencing improves Plasmodium falciparum genotyping in clinical trials of antimalarial drugs. *Sci. Rep.* **9**, 17790 (2019).
166. Rao, P. N. *et al.* A Method for Amplicon Deep Sequencing of Drug Resistance Genes in Plasmodium falciparum Clinical Isolates from India. *J. Clin. Microbiol.* **54**, 1500–1511 (2016).
167. Lin, J. T. *et al.* Using Amplicon Deep Sequencing to Detect Genetic Signatures of Plasmodium vivax Relapse. *J. Infect. Dis.* **212**, 999–1008 (2015).
168. Boyce, R. M. *et al.* Reuse of malaria rapid diagnostic tests for amplicon deep sequencing to estimate Plasmodium falciparum transmission intensity in western Uganda. *Sci. Rep.* **8**, 10159 (2018).
169. Talundzic, E. *et al.* Molecular Epidemiology of Plasmodium falciparum kelch13 Mutations in Senegal Determined by Using Targeted Amplicon Deep Sequencing. *Antimicrob. Agents Chemother.* **61**, (2017).
170. Kim, A., Popovici, J., Menard, D. & Serre, D. Plasmodium vivax transcriptomes reveal stage-specific chloroquine response and differential regulation of male and female gametocytes. *Nat. Commun.* **10**, (2019).

171. Rangel, G. W. *et al.* Plasmodium vivax transcriptional profiling of low input cryopreserved isolates through the intraerythrocytic development cycle. *PLoS Negl. Trop. Dis.* **14**, e0008104 (2020).
172. Sà, J. M., Cannon, M. V., Caleon, R. L., Wellems, T. E. & Serre, D. Single-cell transcription analysis of Plasmodium vivax blood-stage parasites identifies stage- and species-specific profiles of expression. *PLOS Biol.* **18**, e3000711 (2020).

Chapter Two: Population Genomics of *Plasmodium vivax* in Panama to Assess the Risk of Case Importation on Malaria Elimination

2.1 Abstract

Malaria incidence in Panama has plateaued in recent years, despite elimination efforts, with almost all cases caused by *Plasmodium vivax*. Notwithstanding, malaria prevalence remains low (fewer than 1 case per 1000 persons). We used selective whole genome amplification to sequence 59 *P. vivax* samples from Panama. The *P. vivax* samples were collected from two periods (2007-2009 and 2017-2019) to study the population structure and transmission dynamics of the parasite. Imported cases resulting from increased human migration could threaten malaria elimination prospects, and four of the samples evaluated came from individuals with travel history. We explored patterns of recent common ancestry among the samples and observed that a highly genetically related lineage (termed CL1) was dominant among the samples (47 out of 59 samples with good sequencing coverage), spanning the entire period of the collection (2007-2019) and all regions of the country. We also found a second, smaller clonal lineage (termed CL2) of four parasites collected between 2017 and 2019. To explore the regional context of Panamanian *P. vivax* we conducted principal components analysis and constructed a neighbor-joining tree using these samples and samples collected worldwide from a previous study. Three of the four samples with travel history clustered with samples collected from their suspected country of origin (consistent with importation), while one appears to have been a result of local transmission. The small number of Panamanian *P. vivax* samples not belonging to either CL1 or CL2 clustered with samples collected from Colombia, suggesting they represent the genetically similar ancestral *P. vivax* population in Panama or were recently imported from Colombia. The low diversity we observe in Panama indicates that this parasite population has previously been subject to a severe bottleneck and may be eligible for elimination. Additionally, while we confirmed that *P. vivax* is imported to Panama from diverse geographic locations, the lack of impact from imported cases on the parasite population genomic profile suggests that onward

transmission from such cases is limited and that imported cases may not pose a major barrier to elimination.

2.2 Introduction

Malaria is a parasitic disease transmitted by the bite of female *Anopheles* mosquitoes. Malaria parasites cause approximately 219 million cases and 435,000 deaths each year, the vast majority in sub-Saharan Africa. *Plasmodium falciparum,* the most virulent of the six *Plasmodium* species that infect humans (*P. falciparum*, *P. vivax, P. malariae, P. ovale wallikeri, P. ovale curtisi,* and *P. knowlesi*), causes most of these cases (1). Though billions of dollars have been devoted to the control and eradication of malaria caused by *P. falciparum,* comparatively little attention has been given to *P. vivax,* the most prevalent malaria parasite outside Africa (2). The impact of *P. vivax* on human health was once considered minimal, relative to the more virulent *P. falciparum* (1,2). However, recent studies suggest *P. vivax* causes significant global health burden (1,2). The *P. vivax* life cycle includes a dormant liver "hypnozoite" stage. (2) The hypnozoite stage can cause a relapse of malaria weeks to months after the initial infection, thus beginning the cycle of infection and complicating control efforts (2).

Sixty percent of the Central and South American population lives in areas with ongoing malaria transmission, predominantly caused by *P. vivax* (3). The region experiences about 700,000 *P. vivax* cases each year (1). Between 2000 and 2015 the incidence rate of malaria fell 37% globally and 42% in Africa (1). In the Americas, malaria mortality decreased by 72% during the same period (1). Unfortunately, recent evidence suggests that this trend has stalled, and in some countries malaria incidence has even increased (1). Panama eliminated the autochthonous transmission of *P. falciparum* in 2010*,* outside of a small outbreak on the Colombian border in 2015 (4). Since 2010, *P. vivax* has caused almost all malaria cases in

Panama (1,5,6). *P. vivax* cases in Panama have declined precipitously since 2005, from close to 1 case per 1000 persons, to under 0.25 cases per 1000 persons in 2017 (7). However, malaria incidence in Panama has plateaued since 2008. This plateau in incidence could be due to low levels of transmission and/or imported cases that are re-seeding infections.

Human movement leading to parasite migration is a potentially significant epidemiological threat to malaria control in Panama. Parasite importation stemming from human migration is a challenge to elimination programs in other countries around the world (8–10). The unique geographic position of Panama makes it a crossroads for human migration to the United States from South America (5,11). Migrants enter Panama through two paths: through the Darien jungle region on the border with Colombia, and through the Kuna Yala Amerindian reserve ('Comarca') on the Caribbean coast (5,6,11). Previous studies implicate these regions as focal points of ongoing malaria transmission in Panama and suggest this is partly due to imported parasites (5,11). It is estimated that approximately 60,000 continental and extra-continental migrants crossed the southern border of Panama through the Darien jungle region in 2015 and 2016 (12).

To inform effective malaria elimination strategies in Panama, it is critical to understand how the parasite moves throughout the country, uncover pockets of focalized transmission, and differentiate between sustained local infection and case importation as the reason for disease persistence. (4,5). Whole-genome sequencing can help paint a detailed picture of parasite movement and transmission within and between countries (10,12,13). However, *P. vivax* cannot be grown *in vitro*, and the difficulty of sequencing *P. vivax* from clinical samples dominated by host DNA has hindered parasite population studies (14). Recent advances such as hybrid capture (15) and selective whole genome amplification (SWGA) mitigate this problem by allowing for

parasite DNA to be selectively enriched before sequencing (14). Both methods have allowed for population genomic studies of *P. vivax* using samples directly from patients.

In this study, we describe the population genomics of *P. vivax* in Panama over a 12-year time span, with the aim of understanding patterns of genetic variation and recent shared ancestry (relatedness) at different geographical and temporal scales. We found most *P. vivax* cases in Panama belong to a single highly related lineage that has persisted for at least a decade. Furthermore, we observed a second smaller clonal lineage concentrated near the Panamanian-Colombian border. We also found several samples that shared no relatedness with any other sample, which may represent either localized pockets of outbred *P. vivax* transmission or imported cases. Revealing these patterns of relatedness among parasite infections can help inform best strategies for targeting interventions or case investigation methods to increase the likelihood of successful elimination. We discuss these findings and their implications for ongoing elimination efforts of *P. vivax* in Panama. The results obtained from this study will help inform future elimination efforts in Panama and the rest of Meso-America.

2.3 Materials and Methods

Ethics Statement

The Research Bioethics Committee (CBI) of the Gorgas Memorial Institute of Health Studies gave the study ethical approval (Permit: 154/CBI/ICGES/17). Written consent was obtained from infected patients prior to collecting samples.

Sample Collection

We collected 96 *P. vivax* samples from infected consenting volunteers identified through passive or active surveillance by technicians from the Department of Vector Control, Ministry of

Health (MINSA) of Panama. Two groups of DNA samples from infected patients were used in this study: 1) 56 DNA samples collected during 2007-2009 and 2) 40 DNA samples collected during 2017-2019. The 2007-2009 samples were collected as part of an earlier study exploring the genetic diversity of *P. falciparum* and *P. vivax* in Panama (Approved by The National Committee for Research Bioethics of Panama (CNBI): Permit: 468/CNBI/ICGES/06, PI: José E. Calzada). The Gorgas Memorial Institutional Animal Use and Care Committee (CIUCAL) (Permit: 002/CIUCAL-ICCES-2012) approved the use of Aotus *P. vivax* AMRU-1 and SAL-1 infected blood samples as a source of control DNA. Patient blood samples were collected by finger-prick with a lancet and spotted into EBF 903 Five Spot Blood Cards (Eastern Business Forms, INC, SC, USA). The samples were then transported at ambient temperature to the laboratory and stored at –20℃ until processing. Thin and thick blood smears were obtained from patient samples. The blood smears were stained with Giemsa for percent parasite density determination, species identification, and stage differential counts. Each volunteer donated ~150 µL of blood.

Information Survey

We collected demographic, geographic, socioeconomic, and epidemiological information from each study subject using an epidemiological form developed for the Survey123 for ArcGIS online survey program (Esri, Redlands, CA), as allowed under ethical approval.

Malaria Microscopy

Giemsa stained thick and thin blood smears were examined by light microscopy for parasite density determinations, P*lasmodium* species confirmation, and parasite lifecycle stage count. Parasite densities were calculated by quantifying the number of malaria-infected red

blood cells (iRBCs) among 500 – 2000 RBCs on a thin blood smear and expressing the result as % parasitemia (% parasitemia = parasitized RBCs /total RBCs) x 100).

DNA Extraction

We extracted DNA from the filter paper blood spots using the Chelex method as described (16) for samples obtained during 2007-2009 and with the Qiagen DNA mini kit for samples obtained during 2017-2019.

Molecular Confirmation of *P. vivax* Infection

We confirmed *P. vivax* infection for all samples collected during 2017-2019 by amplification of the *P. vivax* PVX_18SrRNA gene using a qRT-PCR assay as described (17).

Selective Whole Genome Amplification and Sequencing

We carried out DNA pre-amplification as described (14). Briefly, the thermocycler was preheated to 35℃. We dispensed aliquots of 37µl of Power SYBR Green Master Mix, plus 3µl phi39 into each PCR tube, next adding DNA, and water to achieve a final volume of 47µl. Thermocycler settings were as follows: 35℃ x 10 min; 34℃ x 10 min; 33℃ x 10 min; 32℃ x 10 min; 31℃ x 10 min; 30℃ x 16 hours; 65℃ x 10 min; and 4 ℃ for infinity. SWGA reaction products were diluted with 50 µl of water. We purified 50 µl of the diluted product using 50 µl AmPURE beads according to the instructions of the manufacturer. We then eluted beads in 30 µl of water. Approximately 60-120 ng/µl of DNA was obtained after bead purification of the SWGA reaction. We measured DNA concentration using Nanodrop quantitation.

Whole-Genome Sequencing

We performed whole-genome sequencing (WGS) on all 96 *P. vivax* samples using Nextera libraries and an Illumina HiSeq X platform. Sample reads were aligned to the P01

reference genome assembly using BWA-MEM, version 0.7 (18). Illumina sequencing reads are available through the NCBI Sequence Read Archive with BioProject accession numbers SAMN15722613–SAMN15722671.

SNP Discovery and Quality Filtering

We marked duplicate reads using the MarkDuplicates tool from Picard tools. We next performed local realignment around indels using the Genome Analysis Toolkit (GATK) RealignerTargetCreator and GATK IndelRealigner (GATK Version 3.5.0). We called variants using GATK HaplotypeCaller using best practices to call and filter single nucleotide polymporphisms (SNPs) and generate individual variant call files (gVCFs) s for each sample. We called variants in two batches, one containing samples collected in 2007-2009 and one containing samples collected in 2017-2019. We performed joint variant calling on the sets separately using GATK GenotypeGVCFs tool with GATK hard filters, including calls in subtelomeric regions. The resulting datasets consisted of 56 samples and 407,554 sites for the 2007-2009 samples, and 40 samples and 171,433 variants for the 2017-2019 samples. We retained samples for analysis if they exhibited a minimum mean read depth of five and had calls at more than 80% of variant sites in the dataset corresponding to their collection period, including those in subtelomeric regions. We calculated and evaluated data quality measures using the VCFtools package and custom R scripts (19). Thirty-five samples from 2007-2009 and 24 samples from 2017-2019 passed these filters and were kept for further analysis.

We next used GenotypeGVCFs tool to construct a joint dataset with the 59 Panamanian samples plus a collection of previously collected global samples (Bioproject numbers PRJNA240356-PRJNA240533 (20)). The joint dataset contained 168 samples and 2,4250,245 variants. We filtered sites based on quality (GQ > 40), passing VQSR truth sensitivity level of

0.99 or greater, missing rate (having a call at that site in > 85% of samples). We also excluded any sites that were not bi-allelic and indels. The joint dataset generated after filtering contained 168 samples and 62,211 sites.

Lastly, we generated a dataset containing SNPs found jointly in 80% of both the 2007-2009 and 2017-2019 samples. We also filtered sites in this dataset by excluding non-biallelic SNPs and based on quality (GQ > 30) and passing GATK filters. The resultant dataset contained 56 samples and 2,335 SNPs for the 2007-2009 samples, and 40 samples and 1,301 SNPs for the 2017-2019 samples. For these samples, we generated a highly filtered variant set containing biallelic SNPs that passed the GATK filters (GQ > 30, truth sensitivity level > 0.99, Mean DP > five) and were called in at least 80% of the samples from both time periods (2007-2009; 2017-2019). Calls from the two sample sets were merged to create a unified dataset of 96 samples and 264 genotyped SNPs.

Determination of Sample Clonality

We estimated sample clonality using the F_{ws} statistic. F_{ws} measures the within-sample genetic diversity (measured by heterozygosity H_w) relative to the overall population genetic diversity (H_s) (21). The underlying theory assumes that a monoclonal (single strain) infection has extremely low genetic diversity relative to overall population genetic diversity. By contrast, a polyclonal (multiple strain) infection has high diversity relative to overall population diversity (compared to a monoclonal infection). By estimating the ratio between within-host diversity and population diversity, we can distinguish between monoclonal and polyclonal infections (21). A sample with an F_{ws} statistic of 0.95 or greater (>=0.95) is considered monoclonal. We calculated F_{ws} using the R package moimix (22).

Analysis of Recent Common Ancestry

We used hmmIBD (23) to estimate the proportion of sites identical by descent (IBD) between sample pairs to ascertain recent common ancestry among Panamanian and Colombian samples collected previously from a global *P. vivax* population study (20). We estimated minor allele frequency (MAF) for IBD inference using the genetically distinct Panamanian samples, a representative sample from each of the two highly related Panamanian clusters, and the Colombian samples. We included Colombian samples to improve MAF estimation given the greater genetic diversity of the Colombian parasite population and presumed historical gene flow with Panama. We subsetted the master dataset file to keep only samples collected in Panama and Colombia. Sites were excluded sites based on minimum and maximum read depth (five and thirty respectively) to ensure that we were using only high-quality SNPs. The input dataset for hmmIBD contained 89 samples (59 Panamanian samples and 30 Colombian samples) and 15,788 variant sites. We then re-formatted the data using a custom perl script for input into hmmIBD along with the MAF estimates. We conducted analysis and visualization of the hmmIBD output using custom R scripts.

Analysis of Population Structure

We employed principal components analysis (PCA) and a neighbor-joining tree to study the population structure of Panamanian samples in the context of the worldwide *P. vivax* population (20). We used a strictly filtered SNP set for PCA, keeping only variants with calls in at least 95% of samples. This input dataset consisted of 168 samples and 2,428 variants. We used the R package SNPRelate to conduct PCA (24). Covariation within the two clusters heavily influenced the PCA of all samples, so we also performed PCA using a single consensus sequence for each cluster.

We used the R packages ape, StAMPP, pegas, and adegenet, (25–28) to generate the neighbor-joining tree and genetic distance statistics. First, we calculated Nei's distance for all pairwise sample combinations using the master dataset consisting of 168 samples and 62,211 sites to generate a distance matrix. The distance matrix was used to generate a tree. We used the bootphylo function in the ape package to bootstrap the dataset 100 times to estimate nodal support. We then visualized the final tree with support values using the FigTree program (29). We used R software (R version 3.6.1) to carry out statistical analysis and data visualization.

2.4 Results

Recent Common Ancestry Analysis Reveals Single Highly Related Lineage of Parasites

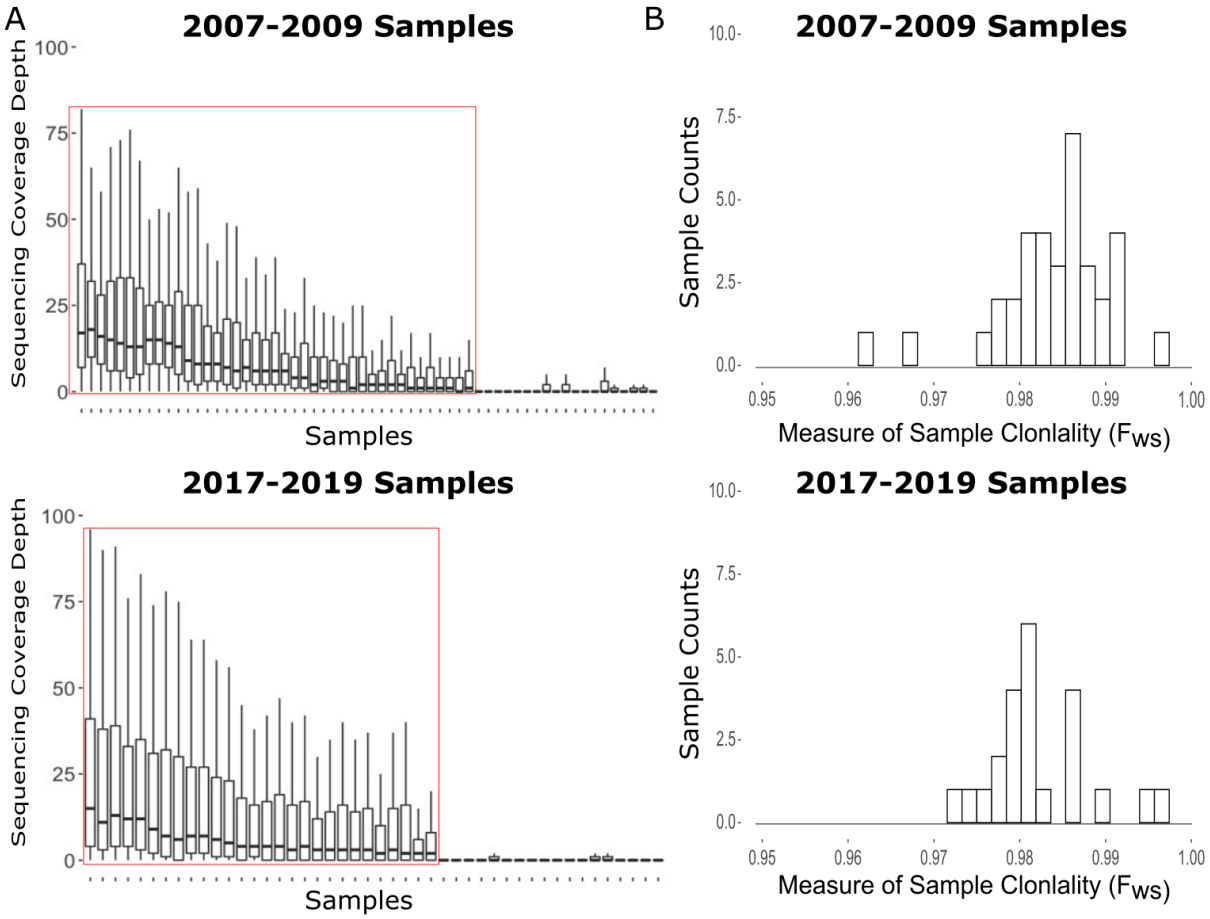


Fig 2.1: Sequencing and Sample Assessment at Variant Sites. A) Distribution of variant site read coverage for each sample stratified by the collection period. Coverage values > 100 were censored for visualization purposes. Samples within the red boxes were kept for analysis. B) Distribution of F_{ws} values for all samples, stratified by the collection period. We interpreted F_{ws} values > 0.95 as evidence of sample monoclonality.

We successfully generated usable sequencing data from 35/56 (58%) Panamanian *P. vivax* samples collected between 2007-2009 and 24/40 (60%) collected between 2017-2019, for a total of 59 samples **(Fig 2.1A).** All Panamanian samples had an F_{ws} statistic greater than 0.95, indicating that they were all monoclonal **(Fig 2.1B).**

We next analyzed the 59 Panamanian genomes in the context of 109 previously published *P. vivax* genomes, generating a filtered dataset consisting of 168 samples and 62,211 high-quality biallelic SNPs.

We used hmmIBD (23) to estimate the proportion of the genome that is identical by descent (IBD) among Panamanian sample pairs to understand patterns of recent common ancestry. IBD measures the proportion of the genome between two individuals that was inherited from a recent common ancestor. Pairwise IBD values closer to 100% indicate very recent common ancestry. We subsetted the dataset to contain only samples collected in Panama and Colombia to estimate pairwise IBD. We strictly filtered sites based on minimum and maximum read depth (five and thirty respectively), resulting in a dataset with 89 samples and 15,788 sites for input into hmmIBD.

We observed a bimodal distribution of pairwise IBD in Panamanian samples, with peaks near zero and 0.95 **(Fig 2.2A).** Forty-seven of the 59 Panamanian samples shared high IBD (>0.875) with each other, indicating very recent common ancestry. Four other Panamanian samples, all collected in the Kuna Yala Province, shared 100% IBD with each other, and 0-10% IBD with any other sample, Panamanian or Colombian **(Figs 2.2A and 2.2B.)**. Another four Panamanian samples exhibited no IBD with each other nor any of the other Panamanian samples. All four of these samples were collected in the Darien jungle region or Kuna Yala, which are the two main points of entry for migrants traveling through Panama. These four samples drive the modal peak of pairwise IBD at zero.

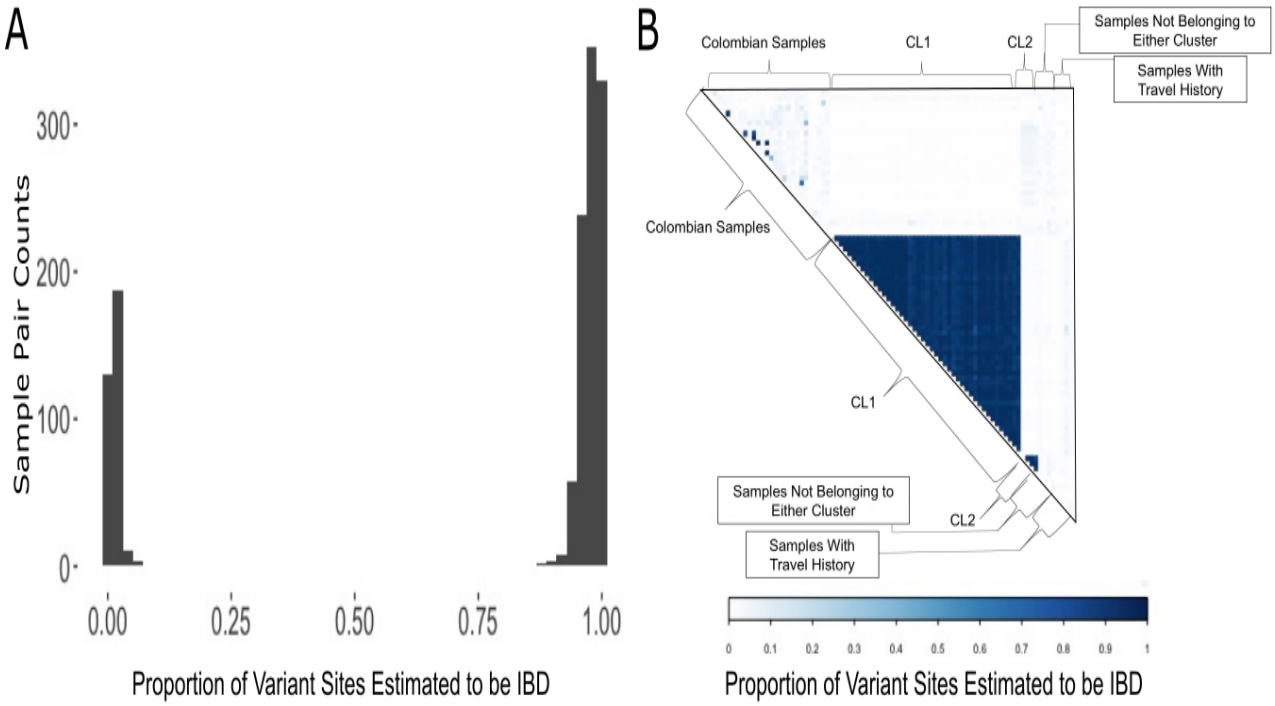


Fig 2.2: IBD analysis of the Panamanian samples. A) The distribution of pairwise IBD estimates among the Panamanian samples. IBD values near zero indicate no recent common ancestry. Values closer to one indicate that the sample pair are clonal or essentially clonal. B) Depicts heatmap of pairwise IBD values for Panamanian and Colombian samples.

The variable degree of relatedness among the 47 samples sharing > 0.85 IBD suggested that data quality potentially impacted the estimation of IBD. We plotted the relationship between IBD and sample quality, measured by the average proportion of high coverage sites in each sample pair, to determine if pairwise sample data quality affected the estimation of IBD **(Supplemental Figure 2.1)**. We defined high coverage sites as sites with greater than 5x coverage (the cutoff for site filtering). We observed that as the average proportion of high coverage sites for sample pairs increased, pairwise IBD estimates correspondingly increased as well **(Supplemental Figure 2.1).** This relationship suggests that poor data quality can lead to underestimation of IBD. It is possible that the majority of the pairwise IBD estimations would be closer to one had the overall sample sequence quality been higher. The prevalent highly

genetically related lineage is referred to henceforth as cluster one (CL1). CL1 samples share an IBD fraction of at least 0.875 with other samples in this cluster. We also concluded that the four samples that shared 100% IBD with each other constituted a second completely clonal lineage, henceforth referred to as cluster two (CL2).

Next, we examined how these two clusters and the other Panamanian samples not belonging to either lineage were geographically distributed in Panama **(Fig. 2.3).** Samples from CL1 were found across Panama. Notably, samples collected from both 2007-2009 and 2017-2019 were found in this lineage. The inclusion of samples from both collection periods demonstrates that this lineage has persisted throughout Panama for at least a decade. We did not find any evidence of structure in the *P. vivax* population by region or relative to the Panama Canal, as was previously observed for *P. falciparum* (4).

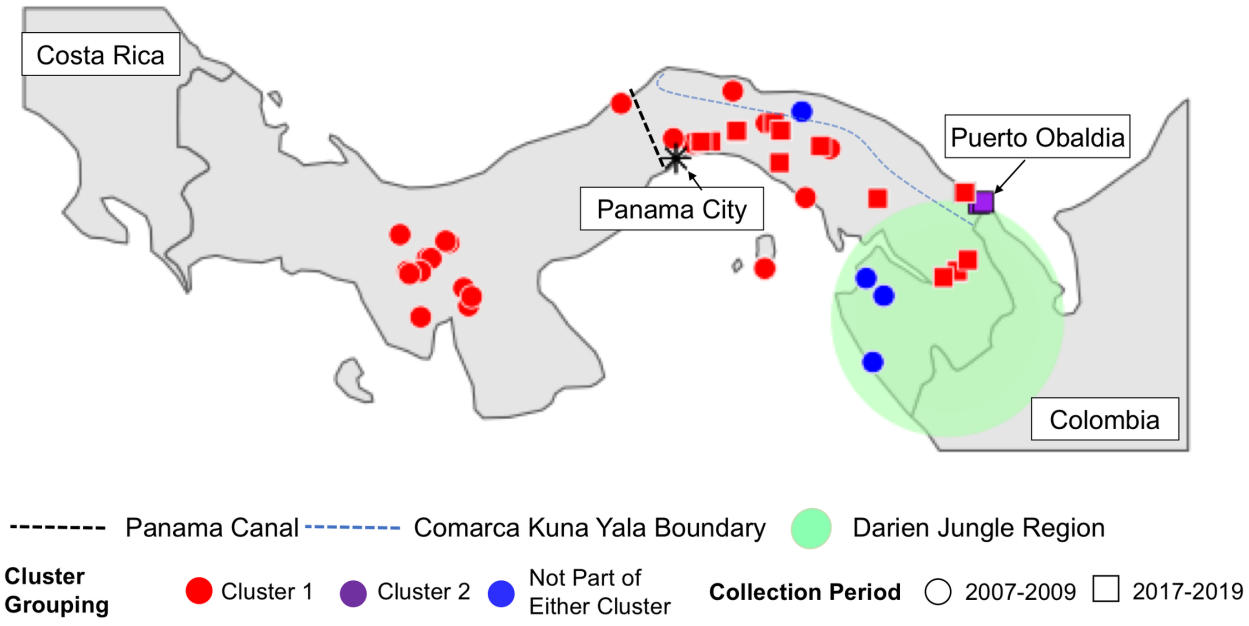


Fig 2.3: Map of Panamanian sample collection sites. Sample colors show which cluster (or neither) each sample belongs too. Shape indicates the sample collection period. The dotted line shows the location of the Panama Canal. The Blue Line shows the border of the Comarca Kuna Yala. The Darien Jungle Region is indicated by the green shaded area.

We only observed samples belonging to CL2 in a specific locality, Puerto Obaldia, in the Kuna Yala Amerindian territory (Comarca) along the Atlantic Coast. We lacked geographic information for one of the four samples in CL2. Additionally, of four samples that shared no recent common ancestry with any sample in the dataset, three were collected in Darien province, along the Colombian border and one was collected in Kuna Yala.

After identifying two highly related lineages in Panama, we explored an approach for determining whether the samples excluded from analysis due to low coverage could belong to one of these lineages. We identified a set of 264 genotyped SNPs that were called in at least 80% of samples across both Panama sample collection periods. We then calculated Nei's standard genetic distance on all pairwise sample comparisons. Most excluded samples across both collection periods (17/21 and 4/16 for the 2007-2009 and 2017-2019 collection periods respectively; a total of 21/37 samples) exhibited very low levels (0-1%) of genetic distance with the CL1 samples and higher genetic distance (0.2-0.25) with the CL2 samples **(Supplemental Figure 2.2).** Seven samples in the 2017-2019 collection period had a high proportion of missing calls for these 264 SNPs, making distance measures uninformative. Three excluded samples from 2007-2009 collected in the Darien Jungle Region had relatively high genetic distance from all other samples in the dataset. Two samples from the 2017-2019 collection period exhibit very low (0-0.1) genetic distance with the CL2 samples, and higher (0.15-0.25) genetic distance with the CL1 samples. The previously observed sample clustering patterns did not change when conducting the analysis with the 264 SNPs.

Exploring the Regional Context of Panamanian *P. vivax*

We built a neighbor-joining tree using the Panamanian samples plus previously sequenced samples (20) **(Fig 2.4A)** to understand the Panamanian *P. vivax* population in a global context. As noted in previous studies (20,30), we observed clusters of samples corresponding to different geographic regions, with a large cluster of Central and South American samples. CL1 and CL2 formed distinct clusters within the Central and South American cluster with 100% bootstrap support. CL2 is situated in a cluster containing samples from Colombia, with 100% bootstrap support at deep nodes. While these four samples are clustered together with 100% support and exhibit short branch length, a long branch connects them to the rest of the Colombian cluster. The Panamanian samples that shared little IBD with either cluster also

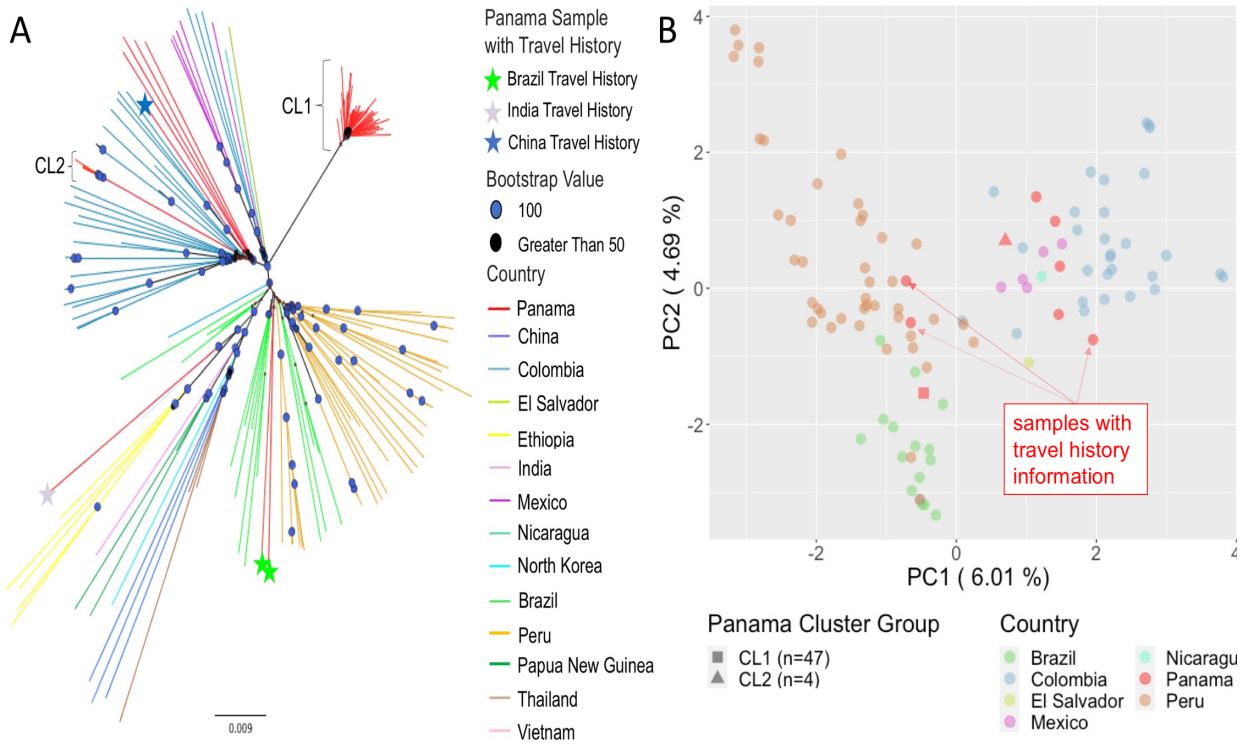


Fig 2.4: Population Structure. A) Neighbor-joining tree for all samples worldwide. Node symbols denote support values: circles indicate 100% support, triangles indicate > 50% support. Branch color indicates the country of collection for each sample. Panamanian samples with travel history are noted with the colored stars. B) PCA of Central and South American samples. Circle color indicates country of collection. Consensus sequences for cluster one and cluster two are noted as a square and triangle respectively. Panamanian samples with travel history are annotated.

grouped with the Colombian samples. These samples appear to share distant ancestry with each other and the rest of the Colombian samples. The samples also formed their own sub-cluster within the Colombian cluster.

PCA conducted with worldwide samples showed tight clustering of all Central and South American samples, with only one Panamanian sample falling outside this Central and South American cluster **(Supplemental Figure 2.3)** PCA restricted to the samples collected from Central and South America is heavily influenced by covariation among samples within the two clusters **(Supplemental Figure 2.4)**. PCA performed with a single consensus sequence

representing each cluster revealed CL1 clusters with samples from Peru and Brazil and CL2 clusters with the Colombian samples **(Fig 2.4B).** All four outbred Panamanian samples that shared no recent ancestry with the other samples also clustered with the Colombian samples. Principal component one differentiated CL1 and samples from Brazil and Peru from CL2 and the rest of the Central and South American samples.

Genomic Data Are Concordant With Travel History in Three Out of Four Cases

Four of the 59 samples had travel history data associated with them. All of these samples with travel history data were collected during the 2007-2009 period. Travel history information suggested that two samples were originally from Brazil, one sample from India, and one sample from China. The two samples with Brazilian travel history fell within the Brazilian cluster on the NJ tree and clustered with Brazilian samples on PCA **(Fig 2.2A and 2.2B)**. The one sample with Indian travel history grouped with the other Indian samples on the NJ tree, and clustered with the other Indian samples via PCA as well (**Fig 2.2A and Supplemental Figure 2.4)** This sample was the only one collected in Panama to fall outside of the Central and South American cluster in the PCA with the worldwide sample set. For the two samples with Brazilian travel history and one sample with Indian travel history, genomic data supported the same country of origin as the travel history information.

The sample with Chinese travel history had a discrepancy between the region of origin suggested by its travel history information and its genomic data. This sample clusters with the Central and South American samples on the worldwide PCA instead of with the samples from China. This sample clustered with the Colombian samples in the PCA conducted with only the Central and South American samples **(Fig 2.4B)**. Similarly, on the NJ tree, this sample fell

within the Colombian cluster with 100% bootstrap support along with the four Panamanian samples that shared zero IBD with other Panamanian samples in the dataset.

2.5 Discussion

Panama is on the cusp of eliminating malaria after several decades of intervention (5). We found extremely high clonality in the Panamanian *P. vivax* population, observing that the majority of the successfully sequenced samples (47/59) belonged to a single highly related lineage, CL1. CL1 has persisted throughout Panama for at least a decade, despite ongoing elimination efforts. Sample contamination could not explain this pattern as samples were collected in two collection periods 10 years apart and extracted, amplified, and sequenced separately. Our study suggests that the Panamanian *P. vivax* population has been through a strong bottleneck due to reduced transmission, resulting in the majority of the population belonging to a single highly related lineage. Similar reductions in clonal diversity of *P. vivax* populations have been observed elsewhere. For example, a study investigating the relationship between *P. vivax* transmission intensity and genetic diversity in Malaysia (31) documented that when there is a decline in parasite transmission, there is an increase in the clonal composition of the population. Several studies of *P. falciparum* genetic diversity and transmission intensity from Senegal (32), Thailand (33,34), and Colombia (34) have also noted the same relationship. Furthermore, there is evidence of persistence and transmission of *P. vivax* clonal lineages in Malaysia (31), and *P. falciparum* clonal lineages in Colombia (35), Ecuador (36), and Haiti (37). Our study demonstrates a similar relationship in Panama between low transmission and extremely low genetic diversity of the *P. vivax* population. Almost all CL1 samples (46/47) share IBD > 0.95 with at least one other CL1 sample, suggesting a substantial fraction of this population is clonal. Several previous studies found that the Central and South American *P.*

vivax populations are distinct from each other (20,30,38). A previous study suggested this population structure is due to multiple founding events after likely European introduction (39). This structure could also be due to genetic drift since founding. However, the Panamanian *P. vivax* population has been through too severe of a bottleneck to help clarify historical causes of this population structure with the present data. Both scenarios point to the need for further longitudinal genomic studies of *Plasmodium* parasites to better understand population dynamics over space and time.

Previous studies have indicated that Panama has focal transmission in indigenous regions (Comarcas) (5,6,11). Malaria transmission in Panama is increasingly concentrated in the Comarcas, with the proportion of total malaria cases in Panama reported from the Comarcas rising from 41.8% in 2005 to 90% in 2019 (40). Prior work shows that low transmission can lead to an increase in clonal population structure (33). The finding that CL1 is distributed ubiquitously throughout Panama is unexpected given the concentration of the malaria epidemic within spatially separated regions of the country. The geographic distribution of CL1 suggests that parasites have historically moved throughout the country, founding new populations, or supplanting small existing ones. Case investigations and understanding human movement patterns throughout Panama will be critical to achieving elimination.

This study had some limitations. We were unable to generate high-quality sequencing data from ~40% of the samples**.** Factors such as differences in DNA extraction techniques used for the two sample collection periods or length of storage of the samples could have affected DNA yield and/or molecular weight, impacting SWGA. Some samples may have had lower coverage due to lower parasitemia. Dissimilarities in coverage between the early and late sample batches could be due to technical factors such as different flowcell loading. We did not find an

association between coverage and geographic location. Low sequencing coverage for some samples may have limited the sensitivity of the F_{ws} statistic to detect polyclonal infections. However, we used a filtered dataset of SNPs for F_{ws} calculation that had a minimum of coverage of 5x and a maximum coverage of 30x to minimize bias from coverage. We also used a small set of 264 SNPs that were called in ~80% of samples to calculate Nei's standard genetic distance to determine if the excluded samples were genetically distant from CL1 or CL2. We found that the majority of the excluded samples were genetically similar to CL1 **(Supplemental Figure 2.2).** This result indicates that our assessment of relatedness within the Panamanian *P. vivax* population is not biased by parasitemia or other factors that could have affected sequencing success. This finding also suggests that we did not miss additional genetically distinct circulating Panamanian *P. vivax* strains and thus did not bias our analysis by excluding these samples. Additionally, all samples that did yield usable sequence data were distributed across almost all localities across Panama. The exception to this was a group of samples collected near the Panamanian-Costa-Rican border from which we were unable to generate usable sequence data. However, most ongoing malaria transmission in Panama occurs East of the Panama Canal, where most of the samples that generated usable sequencing data were collected (6,11). Due to the geographic sampling coverage of regions with ongoing malaria transmission, we believe these data are reflective of the current state of the Panamanian *P. vivax* population. We also lacked geographic collection data for two of the successfully sequenced samples, and they were excluded from the geographic analysis. The lack of geographical data is unlikely to bias our conclusions since these samples came from both different collection periods and regions. The two samples also constitute a small proportion of samples in the final dataset.

Genomic epidemiology can help to support malaria elimination efforts in Panama in multiple ways. First, genomic data can help identify genetically distinct cases that may be imported. Panama sits at the crux of migration paths to the United States, and it is possible that genetically distinct samples collected in Panama represent imported cases. Integrating travel history information with genomic data can help solidify the identification of imported cases. Four samples had patient travel history information, and genomic data supported the presumed country of origin for three of them. The fourth sample was collected from a subject with travel history from China. However, it clustered with Colombian samples on the NJ tree and the PCA, suggesting the infection was likely acquired somewhere in Central or South America, rather than China. Relapsing *P. vivax* infections resulting from dormant hypnozoites could complicate reconciliation of travel history with genomic data if infections were acquired months previously. Further development of tools using a benchmarked set of markers, such as SNP barcodes (41,42), for each *P. vivax* endemic country would help to identify parasite country of origin solely using genomic data.

Second, genomic data will be critical to determine if imported parasites are contributing to local transmission and/or admixing with the local parasite population. For example, we did not observe evidence of admixture between the imported samples from India and Brazil and the samples that comprise CL1, or evidence of onward clonal transmission of the imported samples. Our data cannot distinguish whether CL2 is a native Panamanian parasite lineage or if it has been imported from Colombia. However, the four CL2 samples displayed a genomic and epidemiological pattern consistent with recent local transmission, as all CL2 samples are virtually identical and were collected from the same municipality in 2019.

Overall, the existence of one main parasite genetic lineage exhibiting no recent evidence of outcrossing with imported infections suggests that Panama is ripe for the elimination of *P. vivax*. While case importation remains a threat, the lack of evidence of outcrossing suggests it may not be sufficient to prevent elimination under present circumstances. The potential for genomic data to identify imported cases in Panama will be improved by collecting genomic data from other countries in the region as a population genomics reference. Ongoing genomic surveillance paired with case containment efforts will also be needed to mitigate the risk of outbreaks resulting from imported cases and prevent reversal of the impressive progress that has been recently made towards malaria elimination in Panama.

2.6 References

1. World Health Organization. World Malaria Report 2019. Vol. 1. WHO; 2019.
2. Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, Anstey NM. Vivax malaria: neglected and not benign. Am J Trop Med Hyg. 2007 Dec;77(6 Suppl):79–87.
3. Arevalo-Herrera M, Quiñones ML, Guerra C, Céspedes N, Giron S, Ahumada M, et al. Malaria in selected non-Amazonian countries of Latin America. Acta Trop. 2012 Mar;121(3):303–14.
4. Obaldia N, Baro NK, Calzada JE, Santamaria AM, Daniels R, Wong W, et al. Clonal outbreak of Plasmodium falciparum infection in eastern Panama. J Infect Dis. 2015 Apr 1;211(7):1087–96.
5. Carrera LC, Victoria C, Ramirez JL, Jackman C, Calzada JE, Torres R. Study of the epidemiological behavior of malaria in the Darien Region, Panama. 2015–2017. PLOS ONE. 2019 Nov 15;14(11):e0224508.
6. Calzada JE, Marquez R, Rigg C, Victoria C, De La Cruz M, Chaves LF, et al. Characterization of a recent malaria outbreak in the autonomous indigenous region of Guna Yala, Panama. Malar J. 2015 Nov 17;14:459.
7. World Health Organization. Panama Malaria Country Profile 2017. WHO Malaria Country Profiles [Internet]. 2017;1(1). Available from: https://www.who.int/malaria/publications/country-profiles/profile_pan_en.pdf?ua=1
8. Ruktanonchai NW, DeLeenheer P, Tatem AJ, Alegana VA, Caughlin TT, Zu Erbach-Schoenberg E, et al. Identifying Malaria Transmission Foci for Elimination Using Human Mobility Data. PLoS Comput Biol. 2016 Apr;12(4):e1004846.
9. Marshall JM, Touré M, Ouédraogo AL, Ndhlovu M, Kiware SS, Rezai A, et al. Key traveller groups of relevance to spatial malaria transmission: a survey of movement patterns in four sub-Saharan African countries. Malar J. 2016 Apr 12;15:200.
10. Wesolowski A, Taylor AR, Chang H-H, Verity R, Tessema S, Bailey JA, et al. Mapping malaria by combining parasite genomic and epidemiologic data. BMC Medicine. 2018 Oct 18;16(1):190.
11. Lainhart W, Dutari LC, Rovira JR, Sucupira IMC, Póvoa MM, Conn JE, et al. Epidemic and Non-Epidemic Hot Spots of Malaria Transmission Occur in Indigenous Comarcas of Panama. PLoS Negl Trop Dis [Internet]. 2016 May 16 [cited 2020 Apr 29];10(5). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4868294/
12. José Arcia. Darién, convertido en un río imparable de migrantes irregulares [Internet]. La Estrella de Panamá. 2019 [cited 2020 Apr 29]. Available from: https://www.laestrella.com.pa/nacional/190709/rio-darien-imparable-migrantes-convertido
13. Neafsey DE, Volkman SK. Malaria Genomics in the Era of Eradication. Cold Spring Harb Perspect Med. 2017 Aug;7(8):a025544.
14. Cowell AN, Loy DE, Sundararaman SA, Valdivia H, Fisch K, Lescano AG, et al. Selective Whole-Genome Amplification Is a Robust Method That Enables Scalable Whole-Genome Sequencing of Plasmodium vivax from Unprocessed Clinical Samples. mBio. 2017 07;8(1).
15. Bright AT, Tewhey R, Abeles S, Chuquiyauri R, Llanos-Cuentas A, Ferreira MU, et al. Whole genome sequencing analysis of Plasmodium vivax using whole genome capture. BMC Genomics. 2012 Jun 21;13(1):262.
16. Wooden J, Kyes S, Sibley CH. PCR and strain identification in Plasmodium falciparum. Parasitol Today (Regul Ed). 1993 Aug;9(8):303–5.
17. Obaldia N, Meibalan E, Sa JM, Ma S, Clark MA, Mejia P, et al. Bone Marrow Is a Major Parasite Reservoir in Plasmodium vivax Infection. mBio. 2018 08;9(3).
18. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754–60.
19. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011 Aug 1;27(15):2156–8.
20. Hupalo DN, Luo Z, Melnikov A, Sutton PL, Rogov P, Escalante A, et al. Population genomics studies identify signatures of global dispersal and drug resistance in Plasmodium vivax. Nat Genet.

2016;48(8):953–8.
21. 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 Feb 29;7(2):e32891.
22. Lee S, Harrison A, Tessier N, Tavul L, Miotto O, Siba P, Kwiatkowski D, Müller I, Barry AE and Bahlo M. Assessing clonality in malaria parasites using massively parallel sequencing data. in preparation. 2016;
23. Schaffner SF, Taylor AR, Wong W, Wirth DF, Neafsey DE. hmmIBD: software to infer pairwise identity by descent between haploid genotypes. Malar J. 2018 May 15;17(1):196.
24. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012 Dec 1;28(24):3326–8.
25. Pembleton LW, Cogan NOI, Forster JW. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Molecular Ecology Resources. 2013;13(5):946–52.
26. Paradis E. pegas: an R package for population genetics with an integrated-modular approach. Bioinformatics. 2010 Feb 1;26(3):419–20.
27. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 2019 Feb 1;35(3):526–8.
28. Jombart T, Ahmed I. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics. 2011 Nov 1;27(21):3070–1.
29. Rambaut A, Drummond AJ. FigTree version 1.4. 0. 2012.
30. Rougeron V, Elguero E, Arnathau C, Acuña Hidalgo B, Durand P, Houze S, et al. Human Plasmodium vivax diversity, population structure and evolutionary origin. PLoS Negl Trop Dis [Internet]. 2020 Mar 9 [cited 2020 May 4];14(3). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7082039/
31. Auburn S, Benavente ED, Miotto O, Pearson RD, Amato R, Grigg MJ, et al. Genomic analysis of a pre-elimination Malaysian Plasmodium vivax population reveals selective pressures and changing transmission dynamics. Nature Communications. 2018 Jul 3;9(1):2585.
32. Daniels RF, Schaffner SF, Wenger EA, Proctor JL, Chang H-H, Wong W, et al. Modeling malaria genomics reveals transmission decline and rebound in Senegal. Proc Natl Acad Sci USA. 2015 Jun 2;112(22):7067–72.
33. Nkhoma SC, Nair S, Al-Saai S, Ashley E, McGready R, Phyo AP, et al. Population genetic correlates of declining transmission in a human pathogen. Mol Ecol. 2013 Jan;22(2):273–85.
34. 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 Jul 19;487(7407):375–9.
35. Echeverry DF, Nair S, Osorio L, Menon S, Murillo C, Anderson TJ. Long term persistence of clonal malaria parasite Plasmodium falciparum lineages in the Colombian Pacific region. BMC Genetics. 2013 Jan 7;14(1):2.
36. Sáenz FE, Morton LC, Okoth SA, Valenzuela G, Vera-Arias CA, Vélez-Álvarez E, et al. Clonal population expansion in an outbreak of Plasmodium falciparum on the northwest coast of Ecuador. Malar J [Internet]. 2015 Dec 10 [cited 2020 Jun 2];14. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4676133/
37. Charles M, Das S, Daniels R, Kirkman L, Delva GG, Destine R, et al. Plasmodium falciparum K76T pfcrt Gene Mutations and Parasite Population Structure, Haiti, 2006–2009. Emerg Infect Dis. 2016 May;22(5):786–93.
38. Koepfli C, Rodrigues PT, Antao T, Orjuela-Sánchez P, Van den Eede P, Gamboa D, et al. Plasmodium vivax Diversity and Population Structure across Four Continents. PLoS Negl Trop Dis. 2015;9(6):e0003872.
39. Rodrigues PT, Valdivia HO, de Oliveira TC, Alves JMP, Duarte AMRC, Cerutti-Junior C, et al.

Human migration and the spread of malaria parasites to the New World. Sci Rep [Internet]. 2018 Jan 31 [cited 2020 Oct 2];8. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5792595/

40. Hurtado L, Cumbrera A, Rigg C, Perea M, Santamaría AM, Chaves LF, et al. Long-term transmission patterns and public health policies leading to malaria elimination in Panamá. Malar J [Internet]. 2020 Jul 23 [cited 2020 Aug 3];19. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7376851/
41. Baniecki ML, Faust AL, Schaffner SF, Park DJ, Galinsky K, Daniels RF, et al. Development of a Single Nucleotide Polymorphism Barcode to Genotype Plasmodium vivax Infections. PLoS Negl Trop Dis [Internet]. 2015 Mar 17 [cited 2020 Apr 29];9(3). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4362761/
42. Trimarsanto H, Amato R, Pearson RD, Sutanto E, Noviyanti R, Trianty L, et al. A molecular barcode and online tool to identify and map imported infection with Plasmodium vivax. bioRxiv. 2019 Sep 24;776781.

Chapter 3: Comparative *Plasmodium* Selection Scans Identify *P. vivax* Drug Resistance Candidates Under Likely Antimalarial Driven Selection

3.1 Abstract

P. falciparum and *P. vivax* have been exposed to chemical compounds for more than a hundred years. There are several validated drug resistance genes in *P. falciparum*, and even more candidate loci. However, despite clear evidence of clinical drug resistance to chloroquine (CQ) in *P. vivax*, there is only one validated drug resistance gene (*pvdhfr*), and only a few candidate drug resistance genes mostly based on them being orthologs of *P. falciparum* drug resistance genes. Finding *P. vivax* drug resistance molecular mechanisms and markers is important to understand and track the development of drug resistance in this species. *Plasmodium knowlesi* is a closely related parasite to *P. vivax*, can infect humans and apes. However, *P. knowlesi* is thought to not transmit onward from infected humans. Therefore, *P. knowlesi* has only recently been exposed to anti-malarial drugs, and its genome should not exhibit signs of selection due to pressure from these drugs. I compared the DoS, iHS, and F_{ST} statistics (see definitions below) for all 1:1 orthologs in both species to identify a set of putative drug resistance loci in *P. vivax* to prioritize for validation in the *P. knowlesi* model system to test if they mediate drug resistance. I found that the *P. vivax* orthologs of *P. falciparum* drug resistance displayed signals of selection in several populations. I also compared signals in a set of genes that encode transporters but found no different in selection signals. I highlight drug resistance orthologs and transporter gene outliers from each test. I discuss plans to validate one drug resistance ortholog outlier, *pvpm4,* an ortholog of the *P. falciparum plasmepsinII/III,* genes in a *P. knowlesi* model system.

3.2 Introduction

The parasites *Plasmodium vivax* and *Plasmodium falciparum* cause the majority of human malaria cases[1]. Both species have been treated with antimalarial drugs for more than a hundred years[1]. However, resistance to antimalarials subsequently emerged in both species and continues to threaten ongoing malaria control and elimination efforts[1]. The molecular basis of *P. falciparum* drug resistance is well understood, with several validated drug resistance genes and even more candidate resistance loci[1]. There is comparatively limited knowledge of the molecular basis of *P. vivax* drug resistance, despite this species co-occurring with *P. falciparum* and subsequently being exposed to the same antimalarial compounds for hundreds of years[2–5]. *P. vivax* clinical resistance to chloroquine is well-documented[2,6]. There is also evidence of *P. vivax* resistance to several other antimalarial compounds[2,5–7]. However, *P. vivax* genes that cause chloroquine resistance and possible resistance to other antimalarials have not yet been firmly identified and validated. Identification of the molecular changes in *P. vivax* genes that lead to drug resistance is urgently needed to both understand the molecular basis of drug resistance and track its spread using genomic surveillance. Mutations in *pvdhfr-ts* are the only validated molecular markers of *P. vivax* drug resistance[5,8]. Beyond *pvdhfr-ts*, there are only a handful of candidate drug resistance loci: *pvmdr1, pvcrt,* and *pvmrp1*[6].

Sulfadoxine-pyrimethamine plus (SP) has never been recommended for the treatment of *P. vivax*[8]. However, *pvdhfr-ts* mutations found in *P. vivax* populations in Papua New Guinea have been validated in a yeast model system to cause pyrimethamine resistance and were associated with SP treatment failure in patients[5,8,9]. *P. vivax* and *P. falciparum* are co-endemic in Papua New Guinea. Pyrimethamine-resistant *pvdhfr* alleles are evidence that antimalarials used to treat *P. falciparum* can select for *P. vivax* resistance variants, even if that drug is not used to directly treat *P. vivax*[5,8,9]. There is further evidence of co-selection for SP-resistant *P. vivax* in

India, where a higher frequency of resistant *pvdhfr-ts* alleles is observed in regions where *P. vivax* is co-endemic with *P. falciparum*[4,10]. Conversely, a lower frequency of resistant pv*dhfr-ts* alleles is observed in regions of India where *P. vivax* is the sole malaria-causing parasite[4,10]. Chloroquine (CQ) and primaquine (PQ) remain the standard *P. vivax* treatment in most regions except Papua New Guinea, Malaysia, and Oceania[11]. These regions have high-grade CQ-resistance (CQR) and use artemisinin combination therapies (ACTs) to treat *P. vivax* infections, specifically dihydroartemisinin (DHA) plus piperaquine (PIP) or artemether (AM) plus lumefantrine (LUM)[12]. Identifying the *P. vivax* genes involved in resistance to antimalarial drugs will be critical to ensure effective *P. vivax* malaria treatment recommendations.

The discrepancy in understanding of drug resistance between *P. falciparum* and *P. vivax* stems in part from the lack of a continuous *in vitro* culture system for *P. vivax*[6]. The lack of a continuous *in vitro P. vivax* culture system makes it challenging to interrogate candidate drug resistance variants *in vitro* for molecular validation[6]. Furthermore, differences in life cycles between *P. vivax* and *P. falciparum* can complicate the detection of true *P. vivax* drug resistance. *P. vivax* liver sporozoites can develop into hypnozoites, which can remain dormant for months to years[11]. When hypnozoites become active, they develop into merozoites and continue the infection cycle again. Hypnozoites can currently only be treated with the drugs primaquine (PQ) and tafenoquine (TQ). Relapsing parasites emerging from the hypnozoite stage can make it difficult to determine if persistent infection after treatment is due to resistance, relapse (caused by liver-stage hypnozoites maturing into merozoites), or reinfection by a new strain[13].

Prior studies using population genomic approaches to link *P. vivax* polymorphisms with drug resistance[7,14–19]. The discrepancy between studies is likely due to complications with defining true *P. vivax* drug resistance due to its life cycle and difficulties sequencing parasites

from clinical isolates[6,13]. Alternative approaches using comparative genomics in conjunction with population genomics could help identify and prioritize candidate *P. vivax* drug resistance genes. Comparative genomic studies use genomic information across species to understand conserved biological processes. *Plasmodium* parasite comparative genomic studies have identified possible drug and vaccine targets, the expansion of multi-gene families that may play a role in host adaptation or immune evasion and answered questions about the basic biology of these parasites[20–22]. Comparative population genomic analyses combine divergence-based methods of detecting selection with within-species polymorphism data[23]. Combining divergence and polymorphism datasets has been used to identify loci under strong purifying and recent selection within species and identify how different environmental selection pressures affect evolution between species[23,24]. The rapid generation of *Plasmodium* population genomics data from other members of the *Plasmodium* genus, besides *P. falciparum* and *P. vivax*, presents an opportunity to conduct novel population-level polymorphism/divergence analyses. Particularly, identifying loci that are under selection in one species and not in another could highlight genes under differential environmental pressures. Environmental pressures that differ between species include the pressure of antimalarial drugs, host immune pressure, and adaptation to different mosquito vectors. Genes that may be under environmental-specific pressures that differ between species could be genes that play key roles in drug resistance, immune invasion, host-specific virulence factors, and vector competence. For unculturable pathogens like *P. vivax,* choosing candidate genes to prioritize research focus can be risky, given the cost and time needed to characterize that gene in a model system. Comparative selection scans can help prioritize genes under selection only in *P. vivax* for functional characterization *in vitro*.

P. knowlesi is a closely related parasite to *P. vivax*, which primarily infects long-tailed and pig-tailed macaques but can also infect humans[25–29]. *P. knowlesi* has likely also been exposed to antimalarial compounds, but since it is thought to not transmit onward from infected humans, its genome should not exhibit signs of selection from antimalarial compounds[25,30,31]. We hypothesized that comparing signals of evolution between *P. vivax* and *P. knowlesi* could help identify candidate drug resistance loci in *P. vivax. P. vivax* genes with a signal of selection only in *P. vivax* populations, and not in their *P. knowlesi* orthologs, are under selection pressures unique to *P. vivax*. Antimalarial drug pressure is one of the unique selection pressures acting on *P. vivax* populations. If that same gene is only under selection in *P. vivax* populations and is also expressed in the blood stage, which most antimalarials target, it could be a candidate drug resistance gene. Genes under selection uniquely in *P. vivax* populations that are orthologs of known *P. falciparum* drug resistance genes are also compelling *P. vivax* drug resistance candidates.

In this study, we present a framework for conducting comparative selection scans between two *Plasmodium* species. We use that approach to identify genes under *P. vivax*-specific selection pressures, with a focus on genes that possibly mediate *P. vivax* drug resistance. We used existing *P. vivax* and *P. knowlesi* population genomic datasets to conduct comparative selection scans using three different tests of selection: Direction of Selection (DoS)[32], the integrated Haplotype Score (iHS)[33,34], and the Fixation Index (F_{ST})[35]. These three tests each measure selection using different signals from divergence/polymorphism data. We then identified which *P. vivax* orthologs of known *P. falciparum* drug resistance and *P. vivax* transporter genes were outliers in these tests, and not under selection in the *P. knowlesi* population. We examine how comparative selection scans can prioritize *P. vivax* drug resistance

candidates for *in vitro* characterization in heterologous model systems. We also discuss how this approach can be applied to other datasets or used to prioritize genes that may mediate other phenotypes for *in vitro* characterization. Lastly, we highlight plans to characterize one of the *P. vivax* candidate drug resistance genes, *pvpm4*, which encodes an aspartic protease in the digestive vacuole of the parasite, using a *P. knowlesi* model system.

3.3 Methods

Data

P. vivax data were accessed from the NCBI Sequencing Read Archive with the following accession numbers from Hupalo et al, 2016[36]: PRJNA240356 – PRJNA240533. We also accessed *P. vivax* samples from the European Nucleotide Archive using accession numbers from Pearson et al, 2016[37]. 37 *P. knowlesi* samples were accessed from the European Nucleotide Archive from Assefa et al, 2015[25] using accession numbers SAMEA3503891- SAMEA3503927.

Candidate List Selection

We identified *P. vivax* orthologs of known *P. falciparum* drug resistance genes with the "transform by orthology" function on the PlasmoDB database (https://plasmodb.org/plasmo/app)[38]. We also sought to identify *P. vivax* transporter genes, since transporters play a role in drug resistance in many pathogens[39–42]. We identified *P. vivax* transporter genes using the "search by GO term" tool on the PlasmoDB web server and the GO term for transport activity (GO:0005215), which resulted in an initial list of 106 genes. We then identified which transporters were expressed in the blood stage of the parasite by downloading the transcriptomics expression data from Zhu et al, 2016 from PlasmoDB[43]. We used a custom script to identify which genes were expressed during the intraerythrocytic developmental cycle (IDC) by identifying which genes had expression values above or below the median expression

value at each time point. Genes above the median were classified as expressed, and ones below the median were classified as not expressed. If the gene was expressed at any point during the IDC, it was included for further analysis. After this filtering step, we were left with a final list of 81 transporter genes.

Genome Alignment, Variant Calling, and Sample Filtering:

P. vivax sample reads were aligned to the *P. vivax* P01 reference genome assembly using BWA-MEM, version 0.7[44,45]. *P. knowlesi* sample reads were aligned to the *P. knowlesi* PKH reference genome assembly[46]. We marked duplicate reads using the MarkDuplicates tool from Picard tools.[47] We next performed local realignment around indels using the Genome Analysis Toolkit (GATK) RealignerTargetCreator and GATK IndelRealigner (GATK Version 3.5.0)[48]. We called variants using GATK HaplotypeCaller using best practices to call and filter SNPs and generate gVCFs for each sample. We performed joint variant calling on the collections of *P. vivax* and *P. knowlesi* gVCFs separately using the GATK GenotypeGVCFs tool with GATK hard filters, including calls in subtelomeric regions. We then used the GenotypeGVCFs tool to construct a joint VCF with all samples.

We excluded samples from the analysis if they either missed a high proportion of genotype calls, were polyclonal, or were highly related. We calculated the percentage of sites not called in each sample out of the entire set of variants in accessible regions. Samples missing more than 50% of calls were excluded from further analysis (n=204/745).

Next, we estimated sample clonality using the F_{ws} statistic. F_{ws} measures the within-sample genetic diversity (measured by heterozygosity H_w) relative to the overall population genetic diversity (H_s)[49]. The theory behind F_{ws} assumes that a monoclonal (single strain) infection has extremely low genetic diversity relative to population genetic diversity. Polyclonal

(multiple strain) infections have high diversity relative to population diversity (compared to a monoclonal infection). We can distinguish between monoclonal and polyclonal infections by estimating the ratio between within-host diversity and population diversity[49]. A sample with an F_{ws} statistic of 0.95 or greater (>=0.95) is considered monoclonal. We calculated F_{ws} using the R package moimix40 for all *P. vivax* and *P. knowlesi* samples. Polyclonal samples (n=228/745) were excluded from further analysis.

Finally, we excluded samples that were highly related. We used hmmIBD[50] to estimate the proportion of sites identical by descent (IBD) between sample pairs in each of the populations. We estimated minor allele frequency (MAF) in each population. Next, we subsetted the master VCF file into sub-VCFs for each sample population. Sites were excluded based on the minimum and maximum read depth (five and thirty respectively) to ensure that we were using only high-quality SNPs. Each VCF was re-formatted using a custom Perl script for input into hmmIBD. We conducted analysis and visualization of the hmmIBD output using custom R scripts. Samples with high IBD (IBD > 0.75) were excluded from further analysis (n=46/745).

We excluded 478 out of 745 samples from the dataset after the sample filtering steps. We filtered out low-quality sites in the remaining 267 *P. vivax* samples to ensure a set of trustworthy data to conduct the selection scans. We filtered to keep all biallelic SNPs with a QUAL score of at least 30 and were missing in no more than 25% of the samples in that population. The joint VCF after these filtering steps contained 267 samples and 2,361,838 variants.

P. knowlesi sample alignment, variant calling, and filtering followed the same procedure as above. We used the 37 *P. knowlesi* samples from cluster one as determined in Assefa et al, 2015[25]. No *P. knowlesi* samples were excluded based on missingness, or high relatedness. Eight

P. knowlesi samples (8/37) were excluded for being polyclonal. After filtering out low-quality sites, we were left with a final dataset of 29 samples and 1,903,331sites.

Direction of Selection Test Calculation

We performed the direction of selection (DoS) test using the method outlined by Stoletzki and Eyre-Walker[32]. If brief, this method compares between-species divergence (nonsynonymous (D_N) or synonymous (D_S) divergence) to within-species polymorphism (nonsynonymous (P_N) or synonymous (P_S) polymorphism). The DoS statistic can be used to look for outliers to assess if a gene is A) under selection, and B) if so, under purifying selection or adaptive evolution. The formula for calculating DoS is $D_N/(D_N+D_S)- P_N/(P_N+P_S)$. Negative DoS statistics can indicate either balancing selection or weakened purifying selection. DoS statistics near zero imply no selection acting on that gene. Positive DoS statistics indicate positive selection that has fixed adaptive substitutions in that gene.

We calculated the DoS statistic using a custom Python script that counted the number of synonymous and nonsynonymous substitutions between *P. vivax* and *P. inui,* and *P. knowlesi* and P. *inui*. The script also counted the number of nonsynonymous and synonymous polymorphisms in each species. The script then calculated the DoS statistic after counting D_N, D_S, P_N, and P_S. Genes with both P_N, and P_S counts less than three were excluded from the analysis.

F_{ST} Calculation

We used VCFtools[51] to calculate F_{ST} using the Weir and Cockerham method[52] in one kilobase (kb) non-overlapping sliding windows for each population or population pairwise comparison. We then used a custom Python script to annotate these one kb windows for the

genes that fell within them to use for further analysis. Windows with fewer than five SNPs were excluded from the analysis.

Integrated Haplotype Score

We used the R package *rehh*[53] to calculate the integrated haplotype score (iHS) for each population. We only used monoclonal samples in our analysis and could therefore assume that our samples were phased. We used the consensus allele across all genomes in each of the respective *P. vivax* and *P. knowlesi* datasets as the ancestral allele and used that to calculate iHS across the genome using the *ihh2ihs* function. This approach followed the methods used by Hocking et al and Assefa el al[25,28]. We then used a custom script to annotate the iHS output with gene location information for each species.

Statistical Data Analysis and Visualization

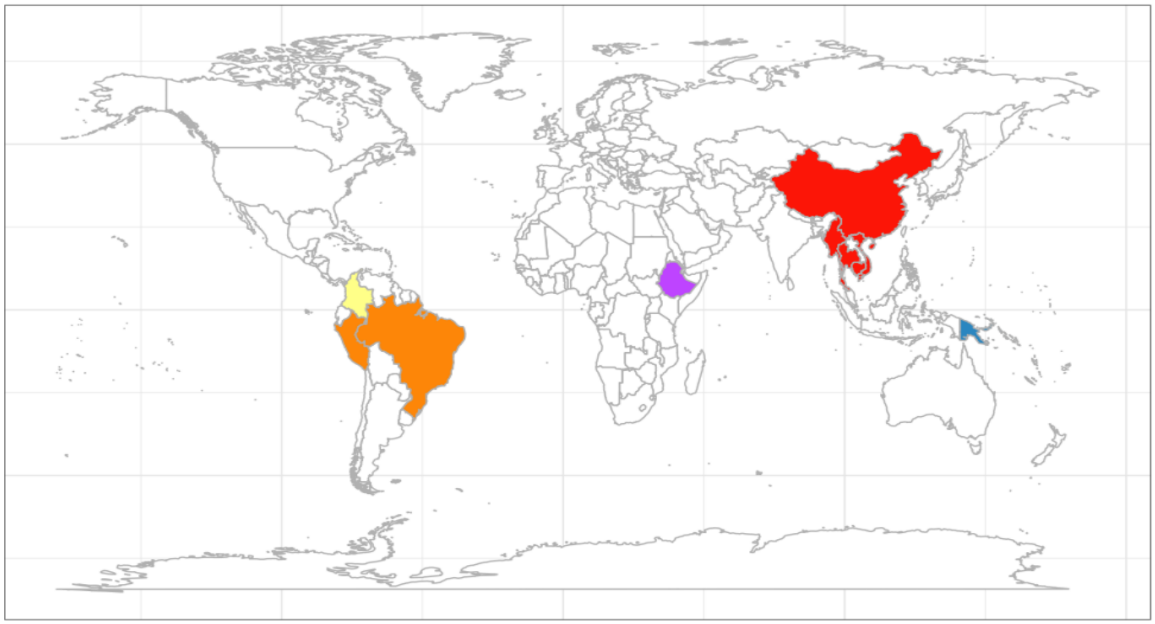
All data analysis and visualizations were conducted using R software (Version 4.1). We tested the hypothesis that the set of drug resistance genes or transporter genes had mean test statistics significantly different from the rest of the population using a bootstrapping approach. We randomly selected a set of genes (either 10 or 81 if the sample set was the drug resistance orthologs or transporter genes respectively) from the population with replacement 10,000 times. We then calculated a p-value for each set of genes in each test of selection by counting the number of occurrences, where a draw had a mean test statistic for that test greater than the mean test statistic for either the set of drug resistance orthologs or transporter genes divided by the number of draws (n=10,000). We considered p-values below 0.01 significant. Data visualization was conducted using the *ggplot2*[54] and *ggrepe*l packages.

3.4 Results:

Variant Calling and Sample Filtering

The initial *P. vivax* dataset contained parasite samples from six major geographic regions: Papua New Guinea (PNG), the Greater Mekong Subregion plus China (GMS plus China), Ethiopia, the Peruvian/Brazilian Amazon (Peru/Brazil), Colombia, and India. We used publicly available sequencing data from Assefa et al. to generate the *P. knowlesi* dataset[25]. The Assefa et al. study sequenced 48 clinical *P. knowlesi* isolates and five laboratory *P. knowlesi* strains for a total of 53 samples[25]. They identified deep population substructure in the *P. knowlesi* population, with the 53 isolates belonging to three distinct clusters[25]. We selected samples that were members of *P. knowlesi* cluster one, as described in Assefa et al, because population substructure can confound calculation and interpretation of test statistics for the selection scans, and because it was the cluster with the largest number of samples (n=37/53)[25]. *P. vivax* and *P. knowlesi* sequences were aligned to the P01 and PKH reference genomes, respectively. Samples missing more than 50% of calls were excluded from the analysis (n=204/745). We used the F_{ws} statistic to assess sample clonality and excluded *P. vivax* samples (n=228/745) and *P. knowlesi* samples (n=8/37) that were polyclonal from further analysis. Polyclonal samples could confound the selection scans by either affecting polymorphism counts or improperly phased sample haplotypes, which would affect the calculation of iHS statistics. Highly related samples were also excluded from analysis, as high relatedness between samples could lead to erroneous signals of selection by over-representing certain alleles or haplotypes. Pairwise IBD for all samples was measured using the hmmIBD program in each of the *P. vivax* populations and the *P. knowlesi* population. Samples were defined as highly related if they had a pairwise comparison, where the fraction of sites identical by descent was greater than 0.75. This cutoff was based on the distribution of pairwise IBD values in all populations (**Supplemental Figure 3.1**). We selected

the sample with the lowest percentage of missing calls in each highly related sample cluster or pair as the representative sample to be included for further analysis. The remaining samples from that cluster or pair were excluded from further analysis (n=46/745). Sample filtering steps are summarized in **Supplemental Figure 3.2**.



Country	CQ	PQ	SP	LUM	AM	AS	MQ	AQ	DHA	QN	PIP	PYN	ART
GMS+China	x	x	x	x	x	x	x	x	x	x	x	x	x
PNG	x	x	x	x	x	x							
Ethiopia	x	x	x	x	x								
Colombia	x	x	x	x	x		x	x	x				
Peru/Brazil	x	x	x			x	x	x		x			

Figure 3.1: Map of Drug Exposure for the *P. vivax* populations. Drug abbreviations are as follows: Chloroquine (CQ), Quinine (QN), Amodiaquine (AQ), Lumefantrine (LMF), Halofantrine (HL), Pyronaridine (PYN), Sulfadoxine-Pyrimethamine (SP), Mefloquine (MQ), Piperaquine (PIP), Artemether (AM), Dihydroartemisinin (DHA), Artesunate (AS), Artemisinin (ART)[55–65].

There were fewer than ten Indian *P. vivax* population samples after sample filtering, which would limit the power to detect selection due to a lack of sufficient polymorphism in that population. Subsequently, all Indian *P. vivax* samples were excluded from analysis because of the low sampling depth in that population. We proceeded with the analysis with 29 *P. knowlesi*

samples and 267 samples from five separate *P. vivax* populations: PNG (n=16), GMS plus China (n=45), Ethiopia (n=14), the Peruvian/Brazilian Amazon (n=169), and Colombia (n=23) (**Figure 3.1**).

Candidate *P. vivax* Drug Resistance Genes Have Been Subjected to Heterogeneous Pressure from Antimalarials.

We compiled a list of candidate *P. vivax* drug resistance genes before computing genome-wide selection statistics. *P. vivax* genes that are orthologs of *P. falciparum* genes that mediate *P. falciparum* drug resistance were selected as one set of orthologous candidate drug resistance genes to investigate (**Table 3.1**). Co-endemic *P. falciparum* populations have been exposed to multiple sources of pressure stemming from different antimalarial treatment regimens. By contrast, chloroquine (CQ) plus primaquine (PQ) is the standard treatment for *P. vivax* in most countries, (**Fig. 3.1**)[55–65].

P. vivax is also exposed to this diverse set of antimalarial drugs, which could select for drug resistance-causing variants (**Fig. 3.1**). The list of candidate genes and compounds to which they confer resistance in *P. falciparum* is displayed in **Table 3.1**[66–84].

Table 3.1: List of *P. vivax* candidate Drug Resistance Genes

***P. vivax* Gene**	***P. falciparum* Ortholog**	**Compounds gene provides resistance to in *P. falciparum***
pvmdr1	*pfmdr1*	CQ, AQ, QN, LUM, HL
pvcrt	*pfcrt*	CQ, AQ, PYN, QN, LUM, HL
pvdhfr	*pfdhfr*	PYR
pvdhps	*pfdhps*	SX
pvpm4	*pfpmII/pfpmIII*	PIP
pvkelch13	*pfkelch13*	AM, ART, AS DHA

Table 3.1 (Continued): List of *P. vivax* candidate Drug Resistance Genes		
pvmdr2	*pfmdr2*	PYR
pvmrp2	*pfmrp2*	CQ, MQ, QUIN
pvaat1	*pfaat1*	CQ
pvdmt1	*pfdmt1*	QN
The full list of the *P. vivax* candidate drug resistance genes. Drug abbreviations are as follows: chloroquine (CQ), quinine (QN), amodiaquine (AQ), lumefantrine (LMF), halofantrine (HL), pyronaridine (PYN). pyrimethamine (PYR), sulfadoxine (SX), mefloquine (MQ), piperaquine (PIP), artemether (AM), dihydroartemisinin (DHA), artesunate (AS), artemisinin (ART).		

P. vivax and *P. falciparum* have several biological differences[85]. The genes that mediate drug resistance in *P. vivax* may be different from the genes that mediate drug resistance in *P. falciparum*. Transporter genes play a role in drug resistance in many pathogens[39–42]. Therefore, genes that encode transporters could be compelling *P. vivax* resistance candidates. We identified a list of genes that encode transporters by searching for genes that were annotated with the Gene Ontology term for transporter (GO:0005215) and filtered out genes that were not expressed in blood-stage parasites. 81 *P. vivax* genes that encode transporters passed this filtering criterion and were included as candidate genes to investigate if they are under differential selection pressure between *P. vivax* and *P. knowlesi*. (**Supplemental Table 3.1**). While many of these transporters currently have no evidence of mediating drug resistance in *P. falciparum*, they may mediate drug resistance in *P. vivax*.

Selection Scans

We next performed two tests of selection to identify genes under differential selection between *P. vivax* and *P. knowlesi,* and one test of selection in just the *P. vivax* population to identify genes under heterogeneous selection pressures.

DoS Highlights Drug Resistance Candidates and Transporter Genes with Evidence of Balancing Selection in *P. vivax*

Comparing the amount of nonsynonymous divergence (D_N) and synonymous divergence (D_S) between species to the amount of nonsynonymous polymorphism (P_N) and synonymous polymorphism (P_S) within species is a well-established method of identifying genes under selection[32,86,87]. The McDonald-Krietman test (MK test), the Neutrality Index (NI), and Direction of Selection (DoS) are commonly applied tests that use divergence and polymorphism data to detect selection[32,86,87]. We chose *Plasmodium inui,* a closely related species to *P. vivax* and *P. knowles*i, as the outgroup to measure D_N and D_S for both species. *P. vivax* displays slightly lower counts of D_N and D_S between itself and *P. inui,* compared to the D_N and D_S counts between *P. knowlesi* and *P. inui* (**Fig. 3.2A**). While *P. inui* is situated at a similar evolutionary distance from both *P. vivax* and *P. knowlesi*, these distances are not the same, and could explain the slight increase in substitution counts for *P. knowlesi*.[88] There are lower nonsynonymous and synonymous polymorphism counts in *P. vivax* genes compared to their *P. knowlesi* orthologs (**Fig. 3.2A**). The observation, coupled with fewer *P. knowlesi* sequences (n=29) than *P. vivax* sequences (n=267) in the dataset, suggests that *P. knowlesi* has a much larger effective population size (N_e) than *P. vivax*. The higher polymorphism counts in *P. knowlesi* genes limit the comparative utility of ratio-based tests, because they have more power to detect selection in *P. knowlesi* genes compared to *P. vivax* genes. We instead used the DoS statistic because DoS is not a ratio of ratios. Therefore, the DoS statistic, unlike estimating the NI or performing the MK

test, is less prone to bias when one of the counts (D_N, D_S, P_N, P_S) is small[32]. The DoS Statistic is calculated using the following formula: $D_N/(D_N+D_S)- P_N/(P_N+P_S)$.

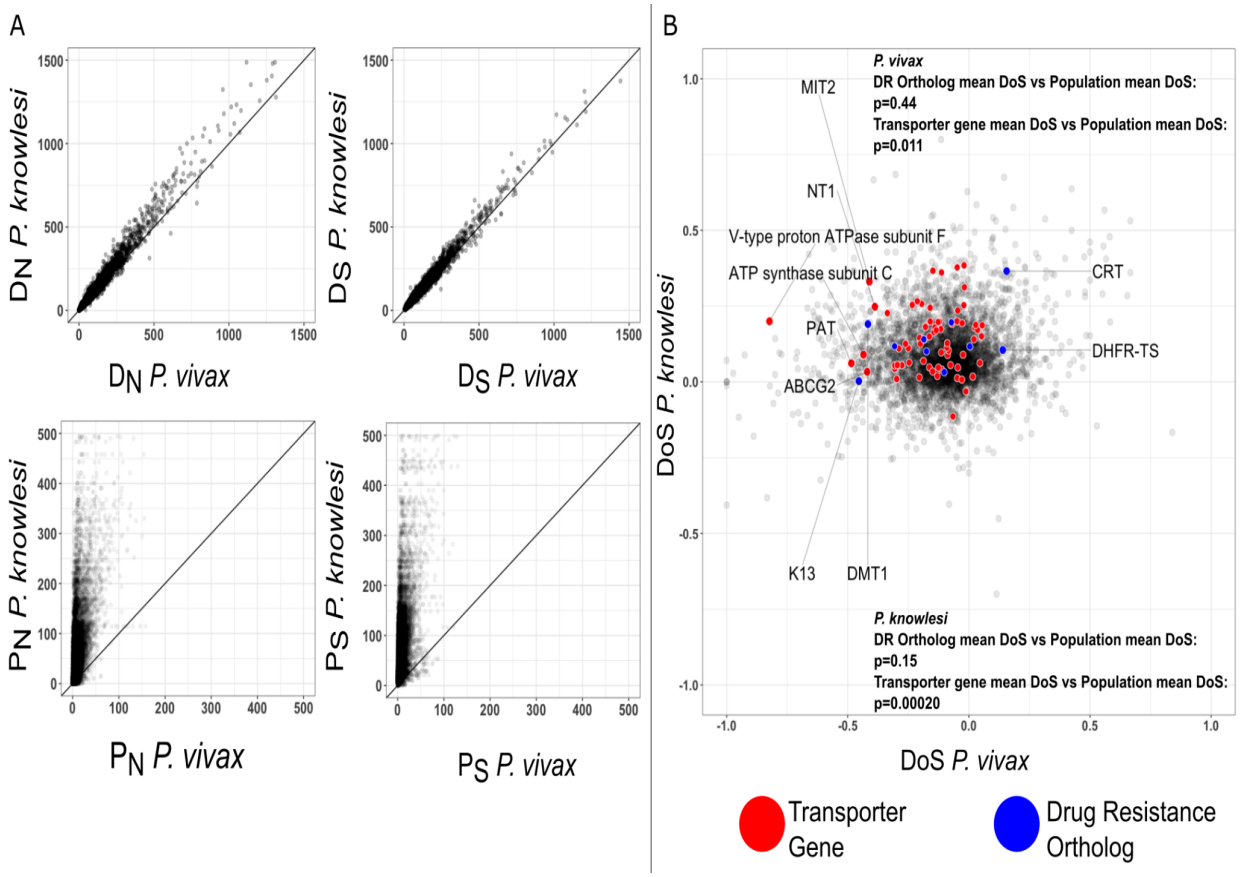


Figure 3.2: Divergence, Polymorphisms, and DoS scores for *P. vivax* vs. *P. knowlesi.* 2A Displays D_N, D_S, P_N, and P_S for *P. vivax* plotted versus *P. knowlesi*. **2B and 2C** display DoS scores for each of the 4104 *P. vivax* genes and their *P. knowlesi* orthologs. **2B** highlights and labels the set of candidate drug resistance genes in blue and transporter genes in red. Drug resistance orthologs and transporter genes in the top 10% or bottom 5% of DoS statistics in *P. vivax and* were between the bottom 5% and top 5% of DoS statistics in *P. knowlesi,* are labeled.

DoS compares the proportion of nonsynonymous divergence (D_N/D_N+D_S) to the proportion of nonsynonymous polymorphism (P_N/P_N+P_S). DoS statistics range from negative one to one. A positive DoS statistic indicates adaptive substitution has fixed advantageous nonsynonymous variants in a gene[32]. A DoS statistic near zero suggests there is no selection

pressure acting on that gene[32]. A negative DoS statistic possibly indicates either weakened purifying selection or balancing selection[32]. Negative DoS statistics result from the higher proportion of $P_{N\ out}$ of total polymorphism counts relative to the proportion of D_N out of total divergent substitution counts. The increased proportion of P_N is evidence of either weakened purifying selection, which is not effectively removing slightly deleterious mutations from the population, or balancing selection, which maintains higher frequencies of nonsynonymous variants than expected via genetic drift[32]. Negative DoS statistics could indicate balancing selection that is acting on drug resistance variants and maintaining them in the population, along with the wild-type allele.

We pooled polymorphism data for all five *P. vivax* subpopulations to calculate DoS statistics for each gene in the global *P. vivax* population. We plotted the DoS statistics for each *P. vivax* gene against its *P. knowlesi* ortholog (n=4104 genes) in **Figure 3.2B**. The distribution of DoS statistics in *P. vivax* is centered slightly below zero, with a mean DoS statistic of -0.10 and a standard deviation of ± 0.18. The smaller N_e of *P. vivax* relative to the N_e of *P. knowlesi* could lead to less efficient purifying selection and explain the negative mean DoS statistic for the population. The *P. knowlesi* DoS statistic distribution is centered slightly above zero, with a mean DoS statistic of 0.085 and a standard deviation of ± 0.12. The positive skew of the *P. knowlesi* DoS distribution is driven by the higher proportion of D_N among substitution counts in most genes, relative to the higher proportion of P_N among polymorphism counts. There was no significant difference in the DoS distribution between *P. vivax* and *P. knowlesi* (**p=0.73, Kruskal-Wallis test**). Divergence and polymorphism counts for all drug resistance orthologs and transporter genes for both *P. vivax* and *P. knowlesi* are displayed in **Supplemental Table 3.2.**

We identified drug resistance candidate genes with outlier DoS statistics in either tail of the DoS distribution of the global *P. vivax* population, and DoS statistics close to zero in *P. knowlesi* populations (**Figure 3.2B**). The mean DoS statistic of the drug resistance candidates did not significantly differ from the population mean in either *P. vivax* (**p= 0.44, bootstrap analysis)** or *P. knowlesi* (**p=0.15, bootstrap analysis**). The mean DoS of drug resistance orthologs in the *P. vivax* population was negative (mean = -0.14), while the mean DoS statistic in the *P. knowlesi* population was positive (mean=0.13). However, two drug resistance orthologs *pvkelch13* and *pvdmt1* had DoS statistics that were among the 5% most negative DoS statistics among all *P. vivax* genes. Additionally, the *pvcrt* and *pvdhfr* DoS statistics were among the top 10% of most positive *P. vivax* DoS statistics. *pkcrt* had a DoS statistic among the top 2.5% of all DoS statistics in *P. knowlesi*. There are very few nonsynonymous polymorphisms in both *pvcrt* and *pkcrt* (four P_N and 14 P_S counts in *pvcrt* compared with three P_N and 43 P_S counts in *pkcrt)*, which resulted in positive DoS statistics in both species. The remainder of the *P. vivax* drug resistance orthologs had DoS statistics between the top 5% and bottom 5% of DoS statistics in both species. DoS statistics for the drug resistance candidates in each subpopulation are presented in **Supplemental Figure 3.3**.

The mean DoS statistic of transporter genes was not significantly different from the *P. vivax* population DoS mean **(p= 0.011, bootstrap analysis**), but was significantly different from the *P. knowlesi* population mean **(p= 0.00020, bootstrap analysis**). The mean DoS statistic of transporter genes in the *P. knowlesi* population was positive (mean = 0.13), while the mean DoS statistic of transporter genes in the *P. vivax* population was negative (mean = -0.15). Several *P. vivax* transporter genes exhibited negative DoS statistics in the global *P. vivax* population, and DoS statistics close to zero in the *P. knowlesi* population (**Figure 3.2B**). PVP01_0940800, which

encodes a V-type proton ATPase subunit, had a negative DoS statistic that was among the bottom 0.05% of all *P. vivax* DoS statistics in the dataset. PVP01_0106100, which encodes an ATP synthase subunit, PVP01_1322800, which encodes an ABC transporter, ABCG2, and PVP01_0418900, which encodes a pantothenate transporter, PAT, all exhibited negative DoS statistics in the bottom 5% of all *P. vivax* DoS statistics. The *P. knowlesi* orthologs of these genes all had DoS statistics near zero. PVP01_1405100, which encodes MIT2, had a DoS statistic among the bottom 5% of DoS in the *P. vivax* population, indicating balancing selection, and a DoS statistic among the top 5% of DoS statistics in the *P. knowlesi* population, indicating positive selection. DoS statistics for the transporter genes in each subpopulation are presented in **Supplemental Figure 3.4**.

Comparative iHS scores Reveal Genes Under Recent Selection Only in *P. vivax*

The integrated Haplotype Score (iHS) identifies regions of the genome under recent positive selection. The high frequency of a derived allele on an ancestral background can indicate recent positive selection. iHS detects recent positive selection by measuring the amount of extended haplotype homozygosity in the area around the ancestral allele vs the derived allele. We defined the ancestral allele as the consensus allele across all sequences at the site, following the methods used by Assefa et al and Hocking et al in their analyses using iHS to identify *P. knowlesi* genes under recent selection[25,28]. We conducted iHS analyses on all five *P. vivax* subpopulations and the *P. knowlesi* population. Next, we identified the highest iHS $-\log_{10}(P$ -values) in each *P. vivax* gene and its *P. knowlesi* ortholog by selecting which SNP in each *P. vivax* gene, and the SNP in its orthologous *P. knowlesi* gene had the highest iHS $-\log_{10}(P$ value). The SNP with the highest iHS $-\log_{10}(P$ value) in that gene was selected as the $-\log_{10}(P$ value) for the entire gene for this analysis. Most *P. vivax* populations, and the *P. knowlesi* population,

did not see significantly different iHS scores for the set of drug resistance genes or transporter genes compared to the rest of the gene-set in each population (**Fig. 3.3**). However, the set of drug resistance genes in the GMS plus China population had significantly higher iHS scores compared to the population **(p=0.0098, bootstrapping analysis)**. Notably, more *P. knowlesi* genes had higher -$\log_{10}$ (*P* values) compared to *P. vivax*. However, there was no significant difference in the iHS -$\log_{10}$ (*P* value) distribution between any *P. vivax* population and *P. knowlesi* (**p>0.5 for all five pairwise comparisons, Kruskal-Wallis test**).

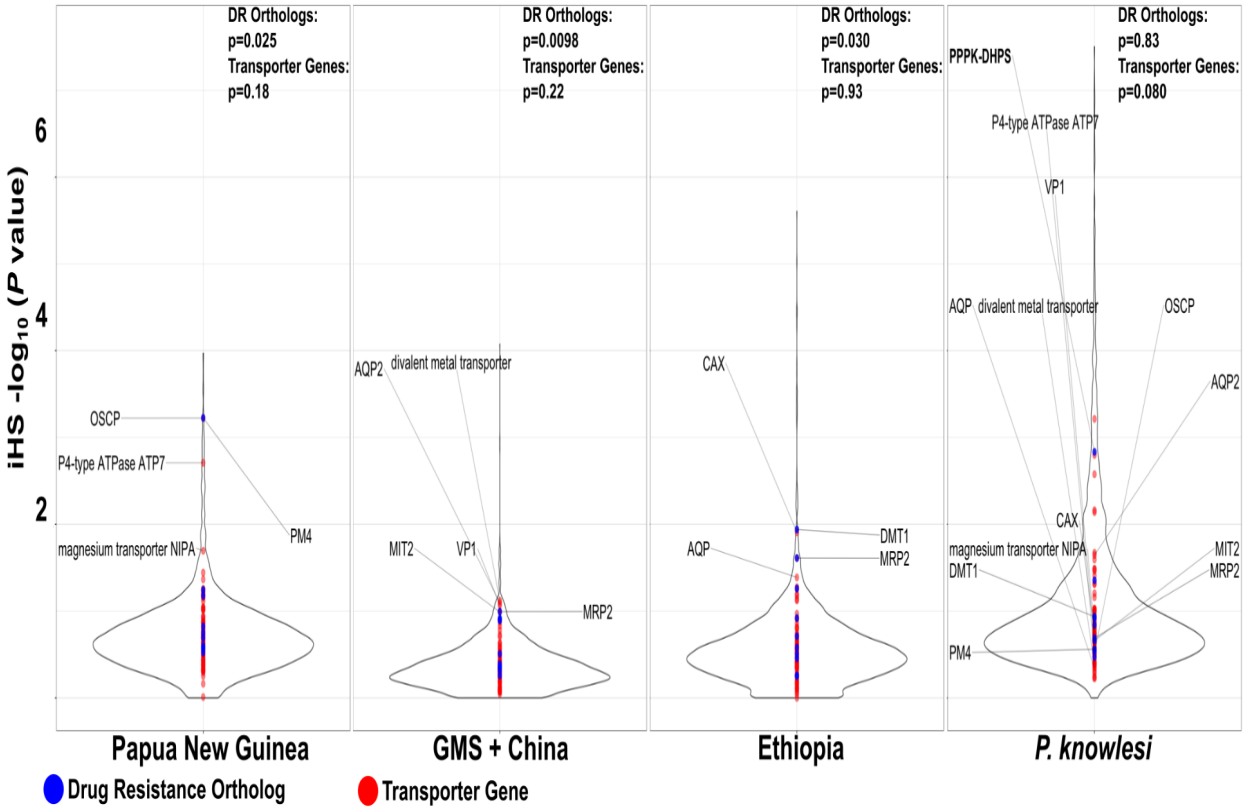


Figure 3.3: Distribution of iHS -$\log_{10}$ (*P* values) for *P. vivax* populations and the *P. knowlesi* population: Displays the distribution of iHS -$\log_{10}$ (*P* values) for *P. vivax* Papua New Guinea, Ethiopia, GMS + China, and *P. knowlesi* populations. Drug resistance genes are highlighted in blue, and transporter genes are highlighted in red. Genes with iHS scores in the top 2% of -$\log_{10}$ (*P* values) in their respective *P. vivax* population are labeled. All genes with iHS scores in the top 2% of -$\log_{10}$ (*P* values) in any of the three displayed *P. vivax* populations are labeled in the *P. knowlesi* distribution.

Several candidate drug resistance genes demonstrated high iHS -$\log_{10}$(*P* values), indicating recent positive selection in certain *P. vivax* populations (**Figure 3.3)**. *plasmepsinIV* (*pvpm4*), the ortholog of the *plasmepsinII* and *plasmepsinIII* genes in *P. falciparum*, had one of the top 10 highest -$\log_{10}$(*P* values) in Papua New Guinea. pfplm2/3 have been implicated in *P. falciparum* resistance to PIP[74]. *pvmdr1* was among the top 10% of -$\log_{10}$(*P* values) in Papua New Guinea **(Figure 3.3)**. *pvdmt1* and *pvmrp2 were* among the top 2%, and *pvmdr2* was among the top 10% -$\log_{10}$(*P* values) in Ethiopia. *pvmrp2* also appeared in the top 2% of -$\log_{10}$(*P* values) in the GMS plus China population. *pvmdr1* and *pvdhfr* had -$\log_{10}$(*P* values) among the top 5% of all GMS plus China genes. **(Table 3.2)**. Very few drug resistance orthologs appeared clear outliers in Peru/Brazil and Colombia **(Supplemental Figure 3.5)**. *pvdhfr* and *pvmdr1* were among the top 5% of -$\log_{10}$(*P* values) in the GMS population. *pvpm4* was among the top 10% of -$\log_{10}$(*P* values) in the Peru/Brazil population. Notably, *pkdhps* in the *P. knowlesi* population had a -$\log_{10}$(*P* value) of 3.54, which was among the top 5% highest iHS -$\log_{10}$(*P* values) in *P. knowlesi*.

Several transporter genes displayed signals of recent selection in *P. vivax* populations, and no signal in the *P. knowlesi* population. PVP01_1411000, which encodes ATP synthase subunit O, and PVP01_0821200, which encodes a P-type ATPase, was among the top 1% of -$\log_{10}$(*P* values) in Papua New Guinea (**Figure 3.3)**. PVP01_1011300, which encodes magnesium transporter NIPA, was among the top 3% of -$\log_{10}$(*P* values) in PNG. PVP01_1145400, which encodes a cation/H+ antiporter (*pvcax*), was also among the top 1% of -$\log_{10}$(*P* values) in Ethiopia (**Figure 3)**. AQP2 and PVP01_1010100, which encodes a divalent metal transporter, were among the top 2% of -$\log_{10}$(*P* values) in the GMS plus China population (**Figure 3.3)**. PVP01_0420400, which encodes a hexose transporter (HT1), PVP01_1266200,

which encodes a member of the major facilitator superfamily (MFS6), and PVP01_1250100, which encodes a V-type H(+)-translocating pyrophosphatase (VP1) were also among the top 1% of -$\log_{10}$ (*P* values) in Peru/Brazil (**Supplemental Figure 3.5)**.

F_{ST} Reveals Candidate Drug Resistance Loci and Transporters Under Heterogeneous Selection Pressure

The F_{ST} statistic compares the frequency of alleles between two populations as a measure of genetic differentiation. F_{ST} statistics closer to one indicate high genetic differentiation, possibly driven by strong selection pressure. F_{ST} statistics near zero suggest free gene flow between those populations. We did not conduct the same analysis between *P. vivax* and *P. knowlesi, because* it would be difficult to determine if F_{ST} statistics closer to one are being driven due to species-specific pressures or simply because they are different species of parasites. *P. vivax* is exposed to heterogeneous drug pressures, meaning different treatment regimens could select for resistant alleles in one *P. vivax* population, but not in others. Selection pressure from drugs could positively select for resistant variants in all populations, but allele frequencies in each population could be different due to differences in N_e. Genes under heterogeneous selection pressure will have higher standard deviations of their F_{ST} statistics across all pairwise population comparisons relative to all *P. vivax* genes. High standard deviations of F_{ST} statistics across several pairwise population comparisons could indicate that a gene is under heterogeneous selection pressure, such as the varying pressure from different antimalarial treatment regimens in different countries. The standard deviation of F_{ST} statistics across population pairwise comparisons may be a useful tool to use to further prioritize genes for validation. We calculated F_{ST} statistics for all pairwise comparisons of the five *P. vivax* populations (n=10 pairwise comparisons) in sliding one kb windows across the entire *P. vivax* genome (n=5,374 genes). We then calculated the standard deviation of F_{ST} statistics for each *P. vivax* gene across all pairwise

comparisons in all windows that contained that gene. The distribution of the standard deviation of F_{ST} statistics for all *P. vivax* genes is displayed in **Figure 3.4.**

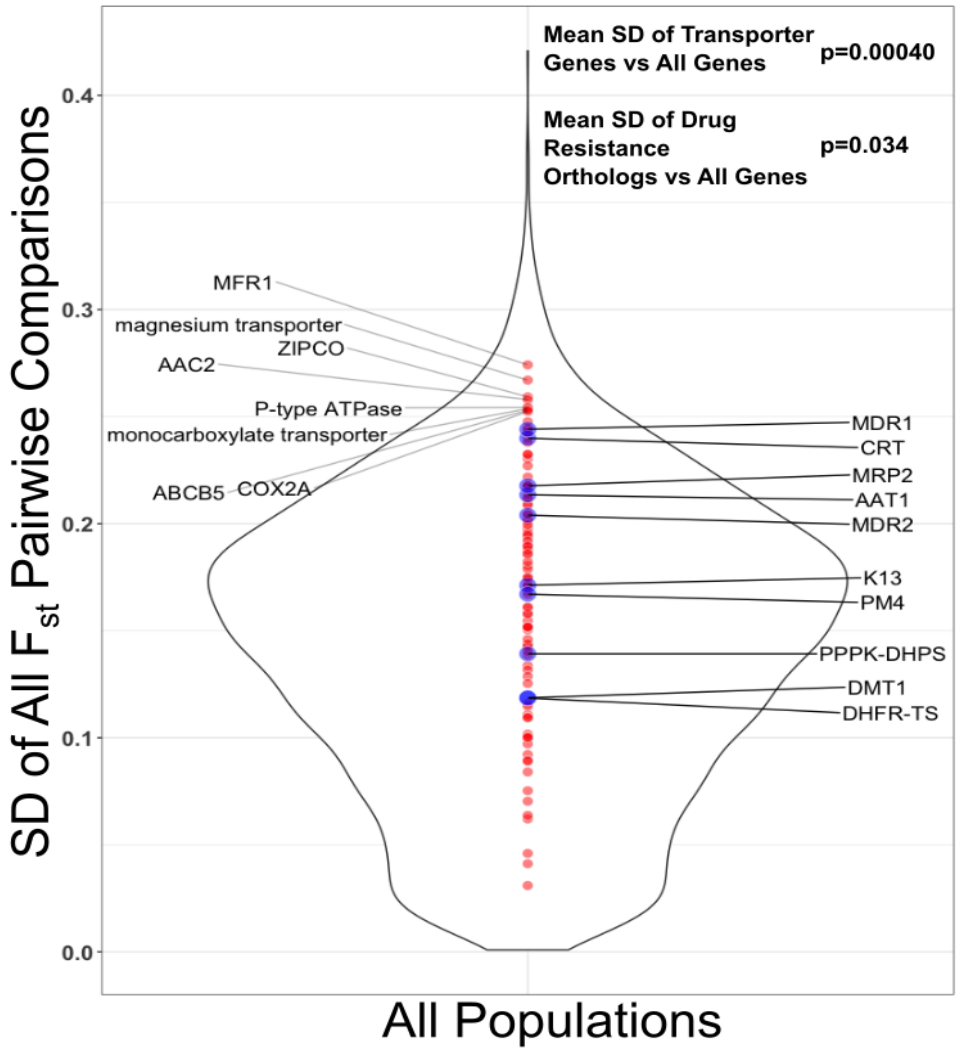


Figure 3.4: Distribution of F_{ST} Standard Deviations (SD) for All 5,374 *P. vivax* Genes. Drug resistance candidates are highlighted in blue and labeled on the right. All transporter genes are highlighted in red, and genes in the top 5% of F_{ST} standard deviations are labeled on the left.

The set of drug resistance candidates did not have significantly greater F_{ST} standard deviations compared to the overall set of genes **(p=0.034, bootstrap analysis)**. However, the *pvmdr1* and *pvcrt* genes had F_{ST} standard deviations in the top 10% of all genes. Both genes are suspected to mediate *P. vivax* resistance because of their well-established drug resistance role in

P. falciparum[6]. Work in Chapter Four of this thesis demonstrates that *pvmdr1* alleles can affect sensitivity to various antimalarials. Copy number variation or increased expression level of *pvcrt* have been linked to *P. vivax* CQ resistant phenotypes. *pvaat1* and pvmrp2 also had F_{ST} standard deviations in the top 20% of genes.

The set of transporter genes had significantly higher F_{ST} standard deviations compared to the rest of the gene set **($p = 4.0 \times 10^{-4}$ bootstrap analysis)**. Several transporter genes had high F_{ST} standard deviations. *pvmrf1*, which encodes the MRF1 gene, had a standard deviation among the top 2% of all F_{ST} standard deviations. MFR1 is a member of the major facilitator superfamily (MFS). *pfmfr3*, another MFS member, can mediate *P. falciparum* multidrug resistance[89]. The genes *pvabcg2, pvzipco, pvaac2, pvcox21*, PVP01_0821200, which encodes a P-type ATPase, PVP01_1239800, which encodes a magnesium transporter, and PVP01_0724900, which encodes a monocarboxylate transporter, all had standard deviations among the top 5% of F_{ST} standard deviations.

Characterizing a Candidate Gene, *plasmepsinIV*, for a Functional Role in *P. vivax* Drug Resistance

Several of the *P. vivax* drug resistance candidate genes were outliers in the three tests of selection. The evidence for each drug resistance candidate that was an outlier in at least one test is summarized in **Table 3.2**. *pvpm4*, one of the candidate drug resistance genes, had one of the top 10 highest iHS $-\log_{10}$ (*P* values) in the Papua New Guinea *P. vivax* population, indicating strong recent selection acting on that gene. *P. vivax* clinical drug resistance to chloroquine in Papua New Guinea is well documented, and there is additional evidence of drug resistance to other antimalarials in the region [5,55,90]. *pvpm4* is a compelling drug candidate for validation because of the evidence of strong recent selection acting on this gene in a region with known *P. vivax* drug resistance. p*vpm4* is an ortholog of the *plasmepsinII* and *plasmepsinIII* genes in *P.*

falciparum. Copy number amplification of *plasmepsinII-III* genes confers resistance to piperaquine[74]. The iHS signal in *pvpm4* could be driven by either copy number amplification or selection on a particular mutation. *pvpm4* is also a compelling candidate, because there is strong evidence of recent selection acting on it in Malaysia, where there is also documented *P. vivax* drug resistance.

Table 3.2: *P. vivax* drug resistance genes with evidence of selection

Gene	DoS-*Pv* (percentile rank)	DoS-*Pk* (percentile rank)	iHS -$\log_{10}$ (*P* values) - *Pv* (percentile rank)	iHS -$\log_{10}$ (*P* values) - *Pk* (percentile rank)	SD of F_{ST} (percentile rank)
pvpm4	0.00385 (76%)	0.117 (68%)	**PNG: 4.0 (100%)**, Ethiopia: 1.40 (86%) GMS+China:0.49 (57%) Peru/Brazil: 1.3 (92%). Colombia:0.65 (37%)	0.700 (23%)	0.167 (56%)
pvmdr1	-0.175 (27%)	0.101 (62%)	PNG: 1.56 (92%), Ethiopia: 0.322 (15%) **GMS+China:1.12 (96%)** Peru/Brazil: 0.888 (73%). Colombia:0.322 (18%)	0.823 (32%)	**0.244 (94%)**
pvdhfr	**0.140 (0.95%)**	0.110 (64%)	PNG: 0.937 (58%), Ethiopia: 0.570 (71%) **GMS+China:1.14 (0.96%)** Peru/Brazil: 0.274 (19%). Colombia:0.561 (31%)	0.678 (21%)	0.119 (31%)

Table 3.2 (Continued): *P. vivax* drug resistance genes with evidence of selection					
pvklech13	**-0.45 (3%)**	0.00260 (18%)	PNG: 0.751 (39%), Ethiopia: 0.579 (40%) GMS+China:0.433 (51%) Peru/Brazil: 0.164 (10%). Colombia:0.378 (18%)	0.588 (15%)	0.17 (60%)
pvmrp2	-0.102 (44%)	0.0321 (30%)	PNG: 1.03 (67%), **Ethiopia: 2.01 (99%) GMS+China:1.24 (98%)** Peru/Brazil: 0.330 (24%). Colombia:1.23 (84%)	0.862 (35%)	0.204 (85%)
pvdmt1	**-0.417 (4%)**	0.190 (85%)	PNG: 0.652 (30%), **Ethiopia: 2.42 (99%)** GMS+China:0.391 (45%) Peru/Brazil: 0.378 (29%). Colombia:1.17 (81%)	1.17 (58%)	0.119 (31%)
pvcrt	**0.156 (95%)**	0.365 (98%)	PNG: 1.47 (91%), Ethiopia: 0.632 (47%) GMS+China:0.634 (75%) Peru/Brazil: 0.349 (26%). Colombia:0.556 (30%)	1.15 (56%)	**0.240 (93%)**
Parentheses indicate percentile rank in that population. Bold cells indicate tests where that gene was an outlier.					

3.5 Discussion:

Comparative Selection Scans Reveal Candidate *P. vivax* Drug Resistance Orthologs and Transporter Genes for Prioritization for *In vitro* Validation

This study demonstrates how comparative selection scans can identify genes under selection pressures unique to *P. vivax* to prioritize for *in vitro* characterization, with a focus on candidate drug resistance genes. The *P. vivax* research community currently prioritizes several *P. vivax* candidate drug resistance genes, simply because they are orthologs of *P. falciparum* drug resistance genes[6]. Orthologs can be a powerful starting point towards identifying candidate resistance genes. However, differences between the biology of *P. vivax* and *P. falciparum* may limit the utility of relying solely on this approach to identify and prioritize *P. vivax* drug resistance genes for research. Comparative evolutionary methods that detect genes under species-specific selection pressures can pinpoint promising drug resistance candidates for further *in vitro* characterization. The drug resistance orthologs genes did not display significantly different test statistics in the DoS test. One explanation for this result is that balancing selection due to drug pressure would need to act on a long timescale to be detected, and *P. vivax* may not have been exposed to drug pressure for a lengthy and consistent enough period of time for this signal to be detected by this test[92]. We also observed few drug resistance orthologs with positive DoS statistics in the *P. vivax* population. The strength of drug selection may also not have been strong enough to fix adaptive substitutions in these genes, which is a possible reason for this observation. However, the drug resistance orthologs displayed significantly higher $-\log_{10}$ (*P* values) in the GMS plus China population and had borderline significantly higher $-\log_{10}$ (*P* values) in Ethiopia and Papua New Guinea, in the iHS test. Concordantly, the drug-resistant orthologs also displayed higher mean F_{ST} standard deviations, though the p-value was just above the significance threshold. The iHS and F_{ST} tests can detect more recently acting selection[33,34,93],

such as from drug pressure, and therefore it was not surprising to observe the synergy in their identification of selection acting on drug resistance genes.

Several individual *P. vivax* drug resistance candidates *pvmdr1*, *pvdhfr*, *pvdmt1*, *pvkelch13, pvmdr2*, and *pvpm4*, were outliers in at least one of the three tests of selection. These drug resistance candidates critically did not display any signals of selection in the *P. knowlesi* population. The *pvdhfr*, and *pvmdr1* genes both have evidence of causing drug resistance in yeast[9] and *P. knowlesi* model systems respectively (shown in Chapter Four of this thesis). This evidence suggests that the comparative selection scans are identifying genuine signals of selection due to drug pressure.

Genes that mediate resistance to antimalarials may differ between *P. vivax* and *P. falciparum*. Identifying other likely drug resistance candidates will be important to inform the prioritization of genes for genetic validation. We chose to include genes that encode transporters for this analysis because of previous work in both *P. falciparum* and pathogenic bacteria, that such genes often mediate drug resistance[40–42]. The transporter gene set displayed significantly different mean F_{ST} standard deviations compared to the mean F_{ST} standard deviation of all *P. vivax* genes. This observation suggests that *P. vivax* transporter genes are subjected to heterogeneous selection pressures, possibly due to their essential roles in transmission, nutrient transport, and possibly drug resistance[94]. The transporter genes did not display significantly different test statistics from other *P. vivax* genes in the iHS and DoS tests. We identified several transporter genes that are compelling candidates for drug resistance characterization since they are uniquely under selection in *P. vivax* populations. *pvabcg2*, and PVP01_0821200, which encodes a P-type ATPase, were outliers in the DoS and F_{ST} and iHS and F_{ST} tests respectively. *pvabcg2* has been found under selection in a previous *P. vivax* population genomic study[95], and

its human ortholog (BCRP) is a known mediator of cancer drug resistance[96]. PVP01_0821200 has also been found under selection in a previous study[97]. However, it should be noted that the selection signal in *pvabcg2* could be from other pressures, such as selection for enhanced transmissibility, rather than drug pressure. Previous work found that *pfabcg2* knockouts in *P. falciparum* led to an increase in the production of gametocytes[98]. Increased production of gametocytes could lead to higher fitness due to the increased likelihood of being transmitted to mosquitoes, which could be advantageous for fitness in low transmission settings[98]. *pvoscp*, which encodes an ATP synthase, had one of the top 10 highest -$\log_{10}$ (*P* values) via the iHS test in Papua New Guinea, suggesting recent positive selection. Additionally, *pvmfr1,* which is a member of the major facilitator superfamily, and PVP01_1239800, which encodes a magnesium transporter, had F_{ST} standard deviations among the top 2% and 3% of all *P. vivax* gene F_{ST} standard deviations respectively. Prioritizing high-confidence candidates for *in vitro* validation will be critical to advance future studies exploring *P. vivax* drug resistance biology. The *P. vivax* drug resistance orthologs and transporter genes with population genomic evidence of selection can be prioritized for functional validation in *P. knowles*i and *P. cynomolgi* model systems. The drug resistance orthologs and transporter genes discussed are all outliers in the tests and warrant further functional characterization either with regards to drug resistance or other phenotypes.

Implications of Signals of Selection in the *P. knowlesi* Population for *P. vivax* Drug Resistance Candidates

We detected signals of selection in the transporter gene set in *P. knowlesi* from the DoS test, and in two drug resistance candidate genes*, pvdhps* and *pvcrt*, in the iHS and DoS tests respectively. The transporter gene set in the *P. knowles*i population had a significantly different DoS mean compared to the rest of the gene set. However, the mean DoS statistic of the transporter gene set and most gene DoS statistics (76/81) in *P. knowlesi* were positive, indicating

that this signal was possibly due to positive selection fixing adaptive substitutions in these genes. Negative DoS statistics could indicate balancing selection that is maintaining drug resistance variants in the population, along with the wild-type allele. The positive mean DoS statistic of the transporter set in *P. knowlesi* indicates the signal is possibly due to selective forces other than drug pressure. *pkdhps* was among the top 5% of $-\log_{10}$(*P* values) in *P. knowlesi*, possibly indicating recent selection on this gene. *dhps* genes in *Plasmodium* parasites are involved in the folate synthesis pathway, which is essential for parasite survival[99]. One possible explanation for this signal of selection could be that the process of *P. knowlesi* adapting to different hosts (human and monkey) selected for *pkdhps* variants to support viability in different host environments. This would imply that *P. knowlesi* is adapting to human populations, and therefore prior observations that this parasite does transmit onward in human populations and the initial hypothesis that the *P. knowles*i genome would not exhibit signs of antimalarial drug selection would be incorrect. If this scenario is true, then *P. knowlesi* would not be the ideal comparator population for selection scans. This hypothesis would warrant further investigation to understand if *P. knowlesi* is becoming a human pathogen, and implications for malaria control and elimination in regions with endemic *P. knowlesi.* However, *dhps* was the only drug resistance candidate gene that was an outlier in either of the selection tests conducted on that population. The drug resistance gene candidate set did not have significantly different test statistic means from the rest of the *P. knowlesi* gene set in either the iHS test or DoS test. This observation suggests that as a group, the drug resistance orthologs are not under selection in *P. knowlesi*. The observation of *dhps* selection in *P. knowles*i could also be a false positive since we did not apply a significance threshold.

pvcrt displayed a signal of adaptive selection from the DoS test in both the *P. vivax* and the *P. knowlesi* populations. Experimental evidence suggests that *pvcrt* possibly mediates chloroquine resistance in *P. vivax via* copy number amplification or increased gene expression[100–102]. It was surprising to observe a strong signal of adaptive selection in both *Plasmodium* species, given the molecular evidence for *pvcrt* possibly mediating *P. vivax* CQR. We would not expect a signal of selection in the *P. knowlesi* population if the signal of selection were driven by drug pressure. DoS statistics for *pvcrt* and *pkcrt* suggest that adaptive substitutions in these genes have been fixed over time. DoS relies solely on polymorphism data in the coding region of the gene. The biology of *crt* and the possible mechanism by which this gene would confer drug resistance could explain the observation of selection in both species, and simultaneously not rule out a drug resistance role for this gene in *P. vivax*. *pvcrt* encodes a transporter that localizes to the membrane of the parasite digestive vacuole[103]. Adaptive substitutions in both species became fixed over time as the parasites adapted to different host populations to facilitate heme digestion. The potential mechanism of *pvcrt* mediated drug resistance suggests increased *pvcrt* copy number or gene expression can result in CQR as suggested by molecular evidence[100–102]. The DoS statistic would not detect selection on copy number variation or changes because the DoS analysis was limited to the coding sequence of each gene and excluded intronic regions, untranslated regions, and promoters. The F_{ST} test includes these regions since it was conducted over one kb non-overlapping sliding windows across the entire genome. The F_{ST} standard deviation scan identified *pvcrt* as one of the *P. vivax* genes with a high F_{ST} standard deviation. Selection pressure from drugs acting on these non-coding regions or copy number variation may explain this signal.

Considerations for Conducting Future Comparative Selection Scans

Many tests of selection rely on polymorphism data as their substrate for analysis. Polymorphism count differences between two species should be considered when choosing which tests to conduct comparative selection scans with and interpreting the results. *P. knowlesi* is more genetically diverse than *P. vivax.* Disparate polymorphism counts between species or populations would affect the density of SNPs in each dataset, and subsequently, the statistical power of the tests to detect true signals of selection in each population[32,104–109] Differences in the effective population size between species could also affect the interpretation of tests statistics. Excess counts of P_N in a population could be evidence of recent selection or due to less efficient purifying selection. We instead chose to describe outliers in the DoS test based on percentile rank, reasoning those genes in the bottom 5% of DoS values would be more likely to have an excess of P_N due to balancing selection, rather than chance alone. We also used the percentile rank approach to describe outliers in our two other tests. The percentile rank approach does not allow us to describe genes as statistically significant outliers. However, the percentile rank approach avoids the dichotomy of categorizing test statistics as "significant" or not" based on the use of significance cutoffs that might otherwise ignore possible signals of selection in both species. A deeper sampling of the *P. vivax* population and greater sequencing depth of *P. vivax* samples can provide higher SNP densities and capture the full extent of *P. vivax* genetic variation to power future comparative power tests of selection.

We chose to limit this analysis to three tests of selection, because several other tests of selection may be redundant to the tests used in this analysis. iHS, along with other commonly used tests of selection, such as Tajima's D or Fay and Wu's H, can detect recent selection and selective sweeps[110,111]. These two tests rely on the site frequency spectrum (SFS) to detect selection[110,111]. By contrast, the iHS test relies on linkage disequilibrium and extended haplotype

homozygosity to infer selection[33,34]. The SFS in *Plasmodium* can be skewed to have an excess of rare variants due to recurrent bottlenecks in the *Plasmodium* life cycle[112,113]. Furthermore, prior work has shown that allele frequencies in *P. falciparum* are more sensitive to selection for increased transmission than within host selection (such as for drug resistance[112]. Parasites that carry alleles selected for their ability to increase transmission probability to mosquitos may also have mutations in drug resistance genes due to chance alone. This scenario would result in founder effects where there are high frequency mutations in candidate drug resistance genes simply because they arose in parasite lineage that also had mutations that enabled them to be more efficiently transmitted. Therefore, mutations in candidate resistance genes at high allele frequency or signals in these genes from a test based on the SFS could lead to erroneous conclusions about an association between mutations in candidate resistance gene and clinical drug resistance. Additionally, the excess of rare variants in *Plasmodium* species due to recurrent bottlenecks in the parasite life cycle can limit the ability to detect positive selection and interpret test statistics if the tests do not take life-cycle history into account[112,113]. Another possible way to identify selection is to use identity by descent to pinpoint regions of the genome under positive selection[114,115]. We only used IBD in this study for sample filtering, but IBD scans could be used in future comparative selection scans to identify regions of the genome that share recent common ancestry, which may indicate positive selection[114,115]. IBD analysis has been used in *P. falciparum* to identify candidate drug resistance loci under selection[115].

Study Limitations, Conclusions, and Future Directions

This study had some limitations. We had limited geographic coverage and sampling depth of the dataset. We subsequently excluded populations with less than 10 parasite samples after quality filtering, which meant we had to exclude the Indian *P. vivax* population. Therefore,

we were not able to investigate the effects of drug pressure in all *P. vivax* populations in the dataset and possibly missed detecting selection acting on certain genes. We would have had more power to detect selection within *P. vivax* subpopulations and the global *P. vivax* population if the dataset had increased sampling depth within the included populations and across a larger geographic range. Additionally, the DoS statistic was likely underpowered to detect selection within the five *P. vivax* subpopulations because our study had uneven sampling numbers for each population, and because of the lack of polymorphism in the global *P. vivax* population relative to the *P. knowlesi* population. We calculated DoS on just the global *P. vivax* population so that we would have enough polymorphism data and subsequent power in the DoS test to have confidence in detecting signals of selection. Pooling all *P. vivax* subpopulations together might mask selection occurring within one population, but not another. However, the paucity of polymorphism data in some of the *P. vivax* subpopulations would mean that the DoS test might not reliably detect genes under selection if applied to just the subpopulations. While using the global *P. vivax* population to calculate the DoS statistic was a conservative approach, we were able to detect selection and avoided the loss of statistical power had we conducted the DoS test just on the *P. vivax* subpopulations. Lastly, it is likely that we missed identifying other possible drug resistance genes because they either mediate drug resistance by a different mechanism of action other than transport function (such as selection on a transcription factor) or because the drug has different targets in *P. vivax* and *P. falciparum*. This analysis suggests that prioritizing genes that are orthologs of known drug resistance genes is a valid approach. Future studies using larger datasets with *P. vivax* samples from other geographic regions would better capture the full extent of selection pressures acting on the *P. vivax* genome. Additionally, the analysis here could be redone with an unbiased approach by prioritizing the genes that are outliers from the

comparative selection scans. This approach would highlight those genes under strong species-specific selection pressure, without prioritizing them based on *a priori* categories. We did not investigate these additional gene categories to focus on conducting a pilot test on the utility of comparative selections to identify candidate drug resistance genes before applying the approach to other categories of genes.

Unfortunately, by the time of writing this thesis, I was not able to generate the laboratory data to determine if *pvpm4* overexpression in *P. knowlesi* mediates reduced susceptibility to several antimalarials (CQ, PIP, LUM, HF, DHA, MDAQ, MQ). I have constructed an overexpression plasmid. I will transfect this construct into *P. knowlesi*, and potentially *P. cynomolgi*, which will episomally over-express *pvpm4*. I will then assay the overexpression *P. knowlesi* and *P. cynomolgi* lines for their drug IC_{50s}. I have also constructed an allelic replacement plasmid that will replace the native *pkpm4* locus with *pvpm4*, and guide RNA plasmids that target the 5' and 3' ends of *pkpm4*. I have constructed a *pvpm4* plasmid that contains the I165V mutation, and I will make a wild-type haplotype without this substitution. I will isolate isogenic clones of these two *pvpm4* lines (*pvpm4*_wt, *pvpmp4*_165V) and assay these two transgenic *P. knowlesi* lines for drug susceptibility. I expect that *pvpm4* will mediate resistance to one of the antimalarial compounds (likely PIP based on the *pvpm4* orthologs of role in PIP-resistance in *P. falciparum*). If I do not observe a reduced drug susceptibility phenotype, that would suggest that selection acting on *pvpm4* is not due to drug pressure and could be due to another environmental pressure.

We present a framework for using comparative selection scans to identify genes uniquely under selection in *P. vivax* in this study and prioritize them for *in vitro* validation. We identified several *P. vivax* candidate genes uniquely under selection in that species. These candidates are

compelling targets for *in vitro* validation and further study with regards to drug resistance. Future comparative selection scans could take an unbiased approach by identifying what *P. vivax* genes are outliers in multiple tests conducted in the *P. vivax* population and not in the comparator population without a predetermined list of genes. Alternatively, one could also curate a list of target genes based on orthology to *P. falciparum* genes with known phenotypes beyond drug resistance (transmissibility, immune evasion, virulence etc.) and identify which of those genes are outliers from these tests. Applying both approaches to conducting comparative selection scans will allow for the prioritization of *P. vivax* candidate genes for functional characterization with regards to drug resistance or other phenotypes.

3.6 References

1. Haldar, K., Bhattacharjee, S. & Safeukui, I. Drug resistance in Plasmodium. *Nat Rev Microbiol* **16**, 156–170 (2018).
2. Price, R. N. *et al.* Global extent of chloroquine-resistant Plasmodium vivax: a systematic review and meta-analysis. *Lancet Infect Dis* **14**, 982–991 (2014).
3. Joy, S., Ghosh, S. K., Achur, R. N., Gowda, D. C. & Surolia, N. Presence of novel triple mutations in the pvdhfr from Plasmodium vivax in Mangaluru city area in the southwestern coastal region of India. *Malar. J.* **17**, 167 (2018).
4. Kaur, S. *et al.* Plasmodium vivax dihydrofolate reductase point mutations from the Indian subcontinent. *Acta Trop* **97**, 174–180 (2006).
5. Hastings, M. D. *et al.* Novel Plasmodium vivax dhfr alleles from the Indonesian Archipelago and Papua New Guinea: association with pyrimethamine resistance determined by a Saccharomyces cerevisiae expression system. *Antimicrob Agents Chemother* **49**, 733–740 (2005).
6. Buyon, L. E., Elsworth, B. & Duraisingh, M. T. The molecular basis of antimalarial drug resistance in Plasmodium vivax. *Int. J. Parasitol. Drugs Drug Resist.* **16**, 23–37 (2021).
7. Suwanarusk *et al.* Amplification of *pvmdr1* associated with multidrug-resistant Plasmodium vivax. *J Infect Dis* **198**, 1558–1564 (2008).
8. Hastings, M. D. *et al.* Dihydrofolate reductase mutations in Plasmodium vivax from Indonesia and therapeutic response to sulfadoxine plus pyrimethamine. *J Infect Dis* **189**, 744–750 (2004).
9. Hastings, M. D. & Sibley, C. H. Pyrimethamine and WR99210 exert opposing selection on dihydrofolate reductase from Plasmodium vivax. *PNAS* **99**, 13137–13141 (2002).
10. Alam, M. T. *et al.* Similar trends of pyrimethamine resistance-associated mutations in Plasmodium vivax and P. falciparum. *Antimicrob Agents Chemother* **51**, 857–863 (2007).
11. Chu, C. S. & White, N. J. The prevention and treatment of Plasmodium vivax malaria. *PLOS Med.* **18**, e1003561 (2021).
12. WHO. *World Malaria Report 2019*. vol. 1 (WHO, 2019).
13. Popovici, J. *et al.* Recrudescence, Reinfection, or Relapse? A More Rigorous Framework to Assess Chloroquine Efficacy for Plasmodium vivax Malaria. *J Infect Dis* **219**, 315–322 (2019).
14. Orjuela-Sánchez, P. *et al.* Analysis of single-nucleotide polymorphisms in the crt-o and mdr1 genes of Plasmodium vivax among chloroquine-resistant isolates from the Brazilian Amazon region. *Antimicrob Agents Chemother* **53**, 3561–3564 (2009).
15. Verzier, L. H., Coyle, R., Singh, S., Sanderson, T. & Rayner, J. C. Plasmodium knowlesi as a model system for characterising Plasmodium vivax drug resistance candidate genes. *PLoS Negl. Trop. Dis.* **13**, e0007470 (2019).
16. Li, J. *et al.* Ex vivo susceptibilities of Plasmodium vivax isolates from the China-Myanmar border to antimalarial drugs and association with polymorphisms in *Pvmdr1* and Pvcrt-o genes. *PLoS Negl. Trop. Dis.* **14**, e0008255 (2020).
17. Suwanarusk *et al.* Chloroquine Resistant Plasmodium vivax: In Vitro Characterisation and Association with Molecular Polymorphisms. *PLOS ONE* **2**, e1089 (2007).
18. Sá *et al.* Plasmodium vivax: allele variants of the mdr1 gene do not associate with chloroquine resistance among isolates from Brazil, Papua, and monkey-adapted strains. *Exp Parasitol* **109**, 256–259 (2005).
19. Marfurt, J. *et al.* Molecular markers of in vivo Plasmodium vivax resistance to amodiaquine plus sulfadoxine-pyrimethamine: mutations in pvdhfr and *pvmdr1*. *J Infect Dis* **198**, 409–417 (2008).
20. Carlton, J. M. *et al.* Comparative genomics of the neglected human malaria parasite Plasmodium vivax. *Nature* **455**, 757–763 (2008).

21. Llinás, M., Bozdech, Z., Wong, E. D., Adai, A. T. & DeRisi, J. L. Comparative whole genome transcriptome analysis of three Plasmodium falciparum strains. *Nucleic Acids Res.* **34**, 1166–1173 (2006).
22. Thompson, J., Janse, C. J. & Waters, A. P. Comparative genomics in Plasmodium: a tool for the identification of genes and functional analysis. *Mol. Biochem. Parasitol.* **118**, 147–154 (2001).
23. Lawrie, D. S. & Petrov, D. A. Comparative population genomics: power and principles for the inference of functionality. *Trends Genet.* **30**, 133–139 (2014).
24. Coleman, M. L. & Chisholm, S. W. Ecosystem-specific selection pressures revealed through comparative population genomics. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 18634–18639 (2010).
25. Assefa, S. *et al.* Population genomic structure and adaptation in the zoonotic malaria parasite Plasmodium knowlesi. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 13027–13032 (2015).
26. Grigg, M. J. *et al.* Individual-level factors associated with the risk of acquiring human Plasmodium knowlesi malaria in Malaysia: a case-control study. *Lancet Planet. Health* **1**, e97–e104 (2017).
27. Singh, B. & Daneshvar, C. Human Infections and Detection of Plasmodium knowlesi. *Clin. Microbiol. Rev.* **26**, 165–184 (2013).
28. Hocking, S. E., Divis, P. C. S., Kadir, K. A., Singh, B. & Conway, D. J. Population Genomic Structure and Recent Evolution of Plasmodium knowlesi, Peninsular Malaysia. *Emerg. Infect. Dis.* **26**, 1749–1758 (2020).
29. Richards, J. & Mueller, I. Identifying the risks for human transmission of Plasmodium knowlesi. *Lancet Planet. Health* **1**, e83–e85 (2017).
30. Grüring, C. *et al.* Human red blood cell-adapted Plasmodium knowlesi parasites: a new model system for malaria research. *Cell Microbiol* **16**, 612–620 (2014).
31. Moon, R. W. *et al.* Adaptation of the genetically tractable malaria pathogen Plasmodium knowlesi to continuous culture in human erythrocytes. *Proc Natl Acad Sci U A* **110**, 531–536 (2013).
32. Stoletzki, N. & Eyre-Walker, A. Estimation of the neutrality index. *Mol. Biol. Evol.* **28**, 63–70 (2011).
33. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A Map of Recent Positive Selection in the Human Genome. *PLoS Biol.* **4**, e72 (2006).
34. Sabeti, P. C. *et al.* Positive natural selection in the human lineage. *Science* **312**, 1614–1620 (2006).
35. Wright, S. The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating. *Evolution* **19**, 395–420 (1965).
36. Hupalo, D. N. *et al.* Population genomics studies identify signatures of global dispersal and drug resistance in Plasmodium vivax. *Nat Genet* **48**, 953–958 (2016).
37. Pearson, R. D. *et al.* Genomic analysis of local variation and recent evolution in Plasmodium vivax. *Nat Genet* **48**, 959–964 (2016).
38. Aurrecoechea, C. *et al.* PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res.* **37**, D539–D543 (2009).
39. Choi, C.-H. ABC transporters as multidrug resistance mechanisms and the development of chemosensitizers for their reversal. *Cancer Cell Int.* **5**, 30 (2005).
40. Martin, R. E., Henry, R. I., Abbey, J. L., Clements, J. D. & Kirk, K. The 'permeome' of the malaria parasite: an overview of the membrane transport proteins of Plasmodium falciparum. *Genome Biol.* **6**, R26 (2005).

41. Putman, M., van Veen, H. W. & Konings, W. N. Molecular Properties of Bacterial Multidrug Transporters. *Microbiol. Mol. Biol. Rev.* **64**, 672–693 (2000).
42. Sanchez, C. P., Dave, A., Stein, W. D. & Lanzer, M. Transporters as mediators of drug resistance in Plasmodium falciparum. *Int. J. Parasitol.* **40**, 1109–1118 (2010).
43. Zhu, L. *et al.* New insights into the Plasmodium vivax transcriptome using RNA-Seq. *Sci. Rep.* **6**, 20498 (2016).
44. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
45. Auburn, S. *et al.* A new Plasmodium vivax reference sequence with improved assembly of the subtelomeres reveals an abundance of pir genes. *Wellcome Open Res.* **1**, 4 (2016).
46. Lapp, S. A. *et al.* PacBio assembly of a Plasmodium knowlesi genome sequence with Hi-C correction and manual annotation of the SICAvar gene family. *Parasitology* **145**, 71–84 (2018).
47. Broad Institute. *PicardTools*. (Broad Institute).
48. Scaling accurate genetic variant discovery to tens of thousands of samples | bioRxiv. https://www.biorxiv.org/content/10.1101/201178v2.
49. Auburn, S. *et al.* Characterization of Within-Host Plasmodium falciparum Diversity Using Next-Generation Sequence Data. *PLOS ONE* **7**, e32891 (2012).
50. Schaffner, S. F., Taylor, A. R., Wong, W., Wirth, D. F. & Neafsey, D. E. hmmIBD: software to infer pairwise identity by descent between haploid genotypes. *Malar J* **17**, 196 (2018).
51. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
52. Weir, B. S. & Cockerham, C. C. Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**, 1358–1370 (1984).
53. Gautier, M., Klassmann, A. & Vitalis, R. rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Mol. Ecol. Resour.* **17**, 78–90 (2017).
54. Wickham, Hadley. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).
55. Pulford, J., Mueller, I., Siba, P. M. & Hetzel, M. W. Malaria case management in Papua New Guinea prior to the introduction of a revised treatment protocol. *Malar. J.* **11**, 157 (2012).
56. Griffing, S. M., Tauil, P. L., Udhayakumar, V. & Silva-Flannery, L. A historical perspective on malaria control in Brazil. *Mem. Inst. Oswaldo Cruz* **110**, 701–718 (2015).
57. Griffing, S. M., Gamboa, D. & Udhayakumar, V. The history of 20th century malaria control in Peru. *Malar. J.* **12**, 303 (2013).
58. Rodríguez, J. C. P., Uribe, G. Á., Araújo, R. M., Narváez, P. C. & Valencia, S. H. Epidemiology and control of malaria in Colombia. *Mem. Inst. Oswaldo Cruz* **106 Suppl 1**, 114–122 (2011).
59. Cui, L. *et al.* Malaria in the Greater Mekong Subregion: heterogeneity and complexity. *Acta Trop.* **121**, 227–239 (2012).
60. Chen, C. Development of antimalarial drugs and their application in China: a historical review. *Infect. Dis. Poverty* **3**, 9 (2014).
61. WHO. *Guidelines for the treatment of malaria - Third edition*. (2018).
62. Silachamroon, U., Krudsood, S., Phophak, N. & Looareesuwan, S. Management of malaria in Thailand. *Korean J. Parasitol.* **40**, 1–7 (2002).
63. Commons, R. J. *et al.* The efficacy of dihydroartemisinin-piperaquine and artemether-lumefantrine with and without primaquine on Plasmodium vivax recurrence: A systematic review and individual patient data meta-analysis. *PLOS Med.* **16**, e1002928 (2019).
64. Goldlust, S. M. *et al.* The decline of malaria in Vietnam, 1991–2014. *Malar. J.* **17**, 226 (2018).

65. Hung, L. Q. *et al.* Control of malaria: a successful experience from Viet Nam. *Bull. World Health Organ.* **80**, 660–666 (2002).
66. Ariey, F. *et al.* A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. *Nature* **505**, 50–55 (2014).
67. Sidhu, A. B. S., Valderramos, S. G. & Fidock, D. A. pfmdr1 mutations contribute to quinine resistance and enhance mefloquine and artemisinin sensitivity in Plasmodium falciparum. *Mol. Microbiol.* **57**, 913–926 (2005).
68. Sidhu, A. B. S., Verdier-Pinard, D. & Fidock, D. A. Chloroquine resistance in Plasmodium falciparum malaria parasites conferred by pfcrt mutations. *Science* **298**, 210–213 (2002).
69. Sidhu, A. B. S. *et al.* Decreasing pfmdr1 Copy Number in Plasmodium falciparum Malaria Heightens Susceptibility to Mefloquine, Lumefantrine, Halofantrine, Quinine, and Artemisinin. *J. Infect. Dis.* **194**, 528–535 (2006).
70. Djimdé, A. *et al.* A molecular marker for chloroquine-resistant falciparum malaria. *N. Engl. J. Med.* **344**, 257–263 (2001).
71. Sá, J. M. *et al.* Geographic patterns of Plasmodium falciparum drug resistance distinguished by differential responses to amodiaquine and chloroquine. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 18883–18889 (2009).
72. Amato, R. *et al.* Genetic markers associated with dihydroartemisinin–piperaquine failure in Plasmodium falciparum malaria in Cambodia: a genotype-phenotype association study. *Lancet Infect. Dis.* **17**, 164–173 (2017).
73. Witkowski, B. *et al.* A surrogate marker of piperaquine-resistant Plasmodium falciparum malaria: a phenotype–genotype association study. *Lancet Infect. Dis.* **17**, 174–183 (2017).
74. Bopp, S. *et al.* Plasmepsin II–III copy number accounts for bimodal piperaquine resistance among Cambodian Plasmodium falciparum. *Nat. Commun.* **9**, 1769 (2018).
75. Dhingra, S. K. *et al.* A Variant PfCRT Isoform Can Contribute to Plasmodium falciparum Resistance to the First-Line Partner Drug Piperaquine. *mBio* **8**, e00303-17 (2017).
76. Agrawal, S. *et al.* Association of a Novel Mutation in the Plasmodium falciparum Chloroquine Resistance Transporter With Decreased Piperaquine Sensitivity. *J. Infect. Dis.* **216**, 468–476 (2017).
77. Madamet, M. *et al.* The Plasmodium falciparum chloroquine resistance transporter is associated with the ex vivo *P. falciparum* African parasite response to pyronaridine. *Parasit. Vectors* **9**, 77 (2016).
78. Cooper, R. A. *et al.* Mutations in transmembrane domains 1, 4 and 9 of the Plasmodium falciparum chloroquine resistance transporter alter susceptibility to chloroquine, quinine and quinidine. *Mol. Microbiol.* **63**, 270–282 (2007).
79. Sisowath, C. *et al.* In vivo selection of Plasmodium falciparum pfmdr1 86N coding alleles by artemether-lumefantrine (Coartem). *J. Infect. Dis.* **191**, 1014–1017 (2005).
80. Sisowath, C. *et al.* The role of pfmdr1 in Plasmodium falciparum tolerance to artemether-lumefantrine in Africa. *Trop. Med. Int. Health TM IH* **12**, 736–742 (2007).
81. Price, R. N. *et al.* Mefloquine resistance in Plasmodium falciparum and increased pfmdr1 gene copy number. *Lancet Lond. Engl.* **364**, 438–447 (2004).
82. Uhlemann, A. C. & Krishna, S. Antimalarial multi-drug resistance in Asia: mechanisms and assessment. *Curr. Top. Microbiol. Immunol.* **295**, 39–53 (2005).
83. Parzy, D. *et al.* Proguanil resistance in Plasmodium falciparum African isolates: assessment by mutation-specific polymerase chain reaction and in vitro susceptibility testing. *Am. J. Trop. Med. Hyg.* **57**, 646–650 (1997).

84. Gil, J. P. *et al.* Detection of atovaquone and Malarone resistance conferring mutations in Plasmodium falciparum cytochrome b gene (cytb). *Mol. Cell. Probes* **17**, 85–89 (2003).
85. Luo, Z., Sullivan, S. A. & Carlton, J. M. The biology of Plasmodium vivax explored through genomics. *Ann. N. Y. Acad. Sci.* **1342**, 53–61 (2015).
86. McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* **351**, 652–654 (1991).
87. Rand, D. M. & Kann, L. M. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. *Mol. Biol. Evol.* **13**, 735–748 (1996).
88. Loy, D. E. *et al.* Out of Africa: origins and evolution of the human malaria parasites Plasmodium falciparum and Plasmodium vivax. *Int. J. Parasitol.* **47**, 87–97 (2017).
89. Rocamora, F. *et al.* PfMFR3: A Multidrug-Resistant Modulator in Plasmodium falciparum. *ACS Infect. Dis.* **7**, 811–825 (2021).
90. Asih, P. B. S. *et al.* Phenotyping clinical resistance to chloroquine in Plasmodium vivax in northeastern Papua, Indonesia. *Int J Parasitol Drugs Drug Resist* **1**, 28–32 (2011).
91. Auburn, S. *et al.* Genomic analysis of a pre-elimination Malaysian Plasmodium vivax population reveals selective pressures and changing transmission dynamics. *Nat. Commun.* **9**, 2585 (2018).
92. Siewert, K. M. & Voight, B. F. Detecting Long-Term Balancing Selection Using Allele Frequency Correlation. *Mol. Biol. Evol.* **34**, 2996–3005 (2017).
93. Pickrell, J. K. *et al.* Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* **19**, 826–837 (2009).
94. Martin, R. E. The transportome of the malaria parasite. *Biol. Rev.* **95**, 305–332 (2020).
95. Benavente, E. D. *et al.* Distinctive genetic structure and selection patterns in Plasmodium vivax from South Asia and East Africa. *Nat. Commun.* **12**, 3160 (2021).
96. Natarajan, K., Xie, Y., Baer, M. R. & Ross, D. D. Role of Breast Cancer Resistance Protein (BCRP/ABCG2) in Cancer Drug Resistance. *Biochem. Pharmacol.* **83**, 1084–1103 (2012).
97. Brashear, A. M. *et al.* Population genomics identifies a distinct Plasmodium vivax population on the China-Myanmar border of Southeast Asia. *PLoS Negl. Trop. Dis.* **14**, e0008506 (2020).
98. Tran, P. N. *et al.* A female gametocyte-specific ABC transporter plays a role in lipid metabolism in the malaria parasite. *Nat. Commun.* **5**, 4773 (2014).
99. Hyde, J. E. Exploring the folate pathway in Plasmodium falciparum. *Acta Trop.* **94**, 191–206 (2005).
100. Sá, J. M. *et al.* Plasmodium vivax chloroquine resistance links to pvcrt transcription in a genetic cross. *Nat. Commun.* **10**, 1–10 (2019).
101. Sà, J. M., Cannon, M. V., Caleon, R. L., Wellems, T. E. & Serre, D. Single-cell transcription analysis of Plasmodium vivax blood-stage parasites identifies stage- and species-specific profiles of expression. *PLOS Biol.* **18**, e3000711 (2020).
102. Melo, G. C. *et al.* Expression levels of pvcrt-o and pvmdr-1 are associated with chloroquine resistance and severe Plasmodium vivax malaria in patients of the Brazilian Amazon. *PLoS ONE* **9**, e105922 (2014).
103. Fidock, D. A. *et al.* Mutations in the *P. falciparum* Digestive Vacuole Transmembrane Protein PfCRT and Evidence for Their Role in Chloroquine Resistance. *Mol Cell* **6**, 861–871 (2000).
104. Zhai, W., Nielsen, R. & Slatkin, M. An Investigation of the Statistical Power of Neutrality Tests Based on Comparative and Population Genetic Data. *Mol. Biol. Evol.* **26**, 273–283 (2009).

105. Willing, E.-M., Dreyer, C. & Oosterhout, C. van. Estimates of Genetic Differentiation Measured by FST Do Not Necessarily Require Large Sample Sizes When Using Many SNP Markers. *PLOS ONE* **7**, e42649 (2012).
106. Liu, X., Saw, W.-Y., Ali, M., Ong, R. T.-H. & Teo, Y.-Y. Evaluating the possibility of detecting evidence of positive selection across Asia with sparse genotype data from the HUGO Pan-Asian SNP Consortium. *BMC Genomics* **15**, 332 (2014).
107. Holsinger, K. E. & Weir, B. S. Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nat. Rev. Genet.* **10**, 639–650 (2009).
108. Morin, P. A., Martien, K. K. & Taylor, B. L. Assessing statistical power of SNPs for population structure and conservation studies. *Mol. Ecol. Resour.* **9**, 66–73 (2009).
109. Kalinowski, S. T. Do polymorphic loci require large sample sizes to estimate genetic distances? *Heredity* **94**, 33–36 (2005).
110. Fay, J. C. & Wu, C. I. Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405–1413 (2000).
111. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
112. Chang, H.-H. & Hartl, D. L. Recurrent bottlenecks in the malaria life cycle obscure signals of positive selection. *Parasitology* **142 Suppl 1**, S98–S107 (2015).
113. Chang, H.-H. *et al.* Malaria life cycle intensifies both natural selection and random genetic drift. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 20129–20134 (2013).
114. Albrechtsen, A., Moltke, I. & Nielsen, R. Natural Selection and the Distribution of Identity-by-Descent in the Human Genome. *Genetics* **186**, 295–308 (2010).
115. Henden, L., Lee, S., Mueller, I., Barry, A. & Bahlo, M. Identity-by-descent analyses for measuring population dynamics and selection in recombining pathogens. *PLOS Genet.* **14**, e1007279 (2018).

Chapter Four: Functional Evidence That *Plasmodium vivax mdr1* Haplotypes Confer Multidrug Resistance

4.1 Abstract:

P. vivax is the most globally distributed malaria parasite species in humans, but important aspects of its biology remain unknown. Chloroquine is the main frontline drug for *P. vivax* treatment. *P. falciparum* Chloroquine resistance (CQR) has spread worldwide. Unfortunately, there is growing evidence that *P. vivax* is also developing resistance to CQR. Understanding of *P. vivax* drug resistance has lagged that of *P. falciparum,* in part because of the lack of a continuous *in vitro* culture system for *P. vivax*. *P. vivax* CQR is thought to be mediated by gene variants encoding the multidrug resistance protein (*pvmdr1*), an orthologue of the known *P. falciparum* resistance gene *pfmdr1*. Despite years of research focusing on *pvmdr1,* no definitive relationship between *pvmdr1* polymorphisms and drug resistance has emerged. *P. knowlesi*, a zoonotic macaque parasite closely related to *P. vivax*, can be cultured, and used as a model system to characterize *P. vivax* genes. We analyzed a *P. vivax* population genomic dataset to identify 23 geographically representative circulating *pvmdr1* polymorphisms plus four additional haplotypes. We then expressed these *pvmdr1* mutants in *P. knowlesi* and assayed them for resistance to various antimalarial compounds. We found that mutations and increased copy number of *pvmdr1* confer resistance to mefloquine, dihydroartemisinin, and lumefantrine. Our results suggest that while *pvmdr1* mutations do not play a role in resistance to chloroquine, they do confer resistance to several other antimalarials, suggesting co-selection of *pvmdr1* from antimalarials used to treat *P. falciparum.* Our results have implications for guiding *P. vivax* treatment policy and molecular surveillance of *P. vivax* drug resistance alleles.

4.2 Introduction:

Malaria caused by *Plasmodium vivax* infection is a major public health issue. Over two-thirds of the world's population is at risk of *P. vivax* infection, causing an estimated 4.5 million

cases in 2020.[1] Increasingly, *P. vivax* is understood as a significant contributor to severe malaria and other adverse malaria-related clinical outcomes[2–4]. *P. vivax* is the most globally distributed malaria parasite species in humans, but important aspects of its biology remain unknown. In particular, knowledge regarding the molecular basis of *P. vivax* drug resistance lags behind that of *Plasmodium falciparum,* the other major human malaria-causing parasite.

The WHO recommends that *P. vivax* is treated with a combination of Chloroquine (CQ) and Primaquine (PQ) in most regions[5]. *P. vivax* differs from *P. falciparum* malaria in its life cycle by its hypnozoite stage, where a subset of parasites in the liver can remain dormant for up to five years after initial infection[2,6]. Primaquine and Tafenoquine (TQ) are the only drugs that can kill hypnozoite stage parasites in the liver[7]. However, PQ and TQ can cause severe hemolysis in patients with G6PD deficiency, precluding its use in certain areas or requiring genetic testing before treatment administration[8]. *P. falciparum* chloroquine resistance (CQR) has spread worldwide, but CQ remains a viable treatment option for *P. vivax*. Worryingly, there is clear evidence that *P. vivax* has developed resistance to CQ in Papua New Guinea, Indonesia, and Malaysia.[8–10]. There is also growing evidence of CQR *P. vivax* spreading in other regions[8–10]. Chloroquine-resistant *P. vivax* has generally been defined as recurrence of *P. vivax* infection by day 28 of treatment, ideally with confirmed presence of whole-blood chloroquine concentrations of greater than 100 nM[8]. Because there is no *in vitro* continuous culture system for *P. vivax*, chemo-genomic approaches currently cannot be used to uncover drug resistance mechanisms[8,11]. CQR *P. vivax* is thought to have emerged in Papua New Guinea and has been reported in other countries[8,12.] Tracking CQ resistance is difficult, because there are no validated molecular markers of CQR[8,12–14].

P. vivax CQR is thought to be mediated by gene variants encoding the multidrug resistance protein (*pvmdr1*), and the chloroquine resistance transporter (*pvcrt*). These genes are orthologs of known drug resistance genes *pfmdr1* and *pfcrt* in *P. falciparum*[8,13] which both encode digestive vacuole transmembrane proteins[15]. *Plasmodium* parasites metabolize hemoglobin, resulting in a buildup of heme in the digestive vacuole that must be detoxified, as the buildup of heme is toxic to the parasite[15–17]. *P. falciparum* (and other *Plasmodium* parasites) convert heme into a non-toxic form, hemozoin to prevent this buildup.[15] Chloroquine inhibits the conversion of heme to hemozoin, causing a toxic buildup of heme and subsequent death of the parasite[15]. The *P. falciparum* genes *pfmdr1* and *pfcrt* encode transporter proteins that respectively pump compounds (such as CQ) in or out of the digestive vacuole[15]. Mutant *pfmdr1* and *pfcrt* alleles are thought to encode changes that either reduce shuttling of the drug into the vacuole or pumping the drug out, causing resistance[15,18]. *pfmdr1* and *pfcrt* play a role in resistance to lumefantrine (LUM) and Mefloquine (MQ) by sequestering these drugs in the digestive vacuole away from their cytoplasmic targets[15]. *pvmdr1* and *pvcrt* are also believed to mediate CQR in *P. vivax*, yet the causal mutations responsible for CQR remain to be elucidated and validated [8,19,20]. Six *pvmdr1* SNPs have been reported in multiple studies at high frequency in regions with reported drug resistance; (relative to the SAL-1 reference) S513R, G698S, M908L, T958M, Y976F, and F1076L[11,21–24]. *Ex vivo P. vivax* assays have found associations between specific *pvmdr1* polymorphisms (Y976F and F1076L) and reduced susceptibility to chloroquine in some, but not in all studies[11,25–29]. These two SNPs are also found in regions without reported CQR, suggesting that they may not in fact mediate CQR, or their ability to do so depends on the genetic background they arise on[30,31]. The T958M, Y976F, F1076L variants (and possibly others) in *pvmdr1* have been shown to arise independently, even within the same population[30]. This observation suggests that

pvmdr1 mutations can arise on different genetic backgrounds and that they have not been subject to a recent selective sweep, such as due to CQ treatment[30]. A genetic cross performed in Aotus monkeys implicated a region on chromosome one that contains *pvcrt* with CQR phenotype[14]. It is unclear whether these discrepancies between studies represent regional differences in parasite diversity, genetic background, and drug selection history or are due to technical variations. However, many *pvmdr1* polymorphisms are being used in molecular epidemiological studies as drug resistance markers without any firm validation that these polymorphisms confer CQR or resistance to other compounds[26,32,33]. Using unvalidated polymorphisms as markers of *P. vivax r*esistance could result in unwarranted changes in national treatment policy. This scenario could lead to unnecessary use of more expensive treatments or switching to a treatment that is already ineffective due to other resistance variants in the parasite population[34].

Reverse genetics is a powerful method to validate molecular changes in drug resistance genes for their capacity of reducing susceptibility to various antimalarials. This has been potently demonstrated for *P. falciparum*, which possesses robust *in vitro* culture, drug assays, and powerful genetics for the functional characterization of the *pfmdr1* and *pfcrt* genes and for the validation of drug targets of novel antimalarials identified in chemical genomics studies[35–40]. A continuous *in vitro* culture system for *P. vivax* does not currently exist, greatly hampering studies of the parasite[41]. Significant improvements have been made to *P. vivax ex vivo* drug susceptibility assays[41]. However, they remain limited to short single cycle assays with relatively poor growth compared to *P. falciparum*[41]. *Ex vivo* drug susceptibility assays also require patient isolates which cannot be well controlled for parasite stage and parasitemia, thereby limiting their accuracy[41]. Improved *ex vivo* assays will significantly improve the ability to determine drug susceptibility in natural isolates and link drug phenotypes to specific resistance loci. Reverse genetic approaches

will be critical to validate candidate drug loci. *In lieu* of a stable culture system, heterologous genetic systems are invaluable for validation studies in *P. vivax*. Prior work using *P. knowlesi* and yeast model systems has validated drug resistance genes and characterized host invasion factors [42–44].

P. knowlesi, a zoonotic macaque parasite, can be continuously cultured *in vitro*. Our group and others adapted *P. knowlesi* culture to efficiently proliferate in human red blood cells[4,35]. *P. knowlesi* is closely related to *P. vivax* phylogenetically, and is genetically tractable, with high transfection efficiencies[42]. Previous work has demonstrated that accurate determinations of *P. knowlesi* drug susceptibility can be conducted[47]. The development of *P. knowlesi* as an *in vitro* system allows for molecular genetic experiments characterizing *P. vivax pvmdr1* polymorphisms with regards to drug resistance. *P. knowlesi* has previously been used to episomally express *pvmdr1* variants[13]. This study did not find an association between the two *pvmdr1* haplotypes tested and reduced drug susceptibility, but it supported the use of *P. knowlesi* as a model system to characterize *P. vivax* candidate drug resistance genes[13].

We used a *P. vivax* population genomics dataset to identify common *pvmdr1* haplotypes worldwide. Our population genomics analysis identified 22 (n=22) circulating *pvmdr1* haplotypes composed of 10 SNPs. We also constructed four parasite lines (n=4) containing individual *pvmdr1* SNPs commonly reported in the literature in the Sal-1 *P. vivax* reference background, and one *pvmdr1* haplotype (n=1) tested in Verzier et al[13] that was not in our population genomic dataset and four *P. falcpairum mdr1* mutations that have been implicated in drug resistance in that species. We then performed an allelic replacement of the native *P. knowlesi pkmdr1* locus with our 27 *pvmdr1* haplotypes and performed drug susceptibility assays. We found that mutations in *pvmdr1* significantly reduced susceptibility to Mefloquine (MQ), Dihydroartemisinin (DHA),

Lumefantrine (LUM), and Halofantrine (HF). We also found that while *pvmdr1* polymorphisms do not lead to high-grade CQ resistance, we observed both increased and decreased susceptibility to CQ. We additionally tested the effect of *pvmdr1* overexpression on drug susceptibility, finding that it confers resistance to DHA, MQ, LUM, and HF. Our findings have significant implications for the molecular surveillance of *P. vivax* drug resistance and treatment policy.

4.3 Methods

Data accession

P. vivax data were accessed from the NCBI Sequencing Read Archive with the following accession numbers listed in Hupalo et al, 2016[48]. We also accessed *P. vivax* samples from the European Nucleotide Archive using accession numbers from Pearson et al, 2016[49].

Sample Alignment and Variant Calling

Sample reads were aligned in parallel to the PO1 reference strain using BWA-mem[50]. Duplicate reads were marked using the *MarkDuplicates* tool from Picard tools[51]. We performed local realignment around indels using GATK RealignerTargetCreator and GATK IndelRealigner [52]. Variants were called in the accessible regions using GATK-haplotype caller (GATK Version 3.5.0) using best practices to filter and identify SNPs and generate gVCFs for each sample. GATKs GenotypeGVCFs tool was next used to construct a joint VCF with these samples. We repeated this with several WGS samples stored locally at the Broad Institute. We called variants against PO1 using GATK Haplotype caller, generating a joint VCF with 339 samples and 23,827,021 sites. We used VCFtools[53] to subset our master VCF into smaller ones containing just *pvmdr1* and *pvcrt*. This filtering resulted in a final VCF of 339 samples and 4333 sites for *pvmdr1*.

We excluded SNPs with a minor allele frequency of less than 5%. 10 SNPs remained after filtering: V221L, D500N, S513R, G698S, L845F, M908L, T958M, Y976F, F1076L, and K1393N.

We subsetted the VCF to just the *pvmdr1* coding region to analyze the haplotypes containing the 10 SNPs of interest. We visualized haplotypes in these genes in R software (R version 3.6.1) using custom scripts and the packages "vcfr"[54] and "ggplot2"[55]. We identified all unique combinations of these 10 SNPs to determine circulating haplotypes using a custom script. We identified 22 unique haplotypes containing various combinations of these 10 SNPs in our dataset.

Additional *pvmdr1* Haplotype Identification

We selected the SNPs S513R, G698S, Y976F, and L1076F to be characterized as single mutations in the Sal-1 *pvmdr1* genomic background since they are commonly reported in the literature and were not present in the Sal-1 background as an existing haplotype in our dataset. We also sought to characterize *pfmdr1* mutations that confer reduced drug susceptibility *in vitro*. We identified SNPs 86Y, 184F, 1042D, and 1246Y from the *pfmdr1* literature to test. We aligned the *pfmdr1* and *pvmdr1* sequences and mapped where these four SNPs aligned in *pvmdr1*. 86Y mapped to position 91, 184F mapped to position 189, 1042D mapped to position 1079, and 1246Y mapped to position 1291 in *pvmdr1*. We then designed primers to introduce these mutations in the ancestral background.

Ancestral Allele Polarization

To infer an ancestral *pvmdr1* haplotype we first downloaded the *mdr1* genomic sequences from *P. vivax*, *P. knowlesi, P. coatneyi*, *P. inui*, and *P. cynomolgi* from the PlasmoDB webserver (Release 56, https://plasmodb.org/plasmo/app)[56]. We then aligned these 5 sequences using the MUSCLE program[57]. We selected the amino acid that was present in the majority of the species as the ancestral SNP for each of the 10 *pvmdr1* SNPs. The amino acid that was present in the minority of the sequences or not present in them at all was defined as the derived allele. The

inferred ancestral haplotype was present in our dataset and was selected as the control line for the drug assays.

Plasmid Construction

To generate the pDC-PkPvMDR gene replacement plasmid, full-length Sal1 MDR was amplified from *P. vivax* Sal1 genomic DNA and cloned into the XhoI and AvrII sites of pDC2-Camp-eGFP-BSD-attP plasmid (kindly provided by Marcus Lee at the Sanger Institute) using Gibson assembly. Primers used are listed in **Table 4.1**. ~1.5 kb of the *P. knowlesi* 5'UTR was amplified from genomic DNA using the primers BE_1 and BE_2 and cloned into the BamHI and XhoI sites . ~1 kb of the *P. knowlesi* 3'UTR was amplified using BE_3 and BE_4 and cloned into the AvrII and ApaI sites.

Table 4.1: Primers to construct *pvmdr1* plasmid		
BE_1	actatagaatactcaagcttgggggatccCCTCGCGGGCATTGGAAGGGG	Fw 5'Pk MDR1 UTR
BE_2	atccttttcatctcgagTTTCAATGGTTATAGCACAGCGGTGGGAGATTGG	Rv 5'Pk MDR1 UTR
BE_3	GTAGcctaggAGGTCGAAAGGGGCCAGCG	Fw 3'Pk MDR1 UTR
BE_4	agcgaattagctaagcatgcgggcccgcggccgcGGTGATGAATAATAGAACGCACAAATGGC	Rv 3'Pk MDR1 UTR
BE_5	agcatcttacgcgtttggtttttgg	Fw from MluI
BE_6	gtaaaaatgacaatgttatttactaaaccggtt	Rv from AgeI
BE_7	cagagatgttcatttgttaaaaaccggtttag	Fw from AgeI

Table 4.1 (Continued): Primers to construct *pvmdr1* plasmid		
BE_8	cgctggccccttttcgacctcctaggCTACTTAGCCAGCTTGAC	Rv from AvrII

To make each of our haplotypes, we amplified *pvmdr1* from the base plasmid and used primers to insert site mutations for SNPs of interest during PCR into the Sal-1 *pvmdr1* genetic background. We then joined PCR fragments in the second round of PCR. We digested the base plasmid using either Mlu1 and Age1 or Age1 and AvrII, depending on the region of *pvmdr1* that was being modified and used Gibson assembly to insert our *pvmdr1* PCR fragment containing the mutation(s) of interest. We repeated this process iteratively to construct all haplotypes. Primers used to introduce each mutation are listed below using and BE_5/BE_6 or BE_7/BE_8 depending on what restriction sites the SNP fell between (**Table 4.2**).

Table 4.2: Primers to mutate *pvmdr1* plasmid		
Description	Forward Primer	Reverse Primer
V221L	aacaagaagCtgaagattaataagaagacg	tcttcttattaatcttcaGcttcttgttgc
D500N	gaacgaaAatggtttttcttctcaaagtg	gaagaaaaaccatTttcgttcgactcc
S513R	tCGCAACAGaTGTAGGGCTAAATGTGC	CCTACAtCTGTTGCGaCTGTTGG
G698S	CCAAAAAATTtcCAACGCcGGAAGCTAC	CTTCCgGCGTTGgaAATTTTTTGG
L845F	gtttgcgctcTtctatgctaagtacg	cttagcatagaAgagcgcaaacagg
M908L	GTTCGAAAATATTcTGTAtCAAGAAATTAGC	AAGCTAATTTCTTGaTACAgAATATTTTCG

Table 4.2 (continued): Primers to mutate *pvmdr1* plasmid		
T958M	TtCTTGTGAGTAtGGTCATGTCATTTTATTTCTGC	GACATGACCaTACTCACAAGaAAGAGCAC
Y976F	CTGACgGGAACGTtCTTCATTTTTATGAGAG	AAAATGAAGaACGTTCCcGTCAGTAC
F1076L	GTGCgCAAcTaTTCATTAACAGTTTTGC	AACTGTTAATGAAtAgTTGcGCACTCTG
K1393N	cggagaaCctcattgagaagacc	tctcaatgagGttctccgaattgg
N86Y *Pv* equivalent	ggtcattatgaagaatatgTacttgggagaaa	ccaagtAcatattcttcataatgacccca
Y184F *Pv* equivalent	gcatttttgggtctatTcatatggtcg	cttaaatagcgaccatatgAatagaccc
N1042D *Pv* equivalent	gcccaattcttcattGacagttttgc	ccagtaggcaaaactgtCaatgaag
D1246Y *Pv* equivalent	gcgattacaacttaaagTatttgagaaac	ggaaaataagtttctcaaatActttaagttg

The pDC-Cas9-PkU6-hdhfr plasmid was generated by amplifying the *P. knowlesi* U6 promoter from genomic DNA using primers BE_9 and BE_10, and the guide chimeric region plus terminator using BE_11 and BE_12 from the pD2-Cas9-PfU6-hdhfr plasmid. The two PCR products were cloned into the pD2-Cas9-PfU6-*hdhfr* plasmid using the BamHI sites in a single Gibson assembly reaction. sgRNAs designed against the 5' and 3' ends of *P. knowlesi mdr1* were cloned into pDC-Cas9-PkU6 using the annealed and phosphorylated primers BE_13/BE_14 (5' end of the gene) and BE_15/BE_16 (3' end of the gene) cloned into the Bbs1 sites by ligation. The primers are listed below in **Table 4.3**.

Table 4.3: Primers used to make guide RNA plasmids		
BE_9	CCTGCAGGTCGACTCTAGAGGATCCGCTAGCgagaacatgatatatagcagaatttaaacgttacc	Fw Pk U6 promoter
BE_10	ctaaaacagGTCTTCtcGAAGACccaataatatactgtaactcagaatatatggatatgc	Rv Pk U6 promoter

Table 4.3 (Continued): Primers used to make guide RNA plasmids		
BE_11	ggGTCTTCgaGAAGACctgttttag	Fw Cas9-chimeric guide region – for Gibson to PkU6
BE_12	AGAATACTCAAGCTTGGGGGGATCCACTAG TGCCTTAAAAACTTCATTATATTTAAAAATT ATTTTATAGG	Rv Chimeric guide – Gibson into cas9 plasmid.
BE_13	TATTgcagtccatagacaataacag	Fw Pk Cas9guide 5' PkMDR1
BE_14	AAACctgttattgtctatggactgc	Rv Pk Cas9guide 5' Pk MDR1
BE_15	TATTgaattgcctctattaagagat	Fw Pk Cas9guide 3' PkMDR1
BE_16	AAACatctcttaatagaggcaattc	Rv Pk Cas9guide 3' PkMDR1

We also constructed an overexpression version of the plasmid by adding a hDHFR sequence into the original plasmid. This plasmid was transfected into *P. knowlesi* without the CRISPR guides and maintained episomally by keeping the parasite line under constant pyrimethamine pressure.

Parasite Culture:

Parasites were cultured in RPMI-1640 media supplemented with 25 mM HEPES, 11.50 mg/l hypoxanthine, 2.42 mM sodium bicarbonate, and 4.31 mg/ml AlbuMAX II, at 37^{o}C in a 1% oxygen, 5% carbon dioxide and 94% nitrogen environment.

Parasite Transfections

We first grew parasites to 5-10% parasitemia. We isolated schizonts using a 60% percoll gradient spun at 900 G for 15 minutes, washed them with complete media, and let them mature at 37°C for one hour. To perform transfections, we combined ~2-3 µg of plasmid in a final volume

of 10 µl with ~2-3 µg of our two guide RNA plasmids in a final volume of 2.5 µl for a total volume of 15 µl. We combined the plasmid mixture with 100 µL of buffer P3 (Lonza). and iRBC pellets (10-20 µl). We electroporated our plasmids into parasites using a Lonza nucleofector machine (FP158 program) and then immediately transferred the plasmid-pellet buffer mixture to 1 ml of 20% hematocrit blood (80% media, 20% rhesus blood) to recover. We let parasites recover for 20-40 minutes at 37°, after which we spun down the mixture and removed the supernatant. We then added the blood back to 10 ml of culture and let parasites grow for 24 hours. After 24 hours, we added pyrimethamine to a final concentration of 100 nM in the culture to select for parasites with the integrated plasmid. We then waited 10-21 days for parasite populations to grow and reappear via smear. We confirmed plasmid integration using species-specific primers Pk_fw: aggaacatgaaaaagtgccattctggg and Pk_rev: CATGGAGACAACAGTGCTGTT and Pv_fw: ctaccgaatcagtacgacaccaac and Pv_rev: tgcagcgactctttaaaggcacc. Clonal parasites were isolated by limiting dilution.

SYBR Green growth inhibition assays

We performed growth inhibition assays as described by van Schalkwyk et al.[47] In brief, to synchronize parasites we isolated schizonts using a 60% percoll gradient, spun at 900g for 15 minutes. We let schizonts reinvade in 10% hematocrit media using leukocyte-depleted blood for five hours. We then removed schizonts that did not reinvade using another 60% percoll gradient and removed the media, schizont layer, and percoll, leaving just the blood pellet.

We counted final parasitemia by staining parasites with SYBR green for 15 minutes, washing parasites, and counting using a MACsquant flow cytometer. For parasites that were included in our final assay, we diluted all parasite lines to a final parasitemia of 0.8% and 1% hematocrit. We plated 40 µl of parasites per well on our 384 well plate with drug dilutions. We

incubated plates under normal growth conditions for 1.5 life cycles (~48 hours) at 37°C. We prepared the SYBR lysis buffer by mixing lysis buffer (0.16% Saponin, 20 mM Tris-HCl, 5 mM EDTA, 1.6% (v/v) Triton X-100, pH 7.4) with SYBR in a 1:1000 ratio. We lysed cells by adding 10 µl of the SYBR lysis buffer to each well. Fluorescence was read on a plate reader (SpectraMax iD5, Molecular Devices) at 490 nm excitation and 520 nm emission after 24 hours of incubation. We performed assays for at least six biological replicates.

Data Analysis and Visualization

We calculated the IC_{50} using GraphPad PRISM software (GraphPad Software). Statistical analysis and visualizations were conducted using R software (Version 4.1) and the *ggplot2* package.

4.4 Results:

Population Genomics Reveal a Diverse Set of Global *pvmdr1* Haplotypes

We first sought to identify a globally representative set of *pvmdr1* haplotypes likely associated with resistance. We aligned *P. vivax* samples to the P01 reference genome and called variants. We then subsetted the master VCF of 339 *P. vivax* samples to the region containing just the *pvmdr1* gene. Next, we calculated minor allele frequency (MAF) for all nonsynonymous SNPs (nSNPs) in *pvmdr1* and excluded all nSNPs with a MAF < 0.05 in the entire population, leaving 10 nSNPs: V221L, D500N, S513R, G698S, L845F, L908M, T958M, Y976F, F1076L, K1393N (**Fig 4.1A).** We next inferred an ancestral *pvmdr1* sequence to use as our control/reference sequence during our assays (**Fig 4.1B**). The ancestral *pvmdr1* sequence represents the pre-drug exposure haplotype. We inferred the ancestral *pvmdr1* sequence by first aligning *pvmdr1* from the P01 reference genome, and the *mdr1* ortholog sequences from *P. knowlesi, P. coatneyi*, *P. inui*, and *P. cynomolgi*. The allele present in the majority of the five species was selected as the ancestral

allele for each of the 10 SNPs. The ancestral haplotype (H1) was present in our population genomics dataset.

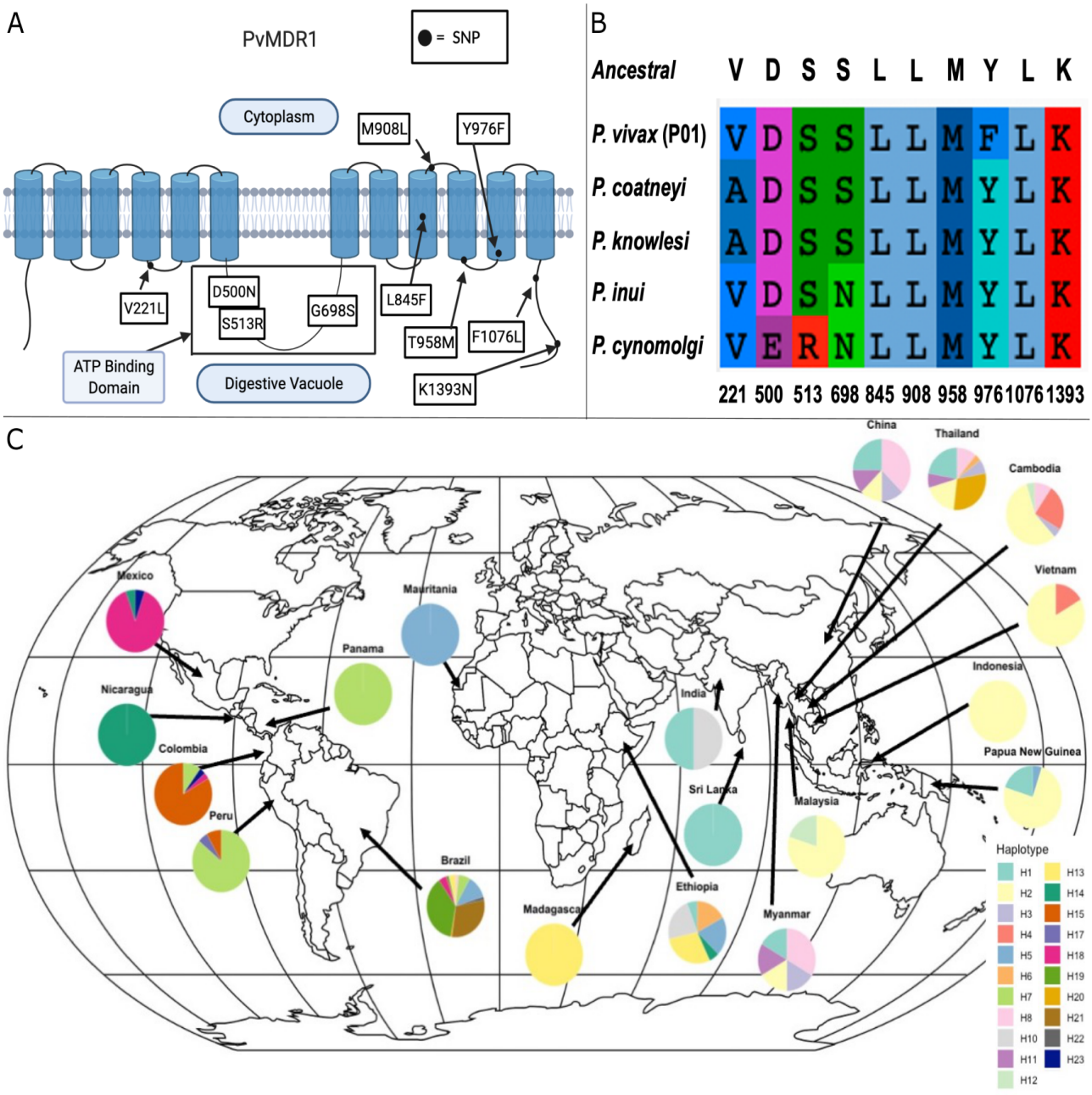


Figure 4.1: *pvmdr1* Mutations and Global Distribution of *pvmdr1* Haplotypes: 4.1A) Displays the 10 *pvmdr1* nSNPs locations on the *pvmdr1* structure. **4.1B)** Displays sequence alignment of the 10 nSNPs across five *Plasmodium* species in the *Plasmodium* clade and ancestral consensus sequence. **4.1C)** Displays the global distribution of *pvmdr1* haplotypes and their frequency in each country.

We then identified all existing *pvmdr1* haplotypes composed of these 10 nSNPs using a custom script to look for all unique existing combinations of the 10 nSNPs in our dataset. We identified 22 unique *pvmdr1* haplotypes, comprising the 10 nSNPs (**Fig 4.1C** and **Table 4.1**). The 22 haplotypes represent a geographically diverse set of *pvmdr1* haplotypes in the circulating worldwide *P. vivax* populations (**Fig 4.1C**). The H9 haplotype was not present in our dataset, possibly due to low sample depth in Indonesia, but was characterized in the study by Verzier et al using episomal overexpression in *P. knowlesi*[58]. The counts of each haplotype by country in the dataset are summarized in **Supplemental Table 4.1.**

We also wanted to characterize the role of *pvmdr1* SNPs commonly reported in the literature. The SNPs S513R, G698S, L908M, M958T, Y976F, and L1076F are commonly reported in the literature as possibly involved in *P. vivax* drug resistance and are also among our 10 SNPs in *pvmdr1* with a MAF > 0.01. The L908M and M958T in the Sal-1 genetic background are among 22 *pvmdr1* haplotypes identified in the population dataset. We wanted to test the other four mutations identified in the literature that were not present in our population as existing haplotypes. We constructed plasmids that contained the 513R, G698S, Y976F, and L1076F SNPs as single mutations in the Sal-1 *pvmdr1* genomic background. We also wanted to test if *pfmdr1* mutations that reduced drug susceptibility in *P. falciparum* also did so in *P. vivax*. We aligned the *pfmdr1* and *pvmdr1* genes and mapped where the 86Y, 184F, 1042C, and 1264D mutations aligned to *pvmdr1*, including the 86Y and 184F mutations combined into a single haplotype. The full list of haplotypes is displayed in **Table 4.4**

Table 4.4: List of *pvmdr1* Haplotypes Assayed											
V221L	**D500N**	**S513R**	**S698G**	**L845F**	**L908M**	**M958T**	**Y976F**	**L1076F**	**K1393N**	**Name**	**SNPs**
V	D	S	S	L	L	M	Y	L	K	H1	Base Haplotype
V	D	S	S	L	L	M	**F**	L	K	H2	H1+976F
V	D	S	S	L	L	M	Y	**F**	K	H3	H1+1076F
V	D	S	S	**F**	L	M	Y	L	K	H4	H1+ 845F
V	D	S	**G**	L	L	M	Y	L	K	H5	H1+698G
V	D	**R**	S	L	L	M	Y	L	K	H6	H1 +513R
V	D	S	**G**	L	L	M	Y	**F**	K	H7	H1+698G+1076F
V	D	**R**	S	L	L	M	Y	**F**	K	H8	H1+513R+1076F
V	D	S	S	L	L	**T**	**F**	L	K	H9	H1+958T+976F
V	D	**R**	**G**	L	L	M	Y	L	K	H10	H1+ 513R+698G
V	D	S	S	L	L	M	Y	**F**	**N**	H11	H1+1076F+ K1393N
V	D	S	S	**F**	L	M	Y	**F**	K	H12	H1+845F+1076F
V	D	**R**	S	L	L	M	**F**	L	K	H13	H1+513R+976F
V	D	S	**G**	L	L	M	**F**	L	K	H14	H1+698G+976F
L	D	S	**G**	L	L	M	Y	**F**	K	H15	H1+221L+698G+ 1076F
V	D	S	S	**F**	L	M	Y	**F**	**N**	H16	H1+845F+1076F+ K1393N
V	D	S	**G**	L	L	**T**	Y	**F**	K	H17	H1+698G+958T+ 1076F
V	D	S	**G**	L	**M**	M	Y	**F**	K	H18	H1+698G+908M+ 1076F
V	**N**	S	**G**	L	L	M	Y	**F**	K	H19	H1+ D500N+698G+ 1076F
V	D	**R**	S	L	L	M	Y	**F**	**N**	H20	H1+513R+ 1076F+1393N
V	**N**	S	**G**	L	L	**T**	Y	**F**	K	H21	H1+500N+698G+ 958T+1076F
V	**N**	S	**G**	L	**M**	M	Y	**F**	K	H22	H1+500N+698G+ 908M+1076F
V	D	S	**G**	L	**M**	**T**	Y	**F**	K	H23	H1+698G+908M+ 958T+1076F
Single Mutations in Sal-1 Background Not in Population Dataset											
V	D	S	**G**	L	**M**	**T**	Y	L	K	H24	H1+698G+908M+ 958T

Table 4.4 (Continued): List of *pvmdr1* Haplotypes Assayed											
V	D	S	S	L	**M**	**T**	Y	**F**	K	H25	H1+908M+958T+ 1076F
V	D	**R**	**G**	L	**M**	**T**	Y	**F**	K	H26	H1+513R+698G+ 908M+958T+ 1076F
V	D	S	**G**	L	**M**	**T**	**F**	**F**	K	H27	H1+698G+908M+ 958T+976F+ 1076F
Single Mutant		Double Mutant		Triple Mutant		Quadruple Mutant			Quintuple Mutant		

Generation of *P. knowlesi* containing the *pvmdr1* in place of the native *pkmdr1*

Next, we constructed plasmids each containing the 27 *pvmdr1* haplotypes. We amplified *pvmdr1* from Sal-1 genomic DNA and used PCR to introduce the mutations iteratively until we constructed all 27 haplotypes. We also constructed plasmids that mapped known *pfmdr1* drug resistance-associated mutations N86Y, Y184F, N1042C, D1246Y onto the *pvmdr1* gene as single mutations, and as an N86Y, 1845F haplotype. Finally, we constructed an overexpression plasmid containing the H1 *pvmdr1* haplotype to mimic the effect of copy number variation. Sanger sequencing confirmed *pvmdr1* plasmid haplotype sequences.

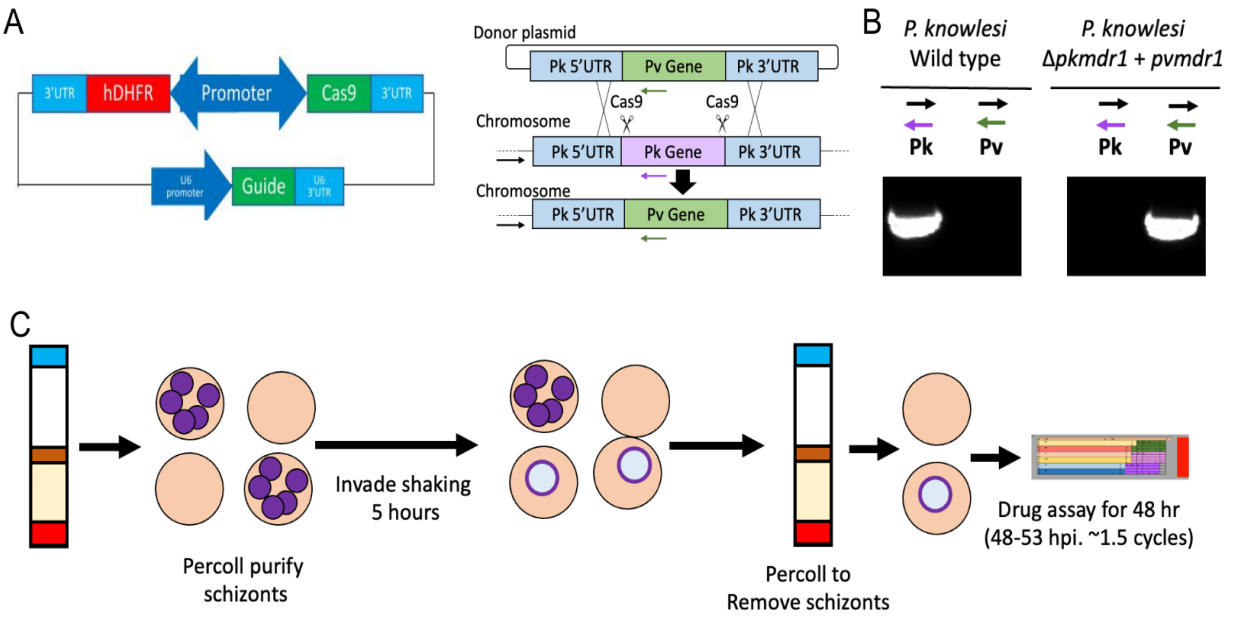


Figure 4.2: Allelic Replacement Strategy and Experimental Design: 4.2A: Displays plasmid constructs for the Cas9 and Guide RNA Plasmid on the left, and the allelic replacement plasmid on the right. **4.2B**: Shows an example of replacement of native *pkmdr1* locus with *pvmdr1*. The left column in each picture displays a PCR reaction with our *P. knowlesi* specific *mdr1* primer set, and the right column displays a PCR reaction with our *P. vivax* specific *mdr1* primer set. **4.2C:** Displays drug assay experimental workflow. First, we isolate schizonts using a percoll gradient and let them reinvade leukfiltered blood for five hours to synchronize parasites. We then removed schizonts that did not invade to ensure parasite synchronization. Parasites were diluted to 0.8% parasitemia, 1% hematocrit and dispensed in a 384-well drug plate and cultured for 48 hours. Parasites were stained with SYBR, and fluorescence was read to measure parasitemia after drug exposure.

Each *pvmdr1* variant plasmid was transfected with two Cas9-sgRNA plasmids targeting the 5' and 3' end of the *pkmdr1* gene (**Fig 4.2A**). We confirmed plasmid integration into the *P. knowlesi* genome using species-specific primers (**Fig 4.2B**). We cloned recombinant parasites by limiting dilution to generate isogenic lines for each haplotype. After isolating isogenic clones, we extracted genomic DNA and sequenced the entire *pvmdr1* gene to ensure that the *pvmdr1* gene in that clonal line was both present and the correct haplotype.

Several *pvmdr1* Haplotypes and Overexpression Confer Reduced Susceptibility to Mefloquine, Halofantrine, Lumefantrine, and Dihydroartemisinin

We synchronized parasite lines and diluted them down to a final parasitemia of 0.8%. We plated parasites on drug plates containing concentration gradients of chloroquine (CQ), dihydroartemisinin (DHA), mefloquine (MFQ), lumefantrine (LUM), ferroquine (FQ), piperaquine (PIP), halofantrine (HF), monodesethylamodiaquine (MDAQ), and pyronaridine (PYN). We let parasites incubate on the plates for 48 hours (~two cycles) and then added lysis buffer with SYBR to measure DNA concentration as a measure of parasite survival at the various drug concentrations (**Fig 4.2C**).

We assayed 24 of the 27 *pvmdr1* haplotypes lines, the one overexpression line, four *pfmdr1* mutants, and the *P. knowlesi* YH1 line (n=30 lines), for their response to CQ, DHA, MFQ, LUM, FQ, PIP, HF, MDAQ, and PYN (**Figure 4.3**). We made transgenic lines containing the H4, H5, and H13 haplotypes, but were unable to isolate isogenic clones, and fully assay them by the time of writing. We included the H1 ancestral haplotype in each assay as the control line to compare fold-changes in other lines against and normalized all lines against H1 for analysis. We also assayed the *P. knowlesi* YH1 line. Each line was assayed with 5-12 biological replicates per drug. H8 was the exception, and only had four biological replicates per drug by the time of writing.

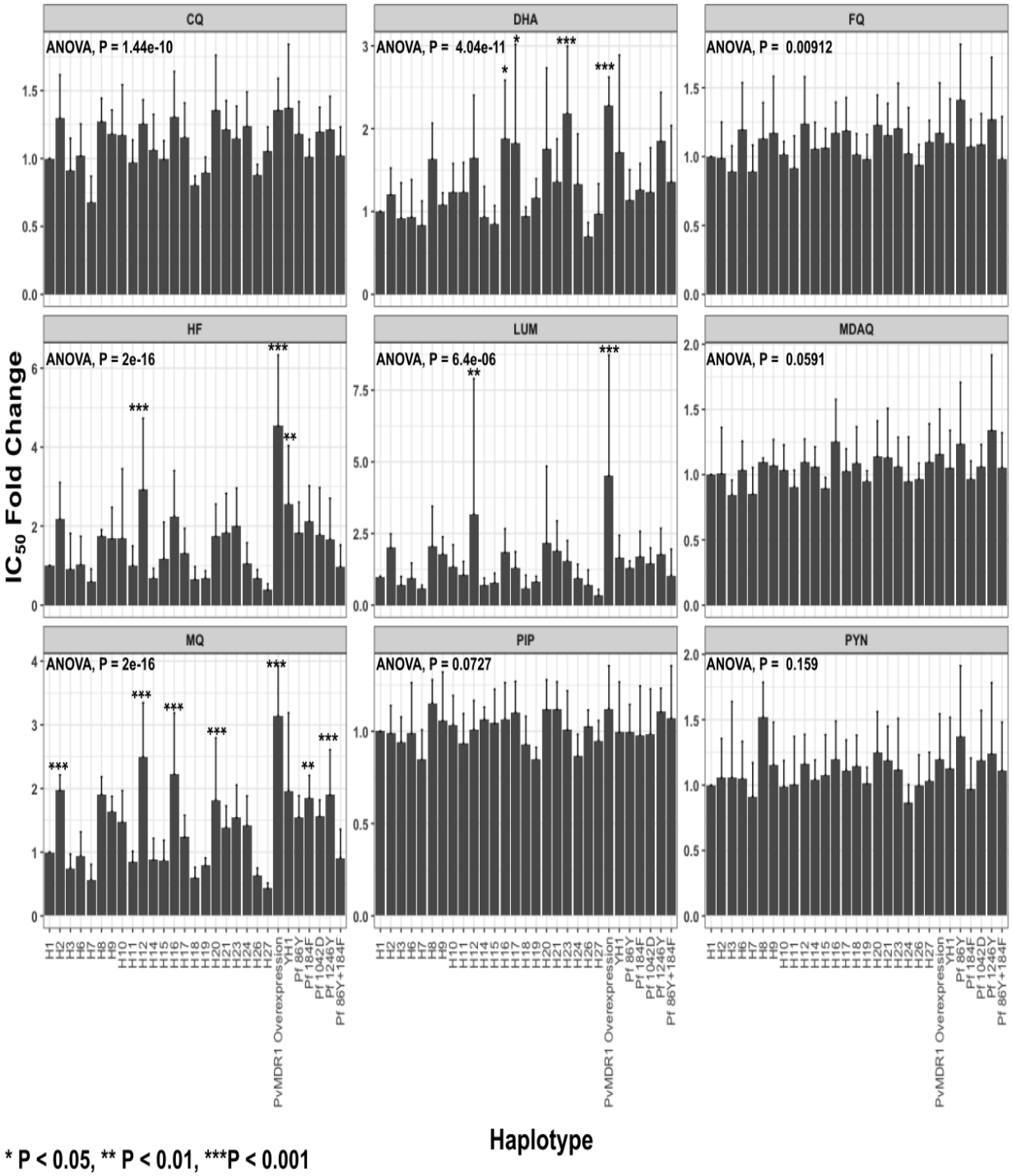


Figure 4.3. Antimalarial IC_{50} Values for All Assayed Lines: Displays IC_{50} values for all 24 assayed lines. Error bars represent standard deviation of all biological replicates. We used an ANOVA test with 30 degrees of freedom to test if there was significant variability across all lines in their response to each drug. We used a pairwise t-test to assess if the mean fold change for each line was different for all pairwise comparisons, using the Bonferroni-Holm correction. *P < .05, **P < .01, and ***P < .001, compared with the H1 ancestral *pvmdr1* line.

pvmdr1 lines, H2, H12, H16, H20, and H23 displayed reduced susceptibility to several of our compounds (**Figure 4.3**). The H2 haplotype displayed a ~2-fold shift in the IC_{50} of LUM, HF, and MQ, though only the fold-increase in MQ was statistically significant. H2 is a single tyrosine to phenylalanine substitution at amino acid position 976 (Y976F) in the ancestral background. The H12 haplotype displayed a statistically significant 2.5-fold increase in the MQ and HF IC_{50}s, and a 3-fold increase in LUM. H12 is a double mutant in the ancestral background, with leucine to phenylalanine substitutions at positions 845 (L845F) and 1076 (L1076F). The H16, H17 and H23 all displayed significant 1.5-2-fold shifts in the IC_{50} of DHA. H16 and H23 also displayed a significant ~1.5-fold and 2-fold increase in MQ IC_{50} respectively, and a ~2-fold increase in HF IC_{50}, though it was not significant. The H16 and H17 lines are triple mutants, and H23 is a quadruple mutant in the ancestral background. H16 contains the L845F, L1076F, and K1393N mutations. H17 and H23 both contain the S698G, M958T, and L1076F mutations. H23 also contains the L908M mutation. The H20 haplotype displayed a significant ~2-fold increase in MQ IC_{50}. H20 constraints the 513R, 1076F, and 1393N mutations. The H1 overexpression line displayed a significant ~4-fold increase in the LUM and HF IC_{50}s, a ~3-fold increase in the IC_{50} of MQ, and a 2-fold shift in the DHA IC_{50}. The H7 haplotype displayed increased sensitivity to several compounds, including a ~2-fold increase in sensitivity to MQ, ~1.5-fold increase in sensitivity to LUM and CQ (**Figure 4.3**), though these fold-decreases were not statistically significant. H7 is a double mutant in the H1 background, with the S698G mutation L1076F mutation. Several other haplotypes also possibly display intermediate sensitivity haplotypes; however, these shifts were not statistically significant possibly due to the variability of the assay.

Some of the *pfmdr1* mutations associated with *P. falciparum* drug resistance mapped onto *pvmdr1* resulted in increased IC_{50}s for many compounds. We noticed a general trend towards

decreased susceptibility towards MQ, LUM, and HF, though not all the increases were statistically significant. The 184F and 1246Y mutations resulted in 2-fold increases in the MQ. The other *pfmdr1* mutations had small increases in the IC_{50}s of DHA, MQ, LF, and HF. We did not observe an increase in the IC_{50} of MDAQ, FQ, PIP, CQ, and PYN in any of the lines expressing the *pfmdr1* SNPs mapped onto the *pvmdr1*.

There was significant variation in the mean IC_{50}s for MQ, LF, HF, DHA, CQ, and FQ across all 30 assayed lines **($P < 0.01$, ANOVA test for all lines)**. MQ, LF, HF, and DHA all had at least two lines with significantly higher IC_{50}s compared to the H1 control. These compounds also had significant variability between other lines in the assay. No lines had significantly higher CQ and FQ IC_{50}s compared to the H1 control. However, there were significant differences in the CQ and FQ IC_{50}s across all lines, suggesting differences in the response to these drugs among the circulating *pvmdr1* haplotypes.

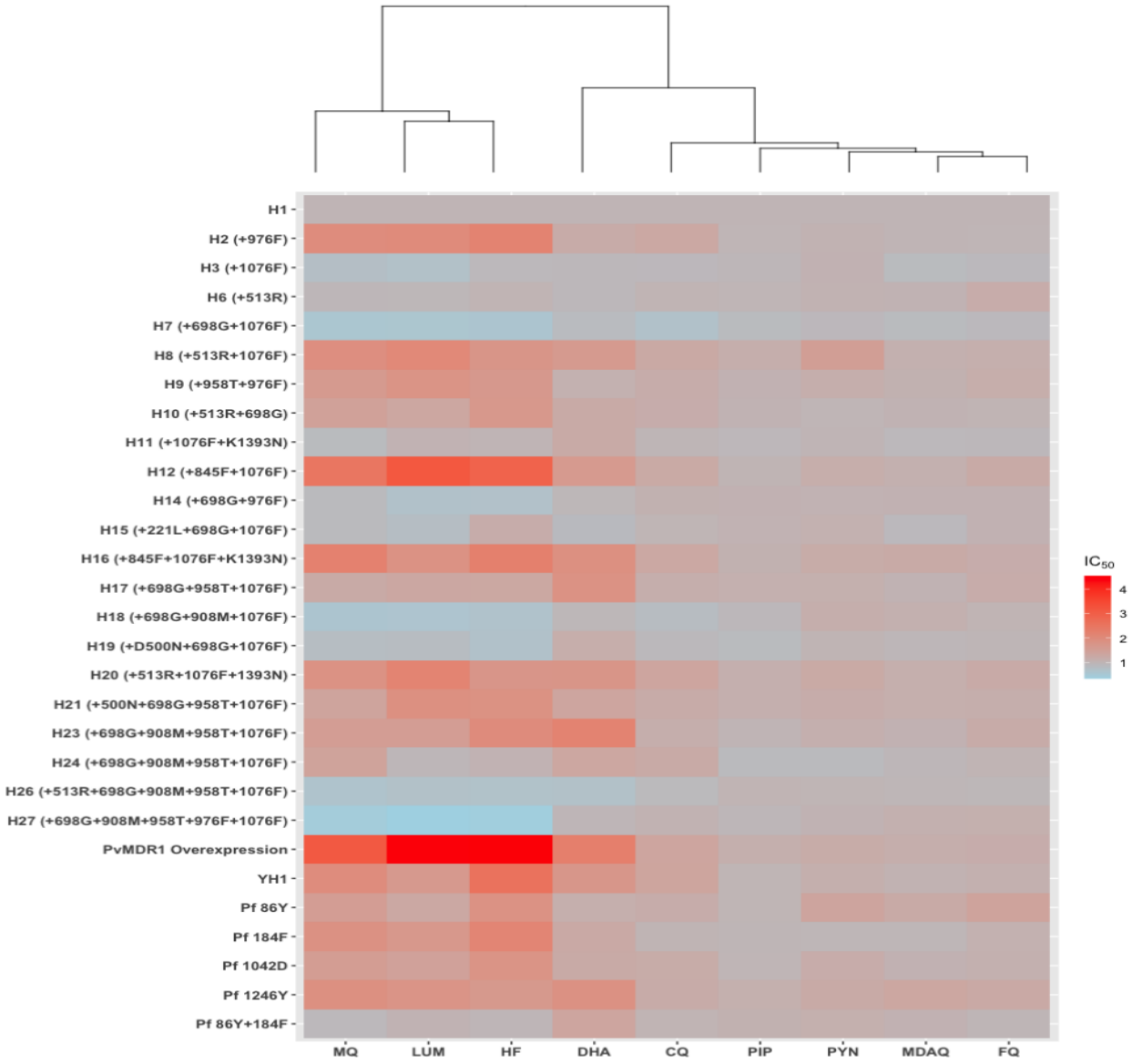


Figure 4.4: Heatmap and Clustering of *pvmdr1* line IC_{50}s: displays heatmap of the nine antimalarial IC_{50} in each line. Dendrogram shows hierarchical clustering of drug IC_{50}s.

We next wanted to test if any of the IC_{50} fold-changes correlated with each other to identify if any of the lines had cross resistance to multiple compounds. We observed IC_{50} increases generally clustered together for MQ, LUM, and HF. This observation suggests cross resistance to MQ, LUM, and HF for all lines that displayed reduced susceptibility to those compounds **(Figure 4.4)**. This observation was supported by hierarchical clustering of the IC_{50} values for each line. We also

observed that increases in DHA IC_{50} did correlate to a lesser extent to other decreases in susceptibility to other antimalarials. We did not observe collateral sensitivity in any line. We observed only small shifts in the IC_{50} of FQ, PYN, MDAQ, and PIP across all lines. Correspondingly, the IC_{50}s for this set of compounds clustered together for all lines.

4.5 Discussion

pvmdr1 Haplotypes Confer Reduced Drug Susceptibility and are Prevalent in the Greater Mekong Subregion and Oceania.

This study provides the first molecular genetic evidence that *pvmdr1* can mediate multidrug resistance in *P. vivax*. The extensive interrogation of globally circulating *pvmdr1* alleles reveals decreases in susceptibility to several widely used antimalarials. The H2, H12, H16, H17, H20, and H23 haplotypes, in addition to the *pvmdr1* overexpression line, all display 2-4-fold resistance to LUM, MQ, and HF (**Table 4.5**). H2 is widespread throughout Southeast Asia, where it was present in 83% of Vietnamese samples, 18.5% of Thailand samples, 75% of Papua New Guinea samples, 17% of the Myanmar samples, 67% of the Malaysia samples, 100% of the Indonesian samples, 12% of the Chinese samples, and 57% of the Cambodian samples. H2 displays a significant increase in MQ $IC_{50,}$ and a ~2-fold shift in the IC_{50} of LUM and HF, though this shift was not statistically significant. The high prevalence of H2 suggests possibly widespread reduced *P. vivax* sensitivities to these drugs throughout Asia and Oceania.

Table 4.5: Lines with a 2-fold increase in IC_{50} of at least one drug.		
Line	**SNPs vs H1**	**Mean IC_{50} fold Change vs H1**
H2	+976F	CQ:1.30, DHA:1.20, **MQ:2.0**, LUM: 2.00, HF: 2.18
H12	+845F +1076F	CQ:1.24, DHA:1.65, **MQ:2.50**, **LUM: 3.18**, **HF: 2.92**
H16	+845F+1076F+1393N	CQ:1.30, **DHA:1.88**, **MQ:2.22,** LUM: 1.83, HF: 2.25
H17	+698G+958T+1076F	CQ:1.16, **DHA:1.82**, MQ:1.23, LUM: 1.32, HF: 1.31
H20	+513R +1076F +1393N	CQ:1.35, **DHA:1.75**, MQ:1.82, LUM: 2.16, HF: 1.74
H23	+698G+908M+958T +1076F	CQ:1.15, **DHA:2.19**, MQ:1.54, LUM: 1.55, HF: 1.99
pvmdr1 overexpression	H1 (no SNPs)	CQ:1.35, **DHA:2.28**, **MQ:3.14**, **LUM: 4.52**, **HF: 4.54**
Compounds with significant fold increases in their $IC_{50\ are}$ bolded. Compounds with IC_{50} fold increases greater than 2 (or 1.5 in the case of DHA) are underlined.		

H12 is rare in our dataset and was only found in one sample each in Malaysia and Cambodia (frequencies of ~5% and ~17% respectively). H20 was found in 30% of the samples from Thailand and less than 1% of samples from Brazil. H16, H17, H23 were also rare, and found in less than 5% of samples from India, Peru, and Mexico, respectively. Despite the observation that several reduced susceptibility haplotypes are rare, they are found distributed around the world. Our dataset may also paint an incomplete picture of the true frequency of these variants, given unequal sampling depth between countries. This work highlights the urgent need for *in vivo* studies exploring links between *pvmdr1* mutations and treatment failure to ACT partner drugs, even if those drugs are not yet recommended as *P. vivax* first-line therapy. There is prior evidence for co-selection of *P. falciparum* and *P. vivax* drug resistance in regions where both species are co-endemic[43,44,59]. SP treatment prescribed for *P. falciparum* has been shown to select for SP-resistant

P. vivax dhfr alleles[43,44,59]. While ACTs are only recommended in a few regions as the first-line *P. vivax* treatment, they have been recommended to treat *P. falciparum* for close to two decades. Consequently, *P. vivax* has been subjected to selection pressure from these same antimalarial compounds. Our study provides evidence that exposure to MQ, LUM, HF, and DHA has been selected for *pvmdr1* alleles with reduced susceptibility to these drugs. Our work emphasizes the importance of understanding the link between circulating *pvmdr1* alleles and their predictive power of ACT treatment failure.

We identified several common SNPs across resistant lines. The H12, H16, H17, H20, and H23 lines had the 1076F mutation **(Table 4.5)**. 1076F alone does not appear to cause resistance, because the H3 haplotype, a single 1076F mutation in the ancestral background, did not demonstrate reduced susceptibility to any compounds. However, the addition of the 845F mutation (H12 Haplotype) or the 698G, 908M, and 958T mutations (H23 Haplotype) resulted in reduced susceptibility to MQ, LUM, and HF for H12 or DHA for H23. Comparing the H23 vs H24, and the H12 vs H3 haplotypes further supports a role for 1076F in the development of drug resistance. H24 and H23 have the same haplotype (698G, 908M, and 958T), except for the amino acid at position 1076 (1076F for H23 and L1076 for H24). H24 had no shift in the HF IC_{50} and displayed reduced susceptibility to DHA relative to the H23 line, while H23 displayed reduced susceptibility to both compounds. This observation suggests a critical role for mutations at position 1076 for the development of multidrug-resistant *pvmdr1* alleles. The 1076F mutation could be critical for the development of reduced drug susceptibility.

The 845F and 513R mutations, in addition to 1076F, may confer reduced susceptibility to several compounds. H12 contains 845F plus 1076F and has reduced susceptibility to DHA, MQ, LUM, and HF. H16 also contains both mutations in a slightly different genetic background, but

also displayed similar decreases in susceptibility to these compounds. Unfortunately, we could not determine if the 845F mutation alone is sufficient to lead to reduced susceptibility phenotypes, because we were unable to assay H17, the single 845F mutation in the ancestral background, by the time of this writing. 513R may also play a role in the development of drug resistance. H20 contains 513R and 1076F and displays a ~1.5-fold shift in DHA IC_{50} and a 2–fold in LUM IC_{50}. The H11 haplotype, which does not contain 513R, but otherwise has the same mutations as H20, displayed no increase in IC_{50} towards any antimalarials. Similarly, the H3 (+1076F in ancestral background) and the H25 (+513R in ancestral background) lines did not display any increases in the IC_{50} of any antimalarials. These data suggest that 513R in concert with 1076F mediates reduced susceptibility to MQ, LUM, and HF. Notably, we did not observe opposing sensitivities in any of our *pvmdr1* lines.

H2 was our only resistant line that did not contain 1076F. H2, which is a single Y976F substitution in the ancestral background, displayed a 2-fold increase in the IC_{50} of MQ, LUM, and HF. Unlike the other reduced drug susceptibility lines, it did not have a shift in the IC_{50} of DHA. Mutations at position 976 had been associated with possible CQ resistance[60]. The H2 haplotype did not display an increased CQ IC_{50}, nor any other line with the 976F mutation. The H9, H14, and H27 lines also had the 976F mutation but did not display increased IC_{50}s to any other antimalarials in the assay. This observation suggests that the association of 976F with reduced antimalarial susceptibility depends on the genetic background it arises in. The H27 line, which has 976F in the SAL-1 (H23) background, displayed increased sensitivity to HF, LUM, and MQ. H27 was not observed in our population genomics dataset, perhaps because of decreased fitness costs due to growth defects or increased drug sensitivity. This observation suggests the 976F mutation can only persist on certain genetic backgrounds. The H13 line also contains a 976F mutation and

the 513R mutation, but we were not able to completely assay this line by the time of writing. H2 is one of the most common haplotypes in Southeast Asia, suggesting it can spread readily and be a useful marker of emerging ACT partner drug resistance in that region.

None of the lines displayed any shift in the IC_{50}s of FQ, PYN, MDAQ, and PIP, and any increase in the IC_{50} of CQ relative to the H1 control. However, there were significant differences in CQ IC_{50} **(P < 0.001=, ANOVA)**, and to a lesser extent FQ IC_{50} **(P < 0.01),** across all 30 assayed lines. These results show that *pvmdr1* mutations may mediate small differences in susceptibility to these two drugs among circulating *P. vivax* lineages. However, no haplotype displayed a clear reduced susceptibility to CQ phenotype (the max fold increase was ~1.35-fold). This result demonstrates that *pvmdr1* alone does not mediate high-grade CQR, and that SNPs in *pvmdr1* should not be used as molecular markers of high-grade CQR. This contrasts with prior work that linked 976F with a 1.7-fold increase in CQ IC_{50}[60]. One explanation for these conflicting observations is that *pvmdr1* alleles containing 976F (or specifically the H2 haplotype) are genetically linked to a gene that mediates high-grade CQR. Another possible explanation is that high grade CQR is multigenic, and may involve resistant *pvmdr1* alleles plus another gene, possibly *pvcr*t. Several lines showed small increases in CQ IC_{50}, suggesting a trend towards small reductions in CQ susceptibility. Mutant *pvmdr1* alleles may therefore play a role in the development of high-grade CQR.

No *pvmdr1* haplotype conferred reduced susceptibility to FQ, PYN, MDAQ, and PIP relative to H1. This observation supports the conclusion that the shifts we observe in the IC_{50}s of DHA, MQ, LUM, and HF relative to H1 are due to genuine biological change that mediate reduced susceptibility specifically to these antimalarials, not to others. If we observed shifts in the IC_{50} relative to H1 for FQ, PYN, MDAQ, and PIP, that might imply that the shifts we observed in DHA,

MQ, LUM, and HF are random noise due to assay variability. Notably, we also did not observe a significant difference **(P > 0.05, ANOVA test)** across all 30 lines in the variability of PYN, MDAQ, and PIP. This result demonstrates that *pvmdr1* alleles do not mediate susceptibility to these compounds and supports the conclusion that the observed increases in IC_{50}s of the other antimalarials are genuine.

pvmdr1 Alleles Antimalarial IC_{50} Increases Are Similar to Changes Seen in *P. falciparum pfmdr1* and Suggests Clinical Importance

There are striking similarities between the IC_{50} increases in MQ, LUM, DHA, and HF due to *pvmdr1* mutations, and previously reported *in vitro* IC_{50} increases observed in *P. falciparum*.[15,39,61,62] Previous studies reported a similar relationship between *pfmdr1* mutations and observed IC_{50} fold-changes to antimalarial compounds[61,62,62,63]. The *pfmdr1* N86 allele results in 2-4 fold increases in the IC_{50}s of LUM and MQ, and a 1.5 fold increase in the IC_{50} of DHA.[15,63] Conversely, the 86Y allele can lead to amodiaquine resistance, and has also been significantly associated with artesunate plus amodiaquine (ASAQ) treatment failure[61,64,65]. The S1034C and D1246Y *pfmdr1* alleles can also result in a ~2 fold shift in parasite resistance of HF and MQ[39]. The *pfmdr1* 1042D allele results in a 2-fold increase in quinine (QN) resistance compared to the N1042 wild-type allele[39]. *pfmdr1* copy number increases have also been associated with increased resistance to MQ, HF, and LUM. *pfmdr1* copy number increases result in a ~2-4-fold increase in the IC_{50} of both MQ and HF i*n vitro*.[66] Conversely, disrupting one of the two *pfmdr1* copies in a *P. falciparum* line with two copies of *pfmdr1* resulted in a 3-fold decrease in MQ IC_{50}, a 4-5-fold decrease in LUM IC_{50}, and a ~2-fold decrease in HF IC_{50}[62]. The similarity between the IC_{50} fold-changes in our *pvmdr1* mutants and overexpression lines to the IC_{50} fold-changes resulting from *pfmdr1* mutations may also help contextualize the impact of *in vitro* resistance to *in vivo* treatment failures.

While *pfmdr1 in vitro* shifts only result in 2-4-fold changes to the IC_{50} of these drugs, work linking the *pfmdr1* genotypes to clinical outcomes suggests that even these modest *in vitro* changes result in significant risk for *in vivo* treatment failure. One study found an increased prevalence of the N86 in *P. falciparum* parasites present in patients with artemether-lumefantrine (AL) treatment failure, compared to the allele frequencies of N86 in the parasites prior to AL treatment[67]. The same study also found a 2-fold increase in the parasite IC_{50} of artemether and ~2.5-fold shift in LUM IC_{50} in *ex vivo* assays of parasites from patients with treatment failure[67]**.** Other studies have shown that the N86 allele is associated with a significant risk of AL treatment failure[64,68–70]. *pfmdr1* CNVs are associated with significantly increased odds of treatment failure in patients treated with AL and MQ[65,71,72]. The 86Y allele is associated with recrudescent infections in patients treated with AQ[73]. A corollary of the relationship between mutations in *pfmdr1* resulting in *in vitro* shifts in drug IC_{50}s and associated with risk of treatment failure, coupled with the similar IC_{50} fold-changes observed in several of the *pvmdr1* mutant lines in our study, suggests that mutations in *pvmdr1* may increase risk for ACT treatment failure for *P. vivax*.

The few studies linking *pvmdr1* polymorphisms to clinical outcomes support a connection between *pvmdr1* polymorphisms and the risk of treatment failure[9,19]. Other work has associated increased *pvmdr1* copy-number with a 2-fold increase in MQ IC_{50} *ex vivo*[45]. Additionally, a clinical trial comparing the efficacy of AL to CQ+PQ found higher treatment efficiency in the CQ arm than the AL arm[75]. The authors posited that one reason for the reduced treatment efficacy in the AL arm was due to the short half-life of LUM, which resulted in faster *P. vivax* relapse[75]. The artemisinin component of ACTs will rapidly kill *P. vivax* parasites, while partner drugs that are eliminated from the body more slowly suppress blood-stage parasites that originate from reactivated hynozoites[76]. LUM has been shown to have the shortest period between treatment and

P. vivax relapse of all ACT partner drugs[76–78]. Therefore, *P. vivax* infected patients treated with LUM will experience relapse more quickly, usually within 20-80 days[76]. A 2-4-fold decrease in *P. vivax* susceptibility to LUM, as seen in several lines in our study, could lead to a faster time to relapse, since hypnozoites will not be suppressed as LUM concentration wanes. Decreased *P. vivax* LUM susceptibility would limit the efficacy of AL treatment. Ethiopia has used AL as the front-line *P. falciparum* treatment regimen since 2003[79]. The use of AL against *P. falciparum* cases could have co-selected for LUM resistant *pvmdr1* variants[79]. Decreased treatment efficacy of AL compared to other ACTs has been documented in other regions as well[78,80].

This work has significant implications for future *P. vivax* treatment policy, particularly considering current efforts exploring the use of different ACT regimens as a universal cure for *P. falciparum* and *P. vivax* [81]. First, in regions where ACTs are used as the front-line therapy for *P. vivax*, surveillance for the H2, H12, H16, H20, and H23 haplotypes will be critical to both monitor the effectiveness of ACTs and track the spread of drug resistance alleles. Genomic surveillance of the *pvmdr1* mutations can help guide treatment policy towards deploying ACT regimens that regional *pvmdr1* haplotypes are more susceptible to. Secondly, for regions considering recommending ACTs as *P. vivax* first-line therapy, genomic surveillance of the resistant *pvmdr1* haplotypes will also help determine which ACT regimens will be most effective in that region. We did not observe resistance to PIP and MDAQ. This evidence suggests that ACT regimens using DHA-PIP or AS-AQ could be more effective than AL for *P. vivax* treatment. Molecular surveillance of resistant *pvmdr1* alleles and other possible *P. vivax* resistance genes will be critical to effectively deploy antimalarial treatment to achieve the elimination of the parasite. Targeted sequencing approaches, such as amplicon sequencing, will help capture the emergence of resistant

haplotypes as they begin to emerge in the population, and explore the full extent of *pvmdr1* variation.

Study Limitations, Conclusions, and Future Directions

This study had some limitations. First, the *pvmdr1* population genomic dataset likely did not capture the full extent of *pvmdr1* genetic variation, because it did not contain samples from all countries with endemic *P. vivax*. Therefore, we did not evaluate the full extent of circulating *pvmdr1* variation in the *P. vivax* population. We also had only one or two samples from some countries, which resulted in a shallow sampling depth that likely does not represent the full extent of *pvmdr1* haplotypes in that country. Another potential limitation is the *P. knowlesi* model system, while a useful tool to study *P. vivax* biology, has differences between its biology and *P. vivax* biology[82]. These differences include the lack of *P. knowlesi* to form hypnozoites, the observation that not all *P. vivax* genes have 1:1 orthologs in *P. knowles*i, and different host and vector preferences[82]. Additionally, the ability of certain *pvmdr1* alleles to confer drug resistance *in vivo* may depend on the genetic background they arise on. Recapitulating the exact genetic background that resistant *pvmdr1* alleles may arise on is difficult in the *P. knowlesi* model system, since we do not know if other possible *P. vivax* resistance genes may interact with *pvmdr1* alleles to confer drug resistance. These observations may explain future possible differences between *in vitro* observations of *pvmdr1* mutations and *ex vivo* resistance observed in the field. However, despite the biological differences between *P. vivax* and *P. knowlesi,* prior work has shown that the protein products of *pvmdr1* sequences expressed in *P. knowles*i localize to the *P. knowlesi* digestive vacuole[13]. This observation suggests *P. knowlesi* is an appropriate system to characterize *pvmdr1* alleles, since *pvmdr1* is functionally similar to its *pkmdr1* ortholog. Additionally, our *P. knowlesi* transgenic lines were viable, suggesting that *pvmdr1* can fulfill the function of *pkmdr1*.

We only observed small 2-3-fold-changes, which made it difficult to accurately determine the significance of the smaller IC_{50} fold changes caused by several of the assayed *pvmdr1* haplotypes. We standardized our assay procedure to limit variability by tightly synchronizing parasites and diluting all parasites to a final parasitemia of 0.8% before plating. However, factors such as the health of parasites entering the assay, previous culturing conditions, slight variations in drug concentration during plate printing, or minor differences in the culturing conditions during the assay from media components may also explain some of this variability. Fitness differences between the lines may also contribute to the variability of the assay. However, since we conducted at least five biological replicates for each of our parasite lines and each drug, we had the statistical power to detect true differences in the IC_{50} between lines. LUM and HF particularly had a lot of variance in the assay, which may explain why the IC_{50} fold-changes of 2 or more relative to H1 for several lines were not significant. The Bonferroni-Holm correction can create a high threshold for significance and is affected by the number of pairwise comparisons. We measured significance by performing pairwise t-tests for all 30 lines (n=900 comparisons), which may have resulted in a too strict cutoff for significance after adjusted for multiple hypothesis testing. It is possible that using another multiple hypothesis testing correct method or repeating the assay for those lines could lead to fold increases in those lines being statistically significant. We chose to describe the IC_{50} fold-increase for those lines with significant increases for that reason.

Future work for this project will explore if increased copy number of different *pvmdr1* alleles in their native genetic *pvmdr1* background further increases the IC_{50} responses towards the set of antimalarials. We have selected the H1 ancestral line, the H7 line that was the most sensitive haplotype, the H2 line, which was the most prevalent resistant allele, and the H12 line, our most resistant haplotype, as the lines to insert a second copy of their *pvmdr1* allele in their respective

pvmdr1 background. We constructed transposon plasmids that contain the H1, H2, H7, and H12 *pvmdr1* haplotypes. We then transfected the H1, H2, H7, and H12 transposon plasmids containing a Blasticidin-S Deaminase resistance selection marker into the H1, H2, H7, and H12 transgenic *P. knowlesi* lines respectively to insert a second copy of that same *pvmdr1* allele into each transgenic line. We are currently using a serial dilution approach to isolate isogenic clones for each line and will assay their drug susceptibility phenotypes once this process is finished. We have also isolated H4, H5, and H13 isogenic clones and are in the process of assaying their drug susceptibilities.

We will also conduct growth fitness assays for our lines with the most sensitive (H7) and resistant (H2, H12, H16, H20, and H23) phenotypes to understand if these haplotypes result in a growth defect. We will use the method described in Small-Saunders et al[83] where we will co-culture each of the lines with a tdTomato tagged H1 as a comparison line in a 1:1 ratio. We will sample each co-culture every two days for 10-12 days and observe the percentage of tdTomato+ parasites by flow cytometry as a measure of parasite growth in competition with the control line. If the percentage of tdTomato+ parasites remain constant over the course of the assay, it would imply no growth defect. If the percentage of tdTomato+ parasites decrease, it would imply that *pvmdr1* allele has a fitness advantage. If the percentage of tdTomato+ parasites increase, it would imply that the mutant *pvmdr1 a*llele confers a growth defect. Finally, we will confirm the level of *pvmdr1* protein expression using Tandem Mass Tag Mass Spectrometry for each line.

This study successfully characterizes a large set of geographically representative *pvmdr1* haplotypes and establishes a role for increased *pvmdr1* copy number increases or overexpression in mediating drug resistance. The findings of this study can be used to support molecular surveillance efforts focused on the resistant *pvmdr1* haplotypes and *pvmdr1* copy number variation. Our work also suggests that future deployment of ACTs against *P. vivax* may not have

maximum effectiveness because years of their usage against *P. falciparum* has selected for *pvmdr1* alleles that can mediate reduced susceptibility to DHA and ACT partners MQ, LUM, and HF. Deployment of ACTs could also select for existing resistant haplotypes already present in the population. Future work is needed to associate the clinical significance of the *in vitro* increases in the DHA, MQ, LUM, and HF IC_{50}s with the risk of treatment failure. However, similar IC_{50} fold changes were observed in *P. falciparum* and linked with a significant risk of treatment failure. This observation suggests these mutations could lead to the risk of MQ or ACT treatment failure using LUM, and further work is needed to clarify if this is indeed the case. This work also supports the use of *P. knowlesi* as a model system to characterize candidate *P. vivax* resistance genes and will further support the use of this model to advance knowledge of *P. vivax* drug resistance biology. Further work is also needed to understand the transmission dynamics and general fitness of resistant vs susceptible *pvmdr1* alleles in high and low transmission settings, and understand the different genetic backgrounds that resistant *pvmdr1* alleles can arise on, Additional work could also be done to study how different *pvdmr1* alleles may affect transmission in the mosquito, given prior work suggesting that some resistant *pfmdr1* alleles can also have increased transmissibility in mosquitos[84]. Finally, we demonstrate that *pvmdr1* alleles alone do not mediate high-grade CQR, and *pvmdr1* SNPs alone should not be used as markers of CQR. Future work is needed to identify and characterize other candidate genes that mediate *P. vivax* CQR.

4.6 References

1. World malaria report 2021. https://www.who.int/publications-detail-redirect/9789240040496.
2. Price, R. N. *et al.* Vivax malaria: neglected and not benign. *Am. J. Trop. Med. Hyg.* **77**, 79–87 (2007).
3. Sharma, R., Suneja, A., Yadav, A. & Guleria, K. Plasmodium vivax Induced Acute Respiratory Distress Syndrome – A Diagnostic and Therapeutic Dilemma in Preeclampsia. *J Clin Diagn Res* **11**, QD03–QD04 (2017).
4. Val, F. *et al.* Are respiratory complications of Plasmodium vivax malaria an underestimated problem? *Malaria Journal* **16**, 495 (2017).
5. WHO. *Guidelines for the treatment of malaria - Third edition*. (2018).
6. Menkin-Smith, L. & Winders, W. T. Malaria (Plasmodium Vivax). in *StatPearls* (StatPearls Publishing, 2020).
7. Llanos-Cuentas, A. *et al.* Tafenoquine versus Primaquine to Prevent Relapse of Plasmodium vivax Malaria. *New England Journal of Medicine* **380**, 229–241 (2019).
8. Price, R. N. *et al.* Global extent of chloroquine-resistant Plasmodium vivax: a systematic review and meta-analysis. *Lancet Infect Dis* **14**, 982–991 (2014).
9. Li, J. *et al.* Ex vivo susceptibilities of Plasmodium vivax isolates from the China-Myanmar border to antimalarial drugs and association with polymorphisms in *Pvmdr1* and Pvcrt-o genes. *PLOS Neglected Tropical Diseases* **14**, e0008255 (2020).
10. Getachew, S. *et al.* Chloroquine efficacy for Plasmodium vivax malaria treatment in southern Ethiopia. *Malar J* **14**, 525 (2015).
11. Orjuela-Sánchez, P. *et al.* Analysis of single-nucleotide polymorphisms in the crt-o and mdr1 genes of Plasmodium vivax among chloroquine-resistant isolates from the Brazilian Amazon region. *Antimicrob. Agents Chemother.* **53**, 3561–3564 (2009).
12. Hupalo, D. N. *et al.* Population genomics studies identify signatures of global dispersal and drug resistance in Plasmodium vivax. *Nat. Genet.* **48**, 953–958 (2016).
13. Verzier, L. H., Coyle, R., Singh, S., Sanderson, T. & Rayner, J. C. Plasmodium knowlesi as a model system for characterising Plasmodium vivax drug resistance candidate genes. *PLOS Neglected Tropical Diseases* **13**, e0007470 (2019).
14. Sá, J. M. *et al.* Plasmodium vivax chloroquine resistance links to pvcrt transcription in a genetic cross. *Nature Communications* **10**, 1–10 (2019).
15. Veiga, M. I. *et al.* Globally prevalent PfMDR1 mutations modulate Plasmodium falciparum susceptibility to artemisinin-based combination therapies. *Nat Commun* **7**, 11553 (2016).
16. Sigala, P. A. & Goldberg, D. E. The peculiarities and paradoxes of Plasmodium heme metabolism. *Annu Rev Microbiol* **68**, 259–278 (2014).
17. Olafson, K. N., Ketchum, M. A., Rimer, J. D. & Vekilov, P. G. Mechanisms of hematin crystallization and inhibition by the antimalarial drug chloroquine. *PNAS* **112**, 4946–4951 (2015).
18. Summers, R. L. *et al.* Diverse mutational pathways converge on saturable chloroquine transport via the malaria parasite's chloroquine resistance transporter. *PNAS* **111**, E1759–E1767 (2014).
19. Barnadas, C. *et al.* Plasmodium vivax Resistance to Chloroquine in Madagascar: Clinical Efficacy and Polymorphisms in *pvmdr1* and pvcrt-o Genes. *Antimicrob Agents Chemother* **52**, 4233–4240 (2008).
20. Hamedi, Y. *et al.* Molecular Epidemiology of *P. vivax* in Iran: High Diversity and Complex

Sub-Structure Using Neutral Markers, but No Evidence of Y976F Mutation at *pvmdr1*. *PLoS ONE* **11**, e0166124 (2016).
21. Vargas-Rodríguez, R. del C. M., da Silva Bastos, M., Menezes, M. J., Orjuela-Sánchez, P. & Ferreira, M. U. Single-Nucleotide Polymorphism and Copy Number Variation of the Multidrug Resistance-1 Locus of Plasmodium vivax: Local and Global Patterns. *Am J Trop Med Hyg* **87**, 813–821 (2012).
22. Brega, S. *et al.* Identification of the Plasmodium vivax mdr-Like Gene (*pvmdr1*) and Analysis of Single-Nucleotide Polymorphisms among Isolates from Different Areas of Endemicity. *J Infect Dis* **191**, 272–277 (2005).
23. Huang, B. *et al.* Molecular surveillance of pvdhfr, pvdhps, and pvmdr-1 mutations in Plasmodium vivax isolates from Yunnan and Anhui provinces of China. *Malar J* **13**, 346 (2014).
24. Joy, S. *et al.* Drug resistance genes: pvcrt-o and pvmdr-1 polymorphism in patients from malaria endemic South Western Coastal Region of India. *Malar J* **17**, 40 (2018).
25. Suwanarusk, R. *et al.* Amplification of *pvmdr1* associated with multidrug-resistant Plasmodium vivax. *J. Infect. Dis.* **198**, 1558–1564 (2008).
26. Marfurt, J. *et al.* Molecular markers of in vivo Plasmodium vivax resistance to amodiaquine plus sulfadoxine-pyrimethamine: mutations in pvdhfr and *pvmdr1*. *J. Infect. Dis.* **198**, 409–417 (2008).
27. Sá, J. M. *et al.* Plasmodium vivax: allele variants of the mdr1 gene do not associate with chloroquine resistance among isolates from Brazil, Papua, and monkey-adapted strains. *Exp. Parasitol.* **109**, 256–259 (2005).
28. Suwanarusk, R. *et al.* Chloroquine Resistant Plasmodium vivax: In Vitro Characterisation and Association with Molecular Polymorphisms. *PLOS ONE* **2**, e1089 (2007).
29. Marques, M. M. *et al.* Plasmodium vivax chloroquine resistance and anemia in the western Brazilian Amazon. *Antimicrob. Agents Chemother.* **58**, 342–347 (2014).
30. Schousboe, M. L. *et al.* Multiple Origins of Mutations in the mdr1 Gene—A Putative Marker of Chloroquine Resistance in *P. vivax*. *PLOS Neglected Tropical Diseases* **9**, e0004196 (2015).
31. Spotin, A. *et al.* Global assessment of genetic paradigms of *Pvmdr1* mutations in chloroquine-resistant Plasmodium vivax isolates. *Trans R Soc Trop Med Hyg* **114**, 339–345 (2020).
32. Rungsihirunrat, K., Sibley, C. H., Mungthin, M. & Na-Bangchang, K. Geographical distribution of amino acid mutations in Plasmodium vivax DHFR and DHPS from malaria endemic areas of Thailand. *Am J Trop Med Hyg* **78**, 462–467 (2008).
33. Tantiamornkul, K., Pumpaibool, T., Piriyapongsa, J., Culleton, R. & Lek-Uthai, U. The prevalence of molecular markers of drug resistance in Plasmodium vivax from the border regions of Thailand in 2008 and 2014. *Int J Parasitol Drugs Drug Resist* **8**, 229–237 (2018).
34. Williams, H. A. The process of changing national malaria treatment policy: lessons from country-level studies. *Health Policy and Planning* **19**, 356–370 (2004).
35. Rottmann, M. *et al.* Spiroindolones, a Potent Compound Class for the Treatment of Malaria. *Science* **329**, 1175–1180 (2010).
36. Baragaña, B. *et al.* A novel multiple-stage antimalarial agent that inhibits protein synthesis. *Nature* **522**, 315–320 (2015).
37. McNamara, C. W. *et al.* Targeting Plasmodium PI(4)K to eliminate malaria. *Nature* **504**, 248–253 (2013).

38. Reed, M. B., Saliba, K. J., Caruana, S. R., Kirk, K. & Cowman, A. F. Pgh1 modulates sensitivity and resistance to multiple antimalarials in Plasmodium falciparum. *Nature* **403**, 906–909 (2000).
39. Sidhu, A. B. S., Valderramos, S. G. & Fidock, D. A. pfmdr1 mutations contribute to quinine resistance and enhance mefloquine and artemisinin sensitivity in Plasmodium falciparum. *Molecular Microbiology* **57**, 913–926 (2005).
40. Sidhu, A. B. S., Verdier-Pinard, D. & Fidock, D. A. Chloroquine resistance in Plasmodium falciparum malaria parasites conferred by pfcrt mutations. *Science* **298**, 210–213 (2002).
41. Buyon, L. E., Elsworth, B. & Duraisingh, M. T. The molecular basis of antimalarial drug resistance in Plasmodium vivax. *International Journal for Parasitology: Drugs and Drug Resistance* **16**, 23–37 (2021).
42. Mohring, F. *et al.* Rapid and iterative genome editing in the malaria parasite Plasmodium knowlesi provides new tools for *P. vivax* research. *Elife* **8**, (2019).
43. Hastings, M. D. & Sibley, C. H. Pyrimethamine and WR99210 exert opposing selection on dihydrofolate reductase from Plasmodium vivax. *PNAS* **99**, 13137–13141 (2002).
44. Hastings, M. D. *et al.* Novel Plasmodium vivax dhfr alleles from the Indonesian Archipelago and Papua New Guinea: association with pyrimethamine resistance determined by a Saccharomyces cerevisiae expression system. *Antimicrob Agents Chemother* **49**, 733–740 (2005).
45. Moon, R. W. *et al.* Adaptation of the genetically tractable malaria pathogen Plasmodium knowlesi to continuous culture in human erythrocytes. *Proc Natl Acad Sci U S A* **110**, 531–536 (2013).
46. Lim, C. *et al.* Expansion of host cellular niche can drive adaptation of a zoonotic malaria parasite to humans. *Nat Commun* **4**, 1638 (2013).
47. van Schalkwyk, D. A., Moon, R. W., Blasco, B. & Sutherland, C. J. Comparison of the susceptibility of Plasmodium knowlesi and Plasmodium falciparum to antimalarial agents. *J. Antimicrob. Chemother.* **72**, 3051–3058 (2017).
48. Hupalo, D. N. *et al.* Population genomics studies identify signatures of global dispersal and drug resistance in Plasmodium vivax. *Nat. Genet.* **48**, 953–958 (2016).
49. Pearson, R. D. *et al.* Genomic analysis of local variation and recent evolution in Plasmodium vivax. *Nat Genet* **48**, 959–964 (2016).
50. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
51. Broad Institute. *PicardTools*. (Broad Institute).
52. Scaling accurate genetic variant discovery to tens of thousands of samples | bioRxiv. https://www.biorxiv.org/content/10.1101/201178v2.
53. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
54. Knaus, B. J. & Grünwald, N. J. vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources* **17**, 44–53 (2017).
55. Wickham, Hadley. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).
56. Aurrecoechea, C. *et al.* PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Research* **37**, D539–D543 (2009).
57. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797 (2004).

58. Verzier, L. H., Coyle, R., Singh, S., Sanderson, T. & Rayner, J. C. Plasmodium knowlesi as a model system for characterising Plasmodium vivax drug resistance candidate genes. *PLOS Neglected Tropical Diseases* **13**, e0007470 (2019).
59. Asih, P. B. S. *et al.* Distribution of Plasmodium vivax pvdhfr and pvdhps alleles and their association with sulfadoxine–pyrimethamine treatment outcomes in Indonesia. *Malaria Journal* **14**, 365 (2015).
60. Suwanarusk *et al.* Chloroquine Resistant Plasmodium vivax: In Vitro Characterisation and Association with Molecular Polymorphisms. *PLOS ONE* **2**, e1089 (2007).
61. Wurtz, N. *et al.* Role of Pfmdr1 in In Vitro Plasmodium falciparum Susceptibility to Chloroquine, Quinine, Monodesethylamodiaquine, Mefloquine, Lumefantrine, and Dihydroartemisinin. *Antimicrobial Agents and Chemotherapy* **58**, 7032–7040 (2014).
62. Sidhu, A. B. S. *et al.* Decreasing pfmdr1 Copy Number in Plasmodium falciparum Malaria Heightens Susceptibility to Mefloquine, Lumefantrine, Halofantrine, Quinine, and Artemisinin. *J Infect Dis* **194**, 528–535 (2006).
63. Duraisingh, M. T. *et al.* The tyrosine-86 allele of the pfmdr1 gene of Plasmodium falciparum is associated with increased sensitivity to the anti-malarials mefloquine and artemisinin. *Mol Biochem Parasitol* **108**, 13–23 (2000).
64. Venkatesan, M. *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* **91**, 833–843 (2014).
65. Picot, S. *et al.* A systematic review and meta-analysis of evidence for correlation between molecular markers of parasite resistance and treatment outcome in falciparum malaria. *Malar J* **8**, 89 (2009).
66. Cowman, A. F., Galatis, D. & Thompson, J. K. Selection for mefloquine resistance in Plasmodium falciparum is linked to amplification of the pfmdr1 gene and cross-resistance to halofantrine and quinine. *PNAS* **91**, 1143–1147 (1994).
67. Dama, S. *et al.* Reduced ex vivo susceptibility of Plasmodium falciparum after oral artemether–lumefantrine treatment in Mali. *Malaria Journal* **16**, 59 (2017).
68. Zeile, I. *et al.* Molecular markers of Plasmodium falciparum drug resistance in southern highland Rwanda. *Acta Tropica* **121**, 50–54 (2012).
69. Baraka, V. *et al.* In Vivo Selection of Plasmodium falciparum Pfcrt and Pfmdr1 Variants by Artemether-Lumefantrine and Dihydroartemisinin-Piperaquine in Burkina Faso. *Antimicrob Agents Chemother* **59**, 734–737 (2015).
70. Malmberg, M. *et al.* Plasmodium falciparum Drug Resistance Phenotype as Assessed by Patient Antimalarial Drug Levels and Its Association With pfmdr1 Polymorphisms. *The Journal of Infectious Diseases* **207**, 842–847 (2013).
71. Duraisingh, M. T. & Cowman, A. F. Contribution of the pfmdr1 gene to antimalarial drug-resistance. *Acta Tropica* **94**, 181–190 (2005).
72. Price, R. N. *et al.* Mefloquine resistance in Plasmodium falciparum and increased pfmdr1 gene copy number. *Lancet* **364**, 438–447 (2004).
73. Humphreys, G. S. *et al.* Amodiaquine and Artemether-Lumefantrine Select Distinct Alleles of the Plasmodium falciparum mdr1 Gene in Tanzanian Children Treated for Uncomplicated Malaria. *Antimicrob Agents Chemother* **51**, 991–997 (2007).
74. Suwanarusk *et al.* Amplification of *pvmdr1* associated with multidrug-resistant Plasmodium vivax. *J. Infect. Dis.* **198**, 1558–1564 (2008).

75. Abreha, T. *et al.* Comparison of artemether-lumefantrine and chloroquine with and without primaquine for the treatment of Plasmodium vivax infection in Ethiopia: A randomized controlled trial. *PLOS Medicine* **14**, e1002299 (2017).
76. White, N. J. Anti-malarial drug effects on parasite dynamics in vivax malaria. *Malaria Journal* **20**, 161 (2021).
77. White, N. J. Determinants of relapse periodicity in Plasmodium vivax malaria. *Malaria Journal* **10**, 297 (2011).
78. Commons, R. J. *et al.* The efficacy of dihydroartemisinin-piperaquine and artemether-lumefantrine with and without primaquine on Plasmodium vivax recurrence: A systematic review and individual patient data meta-analysis. *PLOS Medicine* **16**, e1002928 (2019).
79. Gebreyohannes, E. A., Bhagavathula, A. S., Seid, M. A. & Tegegn, H. G. Anti-malarial treatment outcomes in Ethiopia: a systematic review and meta-analysis. *Malaria Journal* **16**, 269 (2017).
80. Bassat, Q. The Use of Artemether-Lumefantrine for the Treatment of Uncomplicated Plasmodium vivax Malaria. *PLoS Negl Trop Dis* **5**, e1325 (2011).
81. Price, R. N., Commons, R. J., Battle, K. E., Thriemer, K. & Mendis, K. Plasmodium vivax in the Era of the Shrinking *P. falciparum* Map. *Trends in Parasitology* **36**, 560–570 (2020).
82. Grüring, C. *et al.* Human red blood cell-adapted Plasmodium knowlesi parasites: a new model system for malaria research. *Cell. Microbiol.* **16**, 612–620 (2014).
83. Small-Saunders, J. L. *et al.* Evidence for the early emergence of piperaquine-resistant Plasmodium falciparum malaria and modeling strategies to mitigate resistance. *PLOS Pathogens* **18**, e1010278 (2022).
84. Hallett, R. L. *et al.* Combination Therapy Counteracts the Enhanced Transmission of Drug-Resistant Malaria Parasites to Mosquitoes. *Antimicrob Agents Chemother* **48**, 3940–3943 (2004).

Chapter Five: Discussion

5.1: Genomic epidemiology tools will be critical to achieve *P. vivax* elimination and to study *P. vivax* biology

There has been a significant reduction in *P. vivax* cases in recent years, raising the possibility of sustained *P. vivax* elimination in several parts of the world[1]. Genomic epidemiology will be a critical tool to achieve that goal by identifying threats to elimination from imported cases, identifying and tracking drug resistance, and identifying local sources and sinks of *P. vivax* transmission[2–4].

In **Chapter Two,** I demonstrate how advances in the ability to conduct whole genome sequencing from clinical *P. vivax* isolates and genomic epidemiology tools can be applied to assist elimination efforts. I used population genomic and phylogenetic techniques to identify the evolutionary relationship between *P. vivax* samples collected in Panama and *P. vivax* strains worldwide. This approach identified several imported *P. vivax* samples and their general region or country of origin. We did not have the sampling depth worldwide in our dataset to obtain the resolution to pinpoint the exact country for all imported samples. However, this work illustrates how even limited genomic can still be informative by simply identifying cases as imported or not. Identifying imported cases will be critical to achieve elimination by identifying where and how *P. vivax* cases are imported. Future applications of phylogenetics to identify imported cases could take advantage of deeper sampling of *P. vivax* populations worldwide to power phylogenetic epidemiological approaches. A globally representative set of *P. vivax* samples could be used to pinpoint the country or region of origin of imported *P. vivax* cases. For example, we were unable to identify the country of origin for the CL2 lineage, but found it clustered with samples from Colombia. It is possible that future re-analyses of this data with a more globally representative *P. vivax* sample set will better identify their country of origin. Additionally, future genomic epidemiological analyses that compare their samples to a globally

representative *P. vivax* dataset and estimate recent common ancestry using identity-by-descent (IBD) could highlight the region of origin of imported cases and how relatively recently that parasite lineage was imported. These tools will help identify sources of parasite importation and direct public health efforts to staunch their transmission to help achieve elimination.

Furthermore, IBD analysis can illuminate patterns of transmission in a country. In **Chapter Two**, I found that there was one single clonal lineage transmitted throughout Panama. This finding was surprising, given prior work studying *P. falciparum* population structure that showed distinct parasite populations on either side of the Panama Canal[5]. The finding suggests that *P. vivax* is transmitted throughout the country, rather than fracturing into sub populations as might be expected in a low transmission setting[6,7]. This may be because *P. vivax* hypnozoites can reactivate in persons after they have traveled away from the place they were originally infected[8]. The analysis in **Chapter Two** informs the malaria elimination strategy in Panama by supporting the continued need for country-wide surveillance and elimination efforts, instead of efforts targeted on localities with the highest cases. There is another epidemiolocal scenario where IBD could highlight distinct parasite subpopulations transmitted locally. This scenario would support targeted elimination efforts concentrated on those localities. Future uses of IBD analysis can be used to inform malaria control and elimination strategies by inferring transmission dynamics, sources and sinks of parasite transmission, and population structure to best deploy public health resources.

Genomic epidemiology can also be used to identify and track the spread of *P. vivax* drug resistance. Prior to this work, only mutations in *pvdhfr* were validated *P. vivax* markers of drug resistance. This information was not actionable for *P. vivax* control programs, because sulfadoxine-pyrimethamine (SP) had never been recommended for *P. vivax* treatment.

Widespread chloroquine resistant (CQR) *P. vivax* in Papua New Guinea and Indonesia has necessitated a change to use ACTs[9]. Other countries also have evidence of CQR *P. vivax*[9–11]. *pvmdr1* mutations have been used as molecular markers of CQR, despite not being validated in a heterologous model system[12–14]. Use of a not validated molecular marker of drug resistance could lead to unwarranted changes in treatment policy to more expensive ACTs [15]. Work in **Chapter Four** demonstrates that *pvmdr1* does not mediate CQR but validates that it confers reduced susceptibility to dihydroartemisinin (DHA), lumefantrine (LUM), Mefloquine (MQ), and Halofantrine (HF) *in vitro*. This work also highlights several drug resistant *pvmdr1* haplotypes. The work in **Chapter Four** suggests that Artemether plus Lumefantrine (AL) treatment is likely to be less effective than ACTs with Amodiaquine or Piperaquine as the partner drugs. Targeted molecular surveillance of these drug resistant *pvmdr1* haplotypes can be deployed to identify if there is resistance to these compounds and support which ACTs to deploy. The work in **Chapter Four** demonstrates that *pvmdr1* does not mediate high-grade CQR and should not be used as a marker of CQR. Further work is needed to find a validated marker of CQR in *P. vivax*. Identifying a molecular marker of CQR will be critical to support *P. vivax* control and elimination, and to deploy effective treatments.

The decreasing cost and continued ease of deploying sequencing technologies has great potential to support public health efforts combating malaria and other infectious diseases[2,3]. These technological advances could be used to support future cloud-based genomic epidemiology platforms, where a pathogen sequence or set of sequences are uploaded and varied analyses could be carried out on them. I present a use case and analysis workflow in **Chapter Two** that could be used in computational genomic pipelines to rapidly identify if cases are either locally transmitted or imported, and if imported, identify the region of origin. Additionally, these

platforms could identify if parasites carry drug resistance alleles, such as the *pvmdr1* drug resistant haplotypes discussed in **Chapter Four**. This capability will allow for identification of drug resistant parasites, the ability to track the spread of drug resistance, and provide information to recommend appropriate treatment policy. Deploying treatment regimens that drug resistant parasites are still susceptible to will be critical to achieve *P. vivax* elimination. These future platforms will be an important tool to translate genomic epidemiology information into epidemiological intelligence that can be used to support malaria elimination efforts.

Lastly, **Chapter Two** illustrates the potential of selective whole genome amplification (SWGA) to amplify and sequence *P. vivax* from clinical isolates. *P. vivax* can be difficult to sequence directly from clinical samples because of its low parasitemia, which can make isolating *P. vivax* DNA in sufficient quantities for sequencing difficult[16]. Methods such as hybrid-capture and leukocyte depletion can enrich *P. vivax* DNA concentration in samples with low parasitemia, but such methods are expensive and difficult to scale[16,17]. SWGA has enabled higher coverage sequencing of *P. vivax* samples directly from clinical samples, and is more easily scalable and cheaper compared to the aforementioned methods [16]. SWGA can enable genome-wide association studies and identify SNPs associated with host tropism or for drug resistance phenotypes. SWGA could also be applied for genomic surveillance efforts and provide a rich dataset to study relatedness and evolution of *P. vivax* in various transmission settings. However, whole-genome data might not be needed to study parasite relatedness and perform public health surveillance. Amplicon panels, which enable high sequence coverage of select loci, can also be used to achieve these public health surveillance goals as well[18–21]. Continued use and development of SWGA and amplicon panels for application for *P. vivax* genomic surveillance is an exciting and impactful step forward for malaria genomic epidemiology.

5.2 Using Population Genomics, Evolutionary Analysis and Heterologous Genetic Systems Can Allow the Characterization and Validation *of P. vivax* Drug Resistance Genes.

This dissertation provides the first molecular genetic validation of a *P. vivax* drug resistance against antimalarial regimens used to treat *P. vivax* in some regions. The work in **Chapter Four** demonstrates the use of allelic replacement in *P. knowlesi* of a native *P. knowlesi* gene with its *P. vivax* ortholog to characterize its ability to mediate drug resistance. Three resistant *pvmdr1* haplotypes, H2, H12, and H20, are present throughout Southeast Asia. The H2 haplotype, which is a single tyrosine to phenylalanine substitution at amino acid position 976 in the ancestral background, resulted in a 2-fold shift in the MQ, LF, and HF IC_{50}s. The H12 haplotype is a double mutant with leucine to phenylalanine substitutions at positions 845 and 1076. The H20 haplotype is a triple mutant with a serine to arginine mutation at position 513, a leucine to phenylalanine substitution at positions 845 and 1076, and a lysine to asparagine mutation at position 1393. Both haplotypes resulted in 1.5-fold shifts in the DHA IC_{50}, 2-fold shifts for the MQ and HF IC_{50}, and 2-2.5-fold shifts in the LF IC_{50}. We also observed that our *pvmdr1* copy number variation line results in 2-fold shifts in IC_{50}s of MQ, HF, and a ~4-fold in LF IC_{50}. The H17 and H23 (Sal-1 genetic background) haplotypes are found in South America and displayed significantly reduced susceptibility to DHA. The molecular validation that *pvmdr1* haplotypes and copy-number variation can confer reduced drug susceptibility can be used for genomic surveillance of *P. vivax* drug resistance going forward.

This work builds upon previous studies to validate *P. knowlesi* as a model system to characterize *P. vivax* gene function[22,23]. We demonstrate that we can use transgenic *P. knowlesi* lines expressing *P. vivax* genes to reliably characterize those alleles with regards to drug resistance phenotypes. This work helps build a technical foundation for further research using *P. knowlesi* to study the *P. vivax* drug resistance biology. However, model systems for drug

resistance validation are only useful if there are candidate genes to characterize. The work **in Chapter Three** has generated a set of candidate drug resistance genes uniquely under selection pressure in *P. vivax.* I discuss in this chapter why one of the candidate genes, *plasmepsinIV*, is a particularly compelling candidate for characterization, and outline an experimental plan to study if it confers a drug resistance phenotype. The work in this thesis provides an outline of what candidate drug resistance genes to prioritize for functional validation, and the experimental plan and tools to do so in *P. knowlesi*. Future studies can follow the analysis outline discussed in **Chapter Three** to conduct comparative selection scans using more geographically representative *P. vivax* sample sets and identify other compelling candidate drug resistance loci for validation. Understanding of *P. vivax* biology can be rapidly advanced by pairing population genomic and evolutionary analyses with the *P. knowlesi* model system for functional characterization.

5.3 Future Advances in Heterologous Genetic Systems and their Application to Studying *P. vivax* Biology

Further development of the *P. knowlesi* model system to study *P. vivax* will help elucidate the biology of this parasite of public health importance. *P. knowlesi* has been used in other studies to study *P. vivax* genes involved in invasion, and to screen blood stage antigens that could be used as the basis of vaccine development[24,25]. *P. knowlesi* is highly genetically tractable, making it a potent genetic system to characterize *P. vivax* genes[25–27]. There are only two selection markers, hDHFR and blasticidin (BSD), that currently work in *P. knowlesi*. This limits the ability to study the interaction of *P. vivax* genes and how their combinations might affect phenotypes, because we can only express two different *P. vivax* genes at a time in *P. knowlesi* and select for using drugs. The ability to express multiple candidate drug resistance *P. vivax* genes would be a valuable tool to understand the genetic backgrounds where high-grade

chloroquine resistance and drug resistance can arise. Development and validation of selection markers for use in *P. knowlesi* will allow the study of *P. vivax* genetic interactions.

Future work in *P. knowlesi* should take advantage of the genetic tractability of the parasite to conduct pooled screens for phenotypes. Our approach to studying the large array of haplotypes in *pvmdr1* may have been more efficiently studied using a pooled phenotyping approach that utilizes molecular barcodes for each haplotype. In this method, we would have added a unique DNA barcode into each *pvmdr1* haplotype plasmid. We then could have pooled the 27 haplotypes and assayed them after exposure to different concentrations of a compound(s) of interest. We next would collect DNA of the parasite pool after drug exposure for 1-2 *P. knowlesi* life cycles (~36-48 hours) and use next generation sequencing to deconvolute from the pool to identify the relative proportion of parasites lines that are represented in the surviving mixture of parasites. This approach could rapidly screen for resistant lines and select those lines that appear resistant for further characterization of their IC_{50}s in a drug susceptibility assay. This approach could be useful for highly polymorphic drug resistance genes, like *pvmdr1* or *pvmrp1*, to identify drug resistant haplotypes. Beyond applications to studying drug resistance, the development of a pooled barcoding system for *P. knowlesi* could be used to screen for antigens for vaccine development, cell invasion genes to understand host tropism and invasion pathways, and other phenotypes of interest. Coupling the genetic tractability of *P. knowlesi* with pooled phenotype approaches will give the *P. vivax* field a power genetic system for functional characterization.

Recent development of a *P. cynomolgi* continuous culture system also opens further avenues to explore *P. vivax* biology. *P. cynomolgi* is much more closely related to *P. vivax* than P. *knowlesi*, meaning it likes shares more conserved biology with *P. vivax*. One critical

advantage of *P. cynomolgi* over *P. knowlesi* as a model system is that *P cynomolgi* can form hypnozoites. The ability of *P. cynomolgi* to form hypnozoites will allow for the screening of new compounds that target *P. vivax* hypnozoites and enable radical cure. Identifying hypnozoite targeting drugs is important, because it is difficult to achieve radical cure using CQ, MQ, or ACTs without primaquine (PQ) or tafenoquine (TQ). However, these drugs can cause hemolysis in patients G6PD deficiency[28,29] precluding their use in many *P. vivax* endemic regions. As such, successful *P. vivax* elimination will be difficult if there is no drug that can kill hypnozoites without causing hemolysis in G6PD deficient patients[30]. *P. cynomolgi* holds great promise to function as a screening platform for compounds that target the hypnozoite stage[31]. Further development of *P. cynomolgi* CRISPR/Cas9 genetic systems and episomal overexpression systems can allow the screening of *P. vivax* vaccine candidate antigens, drug resistance genes, and invasion genes in a species that more closely resembles *P. vivax* biology. *P cynomolgi* may be less genetically tractable than P. *knowlesi,* but its shared biology will be a compelling additional heterologous model system to study *P. vivax* biology[32].

5.4 Concluding Remarks

The aim of this dissertation was to propel understanding of how genomic epidemiology could help with *P. vivax* elimination efforts, and to provide a blueprint on how to both identify and prioritize candidate drug resistance genes, and to validate those genes so that they can be used for future genomic epidemiology and molecular studies. This thesis builds upon other prior work applying population genomic tools to aid malaria elimination and demonstrates a use case towards applying these tools to help aid elimination programs reach their endgame. Moreover, this thesis provides the first molecular genetic confirmation of a *P. vivax* drug resistance gene, *pvmdr1*, using a *P. knowlesi* model system. This finding is a critical advancement for the field

for several reasons. First, it demonstrates that, contrary to previous suspicions, *pvmdr1* does not mediate *P. vivax* chloroquine resistance (CQR). This is important because prior studies have used the existence of mutations in *pvmdr1* to suggest that there is CQR *P. vivax* circulating in those regions and recommended a possible change in treatment policy. This work demonstrates that *pvmdr1* does not result in high-grade CQR, meaning it is unwarranted to change *P. vivax* treatment policy based on *pvmdr1* mutations. It is possible that there could be linkage of *pvmdr1* with the gene that mediates *P. vivax* CQR, and this scenario warrants further study. This work provides evidence that *pvmdr1* mediate resistance to MQ, LF, HF, and to a lower extent, DHA. This finding is clinically relevant because of considerations shifting treatment policy towards use of ACTs either in regions where there is widespread *P. vivax* CQR or as a universal cure approach where ACTs are the recommended treatment for both *P. vivax* and *P. falciparu*m. This finding suggests that the artemether-lumefantrine (AL) ACT may not be as effective, given that three of our *pvmdr1* haplotypes mediate 2-4-fold shifts in the LUM IC_{50}. The findings in this thesis also support molecular surveillance efforts focused on *pvmdr1* mutations to monitor the risk of partner drug resistance. Additionally, this finding supports the use of DHA-PPQ or AS-AQ to treat *P. vivax* if a switch to an ACT regimen is required. Lastly, this work confirms the utility of P. *knowlesi* as a model system to study *P. vivax* drug resistance. This confirmation, coupled with the list of candidate resistance genes in **Chapter Three,** can set the foundation for future efforts to study *P. vivax* drug resistance.

The work in this thesis, coupled with experimental work in progress at the time of writing, will empower future efforts to elucidate the drug resistance biology of *P. vivax and* provide the field with tools to track its spread. This dissertation will help advance the study of *P.*

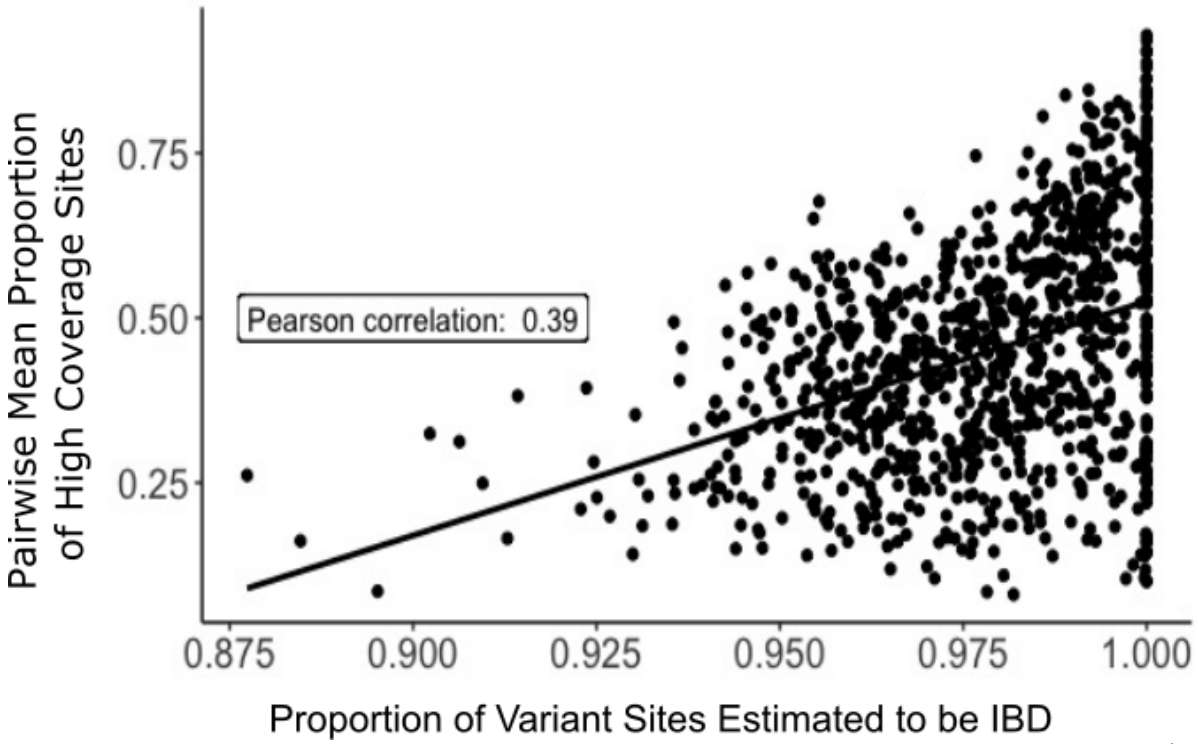
vivax, which supports its successful elimination through knowledge of drug resistant alleles and the deployment of genomic tools to track its transmission.

5.5: References

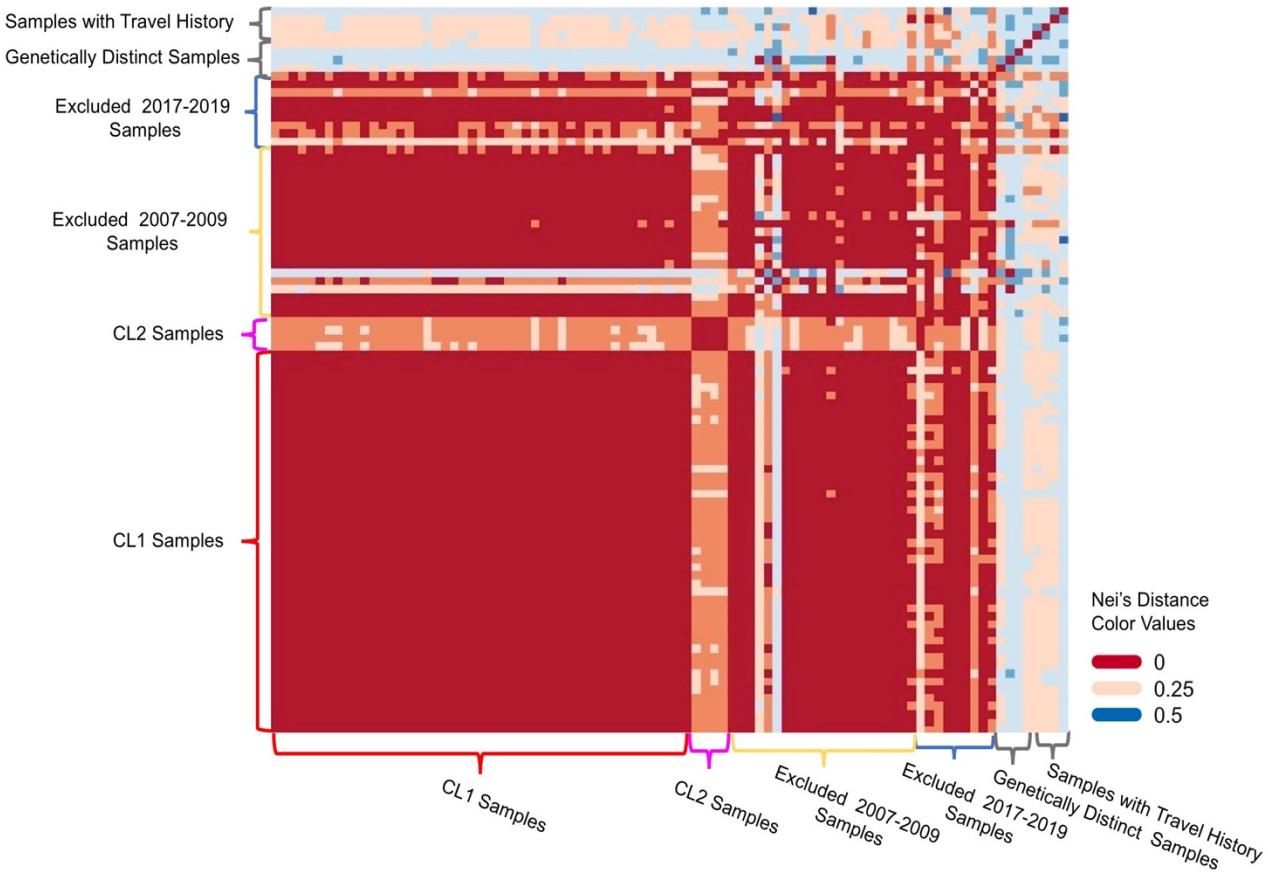
1. World malaria report 2021. https://www.who.int/publications-detail-redirect/9789240040496.
2. Neafsey, D. E. & Volkman, S. K. Malaria Genomics in the Era of Eradication. *Cold Spring Harb Perspect Med* **7**, a025544 (2017).
3. Neafsey, D. E., Taylor, A. R. & MacInnis, B. L. Advances and opportunities in malaria population genomics. *Nat. Rev. Genet.* **22**, 502–517 (2021).
4. Buyon, L. E. *et al.* Population genomics of Plasmodium vivax in Panama to assess the risk of case importation on malaria elimination. *PLoS Negl. Trop. Dis.* **14**, e0008962 (2020).
5. Obaldia, N. *et al.* Clonal outbreak of Plasmodium falciparum infection in eastern Panama. *J Infect Dis* **211**, 1087–1096 (2015).
6. Chenet, S. M., Schneider, K. A., Villegas, L. & Escalante, A. A. Local population structure of Plasmodium: impact on malaria control and elimination. *Malar. J.* **11**, 412 (2012).
7. Dalmat, R., Naughton, B., Kwan-Gett, T. S., Slyker, J. & Stuckey, E. M. Use cases for genetic epidemiology in malaria elimination. *Malar. J.* **18**, 163 (2019).
8. Shanks, G. D. Control and elimination of Plasmodium vivax. *Adv. Parasitol.* **80**, 301–341 (2012).
9. Price, R. N. *et al.* Global extent of chloroquine-resistant Plasmodium vivax: a systematic review and meta-analysis. *Lancet Infect Dis* **14**, 982–991 (2014).
10. Melo, G. C. *et al.* Expression levels of pvcrt-o and pvmdr-1 are associated with chloroquine resistance and severe Plasmodium vivax malaria in patients of the Brazilian Amazon. *PLoS ONE* **9**, e105922 (2014).
11. Getachew, S. *et al.* Chloroquine efficacy for Plasmodium vivax malaria treatment in southern Ethiopia. *Malar J* **14**, 525 (2015).
12. Marfurt, J. *et al.* Molecular markers of in vivo Plasmodium vivax resistance to amodiaquine plus sulfadoxine-pyrimethamine: mutations in pvdhfr and *pvmdr1*. *J Infect Dis* **198**, 409–417 (2008).
13. Rungsihirunrat, K., Muhamad, P., Chaijaroenkul, W., Kuesap, J. & Na-Bangchang, K. Plasmodium vivax Drug Resistance Genes; *Pvmdr1* and Pvcrt-o Polymorphisms in Relation to Chloroquine Sensitivity from a Malaria Endemic Area of Thailand. *Korean J. Parasitol.* **53**, 43–49 (2015).
14. Tantiamornkul, K., Pumpaibool, T., Piriyapongsa, J., Culleton, R. & Lek-Uthai, U. The prevalence of molecular markers of drug resistance in Plasmodium vivax from the border regions of Thailand in 2008 and 2014. *Int J Parasitol Drugs Drug Resist* **8**, 229–237 (2018).
15. Williams, H. A. The process of changing national malaria treatment policy: lessons from country-level studies. *Health Policy Plan.* **19**, 356–370 (2004).
16. Cowell, A. N. *et al.* Selective Whole-Genome Amplification Is a Robust Method That Enables Scalable Whole-Genome Sequencing of Plasmodium vivax from Unprocessed Clinical Samples. *mBio* **8**, (2017).
17. Melnikov, A. *et al.* Hybrid selection for sequencing pathogen genomes from clinical samples. *Genome Biol* **12**, R73 (2011).
18. Gruenberg, M., Lerch, A., Beck, H.-P. & Felger, I. Amplicon deep sequencing improves Plasmodium falciparum genotyping in clinical trials of antimalarial drugs. *Sci. Rep.* **9**, 17790 (2019).
19. Lin, J. T. *et al.* Using Amplicon Deep Sequencing to Detect Genetic Signatures of Plasmodium vivax Relapse. *J Infect Dis* **212**, 999–1008 (2015).
20. Boyce, R. M. *et al.* Reuse of malaria rapid diagnostic tests for amplicon deep sequencing to estimate Plasmodium falciparum transmission intensity in western Uganda. *Sci. Rep.* **8**, 10159 (2018).
21. LaVerriere, E. *et al.* Design and implementation of multiplexed amplicon sequencing panels to serve genomic epidemiology of infectious disease: a malaria case study. 2021.09.15.21263521 (2021) doi:10.1101/2021.09.15.21263521.
22. Verzier, L. H., Coyle, R., Singh, S., Sanderson, T. & Rayner, J. C. Plasmodium knowlesi as a model system for characterizing Plasmodium vivax drug resistance candidate genes. *PLoS Negl. Trop. Dis.* **13**, e0007470 (2019).

23. van Schalkwyk, D. A., Moon, R. W., Blasco, B. & Sutherland, C. J. Comparison of the susceptibility of Plasmodium knowlesi and Plasmodium falciparum to antimalarial agents. *J Antimicrob Chemother* **72**, 3051–3058 (2017).
24. Ndegwa, D. N. *et al.* Using Plasmodium knowlesi as a model for screening Plasmodium vivax blood-stage malaria vaccine targets reveals new candidates. *PLOS Pathog.* **17**, e1008864 (2021).
25. Mohring, F. *et al.* Rapid and iterative genome editing in the malaria parasite Plasmodium knowlesi provides new tools for *P. vivax* research. *Elife* **8**, (2019).
26. Moon, R. W. *et al.* Adaptation of the genetically tractable malaria pathogen Plasmodium knowlesi to continuous culture in human erythrocytes. *Proc Natl Acad Sci U A* **110**, 531–536 (2013).
27. Lim, C. *et al.* Expansion of host cellular niche can drive adaptation of a zoonotic malaria parasite to humans. *Nat Commun* **4**, 1638 (2013).
28. Baird, Battle, K. E. & Howes, R. E. Primaquine ineligibility in anti-relapse therapy of Plasmodium vivax malaria: the problem of G6PD deficiency and cytochrome P-450 2D6 polymorphisms. *Malar J* **17**, (2018).
29. Chu, C. S. *et al.* Haemolysis in G6PD Heterozygous Females Treated with Primaquine for Plasmodium vivax Malaria: A Nested Cohort in a Trial of Radical Curative Regimens. *PLoS Med* **14**, (2017).
30. Thriemer, K., Ley, B. & Seidlein, L. von. Towards the elimination of Plasmodium vivax malaria: Implementing the radical cure. *PLOS Med.* **18**, e1003494 (2021).
31. Deye, G. A. *et al.* Use of a rhesus Plasmodium cynomolgi model to screen for anti-hypnozoite activity of pharmaceutical substances. *Am. J. Trop. Med. Hyg.* **86**, 931–935 (2012).
32. Chua, A. C. Y. *et al.* Robust continuous in vitro culture of the Plasmodium cynomolgi erythrocytic stages. *Nat Commun* **10**, 3635 (2019).

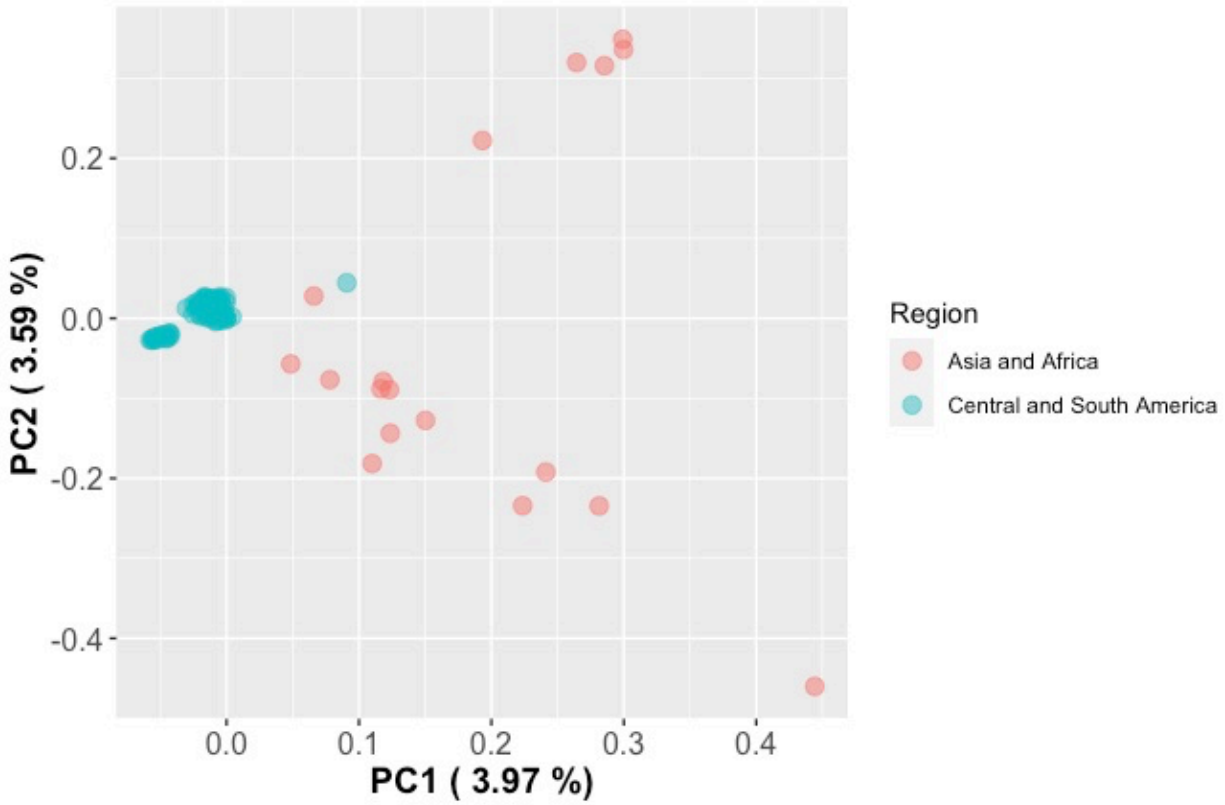
Appendix



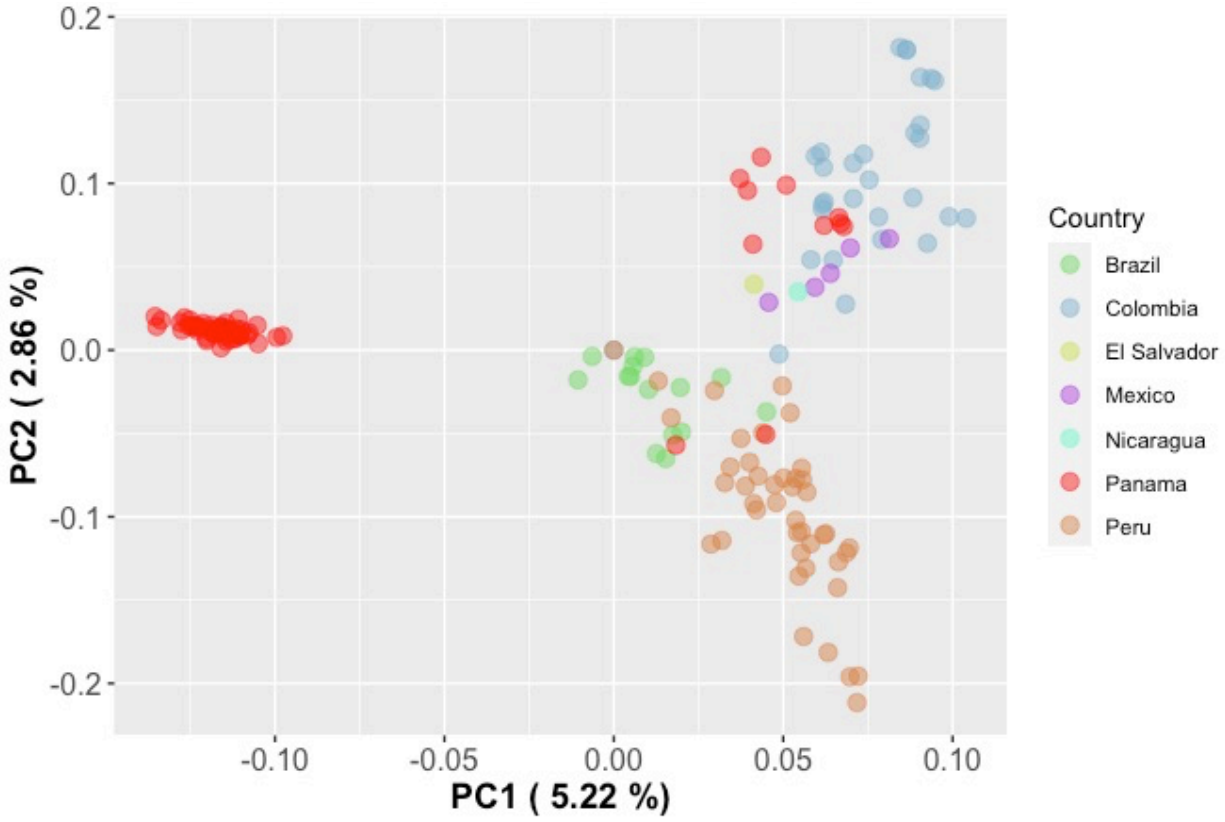
Supplemental Figure 2.1: Pairwise IBD Estimates Increase with Sample Quality. Depicts pairwise IBD estimates for all Panamanian sample pairs with IBD > 0.875 plotted against the mean proportion of high coverage sites (sites with > 5x coverage) in each sample pair. The line indicates a linear regression the box displays the Pearson correlation coefficient between the two axes variables



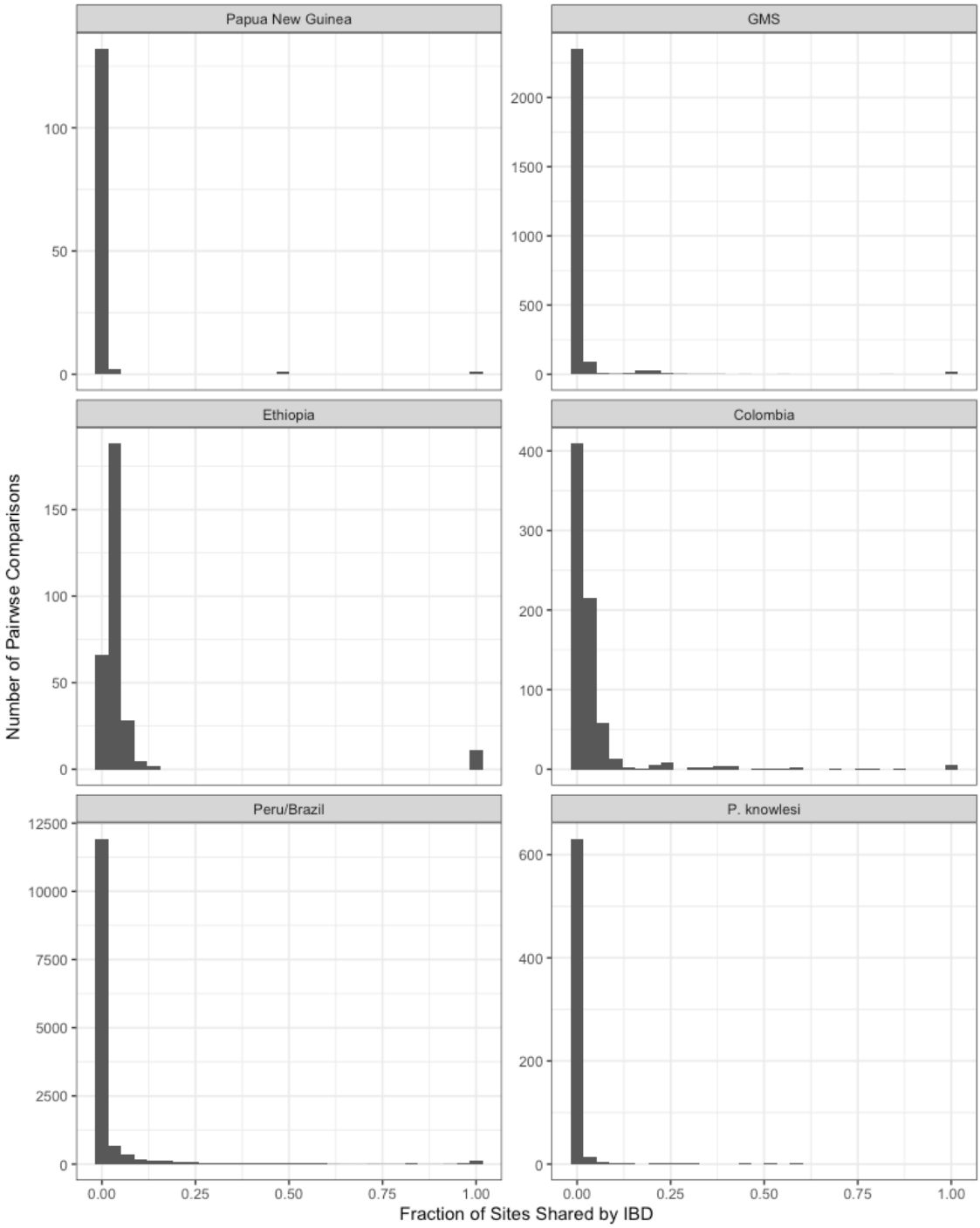
Supplemental Figure 2.2: Annotated heatmap of pairwise Nei's standard distance comparisons between all 2007-2009 and 2017-2019 samples using SNPs that were callable in at least 80% of samples. Each block row and column present a single sample. Brackets indicate sample groups.



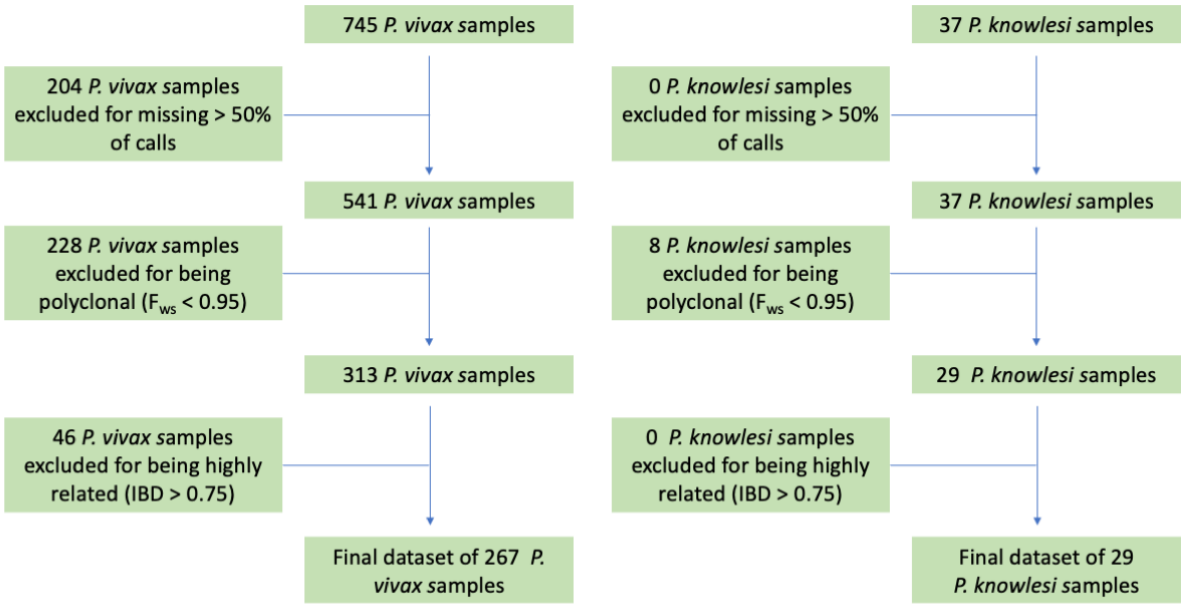
Supplemental Figure 2.3: Principal components analysis of Panama samples and previously collected samples from Central and South America, Asia, and Africa. Samples are colored by the region of origin. Parentheses contain the percentage of variance explained by each principal component.



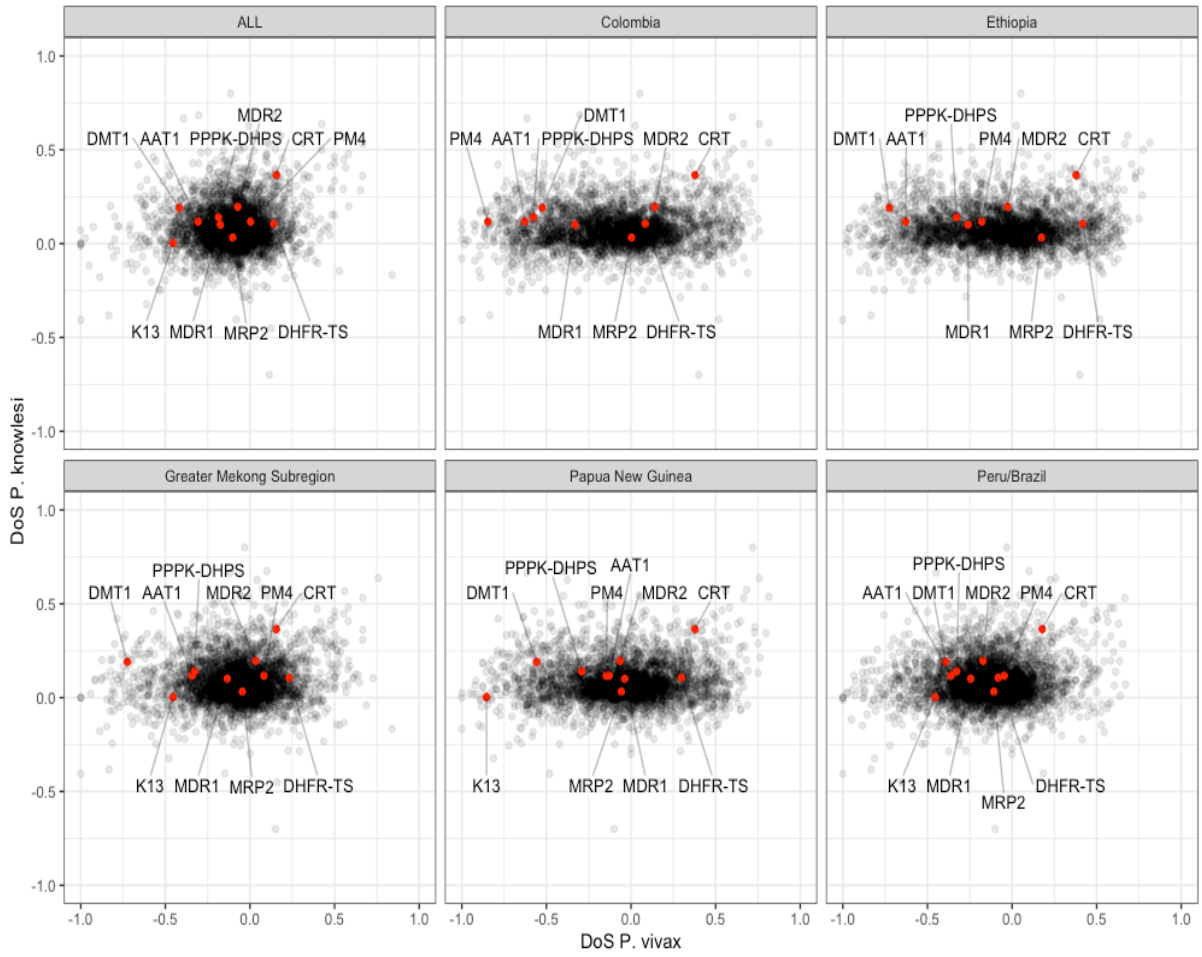
Supplemental Figure 2.4: Principal components analysis of Panama samples and previously collected Central and South American samples. Samples are colored by country of origin. Parentheses contain the percentage of variance explained by each principal component.



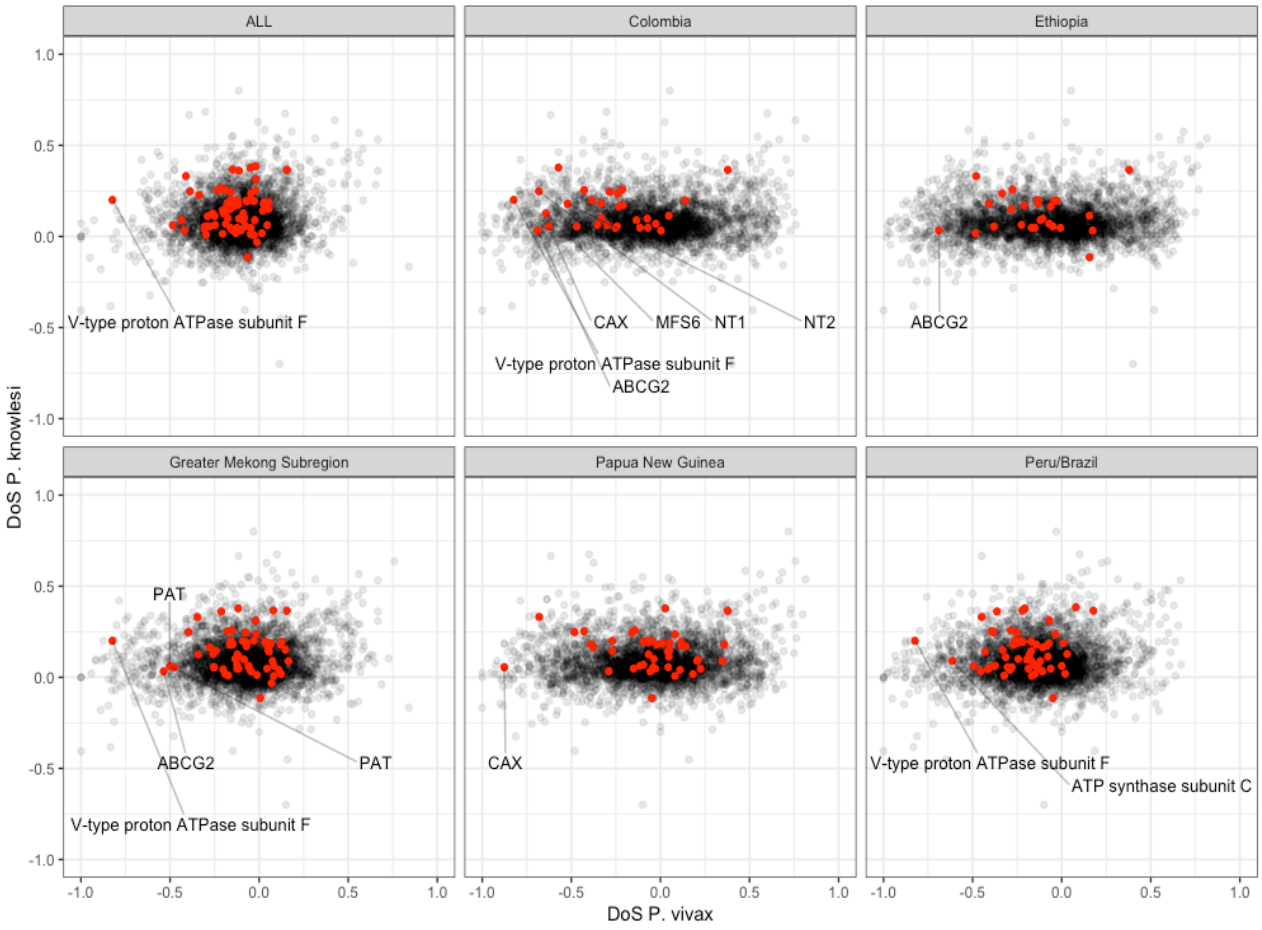
Supplemental Figure 3.1: Displays distribution of pairwise IBD estimates in each *P. vivax* population and the *P. knowlesi* population



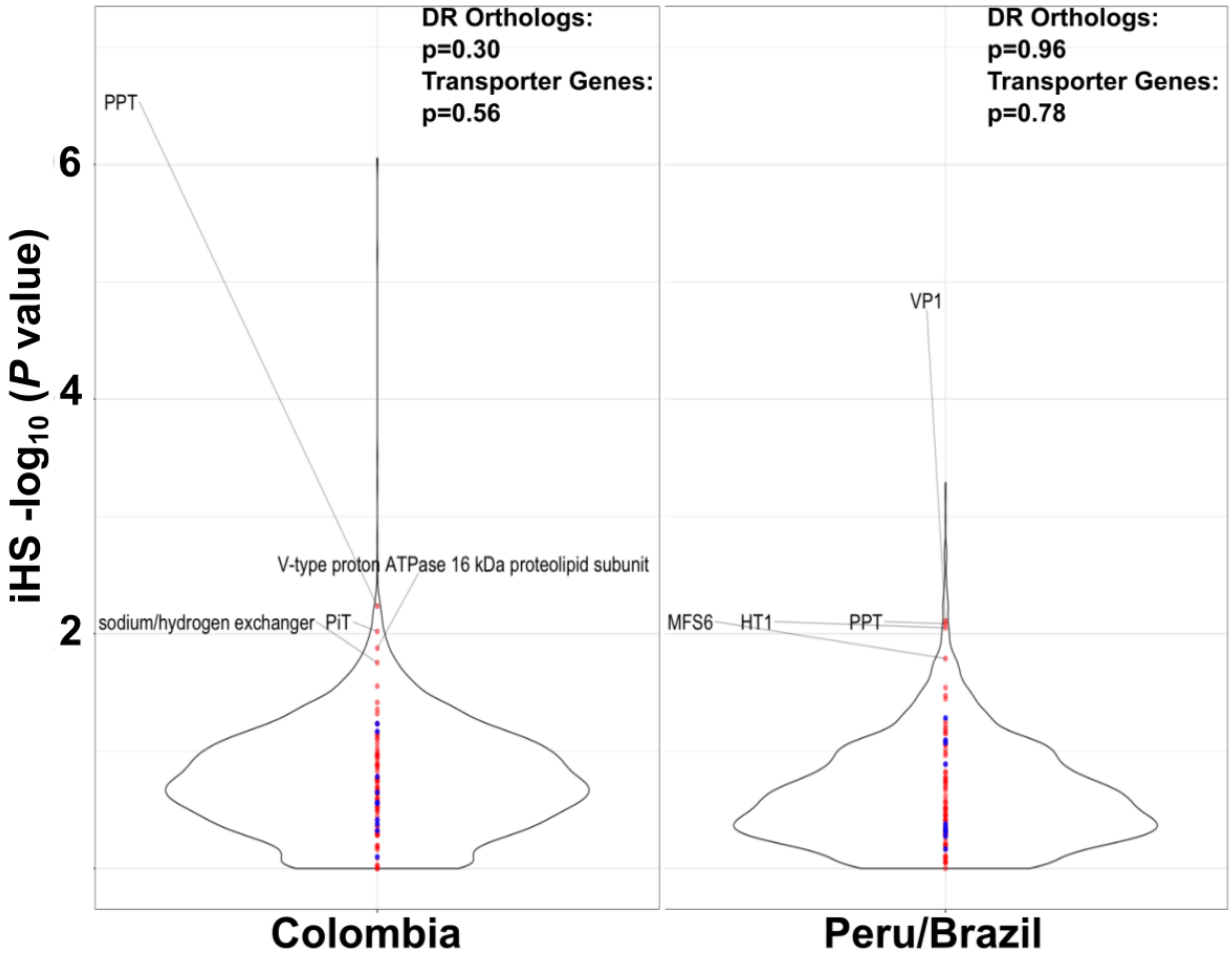
Supplemental Figure 3.2: Outlines the number of samples excluded due to each filtering step.



Supplemental Figure 3.3: *P. vivax* DoS values plotted against *P. knowlesi* DoS values for pooled population and each of the five subpopulations. All drug resistance candidate genes are highlighted and labeled



Supplemental Figure 3.4: *P. vivax* DoS values plotted against *P. knowlesi* DoS values for all pooled populations and each of the five subpopulations. All transporter genes are highlighted in red. Genes with DoS values greater than negative five or larger than five, and that have P_N, or P_S counts of three or more are labeled.



Supplemental Figure 3.5: Distribution of iHS -$\log_{10}$ (*P* values) for *P. vivax* in Colombia and Peru/Brazil Populations. Displays the distribution of iHS -$\log_{10}$ (*P* values) for *P. vivax* Colombia and Peru/Brazil populations. Drug resistance genes are highlighted in blue and transporter genes are highlighted in red. Genes with iHS scores in the top 2% of in their respective *P. vivax* population are labeled. All genes with iHS scores in the top 2% of -$\log_{10}$ (*P* values) in any of the three displayed *P. vivax* populations are labeled in the *P. knowlesi* distribution.

Supplemental Table 3.1: List of Transporter Genes

***P. vivax* Gene ID**	***P. knowlesi* Gene ID**	**Gene Product Description**	**Gene Name**
PVP01_0106100	PKNH_0104400	ATP synthase subunit C	N/A
PVP01_0109000	PKNH_0107300	cytochrome c oxidase subunit 4	COX4
PVP01_0109300	PKNH_0107600	chloroquine resistance transporter	CRT
PVP01_0113600	PKNH_0112000	V-type proton ATPase subunit a	N/A
PVP01_0206600	PKNH_0205300	mitochondrial carrier protein	N/A
PVP01_0208700	PKNH_0207400	V-type proton ATPase subunit C	N/A
PVP01_0209900	PKNH_0208700	novel transporter 1	NPT1
PVP01_0210500	PKNH_0209300	StAR-related lipid transfer protein	N/A
PVP01_0211500	PKNH_0210300	nucleoside transporter 4	NT4
PVP01_0317600	PKNH_0317300	V-type ATPase V0 subunit e	N/A
PVP01_0407900	PKNH_0403800	ATP synthase subunit alpha mitochondrial	N/A
PVP01_0414900	PKNH_0411200	Sec61-gamma subunit of protein translocation	N/A
PVP01_0415200	PKNH_0411500	transporter	N/A
PVP01_0418900	PKNH_0414700	pantothenate transporter	PAT
PVP01_0420400	PKNH_0416200	hexose transporter	HT1
PVP01_0621200	PKNH_0621300	acetyl-CoA transporter	ACT
PVP01_0621700	PKNH_0621800	ADPATP carrier protein 1	AAC1
PVP01_0713100	PKNH_0712600	major facilitator superfamily-related transporter	MFR4
PVP01_0714400	PKNH_0714000	major facilitator superfamily domain-containing protein	MFS2
PVP01_0717900	PKNH_0717500	major facilitator superfamily domain-containing protein	MFS3

Supplemental Table 3.1 (Continued): List of Transporter Genes			
PVP01_0724900	PKNH_0724300	monocarboxylate transporter	N/A
PVP01_0726300	PKNH_0726200	cytochrome c oxidase subunit 5B	COX5B
PVP01_0726500	PKNH_0726400	cytochrome c oxidase subunit 6B	COX6B
PVP01_0733200	PKNH_0733200	V-type proton ATPase subunit E	N/A
PVP01_0820500	PKNH_0821500	ABC transporter I family member 1	ABCI3
PVP01_0821200	PKNH_0822500	P4-type ATPase ATP7	N/A
PVP01_0823500	PKNH_0825200	formate-nitrite transporter	FNT
PVP01_0930500	PKNH_0927800	major facilitator superfamily-related transporter	MFR5
PVP01_0933600	PKNH_0931000	aquaglyceroporin	AQP
PVP01_0939400	PKNH_0936600	guanylyl cyclase alpha	N/A
PVP01_0940800	PKNH_0937900	V-type proton ATPase subunit F	N/A
PVP01_0948300	PKNH_0945600	ATP synthase subunit delta mitochondrial	N/A
PVP01_1003700	PKNH_1002500	phosphoenolpyruvate/phosphate translocator	PPT
PVP01_1010000	PKNH_1009000	MerC domain-containing protein	N/A
PVP01_1010100	PKNH_1009100	divalent metal transporter	N/A
PVP01_1010900	PKNH_1009900	multidrug resistance protein 1	MDR1
PVP01_1011300	PKNH_1010300	magnesium transporter NIPA	N/A
PVP01_1014700	PKNH_1013900	V-type proton ATPase 16 kDa proteolipid subunit	N/A
PVP01_1017500	PKNH_1016800	major facilitator superfamily domain-containing protein	MFS1
PVP01_1028700	PKNH_1028300	UDP-N-acetylglucosamine transporter	N/A
PVP01_1110300	PKNH_1109600	cytochrome c oxidase subunit 2A	COX2A
PVP01_1117400	PKNH_1117400	V-type proton ATPase 21 kDa proteolipid subunit	N/A
PVP01_1132100	PKNH_1133000	mitochondrial import receptor subunit TOM40	TOM40

Supplemental Table 3.1 (Continued): List of Transporter Genes			
PVP01_1134800	PKNH_1135700	major facilitator superfamily-related transporter	MFR1
PVP01_1139800	PKNH_1141100	zinc transporter ZIP1	ZIP1
PVP01_1145400	PKNH_1147200	cation/H+ antiporter	CAX
PVP01_1231400	PKNH_1205600	V-type proton ATPase subunit G	N/A
PVP01_1235600	PKNH_1209800	sulfate transporter	N/A
PVP01_1235700	PKNH_1209900	exported protein 2	EXP2
PVP01_1242100	PKNH_1216800	ATP synthase (C/AC39) subunit	N/A
PVP01_1250100	PKNH_1225000	V-type H(+)-translocating pyrophosphatase	VP1
PVP01_1259100	PKNH_1234200	multidrug resistance protein 2	MDR2
PVP01_1266200	PKNH_1241500	major facilitator superfamily domain-containing protein	MFS6
PVP01_1268000	PKNH_1243300	copper transporter	CTR1
PVP01_1207600	PKNH_1253700	nucleoside transporter 1	NT1
PVP01_1212900	PKNH_1259500	V-type proton ATPase subunit D	N/A
PVP01_1213900	PKNH_1260500	sodium-dependent phosphate transporter	PiT
PVP01_1301500	PKNH_1302100	mitochondrial phosphate carrier protein	PIC
PVP01_0509900	PKNH_1318400	nucleoside transporter 2	NT2
PVP01_1317400	PKNH_1326700	voltage-dependent anion-selective channel protein	N/A
PVP01_1318500	PKNH_1327800	cytochrome c oxidase subunit 2B	COX2B
PVP01_1321100	PKNH_1330500	major facilitator superfamily domain-containing protein	MFS5
PVP01_1322800	PKNH_1332200	ABC transporter G family member 2	ABCG2
PVP01_1327100	PKNH_1336100	copper transporter	CTR2
PVP01_1404400	PKNH_1403700	sodium/hydrogen exchanger	N/A
PVP01_1405100	PKNH_1404400	CorA-like Mg2+ transporter protein	MIT2
PVP01_1407500	PKNH_1406800	V-type proton ATPase subunit H	N/A

Supplemental Table 3.1 (Continued): List of Transporter Genes			
PVP01_1411000	PKNH_1410400	ATP synthase subunit O mitochondrial	OSCP
PVP01_1412300	PKNH_1411700	ATP synthase subunit gamma mitochondrial	N/A
PVP01_1412900	PKNH_1412600	V-type proton ATPase catalytic subunit A	N/A
PVP01_1424600	PKNH_1424500	ATP synthase subunit epsilon mitochondrial	N/A
PVP01_1425000	PKNH_1424900	cation diffusion facilitator family protein	CDF
PVP01_1429900	PKNH_1430300	aquaporin	AQP2
PVP01_1438000	PKNH_1438800	phospholipid-transporting ATPase 2	ATP2
PVP01_1441600	PKNH_1442600	phospholipid-transporting ATPase	N/A
PVP01_1441900	PKNH_1442900	vacuolar iron transporter	VIT
PVP01_1445400	PKNH_1446600	potassium channel K1	K1
PVP01_1447300	PKNH_1448500	multidrug resistance-associated protein 2	MRP2
PVP01_1453300	PKNH_1454900	vacuolar-type H+-translocating inorganic pyrophosphatase	VP2
PVP01_1453800	PKNH_1455400	ATP synthase subunit beta mitochondrial	N/A
PVP01_1458700	PKNH_1460700	S-adenosylmethionine mitochondrial carrier protein	SAMC

Supplemental Table 3.2: Divergence and Polymorphism Counts for All Drug Resistance Orthologs and Transporter Genes

Gene_ID	Name Description	D_N (*Pv*)	D_S (*Pv*)	P_N (*Pv*)	P_S (*Pv*)	DoS (*Pv*)	D_N (*Pk*)	D_S (*Pk*)	P_N (*Pk*)	P_S (*Pk*)	DoS (*Pk*)
PVP01_0621700	AAC1	13	42	6	15	-0.0493506	15	60	0	17	0.2
PVP01_1120000	AAT1	55	93	19	9	-0.3069498	99	89	27	39	0.11750484
PVP01_1322800	ABCG2	82	182	19	7	-0.4201632	74	211	12	41	0.03323403
PVP01_0820500	ABCI3	538	574	61	48	-0.0758201	587	600	155	198	0.05543053
PVP01_0621200	ACT	39	109	6	12	-0.0698198	52	111	6	41	0.19135883
PVP01_0933600	AQP	25	27	3	4	0.0521978	28	37	7	18	0.15076923
PVP01_1429900	AQP2	65	61	20	20	0.01587302	98	86	51	48	0.01745718
PVP01_1242100	ATP synthase (C/AC39) subunit	4	48	3	5	-0.2980769	8	60	4	33	0.00953895
PVP01_0407900	ATP synthase subunit alpha mitochondrial	28	185	7	9	-0.3060446	30	215	3	35	0.04350161
PVP01_1453800	ATP synthase subunit beta mitochondrial	29	166	8	10	-0.2957265	36	173	7	53	0.05558214
PVP01_0106100	ATP synthase subunit C	3	19	4	3	-0.4350649	10	31	2	11	0.09005629
PVP01_0948300	ATP synthase subunit delta mitochondrial	4	11	1	1	-0.2333333	3	20	1	10	0.03952569
PVP01_1424600	ATP synthase subunit epsilon mitochondrial	4	9	1	2	-0.025641	2	11	1	4	-0.0461538
PVP01_1412300	ATP synthase subunit gamma mitochondrial	15	35	5	11	-0.0125	15	45	9	23	-0.03125
PVP01_1438000	ATP2	122	218	17	18	-0.1268908	181	251	37	130	0.1974246
PVP01_1145400	CAX	11	78	4	6	-0.2764045	20	91	5	35	0.05518018
PVP01_1425000	CDF	54	89	5	2	-0.3366633	80	92	15	48	0.22702104
PVP01_1110300	COX2A	24	26	4	4	-0.02	27	26	3	21	0.38443396
PVP01_1318500	COX2B	11	46	2	0	-0.8070175	13	64	6	19	-0.0711688

Supplemental Table 3.2 (Continued): Divergence and Polymorphism Counts for All Drug Resistance Orthologs and Transporter Genes

PVP01_0109000	COX4	11	18	3	3	-0.1206897	14	27	5	30	0.19860627
PVP01_0726300	COX5B	17	32	10	13	-0.0878438	13	49	4	29	0.0884653
PVP01_0726500	COX6B	7	18	6	8	-0.1485714	11	19	0	9	0.36666667
PVP01_0109300	CRT	45	74	4	14	0.15592904	59	78	3	43	0.36543954
PVP01_1268000	CTR1	38	31	20	8	-0.1635611	68	40	24	26	0.14962963
PVP01_1327100	CTR2	22	13	5	1	-0.2047619	24	21	14	13	0.01481481
PVP01_0526600	DHFR	71	99	10	26	0.13986928	71	138	15	49	0.10533792
PVP01_1429500	DHPS	82	112	14	9	-0.1860152	97	106	29	57	0.14062321
PVP01_1010100	divalent metal transporter	71	123	13	20	-0.02796	126	139	20	51	0.19378156
PVP01_1424900	DMT1	30	79	18	8	-0.4170783	47	63	13	42	0.19090909
PVP01_1235700	EXP2	20	44	2	4	-0.0208333	19	40	2	27	0.25306838
PVP01_0823500	FNT	37	82	6	4	-0.2890756	41	96	7	30	0.11008088
PVP01_0939400	guanylyl cyclase alpha	515	682	90	99	-0.0459482	613	752	230	343	0.04768809
PVP01_0420400	HT1	48	139	14	17	-0.1949284	67	92	10	51	0.25744922
PVP01_1445400	K1	286	520	30	34	-0.1139113	287	557	68	212	0.09719025
PVP01_1211100	Kelch13	15	88	6	4	-0.4543689	21	116	11	62	0.00259974
PVP01_1011300	magnesium transporter NIPA	14	85	3	10	-0.0893551	21	93	3	50	0.12760675
PVP01_1010900	MDR1	94	297	32	45	-0.1751752	139	340	33	141	0.10053272
PVP01_1259100	MDR2	301	335	30	25	-0.0721841	301	333	64	166	0.19650254
PVP01_1010000	MerC domain-containing protein	56	25	12	6	0.02469136	72	41	11	13	0.17883481
PVP01_1134800	MFR1	129	224	21	19	-0.1595609	181	218	29	85	0.19924812
PVP01_0713100	MFR4	45	98	7	14	-0.018648	63	91	6	56	0.31231672
PVP01_0930500	MFR5	48	148	27	22	-0.3061224	85	189	11	32	0.05440502
PVP01_1017500	MFS1	34	109	4	10	-0.047952	44	107	17	44	0.0127022
PVP01_0714400	MFS2	267	237	33	13	-0.1876294	291	247	92	103	0.06909732
PVP01_0717900	MFS3	92	123	10	11	-0.0482835	102	110	3	26	0.3776838

Supplemental Table 3.2 (Continued): Divergence and Polymorphism Counts for All Drug Resistance Orthologs and Transporter Genes

PVP01_1266200	MFS6	49	88	13	8	-0.2613834	64	103	17	49	0.12565778
PVP01_1405100	MIT2	40	85	19	7	-0.4107692	47	95	0	38	0.33098592
PVP01_0206600	mitochondrial carrier protein	42	42	10	12	0.04545455	46	62	24	42	0.06228956
PVP01_0724900	monocarboxylate transporter	61	124	8	7	-0.2036036	77	135	12	43	0.14502573
PVP01_1447300	MRP2	464	500	63	45	-0.1020055	590	625	146	176	0.03218056
PVP01_0209900	NPT1	68	133	14	15	-0.1444502	99	134	18	51	0.16402314
PVP01_1207600	NT1	41	88	12	5	-0.3880529	59	92	7	42	0.24787133
PVP01_0509900	NT2	80	87	13	8	-0.1400057	98	99	27	58	0.17981487
PVP01_0211500	NT4	55	116	10	8	-0.2339181	75	114	5	30	0.25396825
PVP01_1411000	OSCP	24	28	1	2	0.12820513	22	30	9	19	0.10164835
PVP01_0821200	P4-type ATPase ATP7	196	447	28	37	-0.1259481	201	390	68	164	0.04699807
PVP01_0418900	PAT	19	135	14	9	-0.485319	29	113	7	42	0.06136821
PVP01_1441600	phospholipid-transporting ATPase	154	344	22	44	-0.0240964	143	372	38	164	0.08955109
PVP01_1301500	PIC	11	31	12	46	0.05500821	15	52	1	26	0.18684356
PVP01_1213900	PiT	76	119	17	31	0.03557692	105	130	22	58	0.17180851
PVP01_1340900	PM4	16	88	3	17	0.00384615	23	89	3	31	0.11712185
PVP01_1003700	PPT	38	52	20	19	-0.0905983	55	69	37	75	0.11319124
PVP01_1458700	SAMC	27	19	4	1	-0.2130435	22	40	4	41	0.26594982
PVP01_0414900	Sec61-gamma subunit of protein translocation	7	14	2	1	-0.3333333	4	15	1	9	0.11052632
PVP01_1404400	sodium/hydrogen exchanger	210	240	31	17	-0.1791667	279	289	75	167	0.18127983
PVP01_0210500	StAR-related lipid transfer protein	157	109	21	7	-0.1597744	177	93	30	43	0.24459665
PVP01_1235600	sulfate transporter	34	128	12	54	0.02805836	41	136	3	66	0.18816016
PVP01_1132100	TOM40	38	103	10	15	-0.1304965	35	101	10	32	0.0192577
PVP01_0415200	transporter	135	271	14	23	-0.0458661	167	268	19	109	0.23547055

Supplemental Table 3.2 (Continued): Divergence and Polymorphism Counts for All Drug Resistance Orthologs and Transporter Genes

PVP01_1028700	UDP-N-acetylglucosamine transporter	139	93	16	5	-0.1627668	143	115	41	40	0.04809073
PVP01_0317600	V-type ATPase V0 subunit e	5	7	0	2	0.41666667	7	18	2	7	0.05777778
PVP01_1014700	V-type proton ATPase 16 kDa proteolipid subunit	12	19	3	3	-0.1129032	13	23	0	8	0.36111111
PVP01_1117400	V-type proton ATPase 21 kDa proteolipid subunit	0	17	1	3	-0.25	2	16	0	18	0.11111111
PVP01_1412900	V-type proton ATPase catalytic subunit A	5	127	3	13	-0.1496212	10	126	2	49	0.03431373
PVP01_0113600	V-type proton ATPase subunit a	87	311	20	23	-0.2465233	88	359	16	104	0.06353468
PVP01_0208700	V-type proton ATPase subunit C	14	113	2	7	-0.111986	11	117	2	42	0.04048295
PVP01_1212900	V-type proton ATPase subunit D	3	27	3	7	-0.2	7	36	1	26	0.12575366
PVP01_0733200	V-type proton ATPase subunit E	10	76	1	4	-0.0837209	8	69	0	18	0.1038961
PVP01_0940800	V-type proton ATPase subunit F	3	14	8	0	-0.8235294	5	20	0	7	0.2
PVP01_1231400	V-type proton ATPase subunit G	2	31	0	0	NA	3	34	0	14	0.08108108
PVP01_1407500	V-type proton ATPase subunit H	5	56	1	8	-0.0291439	8	50	3	20	0.00749625
PVP01_1441900	VIT	7	42	5	13	-0.1349206	14	51	1	21	0.16993007
PVP01_1317400	voltage-dependent anion-selective channel protein	27	77	4	5	-0.1848291	22	83	3	40	0.13975637
PVP01_1250100	VP1	20	122	12	88	0.02084507	34	116	5	53	0.14045977
PVP01_1453300	VP2	143	209	17	19	-0.0659722	129	220	133	142	-0.1140089
PVP01_1139800	ZIP1	27	68	2	3	-0.1157895	36	83	5	34	0.17431588

Supplemental Table 4.1: Haplotype Counts by Country										
	Brazil	Cambodia	China	Colombia	Ethiopia	Gabon	India	Indonesia	Laos	Madagascar
H1	0	0	2	0	2	9	1	0	0	0
H2	0	12	1	0	0	0	0	5	0	0
H3	0	1	1	0	0	0	0	0	0	0
H4	0	5	0	0	0	0	0	0	2	0
H5	15	0	0	0	7	0	0	0	0	0
H6	0	0	0	0	6	0	0	0	0	0
H7	8	0	0	3	0	0	0	0	0	0
H8	2	2	3	0	0	0	0	0	0	0
H10	0	0	0	0	8	0	1	0	0	0
H11	0	0	1	0	0	0	0	0	0	0
H12	0	1	0	0	0	0	0	0	0	0
H13	5	0	0	0	10	0	0	0	0	2
H14	1	0	0	0	2	0	0	0	0	0
H15	1	0	0	25	0	0	0	0	0	0
H16	0	0	0	0	0	0	2	0	0	0
H17	0	0	0	0	0	0	0	0	0	0
H18	5	0	0	1	0	0	0	0	0	0
H19	46	0	0	0	0	0	0	0	0	0
H20	1	0	0	0	0	0	0	0	0	0
H21	37	0	0	0	0	0	0	0	0	0
H22	2	0	0	0	0	0	0	0	0	0
H23	0	0	0	1	0	0	0	0	0	0
	123	21	6	30	33	0	3	5	2	2

Supplemental Table 4.1 (Continued): Haplotype Counts by Country								
	Malaysia	Mauritania	Mexico	Myanmar	Nicaragua	North Korea	Panama	PNG
H1	0	0	0	1	0	0	0	4
H2	4	0	0	1	0	0	0	15
H3	0	0	0	1	0	0	0	0
H4	0	0	0	1	0	1	0	0
H5	0	1	0	0	0	0	0	1
H6	0	0	0	0	0	0	0	0
H7	0	0	0	0	0	0	1	0
H8	0	0	0	2	0	0	0	0
H10	0	0	0	0	0	0	0	0
H11	0	0	0	1	0	0	0	0
H12	1	0	0	0	0	0	0	0
H13	0	0	0	0	0	0	0	0
H14	0	0	1	0	1	0	0	0
H15	0	0	0	0	0	0	0	0
H16	1	0	0	0	0	0	0	0
H17	0	0	0	0	0	0	0	0
H18	0	0	17	0	0	0	0	0
H19	0	0	0	0	0	0	0	0
H20	0	0	0	0	0	0	0	0
H21	0	0	0	0	0	0	0	0
H22	0	0	0	0	0	0	0	0
H23	0	0	1	0	0	0	0	0
	6	1	19	6	1	1	1	16

Supplemental Table 4.1 (Continued): Haplotype Counts by Country				
	Peru	Sri Lanka	Thailand	Vietnam
H1	0	1	6	0
H2	0	0	5	5
H3	0	0	2	0
H4	0	0	0	1
H5	0	0	0	0
H6	0	0	1	0
H7	31	0	0	0
H8	0	0	3	0
H10	0	0	0	0
H11	0	0	2	0
H12	0	0	0	0
H13	0	0	0	0
H14	0	0	0	0
H15	3	0	0	0
H16	0	0	0	0
H17	2	0	0	0
H18	0	0	0	0
H19	0	0	0	0
H20	0	0	8	0
H21	0	0	0	0
H22	0	0	0	0
H23	0	0	0	0
	36	0	21	6