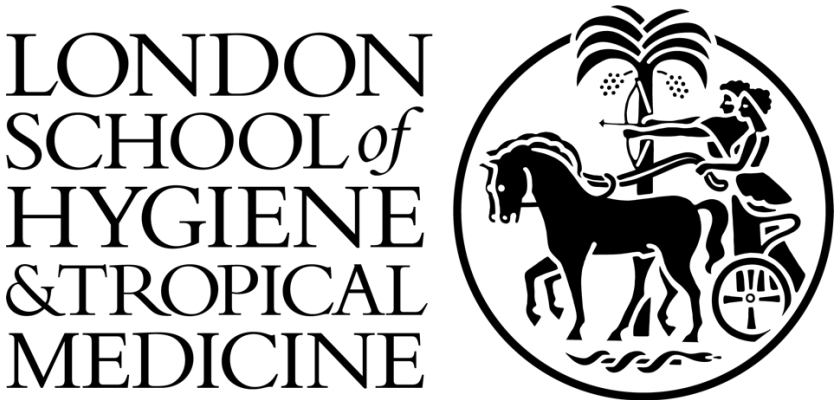


From genome to function: A genomic investigation into understudied populations of the malaria parasites *Plasmodium malariae* and *P. vivax.*

Amy Ibrahim

Thesis submitted in accordance with the requirements for the degree of
Doctor of Philosophy

University of London
February 2022

Department of Infection Biology

Faculty of Infectious and Tropical Diseases

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded by The Medical Research Council

Research group affiliations:
Professor Susana Campino
& Professor Taane Clark

I, Amy Ibrahim, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed:  Date: 23/02/2022

Abstract

Genomics has been a pivotal tool for infectious disease research, especially for malaria where studies into the genome of *Plasmodium falciparum* have provided insights into population dynamics and determinants of drug resistance. There are six human-infective species of *Plasmodium (P. falciparum, P. vivax, P. malariae, P. ovale wallikeri, P. ovale curtisi, P. knowlesi)*, with most deaths due to *P. falciparum* infections, and subsequently most research is focussed on this species. Knowledge of the complexity and variability of the *Plasmodium spp*. genomes, and comparative studies between species, can provide valuable insights into parasite biology, which can aid the development of effective disease control methods. For non-falciparum malaria, the amount of genomic data available is sparse, which this thesis aims to improve, and subsequently provide novel knowledge, specifically regarding two understudied groups of Plasmodium parasites: 1) the first global analysis of *P. malariae* isolates and 2) an investigation into understudied transmission regions of *P. vivax* in Brazil. Whole genome sequencing of *P. malariae* isolates is difficult, partly due to their clinical presentation with low parasitaemias, resulting in difficulties obtaining sufficient DNA for WGS. To overcome this barrier, we initially develop a methodology for selective whole genome amplification (SWGA) of *P. malariae* isolates which is validated using 19 isolates and demonstrate geographical separation of parasites in addition to single nucleotide polymorphisms (SNPs) within orthologs of genes associated with reduced drug susceptibility. Following this initial pilot study, I proceed to create the first large scale global genomic database of *P. malariae* isolates, generating high quality WGS data for 155 isolates spanning 4 continents and 25 countries. Using this database, I confirm that *P. malariae* parasites demonstrate isolation by distance at the continental level between Africa and Asia, with South American isolates related to those from Africa. I identify multiple SNPs within genes associated with drug susceptibility, which can be further validated with *in vitro* studies. Using the SNPs identified in the global genomic database, I develop an *in vitro* system using ortholog replacement through CRISPR-Cas9 mediated genome editing in the culture-adapted *P. knowlesi* A1H1 line to determine the phenotypic effect of genomic variants in the *pmdhfr-ts* gene. I validate this system for testing *dhfr-ts* variants using controls for pyrimethamine sensitive and resistant lines, and further demonstrate a pyrimethamine-resistant *P. malariae* parasite phenotype, based on genomic variants that are seen within the global database. Finally, I investigate *P. vivax* parasites

in the largest genomic database to date (n = 885) spanning 3 continents and 26 countries, with novel sequence data generated for Brazilian isolates across 7 states, 6 of which being previously unstudied regions of *P. vivax* transmission. I demonstrate isolation by distance in *P. vivax* parasites and determine potentially significant SNPs within genes suspected to be involved in drug susceptibility, including some signals of positive selection at these loci.

Overall, this thesis develops the first platform for *P. malariae* genomic research and identifies novel loci for investigation into both *P. malariae* and *P. vivax*, specifically regarding mutations in candidate genes potentially associated with drug resistance. I build on this by translating genomic candidates into functional research and demonstrate the importance of validating genomic signals.

Acknowledgements

Firstly, I would like to thank my primary supervisor Professor Susana Campino, thanks for being an inspiring mentor as well as caring and understanding when I have needed it most. As with any PhD there have been ups and downs and I would like to sincerely thank you for your unwavering support through those. I hope to one day become a supervisor that is as talented, mischievous, and comical as you are. I would also like to thank Professor Taane Clark, my secondary supervisor for his support and advice along the way, and for all the opportunities that you have provided throughout the duration of my studies, of which, there have been so many.

To the rest of the Clark/Campino group: Ernest Diez Benavente, thanks for putting yourself through torture by showing me the ropes and answering endless questions and never being visibly annoyed. Paula Josefina Gómez González (Pepi), it's been a pleasure and a pain to endure this journey together, but we have made it and I am proud of us both, you owe me some bloody teabags! Dan Ward and Matthew Higgins, thanks for being two hugely caring and sensitive souls who have many times listened when it has all got a bit too much. Jody, thanks for being a cheery ray of sunshine that has supported so many students alongside your own research and manages to pull it off with ease. Emilia Manko, thanks for all your help and support. Looking forward to celebrating with you all and the rest of The Hub in the Pumphandle soon.

Additionally, thanks to so many of the staff at LSHTM, especially those based in the malaria lab (380 and 381/382) that have provided help and advice over the course of my PhD. Special thanks to Rob Moon and the rest of the Moon Group, whose collaboration allowed me to progress from descriptive genomics to functional analysis, especially Franziska Mohring whose support and training was pivotal for this chapter to progress. Khalid Beshir and Colin Sutherland, thanks for your invaluable input and mentoring throughout the entire process.

To my amazing friends, those based at LSHTM who have helped me through either with coffee or beers and plenty of laughs along the way. Harry, Rhodri and Pepi, your friendship and care throughout my studies has been immense and invaluable. And those outside of work, with an

exceptional mention to Chloe, Josh, Louis, Immy, Bea, Freds, Nat, Jamie and Rosie for much needed emotional support / distraction.

To my boat, Black Velvet, my witch's cat Billi, and the amazing boating community (especially the boatwomen of the London Waterways – with a special shout out to Bonny). Living on water during my studies has been the most challenging and rewarding experience. Boating has taught me not to be too distraught when all my PCRs haven't worked, or when I have contaminated my cultures, because I could still go home and spend the night with no electricity or water. Strangely this perspective was more calming than it might sound.

To my family, there have been many times when I've doubted if I could complete this, but it was making you proud that pushed me through. Mom, you have been my rock and inspiration for as long as I can remember, this is as much for you as it is for me. Chloe, my twin and my best friend, thanks for caring for me when I've needed it most. Josh, thanks for being there and being 90s and related. Nan and Grandad, thanks for everything that you have done to support me over the course of my life, you have both been incredible and a huge inspiration to me (especially to my Nan who I have no doubt will try and read through this thesis as she has with my papers, essays, and presentations). Dad, thank you for your support and love throughout my life and studies. To all of you, I love you dearly and I wouldn't have been able to do it without any one of you.
Thank you.

I would like to dedicate my thesis to my paternal grandfather, Jaffer Ali Ibrahim, who we unfortunately lost during the first stages of the COVID-19 pandemic. Losing you in such tragic circumstances was the hardest thing that I have been through, and I wish you would have been around to see me complete my studies, as I know you would have been so proud. I hope that I can go forwards with just a fraction of your positivity, humour and generosity, the world was such a better place for your presence in it.
May you rest in eternal peace.

Additional publications

During my PhD, I have contributed to other manuscripts which are not included in this thesis:

1. Higgins M, Ravenhall M, Ward D, Phelan J, **Ibrahim A**, Forrest MS, et al. PrimedRPA: Primer design for recombinase polymerase amplification assays. Bioinformatics. 2019 Feb 15;35(4):682–4.
2. Benavente ED, Gomes AR, De Silva JR, Grigg M, Walker H, Barber BE… **Ibrahim A**…, et al. Whole genome sequencing of amplified Plasmodium knowlesi DNA from unprocessed blood reveals genetic exchange events between Malaysian Peninsular and Borneo subpopulations. Sci Rep [Internet]. 2019;9(1):9873. Available from: http://www.nature.com/articles/s41598-019-46398-z
3. Da Veiga Leal S, Ward D, Campino S, Benavente ED, **Ibrahim A**, Claret T, et al. Drug resistance profile and clonality of Plasmodium falciparum parasites in Cape Verde: the 2017 malaria outbreak. Malar J. 2021 Dec 1;20(1).
4. Nolder D, Stewart L, Tucker J, **Ibrahim A**, Gray A, Corrah T, et al. Failure of rapid diagnostic tests in Plasmodium falciparum malaria cases among travelers to the UK and Ireland: Identification and characterisation of the parasites. Int J Infect Dis [Internet]. 2021 Jul 1 [cited 2022 Feb 12];108:137. Available from: /pmc/articles/PMC8295040/

Table of Contents

Abbreviations

ACT	Artemisinin combination therapy
Amp	Ampicillin
ASSP	Artesunate + sulfadoxine-pyrimethamine combination therapy
ATP	Adenosine triphosphate
bp	Base pairs
BSD	Blasticidin
Cas9	CRISPR associated protein 9
CRISPR	Clustered regularly interspaced short palindromic repeat
CRT	Chloroquine resistance transporter
CTX	Cotrimoxazole
DBP	Duffy Binding Protein
DHA	Dihydroartemisinin
DHFR	Dihydrofolate reductase
DHPS	Dihydropteroate synthase
DNA	Deoxyribonucleic acid
DSB	Double strand break
FACS	Fluorescent activatated cell sorting
FST	Fixation index
Fy-	Duffy negative blood phenotype
Fy+	Duffy positive blood phenotype
G6PDD	Glucose-6-phosphate dehydrogenase deficiency
gDNA	Genomic deoxyribonucleic acid
GIA	Growth inhibitory assay
HK	Hematocrit
HR	Homologous recombination
IBD	Identity by descent
iHS	Integrated haplotype score
INDEL	Insertion/deletion
IPTi	Intermittive preventative treatment in infants

IPTp	Intermittive preventative treatment in pregnant women
iRBC	Infected red blood cell
IRS	Indoor Residual Spraying
ITN	Insecticide treated bed net
kb	Kilobase pair
LB	Luria-Bertani agar
Mb	Megs base pairs
mg	Milligram
ml	Millilitre
mM	Millimolar
MSP	Merozoite surface protein
NHEJ	Non-homologous end joining
nM	Nanomolar
OR	Orthologue replacement
pABA	Para-aminobenzoic acid
PAM	Protospacer adjacent motif
pCas9	Cas9 protein expression plasmid
PCR	Polymerase chain reaction
pDonor	Donor DNA template plasmid for homology directed repair
Pf	Plasmodium falciparum
Pk	Plasmodium knowlesi
Pm	Plasmodium malariae
Poc	Plasmodium ovale curtisi
Pow	Plasmodiu ovale wallikeri
Pv	Plasmodium vivax
RBC	Red blood cell
RDT	Rapid diagnosic tests
RNA	Ribonucleic acid
SA	South Asia
SAM	South America
SEA	Southeast Asia

SNP	Single nucleotide polymorphism
SP	Sulfadoxine Pyrimethamine
SSEA	Southern Southeast Asia
SWGA	Selective whole genome amplification
uM	Micromolar
WGS	Whole genome sequencing
WHO	World Health Organisation
WT	Wild type

CHAPTER 1

1 Introduction

1 Introduction

1.1 Malaria as a global health concern

Malaria is a devastating infectious disease caused by apicomplexan parasites of the genus *Plasmodium,* transmitted by female mosquitoes during blood-feeding on a wide variety of animal hosts. There are 6 commonly occurring human-infective species of *Plasmodium; Plasmodium falciparum, P. vivax, P. malariae, P. ovale curtisi, P. ovale wallikeri,* and *P. knowlesi.* Fever, fatigue, nausea and chills are the most common symptoms of malaria infection, with cyclical fevers occurring at different intervals dependant on the causative *Plasmodium spp.*. There were 241 million estimated malaria cases in 2020 spanning 85 countries, the first increase in malaria incidence since before the millennium, which led to 627, 000 deaths (a 12% increase in comparison to 2019), and is largely due to disruptions in malaria control and services due to the global COVID-19 pandemic [1] (**Figure 1**). The burden of malaria is not distributed evenly across the globe (95% of cases in the African subcontinent [1]), with distribution dependent on many factors, including environmental conditions, allowing for different species of mosquito vector (each with different feeding habits) to inhabit each region. Socioeconomic factors also impact malaria burden, for example poorly insulated, overcrowded housing may make individuals more likely to become infected, and some individuals in rural settings may be less likely to access adequate healthcare facilities [2]. Finally, human biology also impacts malaria transmission, with some genetic traits providing protection against malaria infection, as in the case of sickle cell anaemia [3] that provides protection against severe disease caused by *P. falciparum,* and the Duffy negative (Fy-) phenotype which provides protection against *P. vivax* infections [4].

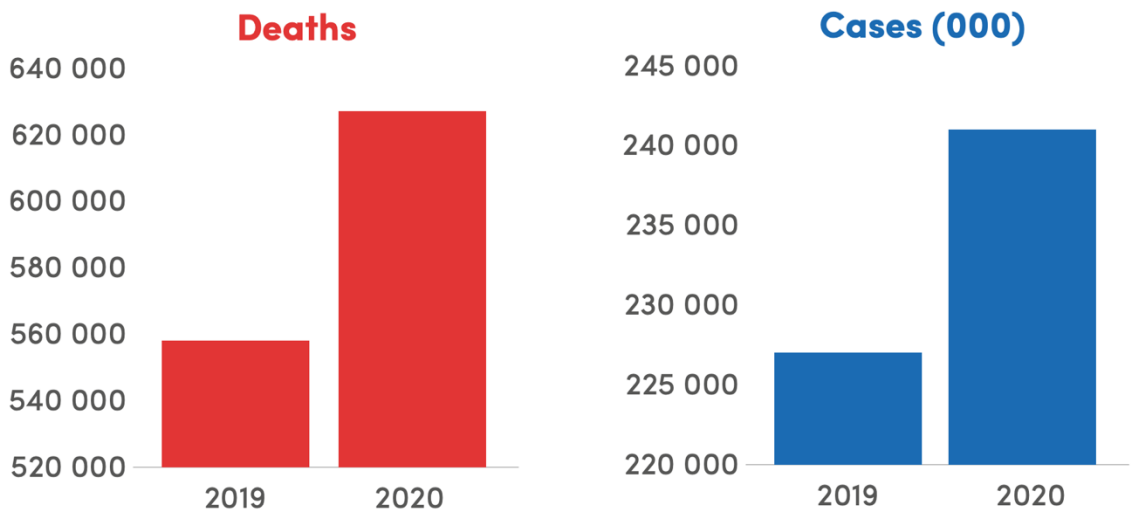


Figure 1. The first increase in malaria incidence and deaths since the beginning of the millennium. The total estimated malaria deaths in 2019 and 2020 (in red) on the left-hand side, and the estimated total malaria cases (in blue) on the right-hand side. Taken from the World Malaria Report 2020 [5].

1.1.1 *P. falciparum*

P. falciparum infections carry the greatest malaria burden and cause differing symptomatic diseases depending on the transmission dynamics in a particular region. The greatest impact of falciparum-malaria is seen in Sub-Saharan Africa, where approximately 95% of malaria cases and 96% of malaria deaths are seen, resulting in exposure to multiple infections at a young age [1]. The high transmission seen in Africa is in part due to the presence of highly anthropophilic mosquito vectors, including the *Anopheles gambiae* complex (*An. gambiae, An. arabiensis, An. merus* and *An. melas*), alongside *An. funestus, An. moucheti* and *An. nili* [6].

Children under the age of five years are the most at risk of severe malaria in high transmission regions such as sub-Saharan Africa. At birth, infants inherit maternal antibodies which provide protection against severe malaria and are thought to last for approximately 5-6 months [7,8]. When protective antibodies wane, children are susceptible to severe disease, and are likely to encounter multiple malaria attacks, leading to children under the age of five years accounting for 80% of malaria deaths in this region [1]. Those who survive multiple attacks will gain a level of protective

immunity (known as clinical immunity), reducing the number of older children and adults with severe malaria pathologies, even though infections still occur.

Following the African continent, the WHO Southeast Asian Region (SEA) suffers the second greatest burden of malaria, with 2% of the global malaria burden in 2020 (83% of this burden was due to malaria incidence in India alone) [1]. Both *P. falciparum* and *P. vivax* are transmitted in SEA, with *P. vivax* infections contributing to 36.3% of malaria incidence in 2020 [1]. Due to reduced malaria transmission, and the differences in prevalent parasite species, the malaria pathology seen in SEA differs from that witnessed in the African continent. In areas of lower *P. falciparum* transmission, such as SEA, children are less likely to be exposed to multiple infections, therefore, immunity is slower to develop, leaving a range of age groups susceptible to severe malaria [9].

Priority areas for malaria transmission in SEA include rainforests and borders between countries, leaving forest and mobile workers as the most at-risk groups. Interactions which increase the likelihood of individuals being near malaria vectors increases malaria transmission, such as work on large infrastructure projects through forested area, commercial work on plantations and large deforestation projects (specifically a factor in SEA which faces some of the greatest deforestation in the world) [10]. There are additional complications to malaria transmission in SEA due to a long history with *P. falciparum* antimalarial drug resistance in this region [11].

Additionally, the WHO Region of the Americas (SAM) is responsible for 0.3% of malaria cases globally, with mostly cases of *P. vivax* (75%), and rare transmission of *P. falciparum* (25%) [1]. Malaria transmission in SAM mainly occurs in the Northern countries surrounding the Amazon rainforest [12]. Similar to SEA, activities which increase the contact between humans and malaria vectors, such as large scale deforestation [13] and gold mining [14], increase the likelihood of malaria transmission.

Severe falciparum-malaria is characterised as dysfunction of vital organs and is due to sequestration of infected red blood cells (iRBCs) to the host vasculature system, caused through the binding interactions of the parasite variable surface antigens expressed and exported to the iRBC cell membrane [15]. Clinical outcomes may include respiratory distress, kidney damage and

cerebral malaria, amongst other presentations [16]. A further complication of falciparum-malaria is seen across the globe in pregnant women, where iRBCs may sequester in the placenta through binding to syncytiotrophoblasts, leading to an increased risk of maternal anaemia, miscarriage, and death [17]. *P. falciparum* erythrocytic development takes 48 hours, however the cyclical fevers associated with falciparum-malaria can be irregular due to less efficient synchronisation of parasites in the human host [18].

P. falciparum infections are commonly diagnosed through light microscopy or rapid diagnostic tests (RDTs), however there are concerns over the efficacy of RDTs due to deletions in the histidine rich protein antigen gene, *Pfhrp2* [19–21], leading to false negative results. The standard first line treatment for falciparum-malaria is a 3-day course of artemisinin-based combination therapy (ACT), which is formulated from an artemisinin derivative (artemether, artesunate or dihydroartemisinin) in combination with a longer acting partner drug (for example lumefantrine or mefloquine), with different formulations favoured in different regions. Artemisinin resistance in *P. falciparum* is widespread in the Greater Mekong Subregion and many regions in South-east Asia [22,23], with recent reports in Africa [24], however, ACTs are still used as a successful treatment due to the long-lasting antimalarial activity of the partner drug. When treatment failures do occur, the recommended treatment is to change to an alternative ACT for 3 days [25].

1.1.2 *P. vivax*

The second most virulent Plasmodium parasite, *P. vivax,* is mainly transmitted within Asia, South America, the Middle East, and Oceania, with a reduced prevalence in Africa due to the presence of the Fy- blood group, which affects the ability of parasites to invade host erythrocytes [4]. It was initially thought that the Fy- blood group provided complete protection against *P. vivax* invasion of red blood cells (RBCs), however it is now known that this is not the case as vivax-malaria has been reported in Fy- individuals in regions in Africa [26,27]. *P. vivax* infections are less pathogenic than *P. falciparum*, however, severe malaria does occur and is commonly associated with anaemia, respiratory difficulties, and rarely, coma [16,28]. Whilst less deaths are due to *P. vivax* infections, more than one third of the world's population are at risk of infection due to its wide distribution, resulting in huge morbidity and economic relevance [29].

Current tools and measures used to control malaria transmission have been developed against *P. falciparum* parasites, and due to innate biological differences between species, are less effective against *P. vivax* parasites, further complicating control of this parasite species. A key difference includes the ability of *P. vivax* parasites to form dormant hypnozoite stages in the human liver, which are not cleared with routine anti-malarials, requiring additional treatment with a hypnozoiticidal drug such as primaquine or tafenoquine, known as a radical cure [30]. Radical cure therapy for vivax-malaria is effective, but requires a prolonged course, challenging adherence, in addition to the inability of using these drugs in pregnant women or children, and in patients with glucose-6-phosphate dehydrogenase deficiency (G6PDD), due to the high risk of haemolytic anaemia [31].

In addition, some *P. vivax* characteristics increase the transmissibility of this parasite species to mosquito vectors, including: The ability of the parasite to develop at lower temperatures within the mosquito midgut, enabling their survival into colder regions than *P. falciparum* [32]; an increased permissibility to a wider range of mosquito vectors [33] and faster production of gametocytes within the human host, increasing the amount of time that a human is infectious to mosquitoes [34]. Additionally, *P. vivax* infections tend to present with lower parasitaemias than with *P. falciparum*, meaning that infections are more difficult to detect and treat, both by light microscopy and RDTs, this coupled with a reduced sensitivity of RDTs increases the chances of missing infections and leads to onward transmission of untreated parasites [35,36]. Further complications include *P. vivax* diagnosis, for instance, most RDTs will only report *P. vivax* infections under the pan-Plasmodium antigen, therefore the exact prevalence of this species is difficult to ascertain [37]. RDTs with a *P. vivax*-specific antigen exist but are not commonplace [38]. Additionally, *P. vivax* parasites tend to invade younger RBCs, which are more commonly found within the bone marrow as opposed to circulating around the body, providing hidden reservoirs of parasites that may not be detected from blood samples [39].

1.1.3 The neglected malaria parasites

1.1.3.1 *P. malariae*

P. malariae infections are found across the tropics and subtropics, with a similar distribution to *P. falciparum*, and are commonly found in mixed infections with other *Plasmodium spp.* [40]. Whilst mostly considered a benign infection, there have been accounts of severe pathologies including nephropathy [40,41] (especially in children [42]), respiratory distress [43], severe anaemia, and fatalities [44]. Previous diagnostics have underestimated the prevalence of *P. malariae*, for example, microscopy is complicated by the presence of mixed infections, with *P. malariae* commonly demonstrating low parasitaemias [40], mixed infections are likely to be missed with the dominant species being diagnosed. Additionally, current RDTs are unable to specifically detect *P. malariae* antigens, instead detecting a pan-Plasmodium antigen, leading to underreporting of *P. malariae*. However, through the advent of molecular screening, it is now evident that *P. malariae* infections are more abundant than shown previously with routine microscopy [45,46], contributing to between 9% and 41% of malaria (in both single-species and mixed infections) in differing studies [45,47–49]. Demonstrating the need for molecular surveillance of *Plasmodium spp.,* a priority that will become more apparent as malaria control programmes continue to reduce the levels of *P. falciparum* infections in a population. When diagnosed, *P. malariae* infections are treated with either chloroquine, or an ACT in areas of high chloroquine resistance.

P. malariae infections are unique in multiple ways including their ability to evade the human immune response, sometimes causing infections lasting decades [40,43,50], in addition to the longer erythrocytic development cycle, resulting in cyclical fevers every 72 hours [40]. Additional complications with *P. malariae* infections are the reported severe pathologies, in addition to both post-treatment [47,51], and post-prophylactic parasitaemia [52], which may be due to an innate parasite factor causing current drugs to be less effective, due to selection pressures or to insufficient treatment duration. It is currently not known if there are drug resistant *P. malariae* parasites, and no molecular markers have been confirmed, with research into this area much needed.

1.1.3.2 *P. ovale spp*.

P. ovale infections were previously thought of as one species, *P. ovale*, however, it is now known that ovale-malaria can be caused by two *P. ovale* species; *P. ovale wallikeri* (*Pow*), and *P. ovale curtisi* (*Poc*) [53]. Both *Poc* and *Pow* infections demonstrate with tertian fevers (49 hour erythrocytic cycle) and have associations with jaundice, anaemia and respiratory difficulties [54,55], further

complicated by the ability to formant dormant hypnozoite stages, akin to *P. vivax* [56]. *Poc* and *Pow* are classified as individual species as they do not genetically recombine, in addition to differing phenotypes, including differences in duration before a relapse infection [53,57], and potential differences in disease presentation [58]. Both *P. ovale spp.* infections are mostly found within Africa, with lower-level transmission in Asia, and rare instances in Europe and Australasia. Within Africa, there is country level differentiation between the prevalence of either species, with *Poc* mostly found in Nigeria, Equatorial Guinea and Gabon, and *Pow* mostly found in Equatorial Guinea, Nigeria and Cameroon [59].

P. ovale spp. infections are likely under-reported due to difficulties in diagnostics, with microscopy the main diagnostic in most of sub-Saharan Africa, where *Poc* and *Pow* are likely found in mixed infections, further confounding microscopy. Additionally, RDTs do not specifically detect *Poc* and *Pow* infections, which will appear under the pan-Plasmodium band, further complicated by the finding that RDTs are less sensitive to both species, but especially *Poc* [60,61], reinforcing the need for continued molecular surveillance. Once diagnosed, *Poc* and *Pow* infections are treated with chloroquine and primaquine (for radical cure), or ACTs and primaquine in regions with high levels of chloroquine resistance [25].

1.1.3.3 Plasmodium knowlesi

P. knowlesi is a zoonotic parasite that is restricted to Southeast Asia (particularly Malaysia) [62], and the main hosts are the long-tailed (*Macaca fascicularis*) or pig-tailed (*M. nemestrina*) macaques [63]. Human infections are commonly seen in forested areas close to the macaque habitats, particularly men who work in farming or clearing vegetation have an increased risk of *P. knowlesi* infection [64,65]. The pathology of *P. knowlesi* is diverse, ranging from asymptomatic to fatal. The erythrocytic cycle takes 24 hours, which can lead to rapid onset of severe disease [66]. Severe knowlesi-malaria presents similarly to severe cases of *P. falciparum* infection[16]. Due to the quick progression of disease, rapid treatment is important, which can be difficult in rural regions where medical facilities are sparse.

P. knowlesi infections are difficult to detect in patients, as microscopy can commonly confuse *P. knowlesi* with *P. malariae* (misdiagnosis also occurs with *P. falciparum* or *P. vivax*). In addition

to this, both RDTs and nested PCR are unable to specifically detect *P. knowlesi* parasites [67], leaving knowlesi-specific molecular assays as the most efficient method, however, these are seldom used in a clinical setting [68]. *P. knowlesi* infections are commonly treated with chloroquine or ACTs in areas with suspected chloroquine resistance [25].

1.2 The life cycle of the malaria parasite

The *Plasmodium* life cycle is complex, with numerous stages both within the invertebrate mosquito vector and the mammalian host. Human infection begins with the bite of an infected female mosquito, as the mosquito feeds on human blood, parasite sporozoites are passed from the mosquito salivary gland into the human bloodstream, where they are transported to the liver and invade hepatocytes. Within hepatocytes, parasites mature into merozoites and multiply by asexual reproduction to form schizonts. Liver development takes between 2 and 7 days depending on the parasite species, where the infection remains asymptomatic (at this stage, *P. vivax* and *P. ovale spp.* can form a dormant hypnozoite stage within the liver which can later cause relapses of infection). Once schizonts have matured, they rupture the hepatocyte releasing individual merozoites into the bloodstream, where merozoites invade erythrocytes and begin the asexual blood stage of malaria infection. Inside erythrocytes, merozoites mature into ring stage parasites which can mature down two different paths: 1) the sexual pathway, where ring stage parasites develop into male and female gametocytes which are taken up from the bloodstream following the bite of a female mosquito, leading to the mosquito life cycle stages or 2) the asexual pathway, which causes human pathology; in the asexual stages, ring stage parasites develop into trophozoites and undergo asexual multiplication to create schizonts. Schizonts rupture releasing multiple merozoites which can invade new erythrocytes and undergo the same stages of development again. It is the rupturing of erythrocytes caused by the asexual replication stages that is associated with anaemia and the cyclical malaria fevers (**Figure 2**). The different *Plasmodium spp.* undergo the asexual cycle at different speeds (**Table 1**), which leads to the differing presentations of cyclical fevers [17].

Table 1. Comparison of basic biological features of human-infective *Plasmodium spp.*

Species	Asexual cycle (hours)	Main host	Reference genome	Genome length (Mb)	GC content (%)	Publicly available WGS data (N)	First-line treatment
P. falciparum	48	Human	Pf3D7	22.8	19.4	7,000	ACT
P. vivax	48	Human	PvP01	29	39.8	1,113	Chloroquine + primaquine*
P. malariae	72	Human	PmUG01	31.9	24.74	195	Chloroquine*
P. ovale curtisi	49	Human	Poc1	34.5	34.5	1	Chloroquine + primaquine*
P. ovale wallikeri	49	Human	Pow1	35.3	35.3	1	Chloroquine + primaquine*
P. knowlesi	24	*Macaca sp.*	A1-H1	23.5	37.5	182	Chloroquine*

*Chloroquine is used in regions where chloroquine resistance is low, if chloroquine resistance is high, or if species determination is inconclusive, then it is recommended to use ACTs instead [25].
Publicly available whole genome sequencing (WGS) data for *P. falciparum* is summarised in the 2021 MalariaGen study [69], publicly available WGS data for *P. vivax* is described in **Chapter 4 including previously published isolates [70], WGS data for *P. malariae* is described in **Chapter 2**, WGS data for *Poc* and *Pow* is previously published [71], and WGS data for *P. knowlesi* was previously published [72,73].

The sexual pathway begins when a subset of ring stage parasites differentiates into male and female gametocytes within the erythrocyte. Gametocytes are ingested by a female mosquito during a bloodmeal and due to the environment within the mosquito midgut, they can further develop into macrogametes (female) and microgametes (male), which can undergo sexual replication to form a diploid zygote that differentiates into an ookinete in the midgut lumen. The ookinete traverses the midgut epithelial cell layer and develops into an oocyst, which produces multiple sporozoites. The oocyst ruptures to release multiple sporozoites (the human-infective parasite stage) which travel to the mosquito salivary gland and may be transferred to a human during the next bloodmeal. Sexual replication only occurs inside the mosquito vector host, and the parasite genome remains haploid for the majority of its life cycle stages apart from a brief diploid stage straight after fertilisation [74] (**Figure 2**).

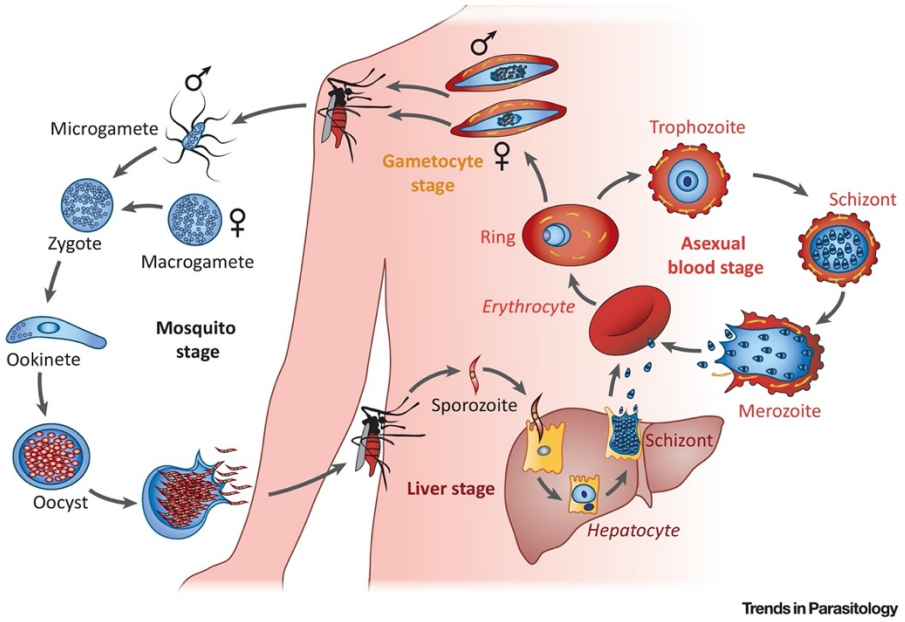


Figure 2. The life cycle of *Plasmodium falciparum* (taken from Maier *et al.* [75]).

1.3 Control measures and malaria progress

There was a substantial increase in both interest and financial support for malaria control in the 1990s, resulting in two large campaigns; the Global Malaria Strategy in 1992 [76] and Roll Back Malaria in 1998 [77]. The result of these campaigns over the course of the past three decades has been numerous interventions to tackle the burden of malaria, including mass indoor residual spraying (IRS) programmes using insecticides to kill mosquito vectors, and mass distribution of bed nets. More recent measures include insecticide treated bed nets (ITNs), to reduce the likelihood of infected mosquito vectors feeding on humans, and rapid and effective malaria treatment using artemisinin combination therapies (ACTs) since the early 2000s [25]. Additionally, seasonal malaria treatment (known as intermittent preventative treatment, IPT) in infants (IPTi), pregnant women (IPTp) and children (IPTc or SMC) is used in some sub-Saharan African countries [78–80], as well as mass drug administration (MDA) where large numbers of a chosen community are treated for malaria regardless of their infection status, however MDA has only been used to date in specific research settings and is thought to work best in regions with low malaria transmission, and the reduction in cases seen is only short term and not sustained without further rounds of treatment [81].

These combined efforts led to a drastic reduction in malaria case incidence (from 81 per 1,000 population at risk in year 2000 to 59 in 2015, and 56 in 2019) (**Figure 3**), resulting in a reduction in malaria deaths in the same period (from 896,000 deaths in 2000 to 562,000 in 2015 and 558, 000 in 2019) [1]. Since year 2000, 23 countries have reported zero indigenous cases of malaria for three consecutive years, with twelve achieving malaria-free status from the World Health Organization (WHO) [1]. There are currently 25 countries registered in the E-2025 initiative set out by the WHO in April 2021, which are aiming to stop indigenous malaria transmission by 2025 [1]. Whilst there has been remarkable progress over the last three decades, it is evident that the current measures are not sufficient to eliminate malaria in many settings, this is further compounded by the threat of resistance to current measures, both within the parasite and vector populations. ITNs were modelled to be the most effective anti-malarial intervention between years 2000 and 2015, responsible for 68% of the averted deaths [82]. Therefore resistance to pyrethroid insecticides (the most widely used chemical on ITNs), which is now found in many regions in Africa, is of great concern, although fortunately nets currently remain an effective control strategy even where insecticide resistance is found [83]. Additionally, the front-line treatment for *P. falciparum* infections is ACT, which has reduced efficacy in many regions in Southeast Asia due to mutations within the kelch propeller gene, *Pfk13* [22]. Additionally, an amino acid substitution in *Pfk13*, R561H, was found in Rwanda, which *in vitro* testing has demonstrated to be associated with artemisinin resistance [84], further challenging malaria control.

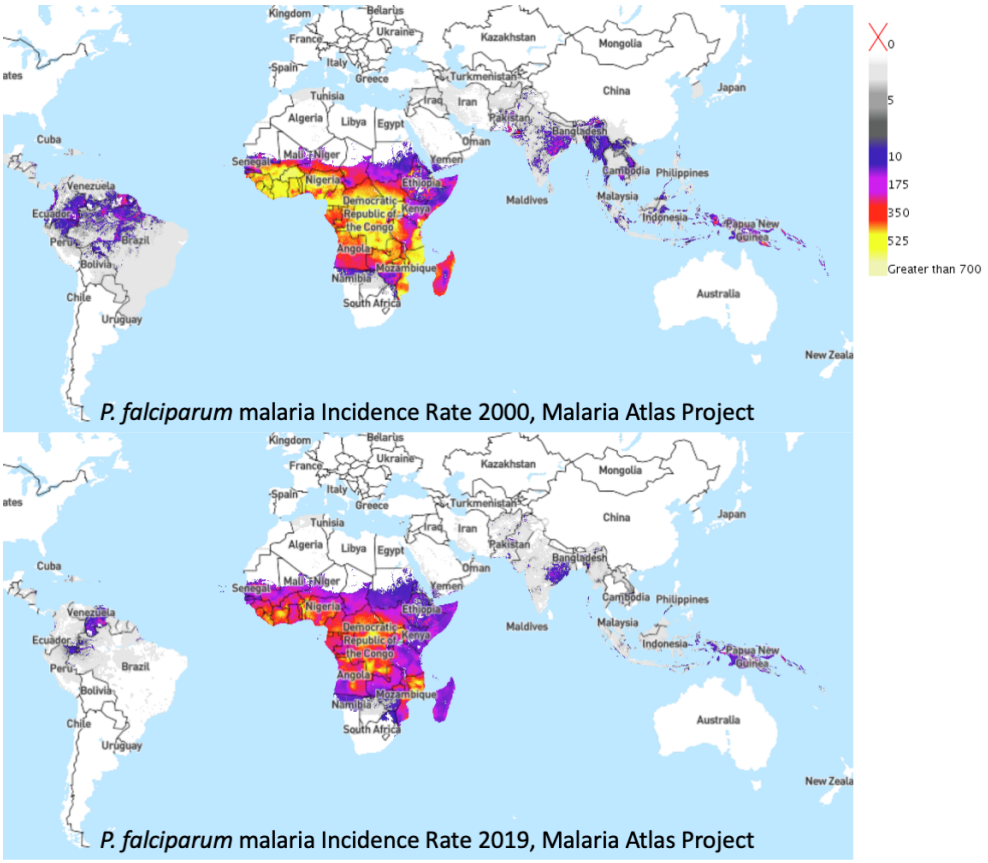


Figure 3. Decrease in incidence rate of *P. falciparum* malaria between years 2000 and 2019. Maps created using open-source data from the Malaria Atlas Project, with the *P. falciparum* incidence rate (cases of *P. falciparum* infections per 1,000 individuals per annum) highlighted using the colour scale [85].

A marker of success for a malaria control programme is a reduction in the incidence of disease, but this brings additional complications to each setting if control measures are not sustained. As malaria case incidence decreases in high transmission areas, immunity to malaria infection in the local population will also decrease, leaving a population more susceptible to severe disease if control measures are not sustained and the population is once again exposed to infection. Therefore, it is vital in settings undergoing malaria control programmes that measures are sustained, and the parasite population is monitored closely. Unfortunately, control measures are commonly halted, with decreased investment in malaria control occurring in many regions, potentially due to reduced infection prevalence (with malaria then seen as a minor threat), or more recently because of the COVID-19 pandemic, which created strain on both drug and net distribution, in addition to strains on healthcare systems [1].

1.4 *P. falciparum* and drug resistance

1.4.1 Chloroquine

Drug resistance has been a major challenge for malaria control since drugs were initially used to combat this disease. Chloroquine, a drug derived from quinine, was an early antimalarial drug used at a mass scale at the start of the 1950s. Unfortunately, towards the end of the decade, *P. falciparum* drug resistance had emerged in both Colombia and the Cambodia-Thailand border, and was present in the African continent in the 1970s [86]. Chloroquine resistance occurs due to single nucleotide polymorphisms (SNPs) in the chloroquine resistance transporter gene (*pfcrt*), with the most significant mutation leading to an amino acid substitution of lysine to threonine (K76T) [87]. The five amino acids at positions 72-76 create an important haplotype associated to the geographical origin of their resistance phenotype. The CVIET haplotype at positions 72-76 represents resistant parasites from Asia and Africa, whereas SVMNT is associated with resistant parasites in Papua New Guinea and South America, the CVMNK haplotype is sensitive to chloroquine [88]. The most widely accepted mechanism of chloroquine action is in the digestive vacuole of the Plasmodium parasite, where chloroquine prevents the degradation of haematin (a toxic product of haem degradation) into haemozoin (non-toxic product of haem degradation), leading to an accumulation of toxic compounds in the parasite digestive vacuole, leading to parasite cell death [89]. Resistance mutations in *pfcrt* are thought to inhibit the transport of chloroquine into the digestive vacuole, reducing the inhibition of haem degradation [90]. Chloroquine resistance is exacerbated by SNPs in the multi-drug resistance transporter gene (*pfmdr1*), including a SNP which leads to amino acid substitution of tyrosine to asparagine at position 86 (Y86N) [91,92], in addition to associations of mefloquine resistance with increased copy numbers of *pfmdr1* [93].

1.4.2 Antifolates

Following chloroquine resistance, a combination of two previously categorised antimalarials, sulfadoxine and pyrimethamine became a widely used antimalarial for the treatment of *P. falciparum* infections, known as SP. SP targets two enzymes within the folate biosynthesis pathway: pyrimethamine targets dihydrofolate reductase (PfDHFR), and sulfadoxine targets

dihydropteroate synthase (PfDHPS). SP resistance was quick to develop and spread, and is complex, involving cumulative SNPs within both genes. Highly resistant parasites commonly demonstrate a triple amino acid substitution within PfDHFR (N51I, C59R and S108N) and either a double (A437G, K540E) or triple (A437G, K540E, A581G) amino acid substitution within PfDHPS, known as quintuple or sextuple mutations respectively [94]. Whilst SP alone is no longer used as a first-line antimalarial for *P. falciparum* infections due to the wide presence of resistant parasites, it is used as the partner drug in ACTs, in addition to routine use for intermittent preventative treatment in infants (IPTi) and pregnant women (IPTp) in many areas, therefore continued surveillance of resistance mutations is essential.

1.4.3 Artemisinin derivatives and combination therapy

Since year 2001, ACTs have been the recommended treatment for *P. falciparum* malaria. ACTs comprise of an artemisinin-derivative, which rapidly kills Plasmodium parasites, combined with a longer-acting partner drug [25]. Artemisinin resistance (defined as slow parasite clearance when using an artemisinin derivative) was rapid to develop in the Greater Mekong region, where the main causative mutation leads to the amino acid substitution of a cysteine to a tyrosine (C580Y) within the kelch propeller domain protein (*pfk13*) [23,95,96]. This is in contrast to the dominant resistance mutation in Myamnar (F446I) [97]. ACT efficacy is further complicated in Southeast Asia due to partner drug resistance, as seen in the case of piperaquine, which is associated with amplifications within the *plasmepsin 2* and *3* genes, polymorphisms within the exonuclease gene, PF3D7_1362500 (E415G) [98], and mutations within *pfcrt* [99–101]. Additionally, mefloquine resistance is present in Southeast Asia and associated with increased copy number of the *pfmdr1* gene [102]. Currently, ACTs are still recommended as the first-line treatment across most of the world, as resistance is not complete, and parasites are usually cleared thanks to the longer acting partner drug (providing that there is no partner drug resistance). For most antimalarial drugs, resistance has developed in the Greater Mekong Subregion, before spreading across the globe, therefore The WHO implemented the Global Plan for Artemisinin Resistance Containment, a strategy to prevent the spread of resistance, which was followed by an action to eliminate all malaria in the Greater Mekong Subregion region by 2030 [103]. Additionally, artemisinin resistance has independently arisen in the African continent, associated with amino acid substitution R561H in Rwanda [84], in addition to substitutions A675V and C469Y in Uganda [24]. The C580Y mutation

within *pfk13* has emerged independently in Guyana, however, it is thought to be associated with a fitness cost to the parasite in this setting [104].

1.5 Non-falciparum resistance

In the case of *P. vivax*, reduced susceptibility to chloroquine was first documented in 1989 [105], and has now been reported in most geographical regions where this parasite is transmitted [106], however the molecular marker for resistance in *P. vivax* is not described, and it is thought to involve a different pathway to *P. falciparum,* with the *P. vivax pvcrt-o* orthologue not implicated [107]. Regardless of growing resistance, chloroquine is still the main antimalarial used to treat *P. vivax* infections globally, with the exception of Indonesia, Sabah and Papua New Guinea where ACTs are recommended [25,30,108]. Mefloquine resistance is another concern when regarding *P. vivax* infections, with increased copy numbers of *pvmdr1* associated with reduced susceptibility to amodiaquine, artesunate and mefloquine [109]. Primaquine and Tafeonquine are currently the only hypnozoiticidal drugs used to prevent relapses for *P. vivax* and *P. ovale spp.* infections, and whilst treatment failures do occur, this may be due to host factors leading to sub-optimal levels of active drug [110], rather than resistance mechanisms within the parasite population. Pyrimethamine resistance, due to SNPs in *pvdhfr* is found within *P. vivax* isolates across the globe, with mutations at positions 58 and 117 potentially implicated [111,112]. Innate parasite differences in *pvdhps* leads to reduced affinity of sulfadoxine binding, and reduced susceptibility [113], therefore SP is not recommended as a first line antimalarial for the treatment of *P. vivax*.

Chloroquine resistance in the other human-infective *Plasmodium spp.* is rarely described, however this may be due to lacking research, and difficulties in performing laboratory-based drug susceptibility tests in parasite species without a stable culture method, rather than a lack of resistant phenotypes. Reports of treatment failures with chloroquine were seen in *P. malariae* infections in Indonesia, in clinical cases where adequate therapeutic levels of chloroquine were seen in the bloodstream, however no genotyping was performed on these isolates to determine a causative molecular marker [114]. Additionally, whilst no molecular markers have been confirmed, reduced susceptibility to ACTs in *P. malariae* isolates is of great concern [47,52,115]. Additionally, point mutations in *pmdhfr-ts* have been previously described, although no resistance phenotype was reported with these isolates [116].

1.6 Genomics to aid malaria control

Plasmodium parasites are continually evolving to adapt to both the human and mosquito hosts, with genetic variation occurring through SNPs, small genome insertions or deletions (indels), duplications and large-scale structural variations. Genetic diversity in *Plasmodium* parasites is due to recombination, with high rates demonstrated in *P. falciparum* [117] . Individual parasite genomes in a given population may be diverse, with the likelihood of polyclonal infections arising in infected individuals in areas of high transmission, with one study demonstrating an individual infected by up to three different clones of Plasmodium from the bite of one or multiple mosquitoes [118].

Parasite genetic diversity is fundamental to the parasite's ability to evade drug treatment, immune, and vaccine clearance, with genetic variants that provide an advantage against such challenges selected for and increased in proportion in a population (providing there is no fitness cost associated with this variant). Genomic surveillance of parasite populations allows us to predict the potential efficacy of certain control measures (e.g., we would not choose to implement a mass drug administration programme in an area with high levels of resistance mutations), in addition to assessing whether current measures are influencing the genome of parasites in a particular setting.

There are reference genomes available for all currently known human-infective *Plasmodium spp.* (**Table 1**), all comprising a nuclear genome of 14 chromosomes, and two organellar genomes, the mitochondria and apicoplast [71,119–121]. *P. falciparum* was the first eukaryotic parasite to have a published whole genome sequence (WGS) in 2002 [119]. Since this date, there are now > 7,000 publicly available sequences which has allowed for large scale population genetics studies [69]. These data have provided important insights into the biology and genetic diversity of *P. falciparum* and signatures of selection have been repeatedly detected around known drug resistance genes. For example, the molecular marker for artemisinin resistance, *pfk13,* and they key associated mutation in SEA (C580Y) was identified through WGS using parasites growing *in vitro* under exposure to escalating artemisinin [95], and also confirmed in field isolates from the Greater Mekong Subregion through WGS alongside patient clinical data on treatment efficacy [122].

There are many different methods available for WGS, each with its own advantages. The Illumina system, which is based on sequencing by synthesis, is used within this project as this method is robust for sequencing of Plasmodium parasites and allows for multiplexing of multiple samples in one run.

Most genomic research into malaria has been focussed on *P. falciparum*, this is in part due to the pathogenicity and resulting research interest into this species, but also due to the availability of an *in vitro* culture, which allows large amounts of parasite DNA to be generated. For non-falciparum malaria, obtaining sufficient parasite DNA from clinical blood samples (contaminated with human DNA) has been a significant challenge, however the use of leukocyte depletion methods [123] and selective whole genome amplification (SWGA) has enabled more efficient sequencing and has led to a growth in sequence data for *P. vivax* [124] and *P. falciparum* isolates [125]. Using SWGA, multiple *P. knowlesi* isolates have been sequenced, allowing for preliminary findings into population structure, which have shown three distinct population subgroups in Malaysia, associated with the isolate location and the macaque host, which undergo complex genomic recombination [72].

The most neglected malaria parasites *P. malariae, Poc* and *Pow* are yet to be investigated at large scale, with <10 publicly available genomes prior to my research project (**Table 1**). Initial insights into the population genomics of *P. malariae* using microsatellite data have demonstrated high levels of genetic diversity, with geographical separation of parasite populations [126,127].This is in contrast to initial investigations of *Poc* and *Pow,* which appear to demonstrate low levels of genetic diversity [53]. However, microsatellite markers only reflect a minority of the genome (<0.1%), and greater levels of information can be obtained from WGS.

Non-falciparum malaria presents a distinct challenge regarding malaria control, whereby many infections are commonly undetectable through common diagnostics, due to low parasitaemias and complex mixed infections, leading to untraceable malaria transmission (this is particularly seen in the case of asymptomatic infections). Molecular based diagnostics are therefore crucial when investigating the prevalence of non-falciparum malaria, however these approaches are more expensive and require advanced laboratory expertise. Persistent infections are common in the case

of *P. malariae,* and relapsing infections with *P. vivax* and *Poc* and *Pow* infections, which, among other factors, create more complex and difficult infections to control in a population.

In summary, we are now in a prime position to perform WGS and genomic diversity studies of the neglected parasites. Several technical roadblocks have been overcome, including the availability of reference genomes and the ability to amplify parasite DNA for WGS, specifically from infections presenting with low parasitaemias. An increase in genomic knowledge of non-falciparum parasites is essential to understand the biology of these species and to enable the design of disease control measures to target these neglected and understudied parasites.

1.7 Thesis structure

This thesis focuses on the investigation of the genomic diversity observed in *P. malariae* and *P. vivax* clinical isolates to improve our understanding of the biology and pathogenesis of these parasite species.

The following five chapters correspond to:

1.7.1 Chapter 2 (published paper)

Before this thesis, only a few genomes were publicly available for *P. malariae*. This is due to the challenge in obtaining enough parasite DNA from low parasitaemia infections observed in individuals and the lack of a stable *in vitro* culture method. In this chapter, I present the development and validation of a methodology for selective whole genome amplification (SWGA) of *P. malariae* DNA from clinical samples, enabling the first investigation into genomic diversity of this parasite species. Using a small data set, I demonstrate geographical separation of parasites in addition to SNPs within orthologs of genes associated with reduced drug susceptibility.

1.7.2 Chapter 3 (manuscript submitted)

In this chapter, I utilise the SWGA method to perform the first global analysis of *P. malariae* genomic diversity, using novel sequence data generated for 155 high quality samples from clinical isolates across four continents. Using this large dataset, I investigate the population structure, identify signatures of selection, and uncover numerous genomic loci for further investigation.

Additionally, I confirm the presence of multiple SNPs within orthologs of genes associated with drug susceptibility.

1.7.3 Chapter 4 (book-style thesis chapter)

P. malariae does not have a long-term *in vitro* culture model to perform drug assays or to genetically manipulate parasites. Therefore, in this chapter I describe an *in vitro* system using ortholog replacement through CRISPR-Cas9 mediated genome editing in the culture-adapted *P. knowlesi* A1H1 line, to establish the phenotypic effect of genomic variants identified in the *P. malariae pmdhfr-ts* gene. I constructed a range of genetically modified parasite lines and demonstrate a pyrimethamine-resistant *P. malariae* parasite phenotype.

1.7.4 Chapter 5 (manuscript submitted)

In this chapter I explore the genomic diversity of the understudied populations of *P. vivax* across Brazil, compared with isolates from other continents in the largest assembled global dataset of *P. vivax* to date. I demonstrate complex population structure of *P. vivax* within Brazil and uncover potentially relevant loci that may associated with drug susceptibility in this parasite species.

1.7.5 Chapter 6

This chapter contains the thesis discussion and potential future directions.

1.8 References

1. World Health Organization. *World Malaria Report 2021*. (2021).
2. Rahimi, A., Kassam, R., Dang, Z. & Sekiwunga, R. Challenges with accessing health care for young children presumed tohave malaria in the rural district of Butaleja, Uganda: a qualitativestudy. *Pharm. Pract. (Granada).* **17**, (2019).
3. Luzzatto, L. Sickle cell anaemia and malaria. *Mediterr. J. Hematol. Infect. Dis.* **4**, (2012).
4. Langhi, D. M. & Bordin, J. O. Duffy blood group and malaria. *Hematology* **11**, 389–398 (2006).
5. WHO. *World Malaria Report 2020*. (2020).
6. Sinka, M. E. *et al.* A global map of dominant malaria vectors. *Parasites and Vectors* **5**, 1–11 (2012).
7. Dobbs, K. R. & Dent, A. E. Plasmodium malaria and antimalarial antibodies in the first year of life. *Parasitology* **143**, 129 (2016).
8. Franks, S. *et al.* Frequent and persistent, asymptomatic Plasmodium falciparum infections in African infants, characterized by multilocus genotyping. *J. Infect. Dis.* **183**, 796–804 (2001).
9. Doolan, D. L., Dobaño, C. & Baird, J. K. Acquired Immunity to Malaria. *Clin. Microbiol. Rev.* **22**, 13 (2009).
10. Suwonkerd, W. *et al.* Vector Biology and Malaria Transmission in Southeast Asia. *Anopheles mosquitoes - New insights into Malar. vectors* (2013). doi:10.5772/56347
11. White, N. J. *et al.* Malaria. *Lancet* **383**, 723–35 (2014).
12. Conn, J. E., Grillet, M. E., Correa, M. & Sallum, M. A. M. Malaria Transmission in South America—Present Status and Prospects for Elimination. *Towar. Malar. Elimin. - A Leap Forw.* (2018). doi:10.5772/INTECHOPEN.76964
13. Vittor, A. Y. *et al.* Linking Deforestation to Malaria in the Amazon: Characterization of the Breeding Habitat of the Principal Malaria Vector, Anopheles darlingi. *Am. J. Trop. Med. Hyg.* **81**, 5 (2009).
14. Castellanos, A. *et al.* Malaria in gold-mining areas in Colombia. *Mem Inst Oswaldo Cruz, Rio Janeiro* **111**, 59–66 (2016).
15. Lee, W. C., Russell, B. & Rénia, L. Sticking for a cause: The falciparum malaria parasites cytoadherence paradigm. *Front. Immunol.* **10**, 1444 (2019).
16. Health Organization, W. Management of severe malaria – A practical handbook. Third edition.
17. Clemence, J. *et al.* Cell biology and immunology of malaria. *Immunol. Rev.* **240**, 297–316 (2011).
18. Cowan, G. O., Friman, G. & Günther, G. *Manson's tropical diseases. Manson's Tropical Diseases* (Saunders, 2009).
19. McCaffery, J. N. *et al.* Plasmodium falciparum pfhrp2 and pfhrp3 gene deletions among patients in the DRC enrolled from 2017 to 2018. *Sci. Reports 2021 111* **11**, 1–8 (2021).
20. Nolder, D. *et al.* Failure of rapid diagnostic tests in Plasmodium falciparum malaria cases among travelers to the UK and Ireland: Identification and characterisation of the parasites. *Int. J. Infect. Dis.* **108**, 137 (2021).
21. Beshir, K. B. *et al.* Plasmodium falciparum parasites with histidine-rich protein 2 (pfhrp2) and pfhrp3 gene deletions in two endemic regions of Kenya. *Sci. Rep.* **7**, (2017).
22. Phyo, A. P. & Nosten, F. The Artemisinin Resistance in Southeast Asia: An Imminent

Global Threat to Malaria Elimination. *Towar. Malar. Elimin. - A Leap Forw.* (2018). doi:10.5772/INTECHOPEN.76519
23. Imwong, M. *et al.* The spread of artemisinin-resistant Plasmodium falciparum in the Greater Mekong subregion: a molecular epidemiology observational study. *Lancet. Infect. Dis.* **17**, 491 (2017).
24. Balikagala, B. *et al.* Evidence of Artemisinin-Resistant Malaria in Africa. *N. Engl. J. Med.* **385**, 1163–1171 (2021).
25. WHO. Guidlines for the treatment of malaria - 3rd edition. **85**, (2015).
26. Ménard, D. *et al.* Plasmodium vivax clinical malaria is commonly observed in Duffy-negative Malagasy people. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 5967–5971 (2010).
27. Golassa, L., Amenga-Etego, L., Lo, E. & Amambua-Ngwa, A. The biology of unconventional invasion of Duffy-negative reticulocytes by Plasmodium vivax and its implication in malaria epidemiology and public health. *Malar. J.* **19**, 1–10 (2020).
28. Anstey, N. M., Russell, B., Yeo, T. W. & Price, R. N. The pathophysiology of vivax malaria. *Trends Parasitol.* **25**, 220–227 (2009).
29. Howes, R. E. *et al.* Global epidemiology of Plasmodium vivax. *Am. J. Trop. Med. Hyg.* **95**, 15–34 (2016).
30. Chu, C. S. & White, N. J. The prevention and treatment of Plasmodium vivax malaria. *PLOS Med.* **18**, e1003561 (2021).
31. Peters, A. L. & Van Noorden, C. J. F. Glucose-6-phosphate Dehydrogenase Deficiency and Malaria: Cytochemical Detection of Heterozygous G6PD Deficiency in Women. *J. Histochem. Cytochem.* **57**, 1003–1011 (2009).
32. Gething, P. W. *et al.* Modelling the global constraints of temperature on transmission of Plasmodium falciparum and P. vivax. (2011). doi:10.1186/1756-3305-4-92
33. Battle, K. E. *et al. The Global Public Health Significance of Plasmodium vivax*. *Advances in Parasitology* **80**, (Elsevier, 2012).
34. Bousema, T. & Drakeley, C. Epidemiology and Infectivity of Plasmodium falciparum and Plasmodium vivax Gametocytes in Relation to Malaria Control and Elimination. *Clin. Microbiol. Rev.* **24**, 377–410 (2011).
35. RN, P., RJ, C., KE, B., K, T. & K, M. Plasmodium vivax in the Era of the Shrinking P. falciparum Map. *Trends Parasitol.* **36**, 560–570 (2020).
36. Ding, X. C. *et al.* Defining the next generation of Plasmodium vivax diagnostic tests for control and elimination: Target product profiles. *Jetsumon Sattabongkot* **10**, (2017).
37. Baird, J. K., Valecha, N., Duparc, S., White, N. J. & Price, R. N. Diagnosis and Treatment of Plasmodium vivax Malaria. *Am. J. Trop. Med. Hyg.* **95**, 35 (2016).
38. Abba, K. *et al.* Rapid diagnostic tests for diagnosing uncomplicated non-falciparum or Plasmodium vivax malaria in endemic countries. *Cochrane database Syst. Rev.* **2014**, (2014).
39. Malleret, B. *et al.* Plasmodium vivax: restricted tropism and rapid remodeling of CD71-positive reticulocytes. *Blood* **125**, 1314–1324 (2015).
40. Collins, W. E. & Jeffery, G. M. Plasmodium malariae: Parasite and disease. *Clinical Microbiology Reviews* (2007). doi:10.1128/CMR.00027-07
41. Badiane, A. S. *et al.* Acute kidney injury associated with Plasmodium malariae infection. *Malar. J.* **13**, 1–5 (2014).
42. Hendrickse, R. G. & Adeniyi, A. Quartan malarial nephrotic syndrome in children. *Kidney Int.* **16**, 64–74 (1979).

43. Marteau, A. *et al.* Severe long-delayed malaria caused by Plasmodium malariae in an elderly French patient. *Malar. J.* **20**, 1–5 (2021).
44. Langford, S. *et al.* Plasmodium malariae Infection Associated with a High Burden of Anemia: A Hospital-Based Surveillance Study. *PLoS Negl. Trop. Dis.* (2015). doi:10.1371/journal.pntd.0004195
45. Scopel, K. K. G., Fontes, C. J. F., Nunes, Á. C., Horta, M. F. & Braga, É. M. High prevalence of Plamodium malariae infections in a Brazilian Amazon endemic area (Apiacás - Mato Grosso State) as detected by polymerase chain reaction. *Acta Trop.* **90**, 61–64 (2004).
46. Lo, E. *et al.* Plasmodium malariae Prevalence and csp Gene Diversity, Kenya, 2014 and 2015. *Emerg. Infect. Dis.* **23**, 601 (2017).
47. Betson, M. *et al.* Detection of persistent Plasmodium spp. infections in Ugandan children after artemether-lumefantrine treatment. *Parasitology* **141**, 1880–1890 (2014).
48. Niño, C. H. *et al.* Plasmodium malariae in the Colombian Amazon region: you don't diagnose what you don't suspect. *Malar. J.* (2016). doi:10.1186/s12936-016-1629-3
49. Doritchamou, J. Y. A. *et al.* Submicroscopic placental infection by non-falciparum Plasmodium spp. *PLoS Negl. Trop. Dis.* **12**, e0006279 (2018).
50. Oseph, J. *et al.* Plasmodium malariae Infection in an Asymptomatic 74-Year-Old Greek Woman with Splenomegaly. *http://dx.doi.org/10.1056/NEJM199802053380605* **338**, 367–371 (2009).
51. Maguire, J. D. *et al.* Chloroquine-resistant Plasmodium malariae in south Sumatra, Indonesia. *Lancet* (2002). doi:10.1016/S0140-6736(02)09336-4
52. Teo, B. H. Y. *et al.* Delayed Onset of Symptoms and Atovaquone-Proguanil Chemoprophylaxis Breakthrough by Plasmodium malariae in the Absence of Mutation at Codon 268 of pmcytb. *PLoS Negl. Trop. Dis.* **9**, 1–12 (2015).
53. Sutherland, C. J. *et al.* Two Nonrecombining Sympatric Forms of the Human Malaria Parasite *Plasmodium ovale* Occur Globally. *J. Infect. Dis.* **201**, 1544–1550 (2010).
54. Kotepui, M., Kotepui, K. U., Milanez, G. D. & Masangkay, F. R. Severity and mortality of severe Plasmodium ovale infection: A systematic review and meta-analysis. *PLoS One* **15**, e0235014 (2020).
55. Lau, Y. L. *et al.* Acute respiratory distress syndrome and acute renal failure from Plasmodium ovale infection with fatal outcome. *Malar. J.* **12**, (2013).
56. Collins, W. E. & Jeffery, G. M. Plasmodium ovale:Parasite and Disease. *Clin. Microbiol. Rev.* **18**, 570 (2005).
57. Nolder, D. *et al.* An observational study of malaria in British travellers: Plasmodium ovale wallikeri and Plasmodium ovale curtisi differ significantly in the duration of latency. *BMJ Open* **3**, 1–6 (2013).
58. Rojo-Marcos, G. *et al.* Comparison of imported Plasmodium ovale curtisi and P. ovale wallikeri infections among patients in Spain, 2005-2011. *Emerg. Infect. Dis.* **20**, 409–416 (2014).
59. Mahittikorn, A., Masangkay, F. R., Kotepui, K. U., Milanez, G. D. J. & Kotepui, M. Comparison of Plasmodium ovale curtisi and Plasmodium ovale wallikeri infections by a meta-analysis approach. *Sci. Reports 2021 111* **11**, 1–15 (2021).
60. Yerlikaya, S., Campillo, A. & Gonzalez, I. J. A Systematic Review: Performance of Rapid Diagnostic Tests for the Detection of Plasmodium knowlesi, Plasmodium malariae, and Plasmodium ovale Monoinfections in Human Blood. *J. Infect. Dis.* **218**, 265–276 (2018).

61. Tang, J. *et al.* Assessment of false negative rates of lactate dehydrogenase-based malaria rapid diagnostic tests for Plasmodium ovale detection. *PLoS Negl. Trop. Dis.* **13**, e0007254 (2019).
62. Cooper, D. J. *et al.* Plasmodium knowlesi Malaria in Sabah, Malaysia, 2015–2017: Ongoing Increase in Incidence Despite Near-elimination of the Human-only Plasmodium Species. *Clin. Infect. Dis. An Off. Publ. Infect. Dis. Soc. Am.* **70**, 361 (2020).
63. Singh, B. & Daneshvar, C. Human infections and detection of plasmodium knowlesi. *Clin. Microbiol. Rev.* **26**, 165–184 (2013).
64. Grigg, M. J. *et al.* Individual-level factors associated with the risk of acquiring human Plasmodium knowlesi malaria in Malaysia: a case-control study Europe PMC Funders Group. *Lancet Planet Heal.* **1**, 97–104 (2017).
65. Moyes, C. L. *et al.* Predicting the geographical distributions of the macaque hosts and mosquito vectors of Plasmodium knowlesi malaria in forested and non-forested areas. *Parasites and Vectors* **9**, 1–12 (2016).
66. Kantele, A. & Jokiranta, T. S. Review of cases with the emerging fifth human malaria parasite, Plasmodium knowlesi. *Clin. Infect. Dis.* **52**, 1356–1362 (2011).
67. Jeremiah, S., Janagond, A. B. & Parija, S. C. Challenges in diagnosis of Plasmodium knowlesi infections. *Trop. Parasitol.* **4**, 25–30 (2014).
68. World Health Organization. Expert consultation on Plasmodium knowlesi to guide malaria elimination strategies. *Expert Consult. Plasmodium knowlesi to Guid. Malar. Elimin. Strateg.* 1–33 (2017).
69. Pearson, R. D. *et al.* An open dataset of Plasmodium falciparum genome variation in 7,000 worldwide samples. *bioRxiv* (2019). doi:10.1101/824730
70. Benavente, E. D. *et al.* Distinctive genetic structure and selection patterns in Plasmodium vivax from South Asia and East Africa. *Nat. Commun. 2021 121* **12**, 1–11 (2021).
71. Rutledge, G. G. *et al.* Plasmodium malariae and P. ovale genomes provide insights into malaria parasite evolution. *Nature* **542**, 101–104 (2017).
72. Benavente, E. D. *et al.* Whole genome sequencing of amplified Plasmodium knowlesi DNA from unprocessed blood reveals genetic exchange events between Malaysian Peninsular and Borneo subpopulations. *Sci. Rep.* **9**, 9873 (2019).
73. 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 (2020).
74. Aly, A. S. I., Vaughan, A. M. & Kappe, S. H. I. Malaria Parasite Development in the Mosquito and Infection of the Mammalian Host. *Annu. Rev. Microbiol.* **63**, 195 (2009).
75. Maier, A. G., Matuschewski, K., Zhang, M. & Rug, M. Plasmodium falciparum. *Trends Parasitol.* **35**, 481–482 (2019).
76. World Health Organization. A global strategy for malaria control. *Geneva* **88**, 604 (1993).
77. Nabarro, D. N. & Tayler, E. M. The 'roll back malaria' campaign. *Science* **280**, 2067–2068 (1998).
78. World Health Organization. WHO Global Malaria Programme WHO Policy Recommendation: Seasonal Malaria Chemoprevention (SMC) for Plasmodium falciparum malaria control in highly seasonal transmission areas of the Sahel sub-region in Africa. (2012).
79. World Health Organization. WHO Policy recommendation on Intermittent Preventive Treatment during infancy with sulphadoxine-pyrimethamine (SP-IPTi) for Plasmodium

falciparum malaria control in Africa. (2010).
80. World Health Organization. A Strategic Framework for Malaria Prevention and Control During Pregnancy in the African Region WORLD HEALTH ORGANIZATION Regional Office for Africa. (2004).
81. Shah, M. P. *et al.* Mass drug administration for malaria. *Cochrane Database Syst. Rev.* **2021**, (2021).
82. Bhatt, S. *et al.* The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature* **526**, 207–211 (2015).
83. Lindsay, S. W., Thomas, M. B. & Kleinschmidt, I. Threats to the effectiveness of insecticide-treated bednets for malaria control: thinking beyond insecticide resistance. *Lancet Glob. Heal.* **9**, e1325–e1331 (2021).
84. Uwimana, A. *et al.* Emergence and clonal expansion of in vitro artemisinin-resistant Plasmodium falciparum kelch13 R561H mutant parasites in Rwanda. *Nat. Med.* **26**, 1602–1608 (2020).
85. Malaria Atlas Project. Malaria Atlas Project - MAP. Available at: https://malariaatlas.org/. (Accessed: 25th July 2022)
86. Payne, D. Spread of chloroquine resistance in Plasmodium falciparum. *Parasitol. Today* **3**, 241–246 (1987).
87. 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 (2000).
88. Tan, L. L., Lau, T. Y., Timothy, W. & Prabakaran, D. Full-length sequence analysis of chloroquine resistance transporter gene in plasmodium falciparum isolates from Sabah, Malaysia. *Sci. World J.* **2014**, (2014).
89. Slater, A. F. G. Chloroquine: Mechanism of drug action and resistance in plasmodium falciparum. *Pharmacol. Ther.* **57**, 203–235 (1993).
90. Ridley, R. G. Malaria: Dissecting chloroquine resistance. *Curr. Biol.* **8**, R346–R349 (1998).
91. Babiker, H. A. *et al.* High-level chloroquine resistance in Sudanese isolates of Plasmodium falciparum is associated with mutations in the chloroquine resistance transporter gene pfcrt and the multidrug resistance Gene pfmdr1. *J. Infect. Dis.* **183**, 1535–1538 (2001).
92. Duraisingh, M. T. & Cowman, A. F. Contribution of the pfmdr1 gene to antimalarial drug-resistance. *Acta Trop.* **94**, 181–190 (2005).
93. Foote, S. J., Thompson, J. K., Cowman, A. F., Kemp, D. J. & Hall, E. Amplification of the Multidrug Resistance Gene in Some Chloroquine-Resistant Isolates of F? falciparum. *Cell* **57**, 921–930 (1989).
94. Chaturvedi, R. *et al.* Geographical spread and structural basis of sulfadoxine-pyrimethamine drug-resistant malaria parasites. *Int. J. Parasitol.* **51**, 505–525 (2021).
95. Ariey, F. *et al.* A molecular marker of artemisinin-resistant Plasmodium falciparum. *Nature* **505**, 50–55 (2016).
96. Takala-Harrison, S. *et al.* Genetic loci associated with delayed clearance of plasmodium falciparum following artemisinin. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 240–245 (2013).
97. Imwong, M. *et al.* Molecular epidemiology of resistance to antimalarial drugs in the Greater Mekong subregion: an observational study. *Lancet Infect. Dis.* **20**, 1470–1480 (2020).

98. 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* **3099**, 164–173 (2017).
99. Ross, L. S. *et al.* Emerging Southeast Asian PfCRT mutations confer Plasmodium falciparum resistance to the first-line antimalarial piperaquine. doi:10.1038/s41467-018-05652-0
100. 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 (2017).
101. Dhingra, S. K. *et al.* A Variant PfCRT Isoform Can Contribute to Plasmodium falciparum Resistance to the First-Line Partner Drug Piperaquine. *MBio* **8**, (2017).
102. Price, R. N. *et al.* Mefloquine resistance in Plasmodium falciparum and increased pfmdr1 gene copy number. *Lancet* **364**, 438 (2004).
103. for the Western Pacific, W. H. O. R. O. *Strategy for malaria elimination in the Greater Mekong Subregion : 2015-2030*. (WHO Regional Office for the Western Pacific, 2015).
104. Mathieu, L. C. *et al.* Local emergence in Amazonia of Plasmodium falciparum k13 C580Y mutants associated with in vitro artemisinin resistance. *Elife* **9**, 1–21 (2020).
105. Rieckmann, K. H., Davis, D. R. & Hutton, D. C. Plasmodium vivax resistance to chloroquine? *Lancet (London, England)* **2**, 1183–1184 (1989).
106. 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).
107. Nomura, T. *et al.* Evidence for different mechanisms of chloroquine resistance in 2 Plasmodium species that cause human malaria. *J. Infect. Dis.* **183**, 1653–1661 (2001).
108. Baird, J. K. Resistance to Therapies for Infection by Plasmodium vivax. *Clin. Microbiol. Rev.* **22**, 508 (2009).
109. Suwanarusk, R. *et al.* Amplification of pvmdr1 Associated with Multidrug-Resistant Plasmodium vivax. (2008). doi:10.1086/592451
110. Bennett, J. W. *et al.* Primaquine Failure and Cytochrome P-450 2D6 in Plasmodium vivax Malaria. *n engl j med* **14**, 369 (2013).
111. Imwong, M. *et al.* Association of Genetic Mutations in Plasmodium vivax dhfr with Resistance to Sulfadoxine-Pyrimethamine: Geographical and Clinical Correlates. **45**, 3122–3127 (2001).
112. 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).
113. 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).
114. Maguire, J. D. *et al.* Chloroquine-resistant Plasmodium malariae in south Sumatra, Indonesia. *Lancet* **360**, 58–60 (2002).
115. Betson, M., Clifford, S., Stanton, M., Kabatereine, N. B. & Stothard, J. R. Emergence of Nonfalciparum Plasmodium Infection Despite Regular Artemisinin Combination Therapy in an 18-Month Longitudinal Study of Ugandan Children and Their Mothers. *J. Infect. Dis.* **217**, 1099–1109 (2018).
116. Tanomsing, N. *et al.* Genetic Variability of Plasmodium malariae dihydropteroate synthase (dhps) in four Asian countries. *PLoS One* **9**, 7–10 (2014).

117. Jiang, H. *et al.* High recombination rates and hotspots in a Plasmodium falciparum genetic cross. *Genome Biol.* **12**, (2011).
118. Ohrt, C., Mirabelli-Primdahl, L., Karnasuta, C., Chantakulkij, S. & Kain, K. C. Distinguishing Plasmodium falciparum treatment failures from reinfections by restrictions fragment length polymorphism and polymerase chain reaction genotyping. *Am. J. Trop. Med. Hyg.* **57**, 430–437 (1997).
119. Gardner, M. J. *et al.* Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* (2002). doi:10.1038/nature01097
120. 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).
121. Benavente, E. D. *et al.* A reference genome and methylome for the Plasmodium knowlesi A1-H.1 line. *Int. J. Parasitol.* (2017). doi:10.1016/j.ijpara.2017.09.008
122. Miotto, O. *et al.* Multiple populations of artemisinin-resistant Plasmodium falciparum in Cambodia. *Nat. Publ. Gr.* **45**, (2013).
123. Venkatesan, M. *et al.* Using CF11 cellulose columns to inexpensively and effectively remove human DNA from Plasmodium falciparum-infected whole blood samples. *Malar. J.* (2012). doi:10.1186/1475-2875-11-41
124. 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**, 1–15 (2017).
125. Oyola, S. O. *et al.* Whole genome sequencing of Plasmodium falciparum from dried blood spots using selective whole genome amplification. *Malar. J.* **15**, 597 (2016).
126. Bruce, M. C., MacHeso, A., McConnachie, A. & Molyneux, M. E. Comparative population structure of Plasmodium malariae and Plasmodium falciparum under different transmission settings in Malawi. *Malar. J.* **10**, 1–12 (2011).
127. BRUCE, M. C., MACHESO, A., GALINSKI, M. R. & BARNWELL, J. W. Characterization and application of multiple genetic markers for Plasmodium malariae. *Parasitology* **134**, 637 (2007).

CHAPTER 2

2 Selective whole genome amplification of *Plasmodium malariae* DNA from clinical samples reveals insights into population structure

RESEARCH PAPER COVER SHEET

SECTION A – Student Details

Student ID Number	**1600466**	**Title**	Miss
First Name(s)	Amy		
Surname/Family Name	Ibrahim		
Thesis Title	From genome to function: A genomic investigation into understudied populations of the malaria parasites *Plasmodium malariae* and *P. vivax*		
Primary Supervisor	Prof. Susana Campino		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Scientific Reports		
When was the work published?	2020		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	N/A		
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	

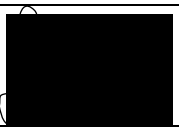
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I undertook laboratory work including the development of a selective whole genome amplification method and preparation of samples for sequencing. I also performed bioinformatic analysis and interpreted the results under the supervision of my supervisors. I wrote the first draft of the manuscript that was then circulated to supervisors and after to co-authors.

SECTION E

Student Signature	
Date	25/02/2022

Supervisor Signature	
Date	February 24, 2022

SCIENTIFIC REPORTS
nature research

OPEN

Selective whole genome amplification of *Plasmodium malariae* DNA from clinical samples reveals insights into population structure

Amy Ibrahim[1], Ernest Diez Benavente[1], Debbie Nolder[1,2], Stephane Proux[3], Matthew Higgins[1], Julian Muwanguzi[1,2], Paula Josefina Gomez Gonzalez[1], Hans-Peter Fuehrer[4], Cally Roper[1], Francois Nosten[3,5], Colin Sutherland[1], Taane G. Clark[1,6] & Susana Campino[1,7]✉

The genomic diversity of *Plasmodium malariae* malaria parasites is understudied, partly because infected individuals tend to present with low parasite densities, leading to difficulties in obtaining sufficient parasite DNA for genome analysis. Selective whole genome amplification (SWGA) increases the relative levels of pathogen DNA in a clinical sample, but has not been adapted for *P. malariae* parasites. Here we design customized SWGA primers which successfully amplify *P. malariae* DNA extracted directly from unprocessed clinical blood samples obtained from patients with *P. malariae*-mono-infections from six countries, and further test the efficacy of SWGA on mixed infections with other *Plasmodium* spp. SWGA enables the successful whole genome sequencing of samples with low parasite density (i.e. one sample with a parasitaemia of 0.0064% resulted in 44% of the genome covered by ≥ 5 reads), leading to an average 14-fold increase in genome coverage when compared to unamplified samples. We identify a total of 868,476 genome-wide SNPs, of which 194,709 are unique across 18 high-quality isolates. After exclusion of the hypervariable subtelomeric regions, a high-quality core subset of 29,899 unique SNPs is defined. Population genetic analysis suggests that *P. malariae* parasites display clear geographical separation by continent. Further, SWGA successfully amplifies genetic regions of interest such as orthologs of *P. falciparum* drug resistance-associated loci (*Pfdhfr*, *Pfdhps*, *Pfcrt*, *Pfk13* and *Pfmdr1*), and several non-synonymous SNPs were detected in these genes. In conclusion, we have established a robust SWGA approach that can assist whole genome sequencing of *P. malariae*, and thereby facilitate the implementation of much-needed large-scale multi-population genomic studies of this neglected malaria parasite. As demonstrated in other Plasmodia, such genetic diversity studies can provide insights into the biology underlying the disease and inform malaria surveillance and control measures.

Malaria, a mosquito-borne disease caused by *Plasmodium* parasites, is a continuing threat to global health. There were an estimated 228 million cases and 405,000 deaths in 2018[1]. The majority of mortality events are due to *P. falciparum* malaria and therefore disease control and elimination efforts have primarily targeted this species.

[1]Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK. [2]Public Health England (PHE) Malaria Reference Laboratory, London School of Hygiene and Tropical Medicine, London, UK. [3]Shoklo Malaria Research Unit, Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Mae Sot, Thailand. [4]Department of Pathobiology, Institute of Parasitology, University of Veterinary Medicine, Vienna, Austria. [5]Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [6]Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK. [7]Department of Infection Biology, London School of Hygiene and Tropical Medicine, London, UK. ✉email: Susana.campino@lshtm.ac.uk

Molecular surveillance has demonstrated that non-falciparum malaria has been underestimated by microscopy diagnosis[2–5], and rapid diagnostic tests (RDT), which are unable to diagnose non-falciparum malaria to the species level[6,7]. Molecular studies are beginning to demonstrate alarmingly high levels (4–24%) of *P. malariae* mono- and co-infections across continents[2,8–12].

P. malariae infections commonly present with mild or no symptoms, however, severe disease, including anaemia, renal pathologies, and splenomegaly[13–17] can occur, complications which can prove fatal[16]. *P. malariae* infections present with quartan fevers with parasites that can remain in the host for decades[13,18,19] . This persistence is a threat to disease elimination strategies[19]. Severe *P. malariae* infections are commonly treated with an Artemisinin Combination Therapy (ACT), similar to *P. falciparum* infections in the same region[18]. The high prevalence of mixed infections with *P. falciparum* and *P. vivax* means that populations of *P. malariae* may have been experiencing substantial drug pressure. Several reports have described *P. malariae* parasites that have not been cleared after treatment with standard antimalarials[8,17,20] or have initiated successful infections despite effective chemoprophylaxis[21], leading to fears of reduced drug efficacy.

Advances in whole genome sequencing (WGS) technologies now allow for large scale genome diversity studies. Such studies in *P. falciparum* and *P. vivax* have provided significant new understanding of the structure of parasite populations, intra- and inter-population genomic diversity, and identified genomic regions under selective pressure, such as drug resistance associated genes[22–25]. However, to date only a few complete genomes have been assembled for *P. malariae* (n = 5; genome size 31.9 Mb)[26,27], which have led to insights into genome structure including species-specific gene expansions, causing the characteristically large genome of *P. malariae*[27]. One expansion of note is a family encoding transmembrane domain proteins, known as *Pm-fam,* containing *fam-m* and *fam-l* genes, which are hypothesised to be involved in host–pathogen interactions and are unique to *P. malariae* parasites[26,27]. To date, investigations of *P. malariae* genetic diversity have used microsatellite data and demonstrated considerable levels of genetic diversity and differences between and within populations from different countries[28,29]. However, microsatellite markers reflect only a minority of the genome (< 0.1%), and further investigation using WGS data is needed to explore genetic diversity and population structure across endemic regions.

A major challenge in performing WGS studies using clinical parasite isolates is the difficulty in obtaining sufficient *Plasmodium* DNA from infected individuals. This is due to low parasite densities and the presence of human DNA from host lymphocytes and other circulating nucleated cells. For *P. malariae,* genome studies are further complicated by the lack of an in vitro culture method for this parasite species. Until now, WGS data for *Plasmodium* parasites has been obtained using DNA extracted from venous blood of clinical cases that were pre-filtered to remove human leukocytes, in order to reduce the amount of co-extracted human DNA[30]. This methodology is efficient when parasite densities are high, however, this is not the case for the majority of *P. malariae* infections, particularly asymptomatic individuals, where this approach would not yield sufficient parasite DNA for WGS. Recently, a selective whole genome amplification (SWGA) strategy has been used to successfully sequence *P. falciparum*, *P. vivax* and *P. knowlesi* genomes from non-filtered blood and from dried blood spots of clinical samples[31–33]. The SWGA method uses oligonucleotide primers that preferentially bind with high frequency to the pathogen DNA, and rarely bind to the host genome[34]. The high fidelity Phi29 polymerase, which works through multiple displacement amplification (MDA), is used to amplify large segments (~70 kb) of DNA, primed by the SWGA oligonucleotides.

The unique but poorly understood characteristics of the *P. malariae* parasite, and the threat of unpredictable drug resistance, indicate a need for better understanding of the biological features of this neglected species. Knowledge of the complexity and variability of the *P. malariae* genome, and comparative studies with the well characterised *P. falciparum* and *P. vivax* genomes[23–25], could provide insights into the biology of this human parasite species. Here, we adapt and validate the SWGA approach for amplification of the *P. malariae* genome, successfully processing and sequencing 19 clinical samples. After selecting 18 high quality samples, we demonstrate that the resulting WGS data can be used to assess genetic diversity in *P. malariae* genes orthologous to known drug resistance markers in other species, and to inform population structure. In doing so, we provide proof-of-principle for large-scale WGS studies using blood samples collected from malaria endemic regions to inform malaria control efforts, and provide new molecular information for development of diagnostics, vaccines and drugs.

Results

SWGA enriches *P. malariae* DNA and increases WGS data coverage. We performed SWGA using a designed primer set (denoted as Pmset1) consisting of five primers (see S1 Table) that preferentially bind the *P. malariae* genome (average binding sites located once every 2.9 kb within the *P. malariae* genome, compared to once every 45.1 kb in the human genome). For successful selective amplification it is essential that the binding sites are in close proximity in the parasite genome and spaced further apart in the human genome[35]. Using two test samples (PM_THA_001 and PM_THA_002), we demonstrate that Pmset1 successfully amplifies the *P. malariae* genome, allowing for higher quality WGS data in comparison to non-amplification (S1 Fig.). Whilst all four samples were sequenced at a similar depth, we observed that amplified samples have a significant increase in coverage, with a mean 18.6-fold increase in the percentage of the genome covered with ≥ 5 reads when compared to non-amplification (S2 Table). The increase in genome coverage seen with SWGA allows for greater detection of SNPs which can be used for downstream population genetics analysis. As a result, there was an 800- to 13,000-fold increase in the number of callable SNPs detected in samples amplified using Pmset1 (S2 Table).

After validation of Pmset1, 17 additional clinical samples were amplified using Pmset1 and underwent WGS. One sample (PM_THA_009), with a low parasitaemia of 0.0016% presented with low coverage after the first sequencing run (27% genome covered ≥ 5 reads), this sample was re-sequenced, and the second run had

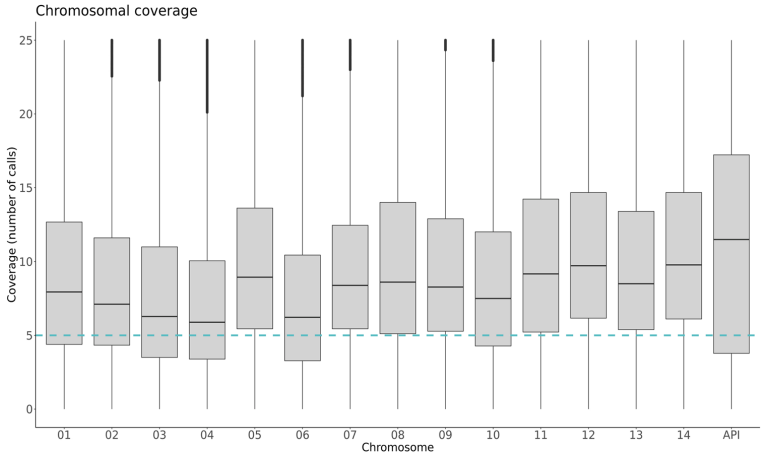


Figure 1. Sequencing coverage by chromosome after SWGA. The distribution of coverage for each position within the genome between 18 high quality samples, among the 14 nuclear chromosomes and the organellar apicoplast (the mitochondrial coverage plot is excluded due to high coverage). The blue horizontal line represents the recommended coverage cut-off point for SNP calling (≥ 5 reads).

better results (44% genome covered ≥ 5 reads) (S2 Table). The two sequencing runs were combined to generate PM_THA_009com (52% genome covered ≥ 5 reads). Excluding the separate runs for PM_THA_009, and one sample with low genome coverage (PM_LBR_003), the remaining samples had an average of 67.4% (± 15%) of the genome covered by ≥ 5 reads (S2 Table). The coverage profile after amplification was uneven, as reported for other Plasmodia[32], but generally, across all chromosomes, reaching coverage above the recommended cut off point for SNP calling (five reads or above) (Fig. 1). Coverage of the mitochondria was variable but consistently high in comparison to other chromosomes (mean: 26-fold coverage). The average chromosomal coverage of the two unamplified samples was much lower, with only 0.82% of the genome with a coverage ≥ 5 reads (S2 Fig).

SWGA is dependent on the initial parasitaemia of a sample. To determine a potential limit of parasitaemia for WGS, a measure of genomic coverage was assessed in nine Thailand samples for which parasitaemia data was available (range of parasitaemia: 0.0004% to 0.2024%). We determine a parasitaemia limit using both microscopy estimates and cycle threshold (CT) values calculated using the qPCR method[36]. We plotted the CT values of each sample alongside the percentage of the genome that was covered by ≥ 5 reads. We determined that a CT value of 30 will lead to an estimate of 50% of the genome covered by ≥ 5 reads (Fig. 2a). Coverage results are unpredictable below this limit, however, as with PM_THA_001, sequence data may be usable below this limit. When using percentage parasitaemia, we verified that all sequence data from parasite densities higher than 0.01% (400 parasites/ul) led to > 50% of the genome covered by five or more reads; this is a lower limit than previously defined for *P. falciparum,* and *P. knowlesi*[32,37] (Fig. 2b). For difficult samples with lower parasitaemia it is possible to improve genome coverage by performing independent SWGA reactions and by increasing sequence data, as observed previously for *P. vivax*[31], and also demonstrated here for PM_THA_009, for which merging data lead to > 50% of genome covered with at least 5 reads (S2 Table).

Determining and excluding hypervariable regions. Many *Plasmodium* species are known to contain large regions of repetitive sequences within the subtelomeres, which is exaggerated in the case of *P. malariae,* leading to an enlarged genome in comparison to other species[27]. We defined the core genome by both excluding regions with > 2.25 SNPs on average per 5 kb window (S3 Fig.) or containing *Pm-fam* genes (Fig. 3, S4 Fig., core genome coordinates are listed in S3 Table), to leave a total core genome size of 23,960,057 bases (81% of the total PmUG01 reference genome).

Genetic diversity and population structure. We investigated the multiplicity of infection (MOI) in all samples using the core genome, initially through determining the proportion of SNPs that were heterozygous, alongside running *estMOI*[38] for each sample which calculates the percentage of the genome that supports a MOI of 1 (S2 Table, S5 Fig.). The samples were *P. malariae* mono-infections, that is, where no other Plasmodium species were detected by qPCR. However, it is possible that > 1 clone of *P malariae* is present in a sample i.e. polyclonal. Using this sample set, three isolates displayed evidence of polyclonal infections (PM_LBR_002, PM_UGA_007 and PM_THA_012). This observation was confirmed by assessing the minor allele frequency (MAF) distribution of these isolates, where they presented with a higher proportion of SNPs with a non-reference MAF

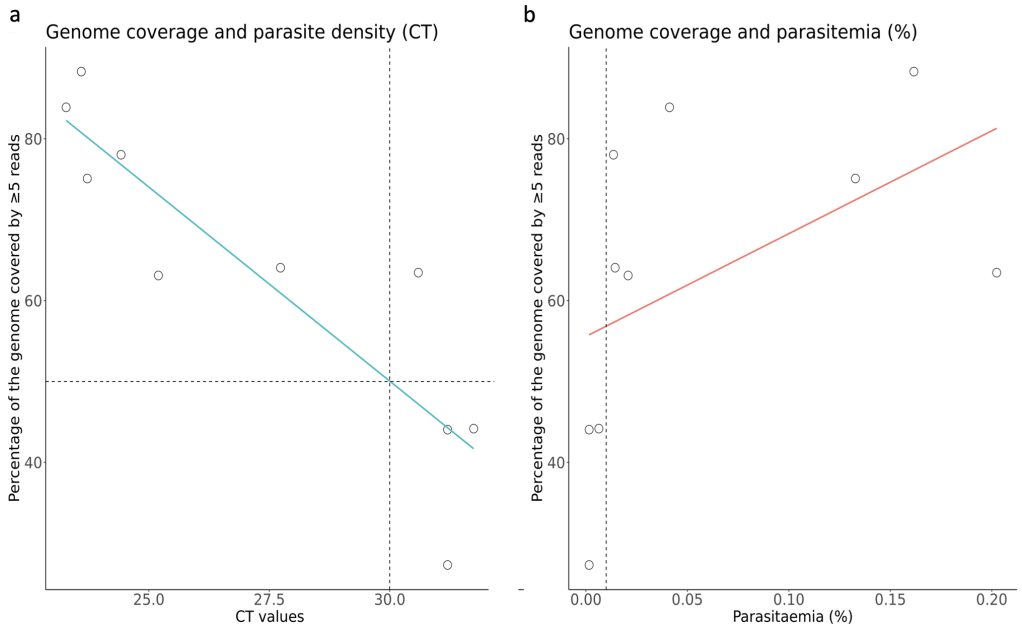


Figure 2. Correlation between parasite density and genome coverage. (**a**) Nine samples (amplified with SWGA approach) were used to assess the effect of parasite density (as measured by CT values obtained from qPCR) on the proportion of the genome covered by five or more reads. Each point demonstrates a single sample, with a lowess line of best fit. The dashed horizontal line represents a cut off of 50% of the genome covered by 5 or above, and the dashed vertical line indicates the suggested CT cut-off of 30. (**b**) The same plot is shown using parasitaemia as the measure of parasite density (percentage of RBCs parasitized). Parasitaemias range from 0.0004% to 0.2024%, and the vertical dashed line represents the suggested parasitaemia cut-off of 0.01%.

in the range 0.2 to 0.8 (S6 Fig.). For these three isolates only the major allele strain in each isolate was used in further population genetics analysis.

A total of 868,476 genome-wide SNPs were found within the 18 high quality samples (average of 48,249 SNPs per sample), of which 194,709 were unique. However, as with other *Plasmodium* spp., the subtelomeric region of the *P. malariae* genome contains large sections of repetitive DNA sequence[27]. These regions are problematic when interpreting WGS data from short-read technologies such as Illumina as short reads are likely to be aligned to incorrect regions along the reference genome, leading to deceptively high coverage and number of SNPs.

After removing hypervariable regions, we analysed the core genome (see S3 Table for coordinates) of 18 samples (≥ 40% of the genome covered by ≥ 5 reads) and identified 29,899 unique SNPs (mean: 5,810 ± 2,229 SNPs per sample) for downstream population genetic analysis. We found that geographically proximal samples displayed less pairwise diversity than geographically separated samples, with parasites from Thailand appearing more closely related to each other than to parasites obtained from Africa. Nucleotide diversities (π) > 3×10^{-4} nucleotide differences per site are only seen when comparing samples between Thailand and Africa, and $\pi < 2 \times 10^{-4}$ was only seen when comparing samples within Thailand or Africa (S4 Table).

A maximum-likelihood tree was constructed using core genome SNP data and demonstrates clear regional separation of *P. malariae* parasites, with samples from the African continent clustering together, and independently from samples originating in Thailand (Fig. 4).

Genetic variation in in orthologs of known *P. falciparum* genes associated with drug resistance. *P. malariae* parasites are commonly subject to antimalarial treatments, therefore we investigated the coverage and prevalence of mutations in orthologs of known *P. falciparum* genes associated with drug resistance (*Pfcrt, Pfdhfr, Pfdhps, Pfk13* and *Pfmdr1;* gene IDs are in S5 Table). SNPs were only found in *Pmdhfr* (n = 3; 2 non-synonymous), *Pmdhps* (n = 5; 1 non-synonymous) and *Pmmdr1* (n = 4, 2 non-synonymous) (Fig. 5, Table 1). SNPs within *Pmdhfr* at positions 1,292,026 and 1,292,193 in chromosome 5 appear to be more common globally than other SNPs, whereas SNPs within *Pmdhps* and *Pmmdr1* appear to be more prevalent in Thailand than Africa (Table 2). All of the non-synonymous mutations found within *Pmdhfr* led to amino acid alterations (F57L, R58S and N114S) at positions that align with known drug-resistance associated positions within the *Pfdhfr* ortholog (C59R and S108N respectively) upon amino acid allignment (Table 1, S7 Fig.)[43]. In addition,

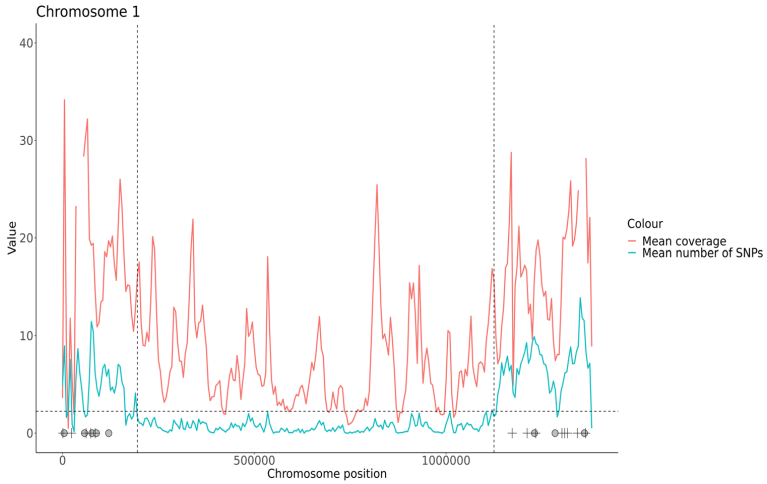


Figure 3. Defining and excluding subtelomeric regions, an example with chromosome 1. The average coverage (pink) and number of SNPs (blue) between all 18 samples for chromosome 1. The black dashed horizontal line demonstrates the previously chosen SNP limit per 5 kb window (as defined in S2 Fig.). Black dashed lines are placed at the suggested hypervariable region cut-off points, where clusters of windows demonstrating > 2.25 SNPs are seen. The midpoints of *Pm-fam* gene families are annotated; *Pm-fam-l* gene positions are denoted by a black plus, whilst *Pm-fam-m* gene positions are denoted with a grey circle. (S3 Fig. for all chromosomes, S3 Table for coordinates).

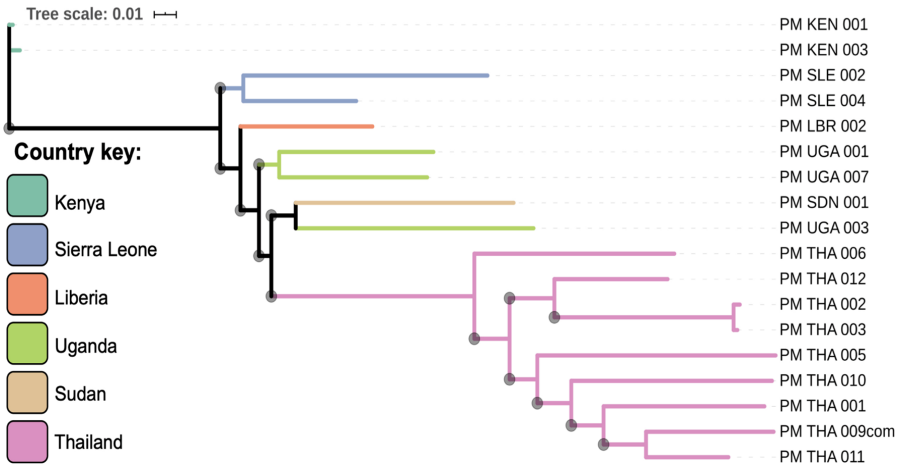


Figure 4. Population structure of *P. malariae* parasites. A maximum likelihood (ML) tree was generated using 29,899 unique SNPs from 18 amplified and sequenced samples (> 40% genome with at least fivefold coverage). The ML tree is unrooted and was generated using *Iqtree*[39] with *Modelfinder* software used to select the best model of substitution[40]. Horizontal branch lengths are drawn to scale demonstrating the number of substitutions per position, and branch bootstrap values (determined using *UFBoot2*[41]) above 50 are denoted with a grey circle at the start of each branch. The tree was visualised in *iTOL*[42], and branches were coloured by country (country codes: KEN = Kenya, SLE = Sierra Leone, LBR = Liberia, UGA = Uganda, THA = Thailand, SDN = Sudan).

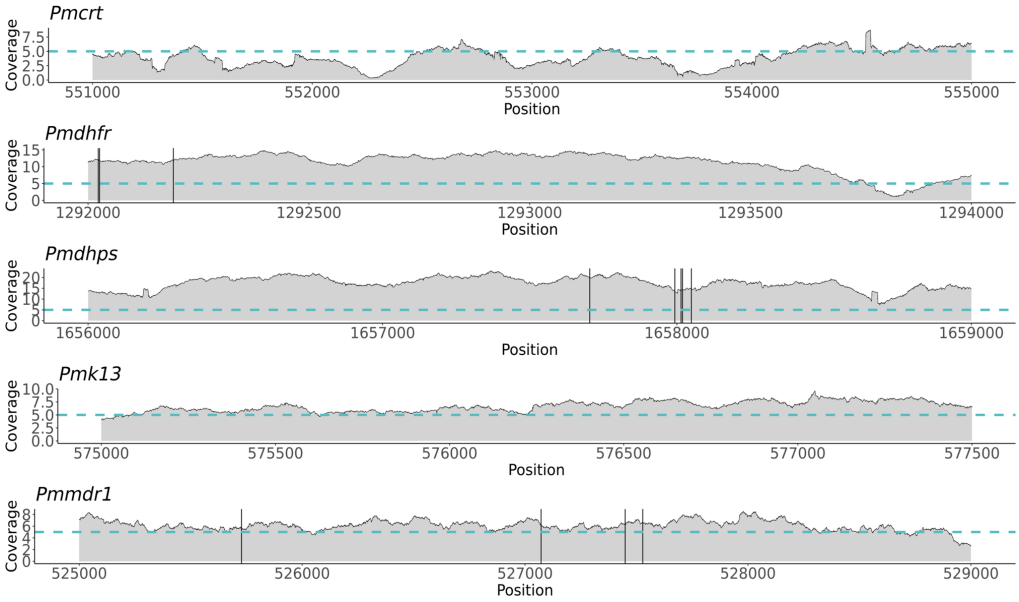


Figure 5. Average sequencing coverage and SNPs found within *P. malariae* orthologues of *P. falciparum* drug resistance associated genes. Average sequencing coverage for 18 samples across 5 genes is represented. The blue horizontal line indicates the coverage of 5 reads and black vertical lines are locations of SNPs (*Pmdhfr:* 1,292,023, 1,292,026 and 1,292,193; *Pmdhps:* 1,657,704, 1,657,993, 1,658,014, 1,658,019 and 1,658,049; *Pmmdr1:* 525,728, 527,072, 527,449 and 527,528).

Gene	Position	Ref	SNP1	Effect	Change in codon	Change in amino acid	Proportion of SNP1	SNP 2	Effect	Change in codon	Change in amino acid	Proportion of SNP2
Pmdhfr	**1,292,023***	**C**	**A**	**NS**	**ttC/ttA**	**F57L***	**0.17**	**G**	**NS**	**ttC/ttG**	**F57L***	**0.11**
	1,292,026*	A	G	S	agA/agG	R58	0.28	**C**	**NS**	**agA/agC**	**R58S***	**0.67**
	1,292,193*	**A**	**G**	**NS**	**aAc/aGc**	**N114S***	**0.78**	/	/	/	/	
Pmdhps	**1,657,704**	**C**	**T**	**NS**	**Gtg/Atg**	**V121M**	**0.06**	/	/	/	/	
	1,657,993	C	T	I	/	/	0.28	/	/	/	/	
	1,658,014	A	G	I	/	/	0.11	/	/	/	/	
	1,658,019	A	T	I	/	/	0.22	C	I	/	/	0.06
	1,658,049	T	A	I	/	/	0.06	/	/	/	/	
Pmmdr1	**525,728***	**T**	**G**	**NS**	**ttA/ttC**	**L1063F***	**0.11**	/	/	/	/	
	527,072	C	T	S	ttG/ttA	L615	0.22	/	/	/	/	
	527,449	**G**	**T**	**NS**	**Ctt/Att**	**L490I**	**0.06**	/	/	/	/	
	527,528	G	A	S	agC/agT	S463	0.33	/	/	/	/	

Table 1. Single nucleotide polymorphisms (SNPs) detected in *Pmdhfr*, *Pmdhps* and *Pmmdr1* genes, and their downstream effects. *Amino acid alterations that lie in close proximity to known *P. falciparum* resistance mutations. SNPs leading to non-synonymous (NS) mutations are in bold, whilst intronic (I) or synonymous (S) mutations are unbolded.

the mutation at position 527,528 within *Pmmdr1* (chromosome 10), which leads to the amino acid substitution L1063F, aligns in close proximity to N1042D in the *Pfmdr1* ortholog that is associated with quinine resistance, and increased mefloquine and artemisinin susceptibility (Table 1, S7 Fig.)[44].

Mixed infections. *P. malariae* parasites are commonly found in mixed infections with other *Plasmodium* spp.[9,11,12]. This provides a further obstacle for WGS, as not only is the human genome a potential contaminant, but also the other *Plasmodium* species present. We used four further unprocessed clinical blood samples from Thailand which were found to be mixed infections after qPCR[36] and underwent SWGA to determine whether

Gene	*Pmdhfr*			*Pmdhps*					*Pmmdr1*			
Position	**1292023***	**1292026***	**1292193***	**1657704**	**1657993**	**1658014**	**1658019**	**1658049**	**525728**	**527072**	**527449**	**527528***
Reference allele	**C**	**A**	**A**	**C**	**C**	**A**	**A**	**T**	**T**	**C**	**G**	**G**
PM_THA_001	/	C	G	/	/	/	/	/	G	–	–	–
PM_THA_002	/	C	G	/	T	/	/	/	/	/	/	/
PM_THA_003	/	C	G	/	T	A/G	A/T	/	N	N	N	N
PM_THA_005	/	C	G	/	T	/	A/T	/	G	/	/	/
PM_THA_006	/	C	G	/	T	A/G	A/T	/	/	N	/	N
PM_THA_009com	/	C	G	/	/	/	A/C	A/T	/	T	/	A
PM_THA_010	/	C	/	/	T	/	A/T	/	/	T	T	A
PM_THA_011	/	C	G	/	/	/	/	/	/	N	/	A
PM_THA_012	/	C	/	/	/	/	/	/	/	T	/	A
PM_KEN_001	A	G	G	/	/	/	/	/	N	/	N	N
PM_KEN_003	A	G	G	/	/	/	/	/	N	N	N	–
PM_LBR_002	/	C	G	C/T	/	/	/	/	/	/	/	A/G
PM_SDN_001	G	G	G	/	/	/	/	/	N	–	N	N
PM_SLE_002	/	C	G	/	/	/	/	/	/	T	/	/
PM_SLE_004	/	C	/	/	/	N	N	N	N	N	N	N
PM_UGA_001	N	N	/	/	N	N	N	N	–	–	N	N
PM_UGA_003	A	G	G	/	/	/	/	/	/	/	/	A
PM_UGA_007	G	G	G	/	/	/	/	/	N	N	N	N

Table 2. Distribution of SNPs in *Pmdhfr*, *Pmdhps* and *Pmmdr1* among 18 samples from Africa and Thailand. / denote Wild-type alleles (the allele observed in the PmUG01 reference genome). N denotes no coverage at this position. *Positions lead to amino acid substitutions that align with drug resistance-associated substitutions observed in *P. falciparum*.

Pmset1 was specific to only the *P. malariae* genome. Each sample contained varying mixtures of other parasite species present and our results suggest that SWGA is likely to work if *P. malariae* is initially the most prevalent parasite in the mixed infection (i.e. has the lowest CT value) (S6 Table). However, when DNA from other species is present at high concentrations, SWGA may not be effective for amplification of *P. malariae* (S6 Table).

Discussion

P. malariae is a neglected malaria parasite with unique features, such as a longer quartan cycle and the ability to persist in the human host for years or decades[13]. Genetic investigation of this parasite may allow us to understand how *P. malariae* is able to cause chronic infections, why there are accounts of *P. malariae* parasites persisting after treatment with ACT, and why some *P. malariae* infections lead to severe outcomes whilst others remain asymptomatic. Malaria parasite genomics can provide important biological insights to understand this disease, but the difficulty of obtaining sufficient parasite DNA for WGS has been a challenge for genomic studies of *P. malariae*. Here we present the first application of SWGA for this species. We have customized the SWGA approach to successfully amplify *P. malariae* DNA extracted directly from unprocessed blood from clinical samples which were obtained from six different countries. In agreement with others[31,32], we have demonstrated that the parasitaemia affects the efficiency of SWGA, and recommend using samples with a percentage parasitaemia > 0.01%, which is a lower threshold than reported for other species[31,32,37]. The WGS data generated from SWGA-treated samples is of high quality with good overall coverage, leading to an average of 67.4% (± 15%) of the genome covered by ≥ 5 reads between the 18 samples assessed in this study. Using these samples, we were able to identify 868,476 total SNPs (average 48,249 SNPs per sample), filtered to 104,583 total SNPs after exclusion of hypervariable regions (average of 5,810 SNPs per sample). This is lower than SNP prevalence documented in *P. knowlesi* (115,995 SNPs per sample including hypervariable regions)[37], yet higher than SNPs found in *P. vivax* (14,463 SNPs per sample before filtering for core genome) after SWGA[31].

It is important to note that differences in the number of SNPs per sample reported could also be due to differences in the method used for variant calling.

A maximum likelihood tree based on SNP data revealed geographic clusters, with clear separation of African and Asian samples. This geographical clustering is consistent with data for *P. falciparum*[45] and *P. vivax* parasites[24,45,46]. Similar geographic clustering was observed in the phylogenetic analysis of SNPs in the circumsporozoite gene from *P. malariae* isolates from Africa and Asia[47]. To improve geographical clustering resolution (i.e. by country), the number of samples investigated needs to be increased. Our data suggests that parasites display isolation by distance, therefore country or multi-country regional analysis of *P. malariae* populations could be used in future studies to identify regions under selection in different populations.

We further demonstrate that SWGA successfully amplifies genes orthologous to those associated with drug resistance in *P. falciparum*, and identify SNPs in *Pmdhfr*, *Pmdhps* and *Pmmdr1*. The effects of these SNPs are unknown, and to date, there are no characterised molecular markers of drug resistance in *P. malariae* parasites, even though treatment failures have been reported[19,48]. Despite this, potential mutations of interest were

found, particularly at positions 1,292,023, 1,292,026 and 1,292,193 in chromosome 5 in the *Pmdhfr* gene. These mutations lead to amino acid substitutions F57L, R58S and N114S respectively, and align almost perfectly with *P. falciparum* amino acid substitutions C59R and S108N which are associated with reduced susceptibility to sulfadoxine/pyrimethamine[49]. The nonsynonymous mutation N114S has been previously reported in two *P. malariae* samples from Thailand and the F57L and R58L mutations have been reported in *P. vivax* samples from several geographical regions[50,51]. In addition, one mutation within *Pmmdr1* at position 525,728 in chromosome 10 leads to amino acid substitution L1063F, which aligns with close proximity to N1042 in the *Pfmdr1* ortholog, associated with reduced susceptibility to quinine and increased susceptibility to mefloquine, halofantrine and artemisinin[44]. It is important to note that whilst treatment failures are seen with *P. malariae* infections, it is not clear whether this is due to mutations within the parasite genome leading to reduced drug efficacy, or perhaps a specific phenotype of this species due to the longer parasite life cycle which may reduce drug absorption[48]; therefore further functional studies are required to determine the effect, if any, of these substitutions.

The subtelomeres, containing the *fam-l* and *fam-m* gene families are of great interest when studying *P. malariae*, as they are unique to this species and are thought to be involved in host-parasite interactions[26]. Unfortunately, sequence analysis of these regions is notoriously difficult using short-read technologies, therefore longer-read sequencing will be needed to further investigate these regions.

In conclusion, the SWGA approach offers a fast, cost effective way to explore the genome diversity of *P. malariae* from unprocessed blood of infected individuals. Further studies should consider the analysis of a larger number of samples from a greater geographical range and different clinical outcomes, in addition to studies investigating the subtelomeric regions with long read technologies. Such studies are necessary to characterize the epidemiology and genetic diversity of *P. malariae* populations, with the potential to provide biological insights for disease control.

Methods

Ethics statement. Isolated from Thailand were collected with ethical approval from the Mahidol Faculty of Tropical Medicine Ethics Committee (Ref: 2015-001.01); PHE-MRL samples are analysed under NHS Ethics approval (#18/LO/0738). In both instances, samples were collected according to relevant guidelines and regulations in both Thailand and the UK, and informed consent was obtained for all subjects over the age of 18 (for subjects under 18 years old, consent was obtained from the appropriate legal guardian.

Sample collection and processing. This project used nine *P. malariae* DNA samples extracted from unprocessed venous blood from infected individuals in Thailand. Parasite density (parasites/µl) determined by microscopy was available for these isolates. Genomic DNA was extracted from frozen unprocessed blood using the QIAamp DNA Blood Mini Kit (Qiagen) or the QIAsymphony DSP DNA Kit in combination with a QIAsymphony SP instrument (Qiagen), according to manufacturer's instructions. As microscopy is prone to human errors, all extracted DNA samples were subject to qPCR as outlined by Shokoples et al.[36] to ensure that only *P. malariae* single species infections were used.

A further ten DNA samples were provided by the Public Health England-Malaria Reference Laboratory (PHE-MRL) at the London School of Hygiene and Tropical Medicine (LSHTM). These samples were sourced from individuals who had reported recent travel to only one country with malaria transmission, including: Kenya (n = 2), Liberia (n = 2), Sierra Leone (n = 2), Sudan (n = 1) and Uganda (n = 3) between 2010 and 2017. PHE-MRL samples are commonly sourced from individuals returning to visit relatives in their original native country. For species identification, PHE-MRL samples perform both a nested PCR[52] and qPCR[36] and are archived according to the species present.

Total DNA concentration for all samples was quantified using a Qubit v2.0 fluorometer (Thermo Fisher Scientific).

Selective whole genome amplification. The *swga* program (www.github.com/eclarke/swga) was used to identify primers that preferentially amplify the *P. malariae* genome[35], using its reference genome (PmUG01, https://plasmodb.org) as the target (foreground), and the human genome (GRCh37; https://grch37.ensembl.org/) as the background. The *swga* program ranks primers dependant on the ratio of foreground genome binding to the background genome binding, combined with the evenness of primer binding along the target genome and generates multiple potential primer sets. The five highest-ranked sets consist of combinations of 4 to 6 oligonucleotides each, with overlapping primers. The set that ranked highest (Pmset1) consisted of five primers: TATGTATA*T*T, TTATTC*G*T, TTCGTT*A*T, TTTTTA*C*G, TATTTC*G*T, that were ordered with a phosphorothioate bond (represented by *) modifications to prevent primer degradation by the exonuclease activity of the Phi29 polymerase. To evaluate the efficacy of Pmset1 for SWGA of the *P. malariae* genome, we tested two samples (PM_THA_001 and PM_THA_002) and sequenced both before and after SWGA.

DNA samples were subject to SWGA following previously published protocols[31,32,37]. All SWGA reactions were carried out in a UV Cabinet for PCR Operations (UV-B-AR, Grant-Bio) to eliminate potential contamination. Briefly, a maximum of 60 ng of gDNA (minimum of 5 ng) was added to a total 50 µl reaction alongside 5 µl of 10 × Phi29 DNA Polymerase Reaction Buffer (New England BioLabs), 0.5 µl of Purified 100 × BSA (New England BioLabs), 0.5 µl of 250 µM Primer mix, 5 µl 10 mM dNTP (Roche), 30 units Phi29 DNA Polymerase (New England BioLabs) and Nuclease-Free Water (Ambion, The RNA Company) to reach a final reaction volume of 50 µl. The reaction was carried out on a thermocycler with the following step-down program: 5 min at 35 °C, 10 min at 34 °C, 15 min at 33 °C, 20 min at 32 °C, 25 min 31 °C, 16 h at 30 °C and 10 min at 65 °C. After successful validation of Pmset1, the remaining samples underwent SWGA as described above. After SWGA, samples were purified using a 1:1 ratio of AMPure XP beads (Beckman-Coulter), following manufacturer's instructions.

Library preparation and WGS. SWGA samples and the unamplified negative controls were sequenced on either an Illumina MiSeq or HiSeq4000 platform. For the MiSeq runs, the QIAseq FX DNA Library Kit (QIAGEN) was used for library preparation according to the manufacturer's protocol, with a 20-min fragmentation step. For the HiSeq4000 runs, samples were prepared using the NEB Next Ultra DNA Library Prep Kit for Illumina (from New England BioLabs Inc., E7370). Library DNA concentration was analysed using a Qubit 2.0 fluorometer. All sequencing reactions were performed using paired (2×) 150 bp reads.

Sequence data analysis. Raw fastq files were trimmed using trimmomatic set to default parameters[53], and aligned to the *P. malariae* UG01 reference genome (PlasmoDB) using *bwa-mem* software[54]. SNPs were identified using the *samtools* software suite (samtools.sourceforge.net)[55] and filtered for quality based on previously described methods[56]. The coverage of each nucleotide position was analysed using *sambamba*[57], which was set to include only SNPs with coverage levels of at least fivefold. Poor quality samples were removed (< 40% of the genome covered by 5 reads) to leave 18 high quality samples. We used estMOI[38] to determine MOI for samples, and the major allele was used when heterozygous SNP calls were found.

Determining and excluding subtelomeric regions. To exclude hypervariable subtelomeric regions the *P. malariae* genome was split into 5 kb segments and the average number of SNPs was calculated. We defined an upper limit for the number of SNPs within each window in order to identify highly polymorphic windows. This SNP limit was used in conjunction with the positions of the *Pm-fam* gene families to define the subtelomeric regions of each chromosome and exclude these from downstream analysis.

Population genetics. To investigate the population structure of *P. malariae* parasites, a distance matrix was created which was based on a matrix of pairwise identity calculated from the SNPs present in each sample. Using the distance matrix, a maximum likelihood tree was produced using *Iqtree*[39] with *Modelfinder*[40] to select the best model of substitution and ultrafast bootstrap analysis[41]. The resulting Newick tree was visualised in iTOL[42]. The nucleotide diversity (π) metric was used to investigate the genetic variability between samples, and was calculated using the *pegas* (v0.10) package[58], which defines nucleotide diversity as the average number of SNPs per position between two sequences.

Drug resistance orthologs. Orthologs of known genes involved in drug resistance in *P. falciparum* were analysed. The SNPs were described using the *snpEff* software[59] which annotates the genes affected, the type of mutation, and if non-synonymous, the amino-acid change that has occurred. The coverage of genes of interest was also analysed using the output file from applying *sambamba* software[57]. The genes investigated and their respective IDs are summarised in S5 Table.

Data availability

All raw sequence data is listed in the European Nucleotide Archive (study accession number PRJEB33837).

Received: 5 March 2020; Accepted: 5 June 2020
Published online: 02 July 2020

References

1. WHO. *World malaria report 2018* (World Health Organization, Geneva, 2018).
2. Scopel, K. K. G., Fontes, C. J. F., Nunes, ÁC., Horta, M. F. & Braga, ÉM. High prevalence of Plamodium malariae infections in a Brazilian Amazon endemic area (Apiacás—Mato Grosso State) as detected by polymerase chain reaction. *Acta Trop.* **90**, 61–64 (2004).
3. Fançony, C. *et al.* Various pfcrt and pfmdr1 genotypes of plasmodium falciparum cocirculate with *P. malariae*, *P. ovale* spp., and *P. vivax* in Northern Angola. *Antimicrob. Agents Chemother.* **56**, 5271–5277 (2012).
4. Mayxay, M., Pukrittayakamee, S., Newton, P. N. & White, N. J. Mixed-species malaria infections in humans. *Trends Parasitol.* **20**, 233–240 (2004).
5. Diversity, G. *et al.* Plasmodium malariae Prevalence. **23**, (2017).
6. Mathison, B. A. & Pritt, B. S. Update on Malaria Diagnostics and Test Utilization. *J. Clin. Microbiol.* **55**, 2009–2017 (2017).
7. Daniels, R. F. *et al.* Evidence of non-Plasmodium falciparum malaria infection in Kédougou, Sénégal. *Malar. J.* **16**, 9 (2017).
8. Betson, M. *et al.* Detection of persistent *Plasmodium* spp. infections in Ugandan children after artemether-lumefantrine treatment. *Parasitology* **141**, 1880–1890 (2014).
9. Niño, C. H. *et al. Plasmodium malariae* in the Colombian Amazon region: you don't diagnose what you don't suspect. *Malar. J.* **15**, 576 (2016).
10. Doritchamou, J. Y. A. *et al.* Submicroscopic placental infection by non-falciparum *Plasmodium* spp. *PLoS Negl. Trop. Dis.* **12**, e0006279 (2018).
11. Camargo-Ayala, P. A. *et al.* High *Plasmodium malariae* prevalence in an endemic area of the Colombian Amazon region. *PLoS ONE* **11**, e0159968 (2016).
12. Zhou, M. *et al.* High prevalence of *Plasmodium malariae* and *Plasmodium ovale* in malaria patients along the Thai–Myanmar border, as revealed by acridine orange staining and PCR-based diagnoses. *Trop. Med. Int. Health* **3**, 304–312 (1998).
13. Collins, W. E. & Jeffery, G. M. Plasmodium malariae: parasite and disease. *Clin Miscrobiol Rev* **20**, 579–592 (2007).
14. Hendrickse, R. G. & Adeniyi, A. Quartan malarial nephrotic syndrome in children. *Kidney Int.* **16**, 64–74 (1979).
15. Badiane, A. S. *et al.* Acute kidney injury associated with *Plasmodium malariae* infection. *Malar. J.* **13**, 1–5 (2014).
16. Langford, S. *et al. Plasmodium malariae* infection associated with a high burden of anemia: a hospital-based surveillance study. *PLoS Negl. Trop. Dis.* **9**, e0004195 (2015).
17. Betson, M., Clifford, S., Stanton, M., Kabatereine, N. B. & Stothard, J. R. Emergence of nonfalciparum plasmodium infection despite regular artemisinin combination therapy in an 18-month longitudinal study of Ugandan children and their mothers. *J. Infect. Dis.* **217**, 1099–1109 (2018).
18. Sutherland, C. J. Persistent parasitism: the adaptive biology of malariae and ovale malaria. *Trends Parasitol.* **32**, 808–819 (2016).

19. Dinko, B., Oguike, M. C., Larbi, J. A., Bousema, T. & Sutherland, C. J. Persistent detection of *Plasmodium falciparum*, *P. malariae*, *P. ovale curtisi* and *P. ovale wallikeri* after ACT treatment of asymptomatic Ghanaian school-children. *Int. J. Parasitol. Drugs Drug Resist.* **3**, 45–50 (2013).
20. Maguire, J. D. *et al.* Chloroquine-resistant Plasmodium malariae in south Sumatra, Indonesia. *Lancet* **360**, 58–60 (2002).
21. Teo, B.H.-Y. *et al.* Delayed onset of symptoms and atovaquone-proguanil chemoprophylaxis breakthrough by *Plasmodium malariae* in the absence of mutation at Codon 268 of pmcytb. *PLoS Negl. Trop. Dis.* **9**, e0004068 (2015).
22. Miotto, O. *et al.* Genetic architecture of artemisinin-resistant *Plasmodium falciparum*. *Nat. Genet.* **47**, 226–234 (2015).
23. Manske, M. *et al.* Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature* **487**, 375–379 (2012).
24. Pearson, R. D. *et al.* Genomic analysis of local variation and recent evolution in *Plasmodium vivax*. *Nat. Genet.* **48**, 959–964 (2016).
25. 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).
26. Rutledge, G. G. *et al. Plasmodium malariae* and *P. ovale* genomes provide insights into malaria parasite evolution. *Nature* **542**, 101–104 (2017).
27. Ansari, H. R. *et al.* Genome-scale comparison of expanded gene families in *Plasmodium ovale wallikeri* and *Plasmodium ovale curtisi* with *Plasmodium malariae* and with other *Plasmodium* species. *Int. J. Parasitol.* **46**, 685–696 (2016).
28. Bruce, M. C., Macheso, A., Galinski, M. R. & Barnwell, J. W. Characterization and application of multiple genetic markers for *Plasmodium malariae*. *Parasitology* **134**, 637–650 (2006).
29. Bruce, M. C., Macheso, A., McConnachie, A. & Molyneux, M. E. Comparative population structure of *Plasmodium malariae* and *Plasmodium falciparum* under different transmission settings in Malawi. *Malar. J.* **10**, 38 (2011).
30. Auburn, S. *et al.* An effective method to purify plasmodium falciparum DNA directly from clinical blood samples for whole genome high-throughput sequencing. *PLoS ONE* **6**, 4–11 (2011).
31. 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**, e02257-e2316 (2017).
32. Oyola, S. O. *et al.* Whole genome sequencing of *Plasmodium falciparum* from dried blood spots using selective whole genome amplification. *Malar. J.* **15**, 597 (2016).
33. Sundararaman, S. A. *et al.* Genomes of cryptic chimpanzee Plasmodium species reveal key evolutionary events leading. *Nat. Commun.* **7**, 1–14 (2016).
34. Leichty, A. R. & Brisson, D. Selective whole genome amplification for resequencing target microbial species from complex natural samples. *Genetics* **198**, 473–481 (2014).
35. Clarke, E. L. *et al.* SWGA: a primer design toolkit for selective whole genome amplification. *Bioinformatics* **33**, 2071–2077 (2017).
36. Shokoples, S. E., Ndao, M., Kowalewska-Grochowska, K. & Yanow, S. K. Multiplexed real-time PCR assay for discrimination of plasmodium species with improved sensitivity for mixed infections. *J. Clin. Microbiol.* **47**, 975–980 (2009).
37. Benavente, E. D. *et al.* Whole genome sequencing of amplified *Plasmodium knowlesi* DNA from unprocessed blood reveals genetic exchange events between Malaysian Peninsular and Borneo subpopulations. *Sci. Rep.* **9**, 9873 (2019).
38. Assefa, S. A. *et al.* estMOI: estimating multiplicity of infection using parasite deep sequencing data. *Bioinformatics* **30**, 1292–1294 (2014).
39. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
40. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
41. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
42. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).
43. Cowman, A. F., Morry, M. J., Biggs, B. A., Cross, G. A. M. & Foote, S. J. Amino acid changes linked to pyrimethamine resistance in the dihydrofolate reductase-thymidylate synthase gene of Plasmodium falciparum. *Proc. Natl. Acad. Sci. USA* **85**, 9109–9113 (1988).
44. 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).
45. Volkman, S. K. *et al.* A genome-wide map of diversity in *Plasmodium falciparum*. *Nat. Genet.* **39**, 113–119 (2007).
46. Neafsey, D. E. *et al.* The malaria parasite *Plasmodium vivax* exhibits greater genetic diversity than *Plasmodium falciparum*. *Nat. Genet.* **44**, 1046–1050 (2013).
47. Saralamba, N. *et al.* Genetic polymorphisms in the circumsporozoite protein of *Plasmodium malariae* show a geographical bias. *Malar. J.* **17**, 269 (2018).
48. Rutledge, G. G. *et al.* Genomic characterization of recrudescent *Plasmodium malariae* after treatment with Artemether/Lumefantrine. *Emerg. Infect. Dis.* **23**, 1300–1307 (2017).
49. McCollum, A. M. *et al.* Antifolate resistance in *Plasmodium falciparum*: multiple origins and identification of novel DHFR alleles. *J. Infect. Dis.* **194**, 189–197 (2006).
50. Auliff, A. *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).
51. Tanomsing, N. *et al.* Genetic analysis of the dihydrofolate reductase-thymidylate synthase gene from geographically diverse isolates of *Plasmodium malariae*. *Antimicrob. Agents Chemother.* **51**, 3523–3530 (2007).
52. Singh, B. *et al.* A genus- and species-specific nested polymerase chain reaction malaria detection assay for epidemiologic studies. *Am. J. Trop. Med. Hyg.* **60**, 687–692 (1999).
53. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
54. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
55. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
56. Campino, S. *et al.* Genomic variation in two gametocyte non-producing *Plasmodium falciparum* clonal lines. *Malar. J.* **15**, 229 (2016).
57. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
58. Paradis, E. Pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* **26**, 419–420 (2010).
59. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms. *SnpEff. Fly (Austin)* **6**, 80–92 (2012).

Acknowledgements

Sequence analysis was performed on the MRC UK funded eMedlab computing resource. AI and PJGG are supported by an MRC LID PhD studentship. TGC is funded by the Medical Research Council UK (Grant no. MR/

M01360X/1, MR/N010469/1, MR/R025576/1, and MR/R020973/1) and BBSRC (Grant no. BB/R013063/1). SC is funded by Medical Research Council UK grants (MR/M01360X/1, MR/R025576/1, and MR/R020973/1). The project was funded by a Research England Bloomsbury SET Project Grant (award reference CCF17-7779). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

S.C. and T.G.C. conceived and directed the project. D.N., S.P., H.P.F. and F.N. organised sample collection and processing. D.N. and J.M. undertook sample processing and DNA extraction for samples from the PHE-MRL. A.I. and S.C. undertook laboratory work including amplification and sequencing. A.I. performed bioinformatic analysis with guidance and training from E.D.B. and under the supervision of S.C. and T.G.C., and together they interpreted the results. Additional advice from M.H., P.J.G.G., C.S., H.P.F. and C.R. was sought during analysis. A.I. wrote the first draft of the manuscript with guidance from E.D.B., T.G.C. and S.C. All authors commented on versions of the manuscript and approved the final manuscript. A.I., E.D.B., S.C. and T.G.C. compiled the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-67568-4.

Correspondence and requests for materials should be addressed to S.C.

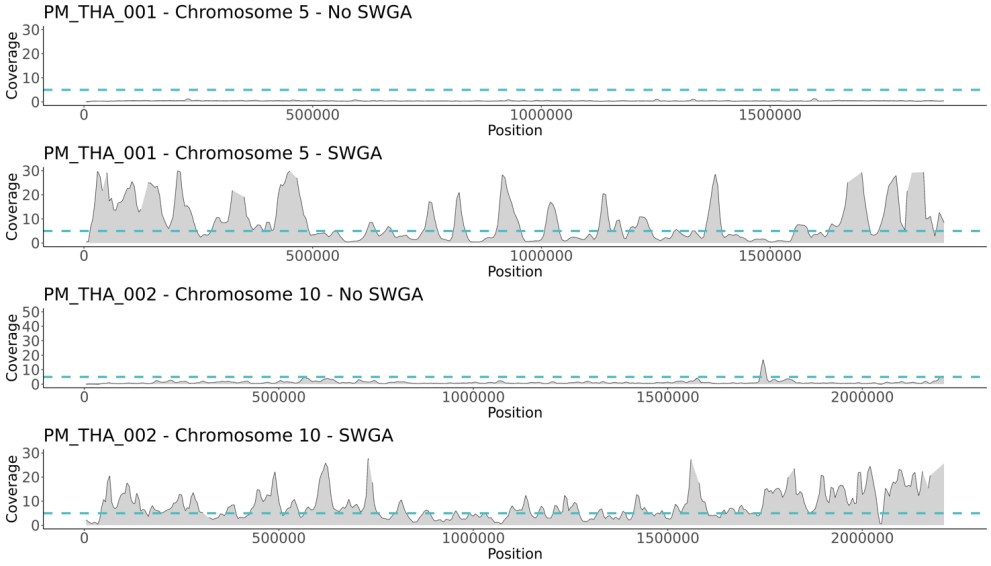
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

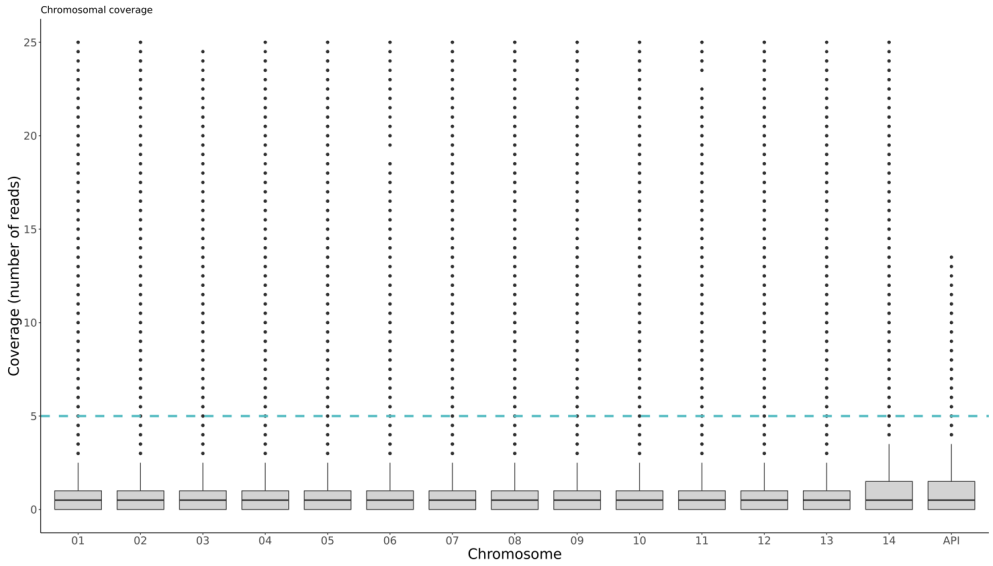
Selective Whole Genome Amplification of *Plasmodium malariae* DNA from clinical samples reveals insights into population structure

Amy Ibrahim[1], Ernest Diez Benavente[1], Debbie Nolder[1,2], Stephane Proux[3], Matthew Higgins[1], Julian Muwanguzi[1,2], Paula Josefina Gomez Gonzalez[1], Hans-Peter Fuehrer[4], Cally Roper[1], Francois Nosten[3,5], Colin Sutherland[1], Taane G. Clark[1,6], Susana Campino[1]

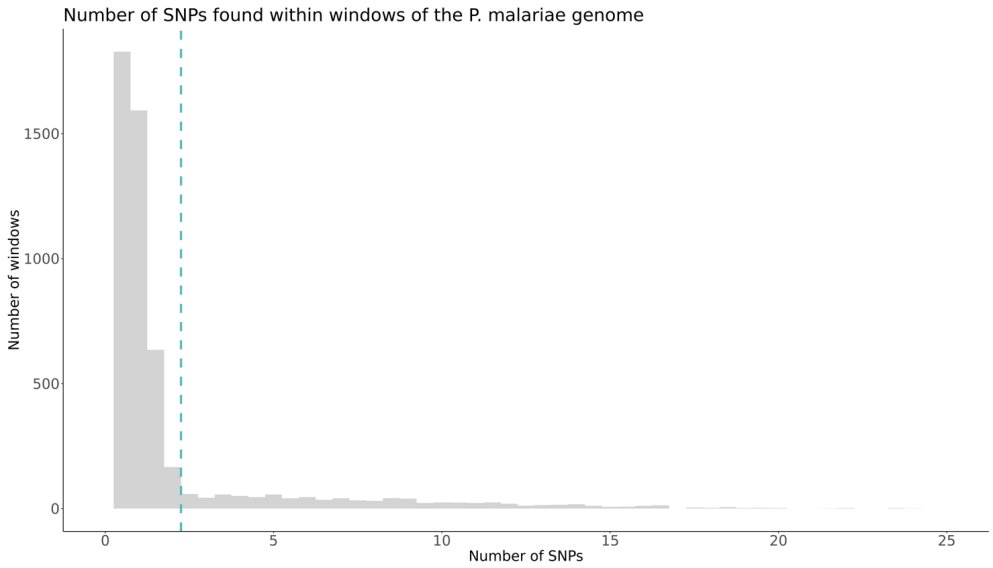
SUPPLEMENTARY INFORMATION



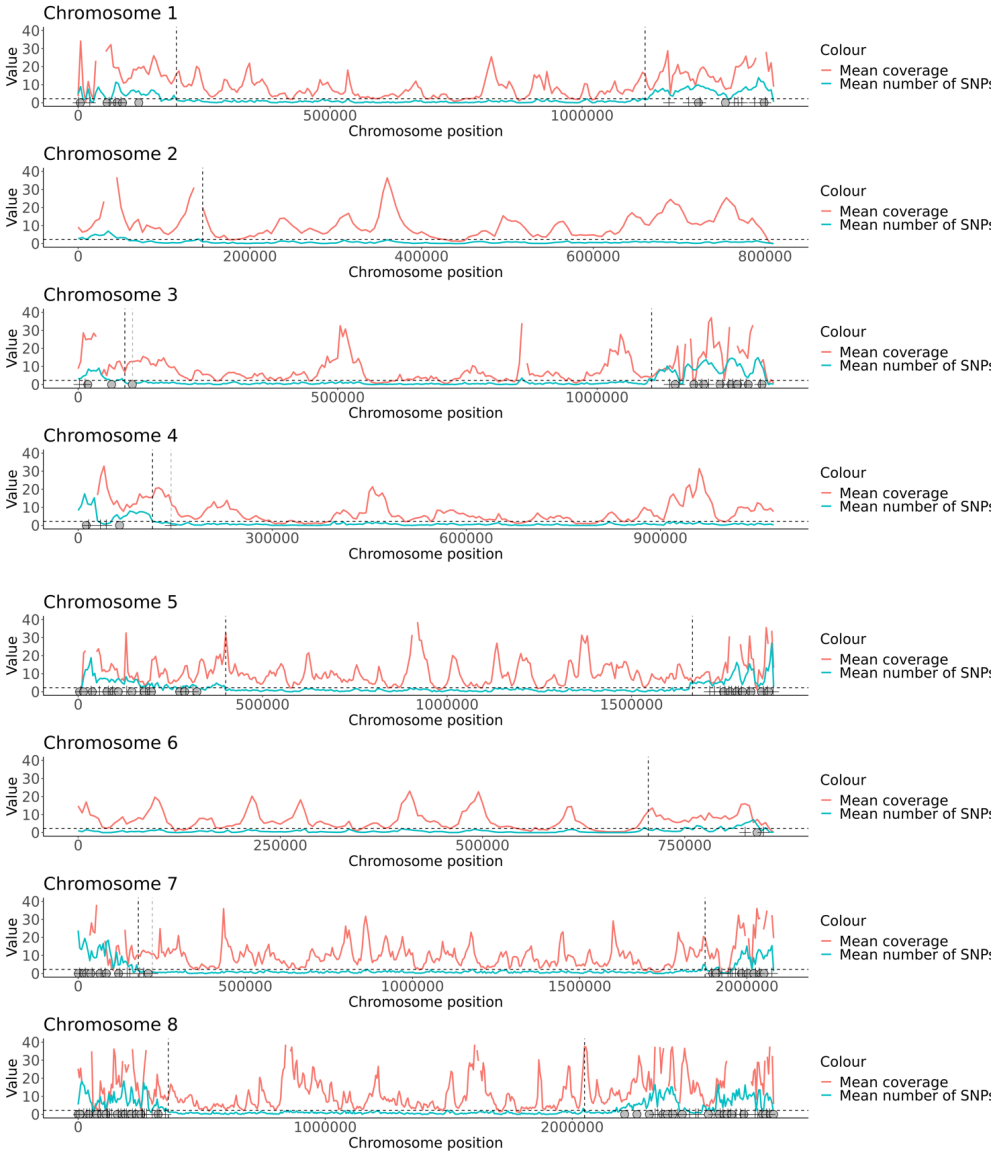
S1 Fig. Genome coverage is increased after SWGA. The top two plots show the coverage (as measured by the number of sequencing reads at each position) of chromosome 5 for sample PM_THA_001 before (No SWGA) and after amplification with Pmset1 (SWGA). The same comparison is shown below with chromosome 10 for sample PM_THA_002. Coverage was calculated using a sliding window of 10 kb. The horizontal y-intercept indicates a coverage of 5-fold which is recommended for confident SNP calling.

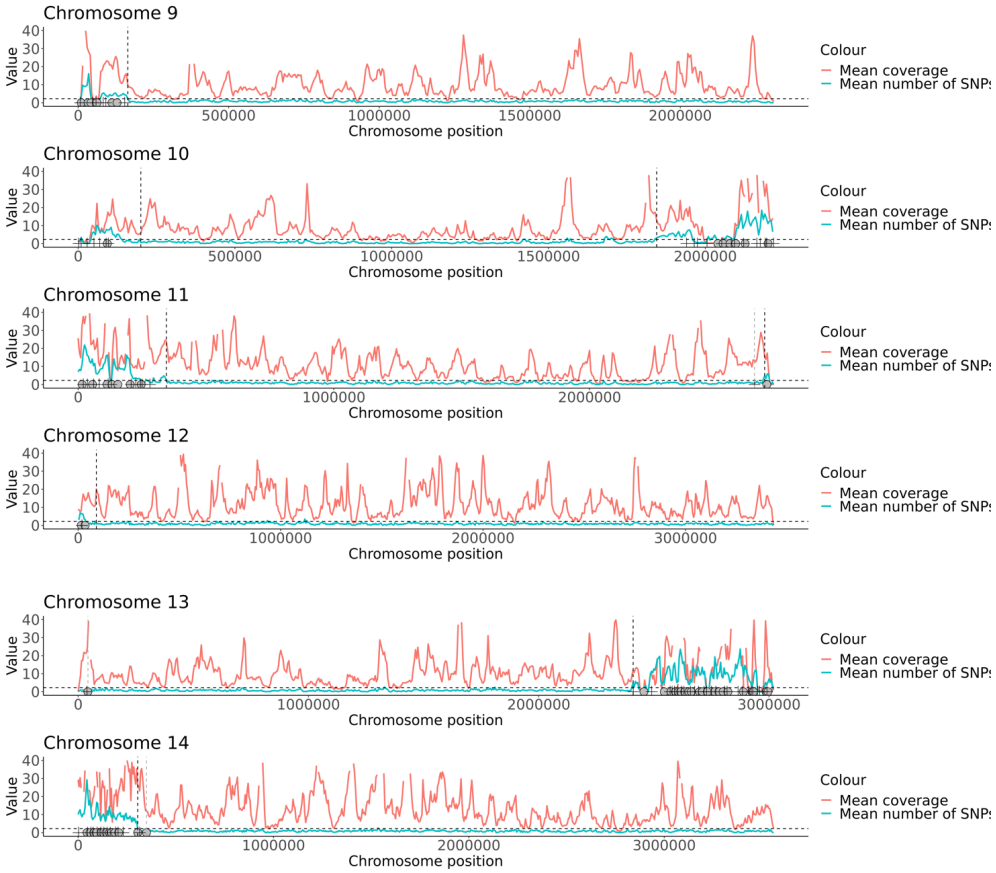


S2 Fig. Sequencing coverage by chromosome without amplification. The distribution of coverage for each position within the genome between two unamplified samples (PM_THA_001 and PM_THA_002), among the 14 nuclear chromosomes and the organellar apicoplast genome. The blue horizontal dashed line represents a cut-off of coverage of 5-fold which is recommended for SNP calling. Outliers are denoted by a black circle, many individual positions will have a coverage of 5-fold, but the majority of the genome falls below this.

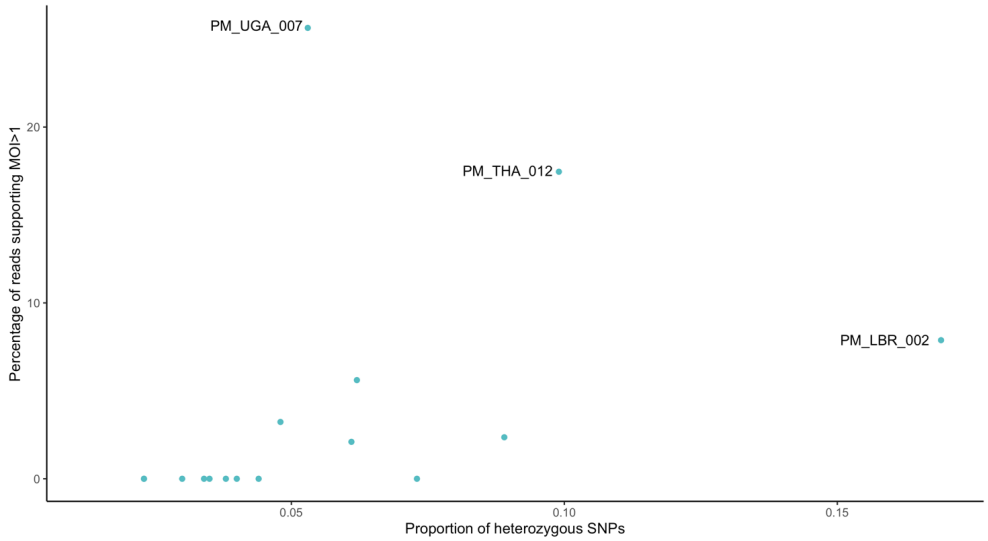


S3 Fig. Number of SNPs seen within 5 kb windows of the *P. malariae* genome. An average of the number of SNPs seen within all 18 samples was generated for 5 kb sliding windows across the entire genome. The distribution of SNPs within each window is shown. The blue dashed x-intercept is at 2.25 SNPs which is used as the limit of SNPs per window for the core genome.

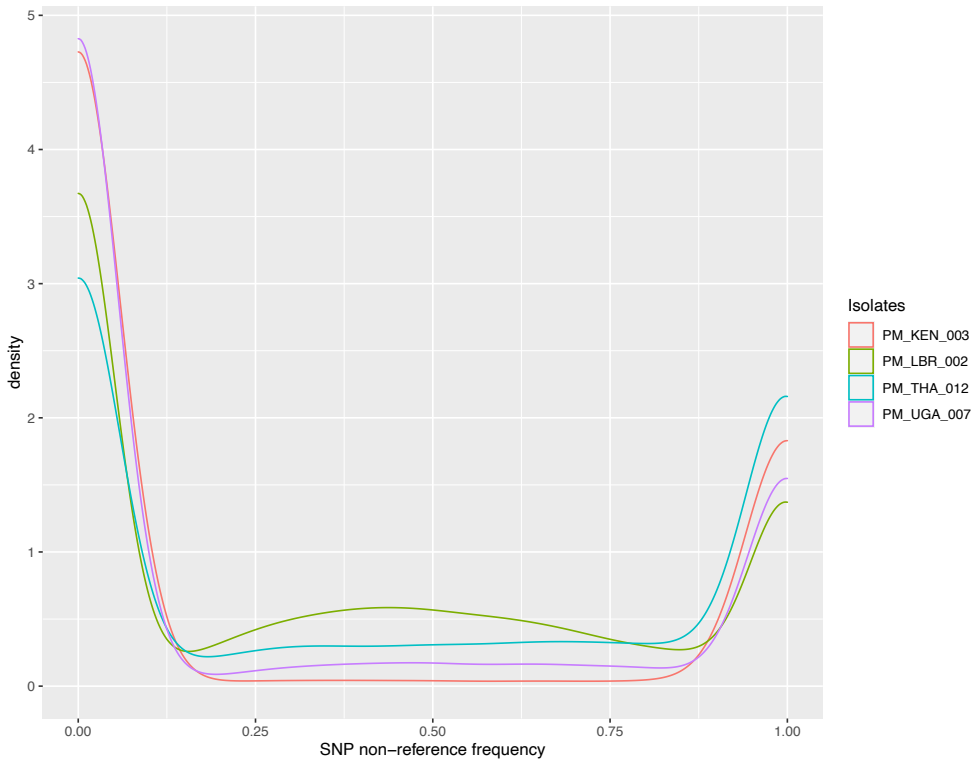

Chromosome 1
Chromosome 2
Chromosome 3
Chromosome 4
Chromosome 5
Chromosome 6
Chromosome 7
Chromosome 8
Value
Chromosome position
Colour
Mean coverage
Mean number of SNPs



S4 Fig. Determining and excluding hypervariable subtelomeric regions. The average coverage (pink) and number of SNPs (blue) between all 18 samples for each nuclear chromosome. The black dashed horizontal line demonstrates the previously chosen SNP limit per 5 kb window (as defined in **S3 Fig.**). Black dashed lines are placed at the suggested hypervariable region cut-off points, where clusters of windows demonstrating >2.25 SNPs are seen. The midpoints of *Pm-fam* gene families are annotated; *Pm-fam-l* gene positions are denoted by a black plus, whilst *Pm-fam-m* gene positions are denoted with a grey circle. There are cases where *Pm-fam* gene families are located outside of the hypothesised hypervariable regions using the 2.25 SNP limit per window, and in these instances, the terminal or initial position of the *Pm-fam* gene is located with a grey dashed line and this is used as the amended hypervariable cut-off position (see **S3 Table** for coordinates).

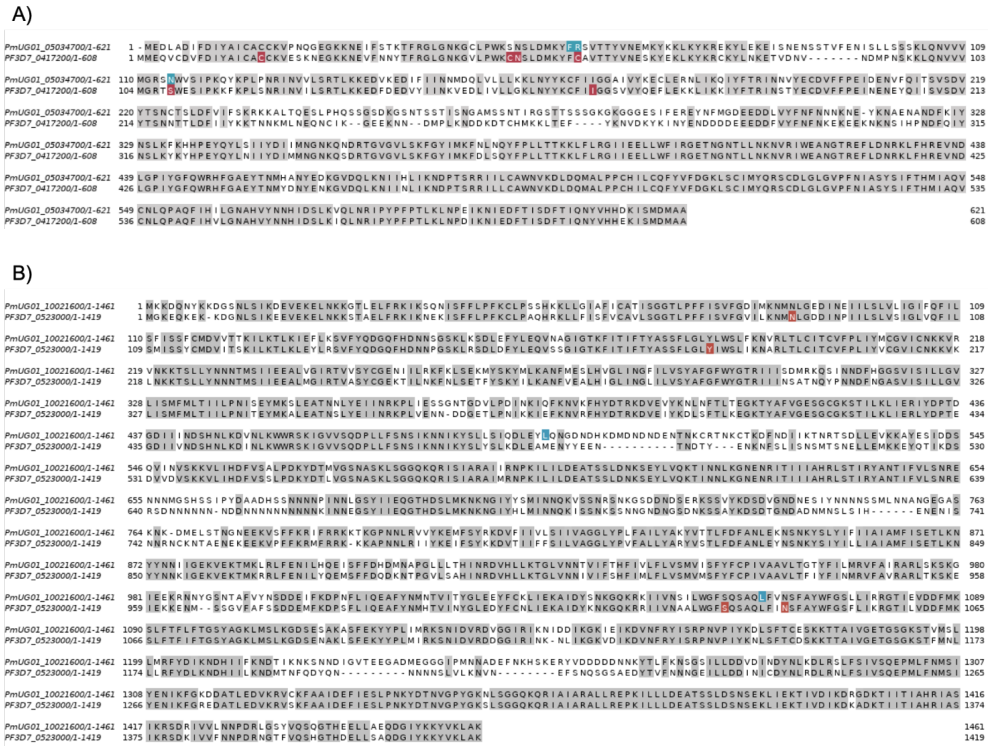


S5 Fig. Determining multiplicity of infection (MOI). EstMoi was ran for all sequenced samples, values were obtained for 18 samples (not including PM_KEN_001, PM_LBR_003 and PM_UGA_001 which did not have sufficient SNP density for estMOI) (**S2 Table**). Three samples appear to separate from the main cluster of samples (PM_LBR_002, PM_THA_012 and PM_UGA_007).



S6 Fig. Density plot of non-reference MAF distribution for isolates with MOI>1.

Distribution (density plot) of non-reference minor allele reference (MAF) for isolates with MOI>1 according to estMOI (PM_LBR_002, PM_THA_012, PM_UGA_007). Isolate PM_KEN_003 did not present evidence of MOI>1 and has been included as reference. The density (y-axis) is scaled so that the area under the curve adds up to 1.



S7 Fig. Substitutions within *P. malariae* orthologs of drug resistance-associated loci in *P. falciparum.*

Amino acid sequences aligned using Clustal Omega. Regions that are conserved between the two positions are highlighted in grey. Positions associated with drug resistance mutations within *P. falciparum* are highlighted in red. Predicted amino acid substitutions found in our *P. malariae* sample set are highlighted in blue. **A)** *PMDHFR* and *PFDHFR.* Both positions 57-58 and 114 within *PMDHFR* lie close to amino acid alterations associated with drug resistance in *P. falciparum (*C59R and S108N/T). **B)** *PMMDR1* and *PFMDR1*. The amino acid substitution at position 1,063 within PmUG01_10021600.1 lies in close proximity to the N1042D substitution associated with quinine resistance.

S1 Table. Pmset1 primers. The sequences primers within Pmset1 are demonstrated. The asterisk in the primer sequence denotes the position of a phosphorothioate bond which inhibits primer degradation. Binding sites within the human genome were calculated using the human GRCh37 genome sequence available at http://grch37.ensembl.org and binding sites within the *P. malariae* genome were calculated using the PmUG01 reference genome available at https://plasmodb.org.

Primer set	Primer sequence	Binding sites in human genome	Binding sites in *P. malariae* genome
Pmset1	TATGTATA*T*T	6054	3057
	TTATTC*G*T	3475	3040
	TTCGTT*A*T	4381	3088
	TTTTTA*C*G	3991	3863
	TATTTC*G*T	3961	3863
	Total Binding Sites	***21862***	***16521***

S2 Table. Sequencing statistics for all samples. Sequence analysis data for all samples. Samples highlighted in grey are comparative samples before SWGA, this is also indicated in the 'Amplification' column. Samples that were excluded due to poor quality and not used in downstream population genetics analysis are labelled with an (E). Accession codes are given for each sample and can be found using the project code (PRJEB37746). Multiplicity of infection (MOI) was calculated using estMOI and the percentage of reads supporting a MOI>1 are given.

Sample	Accession code	Country	Amplified DNA	Pm CT-value (qPCR data)	Parasites/ul (% Parasitaemia)	Total reads	Mapped reads	% Mapped reads	% Genome coverage ≥5	% Genome coverage ≥10	Mean coverage	Mean genic coverage	Median genic coverage	Mean intergenic coverage	Median intergenic coverage	Mean mitochondria coverage	Unfiltered SNPs	Core SNPs filtered	% Genome with MOI>1
PM_THA_001 (E)	/	Thailand	No_SWGA	31.74	256 (0.0064)	6264676	1005761	16.05	0.86	0.12	0.45	0.44	0	0.47	0	/	3	/	/
PM_THA_001	ERS4423439	Thailand	SWGA	31.74	256 (0.0064)	7174584	3154695	43.97	44.17	25.28	8.57	6	3	11	4	23.78	41962	3256	0
PM_THA_002 (E)	/	Thailand	No_SWGA	27.73	576 (0.0144)	5073136	419922	8.28	5.01	0.45	1.25	1.66	1	0.89	0	/	49	/	/
PM_THA_002	ERS4423440	Thailand	SWGA	27.73	576 (0.0144)	8036914	4250708	52.89	64.06	36.16	12.59	16	7	10	6	338.52	43237	6089	0
PM_THA_003	ERS4423441	Thailand	SWGA	30.59	8096 (0.2024)	7852281	4215247	53.68	63.46	40.97	14.63	13	6	16	8	3789.74	45869	6567	2.10
PM_THA_005	ERS4423442	Thailand	SWGA	23.60	6464 (0.1616)	8028153	5735625	71.44	88.31	72.65	22.68	23	18	22	15	1533.33	57182	9998	2.36
PM_THA_006	ERS4423443	Thailand	SWGA	23.72	5312 (0.1328)	5584655	5130545	91.87	75.09	53.23	19.13	14	9	24	12	238.56	56707	6826	0
PM_THA_009a* (E)	ERS4423444	Thailand	SWGA	31.20	64 (0.0016)	3684878	2765795	75.06	44.05	25.70	9.28	7	3	11	4	/	43853	/	0
PM_THA_009b* (E)	ERS4423445	Thailand	SWGA	31.20	64 (0.0016)	2868636	2327144	81.12	27.32	15.84	6.41	4	1	8	2	/	37417	/	0
PM_THA_009_com	ERS4423446	Thailand	SWGA	31.20	64 (0.0016)	6553515	5092953	77.71	52.44	34.66	15.69	11	4	20	6	294.51	48212	4698	0
PM_THA_010	ERS4423447	Thailand	SWGA	23.28	1648 (0.0412)	6934812	4135980	59.64	83.89	62.05	16.49	16	13	17	12	884.78	43804	9042	0

PM_THA_011	ERS4423448	Thailand	SWGA	25.20	16 (0.0004)	6179776	2257408	36.53	63.10	31.03	8.47	8	7	9	6	164.40	43804	5629	0
PM_THA_012	ERS4423449	Thailand	SWGA	24.42	544 (0.0136)	6909366	3264421	47.25	78.03	50.69	12.71	13	10	13	9	414.93	52348	8147	17.46
PM_KEN_001	ERS4423432	Kenya	SWGA	/	/	5709031	3454616	60.51	72.05	44.58	12.03	11	8	13	9	59.60	50899	4820	/
PM_KEN_003	ERS4423433	Kenya	SWGA	/	/	6744405	2186998	32.43	53.11	24.02	7.44	7	5	8	5	28.76	44678	3400	0
PM_LBR_002	ERS4423434	Liberia	SWGA	/	/	7180349	4242520	59.09	86.46	66.06	17.10	17	14	17	13	309.94	63484	8474	7.88
PM_LBR_003 (E)	ERS4423435	Liberia	SWGA	/	/	9338062	5627043	60.26	0.07	0.03	0.19	0	0	0	0	/	25	/	/
PM_SDN_001	ERS4423436	Sudan	SWGA	/	/	5369713	4461837	83.09	72.52	50.32	16.96	13	8	20	11	419.30	54943	5154	0
PM_SLE_002	ERS4423437	Sierra_Leone	SWGA	/	/	8039293	4286570	53.32	84.15	62.52	16.46	17	13	16	12	258.93	53739	7034	5.61
PM_SLE_004	ERS4423438	Sierra_Leone	SWGA	/	/	6933075	1513093	21.82	44.96	14.67	5.35	6	4	5	4	184.48	28204	2554	0
PM_UGA_001	ERS4423451	Uganda	SWGA	/	/	2846843	1666685	58.55	46.56	17.59	5.90	5	4	6	4	50.56	37419	2534	/
PM_UGA_003	ERS4423452	Uganda	SWGA	/	/	6500133	3316826	51.03	81.84	53.76	13.22	13	11	13	10	373.45	53571	6525	3.23
PM_UGA_007	ERS4423453	Uganda	SWGA	/	/	6673599	2481168	37.18	58.43	32.44	9.48	9	6	10	6	87.20	48414	3836	25.64

S3 Table. Identification of core nuclear genome positions. Not all chromosomes have annotated *Pm-fam* genes or hypervariable regions at both ends of the chromosomes, these chromosomes are indicated with an asterisk.

Chromosome	Start	End	Classification
PmUG01_01_v1	0	194999	Hypervariable
PmUG01_01_v1	195000	1125000	Core
PmUG01_01_v1	1125001	1381517	Hypervariable
PmUG01_02_v1*	0	59999	Hypervariable
PmUG01_02_v1*	60000	813519	Core
PmUG01_03_v1	0	89999	Hypervariable
PmUG01_03_v1	90000	1105000	Core
PmUG01_03_v1	1105001	1341933	Hypervariable
PmUG01_04_v1*	0	143680	Hypervariable
PmUG01_04_v1*	143681	1076824	Core
PmUG01_05_v1	0	399999	Hypervariable
PmUG01_05_v1	400000	1665000	Core
PmUG01_05_v1	1665001	1887011	Hypervariable
PmUG01_06_v1*	0	705000	Core
PmUG01_06_v1*	705001	862289	Hypervariable
PmUG01_07_v1	0	221927	Hypervariable
PmUG01_07_v1	221928	1875000	Core
PmUG01_07_v1	1875001	2080590	Hypervariable
PmUG01_08_v1	0	364999	Hypervariable
PmUG01_08_v1	365000	2050000	Core
PmUG01_08_v1	2050000	2818517	Hypervariable
PmUG01_09_v1*	0	164999	Hypervariable
PmUG01_09_v1*	165000	2312276	Core
PmUG01_10_v1	0	199999	Hypervariable
PmUG01_10_v1	200000	1854000	Core
PmUG01_10_v1	1854001	2219074	Hypervariable
PmUG01_11_v1	0	344999	Hypervariable
PmUG01_11_v1	345000	2644881	Core
PmUG01_11_v1	2644882	2721161	Hypervariable
PmUG01_12_v1*	0	89999	Hypervariable
PmUG01_12_v1*	90000	3436769	Core
PmUG01_13_v1	0	42699	Hypervariable
PmUG01_13_v1	42700	2410000	Core
PmUG01_13_v1	2410001	3023685	Hypervariable
PmUG01_14_v1	0	350186	Hypervariable
PmUG01_14_v1	350187	3564280	Core

S4 Table. Pairwise genome-wide nucleotide diversity (π; x10^{-4}). Nucleotide diversity values are coloured to demonstrate samples with high or low diversity in comparison to other samples. Low diversity is shaded in yellow and pale orange, increasing to darker orange for high diversity (yellow for 0 to 1, pale orange for 1 to 2, medium orange for 1 to 2, and dark orange for 2 to 4; all x10^{-4}).

Region		Thailand									East Africa						West Africa		
	Samples	PM_THA_001	PM_THA_002	PM_THA_003	PM_THA_005	PM_THA_006	PM_THA_009com	PM_THA_010	PM_THA_011	PM_THA_012	PM_KEN_001	PM_KEN_003	PM_UGA_001	PM_UGA_003	PM_UGA_007	PM_SDN_001	PM_LBR_002	PM_SLE_002	PM_SLE_004
Thailand	PM_THA_001	/	2.1	2.23	2.29	2.04	1.75	1.99	1.64	1.87	2.75	2.79	2.35	2.91	2.64	2.72	3	3.05	2.41
	PM_THA_002	2.1	/	0.36	2.52	2.16	2.13	2.32	2.04	1.94	3.06	3.1	2.59	3.19	2.91	3.08	3.26	3.33	2.75
	PM_THA_003	2.23	0.36	/	2.52	2.25	2.18	2.34	2.1	1.98	3.12	3.12	2.67	3.21	2.96	3.12	3.25	3.3	2.74
	PM_THA_005	2.29	2.52	2.52	/	2.38	2.21	2.49	2.18	2.1	3.25	3.26	2.95	3.53	3.17	3.37	3.64	3.68	3.06
	PM_THA_006	2.04	2.16	2.25	2.38	/	1.99	2.38	2.17	2.17	2.91	2.98	2.55	3.05	2.85	2.9	3.26	3.27	2.68
	PM_THA_009com	1.75	2.13	2.18	2.21	1.99	/	2.13	1.4	2.07	2.8	2.88	2.43	2.98	2.82	2.83	3.11	3.14	2.62
	PM_THA_010	1.99	2.32	2.34	2.49	2.38	2.13	/	1.94	2.23	3.22	3.24	2.78	3.39	3.11	3.26	3.44	3.55	2.89
	PM_THA_011	1.64	2.04	2.1	2.18	2.17	1.4	1.94	/	2.03	2.91	2.92	2.55	3.07	2.78	2.89	3.08	3.26	2.57
	PM_THA_012	1.87	1.94	1.98	2.1	2.17	2.07	2.23	2.03	/	3.08	3.06	2.67	3.2	2.91	3.09	3.23	3.38	2.72
East Africa	PM_KEN_001	2.75	3.06	3.12	3.25	2.91	2.8	3.22	2.91	3.08	/	0.27	1.87	2.45	2.09	2.26	2.48	2.43	1.89
	PM_KEN_003	2.79	3.1	3.12	3.26	2.98	2.88	3.24	2.92	3.06	0.27	/	1.93	2.4	2.08	2.33	2.5	2.47	1.89
	PM_UGA_001	2.35	2.59	2.67	2.95	2.55	2.43	2.78	2.55	2.67	1.87	1.93	/	2.07	1.86	1.94	2.11	2.17	1.69
	PM_UGA_003	2.91	3.19	3.21	3.53	3.05	2.98	3.39	3.07	3.2	2.45	2.4	2.07	/	2.2	2.45	2.71	2.78	2.19
	PM_UGA_007	2.64	2.91	2.96	3.17	2.85	2.82	3.11	2.78	2.91	2.09	2.08	1.86	2.2	/	2.15	2.34	2.38	1.88
	PM_SDN_001	2.72	3.08	3.12	3.37	2.9	2.83	3.26	2.89	3.09	2.26	2.33	1.94	2.45	2.15	/	2.47	2.59	2.01
West Africa	PM_LBR_002	3	3.26	3.25	3.64	3.26	3.11	3.44	3.08	3.23	2.48	2.5	2.11	2.71	2.34	2.47	/	2.71	2.09
	PM_SLE_002	3.05	3.33	3.3	3.68	3.27	3.14	3.55	3.26	3.38	2.43	2.47	2.17	2.78	2.38	2.59	2.71	/	2.06
	PM_SLE_004	2.41	2.75	2.74	3.06	2.68	2.62	2.89	2.57	2.72	1.89	1.89	1.69	2.19	1.88	2.01	2.09	2.06	/

S5 Table. Drug resistance associated genes in *P. falciparum* and their known orthologs in *P. malariae*.

	P. falciparum (3D7)		*P. malariae (UG01)*	
Gene	**Gene ID**	**Paper code**	**Gene ID**	**Paper code**
crt	PF3D7_070900	*Pfcrt*	PmUG01_01020700	*Pmcrt*
dhfr-ts	PF3D7_0417200	*Pfdhfr*	PmUG01_05034700	*Pmdhfr*
pppk-dhps	PF3D7_0810800	*Pfdhps*	PmUG01_14045500	*Pmdhps*
mdr1	PF3D7_052300	*Pfmdr1*	PmUG01_10021600	*Pmmdr1*
kelch13	PF3D7_1343700	*Pfk13*	PmUG01_12021200	*Pmk13*

S6 Table. WGS data from mixed infections. Four samples with mixed infections (multiple species) were sequenced and mapped to their respective reference genomes. The proportion of each parasite present was analysed using qPCR and the resulting cycle thresholds (CTs) are given. The following reference genomes are used: *P. malariae* (PmUG01), *P. vivax* (PvP01, https://plasmodb.org/common/downloads/Current_Release/PvivaxP01/fasta/data/) *P. ovale curtisi* (PocGh01, https://plasmodb.org/common/downloads/Current_Release/PovalecurtisiGH01/fasta/data/), *P. ovale wallikeri* (PowCR01 https://www.ebi.ac.uk/ena/data/view/ERS1452913).

Sample	Mapped_to	Pf_CT	Po_CT	Pm_CRT	Pv_CRT	Parasites/ul (%Parasitemia)	Total_reads	Map_reads	%_Mapped_reads	%_Genome_cov5	%_Genome_cov10
PM_THA_004	PmUG01	/	/	23.48	31.02	1520 (0.038)	7644777	4619930	60.433	85.004	66.919
PM_THA_004	PvP01	/	/	23.48	31.02	1520 (0.038)	7644777	37351	0.489	0.015	0.010
PM_THA_008	PmUG01	/	29.33	31.12	29.26	192 (0.0048)	3679905	305249	3.993	0.072	0.023
PM_THA_008	PocGH01	/	29.33	31.12	29.26	192 (0.0048)	3679905	1540021	41.849	25.825	15.677
PM_THA_008	PowENA	/	29.33	31.12	29.26	192 (0.0048)	3679905	117162	3.184	0.169	0.100
PM_THA_008	PvP01	/	29.33	31.12	29.26	192 (0.0048)	3679905	242561	6.592	0.046	0.024
PM_THA_013	PmUG01	/	33.43	26.31	/	4560 (0.114)	7533858	3928080	52.139	83.985	61.240
PM_THA_013	PocGH01	/	33.43	26.31	/	4560 (0.114)	7533858	106192	1.410	0.035	0.014
PM_THA_013	PowENA	/	33.43	26.31	/	4560 (0.114)	7533858	48638	0.646	0.034	0.018
PM_THA_020	PmUG01	/	26.72	28.33	26.70	4880 (0.122)	6917577	236151	3.414	0.048	0.020
PM_THA_020	PocGH01	/	26.72	28.33	26.70	4880 (0.122)	6917577	2605168	37.660	23.871	20.314
PM_THA_020	PowENA	/	26.72	28.33	26.70	4880 (0.122)	6917577	895565	12.946	28.435	12.095
PM_THA_020	PvP01	/	26.72	28.33	26.70	4880 (0.122)	6917577	391155	5.655	0.046	0.038

CHAPTER 3

3 First global genomic analysis of *Plasmodium malariae* parasites reveals genetic diversity and key mutations in antimalarial drug targets

RESEARCH PAPER COVER SHEET

SECTION A – Student Details

Student ID Number	**1600466**	**Title**	Miss
First Name(s)	Amy		
Surname/Family Name	Ibrahim		
Thesis Title	From genome to function: A genomic investigation into understudied populations of the malaria parasites *Plasmodium malariae* and *P. vivax*		
Primary Supervisor	Prof. Susana Campino		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	N/A		
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Nature Communications

Please list the paper's authors in the intended authorship order:	Amy Ibrahim, Emilia Manko, Debbie Nolder, Steffen Borrmann, Silvia Maria Di Santi, Francois Nosten, Colin J. Sutherland ,Taane G. Clark, Susana Campino
Stage of publication	Submitted

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I undertook laboratory work including whole genome amplification and the preparation of samples for sequencing. I also performed bioinformatic analysis and interpreted the results under the supervision of my supervisors. I wrote the first draft of the manuscript that was then circulated to supervisors and after to co-authors.

SECTION E

Student Signature	[redacted]
Date	25/02/2022

Supervisor Signature	[redacted]
Date	February 24, 2022

First global genomic analysis of *Plasmodium malariae* parasites reveals genetic diversity and key mutations in antimalarial drug targets

Short title: Global genomic diversity of *Plasmodium malariae* parasites

Amy Ibrahim [1], Emilia Manko [1], Debbie Nolder [1,2], Steffen Borrmann[3], Silvia Maria Di Santi [4], Francois Nosten[5], Colin J. Sutherland [1,2], Taane G. Clark[1,6,#], Susana Campino[1,#]

[1] Faculty of Infectious & Tropical Diseases, London School of Hygiene & Tropical Medicine, London, UK

[2] Public Health England Malaria Reference Laboratory, London School of Hygiene & Tropical Medicine, London, UK

[3] Institute for Tropical Medicine, Eberhard Karls University of Tübingen, Tübingen, Germany

[4] School of Medicine, University of São Paulo, Brazil

[5] Shoklo Malaria Research Unit, Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Mae Sot, Tak, Thailand

[6] Faculty of Epidemiology & Population Health, London School of Hygiene & Tropical Medicine, London, UK

[#] Corresponding authors

E-mail: Susana.campino@lshtm.ac.uk and Taane.clark@lshtm.ac.uk

ABSTRACT

Malaria caused by *Plasmodium malariae* is a neglected tropical disease with a wide geographic distribution and contributing up to 40% of cases in regions of Africa and South America. This parasite species has unique biological and clinical traits, such as the quartan life cycle, an ability to persist in the blood causing chronic infections, and associations with severe disease including nephropathologies and anaemia. These unique features may underly the observed reduced parasite drug response and drive malaria recurrence, representing a challenge for disease elimination. To provide much needed insights into *P. malariae* biology, we explore the genomic diversity of this parasite across 4 endemic regions, spanning 25 countries. Through using selective whole genome amplification, we obtained whole genome sequencing data for 235 clinical isolates of *P. malariae* and identified 1,288,675 genome-wide SNPs, filtered to 221,656 SNPs within the core nuclear genome, leaving 143,201 filtered high-quality SNPs. In the first global genomic analysis of *P. malariae*, we determine population structure, demonstrating a clear separation in parasites from Africa and Asia, as well as shared ancestry between South American and African parasites. We identify signals of selection in genes associated with the natural human immune response (*pmmsp3*) in addition to orthologs of genes associated with artemisinin recrudescence (*pmeIF2a*). Finally, we identify mutations within orthologs of genes associated with drug susceptibility in *P. falciparum*, including amino acid substitutions in the DHFR-TS, CRT, MDR2, K13, CORONIN, PI4K, MRP1 and UBP1 proteins Our study provides the first evaluation of global *P. malariae* population structure and genomic diversity, and creates a valuable resource for elucidating important insights into the biology and pathogenesis of this neglected pathogen.

Word count: 270

INTRODUCTION

Plasmodium malariae is a neglected causative agent of human malaria, with limited research and poor understanding of the unique biology of this disease. Whilst typically rare and benign in comparison to *P. falciparum*, an increased awareness of this neglected *Plasmodium spp.* present within populations is crucial for malaria elimination. *P. malariae* infections are unique, displaying a quartan malaria fever pattern and are able to persist asymptomatically in the blood for years or decades after exposure [12,3]. Although rarely life threatening, substantial illness such as anaemia, splenomegaly and the development of nephrotic syndrome have been described [4–8]. *P. malariae* infections are challenging to diagnose due to low parasitaemias and their complexity, typically involving other *Plasmodium spp.* parasites [9]. However, molecular detection methods are revealing that *P. malariae* infections are much more common than previously estimated, reaching 20% of cases in the Thai-Myanmar border and up to 40% in Nigeria and Colombia [10,11 12]. *P. malariae* is considered the second most common source of human malaria in the African continent, but is also found in South America, Asia and Oceania [13 14]. While usually treatable, concerns over the reduced susceptibility of this "persistent" species to standard antimalarial drugs have been highlighted. Well-documented evidence from imported cases[1,15] and studies in Africa[16,17] and Indonesia[18] have indicated an ability to survive artemisinin-based combination therapy (ACT) and chloroquine treatment [19].

The unique features of *P. malariae* biology, including the longer erythrocytic cycle which may reduce drug susceptibility, the presentation of chronic infections and untraceable transmission that can drive malaria recurrence, represent a challenge to disease elimination. With current control methods focused on falciparum malaria, *P. malariae* parasites are likely to persist long beyond *P. falciparum* decline as seen in Gabon and eastern Tanzania [4,20]. These insights highlight the need for an improved understanding of the underlying biology and pathogenesis of this neglected parasite, especially to ensure that better treatments and control measures are available to assist the elimination of all human malaria. As demonstrated for *P. falciparum* and *P. vivax*, large-scale analysis of genome diversity of field isolates has enabled investigations into parasite population genetics, transmission dynamics, and identified genomic regions under selective pressure, such as drug resistance associated genes [21–25]. However, to date only a few genomes have been generated

for neglected malaria parasites[13,26,27]. Due to low parasitaemias and the lack of *in vitro* culturing approaches, these parasite genomes are difficult to sequence. We have recently developed and validated a methodology for selective whole genome amplification (SWGA) of DNA of *P. malariae* isolates from clinical samples [27], making it now possible to explore the genomic diversity of this species at a larger scale. Here, we apply SWGA and whole genome sequencing methods to DNA sourced from >200 clinical isolates across four continents, to investigate the complexity and variability of *P. malariae* genomes (size ~31.9 Mb), their population structure, and identify genomic regions under selective pressure.

RESULTS

Genomic database, multiclonality and nucleotide diversity

We performed SWGA and WGS on 235 clinical isolates spanning 31 countries. After stringent filtering for high quality samples (removing samples with missing data for >50% of the SNP positions) and SNPs (removing positions with missing calls in >10% of all isolates), we generated a final genomics database of high quality isolates (n = 155), containing samples obtained across Africa (n = 130, Algeria, Angola, Cameroon, Congo, Equatorial Guinea, Gabon, Ghana, Guinea, Ivory Coast, Kenya, Liberia, Malawi, Mali, Mozambique, Nigeria, Senegal, Sierra Leone, Sudan, Tanzania, Uganda), Asia (n = 17, Thailand, Malaysia, Indonesia), Oceania (n = 2, Papua New Guinea, Papua Indonesia) and Brazil (n = 6). Using this filtered data set, a total of 143,201 unique SNPs within the core genome were identified (see **S1 Table** for list of isolates; **S1 Figure** for distribution of filtered isolates). We investigated within host parasite diversity using the F_{WS} metric [28], and demonstrated the lowest levels of diversity within Brazilian isolates (average F_{WS} score 0.961, n = 6, all isolates with $F_{WS} > 0.85$), suggesting a lower possibility of being infected with multiple parasite strains in this country, and less likely parasite inbreeding. Lower F_{WS} scores, suggestive of higher within host parasite diversity, were observed across Africa (mean $F_{WS} = 0.896$, with 84% of isolates with $F_{WS} > 0.85$) and Asia (mean $F_{WS} = 0.806$, with 75% isolates with $F_{WS} > 0.85$). Oceania demonstrates the lowest F_{WS} scores (mean $F_{WS} = 0$), however there are only two isolates in this continent (ERR1355083 and PM_PNG_001) which may skew the results (**S2 Table, S2 Figure).**

Global population structure of *P. malariae*

The filtered dataset of 155 isolates and 143,201 unique SNPs in the core genome (excluding hypervariable and subtelomeric regions) was used to investigate population structure. Both principal component analysis (PCA) and a maximum likelihood (ML) tree show independent clustering of Asian and African isolates. Within Africa, population structure is complex, with no obvious clustering related with geographical positioning across the continent (**Figure 1**). The Brazilian isolates show a close relationship with the African population, which has also been demonstrated in *P. falciparum* genomic studies [29].

Admixture analysis demonstrated that there are most likely 4 ancestral populations across the 155 isolates (**S3 Figure**), with one clear and distinct Asian ancestral population (K4), one clear South American ancestral population (K1), and 3 ancestral populations within Africa (K1, K2 and K3) (**Figure 1**). Isolates from Brazil and Nigeria share a large proportion of the K1 ancestral population which is seen across many countries in Africa in smaller proportions, which may be suggestive of an ancestral African population. Uganda has one predicted ancestral population (K2) which is common in Ghana, Kenya, and Malawi. Of note is Angola, which appears to be distinct from other African countries, with most samples aligning to population K3. Whilst it is currently difficult to ascertain fine scale geographic separation within Africa, it is evident that there are differences in the parasite populations between these regions (**Figure 1**).

Population allele frequency differences using the F_{ST} fixation index

We investigated the SNP positions driving the population separation between African (n = 130) and Asian (n = 14) isolates using the fixation index (F_{ST}). A total of 167 SNPs were found to be highly differentiating between the two populations, of which, 84 lead to nonsynonymous amino acid substitutions in multiple genes (perfect differentiation, F_{ST} = 1, **S3 Table**, **Table 1**). Highly differentiating SNPs were seen in genes linked to parasite surface proteins such as the circumsporozoite protein (*pmcsp*), the parasite antigen used in the development of the only approved malaria vaccine, Mosquirix [30], and a merozoite surface protein (*pmmsp9*). The analysis revealed a candidate multi-drug resistance gene (*pmmrp2*), which has a functional orthologue in *P. falciparum*, and is linked to a class of ABC transporter proteins, widely regarded as associated with drug transport and their respective efficacy [31]. A further differentiating SNP is found within the gene encoding the cystine modular protein (*pmcrmp1*), which has been shown in *P. berghei* to

be associated with mosquito development stages, with differences in continents likely due to the presence of different mosquito vectors in different continents [32] (**Table 1**). In this study we were interested in SNPs driving separation between the two populations, therefore we only investigated SNPs with perfect differentiation, which are likely to suggest regions under differing directional selection. Looking at SNPs with lower FST scores can also be biologically interesting, and may be indicative of regions undergoing balancing selection, whereby it is beneficial to have multiple variant alleles at a particular locus. Lower FST scores may also be indicative of the same selective pressure acting on two different populations, for example when the same drug is being used in two countries, leading to selection of an allele that mediates reduced susceptibility to the drug.

Regions of homology using identity by decent

The global dataset of monoclonal isolates (n = 127, **S1 Table**) was investigated for regions of genetic homology within continents using an identity-by-descent (IBD) metric. Brazil displayed the most homogenous parasite population (mean IBD: 0.133, range: 0.014 – 0.470), likely because the isolates were sourced from a similar geographic region, namely, the southeast coast of Brazil in the states of São Paulo and Santa Catarina. At a continental level, Asia (mean: 0.105, range: 0.006 – 0.948) had higher IBD than Africa (mean: 0.037, range: 0-1), which was more heterogenous (**S4 Figure, S5 Figure).** We scanned the genome in 10 kbp sliding windows to identify signals of high IBD between continents, which revealed signals in Africa, Asia, and Brazil in two regions of chromosome 10 where a bromodomain protein (*pmdbp1*) is encoded (Africa 0.171, Asia 0.235, Brazil 0.284), an orthologue of a chromatin binding protein associated with erythrocyte invasion in *P. falciparum* [33]. There was a further signal across the three populations encompassing an acetyl-coA transporter (*pmact*), whose ortholog has been implicated in multi-drug resistance in *P. falciparum i*nfections and may be involved in broad mechanisms of antimalarial drug resistance [34]. There were multiple Brazilian specific signals of homology (**S5 Figure**), including two regions in chromosome 5 (IBD > 0.2) where a gene associated with both asexual and sporozoite development encoding protein ligase B (*pmlipb*) is found [35], a signal within chromosome 7 (IBD = 0.28) involving a gene encoding an ATP-dependant protease subunit essential for parasite growth and development (*pmcipy*) [36], and two regions of homology within chromosome 12 (IBD = 0.3) encompassing a gene encoding an ABC transporter (*pmabcb6*), potentially involved in drug resistance [37] (**S4 Table, S5 Figure**).

Regions under selection

We used the single population iHS metric to investigate signals of positive selection within monoclonal parasite populations ($F_{WS} > 0.85$, n = 127, **S1 Table**) at the continent level (Africa, n = 108; Asia, n = 14). For the African continent, signals of selection (iHS $P < 10^{-4}$) were observed in genes associated with DNA replication and cellular processes (e.g., *pmphd1*)[38], surface proteins such as *pmmsp3* whose *P. falciparum* ortholog has been associated with the natural immune response in clinical isolates[39], and putative gamete antigens 27/25 (PmUG01_14017400/17600). Other hits included the eukaryotic initiation factor-2a (*pmeIF2a*) whose *P. falciparum* ortholog has been associated with recrudescence after treatment with artemisinin [40], and Plasmodium helical interspersed subtelomeric domain proteins (*pmPHIST*) which are thought to be associated with interactions between the parasite and the host erythrocyte [41]. Asian signals of selection were seen within Plasmodium genes of unknown function (PUFs), a ring finger protein (*pmrnf1*) and an ATP-dependant RNA helicase (*pmdbp6*) (**S5 Table**). Similarly, the Rsb score was used to determine differenntial signals of positive selection between African and Asian populations. These signals included PmUG01_13022300, an ortholog of a *P. falciparum* gene associated with ookinete gliding motility (*pfcdc*50) [42], a PUF on chromosome 10 (PmUG01_13022300), and a gene encoding the TRAP-like protein (TREP) associated with sporozoite gliding motility in *P. falciparum* [43] (**S6 Table**).

Identification of key mutations in drug-resistance associated candidate loci

Whilst there are no confirmed molecular markers for drug resistance in *P. malariae,* treatment failures have been recorded with artemisinin-based therapies[16,17] and chloroquine[19]. SNPs have previously been described in *P. malariae* orthologs of genes associated with drug susceptibility including *pmdhfr* [44], *pmdhps* and *pmmdr1* [27] (see **S7 Table** for a list). Overall, in the global dataset, 286 missense SNPs were identified across 15 genes putatively associated with drug susceptibility in Plasmodium parasites (**S8 Table**). High frequency SNPs (>10% in a population or globally), include two fixed SNPs, one within the ubiquitin hydrolase (*pmubp1*) associated with artemisinin resistance in *P. falciparum* [45] (I2307S), and another in *pmmdr2* (V1028) (**Table 2**). Additionally, we identified two missense high frequency SNPs within the chloroquine resistance transporter *pmcrt* (E278D: Asia 27%; G383V: Brazil 67%), whose ortholog in *P. falciparum* is associated

with chloroquine resistance [46] (**Table 2, S8 Table).** Three high frequency SNPs are found in *pmdhfr* (A15S: Brazil 33%; F57L: Africa 21%; N114S: Global 71%) (**Table 2**), in addition to 12 less frequent SNPs (**S8 Table**). *Pmdhfr* positions align with resistance-associated amino acid codons in the *P. falciparum pfdhfr* gene, including F57L, R58S and N114S [27], suggestive of functional effects. High frequency SNPs are found within ABC transporter genes, including *pmmrp1* (D1630N: Global 72%), *pmmrp2*, (A1519V, T679S, I631V: Global > 50%), and *pmmdr2* (V1028A: Global 100%; V426A; Global 70%) (**Table 2**). The phosphatidylinositol-4-OH kinase (*pmpi4k*) gene is the target of a novel class of imidazopyrazine antimalarials [47], with 4 of 26 SNPs detected at high frequency (E1087K, A261S, G255D, G241D: Global > 10%) (**S8 Table**). A total of 6 non-synonymous SNPs were found within the *pmk13* kelch propeller (M156V, Y372H, V404I, T505V, F507Y, V625I), the ortholog of the molecular marker for artemisinin resistance [48]. A further four SNPs were identified within *Pmcoronin*, an actin-binding protein associated with artemisinin resistance in *P. falciparum* [49], but all had low frequency (<5%) (**S8 Table**).

DISCUSSION

The genomic diversity and population structure of *P. malariae* parasites is poorly understood. Here, we generated and analysed the largest collection of *P. malariae* isolates (n = 235) using whole genome sequence data, with representation across 26 countries. Previously, the largest study of *P. malariae* genomic sequence data investigated 18 isolates, with this work increasing the number of high-quality SNPs in the core nuclear genome 7.4-fold (from 29,899 to 221,656 SNPs).

Isolates from Brazil had less complex infections, potentially due to their originating source being low transmission regions (São Paulo and St. Catarina). Lower F_{WS} scores or greater complexity were observed for Africa and Thailand, however, the majority (>75% isolates) had F_{WS} scores >85%, thus most involved infections with fewer genotypes, lowering the risks of parasite out-crossing. F_{WS} scores did not correlate with whether the parasite isolate came from a traveller returning to the UK, or whether they were obtained from local medical facilities in malaria endemic regions.

Population genetics analysis demonstrated that the Asian parasite population is distinct from parasites in Africa and South America. Admixture analysis revealed a distinct Asian ancestral

population (K4) and suggests shared ancestry between isolates from Brazil and Africa, specifically those in Nigeria (K1). Across Africa, the population structure was complex, where Admixture analysis highlighted two common ancestral population present in all African countries (K1 and K2), but also a distinct parasite population found within Angola (K3), which is seen in some other African countries but at a much lower proportion. Africa demonstrates a more complex ancestral population, with three populations seen, which is consistent with initial genomic studies of *P. falciparum* [28], and subsequent investigations with much larger numbers of isolates were pivotal in improving resolution[24]. Additionally, isolates from Thailand were all gathered from one collaborator; therefore, this single ancestral population is unlikely to be respective of the whole country, with further sampling in this region needed. Thus, reinforcing the need to continue to investigate *P. malariae* genomic diversity using isolates from systematically collected populations, facilitating larger samples sizes and temporal analysis.

We investigated the SNPs driving differences between populations using the F_{ST} metric and found 167 SNPs (84 nonsynonymous amino acid substitutions) with perfect differentiation (FST = 1) between African and Asian parasites in genes encoding parasite surface proteins, alongside genes associated with drug susceptibility and mosquito stages. In this study, we were interested in differential pressures on parasites in separate regions, therefore we only investigated SNPs with perfect differentiation between populations ($F_{ST} = 1$), however, lower FST scores may be indicative of different selective pressures for example balancing selection whereby multiple variants are kept at a specific locus. The sample sizes for continent populations in this study is a confounding factor regarding F_{ST} analysis (Africa n = 130, Asia n = 17), which could be addressed by a subsampling approach whereby randomly picked smaller subsets of the larger population group (Africa) are assessed individually against the Asian group and an overall average F_{ST} score is given.

Identity-by-descent (IBD) analysis was used to further understand the structure and selective pressures present within these parasite populations. As expected, Brazilian isolates sourced in relatively close geographical proximity demonstrated the highest levels of IBD, followed by Asia and then Africa, with global signals of homology encompassing genes potentially involved in erythrocyte invasion (*pmdbp1*) and drug susceptibility (*pmact*), suggesting similar selective pressures against these genes across the globe.

The extent and nature of antimalarial resistance within *P. malariae* parasites is understudied, even though these parasites are usually present in co-infections, and continually exposed to antimalarial drugs used against the dominant malaria parasites *P. falciparum* and *P. vivax*. Possible treatment failures have been recorded after ACT and chloroquine treatments. Therefore, we further explored the presence of missense SNPs in orthologs of genes associated with drug susceptibility in other Plasmodium species. High frequency missense SNPs were present in many genes of which *P falciparum* orthologues are involved in drug resistance. These include the chloroquine resistance transporter (*pmcrt*), *pmdhfr* (associated with antifolate resistance), and other orthologous genes linked to artemisinin resistance, such as ubiquitin hydrolase (*pmubp1*) and the *pmk13* kelch propeller. To further understand any selection acting upon the loci identified here, phenotypic information from drug susceptibility studies linked with genetic variants at these loci is needed. Although a *P. malariae in vitro* culture system is still not available, experimental strategies such as ortholog replacement (or episomal gene expression, as seen in P. vivax studies [50]) can be performed using a *P. knowlesi* model system, as seen in the case of studies into *P. vivax* invasion ligands [51]. *P. malariae* candidate alleles can replace endogenous loci in *P. knowlesi*, allowing *in vitro* assessment of drug susceptibility or other phenotypes such as parasite invasion and development in erythrocytes.

Overall, our study has provided the largest dataset of whole genome data for the neglected *P. malariae* parasite, with insights into genomic diversity in previously uncharacterised populations around the globe. Potential confounding factors of this study include the sample sizes for comparative analyses between African (n = 130) and Asian (n = 17) populations, in addition to the large time frame in which isolates were collected (isolates collected between 2001 and 2020). To account for uneven population sizes, a subsampling approach could have been used where random subsets of the larger population are taken and compared against isolates from Asia and an overall average statistic is given. The large time frame of sample collection could create variation within each population that we have not investigated, due to different selection pressures on parasites between 2001 and 2020 (for example differential drug pressures). This data set could be further investigated to see whether changes in drug regimens have affected parasite populations in each country, and as the *P. malariae* database continues to grow, analysis could be segregated by year

of sample collection to investigate temporal changes on parasite populations. Additionally, we have two different types of clinical isolates in our study, those collected from individuals living in endemic countries, and those from individuals who have travelled to endemic countries and returned to the UK presenting with malaria. It is likely that individuals living in endemic areas are exposed to a greater number of parasite strains than individual travelling to the same country [52], therefore estimates of monoclonality from these two different sample types may be skewed. However, in this study, we did not notice any difference in F_{WS} between travellers and non-travellers.

Although, this is the largest WGS study of *P. malariae* produced so far, further large-scale investigations are needed to provide greater resolution across malaria endemic populations. In lieu of this work, we have provided an invaluable resource that can be exploited to understand the unique biology traits of the *P. malariae* parasite and support functional studies, ultimately informing tools to assist malaria control and elimination.

METHODS

Ethics statement

Isolates from Thailand were obtained with ethical approval from the Mahidol Faculty of Tropical Medicine Ethics Committee (Ref: 2015-001.01). PHE-MRL samples were sourced with NHS UK Ethics approval (#18/LO/0738). In all instances, samples were collected according to relevant guidelines, and informed consent was obtained for all individuals.

Sample collection and processing

For this study, we used a total of 235 clinical isolates from *P. malariae* single-species infections including 229 isolates where we generated novel sequence data, in addition to sequence data from 6 previously published isolates from Malaysia (n = 1), Papua Indonesia (1), Guinea (3) and Mali (1) [13]. Samples were obtained from 31 countries (Algeria 1, Angola 18, Bangladesh 2, Brazil 13, Cameroon 7, Central African Republic 1, Colombia 2, Congo 8, Democratic Republic of the Congo 1, Equatorial Guinea 1, Gabon 17, Ghana 15, Guinea 5, Indonesia 1, Ivory Coast 1, Kenya 13, Liberia 3, Malawi 6, Malaysia 1, Mali 1, Mozambique 1, Nigeria 33, Papua New Guinea 1, Senegal 1, Sierra Leone 10, Sudan 10, Tanzania 26, Thailand 19, The Gambia 1, Uganda 15,

Venezuela 1). All isolates were collected between years 2001 to 2020. Isolates obtained from the PHE-MRL were sourced from individuals who had previously returned to the UK with clinical symptoms of malaria after travelling to one known country with malaria transmission, and were previously screened using both a nested PCR [53] and qPCR [54] for species identification. DNA for all samples was extracted from venous blood samples using the QIAamp DNA Blood Mini Kit (Qiagen) according to manufacturer's instructions. All DNA samples were screened to confirm single-species *P. malariae* infections using a qPCR assay. All samples were quantified using the Qubit 3.0 fluorometer.

Selective Whole Genome Amplification and Sequencing

All DNA samples underwent selective whole genome amplification (SWGA) to increase the relative levels of *P. malariae* DNA within the sample, using the primers: TATGTATA*T*T, TTATTC*G*T, TTCGTT*A*T, TTTTTA*C*G, TATTTC*G*T (asterisk denoting a phosphorothioate bond) as previously described by Ibrahim et al. [27]. Up to 60 ng of DNA was amplified in a 50 µl total reaction alongside 1X Phi29 DNA Polymerase Reaction Buffer (New England Biolabs), 1X BSA (New England Biolabs), 2.5 µM primer mix (Sigma), 1mM dNTP (Roche), Nuclease-Free Water (Ambion, The RNA Company) and 30 units Phi29 DNA Polymerase (New England BioLabs). Reactions were carried out in a UV Cabinet for PCR Operations (UV-B-AR, Grant Bio) to minimise the risk of contamination. Amplified samples were purified using a 1:1 ratio of AMPure XP beads (Beckman-Coulter) following manufacturer's instructions. Sequencing libraries were prepared from clean DNA samples using either the QIAseq FX DNA Library Kit (QIAGEN) (with a 20-minute fragmentation step), or the NEB Next Ultra DNA Library Prep Kit for Illumina (New England Biolabs, E7370) for sequencing on the Illumina MiSeq or Illumina HiSeq4000 (150 bp paired reads), respectively. Sequencing was facilitated by The Applied Genomics Centre, LSHTM.

Bioinformatic analysis

Fastq files outputted from the Illumina sequencing platforms were processed using trimmomatic (v0.39) to remove poor quality sequence data at the ends of reads using the following parameters: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:36. Trimmed reads were then mapped to the PmUG01 reference genome (accessible via https://plasmodb.org) [13] using bwa-mem

software (v0.7.12). The resulting bam files were processed using the samtools (v1.9) functions fixmate and markdup before using GATK's BaseRecalibrator and applyBQSR functions with a training set of 7,415 high quality SNP positions from previous work [27] to produce improved bam files. SNPs and indels were identified using GATK's HaplotypeCaller (v 4.1.4.1) with default parameters using the -ERC GVCF option, to produce a combined VCF file with variants for all isolated within the database, which was filtered to include only SNPs within the core genome (as previously described [27]). The combined VCF file was further filtered to remove SNPs with a Variant Quality Score Log-Odds (VQSLOD) less than zero, in addition to individual SNP positions with missing calls in >10% isolates, and individual samples with missing data at >50% of the SNP positions. Filtering produced a final filtered database comprised of 171 isolates and 221,656 unique SNPs. Isolates were individually assessed using centrifuge software to check for contamination with other species [55], and those with >90% assigned to *P. malariae* were kept for further analysis, removing 16 isolates to leave a database of 155 isolates spanning 25 countries (Africa n = 130, Asia n = 17, South America n = 6, Oceania n = 2), with a total of 143,201 unique SNPs. In some analysis, we aimed to investigate mutations in candidate genes of interest. This was performed on the centrifuge filtered 191 isolates and 1,152,341 SNPs.

Population Genetics

Isolates were grouped at the country level and assessed for multiclonality using the miomix package (v0.0.2.9001) to calculate the F_{WS} score [28]. Only biallelic SNPs within the core genome were used in F_{WS} calculations. Isolates with an F_{WS} score > 0.85 (n = 127) (**S1 Table, S2 Table**) were designated as monoclonal for use in robust population genetic analysis (for homology, IBD and selection analyses), this is a lower threshold than commonly used (usually FWS > 0.95 is used to designate monoclonality in Plasmodium parasites [28]) to conserve the number of isolates available in the dataset. The previously filtered global genomic dataset (155 isolates, 143,201 SNPs, not filtered using F_{WS}) was investigated for population structure, using only biallelic SNPs in the core genome. Principal component analyses (PCAs) were performed on pairwise SNP differences between isolates using the qqman package [56]. Iqtree was used to generate a maximum likelihood tree of all isolates using the 143,201 filtered SNP positions with the GTR+F+R4 substitution model as assigned with ModelFinder. Trees were tested with SH-like aLRT with 1000 replicates, and bootstrapping was performed using UFBoot2 with 900 iterations [57,58]. Proposed

ancestral populations were investigated in the global dataset of 155 isolates (not filtered using F_{WS}), using SNPs with a linkage disequilibrium coefficient ≤ 0.1, using the Admixture package (v1.3.0), with the most likely number of ancestral populations estimated using the cross-validation error score (as calculated using an average from 10 replicates) [59]. Only countries with >5 isolates were visualised in the admixture plot. After visualising genetic separation of parasite populations, we investigated the SNP positions driving differences in allele frequencies between populations using the fixation index (F_{ST}), calculated using the vcftools package [60].

Signals of positive selection within the monoclonal genomic database were calculated using the rehh package (v3.2.1) developed in R [61]. The integrated haplotype score (iHS) was calculated within population and the Rsb score was calculated to identify signals of selection comparing two different populations. Both scores were calculated using monoclonal isolates (F_{WS} > 0.85) with groups of > 10 isolates. Signals of homology at the country and continent level were investigated within monoclonal isolates through screening for identity-by-descent (IBD) using the hmmIBD package (version 2.0.4) [62]. Default parameters were used within hmmIBD, with the recombination rate based on previous studies with *P. falciparum* [62].

TABLES

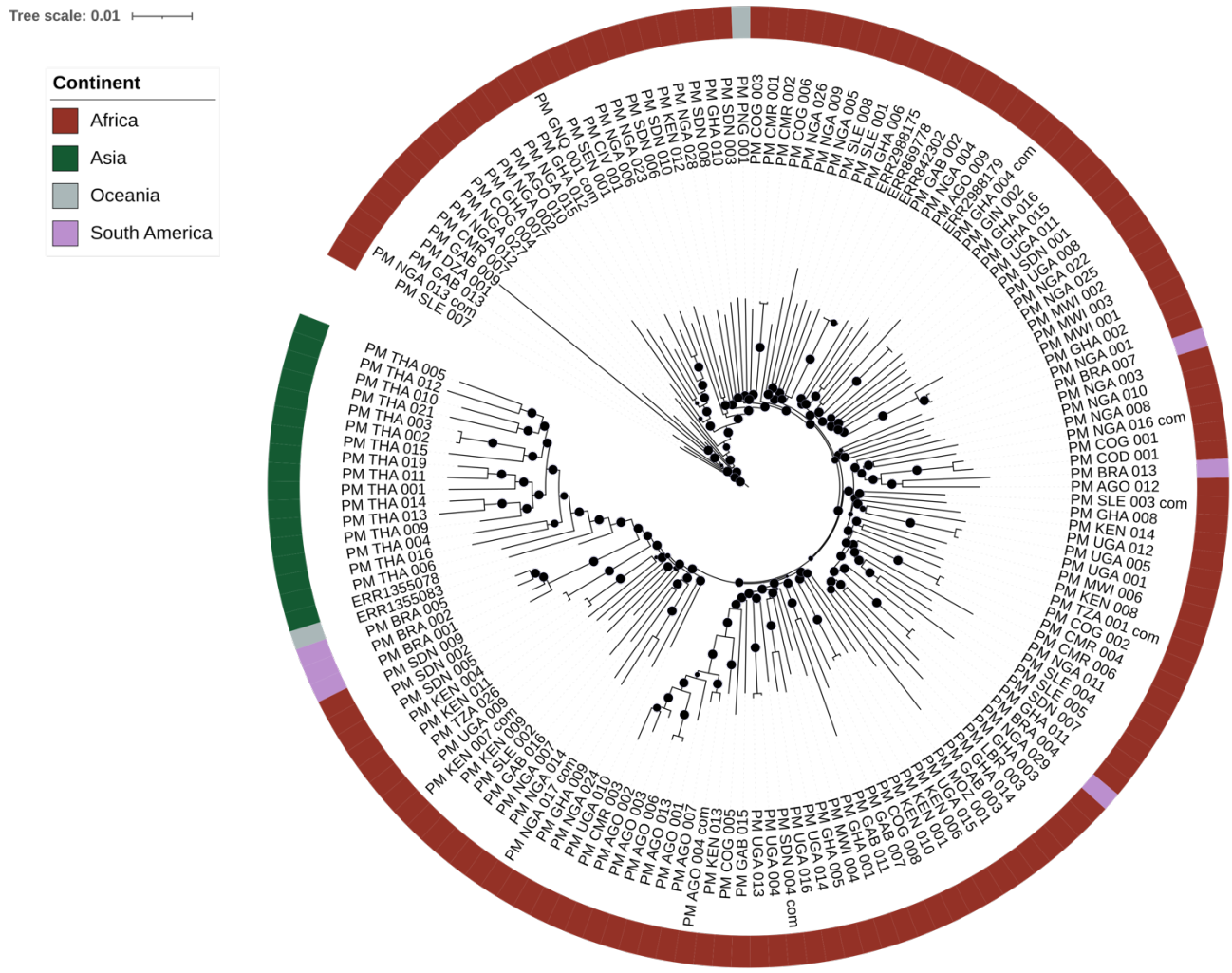
Table 1. Highly differentiating SNPs between African (n = 130) and Asian (n = 17) parasite populations.

Chromosome	Position	FST	Reference	Alternate	Effect	Gene name	Gene ID	Amino acid change	Nucleotide change
1	792521	1	G	C	missense	PmUG01_01025700	PmUG01_01025700	508R>508S	792521G>C
2	515967	1	T	A	missense	NT4	PmUG01_02020400	226N>226K	515967T>A
2	638779	1	G	T	missense	PmUG01_02022700	PmUG01_02022700	2135H>2135N	638779G>T
2	694942	1	G	A	missense	PmUG01_02024100	PmUG01_02024100	465G>465E	694942G>A
3	104471	1	A	T	missense	PmUG01_03012000	PmUG01_03012000	119I>119N	104471A>T
3	551583	1	G	T	missense	PmUG01_03020800	PmUG01_03020800	909W>909L	551583G>T
3	552153	1	A	C	missense	PmUG01_03020800	PmUG01_03020800	1099N>1099T	552153A>C
4	256898	1	A	C	missense	PmUG01_04014200	PmUG01_04014200	924K>924Q	256898A>C
4	317346	1	A	T	missense	ATG11	PmUG01_04015700	1316Y>1316N	317346A>T
4	811445	1	T	C	missense	PmUG01_04025500	PmUG01_04025500	1266D>1266G	811445T>C
4	823561	1	C	G	missense	PmUG01_04025700	PmUG01_04025700	235N>235K	823561C>G
5	436269	1	T	A	missense	GAMA	PmUG01_05017200	254L>254M	436269T>A
6	363359	1	C	G	missense	PRP5	PmUG01_06016200	334G>334R	363359C>G
6	408635	1	G	C	missense	PmUG01_06017100	PmUG01_06017100	257S>257W	408635G>C
6	419801	1	A	C	missense	CAF40	PmUG01_06017300	569I>569S	419801A>C
7	637290	1	C	T	missense	PmUG01_07021200	PmUG01_07021200	565R>565K	637290C>T
7	660470	1	T	C	missense	SET4	PmUG01_07021600	771N>771S	660470T>C
7	729550	1	T	C	missense	CRMP1	PmUG01_07022900	1963S>1963P	729550T>C
7	731085	1	G	A	missense	CRMP1	PmUG01_07022900	2474M>2474I	731085G>A
7	776253	1	C	T	missense	PmUG01_07024000	PmUG01_07024000	229A>229V	776253C>T
7	789061	1	A	G	missense	RPN8	PmUG01_07024400	114K>114R	789061A>G
7	835623	1	C	G	missense	PmUG01_07025700	PmUG01_07025700	196V>196L	835623C>G
8	1122265	1	G	A	missense	RIPR	PmUG01_08032700	1018D>1018N	1122265G>A
8	1122538	1	A	G	missense	RIPR	PmUG01_08032700	1109R>1109G	1122538A>G
8	1229600	1	C	T	missense	PmUG01_08034700	PmUG01_08034700	22D>22N	1229600C>T
8	1394343	1	G	A	missense	PmUG01_08037800	PmUG01_08037800	192D>192N	1394343G>A
8	1755210	1	C	T	missense	PmUG01_08045200	PmUG01_08045200	131G>131E	1755210C>T
8	1755319	1	G	T	missense	PmUG01_08045200	PmUG01_08045200	95Q>95K	1755319G>T
8	1755323	1	G	T	missense	PmUG01_08045200	PmUG01_08045200	93N>93K	1755323G>T
8	1757220	1	G	A	missense	PmUG01_08045300	PmUG01_08045300	772R>772W	1757220G>A
8	2024675	1	C	T	missense	PmUG01_08051100	PmUG01_08051100	2682S>2682L	2024675C>T
8	2049708	1	A	C	missense	CSP	PmUG01_08051600	283K>283Q	2049708A>C
8	2049745	1	A	G	missense	CSP	PmUG01_08051600	295E>295G	2049745A>G
9	402441	1	G	A	missense	ApiAP2	PmUG01_09017400	1084P>1084S	402441G>A
9	672797	1	C	G	missense	PmUG01_09024200	PmUG01_09024200	86A>86G	672797C>G
9	673072	1	C	A	missense	PmUG01_09024200	PmUG01_09024200	178P>178T	673072C>A
9	673075	1	A	T	missense	PmUG01_09024200	PmUG01_09024200	179K>179F	673075A>T+673076A>T+
9	673076	1	A	T	missense	PmUG01_09024200	PmUG01_09024200	179K>179F	673075A>T+673076A>T+
9	673077	1	A	T	missense	PmUG01_09024200	PmUG01_09024200	179K>179F	673075A>T+673076A>T+
9	673603	1	T	A	missense	PmUG01_09024200	PmUG01_09024200	355Y>355N	673603T>A
9	1004676	1	G	A	missense	PmUG01_09031400	PmUG01_09031400	192H>192Y	1004676G>A
9	1072891	1	G	A	missense	CDPK7	PmUG01_09032300	637R>637K	1072891G>A
9	1073347	1	A	T	missense	CDPK7	PmUG01_09032300	789N>789I	1073347A>T
9	1821204	1	T	A	missense	PmUG01_09048100	PmUG01_09048100	1439K>1439N	1821204T>A
9	2061398	1	A	G	missense	PmUG01_09052100	PmUG01_09052100	78H>78R	2061398A>G
9	2250043	1	A	T	missense	PmUG01_09056100	PmUG01_09056100	1170I>1170L	2250043A>T
9	2294640	1	G	C	missense	PmUG01_09057000	PmUG01_09057000	421N>421K	2294640G>C
10	562403	1	T	A	missense	PmUG01_10022200	PmUG01_10022200	6349N>6349I	562403T>A
10	566074	1	T	C	missense	PmUG01_10022200	PmUG01_10022200	5125I>5125M	566074T>C
11	811956	1	A	C	missense	GCbeta	PmUG01_11025500	2606I>2606R	811956A>C
11	814323	1	G	T	missense	GCbeta	PmUG01_11025500	1817A>1817E	814323G>T
11	938450	1	C	T	missense	PmUG01_11028000	PmUG01_11028000	136T>136M	938450C>T
11	1290491	1	T	G	missense	PK4	PmUG01_11035400	1259N>1259T	1290491T>G
11	1443377	1	T	C	missense	SMS1	PmUG01_11038600	108I>108V	1443377T>C
11	1510793	1	A	T	missense	TKL4	PmUG01_11039800	2181K>2181I	1510793A>T
11	2018425	1	G	A	missense	P12	PmUG01_11050500	127E>127K	2018425G>A+2018427A>
12	1026365	1	T	C	missense	PmUG01_12032400	PmUG01_12032400	412K>412R	1026365T>C
12	1226733	1	A	C	missense	PmUG01_12036500	PmUG01_12036500	1329E>1329D	1226733A>C
12	1250275	1	G	A	missense	PmUG01_12037000	PmUG01_12037000	679A>679V	1250275G>A
12	1252779	1	C	T	missense	PmUG01_12037000	PmUG01_12037000	3C>3Y	1252779C>T
12	2092729	1	C	A	missense	PmUG01_12054500	PmUG01_12054500	2022S>2022R	2092729C>A
12	2098884	1	A	T	missense	PmUG01_12054500	PmUG01_12054500	4074N>4074I	2098884A>T
12	2338091	1	C	T	missense	PmUG01_12060500	PmUG01_12060500	9T>9I	2338091C>T
12	2543395	1	C	T	missense	RON2	PmUG01_12065000	1728R>1728K	2543395C>T
12	2937163	1	G	C	missense	PARN	PmUG01_12073500	117G>117R	2937163G>C
12	3071772	1	T	G	missense	MFS6	PmUG01_12076200	265N>265H	3071772T>G
12	3112832	1	G	C	missense	PmUG01_12077000	PmUG01_12077000	79G>79R	3112832G>C
13	2293168	1	T	G	missense	PmUG01_13054900	PmUG01_13054900	554K>554N	2293168T>G
14	403332	1	G	A	missense	PmUG01_14016200	PmUG01_14016200	89S>89N	403332G>A
14	1615408	1	T	G	missense	PmUG01_14044700	PmUG01_14044700	382I>382S	1615408T>G
14	1617405	1	A	G	missense	PmUG01_14044700	PmUG01_14044700	1048N>1048D	1617405A>G
14	1632300	1	T	A	missense	CAF1	PmUG01_14045000	1247I>1247F	1632300T>A
14	1925559	1	C	T	missense	PmUG01_14053000	PmUG01_14053000	2335A>2335T	1925559C>T
14	2110963	1	T	G	missense	ApiAP2	PmUG01_14056700	154H>154Q	2110963T>G
14	2136305	1	A	C	missense	AP2-G	PmUG01_14056900	1916S>1916R	2136305A>C
14	2439329	1	A	T	missense	MSP9	PmUG01_14062900	425V>425D	2439329A>T
14	2440518	1	T	C	missense	MSP9	PmUG01_14062900	29T>29A	2440518T>C
14	2477548	1	A	T	missense	MRP2	PmUG01_14063400	1215L>1215I	2477548A>T
14	2814078	1	G	T	missense	NOT4	PmUG01_14069500	782A>782S	2814078G>T
14	3156801	1	T	C	missense	DOC2	PmUG01_14077100	1213E>1213G	3156801T>C
14	3164993	1	T	C	missense	PmUG01_14077200	PmUG01_14077200	578D>578G	3164993T>C
14	3189891	1	T	G	missense	PmUG01_14077700	PmUG01_14077700	568K>568Q	3189891T>G
14	3191473	1	T	A	missense	PmUG01_14077700	PmUG01_14077700	40R>40S	3191473T>A
14	3362924	1	T	A	missense	PmUG01_14081900	PmUG01_14081900	2076N>2076K	3362924T>A

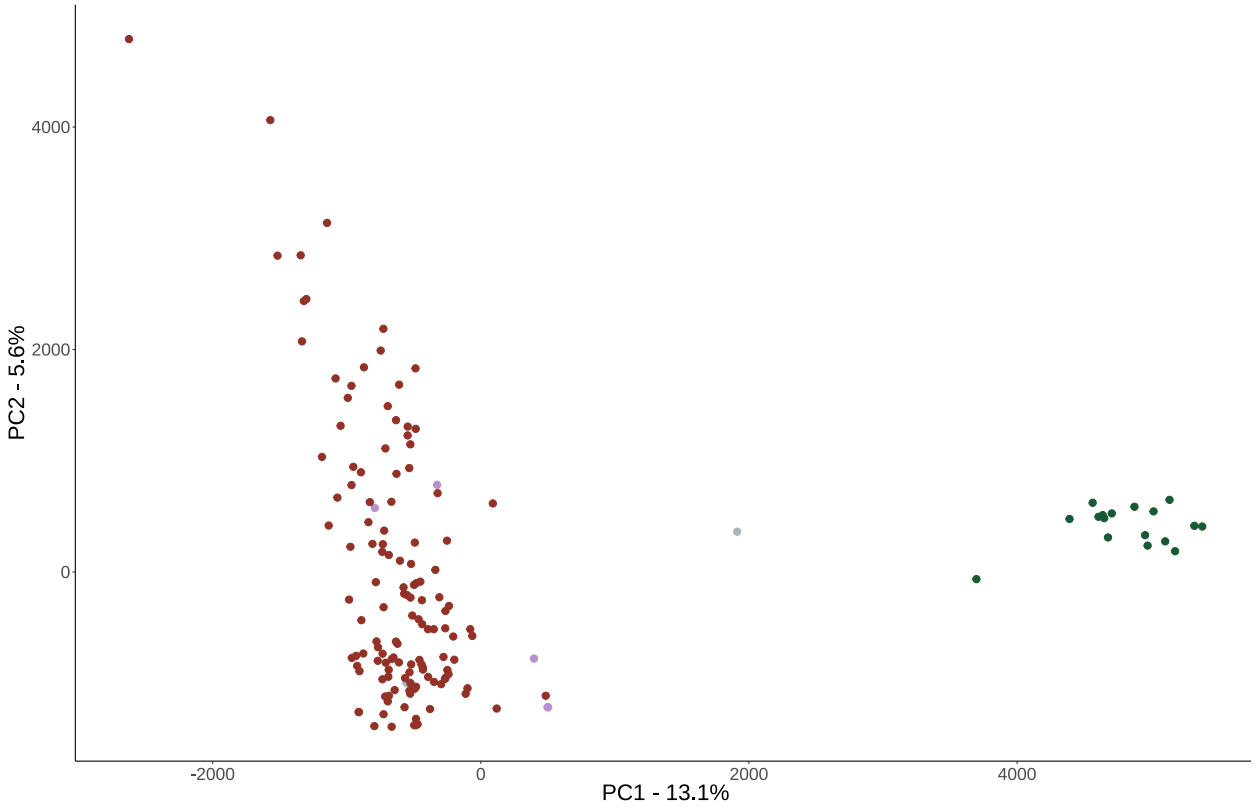
Table 2. Frequent SNPs found within orthologs of genes associated with drug susceptibility in the *P. malariae*

Position	Reference	Effect	Gene name	Gene ID	Amino acid	Nucleotide 1	Nucleotide 2	Global SNP Freq (n = 195)	Africa Freq SNPs (n = 163)	Asia Freq SNPs (n = 22)	Oceania Freq SNPs (n = 2)	SAM Freq SNPs (n = 8)
553440	A	missense	CRT	PmUG01_01020700	278E>278D	553440A>T		0.06	0.01	0.37	0	0
554652	G	missense	CRT	PmUG01_01020700	383G>383V	554652G>T		0.02	0.01	0	0	0.67
109509	C	missense	MRP1	PmUG01_02011900	1630D>1630N	109509C>T		0.72	0.7	0.76	1	1
112215	A	missense	MRP1	PmUG01_02011900	728S>728A	112215A>C		0.02	0	0.15	0	0
446956	T	missense	UBP1	PmUG01_02019300	3405D>3405G	446956T>C		0.04	0.04	0	0	0.5
450665	A	missense	UBP1	PmUG01_02019300	2307I>2307S	450665A>C		1	1	1	1	1
452184	A	missense	UBP1	PmUG01_02019300	1801L>1801V	452184A>C		0.08	0.11	0	0	0
452198	C	missense	UBP1	PmUG01_02019300	1796G>1796E	452198C>T		0.06	0.01	0.62	0	0
452210	C	missense	UBP1	PmUG01_02019300	1792G>1792E	452210C>T		0.04	0	0.4	0	0
452792	T	missense	UBP1	PmUG01_02019300	1598N>1598S	452792T>C		0.44	0.38	0.93	0	0.2
452841	G	missense	UBP1	PmUG01_02019300	1582H>1582Y	452841G>A		0.02	0	0.14	0	0
454431	G	missense	UBP1	PmUG01_02019300	1052H>1052E	454429A>T	454431G>C	0.1	0.12	0	0	0
456030	T	missense	UBP1	PmUG01_02019300	519I>519V	456030T>C		0.05	0.01	0.29	0	0
1291895	G	missense	DHFR-TS	PmUG01_05034700	15A>15S	1291895G>T		0.05	0.04	0	0	0.33
1292023	C	missense	DHFR-TS	PmUG01_05034700	57F>57L	1292023C>G	1292023C>A	0.17	0.21	0	0	0
1292193	A	missense	DHFR-TS	PmUG01_05034700	114N>114S	1292193A>G		0.71	0.67	0.84	1	1
297598	C	missense	PI4K	PmUG01_06015100	1087E>1087K	297598C>T		0.2	0.19	0.36	0	0
300076	C	missense	PI4K	PmUG01_06015100	261A>261S	300074T>A	300076C>A	0.43	0.47	0.3	0	0
300093	C	missense	PI4K	PmUG01_06015100	255G>255D	300092T>G	300093C>T	0.12	0.12	0.11	0	0
300135	C	missense	PI4K	PmUG01_06015100	241G>241D	300134T>G	300135C>T	0.29	0.3	0.25	0	0
525728	T	missense	MDR1	PmUG01_10021600	1063L>1063F	525728T>G		0.01	0	0.1	0	0
527449	G	missense	MDR1	PmUG01_10021600	490L>490I	527449G>T		0.01	0	0.05	0.5	0
1140672	C	missense	Kelch10	PmUG01_10033800	339A>339V	1140672C>T		0.01	0	0	0	0.17
2740223	T	missense	MDR2	PmUG01_12069100	426V>426A	2740223T>C		0.7	0.72	0.72	0.5	0.5
2740398	A	missense	MDR2	PmUG01_12069100	484L>484F	2740398A>T		0.07	0.08	0	1	0
2741992	G	missense	MDR2	PmUG01_12069100	1016V>1016F	2741992G>T		0.47	0.52	0.1	0	0.8
2742029	T	missense	MDR2	PmUG01_12069100	1028V>1028A	2742029T>C		1	1	1	1	1
2742326	A	missense	MDR2	PmUG01_12069100	1127N>1127S	2742326A>G		0.27	0.21	0.74	0	0
2742583	A	missense	MDR2	PmUG01_12069100	1213N>1213Y	2742583A>T		0.03	0	0.25	0	0
2742773	G	missense	MDR2	PmUG01_12069100	1276S>1276N	2742773G>A		0.01	0	0.11	0	0
2742775	G	missense	MDR2	PmUG01_12069100	1277G>1277S	2742775G>A		0.01	0	0.11	0	0
2742779	A	missense	MDR2	PmUG01_12069100	1278N>1278S	2742779A>G		0.01	0	0.11	0	0
666353	C	missense	ATP4	PmUG01_13021900	729E>729K	666353C>T		0.01	0	0	0	0.17
666354	A	missense	ATP4	PmUG01_13021900	728N>728K	666354A>T		0.01	0	0	0	0.17
666359	C	missense	ATP4	PmUG01_13021900	727E>727K	666359C>T		0.01	0	0	0	0.17
666360	A	missense	ATP4	PmUG01_13021900	726N>726K	666360A>T		0.01	0	0	0	0.17
668505	A	missense	ATP4	PmUG01_13021900	11N>11K	668505A>T		0.07	0	0.52	0.5	0
1656941	G	missense	PPPK-DHPS	PmUG01_14045500	375A>375G	1656941G>C		0.13	0.15	0	0	0.17
1657487	C	missense	PPPK-DHPS	PmUG01_14045500	193R>193I	1657487C>A		0.13	0.16	0	0	0
1936646	A	missense	AP2-MU	PmUG01_14053100	299N>299Y	1936646A>T		0.08	0	0.7	0	0
2475164	A	missense	MRP2	PmUG01_14063400	2009I>2009M	2475164A>C		0.09	0.11	0	0	0
2475264	G	missense	MRP2	PmUG01_14063400	1976T>1976I	2475264G>A		0.1	0	0.7	0	0
2475649	T	missense	MRP2	PmUG01_14063400	1848T>1848A	2475649T>C		0.13	0.16	0	0	0
2475782	A	missense	MRP2	PmUG01_14063400	1803D>1803E	2475782A>T	2475782A>C	0.13	0.03	0.88	0	0
2476635	G	missense	MRP2	PmUG01_14063400	1519A>1519V	2476635G>A		0.52	0.46	0.94	1	0.5
2477548	A	missense	MRP2	PmUG01_14063400	1215L>1215I	2477548A>T		0.09	0	0.81	0	0
2479063	A	missense	MRP2	PmUG01_14063400	710Y>710N	2479063A>T		0.05	0	0.43	0	0
2479155	G	missense	MRP2	PmUG01_14063400	679T>679S	2479155G>C		0.64	0.58	1	1	0.67
2479300	C	missense	MRP2	PmUG01_14063400	631V>631I	2479300C>T		0.67	0.65	0.88	1	0

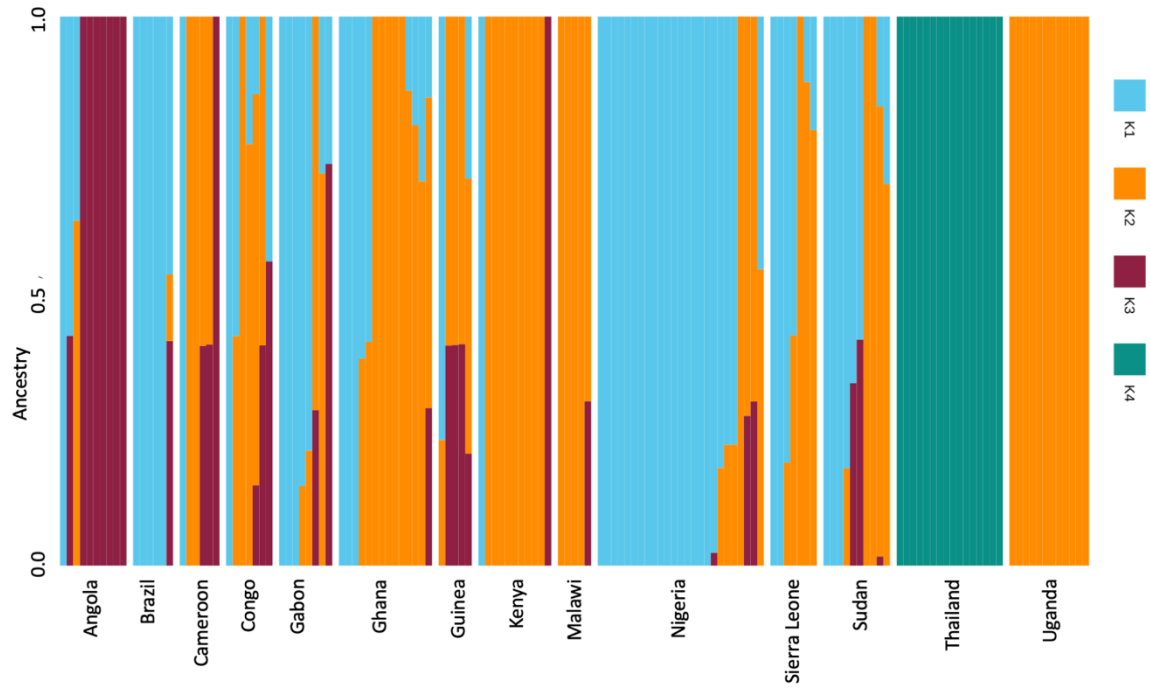
A)


Tree scale: 0.01
Continent
Africa
Asia
Oceania
South America

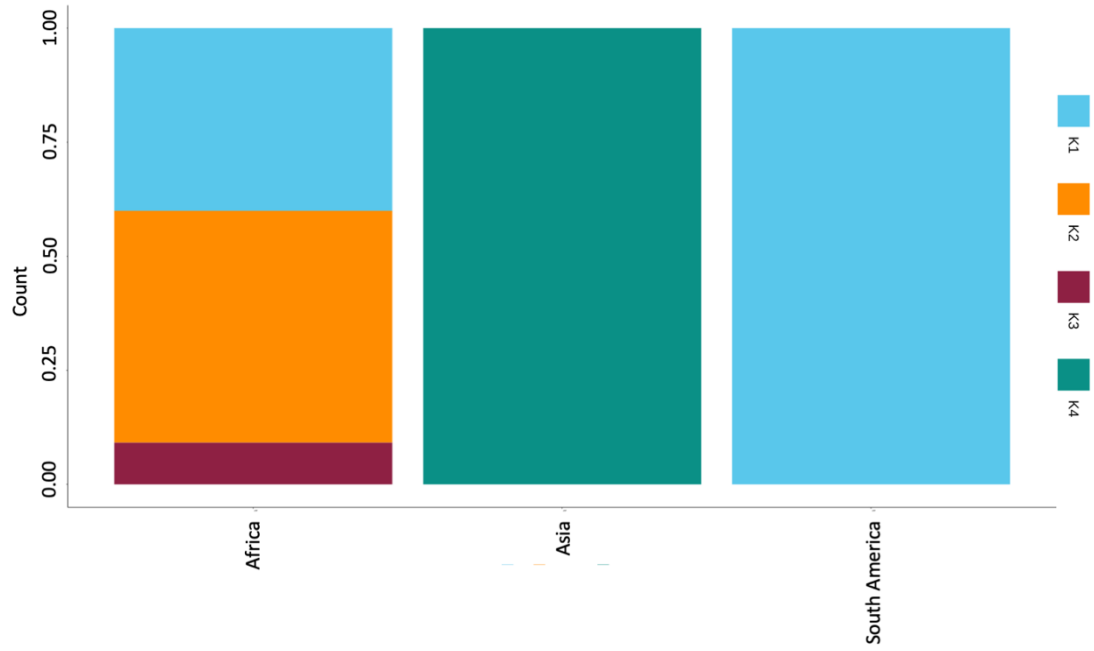
B)



C)


Ancestry
0.0
0.5
1.0
Angola
Brazil
Cameroon
Congo
Gabon
Ghana
Guinea
Kenya
Malawi
Nigeria
Sierra Leone
Sudan
Thailand
Uganda
K1
K2
K3
K4

D)


Count
0.00
0.25
0.50
0.75
1.00
Africa
Asia
South America
K1
K2
K3
K4

FIGURES

Figure 1

Global population structure of *P. malariae* isolates shows continent level separation.

(A) Maximum-likelihood tree of high-quality SNPs from 155 filtered isolates calculated using iqtree [57,58]. Bootstrapping scores >50 are highlighted on the branches by a black circle, and isolates are annotated with their continent of origin (in the inner circular track); **(B)** Principal Components Analysis (PCA) plot of the 155 isolates, with colours based on continent as in (**A**); **(C)** Admixture prediction subpopulation per sample, per country visualised using a bar plot, only including countries with >5 isolates; **(D)** Admixture bar plot illustrating the population structure (*K*) only including countries with >5 isolates, highlighting a distinction between Asian countries, Africa and South America.

REFERENCES

1. Rutledge, G. G. *et al*. Genomic characterization of recrudescent plasmodium malariae after treatment with artemether/lumefantrine. *Emerging Infectious Diseases* (2017). doi:10.3201/eid2308.161582
2. Sutherland, C. J. Persistent Parasitism: The Adaptive Biology of Malariae and Ovale Malaria. *Trends Parasitol*. **32**, 808–819 (2016).
3. Grande, R., Antinori, S., Meroni, L., Menegon, M. & Severini, C. CASE REPORT A case of Plasmodium malariae recurrence : recrudescence or reinfection ? *Malar. J*. 1–9 (2019). doi:10.1186/s12936-019-2806-y
4. Betson, M., Clifford, S., Stanton, M., Kabatereine, N. B. & Stothard, J. R. Emergence of Nonfalciparum Plasmodium Infection Despite Regular Artemisinin Combination Therapy in an 18-Month Longitudinal Study of Ugandan Children and Their Mothers. **217**, 1099–1109 (2018).
5. Hendrickse, R. G. & Adeniyi, A. Quartan malarial nephrotic syndrome in children. *Kidney Int*. **16**, 64–74 (1979).
6. Gilles, H. M. & Hendrickse, R. G. Nephrosis in Nigerian children. Role of Plasmodium malariae, and effect of antimalarial treatment. *Br. Med. J*. **2**, 27–31 (1963).
7. Langford, S. *et al*. Plasmodium malariae Infection Associated with a High Burden of Anemia: A Hospital-Based Surveillance Study. *PLoS Negl. Trop. Dis*. (2015). doi:10.1371/journal.pntd.0004195
8. Douglas, N. M. *et al*. Major burden of severe anemia from non-falciparum malaria species in Southern Papua: a hospital-based surveillance study. *PLoS Med*. **10**, 1–17 (2013).
9. Collins, W. E. & Jeffery, G. M. Plasmodium malariae: Parasite and disease. *Clinical Microbiology Reviews* (2007). doi:10.1128/CMR.00027-07
10. Abdulraheem, M. A., Ernest, M., Ugwuanyi, I. & Abkallo, H. M. High prevalence of Plasmodium malariae and Plasmodium ovale in co-infections with Plasmodium falciparum in asymptomatic malaria parasite carriers in Southwest Nigeria Malaria Unit , Department of Pathology , Institute of Tropical Medicine , Nagasaki Unive. *medRxiv Prepr*. (2021).
11. Niño, C. H. *et al*. Plasmodium malariae in the Colombian Amazon region: you don't diagnose what you don't suspect. *Malar. J*. (2016). doi:10.1186/s12936-016-1629-3
12. Zhou, M. *et al*. High prevalence of Plasmodium malariae and Plasmodium ovale in malaria patients along the Thai-Myanmar border, as revealed by acridine orange staining and PCR-based diagnoses. *Trop. Med. Int. Health* **3**, 304–12 (1998).
13. Rutledge, G. G. *et al*. Plasmodium malariae and P. ovale genomes provide insights into malaria parasite evolution. *Nature* **542**, 101–104 (2017).
14. Scopel, K. K. G., Fontes, C. J. F., Nunes, Á. C., Horta, M. F. & Braga, É. M. High prevalence of Plamodium malariae infections in a Brazilian Amazon endemic area (Apiacás - Mato Grosso State) as detected by polymerase chain reaction. *Acta Trop*. **90**, 61–64 (2004).
15. Islam, S. & Hai, F. Recrudescing Plasmodium malariae infection despite appropriate treatment in an immigrant toddler. *Paediatr Int Child Heal*. **38**, 290–293 (2018).
16. Betson, M. *et al*. Detection of persistent Plasmodium spp. infections in Ugandan children after artemether-lumefantrine treatment. *Parasitology* **141**, 1880–1890 (2014).

17. Dinko, B., Oguike, M. C., Larbi, J. A., Bousema, T. & Sutherland, C. J. Persistent detection of Plasmodium falciparum, P. malariae, P. ovale curtisi and P. ovale wallikeri after ACT treatment of asymptomatic Ghanaian school-children. *Int. J. Parasitol. Drugs Drug Resist.* **3**, 45–50 (2013).
18. Recurrence of Plasmodium malariae and P. falciparum following treatment of uncomplicated malaria in North Sumatera with dihydroartemisinin-piperaquine or artemether-lumefantrine | Open Forum Infectious Diseases | Oxford Academic.
19. Maguire, J. D. *et al*. Chloroquine-resistant Plasmodium malariae in south Sumatra, Indonesia. *Lancet* (2002). doi:10.1016/S0140-6736(02)09336-4
20. Id, V. Y. *et al*. Persistent transmission of Plasmodium malariae and Plasmodium ovale species in an area of declining Plasmodium falciparum transmission in eastern Tanzania. 1–16 (2019).
21. Pearson, R. D. R. *et al*. Genomic analysis of local variation and recent evolution in Plasmodium vivax. *Nat. Genet.* **48**, 959–964 (2016).
22. Benavente, E. D. *et al*. Distinctive genetic structure and selection patterns in Plasmodium vivax from South Asia and East Africa. *Nat. Commun. 2021 121* **12**, 1–11 (2021).
23. Miotto, O. *et al*. Multiple populations of artemisinin-resistant Plasmodium falciparum in Cambodia. *Nat. Genet.* **45**, 648–55 (2013).
24. Gen, M. *et al*. An open dataset of *Plasmodium falciparum* genome variation in 7,000 worldwide samples. *Wellcome Open Res. 2021 642* **6**, 42 (2021).
25. Neafsey, D. E. *et al*. The malaria parasite Plasmodium vivax exhibits greater genetic diversity than Plasmodium falciparum. *Nat. Genet.* **44**, 1046–50 (2012).
26. Ansari, H. R. *et al*. Genome-scale comparison of expanded gene families in Plasmodium ovale wallikeri and Plasmodium ovale curtisi with Plasmodium malariae and with other Plasmodium species. *Int. J. Parasitol.* (2016). doi:10.1016/j.ijpara.2016.05.009
27. Ibrahim, A. *et al*. Selective whole genome amplification of Plasmodium malariae DNA from clinical samples reveals insights into population structure. *Sci. Rep.* **10**, 1–11 (2020).
28. Manske, M. *et al*. Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. *Nature* **487**, 375–379 (2012).
29. Yalcindag, E. *et al*. Multiple independent introductions of Plasmodium falciparum in South America. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 511–516 (2012).
30. Laurens, M. B. RTS,S/AS01 vaccine (Mosquirix™): an overview. *Hum. Vaccin. Immunother.* **16**, 480 (2020).
31. Gil, J. P. & Fançony, C. Plasmodium falciparum Multidrug Resistance Proteins (pfMRPs). *Front. Pharmacol.* **12**, 2896 (2021).
32. Thompson, J. *et al*. Plasmodium cysteine repeat modular proteins 1-4: complex proteins with roles throughout the malaria parasite life cycle. *Cell. Microbiol.* **9**, 1466–1480 (2007).
33. Josling, G. A. *et al*. A Plasmodium Falciparum Bromodomain Protein Regulates Invasion Gene Expression. *Cell Host Microbe* **17**, 741–751 (2015).
34. Lim, M. Y. X. *et al*. UDP-galactose and acetyl-CoA transporters as Plasmodium multidrug resistance genes. *Nat. Microbiol.* **1**, (2016).
35. Biddau, M. *et al*. Plasmodium falciparum LipB mutants display altered redox and carbon metabolism in asexual stages and cannot complete sporogony in Anopheles mosquitoes. *Int. J. Parasitol.* **51**, 441–453 (2021).
36. Jain, S. *et al*. The prokaryotic ClpQ protease plays a key role in growth and development

of mitochondria in Plasmodium falciparum. *Cell. Microbiol.* **15**, 1660–1673 (2013).
37. Koenderink, J. B., Kavishe, R. A., Rijpma, S. R. & Russel, F. G. M. The ABCs of multidrug resistance in malaria. *Trends Parasitol.* **26**, 440–446 (2010).
38. Ngwa, C. J., Farrukh, A. & Pradel, G. Zinc finger proteins of Plasmodium falciparum. (2021). doi:10.1111/cmi.13387
39. Deshmukh, A. *et al*. Plasmodium falciparum MSP3 Exists in a Complex on the Merozoite Surface and Generates Antibody Response during Natural Infection. (2018). doi:10.1128/IAI.00067-18
40. Zhang, M. *et al*. Inhibiting the Plasmodium eIF2α Kinase PK4 Prevents Artemisinin-Induced Latency. *Cell Host Microbe* **22**, 766-776.e4 (2017).
41. Oberli, A. *et al*. A Plasmodium falciparum PHIST protein binds the virulence factor PfEMP1 and comigrates to knobs on the host cell surface. *FASEB J.* **28**, 4420 (2014).
42. Gao, H. *et al*. ISP1-Anchored Polarization of GCb/CDC50A Complex Initiates Malaria Ookinete Gliding Motility Article ISP1-Anchored Polarization of GCb/CDC50A Complex Initiates Malaria Ookinete Gliding Motility. *Curr. Biol.* **28**, 2763-2776.e6 (2018).
43. Lacroix, C. & Ménard, R. TRAP-like protein of Plasmodium sporozoites: linking gliding motility to host-cell traversal. *Trends Parasitol.* **24**, 431–434 (2008).
44. Tanomsing, N. *et al*. Genetic analysis of the dihydrofolate reductase-thymidylate synthase gene from geographically diverse isolates of Plasmodium malariae. *Antimicrob. Agents Chemother.* **51**, 3523–3530 (2007).
45. Henrici, R. C., van Schalkwyk, D. A. & Sutherland, C. J. Modification of pfap2μ and pfubp1 Markedly Reduces Ring-Stage Susceptibility of Plasmodium falciparum to Artemisinin In Vitro. *Antimicrob. Agents Chemother.* **64**, (2019).
46. Tan, L. L., Lau, T. Y., Timothy, W. & Prabakaran, D. Full-length sequence analysis of chloroquine resistance transporter gene in plasmodium falciparum isolates from Sabah, Malaysia. *Sci. World J.* **2014**, (2014).
47. McNamara, C. W. *et al*. Targeting Plasmodium phosphatidylinositol 4-kinase to eliminate malaria Europe PMC Funders Group. *Nature* **504**, 248–253 (2013).
48. Ariey, F. *et al*. A molecular marker of artemisinin-resistant Plasmodium falciparum. *Nature* **505**, 50–55 (2016).
49. Demas, A. R. *et al*. Mutations in Plasmodium falciparum actin-binding protein coronin confer reduced artemisinin susceptibility. *Proc. Natl. Acad. Sci.* **115**, 12799–12804 (2018).
50. 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**, 1–23 (2019).
51. 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).
52. Diez Benavente, E. *et al*. Distinctive genetic structure and selection patterns in Plasmodium vivax from South Asia and East Africa - under review - to update reference. *Nat. Commun.* **XX**, (2021).
53. Singh, B. *et al*. A genus- and species-specific nested polymerase chain reaction malaria detection assay for epidemiologic studies. *Am. J. Trop. Med. Hyg.* **60**, 687–692 (1999).
54. Shokoples, S. E., Ndao, M., Kowalewska-Grochowska, K. & Yanow, S. K. Multiplexed real-time PCR assay for discrimination of Plasmodium species with improved sensitivity for mixed infections. *J. Clin. Microbiol.* (2009). doi:10.1128/JCM.01858-08

55. Kim, D., Song, L., Breitwieser, F. P. & Salzberg, S. L. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res*. **26**, 1721–1729 (2016).
56. Turner, S. D. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open Source Softw*. **3**, 731 (2018).
57. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol*. **32**, 268–274 (2015).
58. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular biology and evolution. *Mol. Biol. Evol*. **35**, 518–522 (2018).
59. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. **19**, 1655–1664 (2009).
60. Danecek, P. *et al*. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
61. 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).
62. 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).

SUPPLEMENTARY INFORMATION

First global genomic analysis of *P. malariae* parasites reveals genetic diversity and key mutations in drug targets

Short title: Global genomic diversity of *Plasmodium malariae*

Tables

S1 Table. Complete *P. malariae* metadata

Sample ID	Continent	Country	State	Year	FWS	Centrifuge Pm	Pipeline QC Pass	Final database	Monoclonal database	Total reads	Proportion reads mapped to PmUG01	% Genome covered >5 reads
ERR1355078	Asia	Malaysia			0.00	0.99995	TRUE	TRUE	FALSE	2528847	0.89	91.21
ERR1355083	Oceania	Papua Indonesia			0.00	0.99928	TRUE	TRUE	FALSE	384031	0.53	1.34
ERR2988175	Africa	Guinea		2013	0.99	0.99979	TRUE	TRUE	TRUE	3857472	0.54	86.87
ERR2988179	Africa	Mali	Faladje		0.00	0.99964	TRUE	TRUE	FALSE	4137231	0.93	96.80
ERR842302	Africa	Guinea		2013	0.99	0.99980	TRUE	TRUE	TRUE	3857364	0.51	82.68
ERR869778	Africa	Guinea		2013	0.99	0.99979	TRUE	TRUE	TRUE	6209191	0.51	94.10
PM_AGO_001	Africa	Angola			0.99	0.99710	TRUE	TRUE	TRUE	7049329	0.68	85.84
PM_AGO_002	Africa	Angola			0.99	0.99173	TRUE	TRUE	TRUE	6831072	0.43	65.83
PM_AGO_003	Africa	Angola			0.98	0.99722	TRUE	TRUE	TRUE	6606859	0.67	86.37
PM_AGO_004_com	Africa	Angola		2014	0.99	0.96013	TRUE	TRUE	TRUE	9648872	0.18	39.82
PM_AGO_006	Africa	Angola		2001	1.00	0.99979	TRUE	TRUE	TRUE	6965388	0.88	37.62
PM_AGO_007	Africa	Angola		2005	0.99	0.99926	TRUE	TRUE	TRUE	5446011	0.82	80.52
PM_AGO_009	Africa	Angola		2019	1.00	0.99792	TRUE	TRUE	TRUE	4449897	0.69	70.37
PM_AGO_010	Africa	Angola		2019	0.80	0.99879	TRUE	TRUE	FALSE	6011635	0.78	87.36
PM_AGO_012	Africa	Angola		2017	0.90	0.99973	TRUE	TRUE	TRUE	11504300	0.88	88.14
PM_AGO_013	Africa	Angola		2017	1.00	0.99901	TRUE	TRUE	TRUE	10955507	0.85	67.46
PM_BRA_001	South_America	Brazil	Sao Paulo		1.00	0.98889	TRUE	TRUE	TRUE	3283135	0.33	23.06
PM_BRA_002	South_America	Brazil	Sao Paulo		0.99	0.99885	TRUE	TRUE	TRUE	4191221	0.78	68.31
PM_BRA_004	South_America	Brazil	Sao Paulo		1.00	0.99887	TRUE	TRUE	TRUE	5927555	0.73	69.40
PM_BRA_005	South_America	Brazil	Sao Paulo		0.99	0.99915	TRUE	TRUE	TRUE	4928608	0.82	75.17
PM_BRA_007	South_America	Brazil	Sao Paulo		0.99	0.99513	TRUE	TRUE	TRUE	5389919	0.47	49.19
PM_BRA_013	South_America	Brazil	Santa Catarina	2019	0.80	0.99933	TRUE	TRUE	FALSE	4011831	0.82	64.17
PM_CIV_001	Africa	Ivory_Coast		2015	0.00	0.98976	TRUE	TRUE	FALSE	6713310	0.39	66.07
PM_CMR_001	Africa	Cameroon			1.00	0.99603	TRUE	TRUE	TRUE	3827977	0.66	60.46
PM_CMR_002	Africa	Cameroon			0.98	0.99856	TRUE	TRUE	TRUE	4901480	0.84	87.87
PM_CMR_003	Africa	Cameroon			0.98	0.99806	TRUE	TRUE	TRUE	4165541	0.70	52.38
PM_CMR_004	Africa	Cameroon			0.99	0.95169	TRUE	TRUE	TRUE	30057145	0.18	51.15
PM_CMR_006	Africa	Cameroon			0.99	0.99332	TRUE	TRUE	TRUE	29555524	0.54	93.96
PM_CMR_007	Africa	Cameroon			0.98	0.99292	TRUE	TRUE	TRUE	30805384	0.51	97.69
PM_COD_001	Africa	DRC			0.00	0.99039	TRUE	TRUE	FALSE	5858851	0.37	50.25
PM_COG_001	Africa	Congo			0.93	0.98513	TRUE	TRUE	TRUE	5220000	0.31	46.91
PM_COG_002	Africa	Congo			0.95	0.99302	TRUE	TRUE	TRUE	5468561	0.49	74.61
PM_COG_003	Africa	Congo			1.00	0.98132	TRUE	TRUE	TRUE	6062412	0.27	41.50
PM_COG_004	Africa	Congo			0.98	0.99498	TRUE	TRUE	TRUE	6036723	0.53	81.01
PM_COG_005	Africa	Congo			0.84	0.99429	TRUE	TRUE	FALSE	7357853	0.61	85.91
PM_COG_006	Africa	Congo			0.87	0.96555	TRUE	TRUE	TRUE	5843079	0.31	34.54
PM_COG_008	Africa	Congo			1.00	0.99441	TRUE	TRUE	TRUE	25806301	0.56	95.30
PM_DZA_001	Africa	Algeria			0.00	0.99252	TRUE	TRUE	FALSE	11400716	0.43	81.51
PM_GAB_002	Africa	Gabon			0.96	0.99710	TRUE	TRUE	TRUE	39133287	0.69	98.45
PM_GAB_003	Africa	Gabon			0.43	0.99849	TRUE	TRUE	FALSE	35370027	0.98	98.96
PM_GAB_007	Africa	Gabon			0.86	0.97883	TRUE	TRUE	TRUE	29449870	0.26	82.31
PM_GAB_009	Africa	Gabon			0.99	0.97124	TRUE	TRUE	TRUE	33854504	0.64	97.68
PM_GAB_011	Africa	Gabon			0.73	0.99813	TRUE	TRUE	FALSE	35106096	0.76	98.66
PM_GAB_013	Africa	Gabon			0.99	0.99888	TRUE	TRUE	TRUE	37415842	0.96	97.58
PM_GAB_015	Africa	Gabon			0.99	0.99672	TRUE	TRUE	TRUE	34522893	0.67	97.59
PM_GAB_016	Africa	Gabon			0.99	0.99640	TRUE	TRUE	TRUE	34830139	0.96	97.92
PM_GHA_001	Africa	Ghana			0.99	0.99102	TRUE	TRUE	TRUE	6569115	0.43	65.58
PM_GHA_002	Africa	Ghana		2014	0.99	0.98884	TRUE	TRUE	TRUE	4073535	0.38	48.21
PM_GHA_003	Africa	Ghana		2014	0.77	0.99368	TRUE	TRUE	FALSE	5615395	0.50	77.16
PM_GHA_004_com	Africa	Ghana			1.00	0.91136	TRUE	TRUE	TRUE	9754164	0.15	8.35
PM_GHA_005	Africa	Ghana		2015	0.99	0.98036	TRUE	TRUE	TRUE	5669951	0.26	37.80
PM_GHA_006	Africa	Ghana		2015	0.99	0.98948	TRUE	TRUE	TRUE	6505780	0.37	67.51
PM_GHA_007	Africa	Ghana		2013	0.99	0.98838	TRUE	TRUE	TRUE	5084519	0.33	47.64
PM_GHA_008	Africa	Ghana		2017	0.81	0.98586	TRUE	TRUE	FALSE	6275118	0.29	38.59
PM_GHA_009	Africa	Ghana		2013	0.99	0.98523	TRUE	TRUE	TRUE	5724572	0.30	46.94
PM_GHA_010	Africa	Ghana		2011	0.99	0.99833	TRUE	TRUE	TRUE	6614123	0.75	87.67
PM_GHA_011	Africa	Ghana			1.00	0.98824	TRUE	TRUE	TRUE	6828425	0.35	57.19
PM_GHA_012	Africa	Ghana		2019	0.99	0.99895	TRUE	TRUE	TRUE	4721775	0.66	64.67
PM_GHA_014	Africa	Guinea		2019	0.92	0.99583	TRUE	TRUE	TRUE	11613147	0.67	95.75

Sample ID	Continent	Country	State	Year	FWS	Centrifuge Pm	Pipeline QC Pass	Final database	Monoclonal database	Total reads	Proportion reads mapped to PmUG01	% Genome covered >5 reads
PM_GHA_015	Africa	Ghana		2019	0.98	0.99657	TRUE	TRUE	TRUE	28245922	0.45	97.47
PM_GHA_016	Africa	Ghana		2019	0.98	0.99083	TRUE	TRUE	TRUE	28245922	0.45	97.47
PM_GIN_002	Africa	Guinea			0.78	0.98605	TRUE	TRUE	FALSE	12804563	0.63	97.56
PM_GNQ_001_com	Africa	Equitorial_Guinea			0.00	0.98749	TRUE	TRUE	FALSE	9570219	0.12	9.56
PM_KEN_001	Africa	Kenya			1.00	0.90096	TRUE	TRUE	TRUE	5580961	0.62	74.03
PM_KEN_004	Africa	Kenya		2015	1.00	0.98630	TRUE	TRUE	TRUE	7511147	0.38	62.50
PM_KEN_006	Africa	Kenya		2018	1.00	0.98805	TRUE	TRUE	TRUE	5478600	0.35	58.75
PM_KEN_007_com	Africa	Kenya		2014	0.99	0.98624	TRUE	TRUE	TRUE	8284616	0.22	42.48
PM_KEN_008	Africa	Kenya		2015	1.00	0.96697	TRUE	TRUE	TRUE	5110003	0.28	35.50
PM_KEN_009	Africa	Kenya		2014	0.99	0.97993	TRUE	TRUE	TRUE	5509448	0.46	72.26
PM_KEN_010	Africa	Kenya		2014	1.00	0.99304	TRUE	TRUE	TRUE	6213261	0.34	48.67
PM_KEN_011	Africa	Kenya			1.00	0.98500	TRUE	TRUE	TRUE	4966661	0.33	28.25
PM_KEN_012	Africa	Kenya		2019	1.00	0.99158	TRUE	TRUE	TRUE	2917984	0.24	11.42
PM_KEN_013	Africa	Kenya		2019	1.00	0.97962	TRUE	TRUE	TRUE	4052808	0.56	54.13
PM_KEN_014	Africa	Kenya			0.99	0.99641	TRUE	TRUE	TRUE	37007927	0.48	98.20
PM_LBR_003	Africa	Liberia		2017	0.42	0.99468	TRUE	TRUE	FALSE	6747461	0.60	82.09
PM_MOZ_001	Africa	Mozambique		2013	0.00	0.99560	TRUE	TRUE	FALSE	6583376	0.54	86.19
PM_MWI_001	Africa	Malawi			0.99	0.99475	TRUE	TRUE	TRUE	2578010	0.76	50.92
PM_MWI_002	Africa	Malawi		2017	0.73	0.99717	TRUE	TRUE	FALSE	7311031	0.29	51.05
PM_MWI_003	Africa	Malawi		2017	0.98	0.98101	TRUE	TRUE	TRUE	6784931	0.42	67.25
PM_MWI_004	Africa	Malawi		2012	1.00	0.99003	TRUE	TRUE	TRUE	5766117	0.15	13.82
PM_MWI_006	Africa	Malawi		2013	0.97	0.98123	TRUE	TRUE	TRUE	6542788	0.70	91.65
PM_NGA_001	Africa	Nigeria		2017	0.94	0.99765	TRUE	TRUE	TRUE	3653021	0.86	69.00
PM_NGA_002	Africa	Nigeria		2017	0.99	0.99877	TRUE	TRUE	TRUE	6481789	0.80	86.20
PM_NGA_003	Africa	Nigeria		2017	0.99	0.99800	TRUE	TRUE	TRUE	6791663	0.59	78.68
PM_NGA_004	Africa	Nigeria		2017	1.00	0.99575	TRUE	TRUE	TRUE	9281868	0.53	83.20
PM_NGA_005	Africa	Nigeria		2017	1.00	0.99395	TRUE	TRUE	TRUE	7625006	0.59	85.04
PM_NGA_006	Africa	Nigeria		2017	0.99	0.99585	TRUE	TRUE	TRUE	7185201	0.72	89.90
PM_NGA_007	Africa	Nigeria		2017	1.00	0.99769	TRUE	TRUE	TRUE	6968978	0.31	50.72
PM_NGA_008	Africa	Nigeria		2017	1.00	0.98306	TRUE	TRUE	TRUE	7070117	0.41	66.59
PM_NGA_009	Africa	Nigeria			0.79	0.99092	TRUE	TRUE	FALSE	6971540	0.44	68.63
PM_NGA_010	Africa	Nigeria		2014	0.99	0.97692	TRUE	TRUE	TRUE	6399757	0.35	63.07
PM_NGA_011	Africa	Nigeria		2013	0.65	0.98852	TRUE	TRUE	FALSE	12702180	0.57	96.61
PM_NGA_012	Africa	Nigeria		2013	0.99	0.99503	TRUE	TRUE	TRUE	5680729	0.49	74.94
PM_NGA_013_com	Africa	Nigeria			0.99	0.99376	TRUE	TRUE	TRUE	13392987	0.55	54.19
PM_NGA_014	Africa	Nigeria		2017	0.86	0.97889	TRUE	TRUE	TRUE	12307176	0.35	84.91
PM_NGA_015	Africa	Nigeria		2010	0.99	0.98804	TRUE	TRUE	TRUE	11805629	0.43	86.56
PM_NGA_016_com	Africa	Nigeria			1.00	0.99079	TRUE	TRUE	TRUE	11569669	0.77	59.17
PM_NGA_017_com	Africa	Nigeria			0.78	0.99486	TRUE	TRUE	FALSE	9696400	0.34	63.36
PM_NGA_022	Africa	Nigeria		2019	1.00	0.99769	TRUE	TRUE	TRUE	4106672	0.59	60.32
PM_NGA_023	Africa	Nigeria		2019	0.99	0.99666	TRUE	TRUE	TRUE	3225761	0.77	62.37
PM_NGA_024	Africa	Nigeria		2019	0.98	0.99883	TRUE	TRUE	TRUE	4859073	0.69	74.38
PM_NGA_025	Africa	Nigeria		2019	0.99	0.99773	TRUE	TRUE	TRUE	3846842	0.72	62.69
PM_NGA_026	Africa	Nigeria		2019	0.99	0.99840	TRUE	TRUE	TRUE	5670567	0.79	82.23
PM_NGA_027	Africa	Nigeria		2019	0.99	0.99896	TRUE	TRUE	TRUE	3558009	0.62	54.76
PM_NGA_028	Africa	Nigeria		2019	1.00	0.99739	TRUE	TRUE	TRUE	4838431	0.92	84.62
PM_NGA_029	Africa	Nigeria		2018	1.00	0.99985	TRUE	TRUE	TRUE	5078249	0.79	66.76
PM_PNG_001	Oceania	Papua_New_Guinea		2014	0.00	0.99890	TRUE	TRUE	FALSE	5828705	0.46	70.59
PM_SDN_001	Africa	Sudan		S1085	0.99	0.99152	TRUE	TRUE	TRUE	5421951	0.67	79.84
PM_SDN_002	Africa	Sudan			0.98	0.99711	TRUE	TRUE	TRUE	5317440	0.84	74.24
PM_SDN_003	Africa	Sudan			0.99	0.99835	TRUE	TRUE	TRUE	11498996	0.56	90.30
PM_SDN_004_com	Africa	Sudan			0.99	0.99484	TRUE	TRUE	TRUE	10340044	0.15	31.14
PM_SDN_005	Africa	Sudan			0.99	0.94707	TRUE	TRUE	TRUE	6476966	0.26	45.36
PM_SDN_006	Africa	Sudan			0.99	0.97818	TRUE	TRUE	TRUE	5648696	0.65	77.84
PM_SDN_007	Africa	Sudan			0.78	0.99706	TRUE	TRUE	FALSE	6411224	0.53	83.53
PM_SDN_008	Africa	Sudan		??	0.97	0.99525	TRUE	TRUE	TRUE	12781618	0.27	54.98
PM_SDN_009	Africa	Sudan		2019	1.00	0.97898	TRUE	TRUE	TRUE	4447668	0.84	67.91
PM_SDN_010	Africa	Sudan			0.98	0.99923	TRUE	TRUE	TRUE	35386513	0.57	97.85
PM_SEN_001	Africa	Senegal			0.00	0.97619	TRUE	TRUE	FALSE	30974703	0.80	98.26
PM_SLE_001	Africa	Sierra_Leone		2015	0.99	0.99855	TRUE	TRUE	TRUE	7590852	0.54	79.19

Sample ID	Continent	Country	State	Year	FWS	Centrifuge Pm	Pipeline QC Pass	Final database	Monoclonal database	Total reads	Proportion reads mapped to PmUG01	% Genome covered >5 reads
PM_SLE_002	Africa	Sierra_Leone		2014	1.00	0.99551	TRUE	TRUE	TRUE	6415798	0.15	12.23
PM_SLE_003_com	Africa	Sierra_Leone			0.92	0.93697	TRUE	TRUE	TRUE	13111050	0.32	41.57
PM_SLE_004	Africa	Sierra_Leone		2014	0.99	0.97477	TRUE	TRUE	TRUE	6807806	0.64	84.82
PM_SLE_005	Africa	Sierra_Leone		2013	0.86	0.99683	TRUE	TRUE	TRUE	5214100	0.26	37.32
PM_SLE_007	Africa	Sierra_Leone			0.99	0.98014	TRUE	TRUE	TRUE	6075627	0.58	85.03
PM_SLE_008	Africa	Sierra_Leone			0.87	0.99567	TRUE	TRUE	TRUE	6014784	0.26	43.96
PM_THA_001	Asia	Thailand		1999	0.99	0.98594	TRUE	TRUE	TRUE	6814574	0.46	47.39
PM_THA_002	Asia	Thailand		1999	0.99	0.98500	TRUE	TRUE	TRUE	8273160	0.58	60.14
PM_THA_003	Asia	Thailand		1999	0.98	0.99214	TRUE	TRUE	TRUE	7518400	0.55	60.34
PM_THA_004	Asia	Thailand		1999	0.98	0.99643	TRUE	TRUE	TRUE	7201736	0.69	85.39
PM_THA_005	Asia	Thailand		1998	0.99	0.99780	TRUE	TRUE	TRUE	7593660	0.73	87.15
PM_THA_006	Asia	Thailand		1999	0.98	0.99828	TRUE	TRUE	TRUE	5340657	0.93	76.07
PM_THA_009	Asia	Thailand		2000	1.00	0.98319	TRUE	TRUE	TRUE	2814407	0.83	28.70
PM_THA_010	Asia	Thailand		1999	0.98	0.99764	TRUE	TRUE	TRUE	6462838	0.61	82.06
PM_THA_011	Asia	Thailand		2000	0.89	0.99599	TRUE	TRUE	TRUE	5662474	0.37	58.05
PM_THA_012	Asia	Thailand		2000	0.79	0.98811	TRUE	TRUE	FALSE	6435479	0.48	74.97
PM_THA_013	Asia	Thailand		2000	0.88	0.99292	TRUE	TRUE	TRUE	7021119	0.54	81.99
PM_THA_014	Asia	Thailand		1999	0.99	0.99426	TRUE	TRUE	TRUE	5669787	0.43	64.54
PM_THA_015	Asia	Thailand		1999	0.97	0.99136	TRUE	TRUE	TRUE	12162299	0.54	90.96
PM_THA_016	Asia	Thailand		2000	0.85	0.99460	TRUE	TRUE	FALSE	11610407	0.25	67.47
PM_THA_019	Asia	Thailand		1999	0.96	0.99809	TRUE	TRUE	TRUE	6273339	0.78	79.89
PM_THA_021	Asia	Thailand		2014	0.99	0.99364	TRUE	TRUE	TRUE	5730101	0.91	78.51
PM_TZA_001_com	Africa	Tanzania		2014	0.78	0.99997	TRUE	TRUE	FALSE	10737862	0.21	53.79
PM_TZA_026	Africa	Tanzania			0.70	0.99967	TRUE	TRUE	FALSE	10039978	0.77	95.45
PM_UGA_001	Africa	Uganda			0.99	0.99812	TRUE	TRUE	TRUE	2748010	0.60	48.27
PM_UGA_004	Africa	Uganda		2013	0.97	0.99426	TRUE	TRUE	TRUE	11283745	0.21	52.87
PM_UGA_005	Africa	Uganda		2014	0.98	0.97074	TRUE	TRUE	TRUE	11830685	0.33	77.07
PM_UGA_008	Africa	Uganda		2010	0.96	0.98884	TRUE	TRUE	TRUE	5880137	0.63	80.89
PM_UGA_009	Africa	Uganda		2011	1.00	0.99628	TRUE	TRUE	TRUE	6311243	0.50	72.39
PM_UGA_010	Africa	Uganda		2012	0.99	0.99383	TRUE	TRUE	TRUE	6735417	0.24	33.21
PM_UGA_011	Africa	Uganda		2013	0.99	0.97435	TRUE	TRUE	TRUE	6254963	0.69	80.75
PM_UGA_012	Africa	Uganda		2013	0.99	0.99745	TRUE	TRUE	TRUE	6113470	0.74	81.83
PM_UGA_013	Africa	Uganda		2013	0.97	0.99802	TRUE	TRUE	TRUE	7028198	0.55	80.60
PM_UGA_014	Africa	Uganda			1.00	0.99514	TRUE	TRUE	TRUE	33806217	0.80	97.98
PM_UGA_015	Africa	Uganda			1.00	0.99849	TRUE	TRUE	TRUE	28430706	0.78	98.30
PM_UGA_016	Africa	Uganda			1.00	0.99819	TRUE	TRUE	TRUE	38316812	0.75	98.61
PM_AGO_005_com	Africa	Angola		2015	NA	0.00442	FALSE	FALSE	FALSE	9894126	0.34	13.01
PM_AGO_008	Africa	Angola		2010	NA	0.48675	FALSE	FALSE	FALSE	4912445	0.44	0.38
PM_AGO_011	Africa	Angola		2017	NA	0.69008	FALSE	FALSE	FALSE	7688972	0.05	1.05
PM_BGD_001	Asia	Bangladesh			NA	0.23738	FALSE	FALSE	FALSE	956531	0.63	0.14
PM_BGD_002	Asia	Bangladesh			NA	0.01345	FALSE	FALSE	FALSE	3591714	0.95	0.65
PM_BRA_003	South_America	Brazil			NA	0.29210	FALSE	FALSE	FALSE	5199914	0.19	0.23
PM_BRA_006	South_America	Brazil			NA	0.17427	FALSE	FALSE	FALSE	5656789	0.19	0.45
PM_BRA_008	South_America	Brazil			NA	0.44670	FALSE	FALSE	FALSE	2841718	0.24	0.21
PM_BRA_009	South_America	Brazil			NA	0.13716	FALSE	FALSE	FALSE	4986465	0.05	0.32
PM_BRA_010	South_America	Brazil			NA	0.57960	FALSE	FALSE	FALSE	5177880	0.20	0.80
PM_BRA_011	South_America	Brazil			NA	0.35073	FALSE	FALSE	FALSE	5061121	0.17	0.31
PM_BRA_012	South_America	Brazil			NA	0.16132	FALSE	FALSE	FALSE	4175569	0.17	0.31
PM_CAF_001	Africa	entral_African_Republic		2015	0.00	NA	TRUE	FALSE	FALSE	5280068	0.22	19.20
PM_CMR_005	Africa	Cameroon			NA	0.90127	FALSE	FALSE	FALSE	31843594	0.13	35.89
PM_COG_007	Africa	Congo		2019	NA	0.93393	FALSE	FALSE	FALSE	4940735	0.12	5.75
PM_COL_001	South_America	Colombia			NA	0.86428	FALSE	FALSE	FALSE	1992509	0.86	0.10
PM_COL_002	South_America	Colombia			NA	0.93319	FALSE	FALSE	FALSE	1861753	0.85	0.09
PM_GAB_001	Africa	Gabon			NA	0.97552	FALSE	FALSE	FALSE	6549129	0.62	4.40
PM_GAB_006	Africa	Gabon			0.82	0.77843	TRUE	FALSE	FALSE	30512533	0.30	81.12
PM_GAB_008	Africa	Gabon			NA	0.92548	FALSE	FALSE	FALSE	29215319	0.14	29.05
PM_GAB_010	Africa	Gabon			NA	0.53188	FALSE	FALSE	FALSE	38349459	0.09	26.67
PM_GAB_012	Africa	Gabon			NA	0.47094	FALSE	FALSE	FALSE	14363947	0.16	18.93
PM_GAB_014	Africa	Gabon			NA	0.63082	FALSE	FALSE	FALSE	36267014	0.12	25.70
PM_GAB_017	Africa	Gabon			NA	0.00132	FALSE	FALSE	FALSE	12132932	0.20	10.74

Sample ID	Continent	Country	State	Year	FWS	Centrifuge Pm	Pipeline QC Pass	Final database	Monoclonal database	Total reads	Proportion reads mapped to PmUG01	% Genome covered >5 reads
PM_GAB_018	Africa	Gabon			0.67	0.79258	TRUE	FALSE	FALSE	28173077	0.60	95.71
PM_GAB_019	Africa	Gabon			NA	0.00006	FALSE	FALSE	FALSE	40983130	0.09	14.65
PM_GIN_001	Africa	Ghana		2019	0.98	NA	TRUE	FALSE	FALSE	34354177	0.37	97.33
PM_GMB_001	Africa	The_Gambia		2015	0.00	NA	TRUE	FALSE	FALSE	5786369	0.35	54.48
PM_KEN_002	Africa	Kenya		2017	NA	0.99428	FALSE	FALSE	FALSE	4675487	0.85	8.43
PM_KEN_003	Africa	Kenya		2014	1.00	0.00188	TRUE	FALSE	FALSE	4998354	0.32	46.16
PM_LBR_001	Africa	Liberia		2013	NA	0.97876	FALSE	FALSE	FALSE	12252833	0.73	0.24
PM_LBR_002	Africa	Liberia			0.99	0.56918	TRUE	FALSE	FALSE	5014206	0.62	72.93
PM_MWI_005_com	Africa	Malawi		2014	0.98	NA	TRUE	FALSE	FALSE	9877546	0.30	67.36
PM_NGA_018_com	Africa	Nigeria		2010	NA	0.98831	FALSE	FALSE	FALSE	2889318	0.61	16.35
PM_NGA_019	Africa	Nigeria		2011	0.99	NA	TRUE	FALSE	FALSE	7462546	0.71	87.70
PM_NGA_020	Africa	Nigeria			NA	0.99783	FALSE	FALSE	FALSE	5289241	0.18	0.35
PM_NGA_021	Africa	Nigeria		2019	1.00	0.28397	TRUE	FALSE	FALSE	4876374	0.68	71.36
PM_NGA_030	Africa	Nigeria		2017	NA	0.99895	FALSE	FALSE	FALSE	8825462	0.04	0.66
PM_NGA_031	Africa	Nigeria			1.00	0.38764	TRUE	FALSE	FALSE	29237661	0.69	98.71
PM_NGA_032	Africa	Nigeria			NA	0.99693	FALSE	FALSE	FALSE	30223098	0.07	6.46
PM_NGA_033	Africa	Nigeria			0.71	0.18445	TRUE	FALSE	FALSE	31800892	0.83	98.90
PM_SLE_009	Africa	Sierra_Leone		2015	NA	0.97905	FALSE	FALSE	FALSE	1951614	0.19	0.30
PM_SLE_010	Africa	Sierra_Leone		2013	NA	0.86304	FALSE	FALSE	FALSE	6600545	0.65	0.41
PM_SLE_011_com	Africa	Sierra_Leone			0.90	0.79694	TRUE	FALSE	FALSE	11157510	0.29	68.16
PM_THA_007_com	Asia	Thailand		2000	NA	0.99971	FALSE	FALSE	FALSE	13477934	0.20	17.10
PM_THA_017_com	Asia	Thailand		2000	NA	0.97721	FALSE	FALSE	FALSE	0	N/A	0.00
PM_THA_018	Asia	Thailand		1999	0.97	NA	TRUE	FALSE	FALSE	6258265	0.73	82.47
PM_TZA_002	Africa	Tanzania		2019?-Check with Robe	NA	0.96664	FALSE	FALSE	FALSE	4863852	0.57	19.77
PM_TZA_003	Africa	Tanzania			NA	0.99731	FALSE	FALSE	FALSE	4540402	0.78	21.16
PM_TZA_004	Africa	Tanzania			NA	0.99985	FALSE	FALSE	FALSE	5042942	0.68	13.86
PM_TZA_005	Africa	Tanzania			NA	0.99665	FALSE	FALSE	FALSE	4802211	0.51	14.78
PM_TZA_006	Africa	Tanzania			NA	0.99908	FALSE	FALSE	FALSE	4639184	0.68	7.15
PM_TZA_007	Africa	Tanzania			NA	0.99671	FALSE	FALSE	FALSE	4668260	0.74	4.33
PM_TZA_008	Africa	Tanzania			NA	0.99283	FALSE	FALSE	FALSE	4907259	0.43	16.37
PM_TZA_009	Africa	Tanzania			NA	0.47288	FALSE	FALSE	FALSE	4742678	0.63	20.26
PM_TZA_010	Africa	Tanzania			NA	0.99907	FALSE	FALSE	FALSE	7255539	0.87	6.11
PM_TZA_011	Africa	Tanzania			NA	0.91219	FALSE	FALSE	FALSE	2851312	0.32	0.62
PM_TZA_012	Africa	Tanzania			NA	0.16192	FALSE	FALSE	FALSE	3567508	0.69	18.06
PM_TZA_013	Africa	Tanzania			NA	0.95240	FALSE	FALSE	FALSE	5886153	0.59	20.07
PM_TZA_014	Africa	Tanzania			NA	0.78319	FALSE	FALSE	FALSE	5891092	0.51	0.55
PM_TZA_015	Africa	Tanzania			NA	0.36655	FALSE	FALSE	FALSE	5862179	0.96	0.98
PM_TZA_016	Africa	Tanzania			NA	0.99394	FALSE	FALSE	FALSE	5644023	0.47	9.37
PM_TZA_017	Africa	Tanzania			NA	0.32011	FALSE	FALSE	FALSE	5495792	0.73	27.11
PM_TZA_018	Africa	Tanzania			NA	0.99866	FALSE	FALSE	FALSE	5366766	0.80	25.84
PM_TZA_019	Africa	Tanzania			NA	0.99538	FALSE	FALSE	FALSE	4068145	0.62	8.06
PM_TZA_020	Africa	Tanzania			NA	0.81175	FALSE	FALSE	FALSE	5148224	0.75	18.34
PM_TZA_021	Africa	Tanzania			NA	0.99886	FALSE	FALSE	FALSE	2982006	0.34	6.61
PM_TZA_022	Africa	Tanzania			NA	0.99667	FALSE	FALSE	FALSE	6488036	0.53	0.21
PM_TZA_023	Africa	Tanzania			NA	0.87825	FALSE	FALSE	FALSE	4831553	0.76	11.62
PM_TZA_024	Africa	Tanzania			NA	0.99451	FALSE	FALSE	FALSE	4298523	0.74	20.03
PM_TZA_025	Africa	Tanzania			NA	0.99950	FALSE	FALSE	FALSE	4822875	0.74	20.07
PM_UGA_002	Africa	Uganda		2017	NA	0.99413	FALSE	FALSE	FALSE	7507652	0.10	8.39
PM_UGA_003	Africa	Uganda		2014	1.00	0.84610	TRUE	FALSE	FALSE	5959961	0.53	79.24
PM_UGA_006	Africa	Uganda			NA	0.98452	FALSE	FALSE	FALSE	5919384	0.11	6.21
PM_VEN_001	South_America	Venezuela			NA	0.99783	FALSE	FALSE	FALSE	6357013	0.92	0.41
PM0015_lshtm	Africa	Angola			1.00	0.18190	TRUE	FALSE	FALSE	2593658	0.29	6.91
PM0018_lshtm	Africa	Angola			NA	0.07677	FALSE	FALSE	FALSE	3684523	0.99	0.21
PM0020_lshtm	Africa	Angola			NA	0.84630	FALSE	FALSE	FALSE	2723061	0.18	0.15
PM0028_lshtm	Africa	Angola_and_STP			NA	0.98252	FALSE	FALSE	FALSE	3026866	0.28	3.41
PM0036_lshtm	Africa	Angola			NA	0.92259	FALSE	FALSE	FALSE	3096809	0.20	0.61

S2 Table. Summary of total isolates used in sequence data generation for the global database, before and after filtering.

Continent	Country	Pre QC (n = 251)	QC Pass (n = 158)	Monoclonal* (n = 133)
Africa	Algeria	1	1	0
	Angola	18	10	10
	Cameroon	7	6	6
	Central African Republic	1	0	0
	Congo	8	7	7
	Democratic Republic of the Congo	1	1	0
	Equatorial Guinea	1	1	0
	Gabon	19	8	6
	Ghana	15	14	13
	Guinea	5	5	4
	Ivory Coast	1	1	0
	Kenya	22	11	11
	Liberia	3	1	0
	Malawi	6	5	4
	Mali	1	1	0
	Mozambique	1	1	0
	Nigeria	33	25	22
	Senegal	1	1	0
	Sierra Leone	10	7	7
	Sudan	10	10	9
	Tanzania	26	2	0
	The Gambia	1	0	0
	Uganda	16	12	12
	TOTAL	**207**	**130**	**111**
Asia	Bangladesh	2	0	0
	Indonesia	1	1	0
	Laos	1	1	0
	Malaysia	1	1	0
	Thailand	21	17	16
	Southeast Asia**	1	1	0
	TOTAL	**27**	**21**	**16**
Oceania	Papua New Guinea	1	1	0
	TOTAL	**1**	**1**	**0**
South America	Brazil	13	6	6
	Colombia	2	0	0
	Venezuela	1	0	0
	TOTAL	**16**	**6**	**6**

*Monoclonal = $F_{WS} > 0.85$
**Isolate from Southeast Asia lacking country level information

S3 Table. Highly differentiating SNPs separating African (n = 130) isolates from Asian (n = 17).

Chromosome	Position	FST	Reference	Alternate	Effect	Gene name	Gene ID	Amino acid change	Nucleotide change
1	792521	1	G	C	missense	PmUG01_01025700	PmUG01_01025700	508R>508S	792521G>C
2	515967	1	T	A	missense	NT4	PmUG01_02020400	226N>226K	515967T>A
2	516421	1	C	A	synonymous	NT4	PmUG01_02020400	319G	516421C>A
2	638779	1	G	T	missense	PmUG01_02022700	PmUG01_02022700	2135H>2135N	638779G>T
2	694942	1	G	A	missense	PmUG01_02024100	PmUG01_02024100	465G>465E	694942G>A
3	104471	1	A	T	missense	PmUG01_03012000	PmUG01_03012000	119I>119N	104471A>T
3	551583	1	G	T	missense	PmUG01_03020800	PmUG01_03020800	909W>909L	551583G>T
3	552153	1	A	C	missense	PmUG01_03020800	PmUG01_03020800	1099N>1099T	552153A>C
4	256898	1	A	C	missense	PmUG01_04014200	PmUG01_04014200	924K>924Q	256898A>C
4	317346	1	A	T	missense	ATG11	PmUG01_04015700	1316Y>1316N	317346A>T
4	317347	1	G	T	synonymous	ATG11	PmUG01_04015700	1315S	317347G>T
4	811445	1	T	C	missense	PmUG01_04025500	PmUG01_04025500	1266D>1266G	811445T>C
4	816163	1	C	A	.				
4	816428	1	T	C	.				
4	818545	1	G	A	.				
4	823561	1	C	G	missense	PmUG01_04025700	PmUG01_04025700	235N>235K	823561C>G
4	930684	1	T	C	.				
5	436269	1	T	A	missense	GAMA	PmUG01_05017200	254L>254M	436269T>A
5	635184	1	A	G	intron	DRE2			
5	1499392	1	T	C	synonymous	ApiAP2	PmUG01_05037900	3143G	1499392T>C
6	257731	1	G	A	.				
6	363359	1	C	G	missense	PRP5	PmUG01_06016200	334G>334R	363359C>G
6	408635	1	G	C	missense	PmUG01_06017100	PmUG01_06017100	257S>257W	408635G>C
6	419801	1	A	C	missense	CAF40	PmUG01_06017300	569I>569S	419801A>C
6	449583	1	G	T	.				
6	617021	1	C	T	.				
7	635878	1	G	T	.				
7	637290	1	C	T	missense	PmUG01_07021200	PmUG01_07021200	565R>565K	637290C>T
7	637934	1	G	A	synonymous	PmUG01_07021200	PmUG01_07021200	350I	637934G>A
7	641372	1	G	T	.				
7	653057	1	T	A	.				
7	660470	1	T	C	missense	SET4	PmUG01_07021600	771N>771S	660470T>C
7	684064	1	T	C	synonymous	PmUG01_07021900	PmUG01_07021900	203C	684064T>C
7	710172	1	A	T	intron	PmUG01_07022600			
7	729550	1	T	C	missense	CRMP1	PmUG01_07022900	1963S>1963P	729550T>C
7	731085	1	G	A	missense	CRMP1	PmUG01_07022900	2474M>2474I	731085G>A
7	776253	1	C	T	missense	PmUG01_07024000	PmUG01_07024000	229A>229V	776253C>T
7	786538	1	A	G	.				
7	789061	1	A	G	missense	RPN8	PmUG01_07024400	114K>114R	789061A>G
7	814328	1	T	G	.				
7	835623	1	C	G	missense	PmUG01_07025700	PmUG01_07025700	196V>196L	835623C>G
7	841123	1	G	A	synonymous	PmUG01_07025800	PmUG01_07025800	736L	841123G>A
7	860746	1	C	T	intron	PmUG01_07026100			
7	1262679	1	T	C	.				
7	1268257	1	T	C	.				
8	1104750	1	G	A	.				
8	1122265	1	G	A	missense	RIPR	PmUG01_08032700	1018D>1018N	1122265G>A
8	1122538	1	A	G	missense	RIPR	PmUG01_08032700	1109R>1109G	1122538A>G
8	1173555	1	T	C	.				
8	1229600	1	C	T	missense	PmUG01_08034700	PmUG01_08034700	22D>22N	1229600C>T
8	1388092	1	G	A	synonymous	PmUG01_08037700	PmUG01_08037700	302K	1388092G>A
8	1394343	1	G	A	missense	PmUG01_08037800	PmUG01_08037800	192D>192N	1394343G>A
8	1580233	1	C	T	.				
8	1755210	1	C	T	missense	PmUG01_08045200	PmUG01_08045200	131G>131E	1755210C>T
8	1755319	1	G	T	missense	PmUG01_08045200	PmUG01_08045200	95Q>95K	1755319G>T
8	1755323	1	G	T	missense	PmUG01_08045200	PmUG01_08045200	93N>93K	1755323G>T
8	1757220	1	G	A	missense	PmUG01_08045300	PmUG01_08045300	772R>772W	1757220G>A
8	1794114	1	G	A	synonymous	PmUG01_08046000	PmUG01_08046000	2366N	1794114G>A
8	1808393	1	C	A	.				
8	1900510	1	G	A	synonymous	PmUG01_08048300	PmUG01_08048300	1879Q	1900510G>A
8	1917337	1	G	A	synonymous	PmUG01_08048500	PmUG01_08048500	1423Q	1917337G>A
8	2024675	1	C	T	missense	PmUG01_08051100	PmUG01_08051100	2682S>2682L	2024675C>T
8	2037801	1	G	A	synonymous	PmUG01_08051400	PmUG01_08051400	1049V	2037801G>A
8	2049708	1	A	C	missense	CSP	PmUG01_08051600	283K>283Q	2049708A>C
8	2049740	1	C	T	synonymous	CSP	PmUG01_08051600	293N	2049740C>T
8	2049745	1	A	G	missense	CSP	PmUG01_08051600	295E>295G	2049745A>G
9	375959	1	A	G	.				
9	402441	1	G	A	missense	ApiAP2	PmUG01_09017400	1084P>1084S	402441G>A
9	672797	1	C	G	missense	PmUG01_09024200	PmUG01_09024200	86A>86G	672797C>G
9	673072	1	C	A	missense	PmUG01_09024200	PmUG01_09024200	178P>178T	673072C>A
9	673075	1	A	T	missense	PmUG01_09024200	PmUG01_09024200	179K>179F	673075A>T+673076A>T+673077A>T
9	673076	1	A	T	missense	PmUG01_09024200	PmUG01_09024200	179K>179F	673075A>T+673076A>T+673077A>T
9	673077	1	A	T	missense	PmUG01_09024200	PmUG01_09024200	179K>179F	673075A>T+673076A>T+673077A>T
9	673080	1	T	G	synonymous	PmUG01_09024200	PmUG01_09024200	180G	673080T>G
9	673603	1	T	A	missense	PmUG01_09024200	PmUG01_09024200	355Y>355N	673603T>A
9	683844	1	A	T	intron	PmUG01_09024400			
9	897753	1	G	T	synonymous	PmUG01_09029100	PmUG01_09029100	218G	897753G>T
9	1004676	1	G	A	missense	PmUG01_09031400	PmUG01_09031400	192H>192Y	1004676G>A
9	1072891	1	G	A	missense	CDPK7	PmUG01_09032300	637R>637K	1072891G>A
9	1073347	1	A	T	missense	CDPK7	PmUG01_09032300	789N>789I	1073347A>T
9	1111761	1	A	T	.				
9	1821204	1	T	A	missense	PmUG01_09048100	PmUG01_09048100	1439K>1439N	1821204T>A
9	2061398	1	A	G	missense	PmUG01_09052100	PmUG01_09052100	78H>78R	2061398A>G
9	2250043	1	A	T	missense	PmUG01_09056100	PmUG01_09056100	1170I>1170L	2250043A>T

Chromosome	Position	FST	Reference	Alternate	Effect	Gene name	Gene ID	Amino acid change	Nucleotide change
9	2286499	1	C	T	synonymous	PmUG01_09056900	PmUG01_09056900	451R	2286499C>T
9	2294640	1	G	C	missense	PmUG01_09057000	PmUG01_09057000	421N>421K	2294640G>C
9	2302379	1	C	T	.				
9	2304409	1	G	C	.				
9	2305403	1	C	G	.				
9	2311853	1	C	T	.				
10	562403	1	T	A	missense	PmUG01_10022200	PmUG01_10022200	6349N>6349I	562403T>A
10	566074	1	T	C	missense	PmUG01_10022200	PmUG01_10022200	5125I>5125M	566074T>C
10	584741	1	A	G	.				
10	588628	1	G	A	intron	PmUG01_10022400			
10	1226412	1	C	T	synonymous	PmUG01_10035400	PmUG01_10035400	49E	1226412C>T
11	710969	1	A	C	.				
11	791564	1	T	A	intron	PmUG01_11025200			
11	811956	1	A	C	missense	GCbeta	PmUG01_11025500	2606I>2606R	811956A>C
11	814323	1	G	T	missense	GCbeta	PmUG01_11025500	1817A>1817E	814323G>T
11	819553	1	G	A	intron	GCbeta			
11	869627	1	C	G	.				
11	938450	1	C	T	missense	PmUG01_11028000	PmUG01_11028000	136T>136M	938450C>T
11	1290491	1	T	G	missense	PK4	PmUG01_11035400	1259N>1259T	1290491T>G
11	1443377	1	T	C	missense	SMS1	PmUG01_11038600	108I>108V	1443377T>C
11	1510793	1	A	T	missense	TKL4	PmUG01_11039800	2181K>2181I	1510793A>T
11	1599635	1	G	A	.				
11	1669518	1	A	G	.				
11	2018425	1	G	A	missense	P12	PmUG01_11050500	127E>127K	2018425G>A+2018427A>G
11	2082352	1	T	C	.				
12	539234	1	T	A	.				
12	539240	1	T	A	.				
12	983825	1	G	C	.				
12	991725	1	G	T	.				
12	1026365	1	T	C	missense	PmUG01_12032400	PmUG01_12032400	412K>412R	1026365T>C
12	1226733	1	A	C	missense	PmUG01_12036500	PmUG01_12036500	1329E>1329D	1226733A>C
12	1250275	1	G	A	missense	PmUG01_12037000	PmUG01_12037000	679A>679V	1250275G>A
12	1252779	1	C	T	missense	PmUG01_12037000	PmUG01_12037000	3C>3Y	1252779C>T
12	1472912	1	A	T	.				
12	2092729	1	C	A	missense	PmUG01_12054500	PmUG01_12054500	2022S>2022R	2092729C>A
12	2098884	1	A	T	missense	PmUG01_12054500	PmUG01_12054500	4074N>4074I	2098884A>T
12	2316148	1	T	A	.				
12	2320861	1	G	A	.				
12	2338091	1	C	T	missense	PmUG01_12060500	PmUG01_12060500	9T>9I	2338091C>T
12	2543395	1	C	T	missense	RON2	PmUG01_12065000	1728R>1728K	2543395C>T
12	2890071	1	T	C	.				
12	2937163	1	G	C	missense	PARN	PmUG01_12073500	117G>117R	2937163G>C
12	2968947	1	T	C	.				
12	3069533	1	G	T	.				
12	3071772	1	T	G	missense	MFS6	PmUG01_12076200	265N>265H	3071772T>G
12	3112832	1	G	C	missense	PmUG01_12077000	PmUG01_12077000	79G>79R	3112832G>C
12	3173886	1	G	A	synonymous	PmUG01_12078200	PmUG01_12078200	689H	3173886G>A
12	3307164	1	G	A	.				
13	956585	1	T	A	intron	NOP10			
13	956968	1	A	C	.				
13	1298759	1	G	A	.				
13	2293168	1	T	G	missense	PmUG01_13054900	PmUG01_13054900	554K>554N	2293168T>G
14	403332	1	G	A	missense	PmUG01_14016200	PmUG01_14016200	89S>89N	403332G>A
14	855970	1	C	G	.				
14	856490	1	C	T	.				
14	1436977	1	G	A	.				
14	1615408	1	T	G	missense	PmUG01_14044700	PmUG01_14044700	382I>382S	1615408T>G
14	1616513	1	G	T	synonymous	PmUG01_14044700	PmUG01_14044700	750T	1616513G>T
14	1617405	1	A	G	missense	PmUG01_14044700	PmUG01_14044700	1048N>1048D	1617405A>G
14	1632300	1	T	A	missense	CAF1	PmUG01_14045000	1247I>1247F	1632300T>A
14	1915111	1	A	G	.				
14	1918866	1	A	G	.				
14	1919072	1	A	T	.				
14	1925559	1	C	T	missense	PmUG01_14053000	PmUG01_14053000	2335A>2335T	1925559C>T
14	1928395	1	G	A	synonymous	PmUG01_14053000	PmUG01_14053000	1510T	1928395G>A
14	1929136	1	T	C	synonymous	PmUG01_14053000	PmUG01_14053000	1263E	1929136T>C
14	1962149	1	G	A	synonymous	PmUG01_14053700	PmUG01_14053700	577N	1962149G>A
14	2110963	1	T	G	missense	ApiAP2	PmUG01_14056700	154H>154Q	2110963T>G
14	2136305	1	A	C	missense	AP2-G	PmUG01_14056900	1916S>1916R	2136305A>C
14	2148801	1	A	T	.				
14	2413890	1	C	G	.				
14	2439329	1	A	T	missense	MSP9	PmUG01_14062900	425V>425D	2439329A>T
14	2440518	1	T	C	missense	MSP9	PmUG01_14062900	29T>29A	2440518T>C
14	2477548	1	A	T	missense	MRP2	PmUG01_14063400	1215L>1215I	2477548A>T
14	2491958	1	T	C	synonymous	PmUG01_14063500	PmUG01_14063500	812S	2491958T>C
14	2814078	1	G	T	missense	NOT4	PmUG01_14069500	782A>782S	2814078G>T
14	3144824	1	T	C	.				
14	3156801	1	T	C	missense	DOC2	PmUG01_14077100	1213E>1213G	3156801T>C
14	3164993	1	T	C	missense	PmUG01_14077200	PmUG01_14077200	578D>578G	3164993T>C
14	3189891	1	T	G	missense	PmUG01_14077700	PmUG01_14077700	568K>568Q	3189891T>G
14	3191473	1	T	A	missense	PmUG01_14077700	PmUG01_14077700	40R>40S	3191473T>A
14	3283700	1	A	T	synonymous	BUD13	PmUG01_14079800	384G	3283700A>T
14	3362924	1	T	A	missense	PmUG01_14081900	PmUG01_14081900	2076N>2076K	3362924T>A

S4 Table. Regions of homology within continents using IBD and monoclonal isolates.
Monoclonal isolates ($F_{WS} > 0.85$): Africa, n = 108; Asia, n = 14; South America, n = 5.

Chromosome	Start	End	Africa	Asia	South America	Genes	Gene products
1	1050001	1100000	0.034	0.139	0.097	PmUG01_01030200(SENP2); PmUG01_01030400(NOL10)	sentrin-specific protease 2, putative; nucleolar protein 10, putative
1	1100001	1150000	0.074	0.144	0.108	NA	NA
2	550001	600000	0.009	0.153	0.1	PmUG01_02021500(TIM50); PmUG01_02021600(VPS53); PmUG01_02021900(DNMT)	mitochondrial import inner membrane translocase subunit TIM50, putative; vacuolar protein sorting-associated protein 53, putative; DNA (cytosine-5)-methyltransferase, putative
2	600001	650000	0.009	0.142	0.1	NA	NA
2	700001	750000	0.01	0.143	0.056	NA	NA
2	750001	800000	0.018	0.187	0.096	NA	NA
3	100001	150000	0.061	0.154	0.162	NA	NA
3	250001	300000	0.023	0.113	0.214	PmUG01_03014800(DPAP3); PmUG01_03015000(P41); PmUG01_03015100(DDX51); PmUG01_03015200(SEC24B)	dipeptidyl aminopeptidase 3, putative; 6-cysteine protein P41, putative; ATP-dependent RNA helicase DDX51, putative; protein transport protein Sec24B, putative
3	300001	350000	0.024	0.104	0.25	PmUG01_03015500(PRPF8); PmUG01_03016000(ASP)	pre-mRNA-processing-splicing factor 8, putative; apical sushi protein, putative
4	150001	200000	0.065	0.056	0.1	PmUG01_04012400(LSA3)	liver stage antigen 3, putative
5	650001	700000	0.038	0.042	0.288	PmUG01_05022100(DTC); PmUG01_05022300(TOM7); PmUG01_05022400(LipB); PmUG01_05022700(GCN5)	dicarboxylate/tricarboxylate carrier, putative; mitochondrial import receptor subunit TOM7, putative; lipoate-protein ligase B, putative; histone acetyltransferase GCN5, putative
5	700001	750000	0.027	0.044	0.276	PmUG01_05023000(VPS15); PmUG01_05023200(SNRNP40); PmUG01_05023300(TRP1); PmUG01_05023400(SEC23); PmUG01_05023500(DLC1); PmUG01_05023700(SNRPG); PmUG01_05024000(MED7)	serine/threonine protein kinase VPS15, putative; U5 small nuclear ribonucleoprotein 40 kDa protein, putative; thrombospondin-related protein 1, putative; protein transport protein SEC23, putative; dynein light chain 1, putative; small nuclear ribonucleoprotein G, putative; mediator of RNA polymerase II transcription subunit 7, putative
6	650001	700000	0.036	0.094	0.005	PmUG01_06023200(IspE)	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase, putative
7	200001	250000	0.017	0.161	0.046	NA	NA
7	250001	300000	0.031	0.298	0.1	NA	NA
7	300001	350000	0.019	0.223	0.1	PmUG01_07014900(DDX60); PmUG01_07015400(RPL32)	ATP-dependent RNA helicase DDX60, putative; 60S ribosomal protein L32, putative
7	350001	400000	0.017	0.185	0.1	PmUG01_07015700(PH); PmUG01_07015900(SPC3)	PH domain-containing protein, putative; signal peptidase complex subunit 3, putative
7	500001	550000	0.016	0.066	0.28	PmUG01_07019000(ClpY)	ATP-dependent protease ATPase subunit ClpY, putative
7	1800001	1850000	0.01	0.062	0.22	PmUG01_07047900(ETRAMP)	early transcribed membrane protein
7	1850001	1900000	0.048	0.08	0.138	NA	NA
8	350001	400000	0.035	0.075	0.066	NA	NA
8	1500001	1550000	0.013	0.07	0.275	PmUG01_08039500(FNT); PmUG01_08039600(NUF2); PmUG01_08039800(PPase)	formate-nitrite transporter, putative; kinetochore protein NUF2, putative; inorganic pyrophosphatase, putative
8	1750001	1800000	0.013	0.142	0.043	PmUG01_08045600(EIF3K); PmUG01_08045700(DHX57); PmUG01_08045800(PIESP1)	eukaryotic translation initiation factor 3 subunit K, putative; ATP-dependent RNA helicase DHX57, putative; parasite-infected erythrocyte surface protein, putative
8	1850001	1900000	0.01	0.148	NA	PmUG01_08047300(SF3B1); PmUG01_08047600(PRPF19); PmUG01_08048000(CCT7)	splicing factor 3B subunit 1, putative; pre-mRNA-processing factor 19, putative; T-complex protein 1 subunit eta, putative
8	1900001	1950000	0.008	0.142	NA	PmUG01_08048800(ClpP); PmUG01_08048900(EB1); PmUG01_08049100(RPS12); PmUG01_08049400(CCT2)	ATP-dependent Clp protease proteolytic subunit, putative; microtubule-associated protein RP/EB family, putative; 40S ribosomal protein S12, putative; T-complex protein 1 subunit beta, putative
9	200001	250000	0.015	0.072	0.268	PmUG01_09013400(CK2beta1); PmUG01_09013500(NOT1)	casein kinase II beta chain, putative; CCR4-NOT transcription complex subunit 1, putative
9	1950001	2000000	0.015	0.025	0.279	PmUG01_09050700(DPM1)	dolichol-phosphate mannosyltransferase, putative

Chromosome	Start	End	Africa	Asia	South America	Genes	Gene products
9	2200001	2250000	0.033	0.047	0.055	PmUG01_09055400(LRR10); PmUG01_09055500(ORP1)	leucine-rich repeat protein; oocyst rupture protein 1, putative
9	2250001	2300000	0.032	0.067	0.037	PmUG01_09056700(MAEBL); PmUG01_09056800(PEX22)	membrane associated erythrocyte binding-like protein, putative; peroxisome assembly protein 22, putative
10	1650001	1700000	0.171	0.235	0.284	PmUG01_10044900(BDP1); PmUG01_10045600(SDHA); PmUG01_10046100(MRScyt)	bromodomain protein 1, putative; flavoprotein subunit of succinate dehydrogenase, putative; methionine--tRNA ligase, putative
10	1700001	1750000	0.155	0.167	0.313	PmUG01_10047600(ACT)	acetyl-CoA transporter, putative
10	1750001	1800000	0.015	0.151	0.208	PmUG01_10047900(PyKII); PmUG01_10048300(DYN2); PmUG01_10048400(XPB); PmUG01_10048500(ERH)	pyruvate kinase 2, putative; dynamin-like protein, putative; TFIIH basal transcription factor complex helicase XPB subunit, putative; enhancer of rudimentary homolog, putative
10	1800001	1850000	0.102	0.129	0.183	NA	NA
10	1850001	1900000	0.036	0.006	0.034	NA	NA
11	300001	350000	0.063	0.047	0.055	NA	NA
11	350001	400000	0.051	0.107	0.049	NA	NA
11	700001	750000	0.011	0.145	0.1	PmUG01_11023900(RUVB3); PmUG01_11024200(PCNA1); PmUG01_11024300(GAC)	RuvB-like helicase 3, putative; proliferating cell nuclear antigen 1, putative; glideosome-associated connector, putative
11	850001	900000	0.014	0.157	0.084	PmUG01_11026600(CELF1); PmUG01_11026800(HMGB4)	CUGBP Elav-like family member 1, putative; high mobility group protein B4, putative
12	100001	150000	0.02	0.126	0.3	PmUG01_12012100(MON1); PmUG01_12012200(AQR)	vacuolar fusion protein MON1, putative; intron-binding protein aquarius, putative
12	150001	200000	0.015	0.068	0.3	PmUG01_12012700(UTP15); PmUG01_12012800(ABCB6)	U3 small nucleolar RNA-associated protein 15, putative; ABC transporter B family member 6, putative
12	3350001	3400000	0.036	0.033	0.044	NA	NA
13	50001	100000	0.047	0.134	0.074	NA	NA
13	100001	150000	0.075	0.149	0.139	NA	NA
14	400001	450000	0.057	0.038	0.195	NA	NA
14	450001	500000	0.029	0.068	0.217	PmUG01_14018800(UIS3)	protein UIS3, putative
14	550001	600000	0.011	0.088	0.217	NA	NA
14	600001	650000	0.012	0.098	0.286	PmUG01_14021300(CYC4); PmUG01_14021500(NIF1); PmUG01_14021600(RPB11); PmUG01_14021900(ApiAP2); PmUG01_14022200(MKP1); PmUG01_14022300(S2P)	cyclin, putative; NLI interacting factor-like phosphatase, putative; DNA-directed RNA polymerase II subunit RPB11, putative; AP2 domain transcription factor, putative; mitogen-activated protein kinase phosphatase 1, putative; site-2 protease S2P, putative
14	650001	700000	0.014	0.171	0.063	PmUG01_14023100(RPT4)	26S protease regulatory subunit 10B, putative
14	700001	750000	0.013	0.197	NA	PmUG01_14023800(UTP6); PmUG01_14024000(DBP6)	U3 small nucleolar RNA-associated protein 6, putative; ATP-dependent RNA helicase DBP6, putative
14	800001	850000	0.013	0.142	NA	PmUG01_14025900(PPM6); PmUG01_14026200(GAR1); PmUG01_14026400(VPS18); PmUG01_14026700(OSCP)	protein phosphatase PPM6, putative; H/ACA ribonucleoprotein complex subunit 1, putative; vacuolar protein sorting-associated protein 18, putative; ATP synthase subunit O, mitochondrial, putative
14	850001	900000	0.012	0.147	0.102	PmUG01_14027300(RAB5b); PmUG01_14027700(ISD11); PmUG01_14028100(AP1M1); PmUG01_14028200(RPT1); PmUG01_14028500(M1AAP)	ras-related protein Rab-5B, putative; protein ISD11, putative; AP-1 complex subunit mu-1, putative; 26S protease regulatory subunit 7, putative; M1-family alanyl aminopeptidase, putative
14	1600001	1650000	0.036	0.063	0.1	PmUG01_14044800(CARM1); PmUG01_14045100(EMC1); PmUG01_14045300(CUL1)	histone-arginine methyltransferase CARM1, putative; ER membrane protein complex subunit 1, putative; cullin-1, putative
14	1650001	1700000	0.042	0.046	0.1	PmUG01_14045300(CUL1); PmUG01_14045500(PPPK-DHPS); PmUG01_14045700(DBP1); PmUG01_14045900(AQP2); PmUG01_14046000(PPM5)	cullin-1, putative; hydroxymethyldihydropterin pyrophosphokinase-dihydropteroate synthase, putative; ATP-dependent RNA helicase DBP1, putative; aquaporin, putative; protein phosphatase PPM5, putative
14	1700001	1750000	0.051	0.06	0.021	PmUG01_14046600(RUVB1)	RuvB-like helicase 1, putative
14	1750001	1800000	0.035	0.055	NA	PmUG01_14047400(GAT); PmUG01_14047600(EIF3A); PmUG01_14047800(BDP2); PmUG01_14048400(DLC8); PmUG01_14048500(RPB8)	glycerol-3-phosphate 1-O-acyltransferase, putative; eukaryotic translation initiation factor 3 subunit A, putative; bromodomain protein 2, putative; dynein light chain 1, putative; DNA-directed RNA polymerases I, II, and III subunit RPABC3, putative
14	1800001	1850000	0.032	0.061	NA	PmUG01_14049000(SET5); PmUG01_14049200(EMC6)	histone-lysine N-methyltransferase SET5, putative; ER membrane protein complex subunit 6, putative
14	2950001	3000000	0.016	0.173	0.005	PmUG01_14072500(CWC22); PmUG01_14072800(NSM); PmUG01_14073000(ACS11); PmUG01_14073100(PK2); PmUG01_14073400(ApiAP2)	pre-mRNA-splicing factor CWC22, putative; sphingomyelin phosphodiesterase, putative; acyl-CoA synthetase, putative; protein kinase 2, putative; AP2 domain transcription factor, putative
14	3000001	3050000	0.015	0.151	0.1	PmUG01_14073400(ApiAP2); PmUG01_14073700(GyrB); PmUG01_14073900(FTSH1); PmUG01_14074100(VPS16); PmUG01_14074300(PFS2)	AP2 domain transcription factor, putative; DNA gyrase subunit B, putative; ATP-dependent zinc metalloprotease FTSH 1, putative; vacuolar protein sorting-associated protein 16, putative; polyadenylation factor subunit 2, putative

S5 Table. Signals of selection within African and Asian continents, as determined by iHS.

Monoclonal isolates ($F_{WS} > 0.85$): Africa, n = 108; Asia, n = 14.

Chromosome	Position	iHS	LOGPVALUE	Ref	Alt	Gene name	Continent	Asia allele frequency (n =16)	Africa allele frequency (n = 108)
4	1044771	4.260	4.690	T	C	PmUG01_04030700	Africa	N(9),1(4),0(3)	N(40),0(38),1(22),0.5(8)
4	1045586	4.635	5.447	A	G	PmUG01_04030700	Africa	N(12),0(2),1(2)	0(67),N(37),1(3),0.5(1)
4	1045783	4.615	5.405	T	G	PmUG01_04030700	Africa	N(14),0(2)	0(57),N(42),1(6),0.5(3)
4	1045997	4.932	6.089	T	A	PmUG01_04030700	Africa	1(7),N(7),0(2)	0(59),N(37),1(11),0.5(1)
6	143173	4.198	4.570	G	A	PmUG01_06012400	Africa	0(15),N(1)	0(90),N(15),1(3)
6	675922	-4.010	4.217	C	T	PmUG01_06022600	Africa	N(10),0(4),1(2)	N(46),0(38),1(24)
7	352512	-3.964	4.132	G	T	PmUG01_07015600	Africa	0(12),N(3),1(1)	0(60),N(28),1(20)
7	666500	4.825	5.853	T	C	PmUG01_07021700	Africa	0(13),N(2),0.5(1)	0(99),N(7),0.5(2)
7	1389129	4.725	5.638	T	C	PmUG01_07037600	Africa	0(12),N(4)	0(90),N(16),1(2)
7	1389137	4.335	4.836	G	A	PmUG01_07037600	Africa	0(13),N(3)	0(91),N(15),1(2)
7	1467158	4.376	4.918	A	G	PmUG01_07038900	Africa	0(7),N(7),0.5(1),1(1)	N(59),0(47),1(2)
7	1467184	6.335	9.626	A	G	PmUG01_07038900	Africa	N(9),0(7)	N(67),0(37),1(3),0.5(1)
7	1467194	7.997	14.896	G	A	PmUG01_07038900	Africa	N(13),0(3)	N(87),0(19),1(2)
7	1467202	4.076	4.339	G	A	PmUG01_07038900	Africa	N(11),0(5)	N(80),0(26),1(2)
7	1737026	4.741	5.673	A	T	PmUG01_07045500	Africa	N(10),0(6)	0(53),N(53),1(2)
8	723170	5.845	8.294	T	A	PmUG01_08023900	Africa	0(10),N(6)	0(69),N(37),0.5(1),1(1)
8	759509	5.000	6.242	T	A	PmUG01_08024700	Africa	0(8),N(8)	0(74),N(32),0.5(1),1(1)
8	914192	5.140	6.561	C	T	NIF4	Africa	0(15),0.5(1)	0(95),N(11),0.5(2)
8	939232	5.730	7.998	A	G	PmUG01_08029200	Africa	0(11),N(4),0.5(1)	0(71),N(35),0.5(2)
8	939238	5.384	7.137	A	G	PmUG01_08029200	Africa	0(12),N(4)	0(76),N(30),1(2)
8	939247	4.246	4.662	A	G	PmUG01_08029200	Africa	0(13),N(3)	0(88),N(18),0.5(2)
8	939250	4.258	4.686	A	G	PmUG01_08029200	Africa	0(11),N(4),1(1)	0(89),N(17),0.5(1),1(1)
8	1364361	5.209	6.720	C	T	PmUG01_08037500	Africa	0(14),N(2)	0(96),N(10),0.5(1),1(1)
9	932289	4.616	5.407	T	C	PmUG01_09029800	Africa	0(9),N(7)	0(82),N(24),0.5(1),1(1)
10	487480	-4.743	5.676	C	G	PmUG01_10020700	Africa	0(16)	0(84),1(21),N(2),0.5(1)
10	487486	-4.844	5.896	A	G	PmUG01_10020700	Africa	0(16)	0(84),1(21),N(2),0.5(1)
10	772568	6.208	9.271	G	A	PmUG01_10026600	Africa	0(15),N(1)	0(93),N(13),1(2)
10	1163663	4.840	5.887	A	T	PmUG01_10034400	Africa	0(16)	0(92),N(14),0.5(2)
10	1682559	3.989	4.179	G	A	PmUG01_10045700	Africa	0(9),1(6),N(1)	0(61),1(34),N(13)
11	1606872	5.771	8.104	A	T	PmUG01_11041800	Africa	N(10),0(6)	N(63),0(43),1(2)
11	1606875	6.523	10.163	A	T	PmUG01_11041800	Africa	N(11),0(5)	N(70),0(36),1(2)
11	1606880	6.378	9.747	A	T	PmUG01_11041800	Africa	N(10),0(6)	N(73),0(33),1(2)
11	1606885	4.700	5.585	A	T	PmUG01_11041800	Africa	0(10),N(6)	0(53),N(53),1(2)
11	1606892	6.236	9.349	A	T	PmUG01_11041800	Africa	0(10),N(6)	N(54),0(51),1(3)
11	1606900	4.161	4.499	A	T	PmUG01_11041800	Africa	0(12),N(4)	N(54),0(52),1(2)
11	1606913	5.115	6.502	A	T	PmUG01_11041800	Africa	0(11),N(5)	0(82),N(24),1(2)
11	2047003	5.645	7.781	G	A	EIF3L	Africa	0(13),N(2),0.5(1)	0(94),N(12),0.5(2)
12	400787	5.878	8.381	C	A	WDR92	Africa	N(10),0(6)	N(56),0(50),0.5(1),1(1)
13	122971	4.617	5.409	G	A	PmUG01_13011300	Africa	0(13),N(2),1(1)	0(89),N(17),0.5(1),1(1)
13	122972	4.391	4.948	C	T	PmUG01_13011300	Africa	N(9),0(6),1(1)	N(70),0(32),1(5),0.5(1)
13	684293	-5.998	8.700	A	G	PmUG01_13022300	Africa	N(12),1(4)	N(77),1(27),0(4)
13	764455	-5.304	6.946	A	C	PmUG01_13024100	Africa	1(11),N(5)	1(64),N(41),0(2),0.5(1)
13	810324	4.961	6.154	A	T	PmUG01_13025000	Africa	0(14),N(2)	0(93),N(13),0.5(2)
13	1126178	4.948	6.125	C	T	PmUG01_13031000	Africa	0(10),N(6)	0(62),N(44),0.5(2)

Chromosome	Position	iHS	LOGPVALUE	Ref	Alt	Gene name	Continent	Asia allele frequency (n =16)	Africa allele frequency (n = 108)
14	446101	4.731	5.652	T	A	PmUG01_14017400	Africa	0(15),N(1)	0(95),N(11),1(2)
14	450529	5.153	6.591	C	T	PmUG01_14017600	Africa	N(12),0(4)	N(76),0(30),1(2)
14	450541	5.848	8.302	A	G	PmUG01_14017600	Africa	N(12),0(4)	N(73),0(31),1(4)
14	450547	5.848	8.302	C	T	PmUG01_14017600	Africa	N(12),0(4)	N(73),0(31),1(4)
14	450570	5.746	8.039	C	T	PmUG01_14017600	Africa	N(12),0(4)	N(72),0(30),1(6)
14	450572	5.746	8.039	C	T	PmUG01_14017600	Africa	N(12),0(4)	N(72),0(30),1(6)
14	450574	5.746	8.039	C	A	PmUG01_14017600	Africa	N(12),0(4)	N(72),0(30),1(6)
14	450585	5.746	8.039	T	G	PmUG01_14017600	Africa	N(12),0(4)	N(72),0(30),1(6)
14	450586	5.746	8.039	A	G	PmUG01_14017600	Africa	N(12),0(4)	N(71),0(31),1(6)
14	450592	5.746	8.039	A	G	PmUG01_14017600	Africa	N(12),0(4)	N(72),0(30),1(6)
14	450596	5.746	8.039	T	G	PmUG01_14017600	Africa	N(12),0(4)	N(71),0(31),1(6)
14	450610	4.604	5.383	A	G	PmUG01_14017600	Africa	N(12),0(4)	N(75),0(31),1(2)
14	450618	4.604	5.383	A	C	PmUG01_14017600	Africa	N(12),0(4)	N(74),0(31),1(3)
14	450621	4.604	5.383	A	T	PmUG01_14017600	Africa	N(12),0(4)	N(74),0(31),1(3)
14	450626	4.604	5.383	G	C	PmUG01_14017600	Africa	N(12),0(4)	N(74),0(31),1(3)
14	450628	4.604	5.383	A	T	PmUG01_14017600	Africa	N(12),0(4)	N(74),0(31),1(3)
14	450630	4.604	5.383	C	T	PmUG01_14017600	Africa	N(12),0(4)	N(74),0(31),1(3)
14	450644	5.223	6.753	C	T	PmUG01_14017600	Africa	N(12),0(4)	N(75),0(30),1(3)
14	450651	4.604	5.383	G	C	PmUG01_14017600	Africa	N(12),0(4)	N(74),0(31),1(3)
14	450658	5.223	6.753	T	C	PmUG01_14017600	Africa	N(12),0(4)	N(74),0(30),1(4)
14	450686	4.856	5.922	A	C	PmUG01_14017600	Africa	N(12),0(4)	N(69),0(32),1(7)
14	450696	4.856	5.922	A	G	PmUG01_14017600	Africa	N(11),0(5)	N(69),0(32),1(7)
14	450700	4.856	5.922	G	A	PmUG01_14017600	Africa	N(12),0(4)	N(68),0(33),1(7)
14	450703	4.856	5.922	A	T	PmUG01_14017600	Africa	N(12),0(4)	N(65),0(36),1(7)
14	450708	4.856	5.922	T	G	PmUG01_14017600	Africa	N(12),0(4)	N(66),0(35),1(7)
14	450711	4.856	5.922	A	T	PmUG01_14017600	Africa	N(12),0(4)	N(69),0(32),1(7)
14	450720	4.856	5.922	A	G	PmUG01_14017600	Africa	N(11),0(5)	N(66),0(35),1(7)
14	450725	4.856	5.922	A	G	PmUG01_14017600	Africa	N(12),0(4)	N(73),0(30),1(5)
14	450728	4.856	5.922	C	T	PmUG01_14017600	Africa	N(12),0(4)	N(73),0(30),1(5)
14	450731	4.856	5.922	T	A	PmUG01_14017600	Africa	N(12),0(4)	N(72),0(31),1(5)
4	172772	-5.035	6.320	T	G	PmUG01_04012200	Asia	1(11),0(2),N(2),0.5(1)	1(67),0(33),N(7),0.5(1)
8	1598834	5.941	8.549	C	T	RNF1	Asia	0(14),1(2)	0(99),N(7),0.5(1),1(1)
9	2283754	7.708	13.894	G	A	PmUG01_09056900	Asia	0(7),N(7),0.5(1),1(1)	0(83),N(24),0.5(1)
10	1717410	4.353	4.871	G	A	PmUG01_10046700	Asia	0(8),N(6),0.5(1),1(1)	0(64),N(37),1(5),0.5(2)
12	1583381	5.482	7.376	G	A	PmUG01_12043500	Asia	0(12),0.5(2),N(2)	0(102),N(6)
13	123049	-5.043	6.338	T	A	PmUG01_13011300	Asia	N(11),1(3),0(2)	N(75),0(19),1(14)
14	706715	7.331	12.642	G	A	DBP6	Asia	0(14),0.5(2)	0(102),N(6)
14	3117387	-3.970	4.143	C	T	PmUG01_14076000	Asia	1(13),0(2),N(1)	0(106),N(2)

S6 Table. Signals of positive selection when comparing African and Asian isolates

Monoclonal isolates ($F_{WS} > 0.85$): Africa, n = 108; Asia, n = 14.

Chromosome	Position	RSB	LOGPVALUE	Reference	Alternate	Gene name	Comparison	Asia allele frequency (n = 16)	Africa allele frequency (n = 108)
3	819527	6.00392008	8.715322227	T	A	PmUG01_03025900	AfricaIAsia	N(13),1(2),0(1)	N(68),1(25),0(11),0.5(4)
3	819527	-6.0039201	8.715322227	T	A	PmUG01_03025900	AfricaIAsia	N(13),1(2),0(1)	N(68),1(25),0(11),0.5(4)
6	685163	4.78171984	5.759945751	A	G	NA	AfricaIAsia	N(11),0(3),0.5(1),1(1)	N(50),1(43),0(15)
6	685163	-4.7817198	5.759945751	A	G	NA	AfricaIAsia	N(11),0(3),0.5(1),1(1)	N(50),1(43),0(15)
6	685215	4.83524432	5.876206479	A	C	NA	AfricaIAsia	N(12),1(3),0(1)	N(57),1(39),0(12)
6	685215	-4.8352443	5.876206479	A	C	NA	AfricaIAsia	N(12),1(3),0(1)	N(57),1(39),0(12)
6	685287	4.52736768	5.223858577	G	A	NA	AfricaIAsia	N(10),0(3),1(3)	1(48),N(44),0(16)
6	685287	-4.5273677	5.223858577	G	A	NA	AfricaIAsia	N(10),0(3),1(3)	1(48),N(44),0(16)
6	690484	4.78863854	5.774906318	C	T	NA	AfricaIAsia	N(11),1(3),0(2)	N(44),1(39),0(25)
6	690484	-4.7886385	5.774906318	C	T	NA	AfricaIAsia	N(11),1(3),0(2)	N(44),1(39),0(25)
7	1362320	5.11118406	6.494652096	T	C	NA	AfricaIAsia	N(11),0(2),1(2),0.5(1)	N(49),0(45),1(12),0.5(2)
7	1362320	-5.1111841	6.494652096	T	C	NA	AfricaIAsia	N(11),0(2),1(2),0.5(1)	N(49),0(45),1(12),0.5(2)
8	1273125	9.55430185	20.90510996	T	C	NA	AfricaIAsia	N(11),0(4),0.5(1)	N(60),0(40),0.5(4),1(4)
8	1273125	-9.5543018	20.90510996	T	C	NA	AfricaIAsia	N(11),0(4),0.5(1)	N(60),0(40),0.5(4),1(4)
10	1293856	6.49873298	10.09151883	G	C	PmUG01_10036800	AfricaIAsia	N(13),0(2),1(1)	N(68),0(37),1(3)
10	1293856	-6.498733	10.09151883	G	C	PmUG01_10036800	AfricaIAsia	N(13),0(2),1(1)	N(68),0(37),1(3)
12	2985634	4.84655608	5.900930521	C	G	TREP	AfricaIAsia	N(13),1(2),0(1)	0(66),N(40),0.5(2)
12	2985634	-4.8465561	5.900930521	C	G	TREP	AfricaIAsia	N(13),1(2),0(1)	0(66),N(40),0.5(2)
13	1192166	4.73959435	5.669289006	A	T	NA	AfricaIAsia	N(13),1(2),0(1)	N(78),1(21),0(8),0.5(1)
13	1192166	-4.7395943	5.669289006	A	T	NA	AfricaIAsia	N(13),1(2),0(1)	N(78),1(21),0(8),0.5(1)
14	2769593	5.49430792	7.406441934	A	C	NA	AfricaIAsia	N(13),0(2),1(1)	N(64),0(33),1(10),0.5(1)
14	2769593	-5.4943079	7.406441934	A	C	NA	AfricaIAsia	N(13),0(2),1(1)	N(64),0(33),1(10),0.5(1)
14	3345798	14.5714489	47.36983348	A	T	NA	AfricaIAsia	N(13),0(2),0.5(1)	N(74),0(20),1(11),0.5(3)
14	3345798	-14.571449	47.36983348	A	T	NA	AfricaIAsia	N(13),0(2),0.5(1)	N(74),0(20),1(11),0.5(3)
14	3345799	5.91899439	8.489567572	A	T	NA	AfricaIAsia	N(11),0(3),1(2)	N(70),1(16),0(15),0.5(7)
14	3345799	-5.9189944	8.489567572	A	T	NA	AfricaIAsia	N(11),0(3),1(2)	N(70),1(16),0(15),0.5(7)

S7 Table. Genes used in screen to search for hypothetical resistance-associated mutations

Chromosome	Gene name	Gene ID
1	*crt*	PmUG01_01020700
2	*mrp1*	PmUG01_02011900
2	*ubp1*	PmUG01_02019300
4	*atg11*	PmUG01_04015700
5	*dhfr-ts*	PmUG01_05034700
6	*atg8*	PmUG01_06011300
6	*pi4k*	PmUG01_06015100
8	*atg18*	PmUG01_08028900
10	*k10*	PmUG01_10033800
10	*mdr1*	PmUG01_10021600
10	*pi3k*	PmUG01_10029300
11	*dhodh*	PmUG01_11059500
12	*k13*	PmUG01_12021200
12	*mdr2*	PmUG01_12069100
13	*atp4*	PmUG01_13021900
14	*coronin*	PmUG01_14084400
14	*dhps*	PmUG01_14045500
14	*mrp2*	PmUG01_14063400
14	*ferredoxin*	PmUG01_14034800
14	*ap2mu*	PmUG01_14053100

S8 Table. Missense SNPs found within resistance-associated genes in the *P. malariae* database and their frequencies in each continent.

Using centrifuge filtered isolates only (>0.9, n = 191)

Chromosome	Position	Reference	Effect	Gene name	Gene ID	Amino acid	Nucleotide 1	Nucleotide 2	Global F	Africa F	Asia F	Oceania F	SAM F
PmUG01_01_v1	553440	A	missense	CRT	PmUG01_01020700	278E>278D	553440A>T		0.06	0.01	0.37	0	0
PmUG01_01_v1	554001	G	missense	CRT	PmUG01_01020700	334S>334I	554001G>T		0.01	0.01	0	0	0
PmUG01_01_v1	554652	G	missense	CRT	PmUG01_01020700	383G>383V	554652G>T		0.02	0.01	0	0	0.67
PmUG01_02_v1	109509	C	missense	MRP1	PmUG01_02011900	1630D>1630N	109509C>T		0.72	0.7	0.76	1	1
PmUG01_02_v1	109692	T	missense	MRP1	PmUG01_02011900	1569N>1569D	109692T>C		0.01	0	0.06	0	0
PmUG01_02_v1	109698	T	missense	MRP1	PmUG01_02011900	1567N>1567H	109698T>G		0.01	0	0.06	0	0
PmUG01_02_v1	109707	C	missense	MRP1	PmUG01_02011900	1564V>1564L	109707C>G		0.01	0	0.05	0	0
PmUG01_02_v1	109712	G	missense	MRP1	PmUG01_02011900	1562S>1562C	109712G>C		0.01	0	0.06	0	0
PmUG01_02_v1	109730	C	missense	MRP1	PmUG01_02011900	1556S>1556N	109730C>T		0.01	0	0.05	0	0
PmUG01_02_v1	109745	T	missense	MRP1	PmUG01_02011900	1551K>1551R	109745T>C		0.01	0	0.05	0	0
PmUG01_02_v1	109756	A	missense	MRP1	PmUG01_02011900	1547H>1547Q	109756A>T		0.01	0	0.05	0	0
PmUG01_02_v1	109766	C	missense	MRP1	PmUG01_02011900	1544R>1544K	109766C>T		0.01	0	0.05	0	0
PmUG01_02_v1	109775	C	missense	MRP1	PmUG01_02011900	1541R>1541K	109775C>T		0.01	0	0.05	0	0
PmUG01_02_v1	109799	C	missense	MRP1	PmUG01_02011900	1533S>1533N	109799C>T		0.01	0	0.05	0	0
PmUG01_02_v1	109805	A	missense	MRP1	PmUG01_02011900	1531I>1531T	109804A>C	109805A>G	0.01	0	0.05	0	0
PmUG01_02_v1	109809	C	missense	MRP1	PmUG01_02011900	1530D>1530N	109809C>T		0.01	0	0.05	0	0
PmUG01_02_v1	109826	C	missense	MRP1	PmUG01_02011900	1524C>1524S	109825A>G	109826C>G	0.01	0	0.05	0	0
PmUG01_02_v1	109833	T	missense	MRP1	PmUG01_02011900	1522I>1522L	109833T>G		0.01	0	0.05	0	0
PmUG01_02_v1	110051	A	missense	MRP1	PmUG01_02011900	1449I>1449T	110051A>G		0.01	0.01	0	0	0
PmUG01_02_v1	110360	T	missense	MRP1	PmUG01_02011900	1346Q>1346L	110360T>A		0.01	0	0.05	0	0
PmUG01_02_v1	110514	C	missense	MRP1	PmUG01_02011900	1295V>1295I	110512A>T	110514C>T	0.01	0	0.05	0	0
PmUG01_02_v1	110517	T	missense	MRP1	PmUG01_02011900	1294S>1294D	110516C>T	110517T>C	0.01	0	0.05	0	0
PmUG01_02_v1	110523	A	missense	MRP1	PmUG01_02011900	1292Y>1292H	110523A>G		0.01	0	0.05	0	0
PmUG01_02_v1	110540	T	missense	MRP1	PmUG01_02011900	1286Y>1286F	110540T>A		0.01	0	0.05	0	0
PmUG01_02_v1	110573	T	missense	MRP1	PmUG01_02011900	1275N>1275S	110573T>C		0.01	0	0.05	0	0
PmUG01_02_v1	110624	G	missense	MRP1	PmUG01_02011900	1258A>1258V	110624G>A		0.01	0	0.05	0	0
PmUG01_02_v1	110674	A	missense	MRP1	PmUG01_02011900	1241F>1241L	110674A>T		0.01	0	0.05	0	0
PmUG01_02_v1	110683	T	missense	MRP1	PmUG01_02011900	1238L>1238F	110683T>A		0.01	0	0.05	0	0
PmUG01_02_v1	110697	T	missense	MRP1	PmUG01_02011900	1234I>1234A	110696A>G	110697T>C	0.01	0	0.05	0	0
PmUG01_02_v1	110700	C	missense	MRP1	PmUG01_02011900	1233V>1233I	110700C>T		0.01	0	0.05	0	0
PmUG01_02_v1	110703	A	missense	MRP1	PmUG01_02011900	1232S>1232A	110703A>C		0.01	0	0.05	0	0
PmUG01_02_v1	110717	A	missense	MRP1	PmUG01_02011900	1227F>1227C	110716A>G	110717A>C	0.01	0	0.05	0	0
PmUG01_02_v1	110737	A	missense	MRP1	PmUG01_02011900	1220F>1220L	110737A>C		0.01	0	0.05	0	0
PmUG01_02_v1	110775	T	missense	MRP1	PmUG01_02011900	1208I>1208F	110775T>A		0.01	0	0.05	0	0
PmUG01_02_v1	111157	T	missense	MRP1	PmUG01_02011900	1080R>1080S	111157T>A		0.01	0	0.05	0	0
PmUG01_02_v1	111160	A	missense	MRP1	PmUG01_02011900	1079F>1079L	111160A>T		0.01	0	0.05	0	0
PmUG01_02_v1	111355	A	missense	MRP1	PmUG01_02011900	1014N>1014K	111355A>T		0.01	0.01	0	0	0
PmUG01_02_v1	112100	C	missense	MRP1	PmUG01_02011900	766R>766K	112099C>T	112100C>T	0.01	0	0.06	0	0
PmUG01_02_v1	112130	T	missense	MRP1	PmUG01_02011900	756N>756S	112130T>C		0.01	0	0.05	0	0
PmUG01_02_v1	112164	C	missense	MRP1	PmUG01_02011900	745A>745Y	112163G>T	112164C>A	0.01	0	0.05	0	0
PmUG01_02_v1	112215	A	missense	MRP1	PmUG01_02011900	728S>728A	112215A>C		0.02	0	0.15	0	0
PmUG01_02_v1	112382	C	missense	MRP1	PmUG01_02011900	672G>672D	112382C>T		0.01	0.01	0	0	0
PmUG01_02_v1	112620	G	missense	MRP1	PmUG01_02011900	593Q>593E	112620G>C		0.01	0	0.05	0	0
PmUG01_02_v1	113291	T	missense	MRP1	PmUG01_02011900	369N>369S	113290A>G	113291T>C	0.01	0	0.05	0	0
PmUG01_02_v1	113328	T	missense	MRP1	PmUG01_02011900	357M>357V	113328T>C		0.01	0	0.05	0	0
PmUG01_02_v1	113615	T	missense	MRP1	PmUG01_02011900	261K>261R	113615T>C		0.01	0.01	0	0	0
PmUG01_02_v1	113630	T	missense	MRP1	PmUG01_02011900	256N>256I	113630T>A		0.01	0.01	0	0	0
PmUG01_02_v1	113802	T	missense	MRP1	PmUG01_02011900	199M>199L	113802T>A		0.01	0	0.05	0	0
PmUG01_02_v1	113850	G	missense	MRP1	PmUG01_02011900	183L>183F	113850G>A		0.01	0	0.05	0	0
PmUG01_02_v1	113859	T	missense	MRP1	PmUG01_02011900	180I>180L	113859T>A		0.01	0	0.06	0	0

Chromosome	Position	Reference	Effect	Gene name	Gene ID	Amino acid	Nucleotide 1	Nucleotide 2	Global F	Africa F	Asia F	Oceania F	SAM F
PmUG01_02_v1	113862	A	missense	MRP1	PmUG01_02011900	179F>179L	113862A>G		0.01	0	0.05	0	0
PmUG01_02_v1	113863	T	missense	MRP1	PmUG01_02011900	178L>178F	113863T>G		0.01	0	0.05	0	0
PmUG01_02_v1	113870	G	missense	MRP1	PmUG01_02011900	176A>176V	113869A>C	113870G>A	0.01	0	0.05	0	0
PmUG01_02_v1	114272	T	missense	MRP1	PmUG01_02011900	42N>42S	114272T>C		0.01	0.01	0	0	0
PmUG01_02_v1	444391	C	missense	UBP1	PmUG01_02019300	4118V>4118I	444391C>T		0.01	0.01	0	0	0
PmUG01_02_v1	445404	C	missense	UBP1	PmUG01_02019300	3898G>3898D	445404C>T		0.02	0.02	0	0	0
PmUG01_02_v1	446030	T	missense	UBP1	PmUG01_02019300	3689L>3689F	446030T>G		0.01	0	0.08	0	0
PmUG01_02_v1	446875	G	missense	UBP1	PmUG01_02019300	3432S>3432F	446875G>A		0.01	0.01	0	0	0
PmUG01_02_v1	446956	T	missense	UBP1	PmUG01_02019300	3405D>3405G	446956T>C		0.04	0.04	0	0	0.5
PmUG01_02_v1	447814	C	missense	UBP1	PmUG01_02019300	3119S>3119N	447814C>T		0.01	0	0.07	0	0
PmUG01_02_v1	447912	C	missense	UBP1	PmUG01_02019300	3086R>3086S	447912C>A		0.01	0	0.08	0	0
PmUG01_02_v1	449660	A	missense	UBP1	PmUG01_02019300	2642V>2642E	449660A>T		0.01	0.01	0	0	0
PmUG01_02_v1	449768	C	missense	UBP1	PmUG01_02019300	2606G>2606V	449768C>A		0.03	0.03	0	0	0
PmUG01_02_v1	449775	T	missense	UBP1	PmUG01_02019300	2604T>2604V	449774G>A	449775T>C	0.02	0.02	0	0	0
PmUG01_02_v1	450276	T	missense	UBP1	PmUG01_02019300	2437N>2437Y	450276T>A		0.01	0.01	0	0	0
PmUG01_02_v1	450288	A	missense	UBP1	PmUG01_02019300	2433L>2433I	450286T>G	450288A>T	0.01	0.01	0	0	0
PmUG01_02_v1	450665	A	missense	UBP1	PmUG01_02019300	2307I>2307S	450665A>C		1	1	1	0	0
PmUG01_02_v1	451172	A	missense	UBP1	PmUG01_02019300	2138M>2138K	451171C>T	451172A>T	0.02	0.03	0	0	0
PmUG01_02_v1	451484	C	missense	UBP1	PmUG01_02019300	2034S>2034I	451484C>A		0.02	0.02	0	0	0
PmUG01_02_v1	451512	G	missense	UBP1	PmUG01_02019300	2025R>2025C	451512G>A		0.01	0.01	0	0	0
PmUG01_02_v1	451593	C	missense	UBP1	PmUG01_02019300	1998D>1998Y	451593C>A		0.03	0.04	0	0	0
PmUG01_02_v1	452025	G	missense	UBP1	PmUG01_02019300	1854R>1854C	452025G>A		0.01	0.01	0	0	0
PmUG01_02_v1	452031	C	missense	UBP1	PmUG01_02019300	1852G>1852S	452031C>T		0.01	0.01	0	0	0
PmUG01_02_v1	452124	C	missense	UBP1	PmUG01_02019300	1821D>1821N	452124C>T		0.01	0.01	0	0	0
PmUG01_02_v1	452184	A	missense	UBP1	PmUG01_02019300	1801L>1801V	452184A>C		0.08	0.11	0	0	0
PmUG01_02_v1	452198	C	missense	UBP1	PmUG01_02019300	1796G>1796E	452198C>T		0.06	0.01	0.62	0	0
PmUG01_02_v1	452210	C	missense	UBP1	PmUG01_02019300	1792G>1792E	452210C>T		0.04	0	0.4	0	0
PmUG01_02_v1	452449	G	missense	UBP1	PmUG01_02019300	1712S>1712R	452449G>T		0.03	0.04	0	0	0
PmUG01_02_v1	452733	A	missense	UBP1	PmUG01_02019300	1618Y>1618H	452733A>G		0.02	0.02	0	0	0
PmUG01_02_v1	452792	T	missense	UBP1	PmUG01_02019300	1598N>1598S	452792T>C		0.44	0.38	0.93	0	0.2
PmUG01_02_v1	452841	G	missense	UBP1	PmUG01_02019300	1582H>1582Y	452841G>A		0.02	0	0.14	0	0
PmUG01_02_v1	453050	T	missense	UBP1	PmUG01_02019300	1512Q>1512R	453050T>C		0.02	0.02	0	0	0
PmUG01_02_v1	454427	A	missense	UBP1	PmUG01_02019300	1053M>1053R	454427A>C		0.07	0.09	0	0	0
PmUG01_02_v1	454431	G	missense	UBP1	PmUG01_02019300	1052H>1052E	454429A>T	454431G>C	0.1	0.12	0	0	0
PmUG01_02_v1	454444	A	missense	UBP1	PmUG01_02019300	1047D>1047E	454444A>T		0.08	0.09	0	0	0
PmUG01_02_v1	454518	C	missense	UBP1	PmUG01_02019300	1023G>1023R	454516T>C	454518C>T	0.01	0.01	0	0	0
PmUG01_02_v1	455135	C	missense	UBP1	PmUG01_02019300	817S>817N	455135C>T		0.05	0.05	0.08	0	0
PmUG01_02_v1	455169	C	missense	UBP1	PmUG01_02019300	806G>806C	455169C>A		0.01	0.01	0	0	0
PmUG01_02_v1	455675	T	missense	UBP1	PmUG01_02019300	637K>637R	455675T>C		0.01	0.02	0	0	0
PmUG01_02_v1	455715	C	missense	UBP1	PmUG01_02019300	624V>624I	455715C>T	455715C>A	0.03	0.03	0	0	0
PmUG01_02_v1	455934	T	missense	UBP1	PmUG01_02019300	551S>551C	455934T>A		0.01	0.01	0	0	0
PmUG01_02_v1	456030	T	missense	UBP1	PmUG01_02019300	519I>519V	456030T>C		0.05	0.01	0.29	0	0
PmUG01_02_v1	456437	T	missense	UBP1	PmUG01_02019300	383K>383T	456437T>G		0.01	0	0.05	0	0
PmUG01_02_v1	456892	G	missense	UBP1	PmUG01_02019300	231D>231E	456892G>T		0.01	0.01	0	0	0
PmUG01_02_v1	457261	C	missense	UBP1	PmUG01_02019300	108M>108I	457261C>T		0.01	0.01	0	0	0
PmUG01_02_v1	457433	T	missense	UBP1	PmUG01_02019300	51K>51R	457433T>C		0.01	0	0.06	0	0
PmUG01_05_v1	1291895	G	missense	DHFR-TS	PmUG01_05034700	15A>15S	1291895G>T		0.05	0.04	0	0	0.33
PmUG01_05_v1	1291999	T	missense	DHFR-TS	PmUG01_05034700	49S>49R	1291999T>A		0.01	0	0.05	0	0
PmUG01_05_v1	1292023	C	missense	DHFR-TS	PmUG01_05034700	57F>57L	1292023C>G	1292023C>A	0.17	0.21	0	0	0
PmUG01_05_v1	1292026		missense	DHFR-TS	PmUG01_05034700	58R>58S	1292026A>C		0.82	0.78	1	1	1

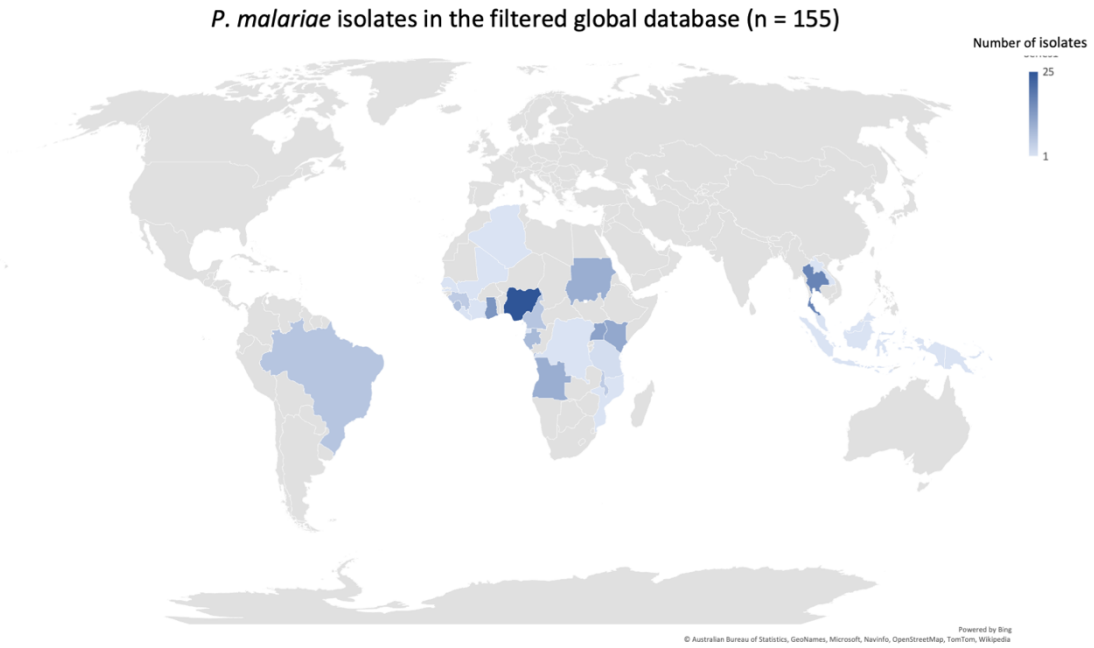
Chromosome	Position	Reference	Effect	Gene name	Gene ID	Amino acid	Nucleotide 1	Nucleotide 2	Global F	Africa F	Asia F	Oceania F	SAM F
PmUG01_05_v1	1292193	A	missense	DHFR-TS	PmUG01_05034700	114N>114S	1292193A>G		0.71	0.67	0.84	1	1
PmUG01_05_v1	1292522	T	missense	DHFR-TS	PmUG01_05034700	224C>224S	1292522T>A		0.01	0.01	0	0	0
PmUG01_05_v1	1293224	C	missense	DHFR-TS	PmUG01_05034700	458H>458Y	1293224C>T		0.01	0.01	0	0	0
PmUG01_05_v1	1293228	C	missense	DHFR-TS	PmUG01_05034700	459A>459D	1293228C>A	1293229C>T	0.01	0.01	0	0	0
PmUG01_05_v1	1293239	G	missense	DHFR-TS	PmUG01_05034700	463D>463N	1293239G>A		0.01	0.01	0	0	0
PmUG01_05_v1	1293272	C	missense	DHFR-TS	PmUG01_05034700	474H>474N	1293272C>A		0.01	0.01	0	0	0
PmUG01_05_v1	1293308	A	missense	DHFR-TS	PmUG01_05034700	486I>486L	1293308A>C		0.01	0.01	0	0	0
PmUG01_05_v1	1293527	A	missense	DHFR-TS	PmUG01_05034700	559I>559V	1293527A>G		0.01	0.01	0	0	0
PmUG01_05_v1	1293575	G	missense	DHFR-TS	PmUG01_05034700	575V>575I	1293575G>A	1293577A>T	0.01	0.01	0	0	0
PmUG01_05_v1	1293628	A	missense	DHFR-TS	PmUG01_05034700	592E>592D	1293628A>T		0.01	0.01	0	0	0
PmUG01_05_v1	1293691	T	missense	DHFR-TS	PmUG01_05034700	613D>613E	1293691T>A		0.01	0.01	0	0	0
PmUG01_06_v1	296253	A	missense	PI4K	PmUG01_06015100	1535L>1535Q	296253A>T		0.01	0.02	0	0	0
PmUG01_06_v1	296383	C	missense	PI4K	PmUG01_06015100	1492D>1492N	296383C>T		0.01	0.01	0	0	0
PmUG01_06_v1	296433	T	missense	PI4K	PmUG01_06015100	1475D>1475G	296433T>C		0.01	0.02	0	0	0
PmUG01_06_v1	296697	A	missense	PI4K	PmUG01_06015100	1387I>1387T	296697A>G		0.01	0.01	0	0	0
PmUG01_06_v1	297526	C	missense	PI4K	PmUG01_06015100	1111E>1111K	297526C>T		0.06	0.07	0	0	0
PmUG01_06_v1	297550	C	missense	PI4K	PmUG01_06015100	1103E>1103K	297550C>T		0.03	0.04	0	0	0
PmUG01_06_v1	297562	T	missense	PI4K	PmUG01_06015100	1099K>1099E	297562T>C		0.06	0.07	0	0	0
PmUG01_06_v1	297567	A	missense	PI4K	PmUG01_06015100	1097V>1097E	297567A>T		0.06	0.07	0	0	0
PmUG01_06_v1	297579	A	missense	PI4K	PmUG01_06015100	1093V>1093E	297579A>T		0.02	0.03	0	0	0
PmUG01_06_v1	297598	C	missense	PI4K	PmUG01_06015100	1087E>1087K	297598C>T		0.2	0.19	0.36	0	0
PmUG01_06_v1	298251	A	missense	PI4K	PmUG01_06015100	869F>869Y	298251A>T		0.01	0.01	0	0	0
PmUG01_06_v1	298259	T	missense	PI4K	PmUG01_06015100	866K>866N	298259T>A		0.01	0.01	0	0	0
PmUG01_06_v1	298529	C	missense	PI4K	PmUG01_06015100	776L>776F	298529C>A		0.03	0.03	0	0	0
PmUG01_06_v1	298842	A	missense	PI4K	PmUG01_06015100	672I>672N	298842A>T		0.01	0.01	0	0	0
PmUG01_06_v1	298897	T	missense	PI4K	PmUG01_06015100	654N>654D	298897T>C		0.01	0.01	0	0	0
PmUG01_06_v1	298924	T	missense	PI4K	PmUG01_06015100	645K>645E	298924T>C		0.01	0.01	0	0	0
PmUG01_06_v1	298930	T	missense	PI4K	PmUG01_06015100	643N>643S	298929T>G	298930T>A	0.01	0.01	0	0	0
PmUG01_06_v1	299047	T	missense	PI4K	PmUG01_06015100	604I>604L	299047T>A		0.01	0.01	0	0	0
PmUG01_06_v1	299052	C	missense	PI4K	PmUG01_06015100	602C>602S	299051G>T	299052C>G	0.01	0.01	0	0	0
PmUG01_06_v1	299090	A	missense	PI4K	PmUG01_06015100	589D>589E	299090A>T		0.01	0.01	0	0	0
PmUG01_06_v1	299805	T	missense	PI4K	PmUG01_06015100	351D>351V	299805T>A		0.01	0.01	0	0	0
PmUG01_06_v1	299958	G	missense	PI4K	PmUG01_06015100	300T>300I	299958G>A		0.01	0.01	0	0	0
PmUG01_06_v1	300076	C	missense	PI4K	PmUG01_06015100	261A>261S	300074T>A	300076C>A	0.43	0.47	0.3	0	0
PmUG01_06_v1	300093	C	missense	PI4K	PmUG01_06015100	255G>255D	300092T>G	300093C>T	0.12	0.12	0.11	0	0
PmUG01_06_v1	300135	C	missense	PI4K	PmUG01_06015100	241G>241D	300134T>G	300135C>T	0.29	0.3	0.25	0	0
PmUG01_06_v1	300246	C	missense	PI4K	PmUG01_06015100	204S>204N	300246C>T		0.02	0.02	0	0	0
PmUG01_10_v1	524859	T	missense	MDR1	PmUG01_10021600	1353N>1353S	524859T>C		0.01	0.01	0	0	0
PmUG01_10_v1	524938	A	missense	MDR1	PmUG01_10021600	1327C>1327S	524938A>T		0.01	0.01	0	0	0
PmUG01_10_v1	525386	T	missense	MDR1	PmUG01_10021600	1177E>1177D	525386T>A		0.01	0	0.05	0	0
PmUG01_10_v1	525728	T	missense	MDR1	PmUG01_10021600	1063L>1063F	525728T>G		0.01	0	0.1	0	0
PmUG01_10_v1	526006	T	missense	MDR1	PmUG01_10021600	971I>971V	526006T>C		0.01	0.01	0	0	0
PmUG01_10_v1	526187	C	missense	MDR1	PmUG01_10021600	910L>910F	526187C>G		0.01	0	0.05	0	0
PmUG01_10_v1	526429	T	missense	MDR1	PmUG01_10021600	830I>830L	526429T>A		0.01	0	0.05	0	0
PmUG01_10_v1	526684	T	missense	MDR1	PmUG01_10021600	745I>745V	526684T>C		0.01	0.02	0	0	0
PmUG01_10_v1	527184	A	missense	MDR1	PmUG01_10021600	578L>578H	527184A>T		0.01	0.02	0	0	0
PmUG01_10_v1	527449	G	missense	MDR1	PmUG01_10021600	490L>490I	527449G>T		0.01	0	0.05	0.5	0
PmUG01_10_v1	527550	C	missense	MDR1	PmUG01_10021600	456R>456T	527550C>G		0.01	0.01	0	0	0
PmUG01_10_v1	527555	C	missense	MDR1	PmUG01_10021600	454W>454C	527555C>G		0.01	0.01	0	0	0
PmUG01_10_v1	528430	C	missense	MDR1	PmUG01_10021600	163V>163I	528430C>T		0.01	0	0.05	0	0

Chromosome	Position	Reference	Effect	Gene name	Gene ID	Amino acid	Nucleotide 1	Nucleotide 2	Global F	Africa F	Asia F	Oceania F	SAM F
PmUG01_10_v1	528446	T	missense	MDR1	PmUG01_10021600	157E>157D	528446T>A		0.01	0	0.05	0	0
PmUG01_10_v1	528475	A	missense	MDR1	PmUG01_10021600	148S>148P	528473A>T	528475A>G	0.01	0	0.05	0	0
PmUG01_10_v1	528493	G	missense	MDR1	PmUG01_10021600	142Q>142E	528491C>T	528493G>C	0.01	0	0.05	0	0
PmUG01_10_v1	528526	C	missense	MDR1	PmUG01_10021600	131E>131K	528526C>T		0.01	0	0.05	0	0
PmUG01_10_v1	528529	T	missense	MDR1	PmUG01_10021600	130I>130L	528527A>T	528529T>A	0.01	0	0.05	0	0
PmUG01_10_v1	528559	C	missense	MDR1	PmUG01_10021600	120V>120I	528559C>T		0.01	0	0.05	0	0
PmUG01_10_v1	528562	C	missense	MDR1	PmUG01_10021600	119V>119T	528561A>G	528562C>T	0.01	0	0.05	0	0
PmUG01_10_v1	528572	A	missense	MDR1	PmUG01_10021600	115F>115L	528572A>T		0.01	0	0.05	0	0
PmUG01_10_v1	528595	T	missense	MDR1	PmUG01_10021600	108I>108V	528593T>C	528595T>C	0.01	0	0.05	0	0
PmUG01_10_v1	528613	T	missense	MDR1	PmUG01_10021600	102I>102V	528613T>C		0.01	0	0.05	0	0
PmUG01_10_v1	528616	A	missense	MDR1	PmUG01_10021600	101L>101G	528615A>C	528616A>C	0.01	0	0.05	0	0
PmUG01_10_v1	1139798	A	missense	Kelch10	PmUG01_10033800	48M>48L	1139798A>T	1139800G>A	0.01	0.01	0	0	0
PmUG01_10_v1	1139832	A	missense	Kelch10	PmUG01_10033800	59Y>59F	1139832A>T		0.01	0.01	0	0	0
PmUG01_10_v1	1140420	A	missense	Kelch10	PmUG01_10033800	255N>255T	1140420A>C		0.01	0.01	0	0	0
PmUG01_10_v1	1140672	C	missense	Kelch10	PmUG01_10033800	339A>339V	1140672C>T		0.01	0	0	0	0.17
PmUG01_10_v1	1141250	C	missense	Kelch10	PmUG01_10033800	532H>532Y	1141250C>T		0.01	0	0.05	0	0
PmUG01_10_v1	1141379	G	missense	Kelch10	PmUG01_10033800	575G>575R	1141379G>C		0.03	0.04	0	0	0
PmUG01_10_v1	1141518	A	missense	Kelch10	PmUG01_10033800	621D>621G	1141518A>G		0.01	0.01	0	0	0
PmUG01_10_v1	1141670	G	missense	Kelch10	PmUG01_10033800	672V>672F	1141670G>T		0.01	0.01	0	0	0
PmUG01_10_v1	1142028	A	missense	Kelch10	PmUG01_10033800	791K>791R	1142028A>G		0.01	0.01	0	0	0
PmUG01_10_v1	1142327	G	missense	Kelch10	PmUG01_10033800	891E>891K	1142327G>A		0.01	0.01	0	0	0
PmUG01_10_v1	1142328	A	missense	Kelch10	PmUG01_10033800	891E>891A	1142328A>C		0.01	0.01	0	0	0
PmUG01_11_v1	2491536	A	missense	DHODH	PmUG01_11059500	20Y>20F	2491536A>T		0.06	0.07	0	0	0
PmUG01_11_v1	2491871	C	missense	DHODH	PmUG01_11059500	132P>132S	2491871C>T		0.01	0.02	0	0	0
PmUG01_11_v1	2493263	A	missense	DHODH	PmUG01_11059500	596R>596G	2493263A>G		0.01	0	0.05	0	0
PmUG01_12_v1	575732	A	missense	Kelch13	PmUG01_12021200	156M>156V	575732A>G		0.01	0	0.05	0	0
PmUG01_12_v1	576380	T	missense	Kelch13	PmUG01_12021200	372Y>372H	576380T>C		0.01	0.01	0	0	0
PmUG01_12_v1	576476	G	missense	Kelch13	PmUG01_12021200	404V>404I	576476G>A	576478T>A	0.01	0.01	0	0	0
PmUG01_12_v1	576779	A	missense	Kelch13	PmUG01_12021200	505T>505V	576779A>G	576780C>T	0.01	0.01	0	0	0
PmUG01_12_v1	576786	T	missense	Kelch13	PmUG01_12021200	507F>507Y	576786T>A		0.01	0.01	0	0	0
PmUG01_12_v1	577139	G	missense	Kelch13	PmUG01_12021200	625V>625I	577139G>A		0.01	0.02	0	0	0
PmUG01_12_v1	2740223	T	missense	MDR2	PmUG01_12069100	426V>426A	2740223T>C		0.7	0.72	0.72	0.5	0.5
PmUG01_12_v1	2740398	A	missense	MDR2	PmUG01_12069100	484L>484F	2740398A>T		0.07	0.08	0	1	0
PmUG01_12_v1	2741953	G	missense	MDR2	PmUG01_12069100	1003G>1003R	2741953G>A		0.07	0.08	0	0	0
PmUG01_12_v1	2741992	G	missense	MDR2	PmUG01_12069100	1016V>1016F	2741992G>T		0.47	0.52	0.1	0	0.8
PmUG01_12_v1	2742029	T	missense	MDR2	PmUG01_12069100	1028V>1028A	2742029T>C		1	1	1	1	1
PmUG01_12_v1	2742326	A	missense	MDR2	PmUG01_12069100	1127N>1127S	2742326A>G		0.27	0.21	0.74	0	0
PmUG01_12_v1	2742474	G	missense	MDR2	PmUG01_12069100	1176M>1176I	2742474G>A		0.01	0.01	0	0	0
PmUG01_12_v1	2742491	A	missense	MDR2	PmUG01_12069100	1182Y>1182C	2742491A>G		0.01	0.01	0	0	0
PmUG01_12_v1	2742583	A	missense	MDR2	PmUG01_12069100	1213N>1213Y	2742583A>T		0.03	0	0.25	0	0
PmUG01_12_v1	2742706	C	missense	MDR2	PmUG01_12069100	1254Q>1254K	2742706C>A		0.01	0.01	0	0	0
PmUG01_12_v1	2742708	G	missense	MDR2	PmUG01_12069100	1254Q>1254H	2742708G>T		0.01	0.01	0	0	0
PmUG01_12_v1	2742729	C	missense	MDR2	PmUG01_12069100	1261S>1261R	2742729C>A		0.01	0.01	0	0	0
PmUG01_12_v1	2742773	G	missense	MDR2	PmUG01_12069100	1276S>1276N	2742773G>A		0.01	0	0.11	0	0
PmUG01_12_v1	2742775	G	missense	MDR2	PmUG01_12069100	1277G>1277S	2742775G>A		0.01	0	0.11	0	0
PmUG01_12_v1	2742779	A	missense	MDR2	PmUG01_12069100	1278N>1278S	2742779A>G		0.01	0	0.11	0	0
PmUG01_12_v1	2742956	G	missense	MDR2	PmUG01_12069100	1337S>1337N	2742956G>A		0.02	0.02	0	0	0
PmUG01_13_v1	664406	T	missense	ATP4	PmUG01_13021900	1378I>1378V	664406T>C		0.01	0.01	0	0	0
PmUG01_13_v1	664631	T	missense	ATP4	PmUG01_13021900	1303K>1303E	664629C>T	664631T>C	0.01	0.01	0	0	0
PmUG01_13_v1	664675	C	missense	ATP4	PmUG01_13021900	1288G>1288A	664675C>G		0.01	0.01	0	0	0

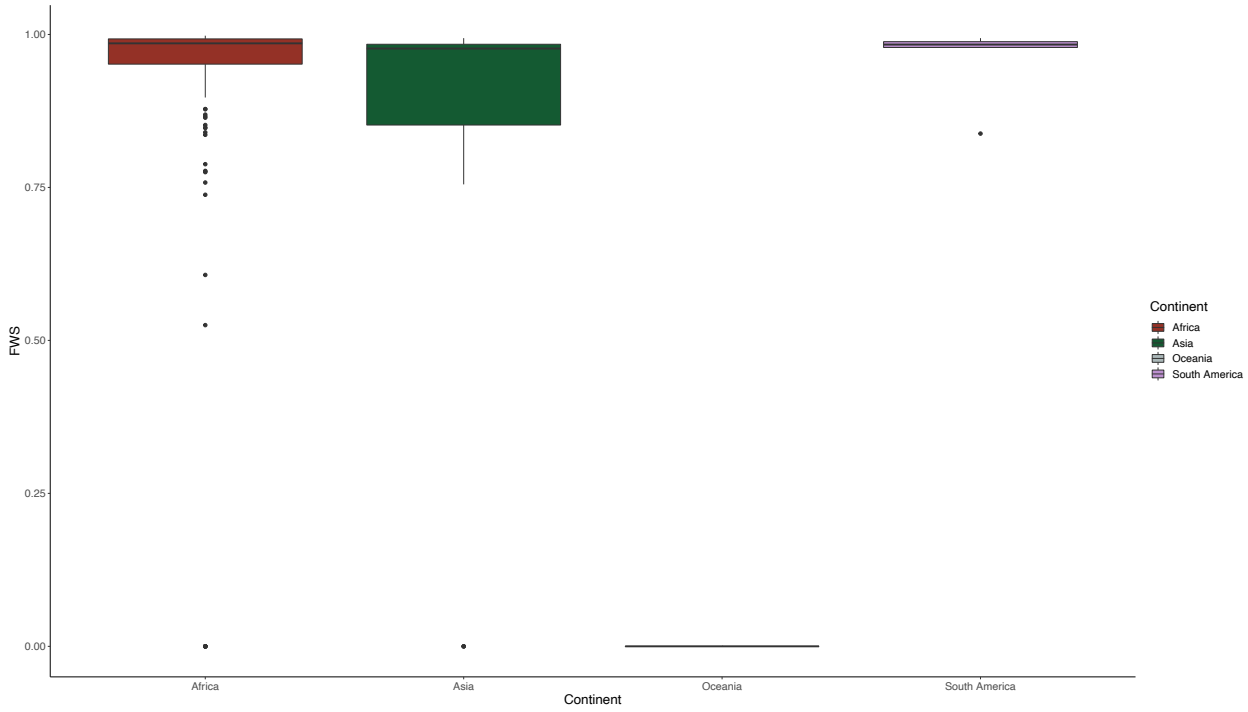
Chromosome	Position	Reference	Effect	Gene name	Gene ID	Amino acid	Nucleotide 1	Nucleotide 2	Global F	Africa F	Asia F	Oceania F	SAM F
PmUG01_13_v1	664689	T	missense	ATP4	PmUG01_13021900	1283I>1283M	664689T>C		0.01	0.01	0	0	0
PmUG01_13_v1	665761	G	missense	ATP4	PmUG01_13021900	926S>926C	665761G>C		0.01	0.01	0	0	0
PmUG01_13_v1	665933	T	missense	ATP4	PmUG01_13021900	869I>869V	665933T>C		0.01	0.01	0	0	0
PmUG01_13_v1	666164	C	missense	ATP4	PmUG01_13021900	792V>792I	666162A>T	666164C>T	0.01	0.01	0	0	0
PmUG01_13_v1	666167	C	missense	ATP4	PmUG01_13021900	791A>791S	666165A>G	666167C>A	0.01	0.01	0	0	0
PmUG01_13_v1	666187	C	missense	ATP4	PmUG01_13021900	784R>784Q	666187C>T		0.01	0.01	0	0	0
PmUG01_13_v1	666207	G	missense	ATP4	PmUG01_13021900	777N>777K	666207G>T		0.01	0.01	0	0	0
PmUG01_13_v1	666353	C	missense	ATP4	PmUG01_13021900	729E>729K	666353C>T		0.01	0	0	0	0.17
PmUG01_13_v1	666354	A	missense	ATP4	PmUG01_13021900	728N>728K	666354A>T		0.01	0	0	0	0.17
PmUG01_13_v1	666359	C	missense	ATP4	PmUG01_13021900	727E>727K	666359C>T		0.01	0	0	0	0.17
PmUG01_13_v1	666360	A	missense	ATP4	PmUG01_13021900	726N>726K	666360A>T		0.01	0	0	0	0.17
PmUG01_13_v1	666506	G	missense	ATP4	PmUG01_13021900	678L>678I	666506G>T		0.01	0.01	0	0	0
PmUG01_13_v1	666511	T	missense	ATP4	PmUG01_13021900	676N>676S	666511T>C		0.01	0.01	0	0	0
PmUG01_13_v1	666521	T	missense	ATP4	PmUG01_13021900	673T>673V	666520G>A	666521T>C	0.01	0.01	0	0	0
PmUG01_13_v1	666590	A	missense	ATP4	PmUG01_13021900	650S>650P	666588C>G	666590A>G	0.01	0.01	0	0	0
PmUG01_13_v1	667061	C	missense	ATP4	PmUG01_13021900	493V>493I	667059T>A	667061C>T	0.01	0.01	0	0	0
PmUG01_13_v1	667099	G	missense	ATP4	PmUG01_13021900	480A>480V	667099G>A		0.01	0.01	0	0	0
PmUG01_13_v1	667140	A	missense	ATP4	PmUG01_13021900	466N>466K	667140A>T		0.01	0.01	0	0	0
PmUG01_13_v1	667159	G	missense	ATP4	PmUG01_13021900	460T>460S	667158A>G	667159G>C	0.01	0.01	0	0	0
PmUG01_13_v1	667169	T	missense	ATP4	PmUG01_13021900	457I>457V	667169T>C		0.01	0.01	0	0	0
PmUG01_13_v1	667172	T	missense	ATP4	PmUG01_13021900	456I>456V	667172T>C		0.01	0.01	0	0	0
PmUG01_13_v1	667181	C	missense	ATP4	PmUG01_13021900	453V>453I	667181C>T		0.01	0.01	0	0	0
PmUG01_13_v1	667213	C	missense	ATP4	PmUG01_13021900	442R>442K	667213C>T		0.01	0.01	0	0	0
PmUG01_13_v1	667226	T	missense	ATP4	PmUG01_13021900	438I>438V	667226T>C		0.01	0.01	0	0	0
PmUG01_13_v1	667336	T	missense	ATP4	PmUG01_13021900	401N>401S	667335A>G	667336T>C	0.01	0.01	0	0	0
PmUG01_13_v1	667373	A	missense	ATP4	PmUG01_13021900	389S>389A	667373A>C		0.01	0.01	0	0	0
PmUG01_13_v1	667385	T	missense	ATP4	PmUG01_13021900	385T>385S	667385T>A		0.01	0.01	0	0	0
PmUG01_13_v1	667387	T	missense	ATP4	PmUG01_13021900	384Y>384L	667386A>C	667387T>A	0.01	0.01	0	0	0
PmUG01_13_v1	667391	C	missense	ATP4	PmUG01_13021900	383D>383N	667391C>T		0.01	0.01	0	0	0
PmUG01_13_v1	667397	G	missense	ATP4	PmUG01_13021900	381P>381A	667395C>T	667397G>C	0.01	0.01	0	0	0
PmUG01_13_v1	667400	T	missense	ATP4	PmUG01_13021900	380I>380V	667398A>T	667400T>C	0.01	0.01	0	0	0
PmUG01_13_v1	667467	T	missense	ATP4	PmUG01_13021900	357E>357D	667467T>A		0.01	0.01	0	0	0
PmUG01_13_v1	667472	C	missense	ATP4	PmUG01_13021900	356V>356F	667470T>A	667472C>A	0.01	0.01	0	0	0
PmUG01_13_v1	667514	T	missense	ATP4	PmUG01_13021900	342I>342L	667514T>G		0.01	0.01	0	0	0
PmUG01_13_v1	667556	T	missense	ATP4	PmUG01_13021900	328I>328V	667556T>C		0.01	0.01	0	0	0
PmUG01_13_v1	667997	C	missense	ATP4	PmUG01_13021900	181A>181S	667997C>A		0.01	0.01	0	0	0
PmUG01_13_v1	667999	C	missense	ATP4	PmUG01_13021900	180R>180T	667999C>G		0.01	0.01	0	0	0
PmUG01_13_v1	668408	A	missense	ATP4	PmUG01_13021900	44Y>44N	668408A>T		0.01	0.01	0	0	0
PmUG01_13_v1	668505	A	missense	ATP4	PmUG01_13021900	11N>11K	668505A>T		0.07	0	0.52	0.5	0
PmUG01_14_v1	1656713	A	missense	PPPK-DHPS	PmUG01_14045500	451I>451T	1656713A>G		0.01	0.01	0	0	0
PmUG01_14_v1	1656842	A	missense	PPPK-DHPS	PmUG01_14045500	408M>408T	1656842A>G		0.02	0.02	0	0	0
PmUG01_14_v1	1656941	G	missense	PPPK-DHPS	PmUG01_14045500	375A>375G	1656941G>C		0.13	0.15	0	0	0.17
PmUG01_14_v1	1656944	G	missense	PPPK-DHPS	PmUG01_14045500	374S>374F	1656944G>A		0.01	0.01	0	0	0
PmUG01_14_v1	1656945	A	missense	PPPK-DHPS	PmUG01_14045500	374S>374H	1656944G>T	1656945A>G	0.01	0.01	0	0	0
PmUG01_14_v1	1657487	C	missense	PPPK-DHPS	PmUG01_14045500	193R>193I	1657487C>A		0.13	0.16	0	0	0
PmUG01_14_v1	1657652	G	missense	PPPK-DHPS	PmUG01_14045500	138T>138I	1657652G>A		0.01	0.01	0	0	0
PmUG01_14_v1	1657704	C	missense	PPPK-DHPS	PmUG01_14045500	121V>121M	1657704C>T		0.01	0.01	0	0	0
PmUG01_14_v1	1936058	G	missense	AP2-MU	PmUG01_14053100	103V>103I	1936058G>A	1936060C>A	0.01	0	0.05	0	0
PmUG01_14_v1	1936443	C	missense	AP2-MU	PmUG01_14053100	231P>231Q	1936443C>A	1936444C>A	0.01	0.01	0	0	0
PmUG01_14_v1	1936451	A	missense	AP2-MU	PmUG01_14053100	234N>234H	1936451A>C		0.01	0.01	0	0	0

Chromosome	Position	Reference	Effect	Gene name	Gene ID	Amino acid	Nucleotide 1	Nucleotide 2	Global F	Africa F	Asia F	Oceania F	SAM F
PmUG01_14_v1	1936646	A	missense	AP2-MU	PmUG01_14053100	299N>299Y	1936646A>T		0.08	0	0.7	0	0
PmUG01_14_v1	1936705	G	missense	AP2-MU	PmUG01_14053100	318M>318I	1936705G>A		0.02	0.03	0	0	0
PmUG01_14_v1	1937129	G	missense	AP2-MU	PmUG01_14053100	460A>460S	1937129G>T		0.02	0.02	0	0	0
PmUG01_14_v1	1937138	G	missense	AP2-MU	PmUG01_14053100	463V>463F	1937138G>T		0.02	0.02	0	0	0
PmUG01_14_v1	2474672	A	missense	MRP2	PmUG01_14063400	2173I>2173M	2474672A>C		0.01	0.02	0	0	0
PmUG01_14_v1	2474712	A	missense	MRP2	PmUG01_14063400	2160L>2160S	2474712A>G		0.01	0.02	0	0	0
PmUG01_14_v1	2475164	A	missense	MRP2	PmUG01_14063400	2009I>2009M	2475164A>C		0.09	0.11	0	0	0
PmUG01_14_v1	2475264	G	missense	MRP2	PmUG01_14063400	1976T>1976I	2475264G>A		0.1	0	0.7	0	0
PmUG01_14_v1	2475649	T	missense	MRP2	PmUG01_14063400	1848T>1848A	2475649T>C		0.13	0.16	0	0	0
PmUG01_14_v1	2475671	A	missense	MRP2	PmUG01_14063400	1840N>1840K	2475671A>T		0.01	0.01	0	0	0
PmUG01_14_v1	2475782	A	missense	MRP2	PmUG01_14063400	1803D>1803E	2475782A>T	2475782A>C	0.13	0.03	0.88	0	0
PmUG01_14_v1	2475929	C	missense	MRP2	PmUG01_14063400	1754M>1754I	2475929C>T		0.01	0.01	0	0	0
PmUG01_14_v1	2476296	T	missense	MRP2	PmUG01_14063400	1632K>1632I	2476296T>A		0.01	0.02	0	0	0
PmUG01_14_v1	2476566	G	missense	MRP2	PmUG01_14063400	1542S>1542L	2476566G>A		0.03	0.03	0	0	0
PmUG01_14_v1	2476599	G	missense	MRP2	PmUG01_14063400	1531T>1531I	2476599G>A		0.01	0.02	0	0	0
PmUG01_14_v1	2476612	C	missense	MRP2	PmUG01_14063400	1527V>1527L	2476612C>A		0.01	0.02	0	0	0
PmUG01_14_v1	2476621	C	missense	MRP2	PmUG01_14063400	1524V>1524I	2476621C>T		0.01	0.02	0	0	0
PmUG01_14_v1	2476635	G	missense	MRP2	PmUG01_14063400	1519A>1519V	2476635G>A		0.52	0.46	0.94	1	0.5
PmUG01_14_v1	2476765	T	missense	MRP2	PmUG01_14063400	1476M>1476L	2476765T>A		0.01	0.02	0	0	0
PmUG01_14_v1	2477059	G	missense	MRP2	PmUG01_14063400	1378L>1378F	2477059G>A		0.01	0.01	0	0	0
PmUG01_14_v1	2477548	A	missense	MRP2	PmUG01_14063400	1215L>1215I	2477548A>T		0.09	0	0.81	0	0
PmUG01_14_v1	2477578	T	missense	MRP2	PmUG01_14063400	1205I>1205L	2477578T>A		0.01	0.01	0	0	0
PmUG01_14_v1	2477839	T	missense	MRP2	PmUG01_14063400	1118I>1118F	2477837T>A	2477839T>A	0.01	0	0.06	0	0
PmUG01_14_v1	2478259	C	missense	MRP2	PmUG01_14063400	978V>978I	2478259C>T		0.01	0.02	0	0	0
PmUG01_14_v1	2478542	G	missense	MRP2	PmUG01_14063400	883S>883R	2478542G>T		0.04	0.04	0	0	0
PmUG01_14_v1	2478705	C	missense	MRP2	PmUG01_14063400	829G>829D	2478705C>T		0.01	0.01	0	0	0
PmUG01_14_v1	2479063	A	missense	MRP2	PmUG01_14063400	710Y>710N	2479063A>T		0.05	0	0.43	0	0
PmUG01_14_v1	2479066	G	missense	MRP2	PmUG01_14063400	709H>709Y	2479066G>A		0.02	0.02	0	0	0
PmUG01_14_v1	2479155	G	missense	MRP2	PmUG01_14063400	679T>679S	2479155G>C		0.64	0.58	1	1	0.67
PmUG01_14_v1	2479284	A	missense	MRP2	PmUG01_14063400	636V>636A	2479284A>G		0.01	0	0.09	0	0
PmUG01_14_v1	2479300	C	missense	MRP2	PmUG01_14063400	631V>631I	2479300C>T		0.67	0.65	0.88	1	0
PmUG01_14_v1	2479326	T	missense	MRP2	PmUG01_14063400	622N>622S	2479326T>C		0.01	0.01	0	0	0
PmUG01_14_v1	3475374	C	missense	Coronin	PmUG01_14084400	135P>135S	3475374C>T		0.01	0.01	0	0	0
PmUG01_14_v1	3475981	C	missense	Coronin	PmUG01_14084400	337T>337N	3475981C>A		0.01	0	0.05	0	0
PmUG01_14_v1	3475990	G	missense	Coronin	PmUG01_14084400	340R>340K	3475990G>A		0.01	0	0.05	0	0
PmUG01_14_v1	3476320	A	missense	Coronin	PmUG01_14084400	450D>450V	3476320A>T		0.03	0.04	0	0	0

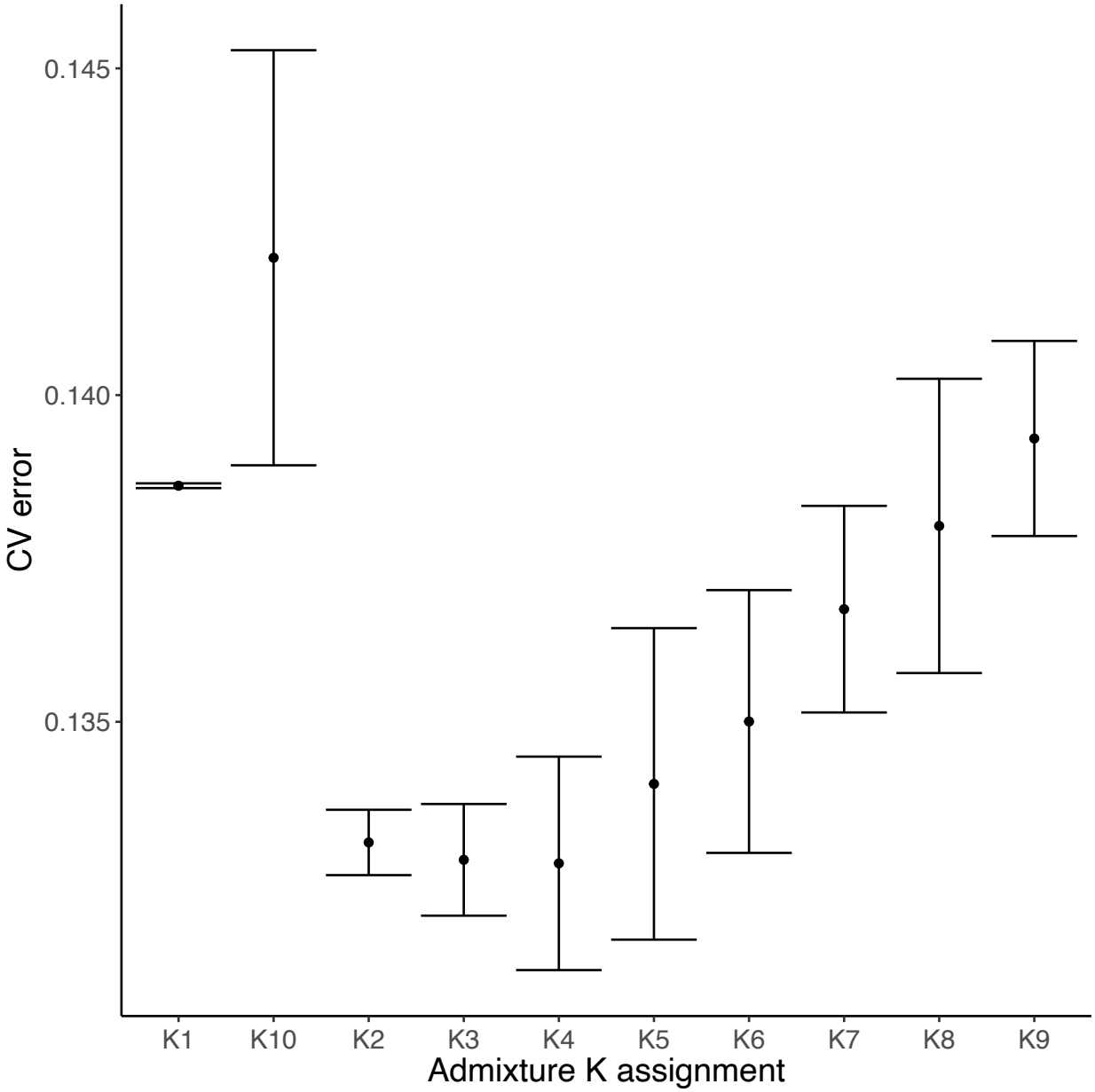
Supplementary Figures



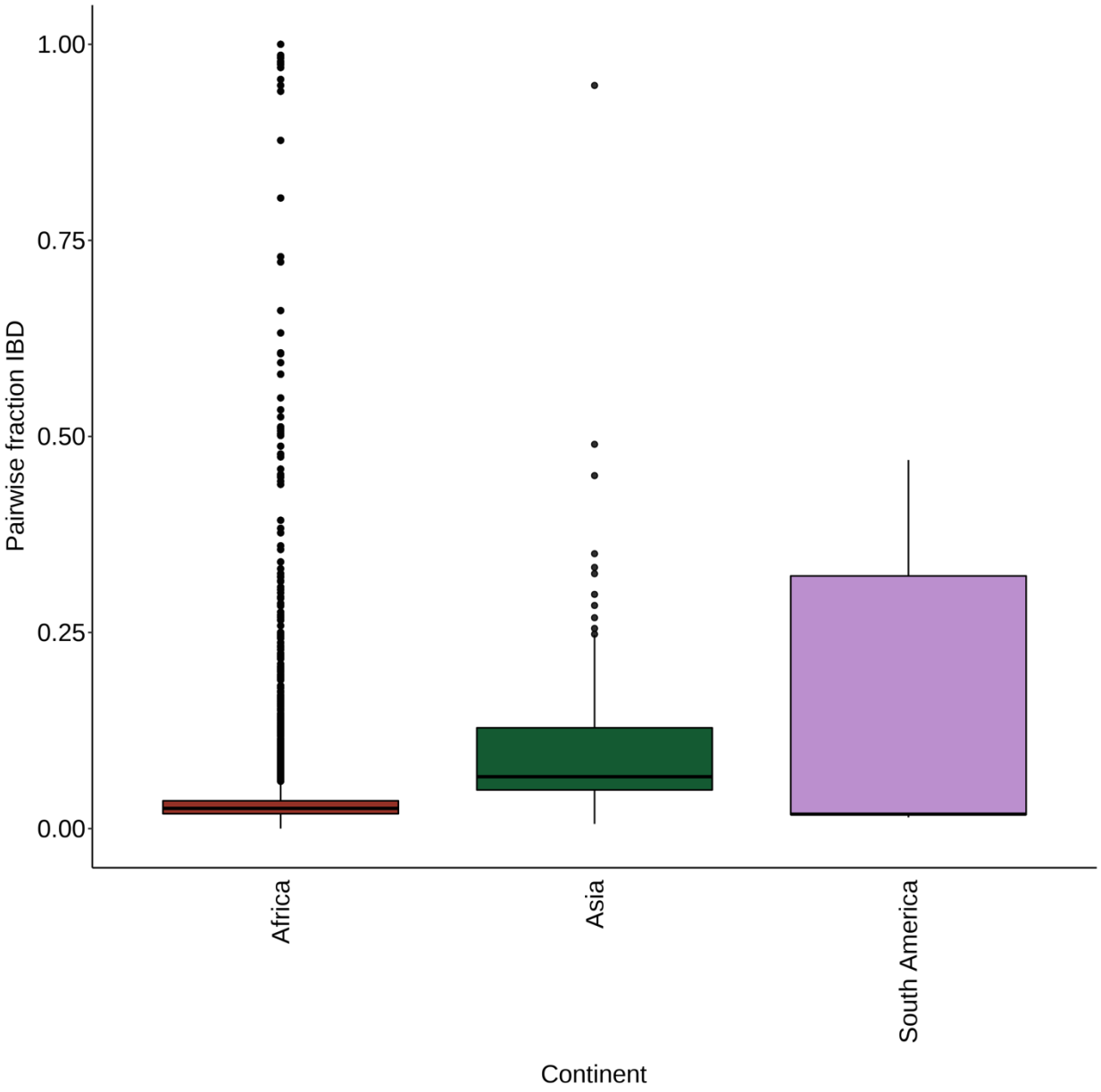
S1 Figure. Distribution of *P. malariae* isolates in the filtered genomics database.
Blue colour scale demonstrates the number of isolates found in each country.



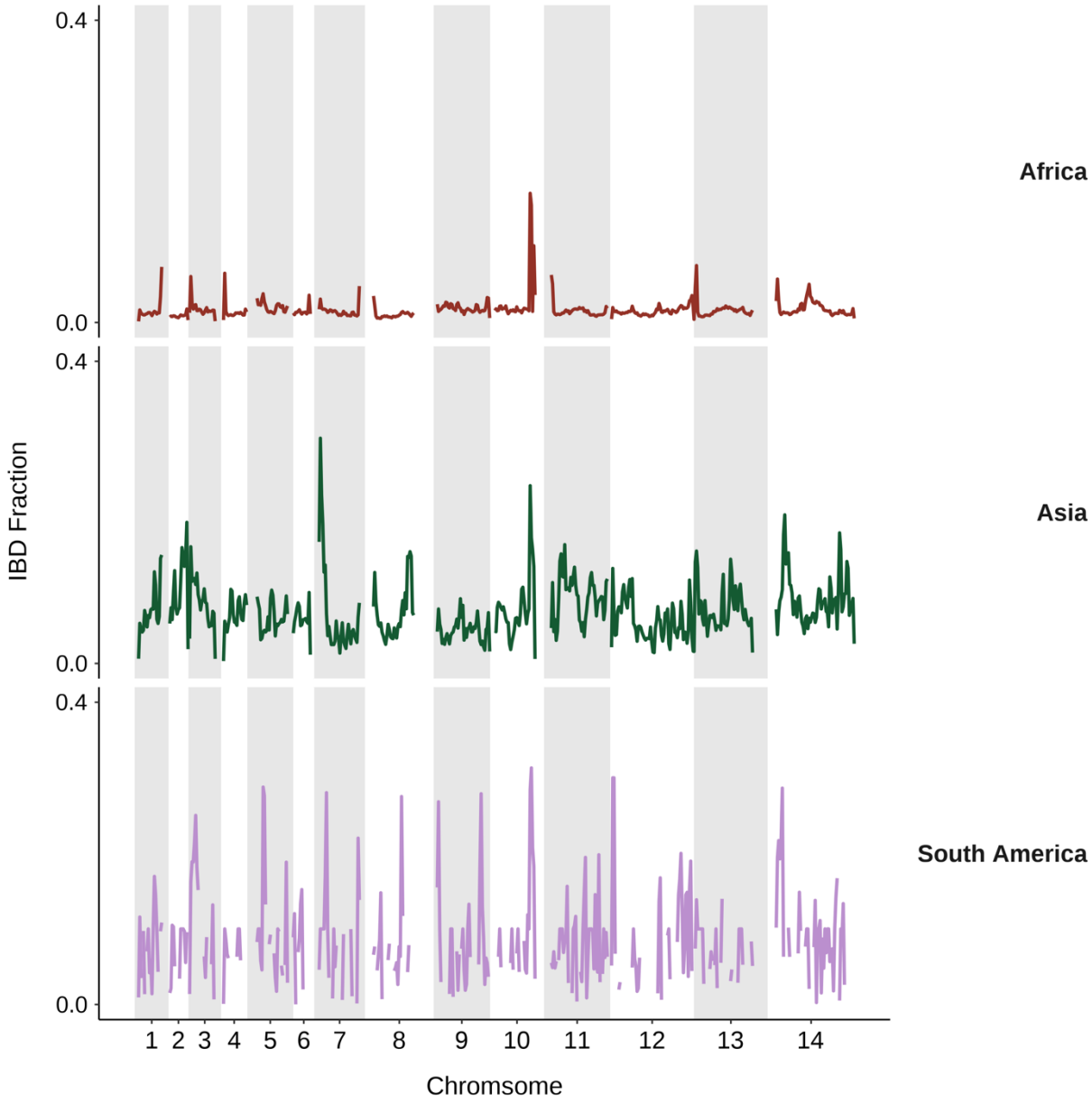
S2 Figure. Multiclonality in isolates in the global database (n = 155).
Boxplot of F_{WS} scores found in each continent in the filtered global dataset. Boxes are coloured corresponding to the continent (Africa = brown, Asia = green, Oceania = grey, South America = lilac).



S3 Figure. CV error scores from Admixture analysis.
Admixture analysis was run for up to 10 possible ancestral populations, with this analysis repeated independently 10 times and the average value for the CV error at each possible K taken. The mean CV error score across the 10 repeats is plotted with standard error bars. The lowest CV error score was seen at 4 ancestral populations (average CV error = 0.132834 ±0.0015).



S4 Figure. Signals of homology within monoclonal parasite populations in the global dataset. Distribution of fraction IBD scores found within each continent in monoclonal isolates (F_{WS} > 0.85; Africa, n = 108; Asia, n = 14; South America, n = 5). Boxplot coloured according to the continent (Africa = brown, Asia = green, South America = lilac). IBD fractions calculated using hmmIBD [62].



S5 Figure. Genome-wide signals of high IBD across the three continents in monoclonal isolates.
Fraction IBD scores found within each continent in monoclonal isolates (F_{WS} > 0.85; Africa, n = 108; Asia, n = 14; South America, n = 5) shown along the *P. malariae* genome. Lines are coloured according to the continent (Africa = brown, Asia = green, South America = lilac). IBD fractions calculated using hmmIBD [62].

CHAPTER 4

4 Functional analysis of *P. malariae dhfr-ts* mutations

4.1 Introduction

4.1.1 Research premise

As discussed in chapter 3, my previous work led to the development of the first global genomic database for *P. malariae* (n = 158 isolates from 26 countries), which, amongst other findings, highlighted multiple mutations within orthologs of genes associated with reduced drug susceptibility in *P. falciparum*. Of note were 15 non-synonymous mutations within *pmdhfr-ts* (PmUG01_05034700), a gene associated with pyrimethamine susceptibility in *P. falciparum* [1,2] and *P. vivax* [3] (**Table 1**).

Whilst there are no confirmed molecular markers for drug resistance in *P. malariae* isolates, there have been multiple reports of post-treatment (with both artemisinin-based [4,5], and chloroquine therapies [6]) and post-prophylactic (using atovaquone-proguanil [7]) parasitaemia, raising concerns over the efficacy of current drugs against this parasite species. Whilst *P. malariae* infections are mostly asymptomatic, they are commonly found in mixed infections with other *Plasmodium spp*., increasing the likelihood of exposure to antimalarial drugs when a patient is treated for the other species present in a mixed infection.

Table 1. Summary of nonsynonymous SNPs found within *pmdhfr-ts*

The presence of SNPs within *pmdhfr-ts* is shown with the allele frequency found in each continent. Downstream amino acid changes are given (AA Change) along with the causative nucleotide change (NT Change 1/NT Change 2).

Chromosome	Gene	Position	Reference	Alternate	AA Change	NT Change 1	NT Change 2	All Freq (N = 195)	African Freq (N = 163)	Asian Freq (N = 22)	SAM Freq (N = 8)	Oceania Freq (N = 2)
PmUG01_05_v1	DHFR-TS	1291895	G	T	15A>15S	1291895G>T		0.05	0.04	0	0.33	0
PmUG01_05_v1	DHFR-TS	1291999	T	A	49S>49R	1291999T>A		0.01	0	0.05	0	0
PmUG01_05_v1	DHFR-TS	1292023	C	G/A	57F>57L	1292023C>A/G		0.17	0.21	0	0	0
PmUG01_05_v1	DHFR-TS	1292026	A	C/G	58R>58S	1292026A>C		0.82	0.78	1	1	1
PmUG01_05_v1	DHFR-TS	1292193	A	G	114N>114S	1292193A>G		0.71	0.67	0.84	1	1
PmUG01_05_v1	DHFR-TS	1292522	T	A	224C>224S	1292522T>A		0.01	0.01	0	0	0
PmUG01_05_v1	DHFR-TS	1293224	C	T	458H>458Y	1293224C>T		0.01	0.01	0	0	0
PmUG01_05_v1	DHFR-TS	1293228	C	A	459A>459D	1293228C>A	1293229C>T	0.01	0.01	0	0	0
PmUG01_05_v1	DHFR-TS	1293239	G	A	463D>463N	1293239G>A		0.01	0.01	0	0	0
PmUG01_05_v1	DHFR-TS	1293272	C	A	474H>474N	1293272C>A		0.01	0.01	0	0	0
PmUG01_05_v1	DHFR-TS	1293308	A	C	486I>486L	1293308A>C		0.01	0.01	0	0	0
PmUG01_05_v1	DHFR-TS	1293527	A	G	559I>559V	1293527A>G		0.01	0.01	0	0	0
PmUG01_05_v1	DHFR-TS	1293575	G	A	575V>575I	1293575G>A	1293577A>T	0.01	0.01	0	0	0
PmUG01_05_v1	DHFR-TS	1293628	A	T	592E>592D	1293628A>T		0.01	0.01	0	0	0
PmUG01_05_v1	DHFR-TS	1293691	T	A	613D>613E	1293691T>A		0.01	0.01	0	0	0

We aimed to investigate differences in pyrimethamine susceptibility between *P. knowlesi*, *P. malariae* and *P. falciparum* using orthologue replacement, in addition to investigating the effects of the mutations found within the global dataset (**Table 1**). To confirm a resistant and sensitive parasite phenotype, we generated transfectant lines where the *pkdhfr* domain was replaced with the 3D7

reference genome sequence for the *pfdhfr* domain (as a sensitive control line), in addition to the *pfdhfr* domain with a triple mutation leading to amino acid substitutions N51I, C59R and S108N, the widespread resistant genotype found in Africa [8]. Additionally, we investigated pyrimethamine susceptibility in the *P. knowlesi* A1-H1 culture adapted parasite line and created a transfectant line where the *pkdhfr* domain was replaced with a recodonized sequence of the *pkdhfr* domain, to determine the effect of the transfection process on the parasite phenotype without any modification of the amino acid sequence after transcription. Finally, we created parasite lines where the *pkdhfr* domain had been replaced with the *pmdhfr* domain, alongside lines containing mutations within *pmdhfr* that were seen within the global dataset. Whilst multiple transfectant lines have been created with combinations of mutations within *pmdhfr,* due to time constraints, we present data for 3 parasite lines with different *pmdhfr* genotypes (**Table 2** for a summary of transfectant parasite lines). These parasite lines were then used in drug susceptibility assays to determine the effect on pyrimethamine susceptibility.

Table 2. A summary of parasite lines

Parasite line	Species DHFR domain	Description	Notes
WT_A1H1	*P. knowlesi*	A1H1 culture adapted line	WT culture with no genetic modification
TF 131_2	*P. knowlesi*	*Pkdhfr-ts* with the DHFR domain recodonised	Transfection control
TF 150	*P. falciparum*	3D7 reference sequence recodonised	*P. falciparum* sensitive line
TF 135_1	*P. falciparum*	3D7 reference sequence (as in TF 150) with substitutions N51I, C59R, S108N	*P. falciparum* resistant line
TF 150_1	*P. malariae*	TF135_3 with mutation 3 and 6	Hypothetical sensitive line with R58S and N114S substitutions
TF 135_3	*P. malariae*	PmUG01 reference sequence recodonised	PmUG01 reference sequence
TF 151	*P. malariae*	TF 135_3 with mutation 5 (F57L)	PmUG01 reference with F57L

4.1.2 Pyrimethamine for malaria treatment

A British research programme focussed on antimalarial compounds during the Second World War identified Proguanil in 1945, a prodrug whose metabolite, chlorocycloguanil, inhibits the Plasmodium DHFR enzyme. The discovery of proguanil led to further investigation into similar compounds, including pyrimethamine, which had been previously been synthesized and investigated for the treatment of tumours [9]. Pyrimethamine is a class II antifolate drug which, when used as an antimalarial treatment, targets and inhibits the Plasmodium enzyme DHFR, disrupting the folate biosynthesis pathway (**Figure 1**) [10]. Pyrimethamine was initially used as a monotherapy until resistance began to emerge in the 1960s, leading to the development of Fansidar/SP, a combination therapy of pyrimethamine and sulfadoxine (a class I antifolate which targets the parasite enzyme dihydropteroate synthase DHPS), which work synergistically, targeting different enzymes in the folate biosynthesis pathway (**Figure 1**) [11].

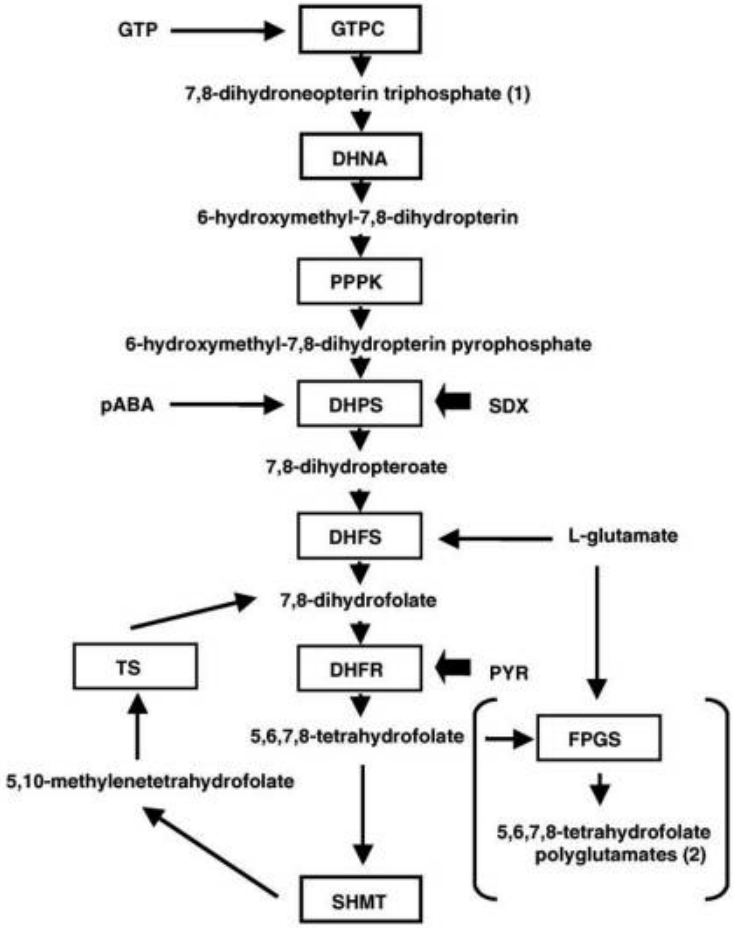


Figure 1. The folate biosynthesis pathway in *Plasmodium falciparum* parasites. The biosynthesis of folates in *Plasmodium spp*. is a complex pathway involving sequential enzymatic processes. The enzymes which are targets for the antifolate drugs, pyrimethamine and sulfadoxine are annotated with an arrow and PYR or SDX respectively, taken from Hyde, J.E 2005 [10].

4.1.2.1 Resistance to pyrimethamine

For *P. falciparum* isolates, SP resistance first developed in South America and Southeast Asia in the 1970s [11,12], with Thailand demonstrating a particularly quick development of resistance, within the same year as SP was first used in 1967 [13]. Treatment failures to both chloroquine and SP were increasing in the African continent by the end of the 1990s, with resulting epidemics at the start of the millennium in Burundi, Sudan and Ethiopia [14], leading to the adoption of ACTs as a first line treatment [15]. Combinations of sequential mutations within the DHFR domain of *pfdhfr-ts* are responsible for pyrimethamine resistance, with the lead causative mutation creating an amino acid substitution of serine to threonine at position 108 (S108N) [16]. Additional nonsynonymous mutations

leading to amino acid substitutions at positions 51, 59 and 164 are known to increase levels of pyrimethamine resistance [1,16], with amino acid substitutions at positions 16 associated with cycloguanil and potentially pyrimethamine resistance [17,18]. A further amino acid substitution at position 50 was highly prevalent in Bolivia, a population with high levels of pyrimethamine resistance and is thought to be associated with reduced susceptibility [1]. The most common pyrimethamine-resistant haplotype in the African continent is the triple mutant, IRN (N51I, C59R, S108N) [19,20], which had a single origin that spread across the continent in contrast to the single point mutation at position 108, which was shown to have risen independently multiple times [21].

Whilst antifolate resistance is widespread in *P. falciparum* parasites in many regions, SP has continued medical relevance through preventative treatment such as intermittent preventative treatment for pregnant women (IPTp) and infants (IPTi), which are still effective preventative treatments in many regions where SP resistance is found [22]. However, the presence of a sextuplet mutation, with three amino acid substitutions in both *pfdhfr* (N51I, C59R, S108N) and *Pfdhps* (A437G, K540E, A581G) was found to reduce the efficacy of IPTp, meaning that continued surveillance into *pfdhfr-ts* mutations is necessary [8,23,24]. Additionally, SP is found within ACTs, in combination with artesunate (artesunate + sulfadoxine-pyrimethamine, ASSP) which is used in some regions in India and the Middle East as well as some East African countries [25].

Whilst most studies regarding pyrimethamine resistance are focussed on *P. falciparum* parasites, *P. vivax* isolates can also display pyrimethamine resistance, which has been reported as early as 1959 [26]. Pyrimethamine resistance in *P. vivax* isolates is due to SNPs within the DHFR domain of the *pvdhfr-ts* gene (PVP01_0526600), with the most relevant mutations leading to amino acid substitutions at three positions (57, 58, 117) [3,27]. The first-line treatment for *P. vivax* infections in most of the world is chloroquine, therefore it is interesting to note that pyrimethamine resistance has arisen without being used as the primary treatment for this infection. In most malarious regions, infections are dynamic and multiple *Plasmodium spp.* are likely to be found, therefore *P. vivax* parasites are likely exposed to pyrimethamine in mixed infections where SP was used to treat *P. falciparum*. Additionally, non-synonymous mutations have been documented within the *P. malariae* ortholog, *pmdhfr-ts,* in isolates from Asia and Africa, with some isolates from Southeast Asia showing amino acid substitutions at locations predicted to be associated with drug susceptibility due to their location in respect to *P. falciparum* and *P. vivax* orthologues. However, phenotypic testing was only possible

for isolates with the mutation at residue 22 and no alteration to drug susceptibility was observed [28]. Whilst chloroquine or ACTs are the recommended therapy for *P. malariae* [29], most infections are mixed, and parasites are likely to have been exposed to pyrimethamine due to past treatment guidelines for *P. falciparum* infections, or due to large scale IPTi and IPTp administration.

Pyrimethamine resistance in both *P. falciparum* and *P. vivax* parasites is due to amino acid substitutions located within the active site of the DHFR enzyme, where pyrimethamine binds [30]. Pyrimethamine is an analog of dihydrofolate and inhibits the folate biosynthesis pathway by acting as a competitive inhibitor, reducing the availability of the active site of DHFR to bind to dihydrofolate. Through this mechanism of action, the folate synthesis pathway is hindered, and parasites die due to a lack of folic acid [10]. Amino acid substitutions in DHFR associated with pyrimethamine resistance are thought to alter the shape of the active site, reducing the affinity of pyrimethamine binding, which allows for increased dihydrofolate binding and continuation of the folate synthesis pathway [30]. Both pyrimethamine and sulfadoxine have long half-lives, which has been suggested to increase the chances of selection for strains with reduced susceptibility, as parasites may be exposed to sub-lethal concentrations of drug for an extended period of time [31].

4.1.2.2 Additional antifolates

In *P. falciparum* isolates, amino acid substitutions S108T and A16V within *pfdhfr-ts* are associated with cycloguanil resistance. Cycloguanil is the active component of the drug proguanil which also targets the folate synthesis pathway. Whilst not commonly used for malaria treatment, proguanil is found in combination with atovaquone in the prophylactic antimalarial Malarone [32]. Additionally, trimethoprim-sulphamethoxazole (cotrimoxazole or CTX), where trimethoprim targets DHFR and sulphamethoxazole competes with para-aminobenzoic acid (pABA), the substrate of DHPS [33], is widely used in the African continent to treat bacterial and protozoan infections, specifically in HIV-positive individuals, which may provide additional drug selection pressures on the folate pathway [34]. Chloroproguanil-dapsone (known as LapDap) is being increasingly used to treat malaria, with the active metabolite, chlorproguanil which binds to DHFR whilst working synergistically with dapsone which binds to DHPS [35].

4.1.3 Genome editing in Plasmodium parasites

4.1.3.1 An overview of genome editing approaches

Genome editing, relying on site-specific nucleases such as zinc-finger nucleases and, more recently, the clustered regularly interspaced short palindromic repeat (CRISPR)-CRISPR-associated protein (Cas) system (CRISPR-Cas) [36,37] is now a widely used tool in malaria research. Site-specific nucleases create a targeted double strand break (DSB) at a specific site within the parasite genome, allowing for genome disruption and manipulation. After a DSB, there are two common processes for DNA repair known as homologous recombination (HR) and non-homologous end joining (NHEJ). HR uses homologous template DNA to repair the broken fragment of DNA, whereas NHEJ attempts to ligate the broken DNA fragments, and is error-prone, commonly leading to small insertions or deletions (INDELs), or premature stop codons and disruption to the affected gene (**Figure 2**) [37]. This project focusses on using the CRISPR-Cas9 system, a naturally occurring system for bacterial immunity [38], to create transfectant lines of *P. knowlesi* parasites with ortholog replacement at the *pkdhfr-ts* locus. In Plasmodium parasites, the NHEJ pathway for genome repair is absent, meaning that DNA repair is reliant on the HR pathway [39].

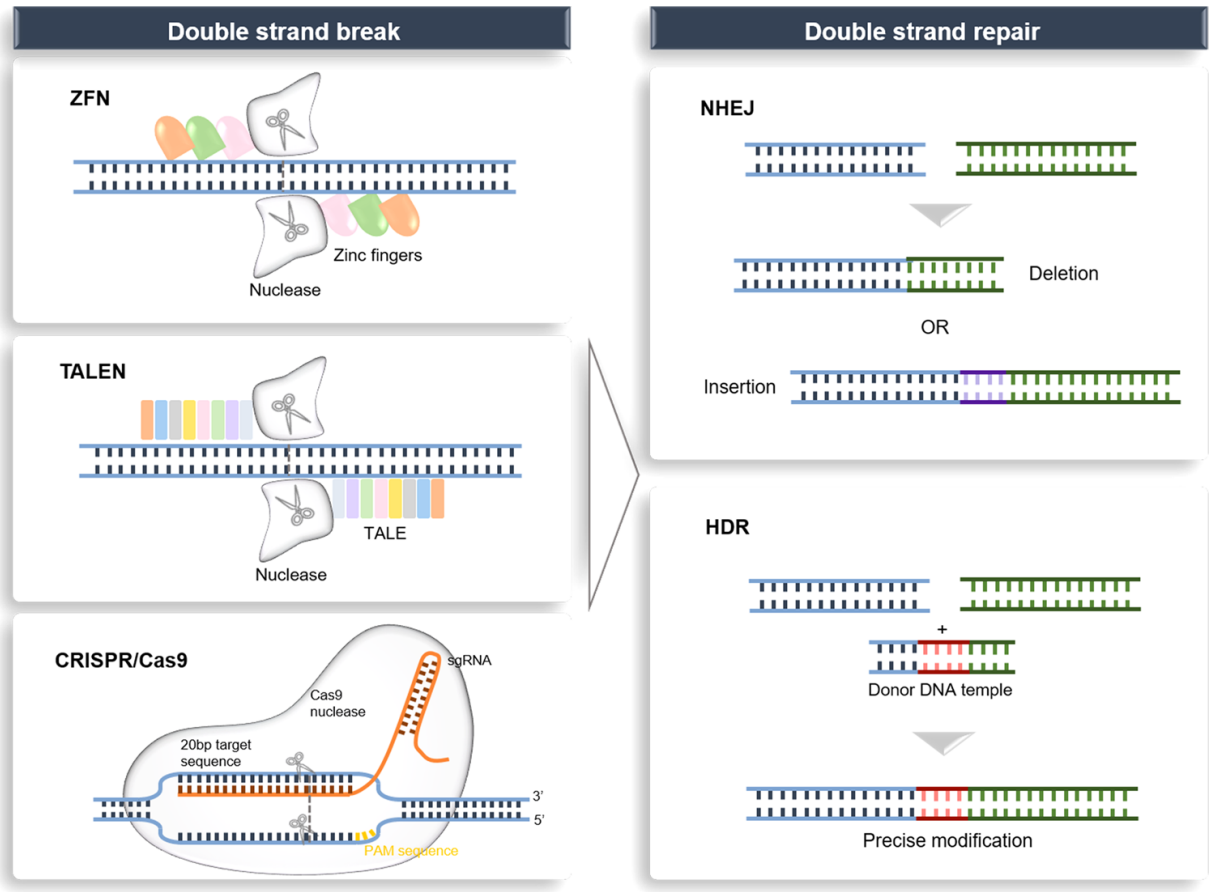


Figure 2. Common methods of genome editing based on site-specific nucleases.
Three differing common methods to induce DSB are shown; ZFN using zinc finger nucleases, TALEN using transcription activator-like effector nucleases, and CRISPR-Cas9 using the clustered regularly interspaced short palindromic repeat–Cas9 system. All methods to induce a DSB lead to the two DNA repair processes; non-homologous end joining (NHEJ) or homologous recombination (HR) when a donor DNA template is present. Taken from Li et al. [37].

The CRISPR-Cas9 system is a naturally occurring immune system in prokaryotes to protect against viral infections which has been utilised in molecular biology to allow for gene editing of organisms. The system is based upon the Cas9 protein, an endonuclease responsible for causing DSBs, which can be targeted to a specific location within the genome via a single guide RNA (sgRNA). When targeting a specific gene (in our case *pkdhfr-ts*), a 20 bp sequence within the gene of interest is chosen and inserted into a cas9 expression plasmid (pCas9) within an RNA expression cassette. When transfected into a parasite culture, the pCas9 plasmid is transcribed, and the guide sequence is expressed as the sgRNA. The sgRNA binds to the complimentary region in the gene of interest, directing the cas9 enzyme to the specific location, where it is able to cause a DSB (**Figure 2**). A protospacer adjacent motif (PAM), which is a 3 bp sequence of 'NGG' where N denotes any

nucleotide, is required for the endonuclease activity of Cas9. The 20 bp guide sequence is designed within the gene of interest directly upstream from a PAM site (**Figure 3**).

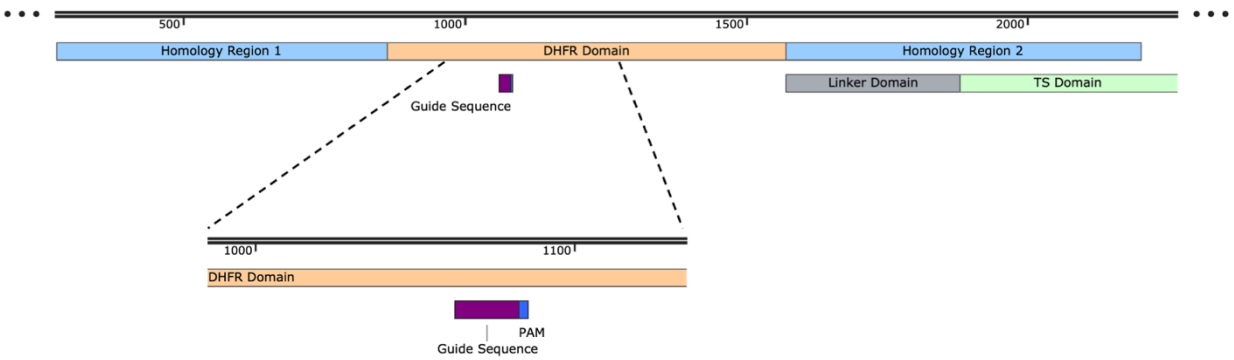


Figure 3. The *P. knowlesi pkdhfr-ts* gene locus.
Map of the *pkdhfr-ts* locus, with the domains, homology regions, guide sequence and PAM site highlighted for designing the donor constructs.

4.1.3.2 CRISPR-Cas9 genome editing in human-infective malaria parasites

The CRISPR-Cas9 system was first utilised in *P. falciparum* isolates in 2014, where resistance to the WR99210 antifolate drug was introduced by integrating a copy of the human DHFR gene (h*dhfr*) into the Plasmodium genome [40]. Plasmodium parasites lack the canonical pathway for NHEJ [39], meaning that DSBs are primarily repaired using HR, and any gene editing procedures using site specific nucleases will require a donor DNA template for DNA repair [36].

P. knowlesi is a zoonotic malaria parasite species which is a major problem for human infections in Southeast Asia, that is genetically closely related to *P. vivax* isolates [41]. In addition to *P. falciparum*, *P. knowlesi* is the only other human-infective Plasmodium species with a stable i*n vitro* culture in human red blood cells (RBCs) [42]. CRISPR-Cas9 genome-editing technology has been utilised in *P. knowlesi* to answer challenging questions into functional biology for Plasmodium parasites [43–45].

The *P. knowlesi* A1-H.1 genome is 24.4 Mb, and previous sequencing of the PKNH reference genome demonstrated a balanced GC content of 37.5% [46] (in comparison to the AT-rich *P. falciparum* with 19.4% GC content), aiding CRISPR-Cas9 mediated genome editing due to the increased likelihood

of appropriate PAM sequences within genes of interest (which can be problematic in the AT-rich *P. falciparum*). Additionally, *P. knowlesi* demonstrates larger merozoites, resulting in higher transfection efficacies when compared to *P. falciparum* [45]. The CRISPR-Cas9 genome editing system is robust in *P. knowlesi* parasites, with a faster generation of clonal parasite lines in comparison to *P. falciparum* due to the shorter erythrocytic life cycle [43,45].

4.2 Project Aims and Objectives

4.2.1 Aims

My aim was to generate stable transgenic parasite lines of *P. knowlesi* isolates that had undergone orthologue replacement at the *dhfr* locus to determine differences in pyrimethamine susceptibility between the different *Plasmodium spp.*, in addition to the effect of mutations within this gene in *P. malariae* isolates. To achieve this aim, the objectives were:

4.2.2 Objectives

- **Objective 1:** Determine potentially relevant mutations within *pmdhfr-ts* using the global dataset of *P. malariae* isolates.
- **Objective 2:** Determine and create DNA constructs for *P. malariae dhrf* loci, in addition to control loci such as *P. falciparum dhfr c*onstructs for both sensitive and resistant genotypes.
- **Objective 3:** Generate a CRISPR-Cas9 guide plasmid (pCas9) to target the *pkdhfr-ts* locus, in addition to donor plasmids (pDonor) that will be used for homology-directed repair.
- **Objective 4:** Work alongside a stable culture of *P. knowlesi* A1H1 parasites that will be used for transfections, whilst learning routine culture methods and aseptic technique.
- **Objective 5:** Transfecting *P. knowlesi* parasites using nucleofection, an electroporation-based methodology for transfections, and selecting transfectant lines using limiting clonal dilution.
- **Objective 6:** Using transfected lines for drug assays with pyrimethamine and dihydroartemisinin (DHA) as a control.

4.3 Methods

4.3.1 Molecular biology

4.3.1.1 Polymerase chain reaction (PCR) and gel electrophoresis

All targeted amplicon DNA amplification for cloning was performed using CloneAmp HiFi polymerase (TaKaRa). The CloneAmp amplification reaction had 32 cycles and was as follows: 10 seconds at 98°C, 15 seconds at 55°C, and an extension of 5 seconds/kb at 72°C. Template DNA never exceeded 100 ng, and 1X CloneAmp HiFi PCR Premix (TaKaRa) was used alongside 5 – 7.5 pmol of each primer (final concentration of each primer = 0.2-0.3 μM). All primers (for both cloning and colony PCRs) were designed using OligoCalc [47] and ordered through IDT. All primers used within this project are listed within **Supplementary Table 1**. DNA amplicons were visualised to determine their size using gel electrophoresis with a 1% agarose gel in Tris-acetate-EDTA (TAE). Gels were stained using SYBR Safe DNA Gel Stain (Thermofisher) and Hyperladder 1kb and Hyperladder 100bp (Bioline) were used for size determination. Gels were routinely ran at 120V for 30 minutes and visualised using a Gel Doc XR+ Gel Documentation System (BIO-RAD).

4.3.1.2 Restriction enzyme digests

DNA was digested and cut into specific fragments through digestion with specifically chosen restriction enzymes (New England Biolabs) according to their specified temperature dependencies. Routine plasmid digests were carried out using 1 μL enzyme per 1 μg plasmid DNA in a total reaction volume of 50 μL. Only enzymes with the same buffer and temperature requirements were used in combination. After enzymatic digestion of DNA and plasmids, reactions were separated on 1% agarose gels and bands of the correct size for the required digestion were visualised using the Safe Imager 2.0 Blue-Light Transilluminator (Invitrogen) and cut out using a scalpel. DNA was extracted from gel bands using the QIAquick Gel Extraction Kit (Qiagen) and quantified using the Nanodrop function on a DS-11 FX Spectrophotometer (DeNovix).

4.3.1.3 Ligation

DNA inserts were ligated into plasmid backbones using T4 DNA ligase (NEB). DNA input concentrations for both the digested plasmid and DNA insert were calculated using the NEBioCalculator (https://nebiocalculator.neb.com/#!/ligation) using a 3:1 Insert:Vector ratio. Ligation reactions were performed using 400 units T4 DNA ligase (NEB) and 1X T4 DNA Ligase Buffer (NEB) and were left to ligate for 10-20 minutes at room temperature.

4.3.1.4 Bacterial transformations

Both Stellar Competent (TaKaRa) and XL10-Gold Ultracompetent (Agilent) *Escherichia coli* cells were used for bacterial transformation via a heat shock protocol according to manufacturer's instructions. For both transformations, 3 μl of ligation reaction was added to 30 μl of competent cell aliquots. Following heat shock, the reaction mixture was spread onto Luria-Bertani (LB) agar plates treated with 100 μg/mL ampicillin and incubated overnight at 37°C.

4.3.1.5 Colony screening

Individual bacterial colonies that had grown on an agar plate overnight were picked and screened for positive ligations via a colony PCR, where primers were designed to amplify a DNA fragment encompassing part of the plasmid backbone in addition to a segment of the integrated locus. Positive PCR bands are indicative of a successfully ligated and integrated insert within the vector backbone. Colony PCR primers were ordered through IDT (**Supplementary Table 1**), and the amplification used 1 X SapphireAmp Fast PCR Master Mix (TaKaRa) alongside 0.1-0.5 uM of each forward and reverse primer.

4.3.1.6 Plasmid purification

Once successful transformed colonies were confirmed by colony PCR, the colony was picked and underwent an overnight bacterial outgrowth in LB medium treated with 100 μg/mL ampicillin in a shaking incubator at 37°C. Plasmid DNA was extracted using either the Plasmid Midi Kit (Qiagen) or HiSpeed Plasmid Midi Kit (Qiagen) depending on the size of the outgrowth culture. Plasmid DNA was quantified using the Nanodrop function on a DS-11 FX Spectrophotometer (DeNovix).

4.3.1.7 Plasmid linearisation

pCas9 plasmids (up to 2μg) were linearized using restriction enzyme digest with 5-10 units of BtgZI (NEB) and 1X CutSmart buffer (NEB) at 60°C for 2 hours. Linearised plasmids were purified after digestion using the QIAquick PCR Purification Kit (Qiagen) according to manufacturer's instructions. pDonor plasmids (up to 2μg) were digested using 5-10 units of PvuI restriction enzyme (NEB) and 1X CutSmart buffer (NEB) at 37°C for 2 hours. Linearised plasmids were purified as previously described with the QIAquick PCR Purification Kit (Qiagen) and quantified using the Nanodrop function on a DS-11 FX Spectrophotometer (DeNovix).

4.3.1.8 Infusion ligation

Guide sequences were incorporated into the pCas9 plasmid through infusion ligation using the In-Fusion HD Cloning Kit Plus (TaKaRa). First, the two guide sequence oligonucleotides (forward and reverse complement, olAI003 and olAI004) were resuspended to a 100 μM concentration in nuclease-free water and 10 μl of each was mixed 2.2 μl of NEB2 buffer. The oligonucleotide-buffer mixture was incubated at 95 °C for 10 minutes in a heat block, before moving the block into a bucket of ice to slowly cool down and allow for the oligonucleotides to anneal. Annealed oligonucleotides were diluted to 0.5 μM with cold Buffer EB (Qiagen). A total of 50 ng linearised pCas9 plasmid (with BgtZI) was mixed with 1 μl diluted annealed oligonucleotides, 1 μl In-Fusion buffer (TaKaRa) and 2 μl H_20 on ice before incubating at 50 °C for 15 minutes. The resulting ligation was transformed into competent cells (Methods 4.3.1.4).

4.3.2 In silico analysis

4.3.2.1 Protein structures and alignments

The crystal structure for *P. falciparum* 3D7 *pfdhfr-ts* complexed with nicotinamide adenine dinucleotide phosphate (NADPH), deoxyuridine monophosphate (dUMP) and pyrimethamine (3QGT) was downloaded from the Protein data bank (PDB, https://www.rcsb.org/) and visualised in PyMOL [48]. The amino acid sequence for *pmdhfr-ts* (PmUG01_05034700) was downloaded from PlasmoDB and the tertiary protein structure was predicted using I-TASSER [49]. The predicted *pmdhfr-ts* structure was aligned to *pfdhfr-ts* in PyMOL and positions of interest were annotated manually.

4.3.2.2 Generating donor DNA constructs for orthologue replacement

Gene sequences for the orthologs of DHFR-TS within *P. malariae, P. falciparum* and *P. knowlesi* were downloaded from the PlasmoDB database (https://plasmodb.org/plasmo/app), using the PmUG01 reference genome for *P. malariae* [50], 3D7 for *P. falciparum* [51] (a known pyrimethamine sensitive lab strain), and the A1H1 reference genome for *P. knowlesi* [52].

The PkA1H1 DHFR-TS (PKA1H_050015200) sequence (including ~1 kb flanking sequence before and after the gene) was visualised in SnapGene v4.3.11, and annotated with the resulting DHFR, linker and TS protein domains according to the associated InterPro Domains listed in PlasmoDB. The

sequence for the DHFR domain, with 500 bp of sequence flanking either side of the domain was selected, and recodonised using the GeneArt Instant Designer interface (by selecting to optimise the sequence by changing codons for the *Mus musculus* host organism). The recodonised sequence was visualised in Snapgene, and manually edited to add restriction enzyme sites either side of the DHFR domain (SpeI site after the first homology region, and an NcoI site before the second homology region) in addition to two SfiI restriction enzyme sites at either side of the homology regions (**Figure 9**).

4.3.2.3 Generating CRISPR-Cas9 plasmids to target the pkdhfr-ts locus

4.3.2.3.1 Identifying target sites in the P. knowlesi genome

One Cas9 guide plasmid (pCas9) was needed for the investigation into DHFR-TS as all ortholog replacements were directed at the same gene, *pkdhfr-ts* (PKA1H_050015200). Target sequences, known as a protospacer adjacent motifs (PAMs), are comprised of 3 nucleotides where a 5' nucleotide is followed by two guanine bases (5'-NGG-3') (**Figure 3**). Target sequences were identified in *pkdhfr-ts* and a 20 bp region upstream from an identified PAM site was selected (the guide sequence) and assessed for its uniqueness across the *P. knowlesi* genome with Protospacer Workbench software (v0.0.1 alpha). Guides with off target binding sites scoring less than 0.3 were accepted. The final chosen guide sequence used in this project was 5'-ccaagtacgagaagttaaag-3'. Once the guide sequence was identified, it was flanked with 15 bp sequences complementary to the overhangs created when the pCas9 plasmid is linearised with BgtZI as described in Mohring *et al*. [43,45]. This 50 bp sequence (5'-TTACAGTATATTATTccaagtacgagaagttaaagGTTTTAGAGCTAGAA-3', lower case = guide sequence, upper case = homology to pCas9 plasmid) was ordered as two oligonucleotide sequences (forward and reverse compliment) (IDT) which were annealed and inserted into the linear pCas9 guide plasmid using infusion ligation after the plasmid was linearised using BgtZI (TaKaRa) (Oligonucleotide sequences in **Supplementary Table 1**).

4.3.2.3.2 Guide plasmid generation

We used the pCas9/sg guide plasmid [43] which contains both Cas9 and sgRNA expression cassettes, in addition to a fusion protein of both positive and negative selection markers (hDHFR-yFCU). The

positive selection marker is human DHFR (hDHFR) and allows for selection of integrated clones through culturing parasites in the presence of pyrimethamine. The negative selection marker is yeast cytosine deaminase/uridyl phosphoribosyl transferase (*yFCU*) which allows for removal of parasites with plasmids through culturing with Ancotil. [43,45]. Due to the nature of the proposed experiment (pyrimethamine drug assays), the hDHFR-yFCU fusion protein was replaced with a blasticidin resistance cassette (BSDr) fusion with yFCU (BSDr-yFCU), using restrictrion enzyme digests with SpeI and AscI (**Figure 9**), allowing for positive selection using blasticidin, to avoid prior exposure of parasite cultures to pyrimethamine, and to avoid the hDHFR domain from remaining in any parasite cultures and affecting the results of the pyrimethamine susceptibility assay. The BSDr-yFCU fusion was amplified from the pl9 plasmid from Avnish Patel & David Baker (unpublished), digested and inserted into the digested pCas9/sg plasmid backbone. After integration of the BSDr cassette into the Cas9 plasmid, the plasmid was linearised using BgtZI (Methods 4.3.1.7) and the guide sequence was fused into the plasmid using infusion ligation (Methods 4.3.1.8) with the In-Fusion HD Cloning Kit (TaKaRa) to create a final circularised plasmid (**Figure 9**).

4.3.2.3.3 Donor plasmid generation

We created 6 donor plasmids (pDonor) to act as the template DNA for HR (plasmids summarised in **Supplementary Table 2**). To create donor plasmids, nucleotide sequences for the *dhfr-ts* genes for respective *Plasmodium spp*. were downloaoded from PlasmoDB and the DHFR domain was annotated as described in 4.3.2.2. The nucleotide sequence for the DHFR domain of *P. malariae, P. falciparum* and *P. knowlesi* was recodonised using GeneArt's Instant Designer with the option to recodonise the nucleotide sequence based on a host preference of *Mus musculus,* and these sequences were used to order synthetic DNA oligonucleotides flanked by 500 base pair regions of the *P. knowlesi* genome either side of the DHFR domain locus. pDonor plasmids were ordered with an ampicillin resistance selectable marker to allow for molecular cloning procedures. Mutations were created using the ordered plasmids through an overlapping PCR with primers outlined in **Supplementary Table 1.** All DHFR domains within plasmids were confirmed for their genotype through Sanger sequencing before being used in transfections.

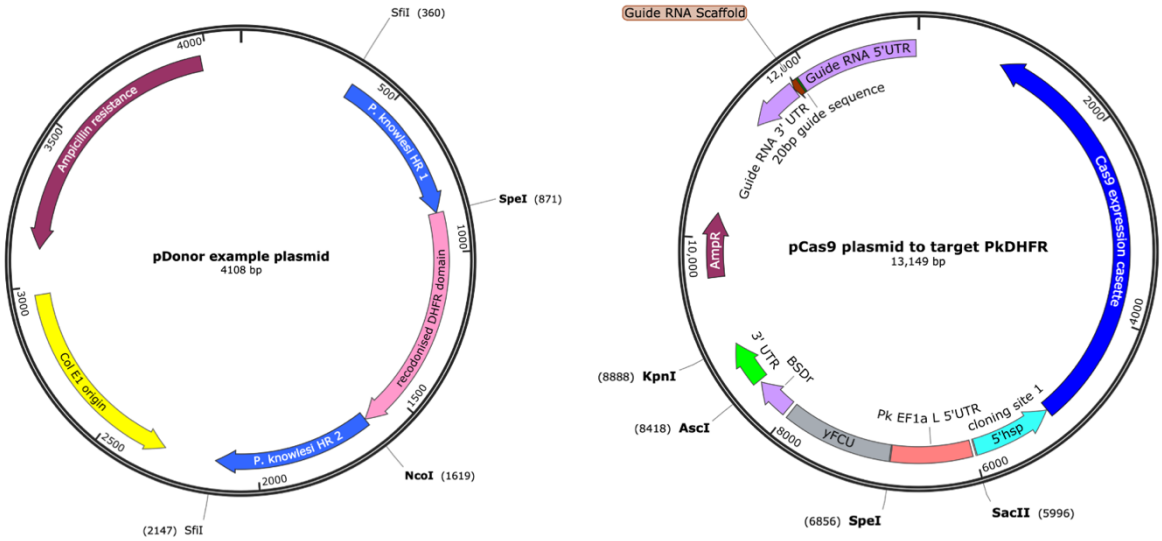


Figure 9. pDonor and pCas9 plasmid maps.
pDonor plasmid (left) containing the *dhfr* domain (pink) flanked with 500 bp homology regions (blue) to the *P. knowlesi* A1H1 reference genome. Ampicilin resistance (burgundy) cassette and the Col E1 origin (yellow) is required for molecular cloning. The pCas9 plasmid (right) contains the expression cassette for the Cas9 protein (blue), the RNA expression cassette (lilac) surrounding the guide RNA sequence (20 bp guide sequence annotated), a blasticidin resistance selectable marker (BSDr), and the 3' and 5' untranslated regions (UTR) for expression of the associated genes.

4.3.3 Parasite cell culture

4.3.3.1 Routine P. knowlesi cell culture

Erythrocytic stages of the laboratory adapted *P. knowlesi* A1-H.1 line were kept under routine culture according to previous guidance [42]. Parasites were maintained in a custom-made RPMI-1640 medium (HEPES Modification, 25 mM HEPES without L-glutamine, Merck, catalogue number: R558), supplemented with 1X L-Glutamine (2 mM) and 10% Horse Serum (v/v), with a 2% haematocrit from human erythrocytes (UK National Blood Transfusion Service) that were confirmed for Duffy positive (Fy+) status. Media was changed every 48 hours, and parasites were kept under 10% parasitaemia (with parasitaemia quantified by light microscopy following Giemsa staining).

4.3.3.2 Synchronisation

Parasites were tightly synchronised using Nycodenz purification as described by Moon *et al*. [42] and Mohring *et al*. [45]. Growth media was removed from parasite cultures, which were subsequently resuspended in media to a 50% haematocrit suspension. Five mls of 55% Nycodenz (Axis-Shield) solution in RPMI (Sigma, catalogue number: R5886) was aliquoted into 15 ml conical tubes and 2 ml of the 50% haematocrit culture was layered on top of the 5 ml 55% Nycodenz, and centrifuged at 900 x g for 12 minutes with low brake/acceleration to separate schizont stages. The resulting top layer of schizonts were transferred to a new conical tube and washed with RPMI before resuspending in 10 mL custom-made RPMI-1640 growth medium with 1 μM compound 2 for 2-3 hours. Compound 2 blocks merozoite egress from late schizonts through inhibition of the Plasmodium cGMP-dependant protein kinase (PKG) and allows for tight synchronization of *P. knowlesi* late schizont stages [45,53]. After 2-3 hours, compound 2 was washed from the schizonts using RPMI and parasites were resuspended in a 2 % haematocrit culture.

4.3.3.3 Transfection of P. knowlesi parasites

Parasites were transfected according to the protocol outlined by Mohring *et al*. [45]. For transfection, DNA to transfect, parasites and blood were prepared prior to nucleofection.

4.3.3.3.1 DNA

~ 20 μg pCas9 plasmid was combined with ~ 60 μg pDonor plasmid (equating to a 1:20 molar ratio) in a total volume of 12 μl in sterile TE buffer (Invitrogen). Resulting mixture of pCas9 and pDonor plasmids was mixed with 100 μl of supplement P3 transfection buffer (Lonza).

4.3.3.3.2 Blood

Fresh Fy+ positive blood was centrifuged at 3000 RPM and for each transfection, 200 μl of 100% haematocrit red blood cells were transferred into individual 1.5 mL reaction tubes (Eppendorf). 500 μl complete growth media was added to each red blood cell aliquot and the tubes were incubated at 37 °C shaking at 550 rpm in a Thermomixer (Eppendorf).

4.3.3.3.3 Parasites

Late stage schizonts were purified from a *P. knowlesi* A1H1 culture using Nycodenz and Compound 2. Parasites were washed using RPMI and for each transfection, 10-20 μl schizonts were transferred to a 1.5 mL reaction tube and resuspended in 100 μl complete growth media. Mature schizonts were incubated at 37 °C for a maximum of 20 minutes to allow for merozoite egress to begin.

4.3.3.3.4 Nucleofection

Once all three components were prepared, schizonts were spun at 845 rcf for 1 minute and the growth media removed. 100 μl of the DNA-transfection buffer mixture generated in step 4.3.3.3.1 was mixed with the schizont pellet (step 4.3.3.3.3) and transferred into a nucleofection cuvette (Lonza), which was placed in the Amaxa 4D-Nucleofector X (experiment settings: Pulse code: FP 158, Solution: Primary Cell P3, Volume: 100 μl) and nucleofection occured following the instrument settings [45].

After each individual transfection, schizonts were immediately transferred from the cuvette into each individual pre-prepared and warmed blood aliquots created in step 4.3.3.3.2 and incubated at 37 °C for 30 minutes shaking at 550 rpm to allow for erythrocyte invasion. Once incubated, the culture was transferred to a 6 well plate where 4.5 mL of complete growth media was added to create a total 5 mL culture with a 4% haematocrit. At this stage, growth media was changed daily, with the addition of the positive selection drug from the following day.

4.3.3.4 Positive selection

Blasticidin was used for positive drug selection to determine successfully transfected parasites, as the pCas9 plasmid contained a BSDr cassette. *P. knowlesi* is 22-fold less susceptible to blasticidin than *P. falciparum* [54], therefore we wanted to use a higher concentration than is routinely used for *P. falciparum* (5 μg/mL [55]). We initially worked with a 5X higher concentration of blasticidin (25 μg/mL) which allowed for selection of transfected parasites. Parasite growth medium supplemented with 25 μg/mL blasticidin was changed daily for 5 days, and parasitaemia was monitored daily. Once parasites began to increase in number (typically on day 3/4), an aliquot of the parasite culture was removed to allow for genomic confirmation of transfected parasites as described in Methods 4.3.3.6.

4.3.3.5 Limiting clonal dilution

After confirmation of transfectant parasites in the bulk culture, limiting clonal dilution in 96 well plates was performed to generate monoclonal transfectant parasite lines [42,43]. Parasite cultures were diluted to approximately 0.3 parasites/100 μl in a 2% haematocrit culture with growth media supplemented with 200 mM L-glutamax (Sigma). 10 ml of diluted culture was plated into a flat-bottom 96 well plate (100 μl per well), and after 7 days the plate was screened for plaques, using an assay adapted from Thomas *et al*. [56]. Cultures from wells positive for single plaques were transferred to 24 well plates and resuspended in 1 mL complete growth media with 2% haematocrit. Aliquots from these 1 ml cultures were taken to confirm genomic integration through genotyping (Methods 4.3.3.6).

4.3.3.6 Screening for positive clones

DNA was extracted from aliquots of parasite bulk parasite cultures, and monoclonal cultures after limiting dilution. Aliquots were taken from cultures and red blood cells were pelleted by centrifugation at 700 x g, the supernatant was removed, and the pellet was washed three times using RPMI. The final RBC pellet was frozen and later genomic DNA was extracted using the Blood genomicPrep Mini Spin Kit (Cytiva).

Once extracted, the genomic DNA is used in a diagnostic PCR to confirm integrated DNA into the *P. knowlesi* genome. Primers were designed to amplify a region spanning at least 100 bp outside of the homology region in the *P. knowlesi* genome and including the modified locus (**Figure 10**). For each transfected isolate, a diagnostic PCR was completed to check for both integration, and the presence of the wild-type locus, in addition to an independent PCR at a different locus in the *P. knowlesi* genome, which should be positive for both wild-type and transfectant lines. Diagnostic PCR primers are summarised in **Supplementary Table 1**, and the amplification methodology is described in Methods 4.3.1.1. DHFR domains were amplified from genomic DNA using CloneAmp and confirmed for integration using sanger sequencing (Eurofins). The indepdant locus is located in the myosin A-tail interacting protein (*pkmtip*) as previously described [42].

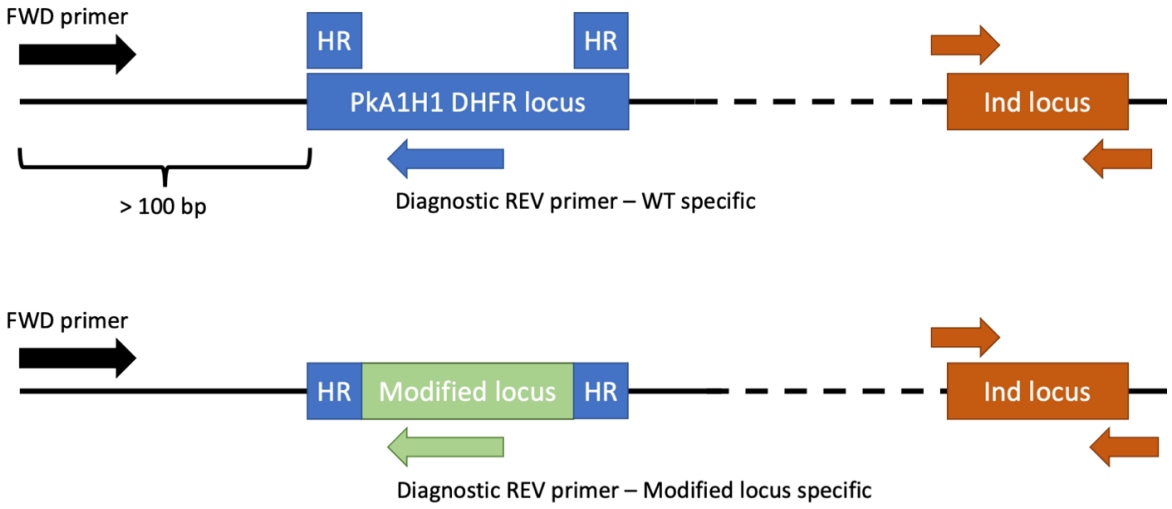


Figure 10. Diagnostic PCR primer design methodology.
A conserved forward primer (FWD primer) was designed upstream of the modified locus within the *P. knowlesi* genome, which is used for all integration PCRs, in addition to species-specific reverse primers (Diagnostic REV primer). This is in addition to an independent locus within the *pkmtip* gene which acts as a positive control. Primers are summarised in **Supplementary Table 1.**

4.3.3.7 Drug assays

Parasite cultures were diluted to 10 mL cultures of 1% parasitaemia in 2% haematocrit before setting up the drug assay. Drug dilutions were prepared and plated into 96 well plates as described below. For all assay plates, 100 μl of 20 μM chloroquine in RPMI was aliquoted into all wells of the first column (A1:H1), alongside 100 μl of complete growth media into every well of the second column (A2:H2) (**Figure 11**). All assays were tested for susceptibility to dihydroartemisinin (DHA), where 200 μl 200nM DHA was aliquoted into the third column in triplicate (D3:F3), and 100 μl growth media was aliquoted into the subsequent columns for that row (D4:D12, E4:E12, F4:F12). DHA was diluted through a two-fold serial dilution by 100 μl taken from the initial well (200 nM) in column 3 and mixed into column 4, and 100 μl from column 4 mixed into column 5 etc. until the final column of the plate. Following drug serial dilutions, 100 μl of the pre-prepared parasite culture was added into each well of the plate, to create a final starting drug concentration of 100 nM.

Serial dilutions were also created for pyrimethamine concentrations, with the starting concentration for pyrimethamine resistant lines (before parasite culture added) as 160 μM, decreasing to 156 nM through a two-fold serial dilution. 100 μl of the pre-prepared parasite culture was added to each well to take the starting drug concentration to 80 μM. For pyrimethamine sensitive lines, the starting

concentration (before parasite culture added) was 200 nM, decreasing to 0.19 nM through a two-fold serial dilution. 100 μl of the pre-prepared parasite culture was added to each well to take the starting drug concentration to 100 nM (**Figure 11**).

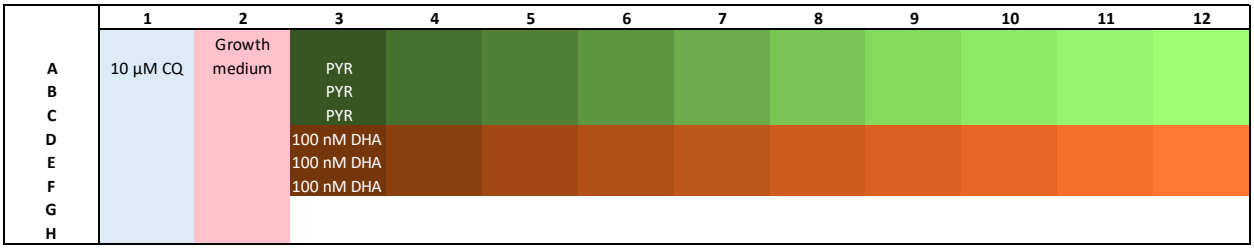


Figure 11. Drug assay plate with final drug concentrations on a 96 well plate.
Chloroquine = CQ, pyrimethamine = PYR, dihydroartemisinin = DHA. Starting concentration of Pyrimethamine for resistant lines (wells A3-C3) was 80000 nM (therefore 160 μM is plated before addition of the parasite culture) decreasing to 156 nM through a two-fold serial dilution. Pyrimethamine starting concentration for sensitive lines started at 100 nM (200 nM plated before adding parasite culture) and decreased to 0.19 nM through a two-fold serial dilution. Starting concentration for dihydroartemisinin in all lines was 100 nM (200 nM plated) decreasing to 0.19 nM through a two-fold serial dilution.

4.4 Results

4.4.1 *In silico* analysis of *PmDHFR* protein structure

In *P. falciparum,* the DHFR enzyme is a bifunctional enzyme coupled with thymidylate synthase (TS), resulting in DHFR-TS. The two enzyme domains are separated by an inactive linker region which affects the catalysis of both enzymes and interacts with the DHFR active site [57], and once expressed, the whole protein forms a homodimer.

4.4.1.1 Amino acid alignments

The global database of *P. malariae* WGS data identified 66 mutations in the *pmdhfr-ts* gene, of these mutations, 15 led to non-synonymous amino acid substitutions in the coding sequence (**Table 1**). Four mutations (S49R, F57L, R58S and N114S) were highlighted as potentially relevant for further investigation due to their proximity to amino acid substitutions associated with resistance in the *P. falciparum* ortholog upon an amino acid alignment (**Figure 4**), however the results presented will focus on the pre-described lines in **Table 2;** the PmUG01 reference sequence, PmUG01 sequence + F57L mutation, and the PmUG01 sequence + R58S and N114S. The mutation at position 15 (A15S)

was not investigated in this study, as mutations at this locus (corresponding to A16 in *P. falciparum*) are associated with cycloguanil resistance rather than pyrimethamine [58].

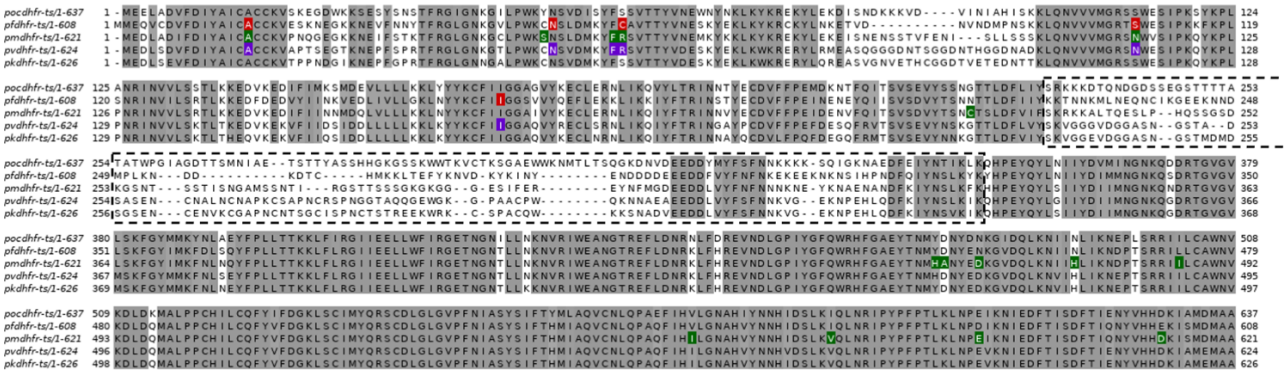


Figure 4. DHFR alignment across *Plasmodium spp.*
ClustalO was used for amino acid alignment of the five human-infective Plasmodium spp. with an available sequence for DHFR-TS in PlasmoDB. The linker domain is outlined with a black dashed box, with positions taken from PlasmoDB. The DHFR domain is at the 5' end of the linker region (upstream) and the TS domain is at the 3' end of the linker region (downstream). Amino acid positions associated with pyrimethamine resistance in *P. falciparum* and *P. vivax* are highlighted in red and purple respectively. Positions found with amino acid substitutions in the *P. malariae* global database are highlighted in green.

4.4.1.2 Tertiary structure analysis

To further investigate the potential relevance of these amino acid positions, we generated a predicted 3D protein structure for *pmdhfr-ts* (blue structure) using the amino acid sequence from PlasmoDB and I-TASSER for structure prediction, and aligned it to the crystal structure of *pfdhfr-ts* (grey structure) bound to pyrimethamine obtained from the Protein Databank (RCSB PDB, https://www.rcsb.org/, structure = 3QGT). The positions associated with pyrimethamine resistance in *P. falciparum* are highlighted in red and overlap in location with the four positions (S49R, F57L, R58S and N114S) found in the *P. malariae* database (highlighted in green), localised around the binding site for pyrimethamine, as indicated by an asterisk next to the small molecule structure of pyrimethamine (**Figure 5**).

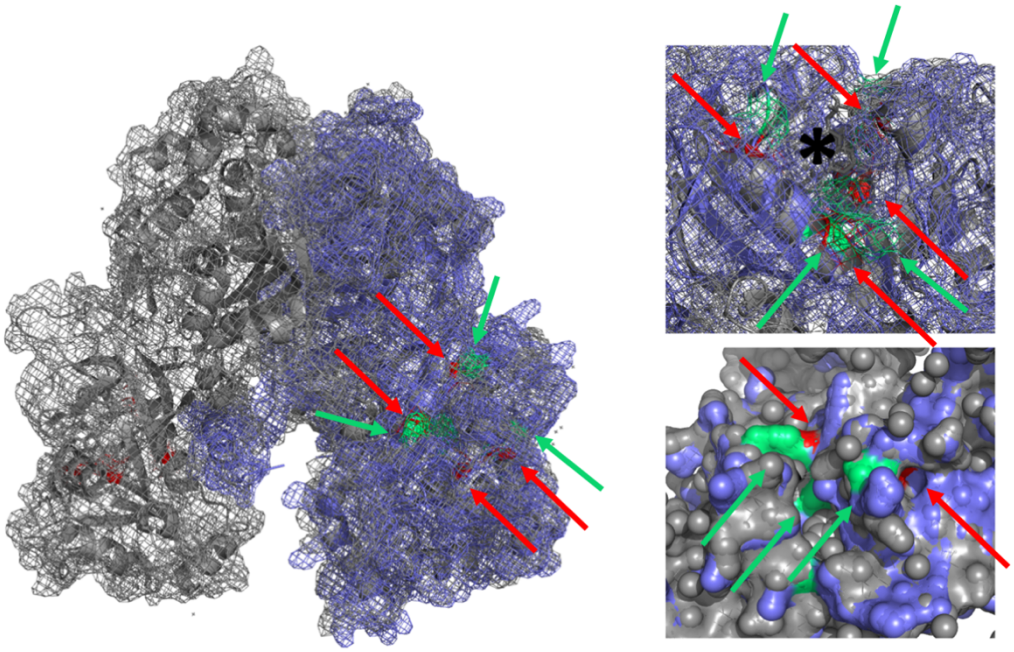


Figure 5. *P. falciparum* and *P. malariae* DHFR-TS protein structures with amino acid residues of potential importance highlighted.
Crystal structure of the 3D7 *pfdhfr-ts* homodimer (grey) downloaded from RCSB PDB, aligned with the predicted structure of *pmdhfr-ts* monomer (structure prediction with I-TASSER [49] and the PmUG01 reference amino acid sequence from PlasmoDB) (blue), visualisation and alignment performed using PyMOL [48]. Amino acid positions associated with pyrimethamine resistance in *P. falciparum* are highlighted in red and labelled with red arrows (51, 59, 108 and 164) and amino acid positions found to be mutated in the *P. malariae* genomics database that were predicted to be relevant (49, 57, 58, 114) are highlighted in green and labelled with green arrows. The crystal structure of *pfdhfr-ts* used has pyrimethamine bound (highlighted by an asterisk) to highlight the active site of the enzyme. Both zoomed in images on the right hand side are taken at the same orientation.

4.4.2 Successful ortholog replacement at the *pkdhfr-ts* locus

As described in Methods 4.3.2.3, we generated one pCas9 plasmid for Cas9 expression targeted at the *pkdhfr-ts* gene locus using a 20 bp guide sequence, and six donor plasmids containing each DHFR domain of interest flanked by 500 bp homology regions to the *P. knowlesi* A1H1 genome, allowing for homologous recombination (**Figure 6**).

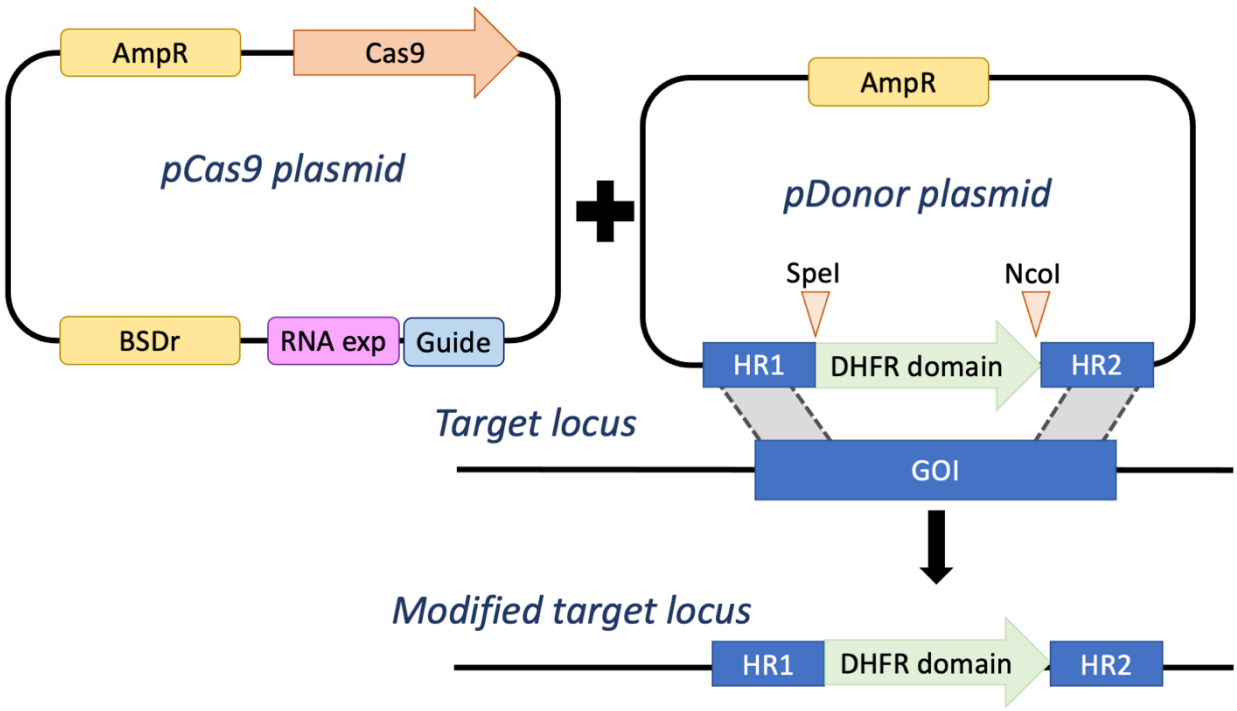


Figure 6. An overview of CRISPR-Cas9 genome editing using a pCas9 plasmid and a pDonor plasmid containing donor DNA for HR.
The pCas9 plasmid contains ampicillin (AmpR) and blasticidin (BSDr) selectable markers in addition to the guide sequence (Guide) and RNA expression cassette (RNA exp), and the expression cassette for the Cas9 endonuclease (Cas9). The pDonor plasmid contains the donor DNA construct (DHFR domain) flanked by regions homologous to the *P. knowlesi* target gene sequence (HR1 and HR2). After transfection into parasites using nucleofection, the Cas9 endonuclease is expressed alongside the sgRNA and guided to the target locus by the guide sequence, where it creates a DSB. The pDonor plasmid is used to repair the DSB through HR.

Wild-type cultures of the *P. knowlesi* A1H1 line were synchronised, and late stage schizonts (27 hour) were used for nucleofection with a pre-mixed combination of pDonor and pCas9 plasmids (plasmids outlined in **Supplementary Table 2**). Resulting cultures were grown under drug selection (with blasticidin) for 1 week, with an initial diagnostic PCR to check for integration into the genome on day 7. Previously, pyrimethamine has been used for positive selection of transfected cultures of *P. knowlesi,* and whilst blasticidin selection is commonplace in *P. falciparum*, it had not been used for *P. knowlesi*. Schalkwyk *et al*. demonstrated that *P. knowlesi* was 22-fold less susceptible to blasticidin than *P. falciparum* [54], therefore we aimed to determine an appropriate concentration of blasticidin to use for positive selection of integrated parasites. In the case of *P. falciparum*, 5 µg/mL has been used for positive selection of CRISPR-Cas9 transfectant parasites [55]. Initially, we attempted to increase

this concentration 5-fold to 25 μg/mL and demonstrated that this concentration was sufficient to select for transfectant *P. knowlesi* isolates, with positive transfected lines for all integrations.

4.4.2.1 Transfectant lines created

Initially, we set out to investigate the differences between pyrimethamine susceptibility in *P. knowlesi, P. malariae* and *P. falciparum*. Three mutations at the *dhfr* locus in *P. falciparum* are known to create pyrimethamine resistance. Therefore, we created two donor plasmids for the *P. falciparum dhfr* domain, one wild type with the 3D7 reference sequence (pyrimethamine sensitive *P. falciparum* line, plasmid 304), and one containing the triple mutation (N51I, C59R, S108N, plasmid 273) to create a pyrimethamine resistant control parasite line. Additionally, we created a donor plasmid for the PmUG01 reference sequence (plasmid 274). After developing these plasmids, we used the plasmid 274 to generate different *P. malariae dhfr* loci containing individual (or combinations) of mutations identified in the genomics database (**Supplementary Table 2**), and this chapter will focus on two of these generated plasmids, 295 which contains mutation 5 (F57L) and 304_1 which contains mutations 3 and 6 (R58S, N114S).

After drug selection, parasite lines were initially screened with a diagnostic PCR to check for integration into the *pkdhfr-ts* locus (as described in Methods 4.3.3.6). After gDNA extraction, and amplification with the diagnostic PCR, DNA was visualised on an agarose gel to verify the presence of the correct size fragment. Primers for each diagnostic PCR and the resulting amplicon size after amplification are summarised in **Table 3**. Following the confirmation of successful integration into the *P. knowlesi* genome, parasite cultures (which likely contain a mix of wild-type parasites in addition to successfully transfected parasites, known as a bulk culture) underwent limiting dilution to obtain a monoclonal successfully transfected parasite line as described in Methods 4.3.3.5. After limiting dilution, parasite cultures were once again screened using the diagnostic PCR, and those with a positive band for the integration and independent PCR (amplification at an unrelated *P. knowlesi* locus, which should be positive in all clones), and a negative result for the wild-type PCR were selected as monoclonal lines and kept in culture for further experiments. The primers used for the diagnostic PCR are listed in **Supplementary Table 1**, with the independent PCR (primers 75 and 76) amplifying a 1043 bp fragment, the *P. malariae* integration PCR (primers olAI034 and olAI037) amplifying a 748 bp fragment, the *P. knowlesi* recodonised *pkdhfr* integration PCR (primers olAI034

and olAI036) amplifying a 754 bp fragment, the *P. falciparum* integration PCR (primers olAI034 and olAI039) amplifying a 749 bp fragment, and the wild type PCR (primers olAI034 and olAI035) amplifying a 746 bp fragment. One monoclonal line was chosen for each integration and grown for subsequent experiments (**Figure 7**). Monoclonal lines were preserved through freezing in liquid nitrogen.

Table 3. Primers and product sizes for diagnostic PCR.

		Integration PCR		Wild-type PCR		Independent PCR	
Parasite line	Description	Primers	Amplicon size (bp)	Primers	Amplicon size (bp)	Primers	Amplicon size (bp)
WT_A1H1	Wild type	All combinations*	NA	olAI034, olAI035	746	75, 76	1043
TF 131_2	PkrDHFR	olAI034, olAI036	754	olAI034, olAI035	NA	75, 76	1043
TF 150	Pf 3D7	olAI034, olAI039	749	olAI034, olAI035	NA	75, 76	1043
TF 135_1	Pf IRN	olAI034, olAI039	749	olAI034, olAI035	NA	75, 76	1043
TF 150_1	Pm sensitive	olAI034, olAI037	748	olAI034, olAI035	NA	75, 76	1043
135_3	PmUG01	olAI034, olAI037	748	olAI034, olAI035	NA	75, 76	1043
151	Pm mutant 5	olAI034, olAI037	748	olAI034, olAI035	NA	75, 76	1043

*To confirm that the integration PCR was specific for integrated loci, each integration combination (olAI034 with; olAI036, olAI039 and olAI037) was tested independently with the wild type *P. knowlesi* A1H1 culture to wield no amplification.

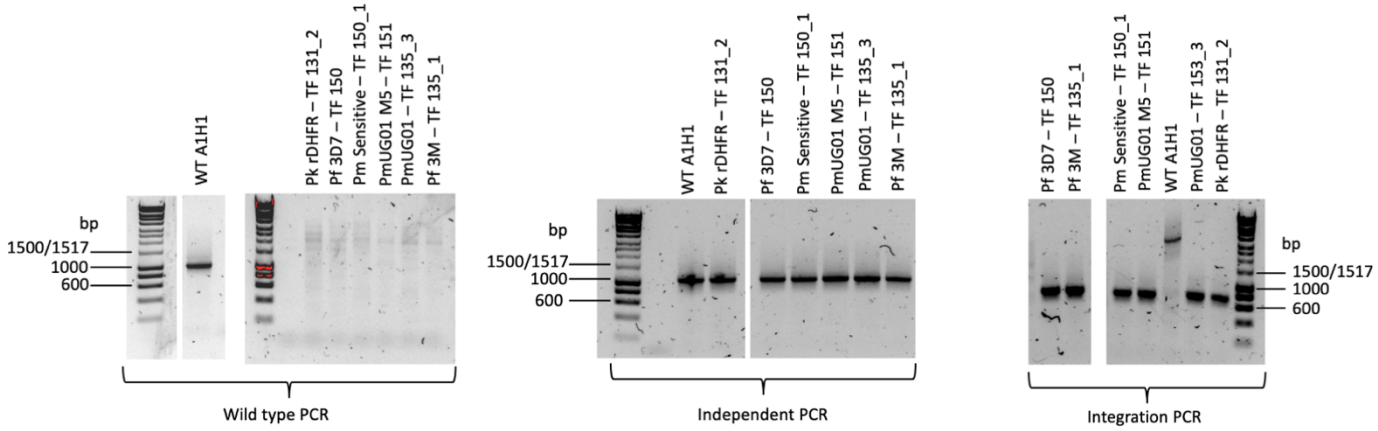


Figure 7. Monoclonal parasite lines confirmed through diagnostic PCR.
Bulk cultures for all integrated parasite lines underwent limiting dilution and were amplified using the diagnostic PCR for a second time. Parasite lines: WT_A1H1 = wild-type *P. knowlesi* culture, TF 131_2 = *P. knowlesi* recodonised *dhfr* domain, TF 150 = Pf 3D7 *dhfr* domain, TF135_1 = Pf *dhfr* domain with triple mutation (N51I, C59R, S108N), TF 150_1 = Pm 'sensitive' *dhfr*, 135_3 = PmUG01 reference sequence, 151 = Pm *dhfr* with mutation 5. Parasite lines were investigated using A) a wild type PCR using primers olAI034 and olAI035, which amplify a 746 bp fragment if a wild-type *P. knowlesi dhfr* locus is present, as seen in the WT lane. B) The independent PCR using primers 75 and 76 which target an independent locus in the *P. knowlesi* genome, amplifying a 1043 bp fragment in all parasite lines and C) The integration PCR which is specific for each species *dhfr*

domain (for *pfdhfr* domains = 749 bp, for *P. knowlesi* recodonised = 754 bp, and for *pmdhfr* domains = 748 bp).

4.4.3 *P. malariae* pyrimethamine resistance phenotype

After confirmation of monoclonal parasite lines, all lines were subject to drug assays with both DHA and pyrimethamine following previously described protocols [54] and as outlined in Methods 4.3.3.7. For all parasite lines, a starting parasitaemia of 1% was used with 1% haematocrit, with parasite cultures exposed to drug pressure for one life cycle (27 hours).

The susceptibility of all 7 parasite lines to DHA was similar, with regression analysis showing no statistically significant difference in EC_{50} when comparing TF 131_2 ($p = 0.18$), n repeats = 3), TF 150_1 ($p = 0.98$, n repeats = 5) and TF 135_3 ($p = 0.17$, n repeats = 3) to the wild type *P. knowlesi* A1H1 culture using regression analysis. There were significant differences between the wild-type line (WT_A1H1) and TF150 ($p = 0.026$, n repeats = 3), TF 135_1 ($p = 0.046$, n repeats = 3), and TF 151 ($p = 0.00437$, n repeats = 3) lines. Whilst there was variation between parasite lines, all EC_{50} data for DHA was below 6 nM, which is similar to previous studies with *P. knowlesi* (DHA EC_{50} of 2.0 nM ± 0.3) and *P. falciparum* (DHA EC_{50} of 4.2 ± 0.5)[54]. It is important to note that EC_{50}'s are not the most accurate way to assay for parasite susceptibility to DHA as resistant parasites are able to undergo dormancy whilst exposed to the drug, and if specifically interested in resistance to DHA, a ring-stage survival assay (RSA) should instead be used. RSAs involve exposing tightly synchronised parasite ring stages to DHA before washing off the drug and keeping the remaining parasites in culture. After 66 hours of culture (in the case of *P. falciparum*) remaining parasites are counted and survival rates can be calculated [59].

Interestingly, regression analysis demonstrated marked differences in the susceptibility of parasite lines to pyrimethamine. The *P. knowlesi* A1H1 lab adapted parasite line demonstrated a mean EC_{50} of 4.20 nM which correlates with previous estimates of 5.10 nM [54]. The 131_2 line transfected with a recodonised *dhfr* domain also demonstrated an EC_{50} of 5.30 nM, with no significant differences between 131_2 and the wild-type susceptibility to pyrimethamine ($p = 0.44$, n repeats = 4). The parasite line transfected with the *P. falciparum* 3D7 *dhfr* domain (TF 150) demonstrated a mean EC_{50} of 25.10 nM (5.9-fold less susceptible than the *P. knowlesi* A1H1 line), which was a statistically different susceptibility ($p = 0.00026$, $n = 2$). Previous work with the 3D7 culture adapted *P.*

falciparum line demonstrated that *P. falciparum* isolates are 10.6-fold less susceptible to pyrimethamine than *P. knowlesi*, and therefore this decrease in susceptibility was expected. The TF 150 line was only 5.9-fold less susceptible to pyrimethamine than *P. knowlesi*, and this may be due to other *P. falciparum* genes affecting the susceptibility. We also demonstrate a *P. malariae* pyrimethamine susceptible line (TF 150_1) which has a mean EC_{50} of 13.5 nM which is statistically significant in comparison to the *P. knowlesi* A1H1 line ($p = 0.0032$, $n = 3$), which is the line demonstrating two mutations seen within the global dataset (in comparison to the PmUG01 reference genome).

The parasite line transfected with the triple mutant *P. falciparum* DHFR domain (TF 135_3) displayed a highly pyrimethamine resistant phenotype with a mean EC50 of 45,878 nM (± 28,247), which was expected due to the high levels of clinical resistance seen when this haplotype is found in clinical settings and is statistically significant using regression analysis ($p = 1.29 \times 10^{-15}$, $n = 5$). Additionally, the parasite line transfected with the *P. malariae* reference sequence for *dhfr-ts* (PmUG01) displayed reduced susceptibility to pyrimethamine (mean EC_{50} = 4, 648 nM, ± 2,480) which was statistically significant ($p = 7.06 \times 10^{-13}$, $n = 3$). The PmUG01 reference sequence encodes an asparagine amino acid at position 114, which aligns with the *P. falciparum* 108 position in amino acid and tertiary structure alignment (**Figure 4, Figure 5**), where the S108N mutation is associated with pyrimethamine resistance [16]. The 114N genotype is seen within the PmUG01 reference strain it is important to note that the 114S genotype is present in 71% of the global population, with the highest presence of the potentially resistant phenotype (114N) in Africa (33%), followed by Asia (16%) and is not seen in isolates from Oceania and South America (**Table 1**). The parasite line with PmUG01 and mutation 5 (F57L), known as PmUG01_M5 (TF 151) also displayed a highly resistant phenotype (EC50 = 30,183 nM ± 8,138), which was statistically significant ($p = 1.09 \times 10^{-13}$, $n = 2$). The parasite line with mutation 5 was 6-fold more resistant than the PmUG01 parasite line, indicating that a SLRN haplotype at positions 49, 57, 58, 114 displays the highest level of resistance in *P. malariae* that has been demonstrated to date. (**Figure 8, Table 4**).

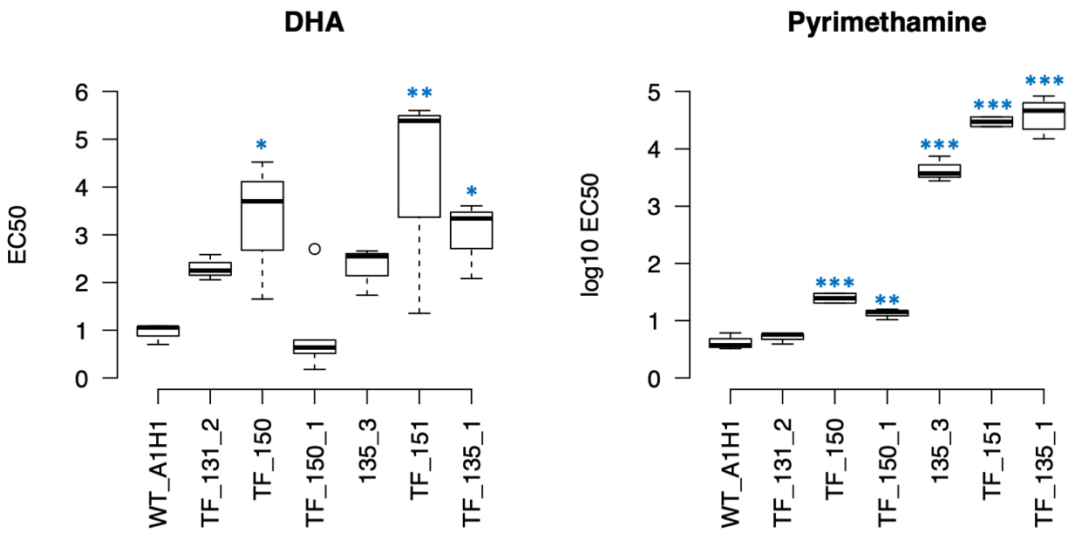


Figure 8. Pyrimethamine resistance in *P. malariae* isolates.
Parasite line susceptibility to both dihydroartemisinin (DHA) and pyrimethamine, with regression analysis comparing each line to WT_A1H1 to investigate significant differences between lines. Significance codes: p = 0-0.001, ***; p = 0.001-0.01, **; p = 0.01-0-.1, *. EC_{50} for DHA is measured in nM, and the log10 EC_{50} is used for pyrimethamine. Clone descriptions: WT_A1H1 = Wild type *P. knowlesi* culture, TF_131_2 = *P. knowlesi* transfected with recodonised reference DHFR sequence, TF_150 = transfected with *P. falciparum* 3D7 DHFR sequence, TF_150_1 = transfected with *P. malariae* hypothetical sensitive sequence, 135_3 = transfected with *P. malariae* reference PmUG01 sequence, TF_151 = transfected with *P. malariae* reference PmUG01 sequence with mutant 5 (F57L), TF_135_1 = transfected with *P. falciparum* DHFR sequence with IRN triple mutation.

Table 4. Susceptibility to DHA and pyrimethamine in transfectant lines.
Mean EC_{50} for both dihydroartemisinin and pyrimethamine is given with the standard deviation (SD), parasite lines described in **Table 2**.

Isolate	Clone name	Drug	Mean EC50	Standard Deviation
Wild type Pk	WT A1H1	DHA	0.948	0.213
Pk recDHFR	TF 131_2		2.3	0.268
Pf 3D7	TF 150		3.29	1.48
Pm sensitive	TF 150_1		0.965	0.998
Pm mutant 5	TF 151		4.11	2.39
Pm UG01	TF 135_3		2.32	0.506
Pf IRN_DHFR	TF 135_1		3.01	0.814
Wild type Pk	WT A1H1	Pyrimethamine	4.2	1.29
Pk recDHFR	TF 131_2		5.3	0.932
Pf 3D7	TF 150		25.1	6.87
Pm sensitive	TF 150_1		13.5	2.74
Pm mutant 5	TF 151		30183	8138
Pm UG01	TF 135_3		4648	2480
Pf IRN_DHFR	TF 135_1		45878	28247

The results shown demonstrate that mutations within the DHFR domain affect pyrimethamine susceptibility in *P. malariae*, as is known with *P. falciparum* and *P. vivax*. We demonstrate and validate a method for *in vitro* analysis of mutations within DHFR and their effect on pyrimethamine susceptibility using ortholog replacement for parasite species that lack an *in vitro* culture method, which previously would have been extremely difficult to assay. We demonstrate that the PmUG01 reference strain of *P. malariae* from Uganda was likely to have been resistant to pyrimethamine, which may be due to an asparagine amino acid at position 114 within *pmdhfr-ts*, a genotype that is most common in African isolates.

4.5 Discussion

Malaria caused by *P. malariae* is greatly understudied, partly due to a lack of research interest, but also the inability to culture this parasite species *in vitro*. Many pivotal findings regarding *P. falciparum* research have been enabled due to whole genome sequencing projects, for instance, large whole genome sequencing (WGS) projects combined with clinical data which allowed for the determination of the genetic marker for artemisinin resistance in SEA [60]. My previous work has enabled easier WGS of *P. malariae* clinical isolates through an SWGA methodology and using this, I was able to generate a genomic database for this parasite species with high quality data for >150

clinical isolates, where I further investigated the presence of SNPs in orthologs of genes known to be associated with drug susceptibility in *P. falciparum*. I focussed on SNPs found within *Pmdhfr-ts*, a bifunctional enzyme involved in folate biosynthesis that is the target of the antimalarial drug pyrimethamine and found SNPs within this gene that aligned with amino acids known to be associated with resistance in *P. falciparum* at a secondary and tertiary protein alignment. Whilst there is no *in vitro* culture method for *P. malariae* isolates, studies using ortholog replacement in *P. knowlesi* parasite cultures have provided useful for investigating basic biology of *P. vivax*, another species which we are unable to culture *in vitro* [61].

Using this methodology for orthologue replacement of the DHFR locus in the *P. knowlesi* A1H1 parasite line, I was able to demonstrate known pyrimethamine resistant parasite phenotypes (using the known genotype for pyrimethamine resistance in a *P. falciparum* ortholog replacement) and known sensitive parasite phenotypes (using the *P. falciparum* 3D7 genotype in ortholog replacement), as well as demonstrating that the reference sequence for *P. malariae* DHFR (PmUG01 reference, PmUG01_05034700) is likely to be resistant to pyrimethamine. Pyrimethamine resistance in *P. malariae* has not yet been confirmed or identified in clinical settings, however specific research into drug susceptibility of *P. malariae* isolates is lacking, and there have been no direct studies regarding pyrimethamine. Whilst this chapter demonstrates a potentially interesting parasite phenotype, it also demonstrates a methodology to further investigate the unique biology of *P. malariae* parasites and opens up the possibility to investigate other drug targets, and potential invasion genes through ortholog replacement in *P. knowlesi* parasite lines.

4.6 References

1. Plowe, C. V. *et al*. Mutations in Plasmodium falciparum dihydrofolate reductase and dihydropteroate synthase and epidemiologic patterns of pyrimethamine-sulfadoxine use and resistance. *J. Infect. Dis*. **176**, 1590–1596 (1997).
2. Peterson, D. S., Walliker, D. & Wellems, T. E. Evidence that a point mutation in dihydrofolate reductase-thymidylate synthase confers resistance to pyrimethamine in falciparum malaria (Plasmodiumfakiparum/drug resistance/folic acid antagonists/genetic linkage analysis/polymerase chain reaction). *Proc. Nati. Acad. Sci. USA* **85**, 9114–9118 (1988).
3. Imwong, M. *et al*. Association of Genetic Mutations in Plasmodium vivax dhfr with Resistance to Sulfadoxine-Pyrimethamine: Geographical and Clinical Correlates. **45**, 3122–3127 (2001).
4. Betson, M. *et al*. Detection of persistent Plasmodium spp. infections in Ugandan children after artemether-lumefantrine treatment. *Parasitology* **141**, 1880–1890 (2014).
5. Betson, M., Clifford, S., Stanton, M., Kabatereine, N. B. & Stothard, J. R. Emergence of Nonfalciparum Plasmodium Infection Despite Regular Artemisinin Combination Therapy in an 18-Month Longitudinal Study of Ugandan Children and Their Mothers. *J. Infect. Dis*. **217**, 1099–1109 (2018).
6. Maguire, J. D. *et al*. Chloroquine-resistant Plasmodium malariae in south Sumatra, Indonesia. *Lancet* (2002). doi:10.1016/S0140-6736(02)09336-4
7. Teo, B. H. Y. *et al*. Delayed Onset of Symptoms and Atovaquone-Proguanil Chemoprophylaxis Breakthrough by Plasmodium malariae in the Absence of Mutation at Codon 268 of pmcytb. *PLoS Negl. Trop. Dis*. **9**, 1–12 (2015).
8. Naidoo, I. & Roper, C. Mapping 'partially resistant', 'fully resistant', and 'super resistant' malaria. *Trends Parasitol*. **29**, 505–515 (2013).
9. Nzila, A. The past, present and future of antifolates in the treatment of Plasmodium falciparum infection. *J. Antimicrob. Chemother*. **57**, 1043–1054 (2006).
10. Hyde, J. E. Exploring the folate pathway in Plasmodium falciparum. *Acta Trop*. **94**, 191 (2005).
11. Gregson, A. & Plowe, C. V. Mechanisms of Resistance of Malaria Parasites to Antifolates.

Pharmacol. Rev. **57**, 117–145 (2005).

12. Heinberg, A. & Kirkman, L. The molecular basis of antifolate resistance in Plasmodium falciparum: looking beyond point mutations. doi:10.1111/nyas.12662
13. Talisuna, A. O., Bloland, P. & D'Alessandro, U. History, Dynamics, and Public Health Importance of Malaria Parasite Resistance. *Clin. Microbiol. Rev.* **17**, 235 (2004).
14. Checchi, F. *et al*. Malaria Epidemics and Interventions, Kenya, Burundi, Southern Sudan, and Ethiopia, 1999–2004. *Emerg. Infect. Dis.* **12**, 1477 (2006).
15. Winstanley, P., Ward, S., Snow, R. & Breckenridge, A. Therapy of falciparum malaria in sub-Saharan Africa: From molecule to policy. *Clin. Microbiol. Rev.* **17**, 612–637 (2004).
16. Cowman, A. F., Morry, M. J., Biggs, B. A., Cross, G. A. M. & Foote, S. J. Amino acid changes linked to pyrimethamine resistance in the dihydrofolate reductase-thymidylate synthase gene of Plasmodium falciparum. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 9109–9113 (1988).
17. Peterson, D. S., Milhoust, W. K. & Wellems, T. E. *Molecular basis of differential resistance to cycloguanil and pyrimethamine in Plasmodium falciparum malaria. Proc. Nati. Acad. Sci. USA* **87**, (1990).
18. Thaithong, S. *et al*. Plasmodium falciparum: gene mutations and amplification of dihydrofolate reductase genes in parasites grown in vitro in presence of pyrimethamine. *Exp. Parasitol.* **98**, 59–70 (2001).
19. Hyde, J. E. Antifolate resistance in Africa and the 164-dollar question. *Trans. R. Soc. Trop. Med. Hyg.* **102**, 301 (2008).
20. Chaturvedi, R. *et al*. Geographical spread and structural basis of sulfadoxine-pyrimethamine drug-resistant malaria parasites. *Int. J. Parasitol.* **51**, 505–525 (2021).
21. Roper, C. *et al*. Antifolate antimalarial resistance in southeast Africa: a population-based analysis. *Lancet* **361**, 1174–1181 (2003).
22. van Eijk, A. M. *et al*. Effect of Plasmodium falciparum sulfadoxine-pyrimethamine resistance on the effectiveness of intermittent preventive therapy for malaria in pregnancy in Africa: a systematic review and meta-analysis. *Lancet Infect. Dis.* **19**, 546–556 (2019).
23. Harrington, W. E. *et al*. Competitive facilitation of drug-resistant Plasmodium falciparum malaria parasites in pregnant women who receive preventive treatment. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 9027–9032 (2009).
24. Gutman, J. *et al*. The A581G Mutation in the Gene Encoding Plasmodium falciparum

Dihydropteroate Synthetase Reduces the Effectiveness of Sulfadoxine-Pyrimethamine Preventive Therapy in Malawian Pregnant Women. *J. Infect. Dis.* **211**, 1997 (2015).

25. Blasco, B., Leroy, D. & Fidock, D. A. Antimalarial drug resistance: linking Plasmodium falciparum parasite biology to the clinic. **23**, (2017).
26. YOUNG, M. D. & BURGESS, R. W. Pyrimethamine resistance in Plasmodium vivax malaria. *Bull. World Health Organ.* **20**, 27 (1959).
27. Eldin De Pécoulas, P., Basco, L. K., Tahar, R., Ouatas, T. & Mazabraud, A. Analysis of the Plasmodium vivax dihydrofolate reductase–thymidylate synthase gene sequence. *Gene* **211**, 177–185 (1998).
28. Tanomsing, N. *et al*. Genetic Analysis of the Dihydrofolate Reductase-Thymidylate Synthase Gene from Geographically Diverse Isolates of Plasmodium malariae. *Antimicrob. Agents Chemother.* **51**, 3523 (2007).
29. WHO. Guidlines for the treatment of malaria - 3rd edition. **85**, (2015).
30. Amusengeri, A., Tata, R. B. & Bishop, Ö. T. Understanding the Pyrimethamine Drug Resistance Mechanism via Combined Molecular Dynamics and Dynamic Residue Network Analysis. *Molecules* **25**, (2020).
31. Nzila, A. M. *et al*. Molecular evidence of greater selective pressure for drug resistance exerted by the long-acting antifolate Pyrimethamine/Sulfadoxine compared with the shorter-acting chlorproguanil/dapsone on Kenyan Plasmodium falciparum. *J. Infect. Dis.* **181**, 2023–2028 (2000).
32. McKeage, K. *et al*. Atovaquone/proguanil: a review of its use for the prophylaxis of Plasmodium falciparum malaria. *Drugs* **63**, 597–623 (2003).
33. Kemnic, T. R. & Coleman, M. Trimethoprim Sulfamethoxazole. *Mayo Clin. Proc.* **74**, 730–734 (2021).
34. Wiktor, S. Z. *et al*. Efficacy of trimethoprim-sulphamethoxazole prophylaxis to decrease morbidity and mortality in HIV-1-infected patients with tuberculosis in Abidjan, Côte d'Ivoire: a randomised controlled trial. *Lancet (London, England)* **353**, 1469–1475 (1999).
35. Sridaran, S. *et al*. Anti-folate drug resistance in Africa: Meta-analysis of reported dihydrofolate reductase (dhfr) and dihydropteroate synthase (dhps) mutant genotype frequencies in African Plasmodium falciparum parasite populations. *Malar. J.* **9**, 1–22 (2010).
36. S Lee, M. C., Lindner, S. E., Lopez-Rubio, J.-J. & Llinás, M. Cutting back malaria:

CRISPR/Cas9 genome editing of Plasmodium. *Brief. Funct. Genomics* **18**, 281–289 (2019).

37. Li, H. *et al*. Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. *Signal Transduct. Target. Ther. 2020 51* **5**, 1–23 (2020).
38. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat. Commun. 2018 91* **9**, 1–13 (2018).
39. Lee, A. H., Symington, L. S. & Fidock, D. A. DNA Repair Mechanisms and Their Biological Roles in the Malaria Parasite Plasmodium falciparum. *Microbiol. Mol. Biol. Rev.* **78**, 469 (2014).
40. Ghorbal, M. *et al*. Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system. *Nat. Biotechnol. 2014 328* **32**, 819–821 (2014).
41. Böhme, U. *et al*. Complete avian malaria parasite genomes reveal features associated with lineage-specific evolution in birds and mammals. *Genome Res.* **28**, 547–560 (2018).
42. 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).
43. 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).
44. Yahata, K. *et al*. Gliding motility of Plasmodium merozoites. (2020). doi:10.1101/2020.05.01.072637
45. Mohring, F., Hart, M., Patel, A., Baker, D. & Moon, R. CRISPR-Cas9 Genome Editing of Plasmodium knowlesi. *Bio-protocol* **10**, (2020).
46. Pain, A. *et al*. The genome of the simian and human malaria parasite Plasmodium knowlesi. *Nature* **455**, 799–803 (2008).
47. Kibbe, W. A. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.* **35**, (2007).
48. Schrödinger, L. *The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC*. (2015).
49. Yang, J. & Zhang, Y. I-TASSER server: new development for protein structure and function predictions. *Web Serv. issue Publ. online* **43**, (2015).
50. Rutledge, G. G. *et al*. Plasmodium malariae and P. ovale genomes provide insights into malaria parasite evolution. *Nature* **542**, 101–104 (2017).

51. Gardner, M. J. *et al*. Genome sequence of the human malaria parasite Plasmodium falciparum. *Nat. 2002 4196906* **419**, 498–511 (2002).
52. Benavente, E. D. *et al*. A reference genome and methylome for the Plasmodium knowlesi A1-H.1 line. *Int. J. Parasitol*. (2017). doi:10.1016/j.ijpara.2017.09.008
53. Collins, C. R. *et al*. Malaria Parasite cGMP-dependent Protein Kinase Regulates Blood Stage Merozoite Secretory Organelle Discharge and Egress. *PLOS Pathog*. **9**, e1003344 (2013).
54. 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).
55. Patel, A. *et al*. Cyclic AMP signalling controls key components of malaria parasite host cell invasion machinery. *PLOS Biol*. **17**, e3000264 (2019).
56. Thomas, J. A. *et al*. Development and Application of a Simple Plaque Assay for the Human Malaria Parasite Plasmodium falciparum. *PLoS One* **11**, e0157873 (2016).
57. Dasgupta, T. & Anderson, K. S. Probing the Role of Parasite-specific, Distant, Structural Regions on Communication and Catalysis in the Bifunctional Thymidylate Synthase-Dihydrofolate Reductase from Plasmodium falciparum NIH Public Access. *Biochemistry* **47**, 1336–1345 (2008).
58. Peterson, D. S., Milhous, W. K. & Wellems, T. E. Molecular basis of differential resistance to cycloguanil and pyrimethamine in Plasmodium falciparum malaria. *Proc. Natl. Acad. Sci. U. S. A*. **87**, 3018–3022 (1990).
59. Zhang, J. *et al*. Overview of the improvement of the ring-stage survival assay-a novel phenotypic assay for the detection of artemisinin-resistant Plasmodium falciparum. *Sci. Press Zool. Res*. **38**, 317–320 (2017).
60. Miotto, O. *et al*. Genetic architecture of artemisinin-resistant Plasmodium falciparum. *Nat. Genet*. **47**, 226–34 (2015).
61. 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).

4.7 Supplementary information

Supplementary table 1. Primers used in this study

Name	Sequence (5'-3')	Purpose	Notes
olAI003	TTACAGTATATTATTccaagtacgagaagttaaagGTTTTAGAGCTAGAA	cas9 guide RNA forward	seed 2 FWD
olAI004	TTCTAGCTCTAAAACctttaacttctcgtacttggAATAATATACTGTAA	cas9 guide RNA reverse	seed 2 REV
olAI015	CATGGGCAGATCCTCCTGGGTGTCAATCCCCAAGCAGTAC	introducing mutation in pm DHFR	Pm mut 3 fwd
olAI016	ATTGACACCCAGGAGGATCTGCCCATGACCACCAC	introducing mutation in pm DHFR	Pm mut 3 rev
olAI030	CAGCCTGGACATGAAGTACTTACGGAGCGTGACCACCTACGT	introducing mutation in pm DHFR	Mut5_FWD
olAI031	GGTCACGCTCCGTAAGTACTTCATGTCCAGGCTGTTAGAC	introducing mutation in pm DHFR	Mut5_REV
olAI034	CCCCTTTGAACACGCAGAGAAGG	transfection diagnostic primers	Pk FWD - outside HR
olAI035	AAGACCCCTAAAGGTTCGCGGG	transfection diagnostic primers	Pk WT rev
olAI036	CCCAGGCCTCTGAAGGTTCTAG	transfection diagnostic primers	Pk recodonised rev
olAI037	CCTCTGAAGGTCTTGGTGCTGAAG	transfection diagnostic primers	Pm rev - for all pm incl mutants
olAI039	GCGGAAGGTGTAGTTGTTGAACACC	transfection diagnostic primers	Pf rev
olAI040	GGACATGAAGTACTTTAGCAGCGTGACCACCTACGTGAACG	Pm DHFR Mut6 FWD	introducing mutation in pm DHFR
olAI041	GTAGGTGGTCACGCTGCTAAAGTACTTCATGTCCAGGCTGTTAGACTTC	Pm DHFR Mut6 REV	introducing mutation in pm DHFR
olAI042	gtattcagcaccaaagtgtctccac	checking constructus in lines	Pk reverse outside HR
olAI043	GCGCAAAAGTACTTCATGTCCAGGCTATTGCACTTCCAAGGCAGCACGC	Pf converting triple mutant to 3D7 WT	Pf REV I51 N muta
olAI044	AATAGCCTGGACATGAAGTACTTTTGCGCCGTGACCACCTACGTGAACG	Pf converting triple mutant to 3D7 WT	Pf FWD I51N muta
olAI045	AATAGCCTGGACATGAAGTACTTTTGCGCCGTGACCACCTACGTGAACG	Pf converting triple mutant to 3D7 WT	Pf REV N108S muta
olAI046	AATAGCCTGGACATGAAGTACTTTTGCGCCGTGACCACCTACGTGAACG	Pf converting triple mutant to 3D7 WT	Pf FWD N108S muta
olAI047	TGCGCCGTGACCACCTACGTGAAC	Pf converting triple mutant to 3D7 WT	PfDHFR FWD R59C
olAI048	AGCTGGTGCGGCCCATGACCAC	Pf converting triple mutant to 3D7 WT	PfDHFR REV R59C
RM76	CCCGGGGCGTTTTCGCGTATCTGCGCTTTTTC	Independent locus for diagnostic PCR - *Pkmtip*	independent - diagnostic
RM77	CCTAGGGGACAATATATCCTCACAGAACAACTTG	Independent locus for diagnostic PCR - *Pkmtip*	independent - diagnostic

Supplementary table 2. Plasmid stock list

Plasmid name	Drug selectable marker	Donor/Cas9	Description	Final transfection construct?	Associated parasite line
260	AmpR	pCas9	Pl12_1	No	N/A
261	AmpR	pCas9	Pl12_2	No	N/A
262	AmpR	pCas9	Pl112_3	No	N/A
276	AmpR	pCas9	Pl12_3 with guide sequence for PkDHFR	Yes	N/A
263	KanR	pDonor	rPkDHFR with homology regions	Yes	131_2
264	AmpR	pDonor	rPfDHFR with pyrimethamine resistance mutations	No	N/A
266	AmpR	pDonor	rPmDHFR domain only	No	N/A
274	AmpR	pDonor	rPmDHFR domain with Pk homology regions	Yes	135_3
304_1	AmpR	pDonor	rPmDHFR domain reverted to hypothesised sensitive line - mutation 3 and 6 - with HR	Yes	150_1
269	AmpR	pDonor	rPmDHFR domain with mutation 3 - with HR	Yes	135_6
295	AmpR	pDonor	rPmDHFR domain with mutation 5 - with HR	Yes	151
292	AmpR	pDonor	rPmDHFR domain with mutation 6 - with HR	Yes	152
304	AmpR	pDonor	rPfDHFR domain - 3D7 sensitive line - with HR	Yes	150
273	AmpR	pDonor	rPfDHFR domain with triple Pyrimethamine resistance mutations - with HR	Yes	135_1

CHAPTER 5

5 Distinct population structure and genetic diversity of *Plasmodium vivax* malaria populations across Brazil and South America

RESEARCH PAPER COVER SHEET

SECTION A – Student Details

Student ID Number	**1600466**	**Title**	Miss
First Name(s)	Amy		
Surname/Family Name	Ibrahim		
Thesis Title	From genome to function: A genomic investigation into understudied populations of the malaria parasites *Plasmodium malariae* and *P. vivax*		
Primary Supervisor	Prof. Susana Campino		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	N/A		
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Nature Communication

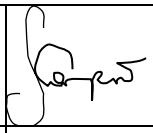
Please list the paper's authors in the intended authorship order:	Amy Ibrahim, Emilia Manko, Jamille G. Dombrowski , Ernest Diez Benavente, Debbie Nolder , Colin J. Sutherland , Francois Nosten , Diana Fernandez, Gabriel Vélez-Tobón, Alberto Tobón Castaño, Anna Caroline C. Aguiar, Dhelio Batista Pereira, Simone dos Santos, Martha Suarez-Mutis, Silvia Maria Di Santi, Andrea Regina de Souza Baptista, Ricardo Luiz Dantas Machado[1], Claudio R. F. Marinho , Taane G. Clark , Susana Campino
Stage of publication	Submitted

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I undertook laboratory work including whole genome amplification and the preparation of samples for sequencing. I also performed bioinformatic analysis and interpreted the results under the supervision of my supervisors. I wrote the first draft of the manuscript that was then circulated to supervisors and after to co-authors.

SECTION E

Student Signature	
Date	25/02/2022

Supervisor Signature	
Date	February 24, 2022

Distinct population structure and genetic diversity of *Plasmodium vivax* malaria populations across Brazil and South America

Short title: Genomic diversity of *Plasmodium vivax* in Brazil

Amy Ibrahim [1,&], Emilia Manko [1,&], Jamille G. Dombrowski [2], Ernest Diez Benavente [1], Debbie Nolder [1,3], Colin J. Sutherland [1,3], Francois Nosten[4, 5], Diana Fernandez[6], Gabriel Vélez-Tobón[6], Alberto Tobón Castaño[6], Anna Caroline C. Aguiar[2], Dhelio Batista Pereira[7], Simone dos Santos[8], Martha Suarez-Mutis[8], Silvia Maria Di Santi[9], Andrea Regina de Souza Baptista[10], Ricardo Luiz Dantas Machado[10], Claudio R. F. Marinho [2], Taane G. Clark[1,11,*#], Susana Campino[1,*,#]

[1] Faculty of Infectious & Tropical Diseases, London School of Hygiene & Tropical Medicine, London, UK

[2] Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil

[3] Public Health England Malaria Reference Laboratory, London School of Hygiene & Tropical Medicine, London, UK

[4] Shoklo Malaria Research Unit, Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Mae Sot, Tak, Thailand

[5] Centre for Tropical Medicine and Global Health, Nuffield Department of Clinical Medicine Research Building, University of Oxford Old Road Campus, Oxford, UK

[6] Grupo Malaria, Facultad de Medicina, Universidad de Antioquia, Antioquia, Colombia

[7] Research Center for Tropical Medicine of Rondonia, Porto Velho, Brazil

[8] Laboratório de Doenças Parasitárias, Institute Oswaldo Cruz - Fiocruz- Rio de Janeiro, Brazil

[9]School of Medicine, University of São Paulo, Brazil

[10]Centro de Investigação de Microrganismos – CIM, Departamento de Microbiologia e Parasitologia Universidade Federal Fluminense, Brazil

[11] Faculty of Epidemiology & Population Health, London School of Hygiene & Tropical Medicine, London, UK

[#] Corresponding authors

E-mail: taane.clark@lshtm.ac.uk, Susana.campino@lshtm.ac.uk

[&] These authors are joint authors; * These authors are joint authors

ABSTRACT

Brazil is a unique and understudied setting for malaria, with complex foci of transmission associated with human and environmental conditions. An understanding of the population genomic diversity of *P. vivax* parasites across Brazil can support malaria control strategies. Through whole genome sequencing of *P. vivax* isolates across 7 states, we compare genetic diversity within country (n = 123), continent (6 countries, n = 315) and globally (26 countries, n = 885). We confirm that South American isolates are distinct, have more ancestral populations than the other global regions, with differentiating mutations in genes under selective pressure linked to antimalarial drugs (*pvmdr1*, *pvdhfr-ts*) and mosquito vectors (*pvcrmp3, pvP45/48, pvP47*). We demonstrate Brazil as a distinct parasite population, with signals of selection including ABC transporter (*PvABCI3*) and PHIST exported proteins. Brazil has a complex population structure, with evidence of *P. simium* infections and Amazonian parasites separating into multiple clusters. Overall, our work provides the first Brazil-wide analysis of *P. vivax* population structure and identifies important mutations which can inform future research and control measures.

Word count: 172

INTRODUCTION

The *Plasmodium vivax* parasite causes the highest malaria burden outside of sub-Saharan Africa [1], with more than one third of the global population at risk due to its wide geographical range [2]. Complications associated with *P. vivax* infections can lead to severe, life-threatening syndromes [3]. *P. vivax* infections underly the majority (>80%) of the >700k annual malaria cases in the Americas, including in South America, where countries surrounding the Amazon rain forest areas, such as Brazil, Colombia, and Venezuela, have hotspots of endemic disease [1]. In South America, malaria transmission dynamics are complicated by *P. vivax* and *P. falciparum* co-infections [1]. Further, significant challenges continue to thwart *P. vivax* control, including the ability of parasites to form dormant hypnozoite stages within the liver, leading to relapses of malaria if not removed using a radical cure of primaquine [4]. Unfortunately, individuals with glucose-6-phosphate dehydrogenase deficiency are unable to take primaquine or tafenoquine due to the risk of severe haemolysis [5]. Control measures are compromised by the presence of sub-microscopic and asymptomatic *P. vivax* infections, leading to untreated human parasite reservoirs [6]. Human settlement and mobility, including through peri-urban expansion, gold mining-related activities, and deforestation in the Amazon, all lead to significantly higher risk of malaria infection [5,7]. There are also gaps in knowledge, including for understudied malaria in pregnancy, and unknown *P. vivax* mutations underlying resistance to the first line antimalarial, chloroquine, to which resistance has emerged in many countries in South America [8], calling for a rapid change in *P. vivax* control strategy.

Brazil has a diverse geographical profile leading to variation in malaria transmission, with foci split into three discrete groups, each with unique settings for transmission. Firstly, the Amazon rainforest in the northeast of Brazil, which accounts for 99% of all malaria cases, and transmission is led by *Anopheles darlingi* and *An. albitarsis complex* mosquitoes. Secondly, the coastal border of Brazil, where transmission is lower and due to the *An. aquasalis* mosquito genus. Finally, the Atlantic Rainforest in the southeast of the country, where transmission is mainly mediated by *An. bellator* and *An. cruzzi* [9]. In this southern region, *P. simium* is transmitted by *An. cruzii,* is highly genetically related to *P. vivax*, and found mainly in non-human primates [10] but human cases have been recorded in São Paulo and Rio de Janeiro[11]. Whilst there is inter-state transmission, importation of malaria from neighbouring countries, such as French Guiana, Guyana, Venezuela, and Peru, plays an important role in malaria transmission in the Brazilian Amazon. Importation events threaten elimination in some geographical areas free of malaria or recording low numbers of autochthonous cases, affecting the extra-Amazonian Region [9]. In addition, human genetics may also contribute to *P. vivax* transmission dynamics in Brazil, with the prevalence of the Duffy negative (Fy-) blood group phenotype, which prevents *P. vivax* invasion of red blood cells, but there is emerging evidence of Duffy-negative individuals positive for *P. vivax* infection [12–14] Furthermore, vivax-malaria in Fy-

individuals may present as an asymptomatic infection that is untreated and acts as a parasite reservoir, complicating malaria eradication [15].

In contrast to the wider transmission of *P. vivax* in Brazil, *P. falciparum* is restricted to hotspot areas, which are mostly constrained to the Amazonian rainforest states of Amazonas and Acre, the two states which account for >45% of all malaria cases [16]. Malaria control measures designed for *P. falciparum* are widely known to be less effective at tackling *P. vivax* infections, due to key differences in parasite biology including the complication of dormant liver stage parasites, viability at wider temperature ranges [17], and parasites that are more permissive to multiple vector species [18]. However, many regions adopt the same vector control and treatment regimens as for *P. falciparum*. In Brazil, chloroquine is still used for treating *P. vivax* infections, even though resistance has already been documented in both *P. falciparum* and *P. vivax* parasites [8,19]. Surveillance of drug resistance is a challenge because reliable molecular assays for *P. vivax* are not yet available, as underlying biological susceptibility mechanisms are poorly understood. Whole genome sequencing (WGS) of *P. vivax* can provide insights into genetic mutations underlying drug resistance and population structure. Studies have shown that parasites within South America display high levels of genetic diversity, comparable to high transmission regions such as Southeast Asia, and arising from patterns of human migration from different populations [20]. South American parasites form a distinct global subpopulation [21–23], with informative barcoding loci linked to mosquito transmission and development (e.g., *pvcrmp* gene family, *pvs47*, *pvs48/45*), possible reduced drug susceptibility (e.g., *pvmdr1*), and red blood cell invasion (e.g., *pvmsp10*, *pvtrap*) [22,23]. *P. vivax* parasites within South America are known to demonstrate general country level separation [20,21], with Brazilian and Peruvian isolates clustering together [20,23].

Brazil is a complex setting to characterise *P. vivax* genetic diversity, due to both the three distinct transmission foci, and the multiple migratory waves of humans carrying malaria. The parasite population structure remains unclear, with the vast majority of currently available WGS data collected from malaria infections in Acre [21,24,25]. Here, we perform a population genomic analysis of the largest WGS dataset for *P. vivax* isolates from 10 regions within Brazil (n = 123) spanning 7 states, position them in a global context (n = 885), characterise the within country and wider regional ancestral and population structure, and identify loci under selective pressure. We find a complex population of parasites within Brazil, with vast genomic diversity in areas of high transmission, and Brazilian specific signals of selection in genes associated with drug susceptibility.

RESULTS

P. vivax isolates and sequencing data

WGS data of Brazilian samples (n = 123) includes isolates from human infections spanning 10 regions (Goianésia do Pará, Novo Repartimento, Itaituba (Pará State), Macapá, Oiapoque (Amapá State), Rio Branco (Acre State), Porto Velho (Rondônia State), Barcelos (Amazonas State), Mato Grosso and São Paulo), building on public WGS data originated from infections in Acre and Rondônia [20,22,25]. WGS data was analysed with 1,113 isolates of *P. vivax* spanning 26 countries, and a total of 3,932,759 SNPs were identified [22–24,26]. After filtering, a final combined "high quality" dataset consisted of 885 isolates with a total of 454,681 unique SNP positions in the core genome of *P. vivax*, excluding the hypervariable regions. The filtered isolates were assigned into regional groups: South America (n = 315, including Brazil (n=123), Colombia, Guyana, Mexico, Panama, Peru), South Asia (n = 114; Afghanistan, Bangladesh, India, Sri Lanka, Pakistan), East Africa (n = 84; Eritrea, Ethiopia, Madagascar, Sudan, Uganda), South East Asia (SEA; n = 286; Cambodia, China, Laos, Myanmar, Thailand, Vietnam), and Southern SEA (SSEA; n = 86; Malaysia, Papua New Guinea, Indonesia, The Philippines) (**S1 Table, S2 Table, S1 Figure**).

Four ancestral populations in South America with a distinct Brazilian parasite population

Both SNP-based maximum likelihood tree and principal component analysis (PCA) on the final dataset (n = 885) revealed the expected regional separation [23] of *P. vivax* parasites with distinct clusters forming for South America, as well as for East Africa, South Asia, SEA, and SSEA (**Figure 1**). An ADMIXTURE analysis suggested that there are ten ancestral populations spread across the five global regions, including four within South America (K2, K3, K9 and K10) and six elsewhere (East Africa K7; South Asia K1; SEA K8, K6 and K4; SSEA K6 and K5) (**Figure 1, S2 Figure**). Within South America (n = 315), a maximum-likelihood tree and PCA analysis based on 102,765 SNPs revealed country-level separation of isolates, including for Brazil (except São Paulo samples), with some minor overlap between Panama and Colombia, and both Panama and Guyana with Brazil (**Figure 2**). There is a high concordance between ADMIXTURE population and country of origin (K3 Brazil, Guyana; K2 Mexico, Colombia; K9 Peru; K10 Panama; **Figure 2)**, with highly clonal clusters for Mexico and Panama consistent with previous studies [21,27]. The samples from São Paulo (n=12) cluster together in a clade separated from the remaining Brazilian samples and close to the Mexican clade (**Figure 2, Figure 3**). These samples could be *P. simium*, being collected in the geographic region where this parasite has been reported, with the majority of them containing two putative *P. simium* barcoding mitochondrial SNPs (T4133C, A4467G) [11]. Deletions in *pvdbp1* and *pvrbp2a* loci reported in *P. simium* but not in *P. vivax* [28] could not be characterised due to poor sequencing coverage at these regions.

Loci informative for South American and Brazilian population differentiation

The fixation index (F_{ST}) was used to identify SNPs driving genetic differentiation between populations, therefore for all F_{ST} analysis, we are investigating SNPs with the highest F_{ST} values (>0.8 across all analyses). SNPs with high F_{ST} values are likely to be demonstrating regions of directional selection in one population, where differential selective pressures are acting on the two comparative populations, causing a difference in the alleles present. Balancing selection, whereby it is preferential to maintain multiple variants at a specific locus, could also be investigated using F_{ST} scores, however for this analysis, lower FST scores would need to be investigated as under balancing selection, alleles are less likely to become fixed at a locus. Additionally, SNPs with lower FST scores may demonstrate regions where the same selective pressure is acting in two different populations.

The isolates from São Paulo that may be *P. simium* infections were removed from further *P. vivax* population genetic analyses due to inconclusive speciation. When comparing South American isolates (n = 303) to other regions, the greatest number of highly differentiating (F_{ST} >= 0.99) SNPs are seen with SSEA (>150 SNPs with F_{ST} >= 0.99) (South Asia 44, Southeast Asia 37, East Africa 22) (**S3 Table**). Across all pairwise regional comparisons, highly differentiating (F_{ST} >= 0.99) SNPs in South America were found in genes potentially involved in gene regulation (*pvjmjc1* [29]), mosquito life stages (*pvcrmp3*, *pvp28*, *pvp47*, *pvp48/45* [30–33]), drug resistance (*pvmdr1* [34]), gliding motility and cell traversal (*pvtrap, pvtlp* [35,36]), and those encoding parasite surface proteins (*pvmsp10* [37]) (**S3 Table**). These genes overlapped with South American-specific differentiating SNPs (F_{ST} > 0.9, vs. non-South American, n = 570; **S4 Table; Table 1**). A nonsynonymous SNP leading to amino acid substitution 698S>698G in *pvmdr1*, fixed in both the Brazilian and South American population, differentiated South American parasites from those in SSEA and SEA (F_{ST} = 1 and 0.99, respectively) in accordance with previous findings[23] (**Table 2, S3 Table**).

Within South America, Brazilian *P. vivax* (n = 111, vs. other South America, n = 192) informative SNPs (F_{ST}>0.8) were found within genes associated with drug susceptibility (*pvmdr1* [34]), gene expression (*pvapiap2* [38]), mosquito life stages (*pvcrmp3* [30]) and a gene encoding reticulocyte binding protein, *pvrbp2a* [39] (position 719K>719E, F_{ST} = 0.815) (**Table 1, S5 Table**). By focusing on country-level pairwise comparisons with Brazil, the highest number of differentiating SNPs (F_{ST} >0.95) were observed against Mexico (62 SNPs), followed by Panama (29 SNPs), Colombia (14 SNPs) and Peru (4 SNPs) (**S6 Table**), consistent with differences in geographical distance and genetic clustering in the maximum likelihood tree and PCA analysis (**Figure 2**). Of note, are putative drug resistance mutations, including within *pvpppk-dhps* (M205I: Brazil 97% vs. Mexico 0%; F_{ST} = 0.95; S. America 65%) previously observed in China [40] and

Thailand [41], and *pvmdr1* (V221L: Panama 92% vs. Brazil 0%;, F_{ST} = 0.97) previously observed in Peru [42] (**S6 Table**).

Distinct populations within Brazil associated with parasite surface proteins and drug resistance loci

All Brazilian regions have similarly low levels of nucleotide diversity (average π *across all states, excluding* São Paulo*, 3.06 x* 10^{-4}), with the lowest diversity seen within São Paulo (π = 0.54 x 10^{-4}) (**S3 Figure**). A SNP-based maximum likelihood tree (n = 70,757 SNPs) of only Brazilian isolates (n = 123) revealed seven distinct clades (C1-C7) (**Figure 3**), including a likely *P. simium* clade with the samples from São Paulo (C5), corresponding to the ADMIXTURE population K9 (**Figure 3, S4 Figure**). Clades C1, C2, C3, C4 and C6 mostly cover isolates from Acre and the Amazonas, with a small number of isolates from Rondônia in clades C2 (n = 1) and C6 (n = 2), and a small number of isolates from Mato Grosso in clades C4 (n = 1) and C6 (n = 2), demonstrating the vast genetic variability of isolates in the Amazon basin. Clade C7 covers isolates from Amapá and Pará located in northern Brazil, with a small number of isolates from Rondônia (n = 2), Acre (n = 2), and Mato Grosso (n = 2) (**Figure 3, S4 Figure**). Two isolates from Acre did not fall into a clade grouping (**Figure 3**). Whilst the population structure within Brazil appears to be complex, it is important to note that excluding those from São Paulo, all other Brazilian isolates clustered together as population K3 in the global ADMIXTURE analysis, which was a distinct population of Brazilian isolates (**Figure 3**).

Informed by the population structure observed, subsequent analysis within Brazil compared different clades as well as two regional groups (A: Amazonas, Acre, Mato Grosso and Rondônia states (n = 88); B: Amapá and Pará states (n = 23)) (**S5 Figure, S2 Table**). F_{ST} scores are heavily impacted by population size, therefore only clades with >15 isolates were compared to each other (excluding clades C2 to C4 from comparisons) (S7 Table). Highly differentiating non-synonymous SNPs (F_{ST} > 0.85) separating clades C1, C6 and C7 were identified (**S7 Table**), including genes associated with; reticulocyte binding (merozoite surface protein, *pvmsp1*) [43], liver stages of infection (*pvlisp2* [44]), and within a Plasmodium-specific ABC transporter (*pvabci3)* whose ortholog has been linked to a drug resistance mechanism in *P. falciparum* [45]. Clades C6 and C7, which are associated with isolates from Acre and Amapá-Pará states respectively only have 11 highly differentiating mutations (F_{ST} >0.85), all synonymous SNPs. When comparing regional groups A and B, the most highly differentiating SNPs were observed within the Plasmodium interspersed repeat gene family (*pvpir*), specifically those found on chromosome 6. *Pvpir* genes are the largest gene family within Plasmodium spp. (found within *P. vivax* as well as simian and rodent malaria parasites), thought to play a role in host red blood cell invasion and immune evasion [46] (**S8 Table**).

Multi-clonality and signals of relatedness and homology within parasite populations

Multi-clonality, as measured by within-sample diversity (F_{WS} metric < 95%), was present in 206 (23.2%) of all isolates, being more common among SEA (35.4%) and SSEA (33.7%), suggesting a higher chance of co-transmission of multiple *P. vivax* strains in these regions (**S1 Table, S6 Figure**). In the Brazilian isolates, multiclonality was more common within clades C4, C6 and C7, in addition to higher likelihood of multiclonality in regional group B than group A (**S7 Figure**). Analysis of identity-by-descent (IBD), to quantify isolate relatedness, was performed at country level on the global dataset of monoclonal isolates. For all IBD analysis, isolates were grouped into countries, and IBD was calculated within country, with no between country comparisons for this metric (n isolates in IBD analysis = 679) (**S1 Table**). Within country IBD analysis revealed that Malaysia (median IBD 0.335), Panama (0.971) and Mexico (0.232) demonstrated the greatest fractions of IBD, with other populations with IBD <0.056 (**S8 Figure, S9 Table**). Across genome-wide sliding windows of 50kbp, there are global patterns of signals of high IBD including a segment on chromosome 5 encompassing *pvdhfr-ts*, a gene associated with pyrimethamine resistance (average IBD fraction across the 679 isolates = 0.168) [47] (**S9 Figure**), which demonstrates a specifically high signal of IBD in South America (0.259) and Brazil (0.122) (**S10 Table**). Additionally, there is a segment of chromosome 10 encompassing *pvmdr1*, a gene associated with multi-drug resistance (average IBD = 0.166) [47], which demonstrates a high signal in East Africa (0.124) and South America (0.276), and specifically within Brazil (0.136). Within Brazil and Mexico there are multiple segments within chromosome 10 with signals of high IBD. Brazil shares a high signal of IBD with Mexico, Panama, Peru, SEA, East Africa, India and Pakistan in chromosome 14 where both *pvdbp1*, a gene associated with erythrocyte invasion [48], and *pvdhps-pppk,* a gene associated with sulfadoxine resistance [47], are found (Brazilian IBD = 0.133) (**Figure 4, S10 Table**).

We investigated patterns of within group IBD for clades C1, C6 and C7 (all with >15 isolates). Clade C1 specific signals of IBD were found within chromosomes 9 (encompassing *pvama1*, a potential vaccine candidate [49]) (IBD = 0.692), and 3 sequential segments within chromosome 14 (positions 2.35Mbp to 2.50Mbp), which included loci encoding the clustered-asparagine-repeat-protein (*pvCARP*), which is associated with the host immune response to malaria infection [50] For clade C6, signals were identified on chromosome 3, 5, 11, 12, 13 and 14 (IBD > 0.3), encompassing loci encoding the GPI-anchored micronemal antigen (*pvGAMA*) on chromosome 5 which is an essential invasion protein in *P. falciparum* infections, suggested as a potential vaccine candidate [51], and loci encoding *pvAPIAP2* associated with gene regulation [38]. Clade 7 specific IBD signals were found in chromosome 14, where both *pvdbp1*, a gene associated with *P. vivax* erythrocyte invasion [48], and *pvpppk-dhps,* a gene associated with sulfadoxine susceptibility [47], are

located (IBD = 0.134) (**S11 Table**). High signals of IBD were observed across all three clades (C1, C6 and C7) within chromosome 14 (average IBD = 0.329), where genes potentially involved in DNA replication are found including *pvsec13* [52]. Signals of within group IBD across the two geographical groupings (A, B) were polarizing, with signals in chromosomes 2 and 5 for Group A including a region encompassing the eukaryotic initiation factor-2α, potentially associated with artemisinin resistance in Plasmodium parasites [53], and within segments of chromosome 14, including *pvpppk-dhps* and *pvrbp1a*, a gene associated with erythrocyte invasion [54] in isolates from Group B (**S12 Table**).

Regions under selection in South American and Brazilian subpopulations

Genome-wide analysis to identify positive selective sweeps using the "single population" integrated haplotype score (iHS) across monoclonal isolates (n = 679), within population groupings, revealed signals involving merozoite surface protein genes (*pvmsp1, pvmsp4, pvmsp5)* in all global regions except for SSEA (**S13 Table**). Within East Africa, South Asia and SEA, signals of positive selection were identified within chromosome 2 in a region that encompasses *pvmrp1*, a gene associated with drug susceptibility [23]. In both South Asia and SEA, signals of positive selection were found within chromosome 5, where the *pvdhfr-ts* gene is located. Loci associated with erythrocyte binding were also identified, including *pvdbp1* in SEA (**S13 Table, S9 Figure**). Across East Africa, South Asia, SEA, and SSEA, analyses detected signals of positive selection within chromosome 3 where *pvlisp2,* linked to parasite development in the liver [44], is found. Signals of positive selection within South America include regions of the genome where multiple Plasmodium Poly-Helical Interspersed Sub-Telomeric (PHIST) proteins are encoded on chromosome 5. These proteins peripherally-localised in infected erythrocytes and in *P. falciparum* are involved in functions such as protein trafficking, membrane rigidity and intercellular signalling [55]. Other loci identified included the leucine-rich repeat (*pvlrr8*) gene , the surface protein *pvmsp1* along with a paralog *pvmsp1p-19* [56] (**S13 Table**). Within South America, we looked for signals of positive selection at the country level (for countries with >10 isolates). Signals were detected in *pvmsp1* within Colombia, Panama, and Peru, in *pvdbp1* within Peru, and *pvlisp2* within both Panama and Peru (**Figure 5, S14 Table**). There were only 5 SNPs detected within Brazil which demonstrated signals of positive selection, with just 2 SNPs in coding regions (Plasmodium exported protein PVP01_0525100, *pvphist*) (**Figure 5, S15 Table**). Within Brazil, signals of positive selection using iHS were detected in chromosomes 8 and 14 in Amazonian clade C6, including the ABC-transporter *pvabci3* (PVP01_082050), whose orthologue is associated with drug resistance in *P. falciparum* [57]. In Clade C7 isolates (associated with Amapá and *Pará* states) candidate regions for positive selection were seen within loci encompassing surface proteins (e.g., *pvmsp1*, *pvmsp4, pvmsp5)*, *pvlisp2, pvlrr8* and *pvdbp* involved in erythrocyte invasion (**S16 Table**).

The between population Rsb method was applied to detect positive selection at both the regional and country level (**S17 Table; S18 Table;** $P<1x10^{-5}$). When comparing South America against the other global regions, multiple SNPs within *pvmsp1*, associated with reticulocyte binding [43], demonstrated signs of positive selection (East Africa 26, South Asia 62 SEA 58, SSEA 20). Similarly, SNPs in the gene encoding PvPHIST exported protein were also detected in all pairwise comparisons, except with SSEA. The surface protein encoding gene *pvmsp5* (chromosome 4) was detected between South America and SEA. Comparisons of Brazil to other South American countries revealed multiple SNPs within *pvlisp2* associated with liver infection [44] (Brazil vs. Panama, Peru), *pvmsp5* (Brazil vs. Mexico, Peru) and *pvmsp1, pvmsp4* and *pvmsp5* (Brazil vs. Peru) (**Figure 6, S18 Table**, **S19 Table**). Within Brazil, two candidate genomic regions were detected between clades C1 and C6, where surface proteins were found (e.g., *pvmsp4, pvmsp5*), in addition to two regions between clades C1 and C7 (including *pvsmp1, pvmsp1p,* and *pvlrr8*). Five candidate genetic regions were identified when comparing clades C6 and C7, which included the *pvlisp2* gene and multiple merozoite surface proteins. Between regional groups A and B, 5 candidate regions were identified, which included the *pvlisp2, pvdbp* and *pvmsp1* genes (**S20 Table**).

In addition to positive selection signals, we investigated genes (with > 5 SNPs) under balancing selection using Tajima's D statistic in all monoclonal isolates (n = 679). As expected, most Tajima's D values for genes across global regions were negative (median: South America -0.437, SEA -1.82, South Asia -0.756, East Africa -0.385, SSEA -0.904), with the most negative value globally occurring in SEA, suggesting population expansion in this region (**S10 Figure**). Within South America, median values for Tajima's D were negative in Brazil (-0.034), Colombia (-0.046) and Panama (-0.330), while positive in Mexico (0.173) and Peru (0.078) indicating a population decrease or a genetic bottleneck (**S11 Figure**). The top 50 genes with the highest and lowest Tajima's D metric in South America are summarised (**S21 Table),** with the most positive including *pvlisp2*, *pvrbp1a*, *pvmsp1*, *pvphist, pvpir,* and *pvcyrpa* also associated with reticulocyte binding [58], suggesting balancing selection. The findings from the same analysis for Brazil overlapped, and includes genes *pvmsp1, pvlisp2, pvrbp1a, pvcyrpa* in addition to loci encoding PvPHIST and PvPIR proteins (**S22 Table**).

Identification of mutations and allele frequencies in *P. vivax* drug resistance candidate genes

Treatment failures have been reported with *P. vivax* infections, however the molecular determinants for reduced drug efficacy are not clearly defined. We investigated the presence of SNPs within orthologs of genes associated with resistance in *P. falciparum,* alongside loci identified by selection metrics from previous population genomics studies [23,24,59] (**Table 2**, **S23 Table**). There are similar patterns of frequencies of SNPs within potential resistance-associated genes between Brazil and the other South American isolates

likely due to similar drug regimens across this region. Of note are SNPs which appear close to fixation within the Brazilian population, found within *pvubp1* (potentially associated with artemisinin resistance in *P. falciparum* [60]), multidrug resistance associated proteins MDR1, MDR2, MRP2, phosphatidylinositol 4-kinase *pvpi4k* (the target of novel antimalarial class imidazopyrazines [61]), *DHODH* (a drug target for DSM265, a novel antimalarial in clinical trials, shown to be less effective against *P. vivax* infections than *P. falciparum* [62]), *ferredoxin – pvfdx* (potentially associated with artemisinin resistance in *P. falciparum* [63]), *pvpppk-dhps* (associated with sulfadoxine resistance [47]), and genes coding for putative ABC transporters (*pvabci3, pvabcg2*), whose orthologues may be associated with antimalarial resistance in *P. falciparum* infections [34,57] (**Table 2**). Some of these mutations are observed in high frequency in South America but have quite different frequencies compared with other global regions, including a missense mutation within *pvmdr1* (698S>689G), which is fixed in all South American isolates, and found in around half of the populations in East Africa and South Asia, but rare and non-existent in SEA and SSEA respectively. Another *pvmdr1* mutation (500D>500N) is also present in Brazil with high frequency (80%) but with lower frequency in wider South America (31%) and not identified in any other continents. Similarly, a missense mutation within *pvabcg2* (457K>457M) is fixed within Brazil and present in South America (69%), but not observed elsewhere (Table 2). *Pvabcg2* encodes an ATP binding cassette (ABC) transporter, which are commonly known to be associated with multiple drug resistance phenotypes in many organisms [34] and is linked to the gametocyte stages of parasite development [64].

DISCUSSION

Whilst *P. vivax* infections pose a serious risk to global health, but genomic analyses of this species, particularly in South America where the parasite is predominant, are scarce in comparison to the more pathogenic *P. falciparum*. Brazil is a unique setting for malaria transmission, with distinct foci relating to the local environments and resultant vector landscapes. To date, all previously published WGS data from Brazil has originated from isolates obtained mainly from Acre and a few from Rondônia, in the north-western region of the country. Here, we provide the first insight into the genomic diversity of *P. vivax* isolates from all three malaria endemic regions in Brazil, spanning seven states, to determine the broader population structure within the country, as well as its position within a continental and global resolution. Using 855 global isolates of *P. vivax* across 26 countries, we placed South America in the global context, demonstrating that they form a distinct population with more ancestral populations than other global regions. The four distinct ancestral South American populations mostly correspond to country groups, in accordance with previous studies demonstrating nation-level separation within this continent [20,23]. Using 123 isolates from Brazil, we demonstrated that the population structure is complex, with samples clustering

across seven distinct clades, clearly separating the Northern states (Amapá and *Pará*) and the highly clonal potential *P. simium* cluster from São Paulo. Isolates from the Amazonian basin fall within five (of the seven) clades, consistent with the high malaria transmission in the large region leading to greater population diversity.

Population genomic analysis was applied to further understand the structure and selection within the parasite populations. In this study, we investigated SNPs that were causing genomic differentiation between populations; therefore, we only investigate SNPs with high F_{ST} scores. High FST scores are likely to demonstrate signals of directional selection in one population that are not occurring in the comparative population, and lower FST scores (which were not investigated in this study) demonstrate regions that may be under balancing selection, or regions whereby the same selective pressure is acting in multiple regions (for example, when the same drug is being used in two different populations). Our results highlight many South American-specific SNPs within genes involved in different parasite life stages and associated with drug resistance. Genes involved in mosquito life stages, such as gametocyte proteins *PVS48/45* and *PVS47*, may be reflective of the different mosquito vectors present in South America compared to other regions. Other studies have also identified mosquito-related proteins under selection in other *P. vivax* endemic regions [23,24,59]. Additionally, South American-specific SNPs were also found within genes encoding parasite surface proteins (e.g., *pvmsp1/4/5*), and drug resistance (e.g., *pvmdr1*). A confounding factor for all F_{ST} analysis is population size, and F_{ST} scores for smaller populations must be interpreted with caution. A potential way to reduce the effect of imbalanced population sizes is to take a subsampling approach whereby multiple smaller subsets of the larger population are assessed against the smaller population individually, and an average F_{ST} score is given.

Several signals of selection and homology were identified in loci associated with drug resistance, specifically within *pvdhfr-ts* and *pvmdr1* across all South American isolates, which may reflect similar selection pressures due to drug regimens within continent. In addition, a Brazilian-specific signal was observed within *pvdhps-pppk*, the determinant of sulfadoxine resistance in *P. falciparum*. Other possible candidates for further investigation linked to antimalarial drugs included *pvmrp1, pvmrp2*, and an ABC transporter I family member (*pvabci3*), revealed as signals of positive selection and/or SNPs fixed in Brazilian samples. The orthologous *pfmrp1* gene in *P. falciparum* is a multidrug-resistance candidate, and has been shown to be under strong selection in across populations, with mutations associated with reduced susceptibility to sulfadoxine-pyrimethamine, chloroquine and mefloquine, and pyronaridine[65–68]. The *ABCI3* protein is a *Plasmodium*-specific ABC family member, and SNP and gene amplification variants in

P. falciparum have recently been shown to confer anti-plasmodial drug resistance across a variety of compounds [57,69].

Determining the downstream effect of SNPs for *P. vivax* research is complicated due to the lack of a routine *in vitro* culture method for this parasite species. It is possible to perform orthologue replacement transgenesis in *P. knowlesi* as this parasite can be cultured in human erythrocytes and is the most closely related species to *P. vivax*. This system allows the functional investigation of the role of genetic variants, such as in drug susceptibility or red blood cell invasion. Brazil-specific SNPs in genes involved in red blood cell invasion (*pvrbp2a*, *pvrbp1*, *pvcyrpa*), and signals of positive selection in PHIST family members were also detected, which may reflect regional-specific host factors on red blood cells. The liver stage *pvlisp2* gene, which differentiates between dormant hypnozoites and early developing parasites [44] was identified when investigating signals of selection across South American populations and within Brazilian clades. Genetic markers in *pvlisp2* can assist the development of drug discovery assays predictive of anti-relapse activity [44].

Overall, our work provides insights into the genomic diversity across all three malaria endemic regions in Brazil, as well as in the broader context of South America and other continents. The results highlight many novel and previously detected genes and mutations, which may reflect ongoing evolutionary interactions with the vector and human hosts in the different regional settings and in response to antimalarial drugs. Our insights will inform functional studies, which can determine the role of the candidate loci during the parasite life cycle and in response to treatment and anti-relapse therapies. Ultimately, this work will assist with the design of much needed tools for infection control, such as diagnostics, ultimately working towards malaria elimination.

METHODS

Whole genome sequence data

A total of 1,113 isolates were analysed, including publicly available (n = 1023) and novel sequence data from Brazil (n = 89). After quality control, the dataset consisted of 885 isolates spanning all regions where *P. vivax* infections are endemic: (i) South America (n = 315: Brazil 123, Colombia 34, Guyana 3, Mexico 20, Panama 46, Peru 89); (ii) East Africa (n = 84; Eritrea 13, Ethiopia 53, Madagascar 4, Sudan 9, Uganda 5); (iii) South Asia (n = 114; Afghanistan 27, Bangladesh 1, India 48, Pakistan 37, Sri Lanka 1); (iv) South East Asia (n = 286; Cambodia 71, China 12, Laos 2, Myanmar 28, Thailand 160, Vietnam 13); and (v) the Western Pacific and southern South East Asia (n = 86; Indonesia (9), Malaysia (50), Papua New Guinea (26), The Philippines (1)) (**S1 Table**). These included newly sequenced isolates (n = 51) and

publicly available data (n= 834) in the final filtered database. Newly sequenced isolates were obtained from whole blood samples from seven states in Brazil (Acre 4; Amapá, 10; Rondônia, 4; Amazonas 3; São Paulo 12; Mato Grosso 5; Pará 13), leading to a total of 123 high quality WGS data from isolates within Brazil spanning all areas of *P. vivax* transmission (**S1 Figure, S2 Table**).

The whole blood samples were obtained from symptomatic malaria patients. All samples were collected with the appropriate ethical approval from relevant authorities, including from Hospital Universitário Antonio Pedro, Universidade Federal Fluminense (ref. CAAE 06214118.2.0000.5243) and Faculdade de Medicina de São José do Rio Preto (CAAE 01774812.2.0000.5415). Informed consent was obtained from all individuals. DNA was extracted from whole blood samples using the QIAamp DNA Blood Mini Kit (Qiagen), quantified using a Qubit (v2.0) fluorometer, and single-species *P. vivax* infections were confirmed using qPCR [70]. Selective whole genome amplification (SWGA) using a set of previously described primers [71] was used to increase the relative levels of *P. vivax* DNA within the sample, allowing for whole genome sequencing (WGS) [72]. Amplified isolates were sequenced using the Illumina MiSeq and HiSeq4000 platforms using paired-150 bp read kits through The Applied Genomics Centre, LSHTM.

Bioinformatic analysis

FASTQ files generated from the Illumina sequencing reads (from both publicly available (n = 1023) and newly sequenced isolates (n = 89)) were trimmed using *TRIMMOMATIC* (version 0.39) with the following parameters: LEADING:3, TRAILING:3, SLIDINGWINDOW:4-20, MINLEN:36 [73]. Trimmed reads were aligned to the PVP01 *P. vivax* reference genome (www.plasmodb.org) using *BWA-MEM* (v0.7.12) [74]. BAM files were processed using *samtools* (v1.10) functions fixmate and markdup. SNPs and indels were determined using *GATK's* HaplotypeCaller (v4.1.4.1) following the recommended settings and default parameters using the option -ERC GVCF [75]. The *GATK* ValidateVariants function was used to validate the genomic VCFS (GVCFs), which were subsequently imported into a GenomicDB using the *GATK* function GenomicsDBImport. *GATK's* GenotypeGVCFs function was used to create a combined VCF including all isolates. A total of 3,932,759 unfiltered SNPs were identified across the 1,113 isolates. Variants within subtelomeric regions and Variant Quality Score Log-Odds (VQSLOD) scores < 0 were removed. A total of 228 isolates with more than 40% of SNPs missing genotype data were excluded from downstream analysis. The final dataset consisted of 885 isolates and 454,681 high quality SNPs used for population genetic analysis. SNPs were annotated with their downstream effect using SnpEff [76].

Population genetic analysis

Multiplicity of infection (MOI) was calculated at a country level using the F_{WS} score implemented in the *moimix* package (https://github.com/bahlolab/moimix), as well as at an individual isolate level using estMOI software[77]. Population structure of isolates was investigated using a principal component analysis (PCA). based on pairwise SNP Manhattan distances between isolates. Maximum-likelihood (phylogenetic) trees were created using *IQTREE* (v2.1.2) [78] on a nucleotide alignment consisting of the high quality isolates SNP positions. Ancestral analysis was performed using the *ADMIXTURE* (v1.3.0) package on matrices of high-quality SNPs with a linkage disequilibrium correlation coefficient ≤ 0.1. *ADMIXTURE* predicts the most likely number of ancestral populations (K) within a dataset using cross-validation error [79]. We calculated the fixation indices (F_{ST}) for SNPs between population groups (at global regional, country and two grouping levels within the Brazilian population; clade and geographic groupings) to investigate alleles driving the differences between populations using the *VCFtools* (v0.1.16) function *--weir-fst-pop* [80]. Nucleotide diversity (Nei and Li π) was calculated genome-wide using VCFtools within each Brazilian state (*Pará*, n = 13; Amapá, n = 10; Mato Grosso, n = 5; Rondônia, n = 5; Acre, n = 74; Amazonas, n = 4; São Paulo, n = 12) using sliding windows of 25 kbp.

Positive and balancing selection and IBD analysis

We screened monoclonal (F_{WS} > 95%) isolates for signals of positive selection at both the regional and country level, with a focus on South American, and specifically Brazilian samples, using the *REHH* package (v3.2.1) in R [81]. The integrated haplotype homozygosity score (iHS) [82] was calculated to identify signals of within population selection, and the Rsb [83] score was calculated to demonstrate signals of selection between two assigned populations. Both measures were calculated at the regional and country level, as well as within Brazil at two different grouping classifications (clade groupings from the phylogenetic tree, and geographical groupings into group A and group B (**S2 Table**)). Candidate regions were identified from iHS and Rsb results using default parameters and a *p*-values of < $1x10^{-4}$ and < $1x10^{-5}$, respectively. Only populations with > 10 isolates and genes with > 5 SNPs were included in analysis. Where there were >10 isolates per country, monoclonal Isolates (F_{WS} > 95%) were screened at the country level for identity-by-descent (IBD), within population groups, using the hmmIBD package with default parameters, including a recombination rate based on previous work in *P. falciparum* [84]. Pairwise comparisons for isolates presenting evidence of IBD was plotted using a sliding window of 50 kbp along the genome location. Signals of selection at the regional level (for populations with >10 isolates), and within Brazil at the gene level (for genes with >5 SNPs), were investigated using the Tajima's D metric, which was calculated using the *PEGAS* package (version 0.14) [85].

Data availability

Raw sequence data is available from the European Nucleotide Archive (see **S2 Table** for accession numbers of novel Brazilian isolates). This data also includes samples from the MalariaGEN P. vivax Genome Variation project, as described elsewhere [24].

ACKNOWLEDGEMENTS

AI is funded by an MRC LiD PhD studentship. TGC is funded by the Medical Research Council UK (Grant no. MR/M01360X/1, MR/N010469/1, MR/R025576/1, and MR/R020973/1) and BBSRC (Grant no. BB/R013063/1). SC is funded by Medical Research Council UK grants (MR/M01360X/1, MR/R025576/1, and MR/R020973/1) and Bloomsbury SET (award reference CCF17-7779). The Shloklo Malaria Research Unit is part of the Mahidol Oxford Research Unit, supported by the Wellcome Trust (Grant no.220211). Further funding support was available by FAPESP (Process 02/ 9546-1), CNPq (Process 302353/2003-8 and process 471605/2011-5) to ARSB and RLDM. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

TGC and SC conceived and directed the project. SS, AC, ATC, DF, GVT, FN, KS, DN, CJS, JD, MSM, DBP, CM, ARSB, RLDM, and SMS organised sample collection and processing. AI and SC undertook laboratory work including sequencing. AI and EM performed bioinformatic analysis under the supervision of SC and TGC, and together they interpreted the results. AI, TGC, and SC wrote the first draft of the manuscript. All authors commented on the results and on the manuscript, and approved the final submission.

COMPETING INTERESTS

The authors have declared that no competing interests exist

TABLES

Table 1. Non-synonymous SNPs with top 20 F_{ST} scores that differentiate *P. vivax* isolates from South America and Brazil.

Region	Chr	Position	Ref	Alt	Gene name	AA change*	Nucleotide change	Fst**
South America	13	337753	A	C	CRMP3	1719K>1719N	337753A>C	0.999
South America	12	327858	G	A	P48/45	418R>418K	327858G>A	0.998
South America	11	1265741	A	T	PVP01_1129500	236N>236F	1265740A>T+1265741A>T	0.998
South America	11	1265741	A	T	PVP01_1129500	236N>236I	1265741A>T	0.998
South America	12	323603	C	T	P47	24L>24F	323603C>T	0.996
South America	11	1276259	A	T	PVP01_1129700	2225T>2225S	1276259A>T	0.993
South America	13	336962	G	A	CRMP3	1456V>1456M	336962G>A	0.992
South America	12	424886	A	T	PVP01_1210400	195R>195W	424886A>T	0.990
South America	14	1322275	C	A	PVP01_1430500	1067L>1067I	1322275C>A	0.986
South America	11	1272806	G	A	PVP01_1129700	1074E>1074K	1272806G>A	0.986
South America	4	652108	G	C	P230p	158L>158V	652108G>C	0.985
South America	7	747984	G	A	PVP01_0716800	575G>575S	747984G>A	0.982
South America	9	1199491	C	T	PVP01_0927300	572E>572K	1199491C>T	0.980
South America	11	1483830	A	G	PVP01_1134800	579K>579R	1483830A>G	0.975
South America	14	2662867	T	A	PVP01_1461600	403I>403L	2662867T>A	0.973
South America	9	892855	G	C	PVP01_0920500	1053Q>1053H	892855G>C	0.973
South America	9	878674	C	G	PVP01_0920200	517G>517A	878674C>G	0.967
South America	11	1262951	G	A	PVP01_1129400	20P>20S	1262951G>A	0.966
South America	8	556253	G	A	PVP01_0813100	150S>150N	556253G>A	0.966
South America	11	1514101	T	C	ApiAP2	319N>319D	1514101T>C	0.962
Brazil	10	481636	C	T	MDR1	500D>500N	481636C>T	0.921
Brazil	12	1621163	C	G	ApiAP2	869R>869G	1621163C>G	0.895
Brazil	13	818665	T	C	PVP01_1317400	39K>39E	818665T>C	0.876
Brazil	13	809067	G	A	PVP01_1317200	1086R>1086Q	809067G>A	0.876
Brazil	5	440493	T	C	NT2	117F>117S	440493T>C	0.875
Brazil	2	377716	C	A	PVP01_0209100	590G>590V	377716C>A	0.869
Brazil	12	1618925	A	G	ApiAP2	123I>123V	1618925A>G	0.868
Brazil	4	530215	T	C	PVP01_0412900	299E>299G	530215T>C	0.860
Brazil	1	716831	A	T	PVP01_0116000	4344L>4344M	716831A>T	0.859
Brazil	12	1860075	C	T	PVP01_1245000	1553A>1553T	1860075C>T	0.853
Brazil	13	810706	G	A	PVP01_1317200	1578G>1578D	810706G>A	0.849
Brazil	11	915559	G	T	PK4	1694T>1694N	915559G>T	0.839
Brazil	6	179243	A	T	PVP01_0604500	441L>441M	179243A>T	0.835
Brazil	9	1366817	C	G	SR1	236E>236Q	1366817C>G	0.832
Brazil	14	2887017	C	T	PVP01_1467700	33A>33T	2887017C>T	0.830
Brazil	14	2153846	G	T	PVP01_1449600	1581P>1581T	2153846G>T	0.820
Brazil	10	490615	C	G	PVP01_1011000	842G>842A	490615C>G	0.818
Brazil	14	115657	A	G	RBP2a	719K>719E	115657A>G	0.815
Brazil	13	336738	C	T	CRMP3	1381P>1381L	336738C>T	0.815
Brazil	7	360367	A	C	PVP01_0706700	544K>544Q	360367A>C	0.815

*AA amino acid; ** Within Region vs. all other isolates

Table 2

Mutations in putative drug resistance genes in Brazil, with reference to other regions across the globe (frequencies from 0 (blue) to 1 (red)).

Chromosome	Position	Reference	Alternate	Gene ID	Gene name	AA Change*	Nucleotide Change	Proportion of isolates containing SNP					
								Brazil	South America	East Africa	South Asia	SEA	SSEA
PvP01_02_v1	158272	A	C	PVP01_0203000	MRP1	218Y>218D	158272A>C	1	0.99	1	1	1	0.98
PvP01_02_v1	158545	C	T	PVP01_0203000	MRP1	127V>127I	158545C>T	1	1	1	0.99	1	0.98
PvP01_02_v1	419360	G	T	PVP01_0210400	UBP1	1967P>1967T	419360G>T	1	1	1	1	1	0.98
PvP01_08_v1	902083	C	T	PVP01_0820500	ABCI3	880V>880I	902083C>T	1	1	0.66	0.59	0.3	0.01
PvP01_10_v1	481042	T	C	PVP01_1010900	MDR1	698S>698G	481042T>C	1	1	0.41	0.56	0.01	0
PvP01_10_v1	1054750	T	A	PVP01_1024200	PI4K	1240N>1240Y	1054750T>A	1	1	0.75	0.98	0.96	0.06
PvP01_10_v1	1056253	T	C	PVP01_1024200	PI4K	739K>739E	1056253T>C	1	1	0.87	0.97	0.96	0.15
PvP01_12_v1	2441608	G	C	PVP01_1259100	MDR2	43V>43L	2441608G>C	1	1	1	1	0.5	0.46
PvP01_13_v1	1034368	A	C	PVP01_1322800	ABCG2	124M>124L	1034368A>C	1	1	1	1	0.16	0
PvP01_13_v1	1034368	A	C	PVP01_1322800	ABCG2	124M>124Q	1034368A>C+1034369T>A	1	1	1	1	0.16	0
PvP01_13_v1	1034884	G	A	PVP01_1322800	ABCG2	296V>296I	1034884G>A	1	1	1	1	0.44	0.01
PvP01_13_v1	1035368	A	T	PVP01_1322800	ABCG2	457K>457M	1035368A>T	1	0.68	0	0	0	0
PvP01_13_v1	1035572	G	C	PVP01_1322800	ABCG2	525S>525T	1035572G>C	1	1	1	1	0.72	0.42
PvP01_14_v1	826313	A	C	PVP01_1419000	FD	147S>147A	826313A>C	1	1	1	1	0.55	0.79
PvP01_14_v1	1071124	G	A	PVP01_1424900	DMT1	247H>247Y	1071124G>A	1	1	1	1	1	0.95
PvP01_14_v1	1071643	G	A	PVP01_1424900	DMT1	74H>74Y	1071643G>A	1	1	1	1	0.98	0.16
PvP01_14_v1	1071664	G	T	PVP01_1424900	DMT1	67L>67I	1071664G>T	1	1	1	1	0.98	0.18
PvP01_14_v1	2052257	C	G	PVP01_1447300	MRP2	1956D>1956H	2052257C>G	1	1	1	1	0.98	0.93
PvP01_14_v1	2053883	A	G	PVP01_1447300	MRP2	1414Y>1414H	2053883A>G	1	1	0.67	0.77	0.14	0.5
PvP01_11_v1	1949796	C	T	PVP01_1145600	DHODH	22T>22I	1949796C>T	0.98	0.98	0.75	0.7	0.07	0
PvP01_02_v1	415490	T	C	PVP01_0210400	UBP1	3218N>3218S	415490T>C	0.97	0.94	0.31	0.29	0.83	0.19
PvP01_14_v1	1271444	C	T	PVP01_1429500	PPPK-DHPS	205M>205I	1271444C>T	0.97	0.65	0.77	0	1	0.1
PvP01_10_v1	830910	G	A	PVP01_1018600	PI3K	193P>193S	830910G>A	0.91	0.82	0.15	0.27	0.05	0
PvP01_08_v1	903349	C	T	PVP01_0820500	ABCI3	458A>458T	903349C>T	0.84	0.79	0.71	0.78	0.29	0.4
PvP01_10_v1	481636	C	T	PVP01_1010900	MDR1	500D>500N	481636C>T	0.79	0.31	0	0	0	0
PvP01_02_v1	418478	C	A	PVP01_0210400	UBP1	2261A>2261S	418478C>A	0.75	0.53	0	0	0	0
PvP01_08_v1	904283	C	T	PVP01_0820500	ABCI3	146R>146Q	904283C>T+904284C>T	0.66	0.83	0.44	0.6	0.11	0.17
PvP01_08_v1	904284	C	T	PVP01_0820500	ABCI3	146R>146Q	904283C>T+904284C>T	0.66	0.83	0.44	0.6	0.11	0.17
PvP01_02_v1	423218	C	T	PVP01_0210400	UBP1	681V>681I	423218C>T	0.6	0.79	1	0.88	0.98	0.64
PvP01_14_v1	2054573	G	A	PVP01_1447300	MRP2	1184P>1184S	2054573G>A	0.59	0.44	0	0	0	0
PvP01_14_v1	2054843	C	A	PVP01_1447300	MRP2	1094A>1094S	2054843C>A	0.58	0.46	0	0	0	0
PvP01_10_v1	827121	C	T	PVP01_1018600	PI3K	1456E>1456K	827121C>T	0.48	0.49	0	0	0	0
PvP01_02_v1	418249	G	C	PVP01_0210400	UBP1	2337P>2337R	418249G>C	0.41	0.23	0.05	0	0	0
PvP01_13_v1	519866	T	A	PVP01_1311100	ATP4	177E>177V	519866T>A	0.4	0.61	0.16	0.5	0.07	0
PvP01_08_v1	898533	G	T	PVP01_0820500	ABCI3	2063P>2063H	898533G>T	0.39	0.3	0	0	0	0
PvP01_08_v1	901460	G	T	PVP01_0820500	ABCI3	1087D>1087E	901460G>T	0.39	0.42	0.22	0	0	0
PvP01_14_v1	1270914	G	C	PVP01_1429500	PPPK-DHPS	382S>382C	1270914G>C	0.37	0.16	0	0	0.01	0
PvP01_02_v1	156208	C	G	PVP01_0203000	MRP1	906E>906Q	156208C>G	0.35	0.32	0.48	0.44	0.09	0.12
PvP01_02_v1	422851	C	T	PVP01_0210400	UBP1	803R>803Q	422851C>T	0.35	0.18	0.01	0	0	0
PvP01_14_v1	2053904	C	G	PVP01_1447300	MRP2	1407E>1407Q	2053904C>G	0.35	0.31	0.04	0.15	0.06	0
PvP01_08_v1	900117	T	A	PVP01_0820500	ABCI3	1535Q>1535L	900117T>A	0.34	0.25	0	0	0	0
PvP01_13_v1	517064	C	T	PVP01_1311100	ATP4	1111S>1111N	517064C>T	0.3	0.51	0	0.11	0	0
PvP01_02_v1	422286	G	T	PVP01_0210400	UBP1	991D>991E	422286G>T	0.29	0.18	0.15	0.41	0.09	0
PvP01_13_v1	517034	G	T	PVP01_1311100	ATP4	1121A>1121D	517034G>T	0.29	0.39	0	0.11	0	0
PvP01_02_v1	154668	C	G	PVP01_0203000	MRP1	1419G>1419A	154668C>G	0.27	0.39	0.2	0.25	0.01	0
PvP01_12_v1	2443022	A	T	PVP01_1259100	MDR2	514Y>514F	2443022A>T	0.24	0.21	0.88	0.95	0.17	0.91
PvP01_02_v1	415407	G	T	PVP01_0210400	UBP1	3246P>3246T	415407G>T	0.23	0.16	0.25	0.35	0.19	0.53
PvP01_10_v1	480261	A	G	PVP01_1010900	MDR1	958M>958T	480261A>G	0.21	0.1	0	0	0	0
PvP01_02_v1	154108	C	T	PVP01_0203000	MRP1	1606A>1606N	154107G>T+154108C>T	0.2	0.09	0	0	0	0
PvP01_14_v1	2909751	T	G	PVP01_1468300	CORONIN	551F>551C	2909751T>G	0.2	0.12	0	0	0	0
PvP01_02_v1	154107	G	T	PVP01_0203000	MRP1	1606A>1606N	154107G>T+154108C>T	0.19	0.09	0	0	0	0

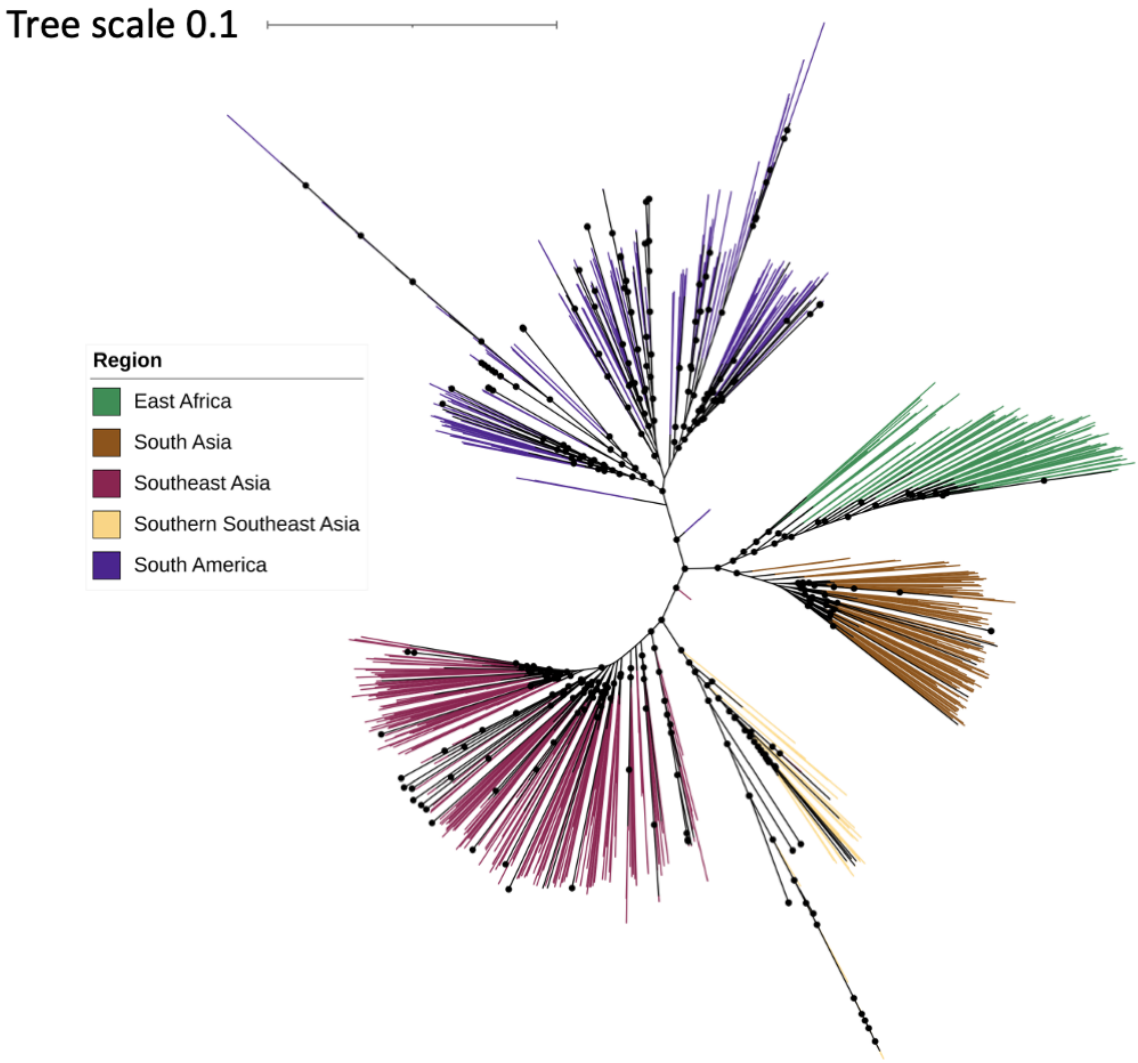
PvP01_02_v1	154294	C	T	PVP01_0203000	MRP1	1544V>1544I	154294C>T	0.19	0.09	0	0	0	0
PvP01_14_v1	2750512	A	C	PVP01_1464000	PKB	272S>272R	2750512A>C	0.19	0.07	0.22	0.79	0.06	0
PvP01_02_v1	424340	C	T	PVP01_0210400	UBP1	307G>307R	424340C>T	0.18	0.1	0	0.03	0.05	0
PvP01_14_v1	2750469	C	T	PVP01_1464000	PKB	287A>287T	2750469C>T	0.16	0.1	0	0	0	0
PvP01_13_v1	516804	C	A	PVP01_1311100	ATP4	1198D>1198Y	516804C>A	0.13	0.16	0	0.03	0.01	0
PvP01_14_v1	2750517	C	T	PVP01_1464000	PKB	271E>271K	2750517C>T	0.13	0.29	0	0	0	0
PvP01_08_v1	901400	A	T	PVP01_0820500	ABCI3	1107H>1107Q	901400A>T	0.12	0.3	0.14	0.22	0.01	0.27
PvP01_13_v1	1802108	T	C	PVP01_1340900	PM4	165I>165V	1802108T>C	0.12	0.53	0.26	0.09	0.62	0.28
PvP01_14_v1	2057861	C	G	PVP01_1447300	MRP2	88E>88Q	2057861C>G	0.12	0.25	0.67	0.7	0.16	0.49
PvP01_02_v1	422965	T	A	PVP01_0210400	UBP1	765H>765L	422965T>A	0.1	0.06	0.29	0.12	0.01	0
PvP01_14_v1	2054372	T	A	PVP01_1447300	MRP2	1251N>1251Y	2054372T>A	0.1	0.1	0	0	0	0
PvP01_02_v1	154492	T	C	PVP01_0203000	MRP1	1478I>1478V	154492T>C	0.06	0.14	0.4	0.22	0.01	0
PvP01_02_v1	416893	G	C	PVP01_0210400	UBP1	2750F>2750L	416893G>C	0.06	0.03	0	0.11	0.06	0
PvP01_02_v1	418197	C	A	PVP01_0210400	UBP1	2354R>2354S	418197C>A	0.06	0.02	0	0	0	0
PvP01_02_v1	418894	A	G	PVP01_0210400	UBP1	2122I>2122T	418894A>G	0.06	0.11	0	0	0	0
PvP01_02_v1	422483	G	A	PVP01_0210400	UBP1	926R>926C	422483G>A	0.06	0.02	0	0.14	0.01	0
PvP01_08_v1	903090	A	C	PVP01_0820500	ABCI3	544F>544C	903090A>C	0.06	0.17	0	0	0	0
PvP01_12_v1	2445106	G	A	PVP01_1259100	MDR2	1209G>1209S	2445106G>A	0.06	0.19	0	0	0	0
PvP01_12_v1	2445238	G	A	PVP01_1259100	MDR2	1253G>1253R	2445238G>A	0.06	0.02	0	0	0	0
PvP01_13_v1	516924	C	T	PVP01_1311100	ATP4	1158E>1158K	516924C>T	0.06	0.17	0.3	0.26	0.56	0.3
PvP01_14_v1	1070602	G	T	PVP01_1424900	DMT1	421L>421I	1070602G>T	0.06	0.03	0	0	0	0
PvP01_14_v1	1070614	T	A	PVP01_1424900	DMT1	417I>417F	1070614T>A	0.06	0.15	0	0	0	0
PvP01_02_v1	424495	C	T	PVP01_0210400	UBP1	255R>255H	424495C>T	0.05	0.02	0	0.01	0	0
PvP01_02_v1	424684	T	C	PVP01_0210400	UBP1	192D>192G	424684T>C	0.05	0.02	0.05	0.03	0.03	0
PvP01_08_v1	900780	T	C	PVP01_0820500	ABCI3	1314K>1314R	900780T>C	0.05	0.15	0.05	0.18	0.31	0.09
PvP01_08_v1	900780	T	C	PVP01_0820500	ABCI3	1314K>1314S	900779T>G+900780T>C	0.05	0.15	0.05	0.18	0.31	0.09
PvP01_10_v1	829671	T	C	PVP01_1018600	PI3K	606I>606V	829671T>C	0.05	0.06	0	0	0	0
PvP01_02_v1	157143	C	T	PVP01_0203000	MRP1	594R>594K	157143C>T	0.04	0.01	0	0	0	0
PvP01_02_v1	416004	C	A	PVP01_0210400	UBP1	3047A>3047S	416004C>A	0.04	0.01	0	0	0.05	0
PvP01_10_v1	829517	C	T	PVP01_1018600	PI3K	657R>657K	829517C>T	0.04	0.04	0	0	0	0
PvP01_14_v1	1270911	C	G	PVP01_1429500	PPPK-DHPS	383G>383A	1270911C>G	0.04	0.44	0.7	0.84	0.08	0.33
PvP01_14_v1	2055095	T	C	PVP01_1447300	MRP2	1010M>1010V	2055095T>C	0.04	0.04	0	0	0	0
PvP01_02_v1	154168	G	A	PVP01_0203000	MRP1	1586H>1586Y	154168G>A	0.03	0.25	0	0.09	0.03	0
PvP01_02_v1	154350	G	A	PVP01_0203000	MRP1	1525T>1525I	154350G>A	0.03	0.05	0	0	0	0
PvP01_02_v1	154486	T	C	PVP01_0203000	MRP1	1480I>1480V	154486T>C	0.03	0.02	0	0	0	0
PvP01_02_v1	155206	A	T	PVP01_0203000	MRP1	1240F>1240I	155206A>T	0.03	0.02	0	0	0	0
PvP01_10_v1	480412	G	T	PVP01_1010900	MDR1	908L>908M	480412G>T	0.03	0.09	0.01	0	0	0
PvP01_10_v1	827999	C	T	PVP01_1018600	PI3K	1163G>1163E	827999C>T	0.03	0.03	0	0.02	0	0
PvP01_12_v1	2441670	A	T	PVP01_1259100	MDR2	63R>63S	2441670A>T	0.03	0.01	0.04	0	0	0
PvP01_12_v1	2442764	T	C	PVP01_1259100	MDR2	428L>428S	2442764T>C	0.03	0.02	0	0	0	0
PvP01_13_v1	517790	T	A	PVP01_1311100	ATP4	869E>869V	517790T>A	0.03	0.01	0	0	0	0
PvP01_14_v1	2056117	C	A	PVP01_1447300	MRP2	669S>669I	2056117C>A	0.03	0.05	0	0.03	0	0.01
PvP01_14_v1	2057278	A	C	PVP01_1447300	MRP2	282M>282R	2057278A>C	0.03	0.07	0	0	0	0
PvP01_14_v1	2058037	G	C	PVP01_1447300	MRP2	29T>29S	2058037G>C	0.03	0.01	0	0	0	0
PvP01_02_v1	416378	C	A	PVP01_0210400	UBP1	2922S>2922I	416378C>A	0.02	0.01	0	0	0	0
PvP01_02_v1	424690	C	T	PVP01_0210400	UBP1	190R>190K	424690C>T	0.02	0.01	0	0	0	0
PvP01_02_v1	424896	G	T	PVP01_0210400	UBP1	121D>121E	424896G>T	0.02	0.01	0	0	0	0
PvP01_05_v1	1077707	A	G	PVP01_0526600	DHFR-TS	116S>116G	1077707A>G	0.02	0.01	0	0	0	0
PvP01_06_v1	351403	C	G	PVP01_0607800	KELCH10	267N>267K	351403C>G	0.02	0.02	0	0	0	0
PvP01_06_v1	351597	A	G	PVP01_0607800	KELCH10	332N>332S	351597A>G	0.02	0.01	0	0	0	0
PvP01_06_v1	352661	G	T	PVP01_0607800	KELCH10	687V>687L	352661G>T	0.02	0.02	0.01	0.01	0	0
PvP01_08_v1	898837	G	T	PVP01_0820500	ABCI3	1962Q>1962K	898837G>T	0.02	0.02	0	0	0	0
PvP01_08_v1	902268	T	C	PVP01_0820500	ABCI3	818D>818G	902268T>C	0.02	0.02	0	0	0	0
PvP01_10_v1	480552	G	T	PVP01_1010900	MDR1	861A>861E	480552G>T	0.02	0.01	0	0.05	0.05	0
PvP01_02_v1	155080	A	T	PVP01_0203000	MRP1	1282L>1282I	155080A>T	0.01	0.08	0	0	0	0
PvP01_02_v1	156089	G	T	PVP01_0203000	MRP1	945F>945L	156089G>T	0.01	0	0	0	0	0

PvP01_02_v1	158818	T	G	PVP01_0203000	MRP1	36K>36Q	158818T>G	0.01	0.01	0	0	0.04	0.04
PvP01_02_v1	416416	G	T	PVP01_0210400	UBP1	2909N>2909K	416416G>T	0.01	0	0	0	0	0
PvP01_02_v1	419181	C	A	PVP01_0210400	UBP1	2026R>2026S	419181C>A	0.01	0.01	0	0	0	0
PvP01_02_v1	420605	C	T	PVP01_0210400	UBP1	1552V>1552I	420605C>T	0.01	0.17	0	0	0	0
PvP01_05_v1	1077534	G	A	PVP01_0526600	DHFR-TS	58R>58K	1077534G>A	0.01	0.07	0	0	0	0
PvP01_10_v1	479133	G	T	PVP01_1010900	MDR1	1334A>1334E	479133G>T	0.01	0	0	0	0	0
PvP01_10_v1	829733	T	C	PVP01_1018600	PI3K	585E>585G	829733T>C	0.01	0.01	0	0	0	0
PvP01_13_v1	519722	T	C	PVP01_1311100	ATP4	225E>225G	519722T>C	0.01	0.04	0	0	0	0
PvP01_13_v1	1035259	C	T	PVP01_1322800	ABCG2	421L>421F	1035259C>T	0.01	0.32	0.99	0.68	0.08	0
PvP01_14_v1	826496	C	T	PVP01_1419000	FD	86G>86R	826496C>T	0.01	0.03	0	0	0	0
PvP01_14_v1	2909573	G	T	PVP01_1468300	CORONIN	492A>492S	2909573G>T	0.01	0	0.05	0.32	0.04	0

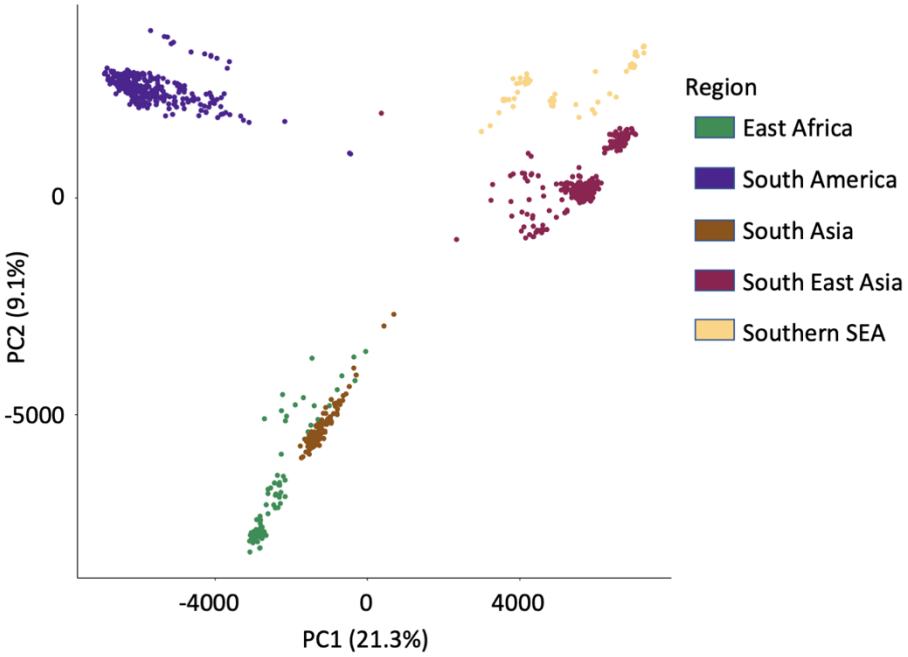
*Resulting amino acid alterations and genes affected were predicted using SnpEff

FIGURES

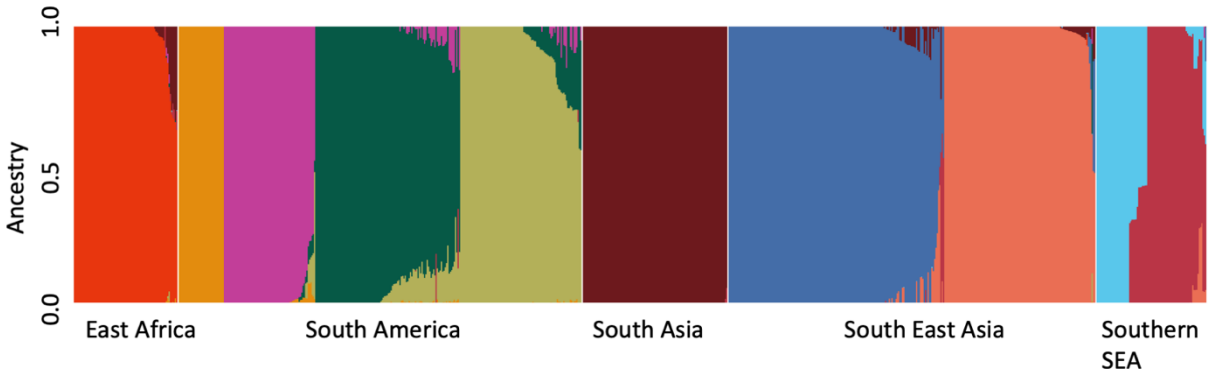
A)


Tree scale 0.1
Region
East Africa
South Asia
Southeast Asia
Southern Southeast Asia
South America

B)


Region
East Africa
South America
South Asia
South East Asia
Southern SEA
0
-5000
PC2 (9.1%)
-4000
0
4000
PC1 (21.3%)

C)


Ancestry
1.0
0.5
0.0
East Africa
South America
South Asia
South East Asia
Southern SEA

D)

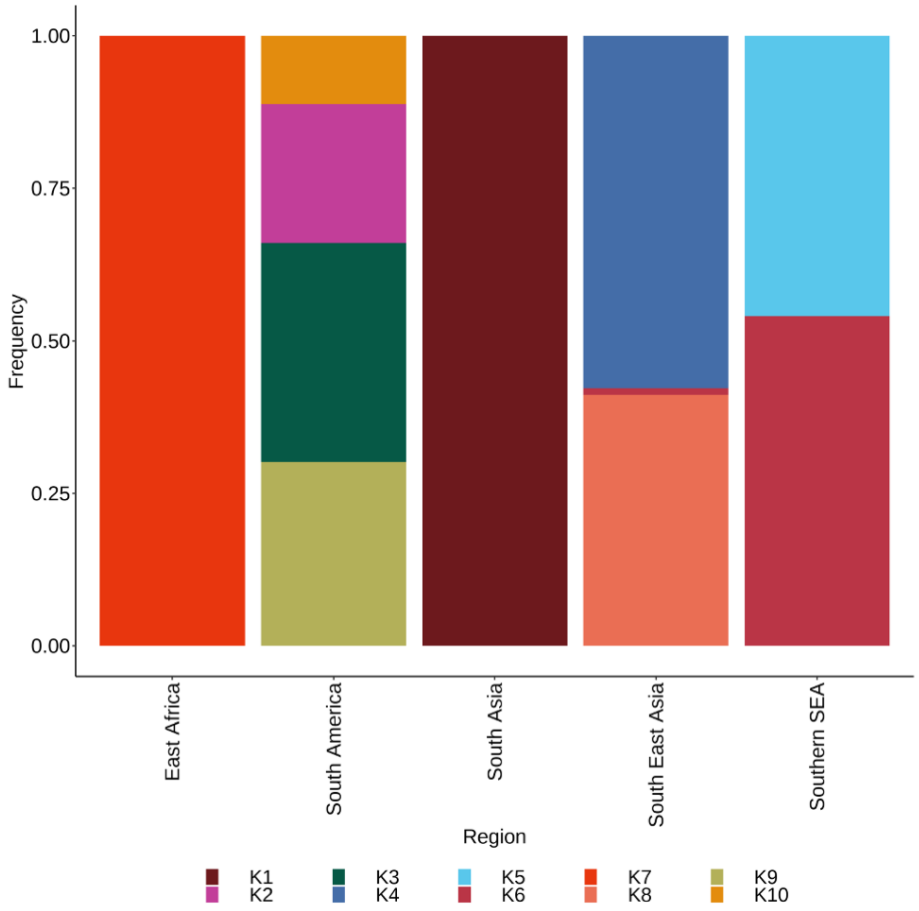


Figure 1

Population structure of 885 *P. vivax* isolates from 26 countries.

A) Maximum likelihood phylogenetic tree generated using IQTREE from the pairwise SNP matrix of the complete global dataset of 885 samples and 454,681 SNPs. IQTREE was run using ModelFinder, tree search, ultrafast bootstrap and SH-aLRT test. Bootstrap scores between 50% and 100% are annotated on the tree branches with a black circle. Branches are colouring according to regional grouping (East Africa, n = 84, green branches; South America, n = 315, purple branches; South Asia, n = 114, brown branches; Southeast Asia (SEA), n = 286, dark pink branches; Southern SEA, n = 86, yellow branches). The phylogenetic tree file was visualised in iTOL with midpoint rooting. **B)** Principal component analysis displaying principal components 1 and 2 of the distance matrix generated using the SNP matrix. Each point represents and individual sample, coloured according with the region assigned in **Figure 1A**. Principal component 1 summarises 21.3% of the total variation whilst principal component 2 summarises 9.1% of

the total variation. **C)** ADMIXTURE analysis of the global dataset predicted a total of 10 ancestral populations spread across each region: East Africa, n = 1; South America, n = 4; South Asia, n = 1; SEA, n = 3; SSEA, n = 2. **D)** Bar plot summarising the number and proportion of each ancestral population within each region.

A)

B)

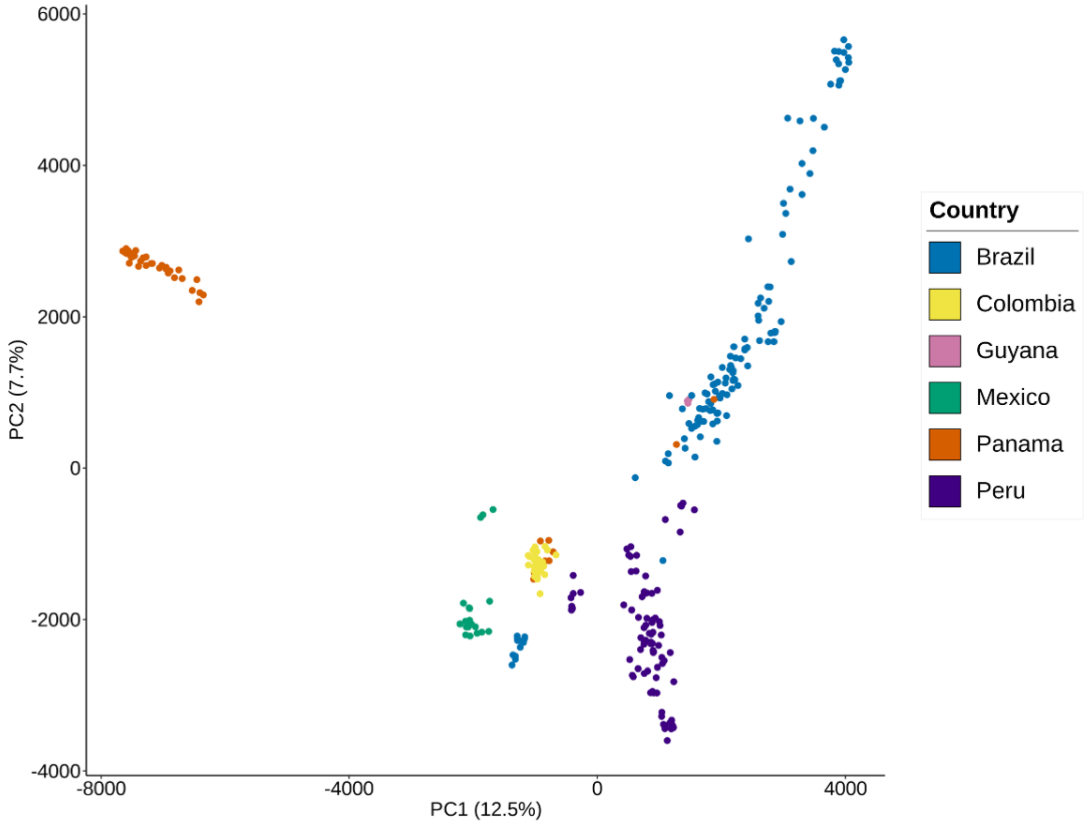
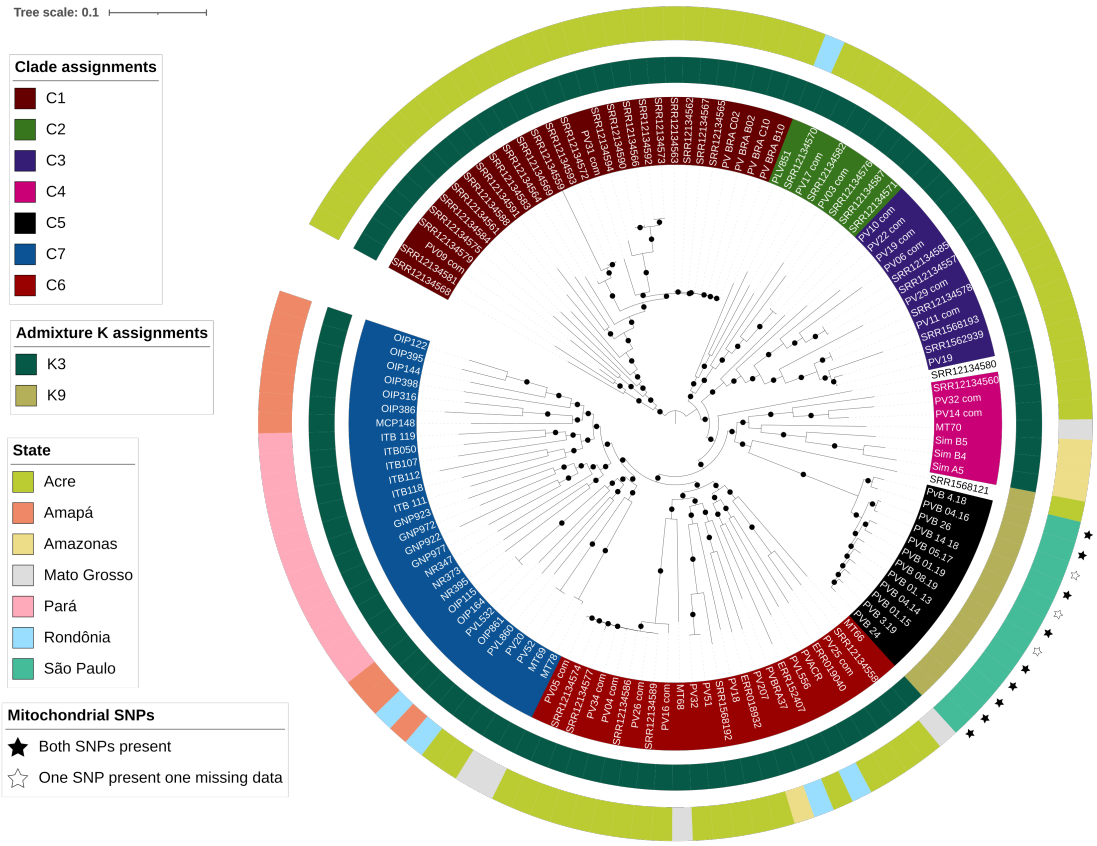


Figure 2

Population structure of South American isolates (n = 315).

A) Maximum likelihood (ML) tree using SNP data (102,765 unique SNPs) from 315 isolates from South America (Brazil, n = 123; Colombia, n = 34; Mexico, n = 20; Panama, n = 46; Peru, n = 89; Guyana, n = 3). The outer circle track is coloured according to the country of each isolate (Brazil - blue, Colombia - yellow, Guyana - pink, Mexico - green, Panama - orange, Peru - purple), and the inner circle track denotes the population assigned to each isolate after ADMIXTURE analysis of the entire global dataset. ADMIXTURE denoted four populations within South America (K2 - pink, K3 - dark green, K9 - light green, K10 - orange). IQTREE was used to generate ML trees using ModelFinder (which calculated GTR+F+R10 as the model with the best fit), tree search, ultrafast bootstrap and SH-aLRT test. **B)** Principal component analysis of the pairwise distance matrix generated using the 102,765 SNP matrix from 315 South American isolates. Each point denotes a sample, which is coloured according to the country, as with the tree in **A)**.

A)


Tree scale: 0.1
Clade assignments
C1
C2
C3
C4
C5
C7
C6
Admixture K assignments
K3
K9
State
Acre
Amapá
Amazonas
Mato Grosso
Pará
Rondônia
São Paulo
Mitochondrial SNPs
Both SNPs present
One SNP present one missing data

B)

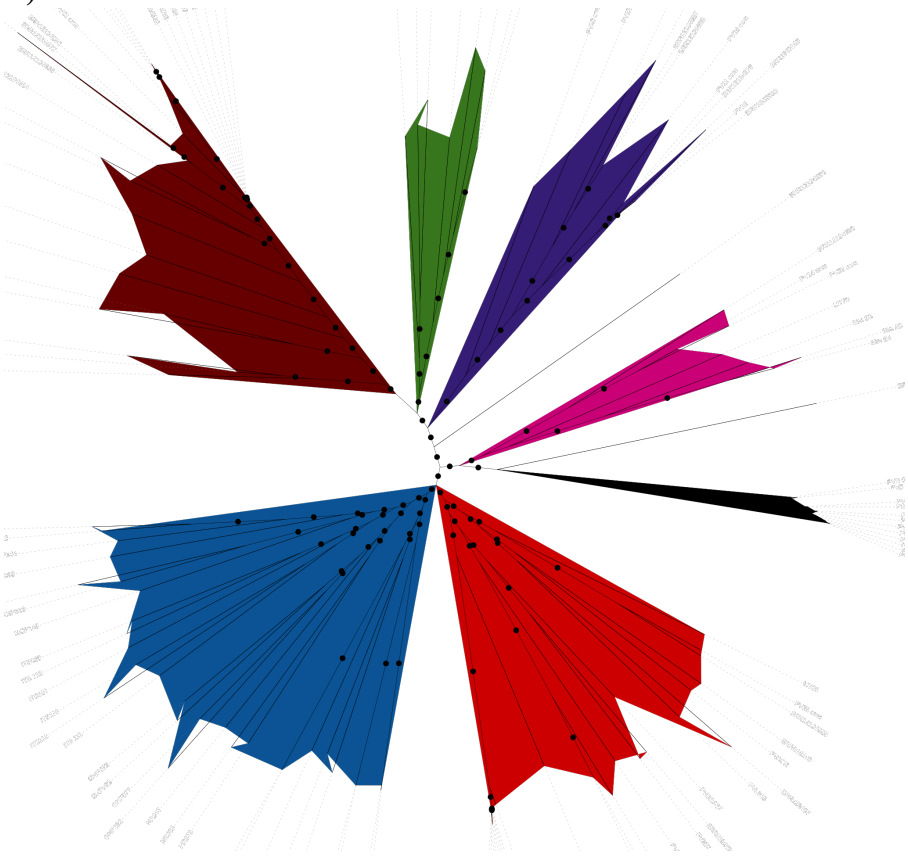


Figure 3

Population structure within Brazil isolates (n=123).

Maximum likelihood tree using SNP data (70,757 unique SNPs) from 123 isolates from seven states within Brazil. **A)** Circular phylogenetic tree in iTOL, with outer colour strip coloured according to the state (*Pará*, n = 13; Amapá, n = 10; Mato Grosso, n = 5; Rondônia, n = 5; Acre, n = 74; Amazonas, n = 4; São Paulo, n = 12), the inner colour strip highlighting the ADMIXTURE population assignments from the global analysis (K3, n = 13; K9, n = 110), and the inner label coloured according to the 7 clades assigned based on the tree topology (C1, n = 29; C2, n = 8; C3, n = 12; C4, n = 7; C5, n = 12; C6, n = 24; C7, n = 29). Two isolates, SRR12134580 and SRR1568121 were not assigned a clade grouping. Isolates containing the two investigated mitochondrial SNPs (T4133C and A4467G in PvP01_MIT_v1) are labelled using a black star on the outer perimeter of the colour track; isolates containing both SNPs have a black filled in star, isolates where one SNP is present, but there is missing data for the other SNP are denoted with a white star with a black outline. **B)** Mid-point rooted visualisation of the same tree in **A)** to demonstrate clade groupings. The maximum likelihood tree for both plots was generated using IQTREE with ModelFinder (which assigned TVM+F+R5 as the model with the best fit), tree search, ultrafast bootstrap and SH-aLRT test. Bootstrap values between 50% and 100% are indicated by a black circle midway along each branch length.

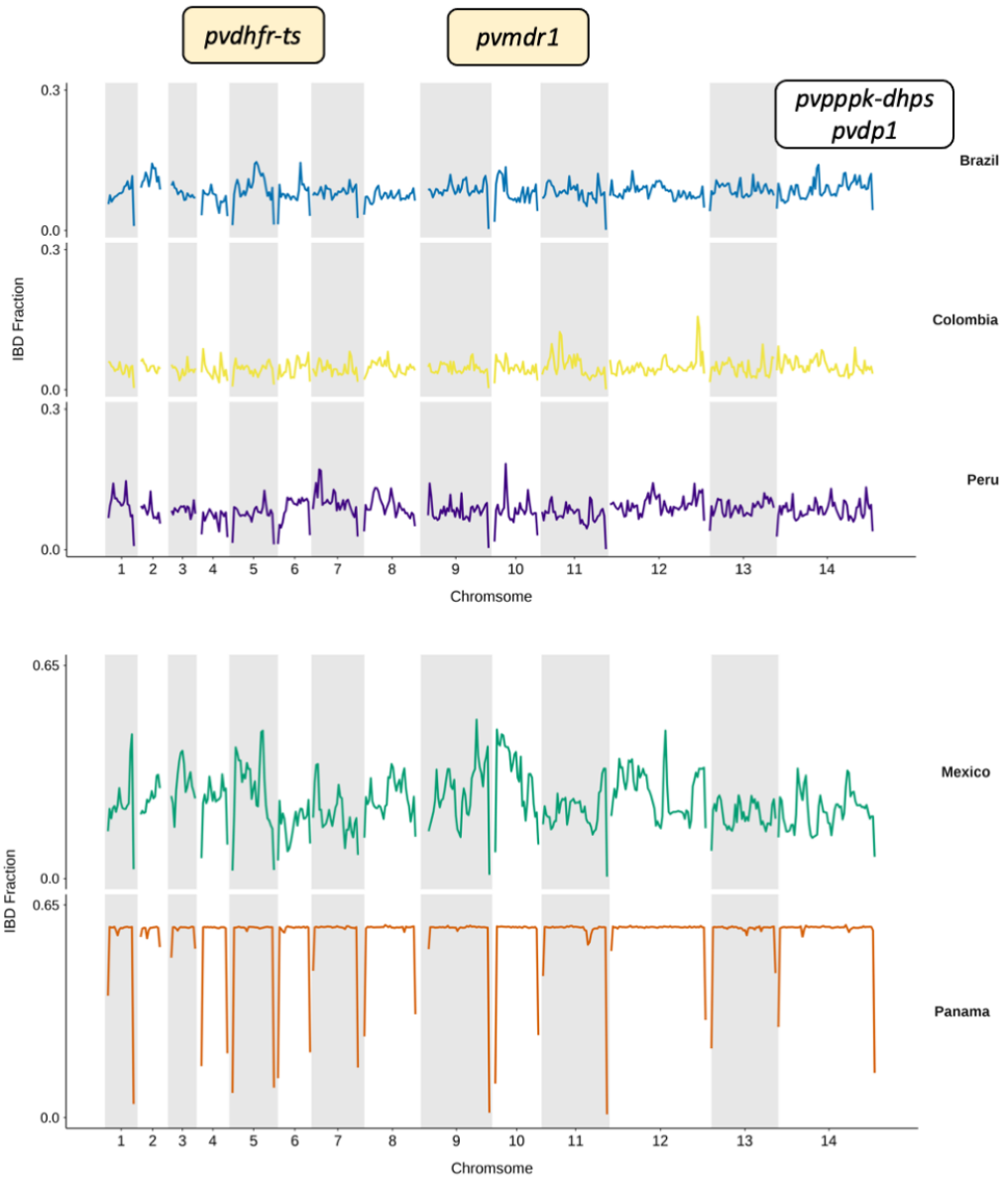


Figure 4

Country level comparisons of identity by descent (IBD) across the whole genome of monoclonal *P. vivax* isolates. IBD fractions along 50 kbp sliding windows across the genome at country level separation. The top 1% of IBD fractions for each country is summarised in **S10 Table**. Genes of interest which demonstrate high signals of IBD are annotated. Where signals of high IBD are conserved across all countries within South America, the gene annotation is at the top of all plots and highlighted in yellow (*pvdhfr-ts* on chromosome 5 and *pvmdr1* on chromosome 10). For country specific signals of high IBD where genes of interest are found, the gene annotation is found above the line graph for each country (*pvpppk-dhps*/*pvdbp1* within chromosome 14 in Brazil).

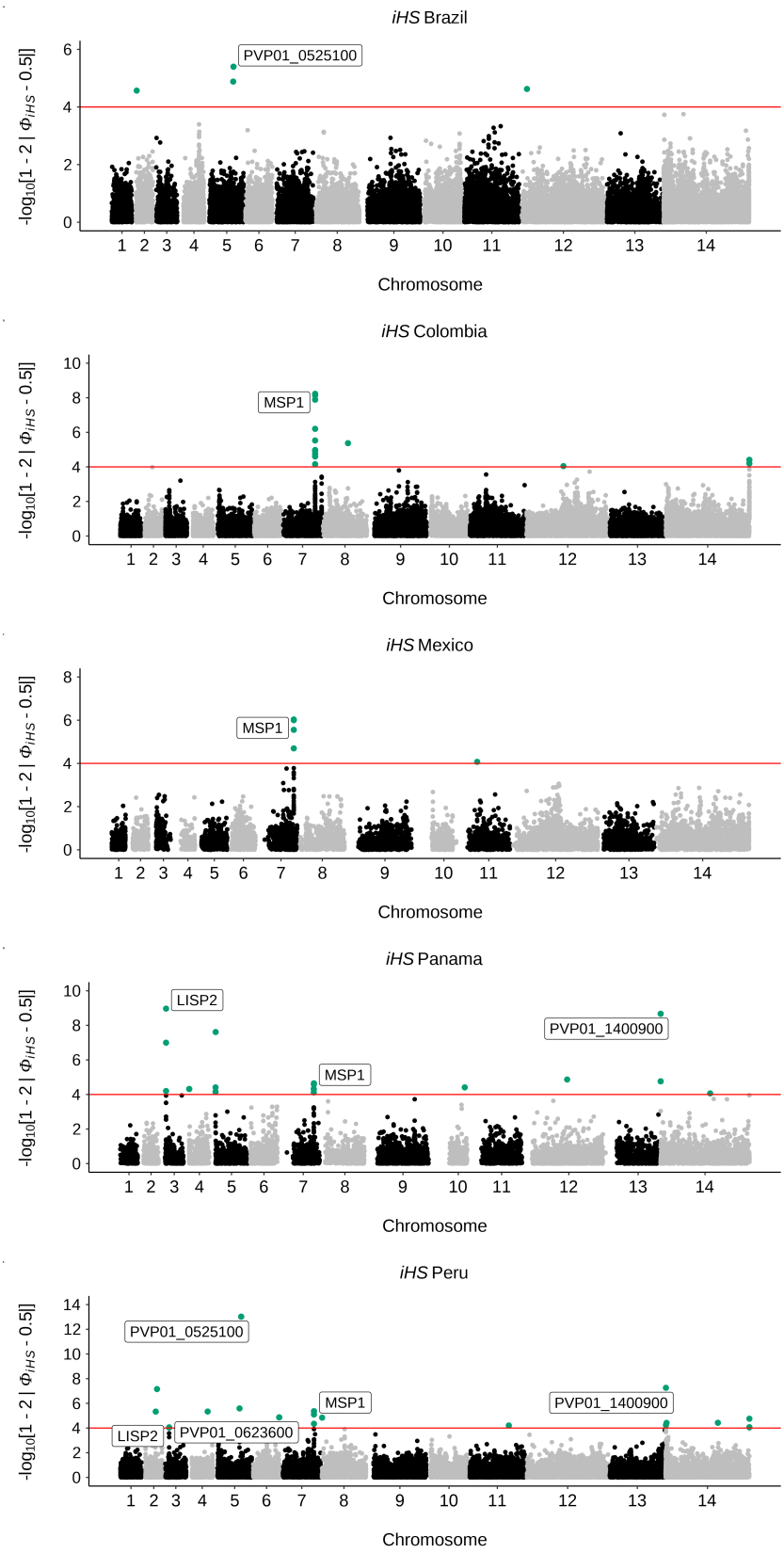
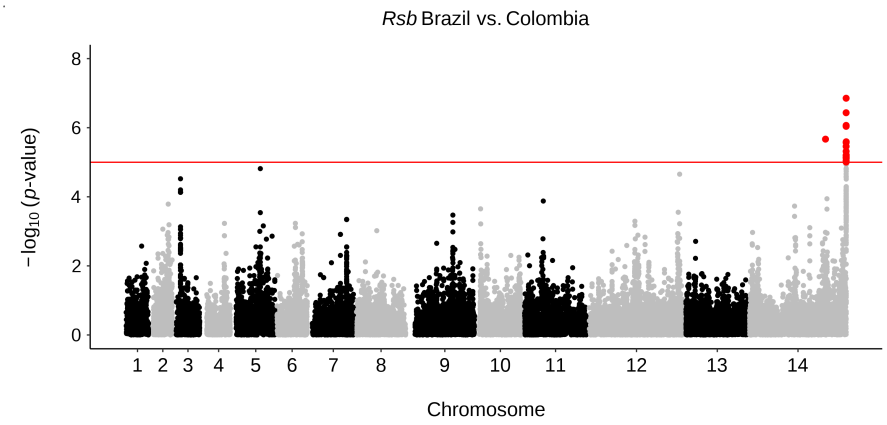
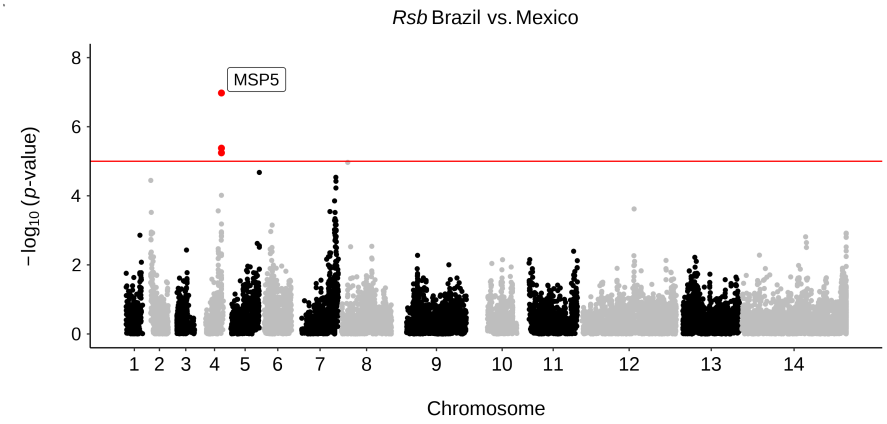

iHS Brazil
PVP01_0525100
-log10[1 - 2 | Φ_iHS - 0.5|]
Chromosome
iHS Colombia
MSP1
-log10[1 - 2 | Φ_iHS - 0.5|]
Chromosome
iHS Mexico
MSP1
-log10[1 - 2 | Φ_iHS - 0.5|]
Chromosome
iHS Panama
LISP2
MSP1
PVP01_1400900
-log10[1 - 2 | Φ_iHS - 0.5|]
Chromosome
iHS Peru
PVP01_0525100
MSP1
PVP01_1400900
LISP2
PVP01_0623600
-log10[1 - 2 | Φ_iHS - 0.5|]
Chromosome

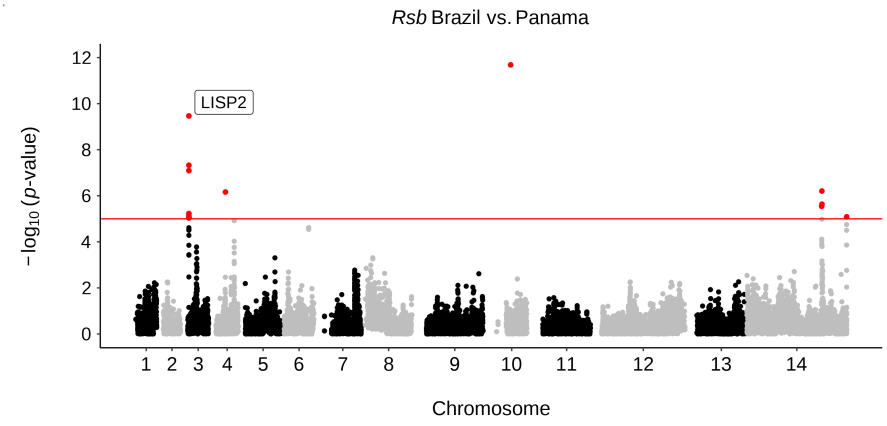
Figure 5

Evidence of selection (iHS) within South American countries.

Genome-wide iHS scores in a Manhattan plot for all countries within South America where there are >10 isolates. SNPs within genes with iHS score of $P < 1x\ 10^{-4}$ are highlighted in green and gene names are annotated for candidate regions of high iHS. Raw outputs of iHS scores, alongside proposed candidate regions for South America and Brazil specifically are summarised in **S14 Table, S15 Table** and **S16 Table**.


Rsb Brazil vs. Colombia
$-\log_{10}$ (*p*-value)
Chromosome


Rsb Brazil vs. Mexico
MSP5
$-\log_{10}$ (*p*-value)
Chromosome


Rsb Brazil vs. Panama
LISP2
$-\log_{10}$ (*p*-value)
Chromosome

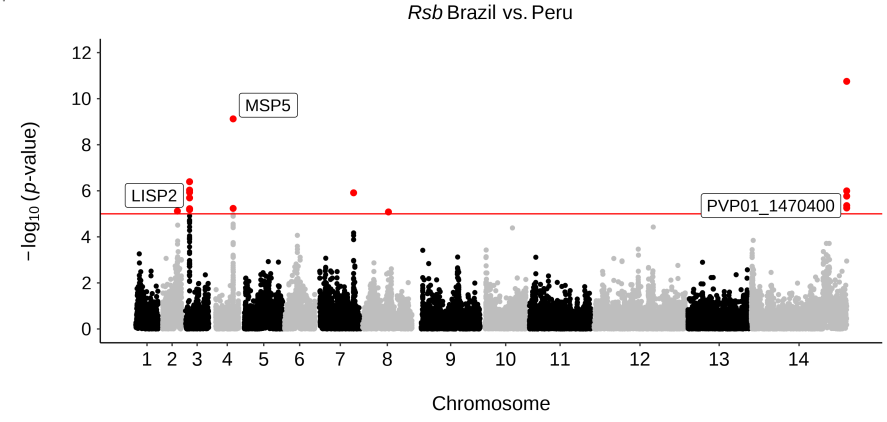

Rsb Brazil vs. Peru
MSP5
LISP2
PVP01_1470400
$-\log_{10}$ (*p*-value)
Chromosome

Figure 6

Evidence of selection between countries in South America (Rsb).

Manhattan plots for genome-wide Rsb analysis for *P. vivax* isolates within South America at the country level. SNPs with $P < 1 \times 10^{-5}$ are highlighted in red, and gene names are annotated for candidate regions of selection. Rsb results output for South America, in addition to within Brazil analyses are summarised in in **S17 Table, S18 Table, S19 Table** and **S20 Table**.

REFERENCES

1. WHO. *World Malaria Report 2020*. (2020).
2. Howes, R. E. *et al*. Global Epidemiology of Plasmodium vivax. *Am. J. Trop. Med. Hyg* **95**, 15–34 (2016).
3. Alexandre, M. A. *et al*. Severe Plasmodium vivax malaria, Brazilian Amazon. *Emerg. Infect. Dis*. **16**, 1611–1614 (2010).
4. World Health Organization. Guidelines for the Treatment of Malaria. Third Edition. (2015).
5. Recht, J. *et al*. Malaria in Brazil , Colombia , Peru and Venezuela : current challenges in malaria control and elimination. *Malar. J*. 1–18 (2017) doi:10.1186/s12936-017-1925-6.
6. Almeida, A. C. G. *et al*. High proportions of asymptomatic and submicroscopic Plasmodium vivax infections in a peri-urban area of low transmission in the Brazilian Amazon. *Parasites and Vectors* **11**, 1–13 (2018).
7. MacDonald, A. J. & Mordecai, E. A. Amazon deforestation drives malaria transmission, and malaria burden reduces forest clearing. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 22212–22218 (2019).
8. Gonçalves, L. A., Cravo, P. & Ferreira, M. U. Emerging Plasmodium vivax resistance to chloroquine in South America: an overview. *Mem. Inst. Oswaldo Cruz* **109**, 534–9 (2014).
9. Carlos, B. C., Rona, L. D. P., Christophides, G. K. & Souza-Neto, J. A. A comprehensive analysis of malaria transmission in Brazil. *Pathog. Glob. Health* **113**, 1–13 (2019).
10. Anete, D. *et al*. Simian malaria in the Brazilian Atlantic forest: first description of natural infection of capuchin monkeys (Cebinae subfamily) by Plasmodium simium. (2011) doi:10.1186/s12936-015-0606-6.
11. Brasil, P. *et al*. Outbreak of human malaria caused by Plasmodium simium in the Atlantic Forest in Rio de Janeiro: a molecular epidemiological investigation. *Lancet. Glob. Heal.* **5**, e1038–e1046 (2017).
12. Carvalho, T. A. A. *et al*. Plasmodium vivax infection in Anajás, State of Pará: No differential resistance profile among Duffy-negative and Duffy-positive individuals. *Malar. J.* **11**, 430 (2012).
13. Zimmerman, P. A. Plasmodium vivax infection in duffy-negative people in Africa. *American Journal of Tropical Medicine and Hygiene* vol. 97 636–638 (2017).
14. Santolamazza, F. *et al*. Insertion polymorphisms of SINE200 retrotransposons within speciation islands of Anopheles gambiae molecular forms. *Malar. J.* **7**, 1–10 (2008).
15. Abou-Ali, R. K. *et al*. Impact of Duffy polymorphisms on parasite density in Brazilian Amazonian patients infected by Plasmodium vivax. *Malar. J.* **18**, 1–9 (2019).
16. De Oliveira Padilha, M. A. *et al*. Comparison of malaria incidence rates and socioeconomic-environmental factors between the states of Acre and Rondônia: A spatio-Temporal modelling

study. *Malar. J.* **18**, 306 (2019).

17. Gething, P. W. *et al.* Modelling the global constraints of temperature on transmission of Plasmodium falciparum and P. vivax. (2011) doi:10.1186/1756-3305-4-92.
18. Battle, K. E. *et al. The Global Public Health Significance of Plasmodium vivax. Advances in Parasitology* vol. 80 (Elsevier, 2012).
19. Marques, M. M. *et al.* Plasmodium vivax chloroquine resistance and anemia in the western brazilian amazon. *Antimicrob. Agents Chemother.* **58**, 342–347 (2014).
20. de Oliveira, T. C. *et al.* Genome-wide diversity and differentiation in New World populations of the human malaria parasite Plasmodium vivax. *PLoS Negl. Trop. Dis.* **11**, e0005824 (2017).
21. 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).
22. Benavente, E. D. *et al.* A molecular barcode to inform the geographical origin and transmission dynamics of Plasmodium vivax malaria. *PLoS Genet.* **16**, e1008576 (2020).
23. Diez Benavente, E. *et al.* Distinctive genetic structure and selection patterns in Plasmodium vivax from South Asia and East Africa. doi:10.1038/s41467-021-23422-3.
24. Pearson, R. D. *et al.* Genomic analysis of local variation and recent evolution in Plasmodium vivax. *Nat. Genet.* **48**, 959–964 (2016).
25. Crippa De Oliveira, T. *et al.* Population genomics reveals the expansion of highly inbred Plasmodium vivax lineages in the main malaria hotspot of Brazil. (2020) doi:10.1371/journal.pntd.0008808.
26. Hupalo, D. N. *et al.* Population genomics studies identify signatures of global dispersal and drug resistance in Plasmodium vivax. *Nat. Genet.* **48**, 953–959 (2016).
27. 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).
28. Mourier, T. *et al.* The genome of the zoonotic malaria parasite Plasmodium simium reveals adaptations to host-switching. *bioRxiv* (2019) doi:10.1101/841171.
29. Cui, L., Fan, Q., Cui, L. & Miao, J. Histone lysine methyltransferases and demethylases in Plasmodium falciparum Corresponding author HHS Public Access. *Int J Parasitol* **38**, 1083–1097 (2008).
30. Douradinha, B. *et al.* Plasmodium Cysteine Repeat Modular Proteins 3 and 4 are essential for malaria parasite transmission from the mosquito to the host. *Malar. J.* **10**, (2011).
31. Tomas, A. M. *et al.* P25 and P28 proteins of the malaria ookinete surface have multiple and partially redundant functions. *EMBO J.* **20**, 3975 (2001).
32. Molina-Cruz, A., Canepa, G. E. & Barillas-Mury, C. Plasmodium P47: a key gene for malaria

transmission by mosquito vectors. *Curr. Opin. Microbiol.* **40**, 168 (2017).

33. Van Dijk, M. R. *et al*. A central role for P48/45 in malaria parasite male gamete fertility. *Cell* **104**, 153–164 (2001).
34. Koenderink, J. B., Kavishe, R. A., Rijpma, S. R. & Russel, F. G. M. The ABCs of multidrug resistance in malaria. *Trends Parasitol.* **26**, 440–446 (2010).
35. Sultan, A. A. *et al*. TRAP Is Necessary for Gliding Motility and Infectivity of Plasmodium Sporozoites. *Cell* **90**, 511–522 (1997).
36. Moreira, C. K. *et al*. The Plasmodium TRAP/MIC2 family member, TRAP-Like Protein (TLP), is involved in tissue traversal by sporozoites. *Cell. Microbiol.* **10**, 1505 (2008).
37. Puentes, A. *et al*. Identifying Plasmodium falciparum merozoite surface protein-10 human erythrocyte specific binding regions. *Biochimie* **87**, 461–472 (2005).
38. Jeninga, M. D., Quinn, J. E. & Petter, M. ApiAP2 Transcription Factors in Apicomplexan Parasites. *Pathogens* **8**, (2019).
39. Malleret, B. *et al*. Plasmodium vivax binds host CD98hc (SLC3A2) to enter immature red blood cells. *Nat. Microbiol. 2021 68* **6**, 991–999 (2021).
40. Ding, S. *et al*. Anti-folate combination therapies and their effect on the development of drug resistance in Plasmodium vivax. *Sci. Reports 2013 31* **3**, 1–6 (2013).
41. Pornthanakasem, W. *et al*. Role of Plasmodium vivax dihydropteroate synthase polymorphisms in sulfa drug resistance. *Antimicrob. Agents Chemother.* **60**, 4453–4463 (2016).
42. Villena, F. E. *et al*. Molecular surveillance of the Plasmodium vivax multidrug resistance 1 gene in Peru between 2006 and 2015. *Malar. J.* **19**, 1–10 (2020).
43. Rodríguez, L. E. *et al*. Plasmodium vivax MSP-1 peptides have high specific binding activity to human reticulocytes. *Vaccine* **20**, 1331–1339 (2002).
44. Gupta, D. K. *et al*. The Plasmodium liver-specific protein 2 (LISP2) is an early marker of liver stage development. *Elife* **8**, (2019).
45. Miguel-Blanco, C. *et al*. The antimalarial efficacy and mechanism of resistance of the novel chemotype DDD01034957. *Sci. Rep.* **11**, (2021).
46. Cunningham, D., Lawton, J., Jarra, W., Preiser, P. & Langhorne, J. The pir multigene family of Plasmodium: Antigenic variation and beyond. *Mol. Biochem. Parasitol.* **170**, 65–73 (2010).
47. 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).
48. Prajapati, S. K. & Singh, O. P. Insights into the invasion biology of Plasmodium vivax. *Front. Cell. Infect. Microbiol.* **3**, (2013).
49. Remarque, E. J., Faber, B. W., Kocken, C. H. M. & Thomas, A. W. Apical membrane antigen 1: a

malaria vaccine candidate in review. *Trends Parasitol*. **24**, 74–84 (2008).

50. WAHLGREN, M. *et al*. Epitopes of the Plasmodium falciparum clustered-asparagine-rich protein (CARP) recognized by human T-cells and antibodies. *Parasite Immunol*. **13**, 681–694 (1991).
51. Arumugam, T. U. *et al*. Discovery of GAMA, a Plasmodium falciparum Merozoite Micronemal Protein, as a Novel Blood-Stage Vaccine Candidate Antigen. *Infect. Immun*. **79**, 4523 (2011).
52. Dahan-Pasternak, N. *et al*. PfSec13 is an unusual chromatin-associated nucleoporin of Plasmodium falciparum that is essential for parasite proliferation in human erythrocytes. *J. Cell Sci*. **126**, 3055–3069 (2013).
53. Zhang, M. *et al*. Inhibiting the Plasmodium eIF2α Kinase PK4 Prevents Artemisinin-Induced Latency. *Cell Host Microbe* **22**, 766-776.e4 (2017).
54. Han, J.-H. *et al*. Identification of a reticulocyte-specific binding domain of Plasmodium vivax reticulocyte-binding protein 1 that is homologous to the PfRh4 erythrocyte-binding domain OPEN. *Nat. Publ. Gr*. (2016) doi:10.1038/srep26993.
55. Tarr, S. J., Moon, R. W., Hardege, I. & Osborne, A. R. A conserved domain targets exported PHISTb family proteins to the periphery of Plasmodium infected erythrocytes. *Mol. Biochem. Parasitol*. **196**, 29 (2014).
56. Cheng, Y. *et al*. The Plasmodium vivax Merozoite Surface Protein 1 Paralog Is a Novel Erythrocyte-Binding Ligand of P. vivax. *Infect. Immun*. **81**, 1585 (2013).
57. Murithi, J. M. *et al*. The Plasmodium falciparum ABC transporter ABCI3 confers parasite strain-dependent pleiotropic antimalarial drug resistance. *Cell Chem. Biol*. (2021) doi:10.1016/J.CHEMBIOL.2021.06.006.
58. Reddy, K. S. *et al*. Multiprotein complex between the GPI-anchored CyRPA with PfRH5 and PfRipr is crucial for Plasmodium falciparum erythrocyte invasion. *Proc. Natl. Acad. Sci. U. S. A*. **112**, 1179–1184 (2015).
59. Benavente, E. D. *et al*. Genomic variation in Plasmodium vivax malaria reveals regions under selective pressure. *PLoS One* **12**, (2017).
60. Henrici, R. C., van Schalkwyk, D. A. & Sutherland, C. J. Modification of pfap2μ and pfubp1 Markedly Reduces Ring-Stage Susceptibility of Plasmodium falciparum to Artemisinin In Vitro. *Antimicrob. Agents Chemother*. **64**, (2019).
61. McNamara, C. W. *et al*. Targeting Plasmodium phosphatidylinositol 4-kinase to eliminate malaria Europe PMC Funders Group. *Nature* **504**, 248–253 (2013).
62. Llanos-Cuentas, A. *et al*. Antimalarial activity of single-dose DSM265, a novel plasmodium dihydroorotate dehydrogenase inhibitor, in patients with uncomplicated Plasmodium falciparum or Plasmodium vivax malaria infection: a proof-of-concept, open-label, phase 2a study. *Lancet*.

Infect. Dis. **18**, 874–883 (2018).

63. Kimata-Ariga, Y. & Morihisa, R. Functional analyses of plasmodium ferredoxin Asp97Tyr mutant related to artemisinin resistance of human malaria parasites. *J. Biochem.* **170**, 521–529 (2021).
64. Tran, P. N. *et al.* A female gametocyte-specific ABC transporter plays a role in lipid metabolism in the malaria parasite. *Nat. Commun. 2014 51* **5**, 1–13 (2014).
65. Veiga, M. I. *et al.* Novel Polymorphisms in Plasmodium falciparum ABC Transporter Genes Are Associated with Major ACT Antimalarial Drug Resistance. *PLoS One* **6**, e20212 (2011).
66. Gupta, B. *et al.* Plasmodium falciparum multidrug resistance protein 1 (pfmrp1) gene and its association with in vitro drug susceptibility of parasite isolates from north-east Myanmar. *J. Antimicrob. Chemother.* **69**, 2110–2117 (2014).
67. Dahlström, S., Veiga, M. I., Mårtensson, A., Björkman, A. & Gil, J. P. Polymorphism in PfMRP1 (Plasmodium falciparum multidrug resistance protein 1) amino acid 1466 associated with resistance to sulfadoxine-pyrimethamine treatment. *Antimicrob. Agents Chemother.* **53**, 2553–2556 (2009).
68. Rijpma, S. R. *et al.* MRP1 mediates folate transport and antifolate sensitivity in Plasmodium falciparum. *FEBS Lett.* **590**, 482–492 (2016).
69. Weiner, J. & Kooij, T. W. A. Phylogenetic profiles of all membrane transportproteins of the malaria parasite highlight new drug targets. *Microb. Cell* **3**, 511 (2016).
70. Shokoples, S. E., Ndao, M., Kowalewska-Grochowska, K. & Yanow, S. K. Multiplexed Real-Time PCR Assay for Discrimination of Plasmodium Species with Improved Sensitivity for Mixed Infections. *J. Clin. Microbiol.* **47**, 975–980 (2009).
71. 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**, e02257-16 (2017).
72. Ibrahim, A. *et al.* Selective whole genome amplification of Plasmodium malariae DNA from clinical samples reveals insights into population structure. *Sci. Rep.* **10**, (2020).
73. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
74. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–60 (2009).
75. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
76. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin).* **6**, 80–92 (2012).

77. Assefa, S. A. *et al*. estMOI: estimating multiplicity of infection using parasite deep sequencing data. *Bioinformatics* **30**, 1292–1294 (2014).
78. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
79. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
80. Danecek, P. *et al*. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
81. 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).
82. 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).
83. Tang, K., Thornton, K. R. & Stoneking, M. A New Approach for Using Genome Scans to Detect Recent Positive Selection in the Human Genome. *PLoS Biol.* **5**, e171 (2007).
84. 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).
85. Paradis, E. Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics* **26**, 419–420 (2010).

SUPPLEMENTARY INFORMATION

SUPPLEMENTARY TABLES

S1 Table. Summary of all isoaltes present within the filtered and unfiltered genomics database for *P. vivax*.

Region	Country	Pre QC (N=1113)	QC Pass (N= 885)	Monoclonal* (N=679)
South America	Brazil	181	123	107
	Colombia	66	34	31
	Guyana	4	3	0
	Mexico	40	20	18
	Panama	58	46	46
	Peru	138	89	75
	TOTAL	**487**	**315**	**277**
East Africa	Eritrea	19	13	10
	Ethiopia	55	53	39
	Madagascar	4	4	1
	Sudan	9	9	9
	Uganda	5	5	5
	TOTAL	**92**	**84**	**64**
South Asia	Afghanistan	31	27	25
	Bangladesh	1	1	0
	India	50	48	39
	Pakistan	38	37	32
	Sri Lanka	1	1	0
	TOTAL	**121**	**114**	**96**
South East Asia (SEA)	Cambodia	71	71	44
	China	12	12	10
	Laos	2	2	1
	Myanmar	36	28	22
	Thailand	167	160	102
	Vietnam	13	13	5
	TOTAL	**301**	**286**	**184**
Southern SEA (SSEA)	Indonesia	9	9	5
	Malaysia	50	50	39
	PNG	51	26	14
	Philippines	2	1	0
	TOTAL	**112**	**86**	**58**

*Monoclonal = $F_{WS} > 0.95$

S2 Table. Isolates from Brazil within the filtered genomics database (n = 123)

Sample	Accession number	State	Specific Site	Year	Date	Country	Fws	Coverage metric 5X*	Number of SNPs	Number of missing sites	Clade assignment	Group Assignment
ERR018932	ERR018932	Acre	Acre	2008	2008	Brazil	0.998	14.83	19082	495	C6	A
ERR019040	ERR019040	Acre	Acre	2008	2008	Brazil	0.903	49.05	19399	561	C6	A
ERR152407	ERR152407	Acre	Acre	2008	2008	Brazil	0.692	5.62	21294	2916	C6	A
PV03_com	ERR3813224	Acre	Cruzeiro do Sul	2015	2015	Brazil	0.995	62.02	19558	294	C2	A
PV04_com	ERR3813225	Acre	Cruzeiro do Sul	2015	2015	Brazil	0.975	22.81	19079	1212	C6	A
PV05_com	ERR3813226	Acre	Cruzeiro do Sul	2015	2015	Brazil	0.997	6.7	11798	10344	C6	A
PV06_com	ERR3813227	Acre	Cruzeiro do Sul	2015	2015	Brazil	0.994	57.7	19488	284	C3	A
PV10_com	ERR3813228	Acre	Cruzeiro do Sul	2015	2015	Brazil	0.994	14.38	17385	3440	C3	A
PV11_com	ERR3813229	Acre	Cruzeiro do Sul	2015	2015	Brazil	0.996	118.66	19557	178	C3	A
PV16_com	ERR3813230	Acre	Cruzeiro do Sul	2015	2015	Brazil	0.996	63.07	19631	161	C6	A
PV17_com	ERR3813231	Acre	Cruzeiro do Sul	2015	2015	Brazil	0.996	41.78	19570	217	C2	A
PV19_com	ERR3813232	Acre	Cruzeiro do Sul	2015	2015	Brazil	0.996	46.88	19596	173	C3	A
PV22_com	ERR3813233	Acre	Cruzeiro do Sul	2015	2015	Brazil	0.994	31.71	19844	204	C3	A
PV25_com	ERR3813234	Acre	Cruzeiro do Sul	2015	2015	Brazil	0.995	12.34	17670	2919	C6	A
PV26_com	ERR3813235	Acre	Cruzeiro do Sul	2015	2015	Brazil	0.995	54.15	19640	182	C6	A
PV29_com	ERR3813236	Acre	Cruzeiro do Sul	2015	2015	Brazil	0.977	114.83	19809	168	C3	A
PV31_com	ERR3813237	Acre	Cruzeiro do Sul	2015	2015	Brazil	0.994	37.88	19733	240	C1	A
PV32_com	ERR3813238	Acre	Cruzeiro do Sul	2015	2015	Brazil	0.998	55.08	19514	178	C4	A
PV34_com	ERR3813239	Acre	Cruzeiro do Sul	2015	2015	Brazil	0.995	11.39	17276	3775	C6	A
PV09_com	ERR5748169	Acre	Cruzeiro do Sul	2015	2015	Brazil	0.999	1.68	2331	28510	C1	A
PV14_com	ERR5748170	Acre	Cruzeiro do Sul	2015	2015	Brazil	0.999	1.96	2264	30754	C4	A
PVBRA37	ERR5748171	Rondonia	Porto Velho			Brazil	1	2.36	12022	12794	C6	A
SRR12134557	SRR12134557	Acre	Mancio Lima	2014	2014-12-05	Brazil	0.996	5.21	17944	3477	C3	A
SRR12134558	SRR12134558	Acre	Mancio Lima	2014	2014-06-30	Brazil	0.999	16.89	19154	495	C6	A
SRR12134559	SRR12134559	Acre	Mancio Lima	2014	2014-10-23	Brazil	0.996	50.49	19840	214	C1	A
SRR12134560	SRR12134560	Acre	Mancio Lima	2014	2014-06-24	Brazil	0.941	31.07	19509	205	C4	A
SRR12134561	SRR12134561	Acre	Mancio Lima	2015	2015-06-01	Brazil	0.994	154.64	19484	275	C1	A
SRR12134562	SRR12134562	Acre	Mancio Lima	2015	2015-04-07	Brazil	0.997	165.92	13808	6882	C1	A
SRR12134563	SRR12134563	Acre	Mancio Lima	2015	2015-04-06	Brazil	0.998	22.69	11838	13997	C1	A
SRR12134564	SRR12134564	Acre	Mancio Lima	2015	2015-03-02	Brazil	0.994	173.19	18307	2027	C1	A
SRR12134565	SRR12134565	Acre	Mancio Lima	2014	2014-06-09	Brazil	0.995	244.01	18498	1926	C1	A
SRR12134566	SRR12134566	Acre	Mancio Lima	2015	2015-05-13	Brazil	0.998	107.68	8912	11337	C1	A
SRR12134567	SRR12134567	Acre	Mancio Lima	2015	2015-01-30	Brazil	0.995	30.94	17636	3086	C1	A
SRR12134568	SRR12134568	Acre	Mancio Lima	2015	2015-02-11	Brazil	0.994	236.95	19614	198	C1	A
SRR12134569	SRR12134569	Acre	Mancio Lima	2014	2014-10-08	Brazil	0.995	30.2	19763	239	C1	A
SRR12134570	SRR12134570	Acre	Mancio Lima	2015	2015-01-22	Brazil	0.997	201.93	18034	2579	C2	A
SRR12134571	SRR12134571	Acre	Mancio Lima	2015	2015-01-21	Brazil	0.996	193.39	16936	3504	C2	A
SRR12134572	SRR12134572	Acre	Mancio Lima	2015	2015-01-16	Brazil	0.994	221.62	19796	194	C1	A
SRR12134573	SRR12134573	Acre	Mancio Lima	2014	2014-12-01	Brazil	0.994	230.51	19709	164	C1	A

SRR12134574	SRR12134574	Acre	Mancio Lima	2018	2018-10-11	Brazil	0.998	58.11	19373	444	C6	A
SRR12134575	SRR12134575	Acre	Mancio Lima	2018	2018-04-24	Brazil	0.999	75.56	19297	325	C1	A
SRR12134576	SRR12134576	Acre	Mancio Lima	2015	2015-01-05	Brazil	0.996	32.87	19554	254	C2	A
SRR12134577	SRR12134577	Acre	Mancio Lima	2016	2016-11-23	Brazil	0.998	2.17	16132	8988	C6	A
SRR12134578	SRR12134578	Acre	Mancio Lima	2016	2016-11-23	Brazil	0.997	45.07	19521	228	C3	A
SRR12134579	SRR12134579	Acre	Mancio Lima	2016	2016-08-31	Brazil	0.995	30.47	19572	230	C1	A
SRR12134580	SRR12134580	Acre	Mancio Lima	2014	2014-09-10	Brazil	0.999	39.49	19489	172	NA	A
SRR12134581	SRR12134581	Acre	Mancio Lima	2016	2016-08-31	Brazil	0.997	17.01	19481	316	C1	A
SRR12134582	SRR12134582	Acre	Mancio Lima	2016	2016-08-25	Brazil	0.993	29.52	19561	200	C2	A
SRR12134583	SRR12134583	Acre	Mancio Lima	2015	2015-06-01	Brazil	0.997	5.14	18462	2709	C1	A
SRR12134584	SRR12134584	Acre	Mancio Lima	2015	2015-03-05	Brazil	0.999	17.26	19374	506	C1	A
SRR12134585	SRR12134585	Acre	Mancio Lima	2014	2014-06-11	Brazil	0.995	78.86	19616	179	C3	A
SRR12134586	SRR12134586	Acre	Mancio Lima	2014	2014-11-26	Brazil	0.991	23.25	19559	304	C6	A
SRR12134587	SRR12134587	Acre	Mancio Lima	2014	2014-06-11	Brazil	0.996	10.53	19236	901	C2	A
SRR12134588	SRR12134588	Acre	Mancio Lima	2014	2014-06-11	Brazil	0.997	2.95	16746	6382	C1	A
SRR12134589	SRR12134589	Acre	Mancio Lima	2016	2016-11-30	Brazil	0.996	44.45	19623	190	C6	A
SRR12134590	SRR12134590	Acre	Mancio Lima	2015	2015-06-05	Brazil	0.995	29.48	19640	235	C1	A
SRR12134591	SRR12134591	Acre	Mancio Lima	2014	2014-08-25	Brazil	0.995	38.24	19627	263	C1	A
SRR12134592	SRR12134592	Acre	Mancio Lima	2014	2014-07-28	Brazil	0.997	13.05	19396	732	C1	A
SRR12134593	SRR12134593	Acre	Mancio Lima	2015	2015-02-12	Brazil	0.995	7.85	19062	1661	C1	A
SRR12134594	SRR12134594	Acre	Mancio Lima	2015	2015-02-03	Brazil	0.997	6.67	18299	2795	C1	A
SRR1562939	SRR1562939	Acre	Placido de Castro	2011	2011	Brazil	0.989	167.13	14241	16062	C3	A
SRR1568121	SRR1568121	Acre	Acrelandia	2011	2011	Brazil	0.764	19.63	19147	5564	NA	A
SRR1568192	SRR1568192	Acre	Placido de Castro	2011	2011	Brazil	0.994	123.83	18863	2697	C6	A
SRR1568193	SRR1568193	Acre	Placido de Castro	2011	2011	Brazil	0.95	32.13	9670	27601	C3	A
PV18	SRR5099324	Acre	Acrelandia	2012	2012	Brazil	0.997	12.31	17176	5687	C6	A
PV19	SRR5099325	Acre	Acrelandia	2012	2012	Brazil	0.998	16.58	17776	4278	C3	A
PV20	SRR5099326	Acre	Acrelandia	2013	2013	Brazil	0.998	13.6	17651	4547	C7	A
PV207	SRR5278291	Amazonas	Remansinho	2013	2013	Brazil	0.922	15.71	18319	4237	C6	A
PV32	SRR5278293	Acre	Acrelandia	2013	2013	Brazil	0.999	14.33	17679	4733	C6	A
PV51	SRR5278295	Acre	Acrelandia	2013	2013	Brazil	0.998	11.16	16878	6766	C6	A
PV52	SRR5278298	Acre	Acrelandia	2013	2013	Brazil	0.998	13.91	17649	4443	C7	A
PVACR	SRR5278301	Acre	Acrelandia	2015	2015	Brazil	0.998	15.1	17683	4678	C6	A
GNP922		Para	Goianésia do Pará	2012	2012	Brazil	0.824	10.13	19456	4067	C7	B
GNP923		Para	Goianésia do Pará	2012	2012	Brazil	0.998	10.92	17594	4024	C7	B
GNP972		Para	Goianésia do Pará	2012	2012	Brazil	0.999	11.23	16754	5332	C7	B
GNP977		Para	Goianésia do Pará	2012	2012	Brazil	0.936	8.18	12916	17191	C7	B
ITB_111		Para	Itaituba	2015	2015	Brazil	0.841	4.42	14732	13931	C7	B
ITB_119		Para	Itaituba	2015	2015	Brazil	0.999	9	17528	4125	C7	B
ITB050		Para	Itaituba	2015	2015	Brazil	0.857	3.13	12661	18327	C7	B

ITB107	Para	Itaituba	2015	2015	Brazil	0.978	8.44	17227	5024	C7	B
ITB112	Para	Itaituba	2015	2015	Brazil	0.705	7.35	19024	7004	C7	B
ITB118	Para	Itaituba	2015	2015	Brazil	0.998	7.8	16553	6145	C7	B
MCP148	Amapa	Macapa	2003	2003	Brazil	0.902	1.83	11766	19415	C7	B
MT66	Mato Grosso	Mato Grosso			Brazil	0.924	9.79	15894	9550	C6	A
MT68	Mato Grosso	Mato Grosso			Brazil	0.997	7.48	16069	7234	C6	A
MT69	Mato Grosso	Mato Grosso			Brazil	0.998	8.51	16394	6903	C7	A
MT70	Mato Grosso	Mato Grosso			Brazil	0.998	8.03	13001	15519	C4	A
MT78	Mato Grosso	Mato Grosso			Brazil	0.998	8.76	8853	30235	C7	A
NR347	Para	Novo Repartimento	2003	2003	Brazil	0.998	9.58	17243	4695	C7	B
NR373	Para	Novo Repartimento	2003	2003	Brazil	0.998	7.45	16808	5632	C7	B
NR395	Para	Novo Repartimento	2003	2003	Brazil	0.998	8.49	16968	5669	C7	B
OIP115	Amapa	Oiapoque	2015	2015	Brazil	0.999	12.48	17369	4682	C7	B
OIP122	Amapa	Oiapoque	2015	2015	Brazil	0.999	4.59	10875	22850	C7	B
OIP144	Amapa	Oiapoque	2015	2015	Brazil	0.998	9.93	13890	14102	C7	B
OIP164	Amapa	Oiapoque	2015	2015	Brazil	0.998	5.88	15800	8319	C7	B
OIP316	Amapa	Oiapoque	2015	2015	Brazil	0.998	12.09	17347	4690	C7	B
OIP386	Amapa	Oiapoque	2015	2015	Brazil	0.811	6.06	17817	7875	C7	B
OIP395	Amapa	Oiapoque	2015	2015	Brazil	0.999	7.97	15953	7517	C7	B
OIP398	Amapa	Oiapoque	2015	2015	Brazil	0.998	14.94	18047	2826	C7	B
OIP861	Amapa	Oiapoque	2015	2015	Brazil	0.998	10.11	16546	6798	C7	B
PLV851	Rondonia	Porto Velho	2004	2004	Brazil	0.999	8.37	17039	5283	C2	A
PV_BRA_B02	Acre	Cruzeiro do Sul	2015	2015	Brazil	0.996	25.36	17858	4836	C1	A
PV_BRA_B10	Acre	Cruzeiro do Sul	2015	2015	Brazil	0.997	22.13	16449	8334	C1	A
PV_BRA_C02	Acre	Cruzeiro do Sul	2015	2015	Brazil	0.998	22.57	17092	6639	C1	A
PV_BRA_C10	Acre	Cruzeiro do Sul	2015	2015	Brazil	0.998	14.79	15331	11144	C1	A
PVB_01._13	Sao Paulo	Riacho Grande	2013	2013	Brazil	0.999	3.03	10136	24753	C5	/
PVB_01.15	Sao Paulo	Caraguatatuba	2015	2015	Brazil	0.998	13.57	17547	4971	C5	/
PVB_01.19	Sao Paulo	Paraibuna	2019	2019	Brazil	0.998	11.82	16478	7423	C5	/
PVB_04.14	Sao Paulo	Juquitiba	2014	2014	Brazil	0.998	9.63	17207	5684	C5	/
PVB_04.16	Sao Paulo	Peruibe	2016	2016	Brazil	0.999	6.33	15140	9478	C5	/
PVB_05.17	Sao Paulo	Registro	2017	2017	Brazil	0.998	4.56	12179	18355	C5	/
PVB_08.19	Sao Paulo	Iporanga	2019	2019	Brazil	0.999	2.87	10805	22459	C5	/
PVB_14_18	Sao Paulo	Salesopolis	2018	2018	Brazil	0.999	4.64	11942	19069	C5	/
PVB_24	Sao Paulo	Juquitiba			Brazil	0.998	20.41	18585	2760	C5	/
PVB_26	Sao Paulo	Maresias			Brazil	0.999	4.93	13522	14152	C5	/
PVB_3.19	Sao Paulo	Miracatu	2019	2019	Brazil	0.999	4.37	12422	17897	C5	/
PvB_4.18	Sao Paulo	Biritiba Mirim	2018	2018	Brazil	0.998	9.05	17083	4821	C5	/
PVL532	Rondonia	Porto Velho	2004	2004	Brazil	0.999	9.69	9496	28426	C7	A
PVL556	Rondonia	Porto Velho	2004	2004	Brazil	0.689	8.37	16112	11748	C6	A
PVL860	Rondonia	Porto Velho	2004	2004	Brazil	0.982	5.23	14131	12986	C7	A
Sim_A5	Amazonas	Barcelos	2010	2010	Brazil	0.854	11.47	18086	7369	C4	A
Sim_B4	Amazonas	Barcelos	2010	2010	Brazil	0.998	18.94	18078	2882	C4	A
Sim_B5	Amazonas	Barcelos	2010	2010	Brazil	0.998	12.57	17550	4124	C4	A

*Coverage metric calculated as the mean coverage of all positions with a minimum of 5 reads, multiplied by the fraction of the genome that matched the threshold of 5 reads

S3 Table. Regional Fst comparisons against South American isolates (n = 303)

Comparison	Chromosome	Position	Ref	Alt	Effect	Gene name	Amino acid change	Nucleotide change	Fst
South America V Southern SEA	PvP01_01_v1	395494	G	C	synonymous	PVP01_0107900	268G	395494G>C	1
South America V Southern SEA	PvP01_01_v1	532879	A	T	missense	PVP01_0111500	170F>170I	532879A>T	1
South America V Southern SEA	PvP01_01_v1	533152	C	A	missense	PVP01_0111500	79V>79L	533152C>A	1
South America V Southern SEA	PvP01_02_v1	327451	G	A	missense	PVP01_0207700	95V>95I	327451G>A	1
South America V Southern SEA	PvP01_03_v1	486502	T	C	missense	PVP01_0311100	972K>972R	>C | 486501T>C+48	1
South America V Southern SEA	PvP01_03_v1	689321	C	G	missense	PVP01_0315600	233Q>233E	689321C>G	1
South America V Southern SEA	PvP01_03_v1	733038	C	T	.	NA	NA	NA	1
South America V Southern SEA	PvP01_03_v1	733040	G	T	.	NA	NA	NA	1
South America V Southern SEA	PvP01_04_v1	413477	T	C	missense	PVP01_0410100	244N>244S	413477T>C	1
South America V Southern SEA	PvP01_04_v1	465641	G	A	intron	PVP01_0411300	NA	NA	1
South America V Southern SEA	PvP01_04_v1	493044	A	G	synonymous	PVP01_0412400	3106Y	493044A>G	1
South America V Southern SEA	PvP01_04_v1	534787	A	G	synonymous	AsnRS	375D	534787A>G	1
South America V Southern SEA	PvP01_04_v1	601916	T	C	*missense	PVP01_0414700	2087L>2087P	601916T>C	1
South America V Southern SEA	PvP01_05_v1	574992	A	T	synonymous	PVP01_0513700	24I	574992A>T	1
South America V Southern SEA	PvP01_05_v1	574998	T	C	synonymous	PVP01_0513700	22K	574998T>C	1
South America V Southern SEA	PvP01_05_v1	928953	T	G	missense	PVP01_0523000	66K>66Q	928953T>G	1
South America V Southern SEA	PvP01_05_v1	988118	A	C	missense	PVP01_0524300	299Q>299P	988118A>C	1
South America V Southern SEA	PvP01_05_v1	1093207	G	C	synonymous	PVP01_0526800	2498L	1093207G>C	1
South America V Southern SEA	PvP01_05_v1	1146235	G	A	.	NA	NA	NA	1
South America V Southern SEA	PvP01_06_v1	203673	A	G	synonymous	PVP01_0604900	299Y	203673A>G	1
South America V Southern SEA	PvP01_06_v1	237667	A	C	missense	PVP01_0605900	591N>591H	237667A>C	1
South America V Southern SEA	PvP01_06_v1	254563	G	C	missense	PVP01_0606300	586S>586R	4563G>C+254564A	1
South America V Southern SEA	PvP01_06_v1	254564	A	G	missense	PVP01_0606300	586S>586R	4563G>C+254564A	1
South America V Southern SEA	PvP01_06_v1	260817	G	C	.	NA	NA	NA	1
South America V Southern SEA	PvP01_06_v1	285512	A	G	missense	PVP01_0606900	2077D>2077G	5512A>G+285513T	1
South America V Southern SEA	PvP01_06_v1	285512	A	G	synonymous	PVP01_0606900	2077D	285512A>G	1
South America V Southern SEA	PvP01_06_v1	286506	G	C	missense	PVP01_0606900	1746T>1746R	286506G>C	1
South America V Southern SEA	PvP01_06_v1	531114	A	G	synonymous	PVP01_0612100	1084Y	531114A>G	1
South America V Southern SEA	PvP01_06_v1	581263	C	T	.	NA	NA	NA	1
South America V Southern SEA	PvP01_06_v1	588969	A	C	*missense	MTRAP	288R>288S	588969A>C	1
South America V Southern SEA	PvP01_06_v1	759043	G	T	.	NA	NA	NA	1
South America V Southern SEA	PvP01_07_v1	550357	G	A	synonymous	GCD10	463N	550357G>A	1
South America V Southern SEA	PvP01_07_v1	608118	C	T	missense	PVP01_0712900	974G>974S	608118C>T	1
South America V Southern SEA	PvP01_07_v1	1182121	A	G	missense	RhopH2	656I>656V	1182121A>G	1
South America V Southern SEA	PvP01_08_v1	91006	C	T	*synonymous	PVP01_0801700	227A	91006C>T	1
South America V Southern SEA	PvP01_08_v1	115456	C	A	.	NA	NA	NA	1
South America V Southern SEA	PvP01_08_v1	336786	T	C	synonymous	HDA2	757P	336786T>C	1
South America V Southern SEA	PvP01_08_v1	450446	G	A	intron	PVP01_0810000	NA	NA	1
South America V Southern SEA	PvP01_08_v1	914227	C	G	missense	PVP01_0820800	488S>488T	914227C>G	1
South America V Southern SEA	PvP01_08_v1	926690	G	T	synonymous	ATPase7	964L	926690G>T	1
South America V Southern SEA	PvP01_08_v1	1011735	A	T	synonymous	PVP01_0822700	331V	1011735A>T	1
South America V Southern SEA	PvP01_08_v1	1074206	C	T	.	NA	NA	NA	1
South America V Southern SEA	PvP01_08_v1	1504842	G	A	missense	PVP01_0835500	208Q>208L	4841A>T+1504842	1
South America V Southern SEA	PvP01_08_v1	1504842	G	A	synonymous	PVP01_0835500	208Q	1504842G>A	1
South America V Southern SEA	PvP01_08_v1	1504843	A	T	missense	PVP01_0835500	209I>209F	1504843A>T	1
South America V Southern SEA	PvP01_09_v1	821669	A	G	intron	PVP01_0918500	NA	NA	1
South America V Southern SEA	PvP01_09_v1	875126	C	G	synonymous	PVP01_0920100	86V	875126C>G	1

South America V Southern SEA	PvP01_09_v1	892891	C	G	synonymous	PVP01_0920500	1065S	892891C>G	1
South America V Southern SEA	PvP01_09_v1	1058053	A	T	.	NA	NA	NA	1
South America V Southern SEA	PvP01_09_v1	1159872	A	C	.	NA	NA	NA	1
South America V Southern SEA	PvP01_09_v1	1198207	T	C	.	NA	NA	NA	1
South America V Southern SEA	PvP01_09_v1	1199717	T	G	missense	PVP01_0927300	496K>496N	1199717T>G	1
South America V Southern SEA	PvP01_09_v1	1202587	A	G	.	NA	NA	NA	1
South America V Southern SEA	PvP01_09_v1	1245258	C	G	*missense	PVP01_0928500	640M>640I	1245258C>G	1
South America V Southern SEA	PvP01_09_v1	1369676	T	C	synonymous	PVP01_0932000	210F	1369676T>C	1
South America V Southern SEA	PvP01_09_v1	1476981	A	C	missense	PVP01_0934700	322N>322H	1476981A>C	1
South America V Southern SEA	PvP01_10_v1	481042	T	C	missense	MDR1	698S>698G	481042T>C	1
South America V Southern SEA	PvP01_10_v1	654836	A	G	synonymous	CCR4	1938E	654836A>G	1
South America V Southern SEA	PvP01_10_v1	685899	T	C	synonymous	PUF1	959G	685899T>C	1
South America V Southern SEA	PvP01_10_v1	1097744	C	T	missense	PVP01_1025100	649V>649I	1097744C>T	1
South America V Southern SEA	PvP01_10_v1	1222882	G	C	missense	PVP01_1028600	398V>398L	1222882G>C	1
South America V Southern SEA	PvP01_10_v1	1226203	T	C	.	NA	NA	NA	1
South America V Southern SEA	PvP01_11_v1	527713	A	C	.	NA	NA	NA	1
South America V Southern SEA	PvP01_11_v1	827140	T	G	intron	PVP01_1119500	NA	NA	1
South America V Southern SEA	PvP01_11_v1	1103201	T	A	.	NA	NA	NA	1
South America V Southern SEA	PvP01_11_v1	1190787	G	C	intron	PVP01_1126900	NA	NA	1
South America V Southern SEA	PvP01_11_v1	1256157	A	G	synonymous	MSP10	145A	1256157A>G	1
South America V Southern SEA	PvP01_11_v1	1261039	T	G	.	NA	NA	NA	1
South America V Southern SEA	PvP01_11_v1	1262951	G	A	missense	PVP01_1129400	20P>20S	1262951G>A	1
South America V Southern SEA	PvP01_11_v1	1265741	A	T	missense	PVP01_1129500	236N>236F	5740A>T+1265741	1
South America V Southern SEA	PvP01_11_v1	1265741	A	T	missense	PVP01_1129500	236N>236I	1265741A>T	1
South America V Southern SEA	PvP01_11_v1	1470109	A	G	synonymous	PVP01_1134400	190N	1470109A>G	1
South America V Southern SEA	PvP01_11_v1	1480435	C	G	.	NA	NA	NA	1
South America V Southern SEA	PvP01_11_v1	1695828	C	T	synonymous	PVP01_1139900	3192L	1695828C>T	1
South America V Southern SEA	PvP01_11_v1	1742207	G	C	missense	PVP01_1141100	388Q>388E	1742207G>C	1
South America V Southern SEA	PvP01_11_v1	2011410	C	G	missense	PVP01_1147100	19K>19N	2011410C>G	1
South America V Southern SEA	PvP01_12_v1	172548	T	G	missense	PVP01_1204200	96T>96P	172548T>G	1
South America V Southern SEA	PvP01_12_v1	192115	T	G	.	NA	NA	NA	1
South America V Southern SEA	PvP01_12_v1	242863	C	T	missense	PVP01_1206200	46A>46T	242863C>T	1
South America V Southern SEA	PvP01_12_v1	338808	G	A	synonymous	PVP01_1208400	3664C	338808G>A	1
South America V Southern SEA	PvP01_12_v1	362687	G	A	.	NA	NA	NA	1
South America V Southern SEA	PvP01_12_v1	429957	C	A	.	NA	NA	NA	1
South America V Southern SEA	PvP01_12_v1	683696	G	A	synonymous	PVP01_1216400	326L	683696G>A	1
South America V Southern SEA	PvP01_12_v1	1190440	G	A	synonymous	YKT6.2	198D	1190440G>A	1
South America V Southern SEA	PvP01_12_v1	1203357	C	T	.	NA	NA	NA	1
South America V Southern SEA	PvP01_12_v1	1374068	T	A	.	NA	NA	NA	1
South America V Southern SEA	PvP01_12_v1	1874995	T	G	missense	PVP01_1245400	274N>274K	1874995T>G	1
South America V Southern SEA	PvP01_12_v1	1937676	T	C	synonymous	PVP01_1247500	109I	1937676T>C	1
South America V Southern SEA	PvP01_12_v1	2358861	G	T	missense	ApiAP2	431Q>431K	2358861G>T	1
South America V Southern SEA	PvP01_12_v1	2463736	G	C	intron	PVP01_1259700	NA	NA	1
South America V Southern SEA	PvP01_13_v1	337753	A	C	missense	CRMP3	1719K>1719N	337753A>C	1
South America V Southern SEA	PvP01_13_v1	756832	G	C	missense	PVP01_1316100	2648L>2648F	756832G>C	1
South America V Southern SEA	PvP01_13_v1	792891	T	C	synonymous	PVP01_1316900	780D	792891T>C	1
South America V Southern SEA	PvP01_13_v1	792893	A	G	missense	PVP01_1316900	781K>781R	792893A>G	1
South America V Southern SEA	PvP01_13_v1	1039560	A	C	synonymous	PVP01_1323000	497L	1039560A>C	1
South America V Southern SEA	PvP01_13_v1	1045389	C	G	missense	PVP01_1323200	96L>96V	1045389C>G	1
South America V Southern SEA	PvP01_13_v1	1433470	C	G	missense	PVP01_1333300	824S>824C	1433470C>G	1

South America V Southern SEA	PvP01_14_v1	446765	T	C	*missense	PVP01_1409500	109K>109E	446765T>C	1
South America V Southern SEA	PvP01_14_v1	556784	C	G	*missense	PVP01_1412600	343I>343M	556784C>G	1
South America V Southern SEA	PvP01_14_v1	762736	G	A	.	NA	NA	NA	1
South America V Southern SEA	PvP01_14_v1	1169796	G	A	.	NA	NA	NA	1
South America V Southern SEA	PvP01_14_v1	1226618	T	G	missense	PVP01_1428700	180L>180V	1226618T>G	1
South America V Southern SEA	PvP01_14_v1	1313512	A	G	intron	JmjC1	NA	NA	1
South America V Southern SEA	PvP01_14_v1	1451192	G	T	.	NA	NA	NA	1
South America V Southern SEA	PvP01_14_v1	1451245	C	T	.	NA	NA	NA	1
South America V Southern SEA	PvP01_14_v1	1847223	G	C	.	NA	NA	NA	1
South America V Southern SEA	PvP01_14_v1	2065363	C	A	missense	PVP01_1447400	608T>608K	2065363C>A	1
South America V Southern SEA	PvP01_14_v1	2065363	C	A	missense	PVP01_1447400	608T>608Q	5362A>C+2065363	1
South America V Southern SEA	PvP01_14_v1	2206669	G	A	synonymous	PVP01_1450700	805F	2206669G>A	1
South America V Southern SEA	PvP01_14_v1	2879231	A	G	missense	VPS26	274I>274V	2879231A>G	1
South America V Southern SEA	PvP01_06_v1	589259	G	C	*missense	MTRAP	385G>385A	589259G>C	0.997391
South America V Southern SEA	PvP01_12_v1	315394	A	G	missense	TOP3	136V>136A	315394A>G	0.997367
South America V Southern SEA	PvP01_03_v1	552932	A	C	synonymous	PVP01_0312700	28I	552932A>C	0.997364
South America V Southern SEA	PvP01_09_v1	1071275	C	A	missense	PVP01_0923900	177Q>177K	1071275C>A	0.997353
South America V Southern SEA	PvP01_12_v1	2148245	C	G	missense	APP	395L>395F	2148245C>G	0.997339
South America V Southern SEA	PvP01_12_v1	2447491	C	T	.	NA	NA	NA	0.997339
South America V Southern SEA	PvP01_14_v1	2862134	C	G	missense	G377	1958L>1958F	2862134C>G	0.997332
South America V Southern SEA	PvP01_05_v1	1278185	A	T	synonymous	PVP01_0530200	180S	1278185A>T	0.997325
South America V Southern SEA	PvP01_12_v1	327858	G	A	missense	P48/45	418R>418K	327858G>A	0.997325
South America V Southern SEA	PvP01_14_v1	2862026	C	G	missense	G377	1994A>1994G	2026C>G+2862027	0.997311
South America V Southern SEA	PvP01_14_v1	2862026	C	G	synonymous	G377	1994A	2862026C>G	0.997311
South America V Southern SEA	PvP01_05_v1	1133836	C	A	*missense	PVP01_0527400	4351T>4351K	1133836C>A	0.99731
South America V Southern SEA	PvP01_07_v1	608178	C	T	missense	PVP01_0712900	954G>954R	608178C>T	0.99731
South America V Southern SEA	PvP01_12_v1	323611	A	G	synonymous	P47	26T	323611A>G	0.997307
South America V Southern SEA	PvP01_14_v1	1702548	T	C	.	NA	NA	NA	0.997307
South America V Southern SEA	PvP01_07_v1	627506	A	C	*missense	NDH2	186D>186E	627506A>C	0.997303
South America V Southern SEA	PvP01_13_v1	1197954	G	A	synonymous	P113	166F	1197954G>A	0.997293
South America V Southern SEA	PvP01_07_v1	1056218	G	T	missense	PVP01_0724600	953L>953I	1056218G>T	0.997282
South America V Southern SEA	PvP01_09_v1	1239132	G	T	missense	PVP01_0928300	223A>223S	1239132G>T	0.997267
South America V Southern SEA	PvP01_09_v1	1063002	T	C	synonymous	CDPK7	1443D	1063002T>C	0.997259
South America V Southern SEA	PvP01_04_v1	387834	T	C	.	NA	NA	NA	0.997255
South America V Southern SEA	PvP01_04_v1	387840	C	A	.	NA	NA	NA	0.997255
South America V Southern SEA	PvP01_08_v1	1372891	T	G	missense	PVP01_0832100	1408K>1408T	1372891T>G	0.997255
South America V Southern SEA	PvP01_03_v1	454900	T	G	missense	RbgA	111K>111Q	454900T>G	0.997252
South America V Southern SEA	PvP01_08_v1	924361	T	C	.	NA	NA	NA	0.997248
South America V Southern SEA	PvP01_12_v1	671184	G	T	*synonymous	PVP01_1216000	456G	671184G>T	0.997248
South America V Southern SEA	PvP01_02_v1	290653	T	G	synonymous	PVP01_0207000	157S	290653T>G	0.997244
South America V Southern SEA	PvP01_05_v1	1071391	A	T	missense	PVP01_0526400	285I>285F	1071391A>T	0.997244
South America V Southern SEA	PvP01_09_v1	815615	C	G	missense	MLH	418V>418L	815615C>G	0.997244
South America V Southern SEA	PvP01_13_v1	1034884	G	A	missense	ABCG2	296V>296I	1034884G>A	0.997244
South America V Southern SEA	PvP01_07_v1	832234	T	A	missense	PVP01_0718800	1530N>1530K	832234T>A	0.99724
South America V Southern SEA	PvP01_08_v1	616600	A	T	missense	PVP01_0814200	904N>904I	616600A>T	0.99724
South America V Southern SEA	PvP01_07_v1	994042	G	C	synonymous	PVP01_0723000	537L	994042G>C	0.997237
South America V Southern SEA	PvP01_05_v1	903893	A	C	.	NA	NA	NA	0.997225
South America V Southern SEA	PvP01_14_v1	2029491	A	C	missense	PVP01_1446900	788E>788D	2029491A>C	0.997221
South America V Southern SEA	PvP01_07_v1	537999	T	C	.	NA	NA	NA	0.997209
South America V Southern SEA	PvP01_11_v1	1302272	A	G	missense	PVP01_1130200	2676D>2676G	1302272A>G	0.997206

South America V Southern SEA	PvP01_08_v1	1528403	G	C	.	NA	NA	NA	0.99719
South America V Southern SEA	PvP01_04_v1	435299	C	T	missense	PVP01_0410600	853R>853K	435299C>T	0.997175
South America V Southern SEA	PvP01_08_v1	1468976	G	C	missense	PVP01_0834700	3493R>3493T	1468976G>C	0.997162
South America V South Asia	PvP01_05_v1	574992	A	T	synonymous	PVP01_0513700	24I	574992A>T	1
South America V South Asia	PvP01_06_v1	670773	C	G	missense	PVP01_0616100	97E>97Q	670773C>G	1
South America V South Asia	PvP01_07_v1	1198184	T	C	missense	PVP01_0728500	609F>609L	1198184T>C	1
South America V South Asia	PvP01_09_v1	1198207	T	C	.	NA	NA	NA	1
South America V South Asia	PvP01_11_v1	1256157	A	G	synonymous	MSP10	145A	1256157A>G	1
South America V South Asia	PvP01_11_v1	1261039	T	G	.	NA	NA	NA	1
South America V South Asia	PvP01_11_v1	1261998	T	A	missense	PVP01_1129300	243L>243Q	1261998T>A	1
South America V South Asia	PvP01_11_v1	1265740	A	T	missense	PVP01_1129500	236N>236F	5740A>T+1265741	1
South America V South Asia	PvP01_11_v1	1265740	A	T	missense	PVP01_1129500	236N>236Y	1265740A>T	1
South America V South Asia	PvP01_11_v1	1265741	A	T	missense	PVP01_1129500	236N>236F	5740A>T+1265741	1
South America V South Asia	PvP01_11_v1	1265741	A	T	missense	PVP01_1129500	236N>236I	1265741A>T	1
South America V South Asia	PvP01_11_v1	1266207	T	G	synonymous	PVP01_1129500	391T	1266207T>G	1
South America V South Asia	PvP01_12_v1	324232	G	T	missense	P47	233M>233I	324232G>T	1
South America V South Asia	PvP01_12_v1	327608	T	G	missense	P48/45	335Y>335D	327608T>G	1
South America V South Asia	PvP01_12_v1	327731	A	G	missense	P48/45	376T>376A	327731A>G	1
South America V South Asia	PvP01_12_v1	327858	G	A	missense	P48/45	418R>418K	327858G>A	1
South America V South Asia	PvP01_13_v1	337753	A	C	missense	CRMP3	1719K>1719N	337753A>C	1
South America V South Asia	PvP01_14_v1	1313512	A	G	intron	JmjC1	NA	NA	1
South America V South Asia	PvP01_12_v1	321295	G	A	.	NA	NA	NA	0.997496
South America V South Asia	PvP01_12_v1	323603	C	T	missense	P47	24L>24F	323603C>T	0.997496
South America V South Asia	PvP01_12_v1	323611	A	G	synonymous	P47	26T	323611A>G	0.997496
South America V South Asia	PvP01_09_v1	1730454	G	C	.	NA	NA	NA	0.997464
South America V South Asia	PvP01_03_v1	553010	C	G	missense	PVP01_0312700	54F>54L	553010C>G	0.995069
South America V South Asia	PvP01_06_v1	668360	G	T	missense	P28	65T>65K	668360G>T	0.995064
South America V South Asia	PvP01_06_v1	712994	G	A	missense	DCP1	337G>337E	712994G>A	0.995042
South America V South Asia	PvP01_12_v1	2475598	C	T	missense	PVP01_1260100	372L>372F	2475598C>T	0.994966
South America V South Asia	PvP01_12_v1	2127660	C	T	.	NA	NA	NA	0.9948
South America V South Asia	PvP01_12_v1	2669791	G	C	missense	TREP	2131R>2131G	2669791G>C	0.994786
South America V South Asia	PvP01_09_v1	1729971	G	C	.	NA	NA	NA	0.994759
South America V South Asia	PvP01_14_v1	1305248	A	G	start_lost	PVP01_1430300	NA	NA	0.99474
South America V South Asia	PvP01_12_v1	909364	C	A	.	NA	NA	NA	0.994661
South America V South Asia	PvP01_09_v1	1729481	A	G	.	NA	NA	NA	0.994618
South America V South Asia	PvP01_08_v1	1546424	C	G	missense	PVP01_0836700	489E>489Q	1546424C>G	0.994514
South America V South Asia	PvP01_14_v1	1314655	T	G	intron	JmjC1	NA	NA	0.994211
South America V South Asia	PvP01_11_v1	1264740	G	T	.	NA	NA	NA	0.99409
South America V South Asia	PvP01_06_v1	669053	A	G	.	NA	NA	NA	0.993895
South America V South Asia	PvP01_14_v1	1314524	C	G	intron	JmjC1	NA	NA	0.993858
South America V South Asia	PvP01_09_v1	1728961	C	T	.	NA	NA	NA	0.993045
South America V South Asia	PvP01_12_v1	770010	A	C	synonymous	TRAP	350A	770010A>C	0.99256
South America V South Asia	PvP01_08_v1	1106577	G	T	.	NA	NA	NA	0.992028
South America V South Asia	PvP01_09_v1	1727777	A	G	.	NA	NA	NA	0.991923
South America V South Asia	PvP01_11_v1	579666	A	G	.	NA	NA	NA	0.991727
South America V South Asia	PvP01_09_v1	1728276	G	T	.	NA	NA	NA	0.991702
South America V South Asia	PvP01_11_v1	1266124	G	T	missense	PVP01_1129500	364G>364W	1266124G>T	0.990087
South America V Southeast Asia	PvP01_05_v1	574992	A	T	synonymous	PVP01_0513700	24I	574992A>T	0.998267
South America V Southeast Asia	PvP01_12_v1	327858	G	A	missense	P48/45	418R>418K	327858G>A	0.998264
South America V Southeast Asia	PvP01_09_v1	1369676	T	C	synonymous	PVP01_0932000	210F	1369676T>C	0.998258

South America V Southeast Asia	PvP01_08_v1	914227	C	G	missense	PVP01_0820800	488S>488T	914227C>G	0.99823
South America V Southeast Asia	PvP01_11_v1	1262951	G	A	missense	PVP01_1129400	20P>20S	1262951G>A	0.998224
South America V Southeast Asia	PvP01_14_v1	556784	C	G	*missense	PVP01_1412600	343I>343M	556784C>G	0.998221
South America V Southeast Asia	PvP01_11_v1	1190787	G	C	intron	PVP01_1126900	NA	NA	0.998201
South America V Southeast Asia	PvP01_11_v1	1448703	G	C	synonymous	PVP01_1133700	1616L	1448703G>C	0.998182
South America V Southeast Asia	PvP01_08_v1	1546424	C	G	missense	PVP01_0836700	489E>489Q	1546424C>G	0.998148
South America V Southeast Asia	PvP01_11_v1	1265741	A	T	missense	PVP01_1129500	236N>236F	5740A>T+1265741	0.998148
South America V Southeast Asia	PvP01_11_v1	1265741	A	T	missense	PVP01_1129500	236N>236I	1265741A>T	0.998148
South America V Southeast Asia	PvP01_06_v1	531114	A	G	synonymous	PVP01_0612100	1084Y	531114A>G	0.998113
South America V Southeast Asia	PvP01_14_v1	1313512	A	G	intron	JmjC1	NA	NA	0.998081
South America V Southeast Asia	PvP01_13_v1	337753	A	C	missense	CRMP3	1719K>1719N	337753A>C	0.998047
South America V Southeast Asia	PvP01_12_v1	323603	C	T	missense	P47	24L>24F	323603C>T	0.996485
South America V Southeast Asia	PvP01_12_v1	323611	A	G	synonymous	P47	26T	323611A>G	0.996479
South America V Southeast Asia	PvP01_14_v1	762736	G	A	.	NA	NA	NA	0.996473
South America V Southeast Asia	PvP01_12_v1	321295	G	A	.	NA	NA	NA	0.996466
South America V Southeast Asia	PvP01_06_v1	531393	C	G	missense	PVP01_0612100	991K>991N	531393C>G	0.996409
South America V Southeast Asia	PvP01_12_v1	338808	G	A	synonymous	PVP01_1208400	3664C	338808G>A	0.996397
South America V Southeast Asia	PvP01_11_v1	1261039	T	G	.	NA	NA	NA	0.99631
South America V Southeast Asia	PvP01_03_v1	553010	C	G	missense	PVP01_0312700	54F>54L	553010C>G	0.994789
South America V Southeast Asia	PvP01_14_v1	2662867	T	A	missense	PVP01_1461600	403I>403L	2662867T>A	0.994652
South America V Southeast Asia	PvP01_12_v1	2127660	C	T	.	NA	NA	NA	0.994536
South America V Southeast Asia	PvP01_09_v1	1729971	G	C	.	NA	NA	NA	0.994485
South America V Southeast Asia	PvP01_11_v1	1256157	A	G	synonymous	MSP10	145A	1256157A>G	0.994479
South America V Southeast Asia	PvP01_09_v1	1728961	C	T	.	NA	NA	NA	0.99349
South America V Southeast Asia	PvP01_12_v1	770010	A	C	synonymous	TRAP	350A	770010A>C	0.99302
South America V Southeast Asia	PvP01_07_v1	747983	G	A	synonymous	PVP01_0716800	574K	747983G>A	0.992933
South America V Southeast Asia	PvP01_07_v1	747984	G	A	missense	PVP01_0716800	575G>575S	747984G>A	0.992933
South America V Southeast Asia	PvP01_08_v1	606745	G	A	missense	PVP01_0814000	69A>69V	606745G>A	0.992799
South America V Southeast Asia	PvP01_10_v1	481042	T	C	missense	MDR1	698S>698G	481042T>C	0.99256
South America V Southeast Asia	PvP01_06_v1	669053	A	G	.	NA	NA	NA	0.991992
South America V Southeast Asia	PvP01_14_v1	2970508	A	G	missense	PVP01_1469500	832C>832R	2970508A>G	0.991935
South America V Southeast Asia	PvP01_14_v1	2662513	G	T	synonymous	PVP01_1461600	521R	2662513G>T	0.99116
South America V Southeast Asia	PvP01_02_v1	211461	C	A	intron	PIP5K	NA	NA	0.990686
South America V Southeast Asia	PvP01_04_v1	601916	T	C	*missense	PVP01_0414700	2087L>2087P	601916T>C	0.990592
South America V East Africa	PvP01_05_v1	574992	A	T	synonymous	PVP01_0513700	24I	574992A>T	1
South America V East Africa	PvP01_05_v1	599497	A	G	synonymous	SRCAP	618P	599497A>G	1
South America V East Africa	PvP01_07_v1	510043	C	T	.	NA	NA	NA	1
South America V East Africa	PvP01_11_v1	1261039	T	G	.	NA	NA	NA	1
South America V East Africa	PvP01_11_v1	1262951	G	A	missense	PVP01_1129400	20P>20S	1262951G>A	1
South America V East Africa	PvP01_11_v1	1448703	G	C	synonymous	PVP01_1133700	1616L	1448703G>C	1
South America V East Africa	PvP01_12_v1	327608	T	G	missense	P48/45	335Y>335D	327608T>G	1
South America V East Africa	PvP01_12_v1	327731	A	G	missense	P48/45	376T>376A	327731A>G	1
South America V East Africa	PvP01_12_v1	327858	G	A	missense	P48/45	418R>418K	327858G>A	1
South America V East Africa	PvP01_12_v1	338808	G	A	synonymous	PVP01_1208400	3664C	338808G>A	1
South America V East Africa	PvP01_13_v1	337753	A	C	missense	CRMP3	1719K>1719N	337753A>C	1
South America V East Africa	PvP01_14_v1	762736	G	A	.	NA	NA	NA	1
South America V East Africa	PvP01_14_v1	1313512	A	G	intron	JmjC1	NA	NA	1
South America V East Africa	PvP01_12_v1	321295	G	A	.	NA	NA	NA	0.997307
South America V East Africa	PvP01_12_v1	323603	C	T	missense	P47	24L>24F	323603C>T	0.997293
South America V East Africa	PvP01_12_v1	323611	A	G	synonymous	P47	26T	323611A>G	0.997293
South America V East Africa	PvP01_11_v1	1265741	A	T	missense	PVP01_1129500	236N>236F	5740A>T+1265741	0.997092
South America V East Africa	PvP01_11_v1	1265741	A	T	missense	PVP01_1129500	236N>236I	1265741A>T	0.997092
South America V East Africa	PvP01_12_v1	2475598	C	T	missense	PVP01_1260100	372L>372F	2475598C>T	0.99454
South America V East Africa	PvP01_09_v1	1729971	G	C	.	NA	NA	NA	0.994313
South America V East Africa	PvP01_09_v1	1728961	C	T	.	NA	NA	NA	0.99277
South America V East Africa	PvP01_12_v1	909364	C	A	.	NA	NA	NA	0.991331

S4 Table. South American (n = 303) specific SNPs in comparison to global dataset (n = 762)

Chromosome	Pos	Ref	Alt	Effect	Gene name	Gene ID	Amino acid change	Nucleotide change	Fst
1	746509	G	A	synonymous	g-tub	PVP01_0116500	130D	746509G>A	0.92266
2	211376	G	C	intron	PIP5K				0.962355
2	211461	C	A	intron	PIP5K				0.968516
2	274834	G	C	missense	PVP01_0206700	PVP01_0206700	1758H>1758Q	274834G>C	0.93388
3	501852	A	C	missense	PVP01_0311200	PVP01_0311200	542V>542G	501852A>C	0.908747
3	542896	G	C	.					0.930594
3	553010	C	G	missense	PVP01_0312700	PVP01_0312700	54F>54L	553010C>G	0.93631
4	308265	C	G	intron	CDPK1				0.902433
4	427365	C	A	.					0.933689
4	652108	G	C	missense	P230p	PVP01_0415900	158L>158V	652108G>C	0.98542
5	574992	A	T	synonymous	PVP01_0513700	PVP01_0513700	24I	574992A>T	0.998831
5	685274	A	T	*missense	PVP01_0516500	PVP01_0516500	460D>460E	685274A>T	0.939088
5	941897	G	A	.					0.913281
5	1065815	G	T	missense	PVP01_0526300	PVP01_0526300	711S>711R	1065815G>T	0.90156
6	203676	A	T	missense	PVP01_0604900	PVP01_0604900	298N>298K	203676A>T	0.907944
6	203682	C	T	synonymous	PVP01_0604900	PVP01_0604900	296R	203682C>T	0.941023
6	286430	C	G	synonymous	PVP01_0606900	PVP01_0606900	1771L	286430C>G	0.920247
6	468862	C	T	missense	PVP01_0610600	PVP01_0610600	235L>235I	68862C>T+468864G:	0.903611
6	582427	G	C	splice_region	PVP01_0613700	PVP01_0613700			0.951656
6	665452	G	A	missense	PVP01_0615800	PVP01_0615800	192A>192T	665452G>A	0.947328
6	665510	G	T	missense	PVP01_0615800	PVP01_0615800	211R>211L	665510G>T	0.958594
6	667777	T	A	.					0.961145
6	668295	C	T	missense	P28	PVP01_0616000	87D>87N	668295C>T	0.957899
6	669053	A	G	.					0.988054
6	670673	G	A	missense	PVP01_0616100	PVP01_0616100	130T>130I	670673G>A	0.956994
6	674658	C	T	.					0.944536
7	747983	G	A	synonymous	PVP01_0716800	PVP01_0716800	574K	747983G>A	0.982149
7	747984	G	A	missense	PVP01_0716800	PVP01_0716800	575G>575S	747984G>A	0.982149
8	556253	G	A	missense	PVP01_0813100	PVP01_0813100	150S>150N	556253G>A	0.965517
8	829137	G	C	missense	PVP01_0818900	PVP01_0818900	657D>657H	829137G>C	0.925952
8	878316	T	C	synonymous	Cap380	PVP01_0819800	3375D	878316T>C	0.902605
8	879514	T	C	.					0.923836
8	1051697	T	C	missense	FNT	PVP01_0823500	252V>252A	1051697T>C	0.913174
8	1255798	T	C	synonymous	PVP01_0828900	PVP01_0828900	1014N	1255798T>C	0.926334
8	1546424	C	G	missense	PVP01_0836700	PVP01_0836700	489E>489Q	1546424C>G	0.912393
9	395581	C	A	synonymous	PVP01_0907400	PVP01_0907400	404P	395581C>A	0.905409
9	582923	A	G	missense	PVP01_0912400	PVP01_0912400	340R>340G	582923A>G	0.903448
9	826837	C	G	.					0.983604
9	878674	C	G	missense	PVP01_0920200	PVP01_0920200	517G>517A	878674C>G	0.966926
9	892855	G	C	missense	PVP01_0920500	PVP01_0920500	1053Q>1053H	892855G>C	0.972531
9	1062948	C	T	synonymous	CDPK7	PVP01_0923700	1425Y	1062948C>T	0.981905
9	1138842	G	C	missense	PVP01_0925500	PVP01_0925500	875K>875N	1138842G>C	0.946152
9	1155592	C	G	missense	mtRNAP	PVP01_0926000	790L>790V	1155592C>G	0.901744
9	1179972	C	A	missense	ATG7	PVP01_0926800	799V>799L	1179972C>A	0.932272
9	1198207	T	C	.					0.972396
9	1199491	C	T	missense	PVP01_0927300	PVP01_0927300	572E>572K	1199491C>T	0.98009
9	1325149	G	A	.					0.965334
9	1404621	G	A	synonymous	PVP01_0932800	PVP01_0932800	42S	1404621G>A	0.902473
9	1726492	C	T	.					0.937025

9	1728961	C	T	.					0.995687
9	1729971	G	C	.					0.996359
9	1730454	G	C	.					0.929572
10	753644	G	A	.					0.918477
10	1229415	A	T	missense	ALP2b	PVP01_1028800	332N>332K	1229415A>T	0.955999
11	672853	G	A	.					0.90844
11	985249	C	T	.					0.927048
11	1075389	G	A	missense	PVP01_1124700	PVP01_1124700	404E>404K	1075389G>A	0.920799
11	1172475	C	T	missense	ApiAP2	PVP01_1126600	764P>764L	1172475C>T	0.912423
11	1190787	G	C	intron	PVP01_1126900				0.95238
11	1246228	T	A	.					0.986413
11	1246557	C	G	.					0.989896
11	1253949	G	A	.					0.979871
11	1256157	A	G	synonymous	MSP10	PVP01_1129100	145A	1256157A>G	0.990139
11	1261039	T	G	.					0.997554
11	1262951	G	A	missense	PVP01_1129400	PVP01_1129400	20P>20S	1262951G>A	0.965648
11	1265741	A	T	missense	PVP01_1129500	PVP01_1129500	236N>236F	65740A>T+1265741A	0.997551
11	1265741	A	T	missense	PVP01_1129500	PVP01_1129500	236N>236I	1265741A>T	0.997551
11	1272806	G	A	missense	PVP01_1129700	PVP01_1129700	1074E>1074K	1272806G>A	0.98554
11	1276259	A	T	missense	PVP01_1129700	PVP01_1129700	2225T>2225S	1276259A>T	0.992816
11	1281911	G	A	synonymous	PVP01_1129900	PVP01_1129900	159D	1281911G>A	0.939999
11	1283422	G	A	synonymous	ACS12	PVP01_1130000	57E	1283422G>A	0.913339
11	1321844	A	G	intron	GPI1				0.933796
11	1395450	C	A	missense	TLP	PVP01_1132600	1232G>1232V	1395450C>A	0.933525
11	1400059	G	C	.					0.916168
11	1440215	G	T	.					0.983608
11	1448703	G	C	synonymous	PVP01_1133700	PVP01_1133700	1616L	1448703G>C	0.992726
11	1448967	C	T	synonymous	PVP01_1133700	PVP01_1133700	1704D	1448967C>T	0.961458
11	1451843	T	A	intron	PVP01_1133700				0.970556
11	1474601	T	C	.					0.970331
11	1483830	A	G	missense	PVP01_1134800	PVP01_1134800	579K>579R	1483830A>G	0.974619
11	1489148	T	C	missense	NAR1	PVP01_1134900	793V>793A	1489148T>C	0.954257
11	1514101	T	C	missense	ApiAP2	PVP01_1135300	319N>319D	1514101T>C	0.961547
11	1548390	A	T	.					0.941292
11	1557314	C	A	missense	P12	PVP01_1136400	125N>125K	1557314C>A	0.935174
11	1625978	G	A	missense	SPT5	PVP01_1138300	418A>418T	1625978G>A	0.917057
11	1688420	C	T	missense	PVP01_1139900	PVP01_1139900	5662A>5662T	1688420C>T	0.903895
11	1695821	A	T	missense	PVP01_1139900	PVP01_1139900	3195F>3195I	1695821A>T	0.94128
12	98194	G	T	missense	PVP01_1202300	PVP01_1202300	202D>202Y	98194G>T	0.940956
12	303164	C	T	.					0.925288
12	314694	G	C	synonymous	TOP3	PVP01_1207700	369L	314694G>C	0.950145
12	321295	G	A	.					0.996452
12	323603	C	T	missense	P47	PVP01_1208000	24L>24F	323603C>T	0.996481
12	323611	A	G	synonymous	P47	PVP01_1208000	26T	323611A>G	0.997652
12	323612	G	A	missense	P47	PVP01_1208000	27E>27K	323612G>A	0.956778
12	327355	C	A	missense	P48/45	PVP01_1208100	250N>250K	327355C>A	0.901433
12	327858	G	A	missense	P48/45	PVP01_1208100	418R>418K	327858G>A	0.997668
12	329689	G	A	synonymous	PVP01_1208200	PVP01_1208200	179K	329689G>A	0.959144
12	338808	G	A	synonymous	PVP01_1208400	PVP01_1208400	3664C	338808G>A	0.992813
12	362687	G	A	.					0.944988

12	367955	G	A	.					0.900331
12	398922	A	T	intron	TRX2				0.937101
12	424886	A	T	missense	PVP01_1210400	PVP01_1210400	195R>195W	424886A>T	0.989535
12	457790	C	T	.					0.906108
12	467311	C	A	synonymous	PVP01_1211000	PVP01_1211000	2447V	467311C>A	0.900469
12	469258	C	T	synonymous	PVP01_1211000	PVP01_1211000	3096N	469258C>T	0.906839
12	524330	T	A	missense	ApiAP2	PVP01_1211900	872M>872K	524330T>A	0.922757
12	652477	T	C	.					0.909435
12	770010	A	C	synonymous	TRAP	PVP01_1218700	350A	770010A>C	0.95335
12	1521747	T	C	missense	PVP01_1238400	PVP01_1238400	152I>152V	1521747T>C	0.914547
12	1591524	C	T	synonymous	PVP01_1239800	PVP01_1239800	64G	1591524C>T	0.930256
12	1591642	C	A	missense	PVP01_1239800	PVP01_1239800	25S>25I	1591642C>A	0.935878
12	1592724	C	A	.					0.923203
12	2127660	C	T	.					0.99152
13	331256	G	T	.					0.90133
13	335184	C	T	missense	CRMP3	PVP01_1307300	863S>863F	335184C>T	0.914863
13	336962	G	A	missense	CRMP3	PVP01_1307300	1456V>1456M	336962G>A	0.992209
13	337753	A	C	missense	CRMP3	PVP01_1307300	1719K>1719N	337753A>C	0.998728
13	821484	G	C	missense	SYN11	PVP01_1317500	229R>229T	821484G>C	0.907139
13	1045555	C	G	missense	PVP01_1323200	PVP01_1323200	151A>151G	1045555C>G	0.940623
13	1169523	A	G	.					0.952526
13	1551996	C	T	.					0.920725
13	1580838	A	T	missense	PVP01_1336500	PVP01_1336500	2198T>2198S	1580838A>T	0.942614
14	556784	C	G	*missense	PVP01_1412600	PVP01_1412600	343I>343M	556784C>G	0.912399
14	762736	G	A	.					0.978863
14	768196	C	T	missense	PVP01_1417600	PVP01_1417600	21S>21F	768196C>T	0.917863
14	934893	C	A	missense	PVP01_1421500	PVP01_1421500	1389L>1389I	934893C>A	0.951479
14	1313512	A	G	intron	JmjC1				0.998739
14	1319511	G	A	missense	PVP01_1430500	PVP01_1430500	403E>403K	1319511G>A	0.908798
14	1322275	C	A	missense	PVP01_1430500	PVP01_1430500	1067L>1067I	1322275C>A	0.986229
14	1354320	T	A	missense	PVP01_1431000	PVP01_1431000	74H>74Q	1354320T>A	0.91254
14	1624579	G	A	.					0.932625
14	1818379	C	G	missense	PVP01_1441700	PVP01_1441700	1531G>1531R	1818379C>G	0.922012
14	1869272	A	G	synonymous	PVP01_1442900	PVP01_1442900	63Y	1869272A>G	0.943637
14	2221345	A	G	.					0.900706
14	2258206	C	T	synonymous	AARP1	PVP01_1451700	2330H	2258206C>T	0.913741
14	2416947	T	G	missense	PVP01_1455600	PVP01_1455600	464N>464H	2416947T>G	0.903222
14	2595663	C	T	.					0.919279
14	2662513	G	T	synonymous	PVP01_1461600	PVP01_1461600	521R	2662513G>T	0.97041
14	2662867	T	A	missense	PVP01_1461600	PVP01_1461600	403I>403L	2662867T>A	0.972793
14	2663010	C	A	missense	PVP01_1461600	PVP01_1461600	355S>355I	2663010C>A	0.934976
14	2663285	C	A	missense	PVP01_1461600	PVP01_1461600	263K>263N	2663285C>A	0.927961

S5 Table. Highly differentiating SNPs (Fst > 0.8) separating Brazilian isolates (n = 111) from the South American population (n = 192)

Allele counts are given where 0 = reference allele, 0.5 = mixed call of reference and alternative, 1 = alternative allele, NA = missing data at this position.

Chromosome	Position	Ref	Alt	Effect	Gene name	Gene ID	Amino acid change	Nucleotide change	Fst	Brazil Allele count (N = 111)
PvP01_01_v1	663102	C	G	intron	PVP01_0114800	PVP01_0114800	NA	NA	0.852314	1(48),NA(42),0(20),0.5(1)
PvP01_01_v1	716831	A	T	missense	PVP01_0116000	PVP01_0116000	4344L>4344M	716831A>T	0.858734	1(86),0(14),NA(7),0.5(4)
PvP01_02_v1	377716	C	A	missense	PVP01_0209100	PVP01_0209100	590G>590V	377716C>A	0.868837	1(63),0(33),NA(15)
PvP01_03_v1	741751	G	C	.	NA	NA	NA	NA	0.852014	1(82),0(18),NA(8),0.5(3)
PvP01_04_v1	530215	T	C	missense	PVP01_0412900	PVP01_0412900	299E>299G	530215T>C	0.859607	1(71),0(26),NA(11),0.5(3)
PvP01_05_v1	440493	T	C	missense	NT2	PVP01_0509900	117F>117S	440493T>C	0.874857	1(84),0(22),NA(4),0.5(1)
PvP01_05_v1	541403	C	T	.	NA	NA	NA	NA	0.818885	1(63),0(34),NA(12),0.5(2)
PvP01_05_v1	572521	G	T	.	NA	NA	NA	NA	0.837211	1(66),0(28),NA(16),0.5(1)
PvP01_05_v1	574671	C	T	synonymous	PVP01_0513700	PVP01_0513700	131L	574671C>T	0.858916	1(88),0(17),0.5(4),NA(2)
PvP01_05_v1	943211	C	A	.	NA	NA	NA	NA	0.864798	1(56),0(34),NA(20),0.5(1)
PvP01_05_v1	1014612	G	C	.	NA	NA	NA	NA	0.869766	NA(108),0(2),1(1)
PvP01_05_v1	1072450	T	G	intron	PUF2	PVP01_0526500	NA	NA	0.802193	1(86),NA(19),0(5),0.5(1)
PvP01_06_v1	179243	A	T	missense	PVP01_0604500	PVP01_0604500	441L>441M	179243A>T	0.834752	1(70),NA(26),0(14),0.5(1)
PvP01_06_v1	248302	A	C	.	NA	NA	NA	NA	0.812195	1(57),0(27),NA(25),0.5(2)
PvP01_06_v1	334452	A	T	.	NA	NA	NA	NA	0.812948	1(66),NA(40),0(5)
PvP01_07_v1	360367	A	C	missense	PVP01_0706700	PVP01_0706700	544K>544Q	360367A>C	0.814614	1(65),0(38),0.5(6),NA(2)
PvP01_07_v1	440429	G	A	synonymous	PVP01_0708600	PVP01_0708600	112L	440429G>A	0.812197	1(52),0(46),NA(11),0.5(2)
PvP01_07_v1	649695	A	G	.	NA	NA	NA	NA	0.809879	1(80),NA(20),0(10),0.5(1)
PvP01_08_v1	461251	C	A	.	NA	NA	NA	NA	0.820703	1(50),NA(31),0(29),0.5(1)
PvP01_08_v1	462131	T	G	.	NA	NA	NA	NA	0.82776	NA(50),1(31),0(30)
PvP01_08_v1	1195569	A	G	synonymous	PVP01_0827200	PVP01_0827200	226K	1195569A>G	0.805411	1(81),0(22),NA(5),0.5(3)
PvP01_08_v1	1510168	C	T	.	NA	NA	NA	NA	0.81788	NA(65),1(46)
PvP01_08_v1	1519458	G	T	.	NA	NA	NA	NA	0.807491	1(70),NA(41)
PvP01_09_v1	553923	C	T	intron	VPS35	PVP01_0911300	NA	NA	0.801111	1(73),0(27),NA(8),0.5(3)
PvP01_09_v1	1366817	C	G	missense	SR1	PVP01_0931900	236E>236Q	1366817C>G	0.83248	1(75),0(23),NA(10),0.5(3)
PvP01_09_v1	1797351	G	A	.	NA	NA	NA	NA	0.826417	1(80),0(22),NA(5),0.5(4)
PvP01_09_v1	1943582	G	T	.	NA	NA	NA	NA	0.801676	1(72),0(25),NA(11),0.5(3)
PvP01_10_v1	256817	C	A	.	NA	NA	NA	NA	0.873925	1(74),NA(20),0(17)
PvP01_10_v1	286019	T	A	.	NA	NA	NA	NA	0.841663	1(82),0(23),NA(5),0.5(1)
PvP01_10_v1	301677	A	G	.	NA	NA	NA	NA	0.891889	1(76),0(17),NA(16),0.5(2)
PvP01_10_v1	306039	G	T	.	NA	NA	NA	NA	0.855725	1(65),0(31),NA(10),0.5(5)
PvP01_10_v1	371891	T	C	.	NA	NA	NA	NA	0.903163	1(87),0(17),0.5(5),NA(2)
PvP01_10_v1	385010	T	C	.	NA	NA	NA	NA	0.825556	1(52),0(44),NA(13),0.5(2)
PvP01_10_v1	481636	C	T	missense	MDR1	PVP01_1010900	500D>500N	481636C>T	0.921208	1(79),0(22),NA(6),0.5(4)
PvP01_10_v1	490615	C	G	missense	PVP01_1011000	PVP01_1011000	842G>842A	490615C>G	0.818444	1(64),0(29),NA(13),0.5(5)
PvP01_10_v1	1111675	T	A	.	NA	NA	NA	NA	0.853279	1(47),0(33),NA(31)
PvP01_10_v1	1122429	G	A	synonymous	TPT	PVP01_1025700	255A	1122429G>A	0.803377	1(64),0(39),NA(5),0.5(3)
PvP01_10_v1	1133901	C	T	.	NA	NA	NA	NA	0.844972	1(54),0(38),NA(18),0.5(1)
PvP01_10_v1	1153322	C	T	.	NA	NA	NA	NA	0.831507	NA(40),1(37),0(34)
PvP01_11_v1	377942	G	T	synonymous	PVP01_1109500	PVP01_1109500	328G	377942G>T	0.839499	1(73),0(31),NA(5),0.5(2)
PvP01_11_v1	915559	G	T	missense	PK4	PVP01_1121300	1694T>1694N	915559G>T	0.839191	1(88),0(17),NA(4),0.5(2)
PvP01_11_v1	965497	T	C	.	NA	NA	NA	NA	0.848	1(57),0(41),NA(12),0.5(1)
PvP01_11_v1	967548	G	C	.	NA	NA	NA	NA	0.856445	1(66),0(37),NA(5),0.5(3)
PvP01_11_v1	1327521	G	T	.	NA	NA	NA	NA	0.801094	1(98),0(6),NA(6),0.5(1)
PvP01_12_v1	649263	C	T	.	NA	NA	NA	NA	0.808474	1(72),0(28),NA(7),0.5(4)
PvP01_12_v1	1187189	T	G	.	NA	NA	NA	NA	0.844182	1(68),0(24),NA(17),0.5(2)

PvP01_12_v1	1523526	A	G	.	NA	NA	NA	NA	0.826027	1(67),0(34),NA(9),0.5(1)
PvP01_12_v1	1618925	A	G	missense	ApiAP2	PVP01_1250900	123I>123V	1618925A>G	0.867577	1(87),0(15),NA(8),0.5(1)
PvP01_12_v1	1621163	C	G	missense	ApiAP2	PVP01_1250900	869R>869G	1621163C>G	0.895036	1(94),NA(9),0(6),0.5(2)
PvP01_12_v1	1701637	C	T	synonymous	PVP01_1241700	PVP01_1241700	820S	1701637C>T	0.809855	1(60),0(40),0.5(6),NA(5)
PvP01_12_v1	1860075	C	T	missense	PVP01_1245000	PVP01_1245000	1553A>1553T	1860075C>T	0.853266	1(83),0(21),0.5(4),NA(3)
PvP01_12_v1	2424762	A	G	missense	PVP01_1258700	PVP01_1258700	979K>979R	2424762A>G	0.800613	1(88),0(13),NA(7),0.5(3)
PvP01_12_v1	2721219	T	A	synonymous	FACT-S	PVP01_1265600	31R	2721219T>A	0.835397	1(74),0(27),NA(8),0.5(2)
PvP01_13_v1	322323	C	T	synonymous	PVP01_1307200	PVP01_1307200	1404F	322323C>T	0.802792	1(70),NA(32),0(9)
PvP01_13_v1	336738	C	T	missense	CRMP3	PVP01_1307300	1381P>1381L	336738C>T	0.814964	1(83),NA(24),0(4)
PvP01_13_v1	809067	G	A	missense	PVP01_1317200	PVP01_1317200	1086R>1086Q	809067G>A	0.875779	1(76),0(23),NA(9),0.5(3)
PvP01_13_v1	810706	G	A	missense	PVP01_1317200	PVP01_1317200	1578G>1578D	810706G>A	0.849214	1(75),0(28),0.5(5),NA(3)
PvP01_13_v1	810706	G	A	splice_region	PVP01_1317200	PVP01_1317200	NA	NA	0.849214	1(75),0(28),0.5(5),NA(3)
PvP01_13_v1	818665	T	C	missense	PVP01_1317400	PVP01_1317400	39K>39E	818665T>C	0.876322	1(67),0(32),NA(12)
PvP01_13_v1	827933	G	A	.	NA	NA	NA	NA	0.8365	1(61),0(28),NA(21),0.5(1)
PvP01_13_v1	914350	G	A	intron	PVP01_1320100	PVP01_1320100	NA	NA	0.826523	1(64),0(25),NA(21),0.5(1)
PvP01_13_v1	1035368	A	T	missense	ABCG2	PVP01_1322800	457K>457M	1035368A>T	0.80736	1(104),NA(6),0.5(1)
PvP01_13_v1	1209703	C	G	missense	GlyRS	PVP01_1328500	75A>75P	1209703C>G	0.802047	1(56),0(36),NA(17),0.5(2)
PvP01_13_v1	1235727	T	C	.	NA	NA	NA	NA	0.902074	1(92),NA(11),0(8)
PvP01_13_v1	1803167	G	A	.	NA	NA	NA	NA	0.813397	1(69),0(35),NA(7)
PvP01_14_v1	115657	A	G	missense	RBP2a	PVP01_1402400	719K>719E	115657A>G	0.815252	1(45),0(43),NA(22),0.5(1)
PvP01_14_v1	121415	A	C	.	NA	NA	NA	NA	0.83404	1(47),0(34),NA(27),0.5(3)
PvP01_14_v1	1690445	C	T	intron	PVP01_1439600	PVP01_1439600	NA	NA	0.80084	1(77),0(29),NA(3),0.5(2)
PvP01_14_v1	1748109	G	T	.	NA	NA	NA	NA	0.810328	1(45),0(42),NA(21),0.5(3)
PvP01_14_v1	2145037	A	T	.	NA	NA	NA	NA	0.924846	NA(67),1(36),0(8)
PvP01_14_v1	2151758	G	A	.	NA	NA	NA	NA	0.904335	1(88),0(13),0.5(6),NA(4)
PvP01_14_v1	2153846	G	T	missense	PVP01_1449600	PVP01_1449600	1581P>1581T	2153846G>T	0.819992	1(91),0(11),NA(6),0.5(3)
PvP01_14_v1	2768649	G	A	missense	PVP01_1464600	PVP01_1464600	314M>314I	2768649G>A	0.806476	1(90),NA(13),0(8)
PvP01_14_v1	2864180	C	T	synonymous	G377	PVP01_1467200	1276K	2864180C>T	0.819134	1(59),0(37),NA(13),0.5(2)
PvP01_14_v1	2864300	T	C	synonymous	G377	PVP01_1467200	1236K	2864300T>C	0.81185	0(93),1(14),0.5(3),NA(1)
PvP01_14_v1	2869957	T	C	.	NA	NA	NA	NA	0.818498	1(73),NA(19),0(17),0.5(2)
PvP01_14_v1	2887017	C	T	missense	PVP01_1467700	PVP01_1467700	33A>33T	2887017C>T	0.829969	1(73),0(26),NA(9),0.5(3)
PvP01_14_v1	2900393	C	T	.	NA	NA	NA	NA	0.928941	1(88),NA(14),0(9)
PvP01_14_v1	2904040	T	A	.	NA	NA	NA	NA	0.817702	1(70),0(32),NA(8),0.5(1)
PvP01_14_v1	2915258	G	A	.	NA	NA	NA	NA	0.82004	1(81),NA(20),0(9),0.5(1)

S6 Table. Highly differentiating SNPs (Fst > 0.95) separating Brazilian isolates (n = 111) from each country within South America (pairwise)

Allele counts are given where 0 = reference allele, 0.5 = mixed call of reference and alternative, 1 = alternative allele, NA = missing data at this position.

Comparison	Chromosome	Position	Reference	Alt	Effect	Gene name	Amino acid change	Nucleotide change	Fst	Allele count Brazil (n = 111)	Allele count comparison country
Brazil VS Colombia	PvP01_08_v1	1508825	A	G	missense	CSP	38N>38G	1508825A>G+1508826A>G	1	0(95),NA(16)	1(31),NA(3)
Brazil VS Colombia	PvP01_08_v1	1508826	A	G	missense	CSP	38N>38G	1508825A>G+1508826A>G	1	0(95),NA(16)	1(31),NA(3)
Brazil VS Colombia	PvP01_08_v1	1510168	C	T	.	NA	NA	NA	1	NA(65),1(46)	0(31),NA(3)
Brazil VS Colombia	PvP01_14_v1	2043057	A	G	intron	PVP01_1447100	NA	NA	1	NA(105),0(6)	NA(33),1(1)
Brazil VS Colombia	PvP01_08_v1	1507050	G	T	.	NA	NA	NA	0.990124	1(68),NA(42),0.5(1)	0(33),NA(1)
Brazil VS Colombia	PvP01_13_v1	1034369	T	A	missense	ABCG2	124M>124Q	1034368A>C+1034369T>A	0.985692	0(110),NA(1)	1(33),0(1)
Brazil VS Colombia	PvP01_07_v1	496565	C	T	missense	CRMP1	407L>407F	496565C>T	0.985364	0(103),NA(7),1(1)	1(34)
Brazil VS Colombia	PvP01_07_v1	496547	G	C	missense	CRMP1	401V>401L	496547G>C	0.985256	0(102),NA(8),1(1)	1(34)
Brazil VS Colombia	PvP01_13_v1	1938055	G	C	missense	PVP01_1344100	33G>33A	1938055G>C	0.98392	0(107),NA(4)	1(22),NA(11),0(1)
Brazil VS Colombia	PvP01_13_v1	1035259	C	T	missense	ABCG2	421L>421F	1035259C>T	0.978212	0(108),NA(2),0.5(1)	1(33),0(1)
Brazil VS Colombia	PvP01_13_v1	1035368	A	T	missense	ABCG2	457K>457M	1035368A>T	0.977584	1(104),NA(6),0.5(1)	0(33),1(1)
Brazil VS Colombia	PvP01_13_v1	1036067	A	G	*missense	ABCG2	690H>690R	1036067A>G	0.97052	0(108),NA(3)	1(32),0(2)
Brazil VS Colombia	PvP01_14_v1	2697430	G	A	missense	PVP01_1462600	372S>372N	2697430G>A	0.961571	0(100),NA(9),0.5(1),1(1)	1(33),0(1)
Brazil VS Colombia	PvP01_14_v1	2906043	C	T	missense	PVP01_1468200	77T>77I	2906043C>T	0.95087	0(104),1(3),NA(3),0.5(1)	1(34)
Brazil VS Mexico	PvP01_14_v1	1049964	C	G	synonymous	PNPase	10G	1049964C>G	1	1(104),NA(7)	0(20)
Brazil VS Mexico	PvP01_14_v1	1693799	C	T	.	NA	NA	NA	1	0(109),NA(2)	1(20)
Brazil VS Mexico	PvP01_13_v1	1231937	C	G	missense	GR	80A>80G	1231937C>G	0.992212	0(108),NA(2),0.5(1)	1(20)
Brazil VS Mexico	PvP01_13_v1	1035368	A	T	missense	ABCG2	457K>457M	1035368A>T	0.991961	1(104),NA(6),0.5(1)	0(20)
Brazil VS Mexico	PvP01_04_v1	434194	G	A	synonymous	PVP01_0410600	1221S	434194G>A	0.984589	0(110),1(1)	1(20)
Brazil VS Mexico	PvP01_13_v1	1034369	T	A	missense	ABCG2	124M>124Q	1034368A>C+1034369T>A	0.983811	0(110),NA(1)	1(19),0(1)
Brazil VS Mexico	PvP01_14_v1	1554295	G	C	missense	PVP01_1437000	442A>442G	1554295G>C	0.983811	0(110),NA(1)	1(19),0(1)
Brazil VS Mexico	PvP01_12_v1	2093206	T	C	synonymous	PVP01_1251200	68G	2093206T>C	0.983686	0(109),NA(2)	1(19),0(1)
Brazil VS Mexico	PvP01_14_v1	1685572	G	A	intron	PVP01_1439400	NA	NA	0.983686	0(109),NA(2)	1(19),0(1)
Brazil VS Mexico	PvP01_13_v1	1036067	A	G	*missense	ABCG2	690H>690R	1036067A>G	0.983558	0(108),NA(3)	1(19),0(1)
Brazil VS Mexico	PvP01_11_v1	1548740	G	T	.	NA	NA	NA	0.983439	1(101),NA(9),0(1)	0(20)
Brazil VS Mexico	PvP01_09_v1	1783999	T	C	.	NA	NA	NA	0.983429	0(107),NA(4)	1(19),0(1)
Brazil VS Mexico	PvP01_12_v1	2091196	A	T	.	NA	NA	NA	0.983028	0(104),NA(7)	1(19),0(1)
Brazil VS Mexico	PvP01_11_v1	914852	T	A	missense	PK4	1930S>1930C	914852T>A	0.98275	0(102),NA(9)	1(19),0(1)
Brazil VS Mexico	PvP01_12_v1	2945024	C	A	.	NA	NA	NA	0.982462	0(100),NA(11)	1(19),0(1)
Brazil VS Mexico	PvP01_12_v1	2330891	T	C	.	NA	NA	NA	0.9817	1(95),NA(16)	0(19),1(1)
Brazil VS Mexico	PvP01_13_v1	1035259	C	T	missense	ABCG2	421L>421F	1035259C>T	0.975177	0(108),NA(2),0.5(1)	1(19),0(1)
Brazil VS Mexico	PvP01_01_v1	724793	T	G	missense	PVP01_0116000	1690I>1690L	724793T>G	0.97402	1(94),NA(15),0(1),0.5(1)	0(20)
Brazil VS Mexico	PvP01_02_v1	669049	G	T	.	NA	NA	NA	0.9671	0(110),1(1)	1(19),0(1)
Brazil VS Mexico	PvP01_12_v1	614088	G	A	missense	PVP01_1214400	310G>310R	614088G>A	0.966662	0(111)	1(18),0(2)
Brazil VS Mexico	PvP01_11_v1	1952727	C	G	synonymous	PVP01_1145700	154T	1952727C>G	0.96659	0(108),NA(2),1(1)	1(19),0(1)
Brazil VS Mexico	PvP01_11_v1	943048	A	G	synonymous	PVP01_1121400	5135T	943048A>G	0.966154	0(109),NA(2)	1(18),0(2)
Brazil VS Mexico	PvP01_12_v1	153050	C	T	intron	PVP01_1203700	NA	NA	0.966154	0(109),NA(2)	1(18),0(2)
Brazil VS Mexico	PvP01_12_v1	2279399	C	G	.	NA	NA	NA	0.966154	0(109),NA(2)	1(18),0(2)
Brazil VS Mexico	PvP01_10_v1	358652	G	A	.	NA	NA	NA	0.966064	0(106),NA(4),1(1)	1(19),0(1)
Brazil VS Mexico	PvP01_13_v1	360018	T	C	.	NA	NA	NA	0.966064	0(106),NA(4),1(1)	1(19),0(1)
Brazil VS Mexico	PvP01_01_v1	903246	A	C	missense	PVP01_0119600	113K>113Q	903246A>C	0.965894	0(108),NA(3)	1(18),0(2)
Brazil VS Mexico	PvP01_02_v1	473659	T	C	missense	NT4	180I>180T	473659T>C	0.965894	0(108),NA(3)	1(18),0(2)
Brazil VS Mexico	PvP01_05_v1	789400	G	A	synonymous	PVP01_0519000	874R	789400G>A	0.965894	0(108),NA(3)	1(18),0(2)
Brazil VS Mexico	PvP01_05_v1	789671	T	C	.	NA	NA	NA	0.965894	0(108),NA(3)	1(18),0(2)
Brazil VS Mexico	PvP01_12_v1	128261	G	A	synonymous	IMC1f	1250D	128261G>A	0.965894	0(108),NA(3)	1(18),0(2)
Brazil VS Mexico	PvP01_11_v1	384043	C	T	missense	PVP01_1109600	958E>958K	384043C>T	0.96563	0(107),NA(4)	1(18),0(2)
Brazil VS Mexico	PvP01_14_v1	1277382	C	T	synonymous	DBP1	569R	1277382C>T	0.96563	0(107),NA(4)	1(18),0(2)
Brazil VS Mexico	PvP01_08_v1	756862	G	A	missense	PVP01_0817500	36S>36N	756862G>A	0.965521	0(104),NA(6),1(1)	1(19),0(1)
Brazil VS Mexico	PvP01_10_v1	498057	G	C	intron	BUD31	NA	NA	0.965362	0(106),NA(5)	1(18),0(2)

Brazil VS Mexico	PvP01_10_v1	1127350	C	A	intron	PVP01_1025800	NA	NA	0.965362	0(106),NA(5)	1(18),0(2)
Brazil VS Mexico	PvP01_11_v1	948283	C	A	missense	PVP01_1121400	6880S>6880R	948283C>A	0.965362	0(106),NA(5)	1(18),0(2)
Brazil VS Mexico	PvP01_13_v1	1557856	C	A	.	NA	NA	NA	0.965362	0(106),NA(5)	1(18),0(2)
Brazil VS Mexico	PvP01_08_v1	715661	G	A	.	NA	NA	NA	0.96509	0(105),NA(6)	1(18),0(2)
Brazil VS Mexico	PvP01_09_v1	1738609	G	A	missense	PVP01_0939900	1452R>1452H	1738609G>A	0.96509	1(105),NA(6)	0(18),1(2)
Brazil VS Mexico	PvP01_11_v1	1662274	T	C	synonymous	PVP01_1139300	1123L	1662274T>C	0.96509	0(105),NA(6)	1(18),0(2)
Brazil VS Mexico	PvP01_14_v1	1277438	T	G	synonymous	DBP1	551R	1277438T>G	0.96509	1(105),NA(6)	0(18),1(2)
Brazil VS Mexico	PvP01_14_v1	2903440	A	T	.	NA	NA	NA	0.96509	0(105),NA(6)	1(18),0(2)
Brazil VS Mexico	PvP01_05_v1	420002	C	G	.	NA	NA	NA	0.964814	0(104),NA(7)	1(18),0(2)
Brazil VS Mexico	PvP01_05_v1	1092910	C	A	missense	PVP01_0526800	2399D>2399E	1092910C>A	0.964814	0(104),NA(7)	1(18),0(2)
Brazil VS Mexico	PvP01_07_v1	778321	T	G	.	NA	NA	NA	0.964533	0(103),NA(8)	1(18),0(2)
Brazil VS Mexico	PvP01_09_v1	1741068	A	G	.	NA	NA	NA	0.964248	0(102),NA(9)	1(18),0(2)
Brazil VS Mexico	PvP01_10_v1	1152398	A	G	.	NA	NA	NA	0.964121	1(90),NA(19),0(2)	0(20)
Brazil VS Mexico	PvP01_01_v1	903137	A	C	missense	PVP01_0119600	76L>76F	903137A>C	0.963663	0(100),NA(11)	1(18),0(2)
Brazil VS Mexico	PvP01_08_v1	691807	A	T	.	NA	NA	NA	0.963663	0(100),NA(11)	1(18),0(2)
Brazil VS Mexico	PvP01_10_v1	1243719	C	A	missense	PVP01_1029000	424P>424T	1243719C>A	0.962911	0(107),NA(4)	1(13),NA(5),0(2)
Brazil VS Mexico	PvP01_14_v1	2930585	C	T	.	NA	NA	NA	0.962749	0(97),NA(14)	1(18),0(2)
Brazil VS Mexico	PvP01_01_v1	880382	G	T	.	NA	NA	NA	0.962434	0(96),NA(15)	1(18),0(2)
Brazil VS Mexico	PvP01_13_v1	334389	A	T	missense	CRMP3	598E>598M	334388G>A+334389A>T	0.961705	1(83),NA(26),0(2)	0(20)
Brazil VS Mexico	PvP01_11_v1	464698	C	A	intron	PIAS	NA	NA	0.959709	1(88),NA(23)	0(18),1(2)
Brazil VS Mexico	PvP01_09_v1	2057365	A	G	missense	PVP01_0947300	483M>483T	2057365A>G	0.955483	0(104),NA(6),0.5(1)	1(18),0(2)
Brazil VS Mexico	PvP01_12_v1	323599	A	C	missense	P47	22L>22F	323599A>C	0.955483	1(104),NA(6),0.5(1)	0(18),1(2)
Brazil VS Mexico	PvP01_12_v1	1233558	G	A	missense	VPS52	459S>459N	1233558G>A	0.955177	0(105),NA(5),0.5(1)	1(17),0(2),NA(1)
Brazil VS Mexico	PvP01_09_v1	1999453	G	A	.	NA	NA	NA	0.952874	0(104),NA(4),1(3)	1(20)
Brazil VS Mexico	PvP01_12_v1	2965686	T	C	missense	CRMP4	380K>380E	2965686T>C	0.952096	0(95),NA(15),0.5(1)	1(18),0(2)
Brazil VS Mexico	PvP01_14_v1	1271444	C	T	missense	PPPK-DHPS	205M>205I	1271444C>T	0.95173	1(101),NA(7),0(3)	0(20)
Brazil VS Mexico	PvP01_10_v1	1133886	C	A	.	NA	NA	NA	0.95053	0(98),NA(10),1(3)	1(20)
Brazil VS Panama	PvP01_06_v1	951011	A	T	.	NA	NA	NA	1	0(74),NA(37)	NA(32),1(14)
Brazil VS Panama	PvP01_09_v1	2125269	A	G	synonymous	MAEBL	1182E	2125269A>G	1	NA(109),0(2)	NA(45),1(1)
Brazil VS Panama	PvP01_10_v1	451770	T	G	.	NA	NA	NA	1	0(82),NA(29)	1(27),NA(19)
Brazil VS Panama	PvP01_12_v1	1050083	T	A	.	NA	NA	NA	1	NA(72),0(39)	NA(45),1(1)
Brazil VS Panama	PvP01_14_v1	1821993	G	A	missense	PVP01_1441700	326S>326L	1821993G>A	1	0(64),NA(47)	NA(43),1(3)
Brazil VS Panama	PvP01_08_v1	871297	C	A	missense	Cap380	1036H>1036N	871297C>A	0.984694	0(104),NA(7)	1(30),NA(15),0(1)
Brazil VS Panama	PvP01_05_v1	1117667	C	G	missense	PVP01_0527300	173D>173E	1117667C>G	0.983174	0(100),NA(11)	1(23),NA(22),0(1)
Brazil VS Panama	PvP01_05_v1	216818	C	T	.	NA	NA	NA	0.980351	0(81),NA(30)	1(24),NA(21),0(1)
Brazil VS Panama	PvP01_06_v1	957685	G	T	missense	PVP01_0623400	180C>180F	957685G>T	0.978627	0(76),NA(35)	NA(24),1(21),0(1)
Brazil VS Panama	PvP01_06_v1	998999	A	G	intron	PVP01_0624100	NA	NA	0.973251	0(66),NA(44),1(1)	NA(37),1(9)
Brazil VS Panama	PvP01_10_v1	497660	C	G	.	NA	NA	NA	0.971264	0(108),NA(3)	1(35),NA(9),0(2)
Brazil VS Panama	PvP01_02_v1	377350	T	C	missense	PVP01_0209100	712H>712R	377350T>C	0.969564	0(101),NA(10)	1(34),NA(10),0(2)
Brazil VS Panama	PvP01_10_v1	482473	C	A	missense	MDR1	221V>221L	482473C>A	0.967475	0(105),NA(6)	1(24),NA(20),0(2)
Brazil VS Panama	PvP01_06_v1	664965	T	A	synonymous	PVP01_0615800	29I	664965T>A	0.95805	0(109),1(1),NA(1)	1(39),NA(5),0(2)
Brazil VS Panama	PvP01_13_v1	413024	G	A	missense	ABCB7	19P>19L	413024G>A	0.957834	0(107),NA(3),1(1)	1(40),NA(4),0(2)
Brazil VS Panama	PvP01_05_v1	599077	G	C	synonymous	SRCAP	478L	599077G>C	0.956778	0(107),NA(4)	1(36),NA(7),0(3)
Brazil VS Panama	PvP01_10_v1	497657	A	T	.	NA	NA	NA	0.956686	0(108),NA(3)	1(35),NA(8),0(3)
Brazil VS Panama	PvP01_11_v1	1130395	T	G	synonymous	SUV3	235S	1130395T>G	0.956623	1(104),NA(6),0(1)	0(39),NA(5),1(2)
Brazil VS Panama	PvP01_02_v1	530263	C	T	synonymous	PVP01_0212400	2R	530263C>T	0.956103	0(106),NA(5)	1(35),NA(8),0(3)
Brazil VS Panama	PvP01_13_v1	823007	G	A	missense	ISC3	772R>772W	823007G>A	0.956103	0(106),NA(5)	1(35),NA(8),0(3)
Brazil VS Panama	PvP01_14_v1	2849071	C	A	missense	THO2	1975A>1975S	2849071C>A	0.956033	0(107),NA(3),1(1)	1(35),NA(9),0(2)
Brazil VS Panama	PvP01_11_v1	1057913	G	T	missense	PVP01_1124300	116N>116K	1057913G>T	0.955298	0(106),NA(5)	1(33),NA(10),0(3)
Brazil VS Panama	PvP01_14_v1	223101	C	T	missense	PVP01_1404700	6447G>6447D	223101C>T	0.954127	0(105),NA(6)	1(31),NA(12),0(3)

Brazil VS Panama	PvP01_14_v1	223110	T	A	missense	PVP01_1404700	6444Q>6444L	223110T>A	0.954127	0(105),NA(6)	1(31),NA(12),0(3)
Brazil VS Panama	PvP01_08_v1	1465539	A	G	synonymous	PVP01_0834700	2347V	1465539A>G	0.952802	1(83),NA(26),0(1),0.5(1)	0(26),NA(19),1(1)
Brazil VS Panama	PvP01_07_v1	496547	G	C	missense	CRMP1	401V>401L	496547G>C	0.95275	0(102),NA(8),1(1)	1(31),NA(13),0(2)
Brazil VS Panama	PvP01_07_v1	496565	C	T	missense	CRMP1	407L>407F	496565C>T	0.952157	0(103),NA(7),1(1)	1(29),NA(15),0(2)
Brazil VS Panama	PvP01_11_v1	500344	C	A	.	NA	NA	NA	0.950905	0(66),NA(43),1(2)	NA(32),1(14)
Brazil VS Panama	PvP01_11_v1	1705391	C	T	missense	PVP01_1139900	5A>5T	1705391C>T	0.950769	0(98),NA(13)	1(29),NA(14),0(3)
Brazil VS Peru	PvP01_05_v1	195410	C	T	synonymous	PVP01_0504500	1296P	195410C>T	1	NA(72),0(39)	NA(87),1(2)
Brazil VS Peru	PvP01_05_v1	1014687	A	T	.	NA	NA	NA	1	NA(110),1(1)	NA(84),0(5)
Brazil VS Peru	PvP01_05_v1	1014688	A	T	.	NA	NA	NA	1	NA(110),1(1)	NA(84),0(5)
Brazil VS Peru	PvP01_05_v1	1014689	A	T	.	NA	NA	NA	1	NA(110),1(1)	NA(84),0(5)
Brazil VS Peru	PvP01_14_v1	2043057	A	G	intron	PVP01_1447100	NA	NA	1	NA(105),0(6)	NA(88),1(1)

S7 Table. Highly differentiating (Fst > 0.85) non-synonymous SNPs between assigned clades within Brazil

Comparison*	Chr	Pos	Ref	Alt	AA Change	Gene	Fst
1 VS 6	7	1218391	C	G	987A>987G	MSP1	1.00
1 VS 6	7	1218398	A	C	989E>989D	MSP1	1.00
1 VS 6	9	924599	C	G	686K>686N	PVP01_0921200	0.94
1 VS 6	13	1811970	G	A	423D>423N	PVP01_1341100	0.92
1 VS 6	14	1335832	G	C	2502G>2502A	PVP01_1430700	0.92
1 VS 6	10	1102717	C	A	138L>138I	SSB	0.91
1 VS 6	5	1014359	G	A	170D>170N	PVP01_0525000	0.91
1 VS 6	7	1218441	A	T	1004I>1004F	MSP1	0.90
1 VS 6	2	608108	G	A	857C>857Y	PVP01_0214300	0.89
1 VS 6	6	957535	A	T	130K>130I	PVP01_0623400	0.89
1 VS 6	9	1322105	C	A	10S>10R	PVP01_0930400	0.89
1 VS 6	11	429273	G	A	1194A>1194V	PVP01_1110800	0.88
1 VS 6	13	1406515	C	T	721P>721S	MCA3	0.88
1 VS 6	6	482846	T	G	1650N>1650H	PVP01_0610700	0.88
1 VS 6	14	1884457	A	C	1227T>1227H	PVP01_1443300	0.87
1 VS 6	14	1884457	A	C	1227T>1227P	PVP01_1443300	0.87
1 VS 6	13	1812447	A	G	582M>582V	PVP01_1341100	0.87
1 VS 6	13	1812582	G	A	627G>627R	PVP01_1341100	0.87
1 VS 6	3	214083	G	A	2256G>2256D	PVP01_0304000	0.87
1 VS 6	5	928680	C	T	157D>157N	PVP01_0523000	0.87
1 VS 6	8	901460	G	T	1087D>1087E	ABCI3	0.87
1 VS 6	1	832361	G	T	1956P>1956T	PVP01_0118300	0.86
1 VS 6	14	2351645	A	C	1405S>1405A	PVP01_1453900	0.86
1 VS 6	2	551804	A	G	365V>365A	PVP01_0213200	0.86
1 VS 6	3	251990	G	A	451P>451H	LISP2	0.86
1 VS 6	3	251991	G	T	451P>451H	LISP2	0.86
1 VS 6	5	886846	C	T	839S>839F	CRK3	0.85
1 VS 6	13	1615972	T	C	1218V>1218A	PREX	0.85
1 VS 6	9	1322086	A	G	4K>4R	PVP01_0930400	0.85
1 VS 6	9	1322087	G	A	4K>4R	PVP01_0930400	0.85
1 VS 7	7	1218391	C	G	987A>987G	MSP1	1.00
1 VS 7	7	1218398	A	C	989E>989D	MSP1	1.00
1 VS 7	2	608108	G	A	857C>857Y	PVP01_0214300	0.93
1 VS 7	5	928617	C	T	178E>178K	PVP01_0523000	0.92
1 VS 7	5	928617	C	T	178E>178N	PVP01_0523000	0.92
1 VS 7	14	1881276	T	G	216H>216Q	PVP01_1443300	0.89
1 VS 7	2	513991	G	T	2502A>2502E	PVP01_0212200	0.89
1 VS 7	7	959051	G	C	233L>233V	PVP01_0722200	0.88
1 VS 7	5	928680	C	T	157D>157N	PVP01_0523000	0.88
1 VS 7	14	2090885	A	T	216M>216K	PVP01_1448100	0.88
1 VS 7	13	1385683	G	C	1144G>1144R	PVP01_1332300	0.88
1 VS 7	5	886711	C	A	794A>794E	CRK3	0.88
1 VS 7	5	886846	C	T	839S>839F	CRK3	0.87
1 VS 7	13	1406515	C	T	721P>721S	MCA3	0.87
1 VS 7	13	1811970	G	A	423D>423N	PVP01_1341100	0.86
1 VS 7	13	1812582	G	A	627G>627R	PVP01_1341100	0.86
1 VS 7	9	1515296	G	A	2052P>2052S	PVP01_0935400	0.85
1 VS 7	14	2422669	T	A	315L>315M	PWP1	0.85
1 VS 7	2	501402	T	A	2635I>2635F	PVP01_0212100	0.85
1 VS 7	2	539283	G	T	480Q>480K	PVP01_0212800	0.85
1 VS 7	6	999468	A	T	152K>152I	PVP01_0624100	0.85

*Clade 1, n = 29; clade 6, n = 24; clade 7, n = 29

S8 Table. Highly differentiating (Fst > 0.85) SNPs between Group A (n = 88) and Group B (n = 23) isolates within Brazil

Chromosome	Position	Fst	Ref	Alt	Effect	Gene name	Gene ID	AA change	NT change
11	2038782	0.9858	A	G	.				
14	3015360	0.969	T	C	synonymous	PVP01_1470500	PVP01_1470500	34S	3015360T>C
11	2025987	0.9677	G	A	.				
6	999468	0.9672	A	T	missense	PIR protein	PVP01_0624100	152K>152I	999468A>T
14	24474	0.9636	T	A	.				
9	2146408	0.9592	A	G	.				
9	2150236	0.9428	T	C	missense	TRAG16	PVP01_0948800	292T>292A	2150236T>C
14	43455	0.9412	G	A	*missense	PVP01_1401100	PVP01_1401100	324V>324I	43455G>A
6	999976	0.9375	T	C	*synonymous	PIR protein	PVP01_0624100	321G	999976T>C
6	999864	0.9322	G	C	*missense	PIR protein	PVP01_0624100	284G>284A	999864G>C
6	999835	0.9286	A	G	*missense	PIR protein	PVP01_0624100	274I>274M	999835A>G
14	2994082	0.9149	A	T	.				
14	34286	0.9148	G	A	.				
2	739086	0.8999	T	C	.				
6	999629	0.8983	A	G	*missense	PIR protein	PVP01_0624100	206N>206D	999629A>G
6	999612	0.8929	A	T	*missense	PIR protein	PVP01_0624100	200K>200M	999612A>T
6	999547	0.8889	A	G	*synonymous	PIR protein	PVP01_0624100	178K	999547A>G
6	999559	0.8846	T	A	*missense	PIR protein	PVP01_0624100	182D>182E	999559T>A
6	999560	0.8846	G	A	*missense	PIR protein	PVP01_0624100	183D>183N	999560G>A
6	999569	0.8846	T	C	*missense	PIR protein	PVP01_0624100	186F>186H	999569T>C+999570T>A
6	999569	0.8846	T	C	*missense	PIR protein	PVP01_0624100	186F>186L	999569T>C
2	745495	0.8824	C	A	intron	PVP01_0217200			
6	999570	0.8824	T	A	*missense	PIR protein	PVP01_0624100	186F>186H	999569T>C+999570T>A
6	999570	0.8824	T	A	missense	PIR protein	PVP01_0624100	186F>186Y	999570T>A
6	999573	0.88	C	A	*missense	PIR protein	PVP01_0624100	187A>187D	999573C>A
13	591178	0.8774	A	G	.				
6	999590	0.875	G	T	*missense	PIR protein	PVP01_0624100	193D>193Y	999590G>T
6	999595	0.875	C	A	*missense	PIR protein	PVP01_0624100	194D>194E	999595C>A
8	856249	0.8699	C	T	splice_region	SPC2	PVP01_0819500		
6	999600	0.8583	G	C	*missense	PIR protein	PVP01_0624100	196S>196T	999600G>C
14	55321	0.8565	G	A	.				

S9 Table. Median of IBD fractions across the whole genome in all countries with > 10 monoclonal isolates

Country	Number of monoclonal isolates*	Median IBD	Standard deviation
Afghanistan	25	0.0121	0.0812
Brazil	107	0.0426	0.174
Cambodia	44	0.0123	0.0569
Colombia	31	0.0462	0.116
Ethiopia	39	0.0561	0.0468
India	39	0.0236	0.0894
Malaysia	39	0.335	0.427
Mexico	18	0.232	0.221
Myanmar	22	0.00698	0.303
Pakistan	32	0.0137	0.0772
Panama	46	0.971	0.481
Papua New Guinea	14	0.00607	0.178
Peru	75	0.0544	0.167

*Monoclonal = $F_{WS} > 0.95$

S10 Table. Highest 1% of pairwise Identity by Descent (IBD) fractions across the global database of 679 monoclonal isolates

			Southeast Asia		Southern SEA		South Asia			East Africa	South America					Genes	Gene products
Chr	Window Start	Window End	Cambodia	Myanmar	Malaysia	PNG	Afghanistan	India	Pakistan	Ethiopia	Brazil	Colombia	Mexico	Panama	Peru		
1	650001	700000	0.014	0.143	0.37	0.033	0.011	0.026	0.011	0.048	0.096	0.053	0.212	0.582	0.147	PVP01_0114600; PVP01_0114700; PVP01_0114800(FIKK); PVP01_0114900; PVP01_0115000; PVP01_0115100; PVP01_0115200; PVP01_0115300(GAMER); PVP01_0115400; PVP01_0115500	conserved protein unknown function; conserved Plasmodium protein unknown function; serine/threonine protein kinase FIKK family; PAP2-like protein putative; conserved Plasmodium protein unknown function; N-acetyltransferase GNAT family putative; conserved Plasmodium protein unknown function; gamete release protein putative; conserved protein unknown function; alpha/beta hydrolase putative
2	100001	150000	0.065	0.149	0.366	0.032	0.017	0.026	0.012	0.026	0.092	0.06	0.196	0.553	0.094	PVP01_0202000; PVP01_0202100; PVP01_0202200(TRAG2); PVP01_0202300; PVP01_0202400; PVP01_0202500; PVP01_0202600; PVP01_0202700; PVP01_0202800; PVP01_0202900	hypothetical protein; Plasmodium exported protein unknown function; tryptophan-rich protein; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein (PHIST) unknown function; 28S ribosomal RNA; 5.8S ribosomal RNA; 18S ribosomal RNA
2	150001	200000	0.094	0.152	0.37	0.033	0.011	0.026	0.01	0.035	0.104	0.066	0.206	0.578	0.088	PVP01_0203000(MRP1); PVP01_0203100; PVP01_0203200(TatD); PVP01_0203300; PVP01_0203400; PVP01_0203500; PVP01_0203600; PVP01_0203700; PVP01_0203800(GEXP19); PVP01_0203900; PVP01_0204000; PVP01_0204100	multidrug resistance-associated protein 1 putative; conserved Plasmodium protein unknown function; TatD-like deoxyribonuclease putative; conserved protein unknown function; eukaryotic translation initiation factor 4E putative; conserved protein unknown function; conserved Plasmodium protein unknown function; UMP-CMP kinase putative; conserved Plasmodium protein unknown function; P-loop containing nucleoside triphosphate hydrolase putative; conserved protein unknown function; kinesin-8 putative
2	450001	500000	0.012	0.105	0.37	0.033	0.008	0.023	0.009	0.013	0.143	0.051	0.266	0.58	0.084	PVP01_0211100; PVP01_0211200; PVP01_0211300; PVP01_0211400; PVP01_0211500(NT4); PVP01_0211600(VPS51); PVP01_0211700; PVP01_0211800; PVP01_0211900; PVP01_0212000(PMS1); PVP01_0212100	ATP-dependent RNA helicase putative; conserved Plasmodium protein unknown function; zinc-carboxypeptidase putative; conserved protein unknown function; nucleoside transporter 4; vacuolar protein sorting-associated protein 51 putative; vacuolar protein sorting-associated protein VTA1 putative; aspartate--tRNA ligase putative; seipin domain-containing protein putative; DNA mismatch repair protein PMS1 putative; conserved Plasmodium membrane protein unknown function
3	100001	150000	0.022	0.129	0.337	0.033	0.028	0.039	0.022	0.055	0.096	0.054	0.252	0.489	0.083	PVP01_0302000; PVP01_0302100; PVP01_0302200; PVP01_0302300; PVP01_0302400; PVP01_0302500; PVP01_0302600; PVP01_0302700; PVP01_0302800; PVP01_0302900	Plasmodium exported protein (PHIST) unknown function pseudogene; Plasmodium exported protein (PHIST) unknown function; tRNA Asparagine; regulator of chromosome condensation putative; pre-mRNA splicing factor putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; ubiquitin specific protease putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function
3	250001	300000	0.017	0.144	0.402	0.034	0.015	0.028	0.025	0.04	0.089	0.045	0.269	0.578	0.084	PVP01_0304700(LISP2); PVP01_0304800(PRPF8); PVP01_0304900; PVP01_0305000; PVP01_0305100; PVP01_0305200; PVP01_0305300(RON1); PVP01_0305400; PVP01_0305500; PVP01_0305600; PVP01_0305700(cGLO2)	liver specific protein 2 putative; pre-mRNA-processing-splicing factor 8 putative; conserved Plasmodium protein unknown function; TMEM33 domain-containing protein putative; lysine decarboxylase putative; conserved protein unknown function; rhoptry neck protein 1; conserved protein unknown function; V-type proton ATPase subunit B putative; sexual stage antigen s16 putative; cytosolic glyoxalase II putative

4	450001	500000	0.008	0.112	0.392	0.033	0.009	0.012	0.023	0.017	0.098	0.028	0.274	0.582	0.081	PVP01_0411000; PVP01_0411100; PVP01_0411200; PVP01_0411300; PVP01_0411400; PVP01_0411500(PK7); PVP01_0411600; PVP01_0411700; PVP01_0411800(SIS1); PVP01_0411900; PVP01_0412000(ATEL1); PVP01_0412100(MATE); PVP01_0412200; PVP01_0412300(SPATR); PVP01_0412400	RTR1 domain-containing protein putative; conserved protein unknown function; conserved protein unknown function; conserved Plasmodium protein unknown function; tetratricopeptide repeat protein putative; protein kinase 7 putative; conserved protein unknown function; conserved protein unknown function; protein SIS1 putative; conserved protein unknown function; arginyl-tRNA--protein transferase putative; multidrug efflux pump putative; SRR1-like protein; secreted protein with altered thrombospondin repeat domain putative; conserved Plasmodium protein unknown function
5	800001	850000	0.013	0.139	0.57	0.035	0.009	0.019	0.009	0.026	0.143	0.043	0.297	0.582	0.088	PVP01_0519500(VPS9); PVP01_0519600(Ub); PVP01_0519700(EIF3G); PVP01_0519800; PVP01_0519900; PVP01_0520000(PGM2); PVP01_0520100(RPT3); PVP01_0520200; PVP01_0520300; PVP01_0520400(USP13); PVP01_0520500(SMC3); PVP01_0520600; PVP01_0520700	vacuolar protein sorting-associated protein 9 putative; ubiquitin putative; eukaryotic translation initiation factor 3 subunit G putative; conserved Plasmodium protein unknown function; ATP synthase-associated protein putative; phosphoglucomutase-2 putative; 26S protease regulatory subunit 6B putative; lysine decarboxylase-like protein putative; 50S ribosomal protein L10 putative; ubiquitin carboxyl-terminal hydrolase 13 putative; structural maintenance of chromosomes protein 3 putative; conserved Plasmodium protein unknown function; calmodulin-like protein
5	850001	900000	0.013	0.143	0.521	0.035	0.007	0.016	0.01	0.03	0.146	0.036	0.223	0.581	0.092	PVP01_0520700; PVP01_0520800; PVP01_0520900; PVP01_0521000; PVP01_0521100(APH); PVP01_0521200; PVP01_0521300; PVP01_0521400(ARO); PVP01_0521500; PVP01_0521600; PVP01_0521700; PVP01_0521800(CRK3); PVP01_0521900; PVP01_0522000(CBP20); PVP01_0522100(GAK); PVP01_0522200; PVP01_0522300	calmodulin-like protein; Rab5-interacting protein putative; conserved Plasmodium protein unknown function; RNA-binding protein putative; acylated pleckstrin-homology domain-containing protein putative; GTP-binding protein putative; conserved Plasmodium protein unknown function; armadillo-domain containing rhoptry protein putative; arsenical pump-driving ATPase putative; conserved protein unknown function; conserved Plasmodium protein unknown function; cdc2-related protein kinase 3 putative; coatomer subunit zeta putative; nuclear cap-binding protein subunit 2 putative; GTP:AMP phosphotransferase putative; conserved Plasmodium protein unknown function; RING zinc finger protein putative
5	1000001	1050000	0.02	0.154	0.523	0.033	0.008	0.034	0.015	0.044	0.117	0.043	0.446	0.579	0.078	PVP01_0524600; PVP01_0524700; PVP01_0524800; PVP01_0524900; PVP01_0525000; PVP01_0525100; PVP01_0525200; PVP01_0525300;	Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function pseudogene; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST)
5	1050001	1100000	0.014	0.185	0.519	0.033	0.013	0.032	0.013	0.084	0.122	0.052	0.451	0.58	0.09	PVP01_0525800(HAT1); PVP01_0525900(MAF1); PVP01_0526000(PHBL); PVP01_0526100; PVP01_0526200(SAR1); PVP01_0526300; PVP01_0526400; PVP01_0526500(PUF2); PVP01_0526600(DHFR-TS); PVP01_0526700; PVP01_0526800	histone acetyltransferase putative; repressor of RNA polymerase III transcription MAF1 putative; prohibitin-like protein PHBL putative; CDGSH iron-sulfur domain-containing protein putative; small GTP-binding protein sar1 putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; mRNA-binding protein PUF2 putative; bifunctional dihydrofolate reductase-thymidylate synthase putative; LETM1-like protein putative; conserved Plasmodium protein unknown function
5	1100001	1150000	0.01	0.13	0.481	0.036	0.01	0.015	0.011	0.033	0.111	0.049	0.305	0.582	0.126	PVP01_0526800; PVP01_0526900; PVP01_0527000; PVP01_0527100; PVP01_0527200(CRK1); PVP01_0527300; PVP01_0527400; PVP01_0527500(SOC1); PVP01_0527600(EIF3M); PVP01_0527700	conserved Plasmodium protein unknown function; memo-like protein; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; cdc2-related protein kinase 1 putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; protein SOC1 putative; eukaryotic translation initiation factor 3 subunit M putative; conserved Plasmodium protein unknown function
6	700001	750000	0.008	0.139	0.388	0.033	0.012	0.027	0.011	0.015	0.145	0.046	0.213	0.581	0.099	PVP01_0616800(I3); PVP01_0616900; PVP01_0617000; PVP01_0617100(RimM); PVP01_0617200(DCP1); PVP01_0617300; PVP01_0617400; PVP01_0617500(PIGA); PVP01_0617600(Der1-2); PVP01_0617700; PVP01_0617800; PVP01_0617900; PVP01_0618000; PVP01_0618100(HPR1); PVP01_0618200(AdoMetDC/ODC)	protein phosphatase inhibitor 3 putative; conserved protein unknown function; conserved Plasmodium protein unknown function; ribosome maturation factor RimM putative; mRNA-decapping enzyme subunit 1 putative; conserved Plasmodium protein unknown function; conserved protein unknown function; phosphatidylinositol N-acetylglucosaminyltransferase subunit A putative; DER1-like protein putative; conserved protein unknown function; conserved Plasmodium protein unknown function; leucine-rich repeat protein; RNA polymerase II-associated protein 1 putative; heptatricopeptide repeat-containing protein HPR1 putative; S-adenosylmethionine decarboxylase/ornithine decarboxylase putative
6	950001	1000000	0.031	0.151	0.422	0.074	0.008	0.045	0.007	0.088	0.092	0.08	0.24	0.581	0.102	PVP01_0623300; PVP01_0623400; PVP01_0623500; PVP01_0623600; PVP01_0623700; PVP01_0623800(DBP); PVP01_0623900(REX4); PVP01_0624000; PVP01_0624100	Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; PIR protein; Plasmodium exported protein (PHIST) unknown function; duffy binding protein; ring-exported protein 4 putative; Plasmodium exported protein unknown function; PIR protein

7	150001	200000	0.009	0.11	0.381	0.033	0.014	0.021	0.013	0.021	0.077	0.03	0.346	0.583	0.143	PVP01_0702500; PVP01_0702600(PH); PVP01_0702700; PVP01_0702800(SPC3); PVP01_0702900; PVP01_0703000; PVP01_0703100(HU); PVP01_0703200(RPA1); PVP01_0703300(CuTP); PVP01_0703400; PVP01_0703500(NUP221)	AP-4 complex subunit epsilon putative; PH domain-containing protein putative; conserved protein unknown function; signal peptidase complex subunit 3 putative; prefoldin subunit 4 putative; ubiquitin specific protease putative; bacterial histone-like protein putative; replication protein A1 small fragment putative; copper-transporting ATPase putative; conserved Plasmodium protein unknown function; nucleoporin NUP221 putative
7	250001	300000	0.008	0.118	0.377	0.033	0.012	0.025	0.015	0.031	0.084	0.067	0.209	0.582	0.172	PVP01_0704200; PVP01_0704300(SEC27); PVP01_0704400; PVP01_0704500(RNaseII); PVP01_0704600(VPS46); PVP01_0704700; PVP01_0704800; PVP01_0704900; PVP01_0705000; PVP01_0705100	conserved Plasmodium protein unknown function; coatomer subunit beta putative; endonuclease/exonuclease/phosphatase domain containing protein; exoribonuclease II putative; vacuolar protein sorting-associated protein 46 putative; conserved Plasmodium protein unknown function; Maf-like protein putative; dynein intermediate light chain putative; arginase putative; zinc finger protein putative
7	300001	350000	0.006	0.111	0.371	0.033	0.011	0.02	0.014	0.015	0.093	0.062	0.21	0.584	0.169	PVP01_0705100; PVP01_0705200(LRR9); PVP01_0705300; PVP01_0705400; PVP01_0705500; PVP01_0705600; PVP01_0705700; PVP01_0705800; PVP01_0705900(ClpY); PVP01_0706000; PVP01_0706100; PVP01_0706200; PVP01_0706300; PVP01_0706400(PDF); PVP01_0706500(PMP1)	zinc finger protein putative; leucine-rich repeat protein; conserved Plasmodium protein unknown function; tubulin-specific chaperone putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; GTPase-activating protein putative; prefoldin-like protein putative; ATP-dependent protease ATPase subunit ClpY putative; conserved protein unknown function; translation initiation factor SUI1 putative; proteasome activator 28 subunit beta putative; ribosomal protein L35 apicoplast putative; peptide deformylase putative; plasma membrane protein 1 putative
7	1400001	1450000	0.013	0.107	0.36	0.033	0.014	0.021	0.011	0.04	0.099	0.047	0.184	0.582	0.091	PVP01_0734200; PVP01_0734300; PVP01_0734400; PVP01_0734500(CLAG7); PVP01_0734600; PVP01_0734700; PVP01_0734800(ETRAMP); PVP01_0734900	Plasmodium exported protein unknown function; conserved Plasmodium protein unknown function; Plasmodium exported protein unknown function; cytoadherence linked asexual protein 7; Plasmodium exported protein unknown function; hypothetical protein; early transcribed membrane protein; Plasmodium exported protein (PHIST) unknown function
8	650001	700000	0.013	0.109	0.479	0.033	0.023	0.024	0.009	0.032	0.081	0.063	0.22	0.589	0.076	PVP01_0814700(ADA2); PVP01_0814800(PHB2); PVP01_0814900; PVP01_0815000; PVP01_0815100; PVP01_0815200; PVP01_0815300(CysRS); PVP01_0815400(METAP1b); PVP01_0815500; PVP01_0815600; PVP01_0815700(HSP60); PVP01_0815800; PVP01_0815900	transcriptional coactivator ADA2 putative; prohibitin 2 putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; FAD synthetase putative; CWC16 domain-containing protein putative; cysteine--tRNA ligase putative; methionine aminopeptidase 1b putative; conserved Plasmodium protein unknown function; nucleotidyltransferase putative; heat shock protein 60 putative; tubulin binding cofactor c putative; ribonucleoside-diphosphate reductase small chain putative
9	600001	650000	0.01	0.208	0.551	0.036	0.011	0.029	0.016	0.021	0.079	0.045	0.217	0.582	0.066	PVP01_0912900; PVP01_0913000(COQ4); PVP01_0913100; PVP01_0913200; PVP01_0913300; PVP01_0913400; PVP01_0913500; PVP01_0913600; PVP01_0913700; PVP01_0913800; PVP01_0913900(PRL)	conserved Plasmodium protein unknown function; ubiquinone biosynthesis protein COQ4 putative; conserved Plasmodium protein unknown function; nucleic acid binding protein putative; RNA transcription translation and transport factor protein putative; ATPase putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; protein tyrosine phosphatase putative
9	650001	700000	0.011	0.208	0.531	0.036	0.009	0.029	0.018	0.03	0.081	0.059	0.239	0.582	0.085	PVP01_0914000; PVP01_0914100(UGT); PVP01_0914200(DSK2); PVP01_0914300; PVP01_0914400; PVP01_0914500(GLO1); PVP01_0914600; PVP01_0914700(MAPK2); PVP01_0914800; PVP01_0914900(ROM1); PVP01_0915000; PVP01_0915100; PVP01_0915200; PVP01_0915300; PVP01_0915400(CLK3)	conserved Plasmodium protein unknown function; UDP-galactose transporter putative; ubiquitin domain-containing protein DSK2 putative; GTP-binding protein putative; conserved Plasmodium protein unknown function; glyoxalase I putative; conserved Plasmodium protein unknown function; mitogen-activated protein kinase 2 putative; dynein light chain Tctex-type putative; rhomboid protease ROM1 putative; GTPase-activating protein putative; RAP protein putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; serine/threonine protein kinase putative
9	700001	750000	0.025	0.21	0.369	0.033	0.013	0.013	0.009	0.016	0.075	0.055	0.242	0.582	0.078	PVP01_0915400(CLK3); PVP01_0915500(G3PDH); PVP01_0915600; PVP01_0915700; PVP01_0915800; PVP01_0915900(SET7); PVP01_0916000; PVP01_0916100(VP3); PVP01_0916200; PVP01_0916300(ApiAP2); PVP01_0916400; PVP01_0916450; PVP01_0916500(DHHC9); PVP01_0916600(RON4); PVP01_0916700	serine/threonine protein kinase putative; glycerol-3-phosphate dehydrogenase putative; conserved protein unknown function; conserved protein unknown function; histone-lysine N-methyltransferase SET7 putative; vivapain-2; vivapain-3; vivapain-2; AP2 domain transcription factor putative; peptidyl-prolyl cis-trans isomerase 11 putative; conserved Plasmodium protein unknown function; palmitoyltransferase DHHC9 putative; rhoptry neck protein 4; serine esterase putative

9	850001	900000	0.009	0.133	0.373	0.033	0.016	0.017	0.009	0.016	0.085	0.054	0.333	0.586	0.089	PVP01_0919300(MLC-B); PVP01_0919400(ARC40); PVP01_0919500; PVP01_0919600; PVP01_0919700; PVP01_0919800; PVP01_0919900(U2AF1); PVP01_0920000; PVP01_0920100; PVP01_0920200; PVP01_0920300; PVP01_0920400(ASF1); PVP01_0920500(SEC16); PVP01_0920600	myosin light chain B putative; actin-related protein 2/3 complex subunit 1 putative; conserved protein unknown function; acyl-CoA-binding protein putative; tRNA m(1)G methyltransferase putative; conserved protein unknown function; splicing factor U2AF small subunit putative; ubiquitin-protein ligase putative; AP-4 complex subunit mu putative; ATP-dependent zinc metalloprotease FTSH putative; conserved Plasmodium protein unknown function; alternative splicing factor ASF-1 putative; protein transport protein SEC16 putative; conserved protein unknown function
9	950001	1000000	0.019	0.139	0.517	0.033	0.021	0.057	0.014	0.064	0.119	0.06	0.281	0.581	0.118	PVP01_0921900(TKL2); PVP01_0922000; PVP01_0922100; PVP01_0922200(EXP1); PVP01_0922300(GCN20); PVP01_0922400; PVP01_0922500; PVP01_0922600; PVP01_0922700(GPI16); PVP01_0922800; PVP01_0922900; PVP01_0923000	tyrosine kinase-like protein putative; WD repeat-containing protein putative; conserved protein unknown function; exported protein 1 putative; protein GCN20 putative; peptidase M16 putative; serine/threonine protein kinase putative; conserved protein unknown function; GPI transamidase component GPI16 putative; JmjC domain-containing protein putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function
9	1450001	1500000	0.014	0.123	0.49	0.033	0.018	0.043	0.014	0.081	0.113	0.043	0.198	0.585	0.072	PVP01_0934000; PVP01_0934100(CDC50B); PVP01_0934200(AMA1); PVP01_0934300; PVP01_0934400; PVP01_0934500; PVP01_0934600; PVP01_0934700; PVP01_0934800(HSP70-3); PVP01_0934900(PDI-11); PVP01_0935000; PVP01_0935100	conserved Plasmodium protein unknown function; LEM3/CDC50 family protein putative; apical membrane antigen 1; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; FHA domain-containing protein putative; RNA (uracil-5-)methyltransferase putative; conserved Plasmodium protein unknown function; heat shock protein 70 putative; protein disulfide isomerase putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function
9	1550001	1600000	0.012	0.11	0.418	0.033	0.012	0.019	0.009	0.015	0.075	0.043	0.222	0.587	0.061	PVP01_0936100; PVP01_0936200; PVP01_0936300; PVP01_0936400; PVP01_0936500; PVP01_0936600;	conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; conserved protein unknown function; condensin-2
9	1750001	1800000	0.01	0.112	0.482	0.033	0.008	0.024	0.015	0.033	0.096	0.054	0.484	0.581	0.087	PVP01_0940100(ApiAP2); PVP01_0940200; PVP01_0940300; PVP01_0940400; PVP01_0940500; PVP01_0940600(EMC8); PVP01_0940700; PVP01_0940800; PVP01_0940900; PVP01_0941000	AP2 domain transcription factor putative; AAA family ATPase putative; conserved Plasmodium protein unknown function; adrenodoxin reductase putative; conserved Plasmodium protein unknown function; ER membrane protein complex subunit 8 putative; carbonic anhydrase putative pseudogene; V-type proton ATPase subunit F putative; conserved Plasmodium protein unknown function; P-loop containing nucleoside triphosphate hydrolase putative
10	150001	200000	0.013	0.109	0.416	0.047	0.016	0.024	0.018	0.035	0.1	0.075	0.454	0.584	0.085	PVP01_1002700; PVP01_1002800; PVP01_1002900(FRM1); PVP01_1003000; PVP01_1003100(DPM3); PVP01_1003200; PVP01_1003300; PVP01_1003400; PVP01_1003500; PVP01_1003600; PVP01_1003700(PPT); PVP01_1003800(SYN6); PVP01_1003900; PVP01_1004000	conserved Plasmodium protein unknown function; SURF1 domain-containing protein putative; formin 1 putative; CPW-WPC family protein; dolichol-phosphate mannosyltransferase subunit 3 putative; conserved Plasmodium protein unknown function; XAP-5 DNA binding protein putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; phosphoenolpyruvate/phosphate translocator putative; SNARE protein putative; conserved Plasmodium protein unknown function; RING zinc finger protein putative
10	300001	350000	0.013	0.106	0.399	0.175	0.011	0.018	0.011	0.036	0.124	0.057	0.44	0.584	0.074	PVP01_1006800(MCM3); PVP01_1006900; PVP01_1007000; PVP01_1007100; PVP01_1007200; PVP01_1007250; PVP01_1007300; PVP01_1007400; PVP01_1007500(WDR55); PVP01_1007600(ARF-GAP); PVP01_1007700	DNA replication licensing factor MCM3 putative; transmembrane emp24 domain-containing protein putative; conserved Plasmodium protein unknown function; conserved protein unknown function; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; WD repeat-containing protein 55 putative; ADP-ribosylation factor GTPase-activating protein putative; conserved Plasmodium protein unknown function
10	350001	400000	0.012	0.106	0.377	0.133	0.013	0.016	0.008	0.021	0.12	0.048	0.436	0.583	0.084	PVP01_1007800; PVP01_1007900(NEK2); PVP01_1008000(IMC1g); PVP01_1008100; PVP01_1008200; PVP01_1008300; PVP01_1008400; PVP01_1008500; PVP01_1008600(SMC6); PVP01_1008700(ACS10); PVP01_1008800	RAP protein putative; NIMA related kinase 2 putative; inner membrane complex protein 1g putative; conserved protein unknown function; RNA methyltransferase putative; WD repeat-containing protein putative; 7-helix-1 protein putative; conserved protein unknown function; structural maintenance of chromosomes protein 6 putative; acyl-CoA synthetase putative; zinc finger protein putative

10	450001	500000	0.037	0.115	0.4	0.041	0.018	0.025	0.023	0.124	0.136	0.077	0.4	0.583	0.184	PVP01_1010100; PVP01_1010200; PVP01_1010300; PVP01_1010400; PVP01_1010500; PVP01_1010600; PVP01_1010700; PVP01_1010800; PVP01_1010900(MDR1); PVP01_1011000; PVP01_1011100(BUD31); PVP01_1011200(SufA)	divalent metal transporter putative; conserved Plasmodium protein unknown function; conserved protein unknown function; dynein light chain Tctex-type putative; DnaJ protein putative; cytochrome c oxidase subunit ApiCOX18 putative; heptatricopeptide repeat-containing protein putative; mitochondrial-processing peptidase subunit alpha putative; multidrug resistance protein 1 putative; zinc finger protein putative; pre-mRNA-splicing factor BUD31 putative; iron-sulfur assembly protein
10	550001	600000	0.012	0.108	0.4	0.074	0.01	0.02	0.012	0.021	0.07	0.035	0.385	0.583	0.071	PVP01_1012000; PVP01_1012100; PVP01_1012200(DDX1); PVP01_1012300; PVP01_1012400; PVP01_1012500; PVP01_1012600; PVP01_1012700; PVP01_1012800; PVP01_1012900; PVP01_1013000(SAHH); PVP01_1013100; PVP01_1013200	conserved Plasmodium protein unknown function; AFG1-like ATPase putative; ATP-dependent RNA helicase DDX1 putative; conserved Plasmodium protein unknown function; ribosomal large subunit pseudouridylate synthase putative; conserved protein unknown function; zinc finger protein putative; conserved protein unknown function; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; adenosylhomocysteinase putative; conserved protein unknown function; CDC73 domain-containing protein putative
10	1000001	1050000	0.017	0.118	0.386	0.041	0.011	0.029	0.009	0.073	0.093	0.06	0.234	0.584	0.084	PVP01_1023000(TCTP); PVP01_1023100; PVP01_1023200; PVP01_1023300; PVP01_1023400; PVP01_1023500; PVP01_1023600; PVP01_1023700(SIP); PVP01_1023800(CYP87); PVP01_1023900; PVP01_1024000	translationally-controlled tumor protein homolog putative; small nucleolar RNA snoR04; WD repeat-containing protein putative; conserved Plasmodium protein unknown function; tRNA Leucine; topoisomerase I putative; PDCD2 domain-containing protein putative; stripes inner membrane complex protein putative; peptidyl-prolyl cis-trans isomerase putative; conserved protein unknown function; conserved Plasmodium protein unknown function
10	1250001	1300000	0.015	0.109	0.519	0.034	0.01	0.014	0.01	0.018	0.078	0.034	0.308	0.582	0.079	PVP01_1029100; PVP01_1029200; PVP01_1029300(CEP120); PVP01_1029400(BCKDHB); PVP01_1029500; PVP01_1029600; PVP01_1029700; PVP01_1029800(DDX27); PVP01_1029900; PVP01_1030000(ATPase3); PVP01_1030100	conserved protein unknown function; conserved Plasmodium protein unknown function; centrosomal protein CEP120 putative; 2-oxoisovalerate dehydrogenase subunit beta mitochondrial putative; MOLO1 domain-containing protein putative; ATP-dependent helicase putative; conserved Plasmodium protein unknown function; ATP-dependent RNA helicase DDX27 putative; AN1-type zinc finger protein putative; cation transporting P-ATPase putative; conserved Plasmodium protein unknown function
10	1300001	1350000	0.013	0.109	0.515	0.035	0.03	0.033	0.022	0.03	0.086	0.052	0.227	0.581	0.075	PVP01_1030100; PVP01_1030200(RPL31); PVP01_1030300(MyoB); PVP01_1030400; PVP01_1030500(ADF1); PVP01_1030600(SRSF12); PVP01_1030700; PVP01_1030800; PVP01_1030900(MSP3.11); PVP01_1031000(MSP3.10); PVP01_1031100(MSP3G); PVP01_1031200(MSP3.9); PVP01_1031300(MSP3.8); PVP01_1031400(MSP3.5)	conserved Plasmodium protein unknown function; 60S ribosomal protein L31 putative; myosin B putative; protein kinase putative; actin-depolymerizing factor 1 putative; serine/arginine-rich splicing factor 12 putative; hypothetical protein; hypothetical protein; merozoite surface protein 3; merozoite surface protein 3; merozoite surface protein 3; merozoite surface protein 3; merozoite surface protein 3
10	1400001	1450000	0.013	0.116	0.589	0.033	0.029	0.037	0.024	0.066	0.099	0.055	0.201	0.577	0.078	PVP01_1032900; PVP01_1033000(VPS11); PVP01_1033100; PVP01_1033200(CAF1); PVP01_1033300(APC3); PVP01_1033400(RAP2); PVP01_1033500; PVP01_1033600(SBP1); PVP01_1033700; PVP01_1033800(TRAG17); PVP01_1033900(TRAG18)	HCNGP-like protein; vacuolar protein sorting-associated protein 11 putative; conserved protein unknown function; chromatin assembly factor 1 subunit A putative; anaphase-promoting complex subunit 3 putative; rhoptry-associated protein 2; Plasmodium exported protein unknown function; skeleton-binding protein 1 putative; Plasmodium exported protein unknown function; tryptophan-rich protein; tryptophan-rich protein
11	350001	400000	0.016	0.108	0.513	0.033	0.008	0.014	0.01	0.026	0.097	0.098	0.157	0.583	0.113	PVP01_1109000; PVP01_1109100; PVP01_1109200; PVP01_1109300;	conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; conserved Plasmodium protein
11	600001	650000	0.012	0.12	0.394	0.033	0.014	0.024	0.017	0.025	0.083	0.124	0.231	0.583	0.072	PVP01_1113900; PVP01_1114000; PVP01_1114100(CCT4); PVP01_1114200(UTP21); PVP01_1114300; PVP01_1114400; PVP01_1114500; PVP01_1114600; PVP01_1114700(GluRS); PVP01_1114800; PVP01_1114900; PVP01_1115000(PK5); PVP01_1115100(ARK3)	patatin-like phospholipase putative; pyrroline-5-carboxylate reductase putative; T-complex protein 1 delta subunit putative; U3 small nucleolar RNA-associated protein 21 putative; conserved protein unknown function; DNA helicase putative; conserved Plasmodium protein unknown function; dynein light chain putative; glutamate--tRNA ligase putative; elongation factor 1-alpha putative; elongation factor 1-alpha putative; protein kinase 5; serine/threonine protein kinase putative

11	650001	700000	0.012	0.114	0.394	0.033	0.014	0.034	0.011	0.035	0.08	0.116	0.243	0.582	0.072	PVP01_1115100(ARK3); PVP01_1115200; PVP01_1115300; PVP01_1115400; PVP01_1115500; PVP01_1115600; PVP01_1115700(TIM23); PVP01_1115800; PVP01_1115900; PVP01_1116000; PVP01_1116100(SF3B5); PVP01_1116200(NIF3)	serine/threonine protein kinase putative; DnaJ protein putative; regulator of chromosome condensation putative; conserved Plasmodium protein unknown function; phosphatase 2A regulatory subunit-related protein putative; ubiquitin-conjugating enzyme E2 putative; mitochondrial import inner membrane translocase subunit TIM23 putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; RWD domain-containing protein
12	1350001	1400000	0.011	0.111	0.415	0.033	0.011	0.019	0.009	0.018	0.089	0.039	0.233	0.583	0.122	PVP01_1233600; PVP01_1233700; PVP01_1233800; PVP01_1233900; PVP01_1234000; PVP01_1234100; PVP01_1234200(PDI-14); PVP01_1234300; PVP01_1234400; PVP01_1234500; PVP01_1234600(HDA1); PVP01_1234700(YIP1); PVP01_1234800(ISY1); PVP01_1234900; PVP01_1235000; PVP01_1235100	DnaJ protein putative; GTPase-activating protein putative; conserved Plasmodium protein unknown function; dihydroorotase putative; HSP20-like chaperone putative; DNA-directed RNA polymerase alpha subunit putative; protein disulfide-isomerase putative; conserved Plasmodium protein unknown function; M1-family alanyl aminopeptidase putative; conserved protein unknown function; histone deacetylase putative; protein transport protein YIP1 putative; pre-mRNA-splicing factor ISY1 putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function
12	1750001	1800000	0.011	0.128	0.4	0.033	0.014	0.029	0.016	0.041	0.095	0.073	0.451	0.584	0.104	PVP01_1242900; PVP01_1243000; PVP01_1243100(CYC1); PVP01_1243200; PVP01_1243300; PVP01_1243400; PVP01_1243500; PVP01_1243600(RFC3); PVP01_1243700; PVP01_1243800(MED20); PVP01_1243900	EF-hand calcium-binding domain-containing protein putative; conserved Plasmodium protein unknown function; cyclin 1 putative; conserved Plasmodium protein unknown function; fam-a protein; DNA-directed RNA polymerase III subunit RPC4 putative; DNA polymerase alpha subunit B putative; replication factor C subunit 3 putative; conserved Plasmodium protein unknown function; mediator of RNA polymerase II transcription subunit 20 putative; conserved Plasmodium protein unknown function
12	1950001	2000000	0.01	0.108	0.743	0.033	0.013	0.02	0.015	0.018	0.086	0.043	0.2	0.582	0.092	PVP01_1247900(DBP5); PVP01_1248000(TVP23); PVP01_1248100; PVP01_1248200; PVP01_1248300; PVP01_1248400(SAS4); PVP01_1248500(ADCL); PVP01_1248600; PVP01_1248700; PVP01_1248800(PET117); PVP01_1248900; PVP01_1249000; PVP01_1249100	ATP-dependent RNA helicase DBP5 putative; golgi apparatus membrane protein TVP23 putative; DNA-directed RNA polymerase III subunit RPC5 putative; exonuclease V mitochondrial putative; zinc finger protein putative; spindle assembly abnormal protein 4 putative; aminodeoxychorismate lyase putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; protein PET117 putative; vivapain-1; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function
12	2000001	2050000	0.008	0.116	0.716	0.033	0.012	0.027	0.019	0.031	0.08	0.046	0.203	0.583	0.073	PVP01_1249200; PVP01_1249300; PVP01_1249400(VPS4); PVP01_1249500; PVP01_1249600; PVP01_1249700(TRX1); PVP01_1249800; PVP01_1249900(SPP); PVP01_1250000; PVP01_1250100(VP1); PVP01_1250200	large ribosomal subunit nuclear export factor putative; conserved Plasmodium protein unknown function; vacuolar protein sorting-associated protein 4 putative; conserved Plasmodium protein unknown function; MA3 domain-containing protein putative; thioredoxin 1 putative; conserved Plasmodium protein unknown function; signal peptide peptidase putative; conserved protein unknown function; V-type H(+)-translocating pyrophosphatase putative; conserved Plasmodium protein unknown function
12	2350001	2400000	0.011	0.113	0.631	0.033	0.009	0.03	0.012	0.037	0.085	0.041	0.181	0.583	0.114	PVP01_1257300(RRP6); PVP01_1257400; PVP01_1257500(ApiAP2); PVP01_1257600(MUS81); PVP01_1257700(BRF1); PVP01_1257800; PVP01_1257900; PVP01_1258000(GEST); PVP01_1258100; PVP01_1258200; PVP01_1258300	exosome complex exonuclease RRP6 putative; conserved protein unknown function; AP2 domain transcription factor AP2-O5 putative; crossover junction endonuclease MUS81 putative; transcription factor IIIb subunit putative; conserved Plasmodium protein unknown function; CLASP domain-containing protein putative; gamete egress and sporozoite traversal protein putative; conserved protein unknown function; Appr-1-p processing domain protein; HSP20-like chaperone putative
12	2450001	2500000	0.01	0.106	0.647	0.033	0.008	0.022	0.009	0.013	0.067	0.04	0.203	0.581	0.088	PVP01_1259200; PVP01_1259300; PVP01_1259400; PVP01_1259500; PVP01_1259600; PVP01_1259700; PVP01_1259800; PVP01_1259900; PVP01_1260000; PVP01_1260100; PVP01_1260200(HDP); PVP01_1260300; PVP01_1260400(CEN2); PVP01_1260500(NUP313); PVP01_1260600(pdhB)	calponin homology domain-containing protein putative; conserved protein unknown function; conserved protein unknown function; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; mitochondrial ribosomal protein S14 precursor putative; conserved protein unknown function; conserved Plasmodium protein unknown function; 40S ribosomal protein S5 putative; glutaminyl-peptide cyclotransferase putative; heme detoxification protein putative; conserved Plasmodium protein unknown function; centrin-2 putative; nucleoporin NUP313 putative; pyruvate dehydrogenase E1 component subunit beta putative
12	2500001	2550000	0.004	0.104	0.659	0.033	0.012	0.023	0.009	0.023	0.072	0.048	0.155	0.585	0.098	PVP01_1260700; PVP01_1260800(LAP); PVP01_1260900; PVP01_1261000; PVP01_1261100(DDX5); PVP01_1261200; PVP01_1261300; PVP01_1261400; PVP01_1261500; PVP01_1261600(CLK1); PVP01_1261700; PVP01_1261800(MAK5)	conserved Plasmodium protein unknown function; M17 leucyl aminopeptidase putative; conserved Plasmodium protein unknown function; U5 spliceosomal RNA; ATP-dependent RNA helicase DDX5 putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; RNA-binding protein putative; conserved Plasmodium protein unknown function; protein serine/threonine kinase-1 putative; mitochondrial ribosomal protein S29 precursor putative; ATP-dependent RNA helicase MAK5 putative

12	2550001	2600000	0.007	0.104	0.656	0.033	0.014	0.026	0.015	0.029	0.082	0.049	0.178	0.581	0.099	PVP01_1261800(MAK5); PVP01_1261900; PVP01_1262000; PVP01_1262100; PVP01_1262200; PVP01_1262300; PVP01_1262400; PVP01_1262500(eIK1); PVP01_1262600; PVP01_1262700(LPAAT); PVP01_1262800; PVP01_1262900	ATP-dependent RNA helicase MAK5 putative; histidine--tRNA ligase putative; prolyl hydroxylase-like protein putative; conserved protein unknown function; fructose 16-bisphosphate aldolase putative; shikimate dehydrogenase putative; conserved Plasmodium protein unknown function; eukaryotic translation initiation factor 2-alpha kinase 1 putative; conserved Plasmodium protein unknown function; 1-acyl-sn-glycerol-3-phosphate acyltransferase putative; calmodulin-like protein; conserved Plasmodium protein unknown function
12	2600001	2650000	0.008	0.107	0.625	0.033	0.025	0.021	0.016	0.036	0.068	0.059	0.281	0.582	0.096	PVP01_1262900; PVP01_1263000; PVP01_1263100(HSP90); PVP01_1263200(CWC24); PVP01_1263300(DPCK); PVP01_1263400; PVP01_1263500(PARN); PVP01_1263600(WLP1); PVP01_1263700(LSM5); PVP01_1263800; PVP01_1263900; PVP01_1264000(SRPK2); PVP01_1264100(SEC7)	conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; heat shock protein 90 putative; pre-mRNA-splicing factor CWC24 putative; dephospho-CoA kinase putative; gamma-tubulin complex component putative; poly(A)-specific ribonuclease PARN putative; WD repeat-containing protein putative; U6 snRNA-associated Sm-like protein LSm5 putative; heptatricopeptide repeat-containing protein putative; conserved Plasmodium protein unknown function; serine/threonine protein kinase putative; protein transport protein SEC7 putative
12	2800001	2850000	0.005	0.116	0.402	0.033	0.009	0.014	0.009	0.028	0.071	0.157	0.335	0.581	0.109	PVP01_1267200; PVP01_1267300; PVP01_1267400; PVP01_1267500(ORP2); PVP01_1267600; PVP01_1267700; PVP01_1267800; PVP01_1267900; PVP01_1268000(CTR1); PVP01_1268100(TPx1)	vesicle-associated membrane protein putative; leucine carboxyl methyltransferase putative; cytochrome c oxidase subunit ApiCOX26 putative; oocyst rupture protein 2 putative; cytochrome b-c1 complex subunit Rieske putative; Sad1/UNC domain-containing protein putative; conserved Plasmodium protein unknown function; DEAD/DEAH box helicase putative; copper transporter putative; thioredoxin peroxidase 1 putative
12	2850001	2900000	0.007	0.104	0.399	0.04	0.009	0.017	0.008	0.017	0.071	0.133	0.334	0.586	0.126	PVP01_1268200; PVP01_1268300; PVP01_1268400; PVP01_1268500(CPSF3); PVP01_1268600(MCA2); PVP01_1268700; PVP01_1268800; PVP01_1268900; PVP01_1269000; PVP01_1269100	OST-HTH associated domain protein putative; DNA primase small subunit putative; conserved Plasmodium protein unknown function; cleavage and polyadenylation specificity factor subunit 3 putative; metacaspase-2 putative; tRNA Methionine; tRNA Selenocysteine; translocation protein SEC62 putative; 50S ribosomal protein L20 putative; conserved Plasmodium protein unknown function
12	2950001	3000000	0.013	0.105	0.384	0.033	0.024	0.038	0.023	0.054	0.094	0.082	0.337	0.581	0.128	PVP01_1270300(CRMP4); PVP01_1270400(CCp1); PVP01_1270500; PVP01_1270600; PVP01_1270700; PVP01_1270800; PVP01_1270900	cysteine repeat modular protein 4 putative; LCCL domain-containing protein; bromodomain protein putative; tubulin epsilon chain putative; conserved Plasmodium protein unknown function; KELT protein; hypothetical protein
13	1650001	1700000	0.01	0.141	0.389	0.033	0.028	0.021	0.008	0.025	0.084	0.097	0.16	0.582	0.093	PVP01_1338200; PVP01_1338300(eIF2gamma); PVP01_1338400; PVP01_1338500(RAP1); PVP01_1338600; PVP01_1338700(CTPS); PVP01_1338800; PVP01_1338900(EMC2); PVP01_1339000(CDS)	conserved protein unknown function; eukaryotic translation initiation factor 2 subunit gamma putative; conserved Plasmodium protein unknown function; rhoptry-associated protein 1; WD repeat-containing protein putative; CTP synthase putative; alpha/beta hydrolase putative; ER membrane protein complex subunit 2 putative; cytidine diphosphate-diacylglycerol synthase putative
14	50001	100000	0.015	0.113	0.473	0.033	0.007	0.055	0.007	0.06	0.1	0.093	0.248	0.578	0.074	PVP01_1401300; PVP01_1401400; PVP01_1401500; PVP01_1401600; PVP01_1401700; PVP01_1401800(TRAG21); PVP01_1401900; PVP01_1402000; PVP01_1402100	Plasmodium exported protein unknown function; cytoadherence linked asexual protein CLAG putative; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; tryptophan-rich protein; lysophospholipase putative; hypothetical protein; Plasmodium exported protein unknown function
14	750001	800000	0.072	0.131	0.359	0.033	0.019	0.021	0.01	0.051	0.067	0.051	0.239	0.553	0.083	PVP01_1417100; PVP01_1417200; PVP01_1417300(SAS6L); PVP01_1417400(PRP40); PVP01_1417500(CCT); PVP01_1417600; PVP01_1417700(SEC20); PVP01_1417800; PVP01_1417900(USP39); PVP01_1418000(MCM4); PVP01_1418100(ApiAP2)	ADP-ribosylation factor putative; conserved Plasmodium protein unknown function; SAS6-like protein putative; pre-mRNA-processing factor 40 putative; choline-phosphate cytidylyltransferase putative; conserved protein unknown function; protein transport protein SEC20 putative; conserved protein unknown function; U4/U6.U5 tri-snRNP-associated protein 2 putative; DNA replication licensing factor MCM4 putative; AP2 domain transcription factor AP2-G3 putative

14	800001	850000	0.068	0.131	0.364	0.033	0.013	0.019	0.01	0.04	0.069	0.051	0.292	0.576	0.071	PVP01_1418100(ApiAP2); PVP01_1418200; PVP01_1418300; PVP01_1418400; PVP01_1418500; PVP01_1418600; PVP01_1418700(RPS19); PVP01_1418800(NOC4); PVP01_1418900; PVP01_1419000(FD); PVP01_1419100(G3PAT); PVP01_1419200; PVP01_1419300(SMC2); PVP01_1419400; PVP01_1419500; PVP01_1419600	AP2 domain transcription factor AP2-G3 putative; conserved Plasmodium protein unknown function; zinc finger protein putative; conserved protein unknown function; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; 40S ribosomal protein S19 putative; nucleolar complex protein 4 putative; conserved protein unknown function; ferredoxin putative; glycerol-3-phosphate 1-O-acyltransferase putative; conserved Plasmodium protein unknown function; structural maintenance of chromosomes protein 2 putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; conserved protein unknown function
14	850001	900000	0.057	0.132	0.362	0.033	0.009	0.035	0.013	0.065	0.068	0.038	0.314	0.587	0.074	PVP01_1419700(SEC63); PVP01_1419800; PVP01_1419900; PVP01_1420000; PVP01_1420100; PVP01_1420200(TRM1); PVP01_1420300; PVP01_1420400; PVP01_1420500; PVP01_1420600	translocation protein SEC63 putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; ubiquitin carboxyl-terminal hydrolase MINDY putative; conserved Plasmodium protein unknown function; tRNA (guanine(26)-N(2))-dimethyltransferase putative; conserved Plasmodium protein unknown function; conserved protein unknown function; ACDC domain-containing protein putative; protein phosphatase PP2A regulatory subunit A putative
14	1150001	1200000	0.021	0.144	0.373	0.033	0.008	0.056	0.01	0.022	0.114	0.065	0.131	0.583	0.084	PVP01_1427000; PVP01_1427100; PVP01_1427200; PVP01_1427300; PVP01_1427400; PVP01_1427500(SNRPC); PVP01_1427600; PVP01_1427700; PVP01_1427800; PVP01_1427900(SSP3); PVP01_1428000; PVP01_1428100; PVP01_1428200	conserved protein unknown function; CS domain protein putative; conserved Plasmodium protein unknown function; conserved protein unknown function; conserved protein unknown function; U1 small nuclear ribonucleoprotein C putative; ubiquitin-conjugating enzyme E2 putative; RNA-binding protein putative; karyopherin alpha putative; sporozoite surface protein 3 putative; trypsin-like serine protease putative; GTPase putative; proteasome activator complex subunit 4 putative
14	1200001	1250000	0.042	0.224	0.37	0.033	0.008	0.052	0.01	0.021	0.089	0.071	0.134	0.582	0.111	PVP01_1428200; PVP01_1428300; PVP01_1428400; PVP01_1428500; PVP01_1428600; PVP01_1428700; PVP01_1428800(CARM1); PVP01_1428900; PVP01_1429000(CAF1)	proteasome activator complex subunit 4 putative; conserved protein unknown function; RNA-binding protein putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; conserved protein unknown function; histone-arginine methyltransferase CARM1 putative; conserved protein unknown function; CCR4-associated factor 1 putative
14	1250001	1300000	0.067	0.198	0.374	0.039	0.009	0.089	0.016	0.049	0.133	0.077	0.221	0.582	0.101	PVP01_1429100(EMC1); PVP01_1429200; PVP01_1429300(CUL1); PVP01_1429400; PVP01_1429500(PPPK-DHPS); PVP01_1429600; PVP01_1429700(DBP1); PVP01_1429800(PPM7); PVP01_1429900(AQP2); PVP01_1430000(PPM5); PVP01_1430100(ABCK1)	ER membrane protein complex subunit 1 putative; mitochondrial carrier protein putative; cullin-1 putative; conserved Plasmodium protein unknown function; hydroxymethyldihydropterin pyrophosphokinase-dihydropteroate synthase putative; conserved Plasmodium protein unknown function; ATP-dependent RNA helicase DBP1 putative; protein phosphatase PPM7 putative; aquaporin putative; protein phosphatase PPM5 putative; ABC1 family putative
14	1300001	1350000	0.078	0.181	0.513	0.042	0.007	0.053	0.011	0.063	0.141	0.051	0.253	0.582	0.076	PVP01_1430100(ABCK1); PVP01_1430200; PVP01_1430300; PVP01_1430400(JmjC1); PVP01_1430500; PVP01_1430600(RUVB1); PVP01_1430700; PVP01_1430800; PVP01_1430900	ABC1 family putative; 50S ribosomal protein L33 putative; acyl-CoA binding protein putative; JmjC domain-containing protein putative; conserved Plasmodium protein unknown function; RuvB-like helicase 1 putative; peptidase family C50 putative; histone acetyltransferase subunit NuA4 putative; conserved protein unknown function
14	1400001	1450000	0.022	0.143	0.583	0.059	0.009	0.023	0.015	0.028	0.074	0.085	0.203	0.585	0.092	PVP01_1431900; PVP01_1432000(MED18); PVP01_1432100; PVP01_1432200; PVP01_1432300; PVP01_1432400(DLC8); PVP01_1432500(RPB8); PVP01_1432600(PRS); PVP01_1432700; PVP01_1432800; PVP01_1432900(PIGO); PVP01_1433000(SET5); PVP01_1433100; PVP01_1433200(EMC6); PVP01_1433300; PVP01_1433400; PVP01_1433500; PVP01_1433600; PVP01_1433700	U1 spliceosomal RNA; mediator of RNA polymerase II transcription subunit 18 putative; conserved Plasmodium protein unknown function; kelch domain-containing protein putative; integral membrane protein GPR180 putative; dynein light chain 1 putative; DNA-directed RNA polymerases I II and III subunit RPABC3 putative; proline--tRNA ligase putative; conserved Plasmodium protein unknown function; conserved protein unknown function; GPI ethanolamine phosphate transferase 3 putative; histone-lysine N-methyltransferase putative; geranylgeranyl transferase type-2 subunit beta putative; ER membrane protein complex subunit 6 putative; conserved Plasmodium protein unknown function; adrenodoxin-type ferredoxin putative; conserved protein unknown function; conserved Plasmodium protein unknown function; conserved protein unknown function
14	1450001	1500000	0.015	0.142	0.439	0.054	0.014	0.023	0.013	0.027	0.079	0.081	0.202	0.583	0.075	PVP01_1433700; PVP01_1433800(Trx-Px2); PVP01_1433900; PVP01_1434000(CYP32); PVP01_1434100; PVP01_1434200; PVP01_1434300; PVP01_1434400; PVP01_1434500; PVP01_1434600(RSPH9); PVP01_1434700(SR10); PVP01_1434800; PVP01_1434900; PVP01_1435000; PVP01_1435100(SRP19); PVP01_1435200; PVP01_1435300(MDV1); PVP01_1435400(CelTOS); PVP01_1435500(PLP2)	conserved protein unknown function; thioredoxin peroxidase 2 putative; SUN domain-containing protein putative; peptidyl-prolyl cis-trans isomerase putative; 10 kDa chaperonin putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; conserved protein unknown function; conserved Plasmodium protein unknown function; radial spoke head protein 9 putative; serpentine receptor putative; serine--tRNA ligase putative; conserved protein unknown function; glycerol-3-phosphate dehydrogenase [NAD(+)] putative; signal recognition particle subunit SRP19 putative; conserved Plasmodium protein unknown function; male development gene 1 putative; cell traversal protein for ookinetes and sporozoites; perforin-like protein 2

14	2100001	2150000	0.014	0.177	0.402	0.033	0.012	0.018	0.018	0.031	0.122	0.029	0.28	0.589	0.106	PVP01_1448200; PVP01_1448300(TLAP1); PVP01_1448400; PVP01_1448500(SPM2); PVP01_1448600(ClpQ); PVP01_1448700; PVP01_1448800; PVP01_1448900(SEC13); PVP01_1449000; PVP01_1449100(RIO1); PVP01_1449200; PVP01_1449300(RAB2); PVP01_1449400; PVP01_1449500	TBC domain-containing protein putative; thioredoxin-like associated protein 1 putative; vacuolar transporter chaperone putative; subpellicular microtubule protein 2 putative; ATP-dependent protease subunit ClpQ putative; WD repeat-containing protein putative; sun-family protein putative; protein transport protein SEC13 putative; pre-mRNA-splicing regulator putative; serine/threonine protein kinase RIO1 putative; conserved Plasmodium protein unknown function; ras-related protein Rab-2 putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function
14	2550001	2600000	0.01	0.1	0.359	0.043	0.009	0.013	0.006	0.014	0.105	0.05	0.224	0.582	0.091	PVP01_1458400; PVP01_1458500; PVP01_1458600; PVP01_1458700; PVP01_1458800(RFC4); PVP01_1458900(DBP9); PVP01_1459000; PVP01_1459100(QSOX); PVP01_1459200; PVP01_1459300; PVP01_1459400; PVP01_1459500; PVP01_1459600; PVP01_1459700; PVP01_1459800	ribosome-binding factor A putative; RNA-binding protein putative; conserved Plasmodium protein unknown function; mitochondrial carrier protein putative; replication factor C subunit 4 putative; ATP-dependent RNA helicase DBP9 putative; tetratricopeptide repeat protein putative; sulfhydryl oxidase putative; conserved Plasmodium protein unknown function; queuine tRNA-ribosyltransferase putative; conserved Plasmodium protein unknown function; GAS8-like protein putative; conserved protein unknown function; protein farnesyltransferase subunit alpha putative; 40S ribosomal protein S17 putative
14	2850001	2900000	0.007	0.112	0.43	0.208	0.008	0.013	0.007	0.021	0.09	0.049	0.182	0.584	0.1	PVP01_1466900(THO2); PVP01_1467000(KsgA2); PVP01_1467100(FCF2); PVP01_1467200(G377); PVP01_1467300; PVP01_1467400(VPS26); PVP01_1467500; PVP01_1467600; PVP01_1467700; PVP01_1467800; PVP01_1467900; PVP01_1468000	THO complex subunit 2 putative; ribosomal RNA small subunit methyltransferase A2 putative; rRNA-processing protein FCF2 putative; osmiophilic body protein G377 putative; CSC1-like protein putative; vacuolar protein sorting-associated protein 26 putative; conserved Plasmodium protein unknown function; AP-3 complex subunit sigma putative; translation initiation factor eIF-2B subunit beta putative; transcription initiation factor IIA subunit 2 putative; DNA repair protein rhp16 putative; conserved protein unknown function
14	2900001	2950000	0.005	0.109	0.492	0.22	0.008	0.016	0.013	0.021	0.112	0.043	0.189	0.582	0.088	PVP01_1468000; PVP01_1468100; PVP01_1468200; PVP01_1468300; PVP01_1468400(TMK); PVP01_1468500; PVP01_1468600(DRS1); PVP01_1468700; PVP01_1468800(aTrpRS); PVP01_1468900; PVP01_1469000; PVP01_1469100; PVP01_1469200(RON3); PVP01_1469300	conserved protein unknown function; HSP20-like chaperone putative; conserved Plasmodium protein unknown function; coronin putative; thymidylate kinase putative; conserved Plasmodium protein unknown function; ATP-dependent RNA helicase DRS1 putative; conserved Plasmodium protein unknown function; tryptophan--tRNA ligase putative; conserved Plasmodium protein unknown function; tRNA Valine; tRNA Glutamine; rhoptry neck protein 3 putative; conserved Plasmodium protein unknown function

S11 Table. Highest 1% of pairwise Identity by Descent (IBD) fractions across clades within the monoclonal isolates of Brazil

Chr	Window start	Window end	C1	C6	C7	Genes	Gene products
3	600001	650000	0.324	0.344	0.023	PVP01_0313800; PVP01_0313900(EST); PVP01_0314000(CRMP2); PVP01_0314100; PVP01_0314200; PVP01_0314300	dynein heavy chain putative; exported serine/threonine protein kinase putative; cysteine repeat modular protein 2 putative; 30S ribosomal protein S8 putative; prefoldin subunit 3 putative; conserved Plasmodium protein unknown function
5	250001	300000	0.327	0.325	0.026	PVP01_0505500; PVP01_0505600(GAMA); PVP01_0505700; PVP01_0505800(FT1); PVP01_0505900; PVP01_0506000(OXA1); PVP01_0506100; PVP01_0506200; PVP01_0506300; PVP01_0506400(ROM3); PVP01_0506500(PDI8); PVP01_0506600(SET3); PVP01_0506700	conserved protein unknown function; GPI-anchored micronemal antigen; conserved protein unknown function; folate transporter 1 putative; translation initiation factor eIF-2B subunit alpha putative; mitochondrial inner membrane protein OXA1 putative; conserved protein unknown function; leucine--tRNA ligase putative; conserved Plasmodium protein unknown function; rhomboid protease ROM3 putative; protein disulfide isomerase putative; SET domain protein putative; magnesium transporter putative
7	250001	300000	0.407	0.182	0.078	PVP01_0704200; PVP01_0704300(SEC27); PVP01_0704400; PVP01_0704500(RNaseII); PVP01_0704600(VPS46); PVP01_0704700; PVP01_0704800; PVP01_0704900; PVP01_0705000; PVP01_0705100	conserved Plasmodium protein unknown function; coatomer subunit beta putative; endonuclease/exonuclease/phosphatase domain containing protein; exoribonuclease II putative; vacuolar protein sorting-associated protein 46 putative; conserved Plasmodium protein unknown function; Maf-like protein putative; dynein intermediate light chain putative; arginase putative; zinc finger protein putative
9	1400001	1450000	0.723	0.237	0.024	PVP01_0932700; PVP01_0932800; PVP01_0932900; PVP01_0933000; PVP01_0933100; PVP01_0933200; PVP01_0933300; PVP01_0933400(PRP38A); PVP01_0933500; PVP01_0933600(AQP); PVP01_0933700(GCVH); PVP01_0933800; PVP01_0933900; PVP01_0934000	conserved Plasmodium protein unknown function; ubiquitin-like protein putative; RNA-binding protein putative; T-complex protein 1 subunit alpha putative; nucleic acid binding protein putative; conserved Plasmodium protein unknown function; amino acid transporter putative; pre-mRNA-splicing factor 38A putative; 50S ribosomal protein L2 putative; aquaglyceroporin putative; glycine cleavage system H protein putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function
9	1450001	1500000	0.692	0.237	0.052	PVP01_0934000; PVP01_0934100(CDC50B); PVP01_0934200(AMA1); PVP01_0934300; PVP01_0934400; PVP01_0934500; PVP01_0934600; PVP01_0934700; PVP01_0934800(HSP70-3); PVP01_0934900(PDI-11); PVP01_0935000; PVP01_0935100	conserved Plasmodium protein unknown function; LEM3/CDC50 family protein putative; apical membrane antigen 1; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; FHA domain-containing protein putative; RNA (uracil-5-)methyltransferase putative; conserved Plasmodium protein unknown function; heat shock protein 70 putative; protein disulfide isomerase putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function
11	1000001	1050000	0.511	0.332	0.042	PVP01_1122600; PVP01_1122700; PVP01_1122800; PVP01_1122900; PVP01_1123000; PVP01_1123100; PVP01_1123200; PVP01_1123300; PVP01_1123400; PVP01_1123500; PVP01_1123600; PVP01_1123700; PVP01_1123800; PVP01_1123900(PAP); PVP01_1124000	mitochondrial ribosomal protein L46 precursor putative; pyruvate kinase putative; ATPase putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; CRAL/TRIO domain-containing protein putative; 3-oxoacyl-[acyl-carrier-protein] synthase i/ii putative; conserved Plasmodium protein unknown function; putative oxidoreductase short-chain dehydrogenase family putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; conserved protein unknown function; poly(A) polymerase PAP putative; conserved Plasmodium protein unknown function
12	2350001	2400000	0.368	0.33	0.04	PVP01_1257300(RRP6); PVP01_1257400; PVP01_1257500(ApiAP2); PVP01_1257600(MUS81); PVP01_1257700(BRF1); PVP01_1257800; PVP01_1257900; PVP01_1258000(GEST); PVP01_1258100; PVP01_1258200; PVP01_1258300	exosome complex exonuclease RRP6 putative; conserved protein unknown function; AP2 domain transcription factor AP2-O5 putative; crossover junction endonuclease MUS81 putative; transcription factor IIIb subunit putative; conserved Plasmodium protein unknown function; CLASP domain-containing protein putative; gamete egress and sporozoite traversal protein putative; conserved protein unknown function; Appr-1-p processing domain protein; HSP20-like chaperone putative
13	1450001	1500000	0.481	0.329	0.032	PVP01_1333600; PVP01_1333700; PVP01_1333800; PVP01_1333900(UDG); PVP01_1334000; PVP01_1334100(SMB1); PVP01_1334200; PVP01_1334300(Pgt1); PVP01_1334400(ABCk2); PVP01_1334500(PP1); PVP01_1334600	RNA-binding protein Nova-1 putative; DNA-directed RNA polymerases I and III subunit RPAC2 putative; conserved protein unknown function; uracil-DNA glycosylase putative; ATP-dependent protease putative; small nuclear ribonucleoprotein-associated protein B putative; ubiquitin carboxyl-terminal hydrolase putative; mRNA-capping enzyme subunit alpha putative; atypical protein kinase ABC-1 family putative; serine/threonine protein phosphatase PP1 putative; 60S ribosomal protein L10 putative

14	1250001	1300000	0.246	0.26	0.134	PVP01_1429100(EMC1); PVP01_1429200; PVP01_1429300(CUL1); PVP01_1429400; PVP01_1429500(PPPK-DHPS); PVP01_1429600; PVP01_1429700(DBP1); PVP01_1429800(PPM7); PVP01_1429900(AQP2); PVP01_1430000(PPM5); PVP01_1430100(ABCK1)	ER membrane protein complex subunit 1 putative; mitochondrial carrier protein putative; cullin-1 putative; conserved Plasmodium protein unknown function; hydroxymethyldihydropterin pyrophosphokinase-dihydropteroate synthase putative; conserved Plasmodium protein unknown function; ATP-dependent RNA helicase DBP1 putative; protein phosphatase PPM7 putative; aquaporin putative; protein phosphatase PPM5 putative; ABC1 family putative
14	1300001	1350000	0.27	0.264	0.116	PVP01_1430100(ABCK1); PVP01_1430200; PVP01_1430300; PVP01_1430400(JmjC1); PVP01_1430500; PVP01_1430600(RUVB1); PVP01_1430700; PVP01_1430800; PVP01_1430900	ABC1 family putative; 50S ribosomal protein L33 putative; acyl-CoA binding protein putative; JmjC domain-containing protein putative; conserved Plasmodium protein unknown function; RuvB-like helicase 1 putative; peptidase family C50 putative; histone acetyltransferase subunit NuA4 putative; conserved protein unknown function
14	2100001	2150000	0.578	0.241	0.076	PVP01_1448200; PVP01_1448300(TLAP1); PVP01_1448400; PVP01_1448500(SPM2); PVP01_1448600(ClpQ); PVP01_1448700; PVP01_1448800; PVP01_1448900(SEC13); PVP01_1449000; PVP01_1449100(RIO1); PVP01_1449200; PVP01_1449300(RAB2); PVP01_1449400; PVP01_1449500	TBC domain-containing protein putative; thioredoxin-like associated protein 1 putative; vacuolar transporter chaperone putative; subpellicular microtubule protein 2 putative; ATP-dependent protease subunit ClpQ putative; WD repeat-containing protein putative; sun-family protein putative; protein transport protein SEC13 putative; pre-mRNA-splicing regulator putative; serine/threonine protein kinase RIO1 putative; conserved Plasmodium protein unknown function; ras-related protein Rab-2 putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function
14	2150001	2200000	0.567	0.311	0.109	PVP01_1449500; PVP01_1449600; PVP01_1449700; PVP01_1449800(PRP2); PVP01_1449900; PVP01_1450000; PVP01_1450100; PVP01_1450200(aFRS); PVP01_1450300(CPN60); PVP01_1450400(LPD1)	conserved Plasmodium protein unknown function; amino acid transporter putative; mitosis protein dim1 putative; pre-mRNA-splicing factor ATP-dependent RNA helicase PRP2 putative; conserved protein unknown function; conserved Plasmodium protein unknown function; WD repeat-containing protein putative; phenylalanine--tRNA ligase putative; 60 kDa chaperonin putative; dihydrolipoyl dehydrogenase mitochondrial putative
14	2350001	2400000	0.677	0.235	0.035	PVP01_1453900; PVP01_1454000(XAB2); PVP01_1454100; PVP01_1454200(CARP); PVP01_1454300; PVP01_1454400; PVP01_1454500; PVP01_1454600; PVP01_1454700; PVP01_1454800; PVP01_1454900; PVP01_1455000; PVP01_1455100(UBA2); PVP01_1455200	conserved Plasmodium protein unknown function; pre-mRNA-splicing factor SYF1 putative; vesicle transport v-SNARE protein VTI1 putative; clustered-asparagine-rich protein putative; conserved Plasmodium protein unknown function; conserved protein unknown function; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; p25-alpha family protein putative; S-adenosyl-methyltransferase mraW putative; protein-S-isoprenylcysteine O-methyltransferase putative; conserved Plasmodium protein unknown function; SUMO-activating enzyme subunit 2 putative; conserved Plasmodium protein unknown function
14	2400001	2450000	0.695	0.234	0.022	PVP01_1455200; PVP01_1455300; PVP01_1455400; PVP01_1455500; PVP01_1455600; PVP01_1455700(PWP1); PVP01_1455800; PVP01_1455900(HUB1); PVP01_1456000; PVP01_1456100; PVP01_1456200; PVP01_1456300; PVP01_1456400(CWC22); PVP01_1456500; PVP01_1456600	conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; hydrolase putative; conserved Plasmodium protein unknown function; periodic tryptophan protein 1 putative; conserved protein unknown function; ubiquitin-like modifier HUB1 putative; conserved Plasmodium protein unknown function; COPI associated protein putative; calcyclin binding protein putative; conserved Plasmodium protein unknown function; pre-mRNA-splicing factor CWC22 putative; conserved protein unknown function; conserved Plasmodium protein unknown function
14	2450001	2500000	0.691	0.248	0.021	PVP01_1456600; PVP01_1456700; PVP01_1456800; PVP01_1456900(ACS11); PVP01_1457000(PK2); PVP01_1457100; PVP01_1457200; PVP01_1457300(ApiAP2); PVP01_1457400	conserved Plasmodium protein unknown function; sphingomyelin phosphodiesterase putative; BTB/POZ domain-containing protein putative; acyl-CoA synthetase putative; protein kinase 2 putative; HD superfamily phosphohydrolase protein putative; 50S ribosomal protein L23 putative; AP2 domain transcription factor putative; conserved Plasmodium protein unknown function

S12 Table. Highest 1% of pairwise Identity by Descent (IBD) fractions across population groups (A and B) within the monoclonal isolates of Brazil

Chr	Window start	Window end	West Central Brazil / A	Northern Brazil / B	Genes	Gene products
2	450001	500000	0.199	0.032	PVP01_0211100; PVP01_0211200; PVP01_0211300; PVP01_0211400; PVP01_0211500(NT4); PVP01_0211600(VPS51); PVP01_0211700; PVP01_0211800; PVP01_0211900; PVP01_0212000(PMS1); PVP01_0212100	ATP-dependent RNA helicase putative; conserved Plasmodium protein unknown function; zinc-carboxypeptidase putative; conserved protein unknown function; nucleoside transporter 4; vacuolar protein sorting-associated protein 51 putative; vacuolar protein sorting-associated protein VTA1 putative; aspartate--tRNA ligase putative; seipin domain-containing protein putative; DNA mismatch repair protein PMS1 putative; conserved Plasmodium membrane protein unknown function
2	550001	600000	0.189	0.021	PVP01_0213100; PVP01_0213200; PVP01_0213300; PVP01_0213400; PVP01_0213500; PVP01_0213600; PVP01_0213700(eIF2alpha); PVP01_0213800; PVP01_0213900(ALP3)	proteasome subunit alpha type-5 putative; mTERF domain-containing protein putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; cation transporting ATPase putative; conserved Plasmodium protein unknown function; eukaryotic translation initiation factor 2 subunit alpha putative; conserved Plasmodium membrane protein unknown function; actin-like protein putative
5	800001	850000	0.198	0.036	PVP01_0519500(VPS9); PVP01_0519600(Ub); PVP01_0519700(EIF3G); PVP01_0519800; PVP01_0519900; PVP01_0520000(PGM2); PVP01_0520100(RPT3); PVP01_0520200; PVP01_0520300; PVP01_0520400(USP13); PVP01_0520500(SMC3); PVP01_0520600; PVP01_0520700	vacuolar protein sorting-associated protein 9 putative; ubiquitin putative; eukaryotic translation initiation factor 3 subunit G putative; conserved Plasmodium protein unknown function; ATP synthase-associated protein putative; phosphoglucomutase-2 putative; 26S protease regulatory subunit 6B putative; lysine decarboxylase-like protein putative; 50S ribosomal protein L10 putative; ubiquitin carboxyl-terminal hydrolase 13 putative; structural maintenance of chromosomes protein 3 putative; conserved Plasmodium protein unknown function; calmodulin-like protein
5	850001	900000	0.198	0.033	PVP01_0520700; PVP01_0520800; PVP01_0520900; PVP01_0521000; PVP01_0521100(APH); PVP01_0521200; PVP01_0521300; PVP01_0521400(ARO); PVP01_0521500; PVP01_0521600; PVP01_0521700; PVP01_0521800(CRK3); PVP01_0521900; PVP01_0522000(CBP20); PVP01_0522100(GAK); PVP01_0522200; PVP01_0522300	calmodulin-like protein; Rab5-interacting protein putative; conserved Plasmodium protein unknown function; RNA-binding protein putative; acylated pleckstrin-homology domain-containing protein putative; GTP-binding protein putative; conserved Plasmodium protein unknown function; armadillo-domain containing rhoptry protein putative; arsenical pump-driving ATPase putative; conserved protein unknown function; conserved Plasmodium protein unknown function; cdc2-related protein kinase 3 putative; coatomer subunit zeta putative; nuclear cap-binding protein subunit 2 putative; GTP:AMP phosphotransferase putative; conserved Plasmodium protein unknown function; RING zinc finger protein putative
5	900001	950000	0.193	0.044	PVP01_0522400; PVP01_0522500; PVP01_0522600; PVP01_0522700; PVP01_0522800; PVP01_0522900; PVP01_0523000; PVP01_0523100; PVP01_0523200; PVP01_0523300; PVP01_0523400	60S ribosomal protein L15 putative; RNA-binding protein putative; hypothetical protein; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function pseudogene; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function

7	250001	300000	0.109	0.129	PVP01_0704200; PVP01_0704300(SEC27); PVP01_0704400; PVP01_0704500(RNaseII); PVP01_0704600(VPS46); PVP01_0704700; PVP01_0704800; PVP01_0704900; PVP01_0705000; PVP01_0705100	conserved Plasmodium protein unknown function; coatomer subunit beta putative; endonuclease/exonuclease/phosphatase domain containing protein; exoribonuclease II putative; vacuolar protein sorting-associated protein 46 putative; conserved Plasmodium protein unknown function; Maf-like protein putative; dynein intermediate light chain putative; arginase putative; zinc finger protein putative
11	900001	950000	0.125	0.105	PVP01_1120900; PVP01_1121000; PVP01_1121100; PVP01_1121200(CEPT); PVP01_1121300(PK4); PVP01_1121400	DNA methyltransferase 1-associated protein 1 putative; conserved Plasmodium protein unknown function; protease putative; choline/ethanolaminephosphotransferase putative; eukaryotic translation initiation factor 2-alpha kinase putative; HECT-domain (ubiquitin-transferase) putative
13	950001	1000000	0.13	0.126	PVP01_1321000; PVP01_1321100(MFS5); PVP01_1321200(WBP11); PVP01_1321300; PVP01_1321400; PVP01_1321500(RRP41); PVP01_1321600; PVP01_1321700(MIT3); PVP01_1321800(MSH2-1); PVP01_1321900; PVP01_1322000; PVP01_1322100(Sel4)	proliferation-associated protein 2g4 putative; major facilitator superfamily domain-containing protein putative; WW domain-binding protein 11 putative; conserved Plasmodium protein unknown function; leucine-rich repeat protein; exosome complex component RRP41 putative; conserved Plasmodium protein unknown function; CorA-like Mg2+ transporter protein putative; DNA mismatch repair protein MSH2 putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; selenoprotein putative
14	1250001	1300000	0.142	0.179	PVP01_1429100(EMC1); PVP01_1429200; PVP01_1429300(CUL1); PVP01_1429400; PVP01_1429500(PPPK-DHPS); PVP01_1429600; PVP01_1429700(DBP1); PVP01_1429800(PPM7); PVP01_1429900(AQP2); PVP01_1430000(PPM5); PVP01_1430100(ABCK1)	ER membrane protein complex subunit 1 putative; mitochondrial carrier protein putative; cullin-1 putative; conserved Plasmodium protein unknown function; hydroxymethyldihydropterin pyrophosphokinase-dihydropteroate synthase putative; conserved Plasmodium protein unknown function; ATP-dependent RNA helicase DBP1 putative; protein phosphatase PPM7 putative; aquaporin putative; protein phosphatase PPM5 putative; ABC1 family putative
14	1300001	1350000	0.157	0.115	PVP01_1430100(ABCK1); PVP01_1430200; PVP01_1430300; PVP01_1430400(JmjC1); PVP01_1430500; PVP01_1430600(RUVB1); PVP01_1430700; PVP01_1430800; PVP01_1430900	ABC1 family putative; 50S ribosomal protein L33 putative; acyl-CoA binding protein putative; JmjC domain-containing protein putative; conserved Plasmodium protein unknown function; RuvB-like helicase 1 putative; peptidase family C50 putative; histone acetyltransferase subunit NuA4 putative; conserved protein unknown function

S13 Table. Candidate regions of positive selection across the global dataset using iHS

Chr	Start	End	Subpopulation	Number of markers	Mean iHS markers	Max iHS	Number of iHS extreme markers	Percentage of extreme iHS markers	Mean iHS of extreme markers	Genes in the region	Products
2	130000	160000	East Africa	109	0.676	5.341	2	1.83	4.722	PVP01_0202600; PVP01_0202700; PVP01_0202800; PVP01_0202900; PVP01_0203000(MRP1)	Plasmodium exported protein (PHIST) unknown function; 28S ribosomal RNA; 5.8S ribosomal RNA; 18S ribosomal RNA; multidrug resistance-associated protein 1 putative
3	240000	260000	East Africa	92	0.86	14.177	2	2.17	11.921	PVP01_0304600(MAg-1); PVP01_0304700(LISP2); PVP01_0304800(PRPF8)	ag-1 blood stage membrane protein homologue; liver specific protein 2 putative; pre-mRNA-processing-splicing factor 8 putative
4	750000	780000	East Africa	171	0.747	7.633	4	2.34	6.161	PVP01_0418100(IscA1); PVP01_0418200; PVP01_0418300(MSP4); PVP01_0418400(MSP5); PVP01_0418500(ADSL); PVP01_0418600; PVP01_0418700; PVP01_0418800(AROM)	iron-sulfur assembly protein putative; conserved Plasmodium protein unknown function; merozoite surface protein 4; merozoite surface protein 5; adenylosuccinate lyase; DNA-directed RNA polymerase III subunit RPC10 putative; conserved Plasmodium protein unknown function; pentafunctional AROM polypeptide putative
5	1000000	1030000	East Africa	73	0.709	5.717	4	5.48	5.056	PVP01_0524600; PVP01_0524700; PVP01_0524800; PVP01_0524900; PVP01_0525000; PVP01_0525100; PVP01_0525200; PVP01_0525300	Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function pseudogene; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function
5	1000000	1030000	South America	99	1.264	6.983	12	12.12	4.521	PVP01_0524600; PVP01_0524700; PVP01_0524800; PVP01_0524900; PVP01_0525000; PVP01_0525100; PVP01_0525200; PVP01_0525300	Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function pseudogene; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function
7	1200000	1230000	South America	306	1.03	9.875	10	3.27	6.379	PVP01_0728600; PVP01_0728700(LRR8); PVP01_0728800(MSP1P); PVP01_0728900(MSP1); PVP01_0729000; PVP01_0729100(DGK1)	heptatricopeptide repeat-containing protein putative; leucine-rich repeat protein; merozoite surface protein 1 paralog; merozoite surface protein 1; zinc finger protein putative; diacylglycerol kinase putative
1	560000	590000	South Asia	90	0.772	5.346	3	3.33	5.064	PVP01_0112300; PVP01_0112400; PVP01_0112500(TyrRS); PVP01_0112600(RPN10); PVP01_0112700(DegP); PVP01_0112800; PVP01_0112900; PVP01_0113000(COQ10)	AP-3 complex subunit delta putative; conserved Plasmodium protein unknown function; tyrosine--tRNA ligase putative; 26S proteasome regulatory subunit RPN10 putative; serine protease DegP putative; conserved Plasmodium protein unknown function; proteasome subunit alpha type-6 putative; coenzyme Q-binding protein COQ10 homolog mitochondrial putative
2	130000	160000	South Asia	110	0.653	6.879	3	2.73	6.415	PVP01_0202600; PVP01_0202700; PVP01_0202800; PVP01_0202900; PVP01_0203000(MRP1)	Plasmodium exported protein (PHIST) unknown function; 28S ribosomal RNA; 5.8S ribosomal RNA; 18S ribosomal RNA; multidrug resistance-associated protein 1 putative
3	240000	270000	South Asia	186	1.633	12.162	29	15.59	6.183	PVP01_0304600(MAg-1); PVP01_0304700(LISP2); PVP01_0304800(PRPF8); PVP01_0304900; PVP01_0305000	ag-1 blood stage membrane protein homologue; liver specific protein 2 putative; pre-mRNA-processing-splicing factor 8 putative; conserved Plasmodium protein unknown function; TMEM33 domain-containing protein putative
5	310000	340000	South Asia	102	0.493	4.607	2	1.96	4.467	PVP01_0507000; PVP01_0507100; PVP01_0507200; PVP01_0507300(IF2c); PVP01_0507400(DBP10); PVP01_0507500; PVP01_0507600(RACK1); PVP01_0507700(VAMP7); PVP01_0507800(UBE4B)	conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; translation initiation factor IF-2 putative; ATP-dependent RNA helicase DBP10 putative; conserved Plasmodium protein unknown function; receptor for activated c kinase putative; SNARE protein putative; ubiquitin conjugation factor E4 B putative
5	1000000	1030000	South Asia	130	0.701	9.028	2	1.54	7.083	PVP01_0524600; PVP01_0524700; PVP01_0524800; PVP01_0524900; PVP01_0525000; PVP01_0525100; PVP01_0525200; PVP01_0525300	Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function pseudogene; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function

5	1070000	1090000	South Asia	47	0.789	5.402	2	4.26	4.982	PVP01_0526400; PVP01_0526500(PUF2); PVP01_0526600(DHFR-TS); PVP01_0526700; PVP01_0526800	conserved Plasmodium protein unknown function; mRNA-binding protein PUF2 putative; bifunctional dihydrofolate reductase-thymidylate synthase putative; LETM1-like protein putative; conserved Plasmodium protein unknown function
6	170000	200000	South Asia	73	0.78	5.735	3	4.11	5.134	PVP01_0604300; PVP01_0604400; PVP01_0604500; PVP01_0604600; PVP01_0604700; PVP01_0604800	conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; zinc finger protein putative; 60S ribosomal protein L30e putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function
6	820000	840000	South Asia	203	0.637	4.769	2	0.99	4.485	PVP01_0620200; PVP01_0620300; PVP01_0620400; PVP01_0620500; PVP01_0620600; PVP01_0620700	conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; hypothetical protein; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; hypothetical protein
7	580000	640000	South Asia	142	0.538	4.5	4	2.82	4.182	PVP01_0712300(RRSapi); PVP01_0712400; PVP01_0712500; PVP01_0712600; PVP01_0712700; PVP01_0712800; PVP01_0712900; PVP01_0713000(ELF1); PVP01_0713100(MFR4); PVP01_0713200; PVP01_0713300; PVP01_0713400(NDH2); PVP01_0713500(UBC9); PVP01_0713600; PVP01_0713700; PVP01_0713800(PFK9)	arginine--tRNA ligase putative; pseudouridylate synthase putative; conserved Plasmodium protein unknown function; phospholipid or glycerol acyltransferase putative; met-10+ like protein putative; conserved protein unknown function; conserved Plasmodium protein unknown function; transcription elongation factor 1 putative; major facilitator superfamily-related transporter putative; GINS complex subunit Psf3 putative; BSD-domain protein putative; type II NADH:ubiquinone oxidoreductase putative; SUMO-conjugating enzyme UBC9 putative; ribonuclease H2 subunit C putative; heptatricopeptide repeat-containing protein putative; ATP-dependent 6-phosphofructokinase putative
7	820000	850000	South Asia	87	0.634	7.947	4	4.6	6.026	PVP01_0718500; PVP01_0718600; PVP01_0718700; PVP01_0718800; PVP01_0718900; PVP01_0719000; PVP01_0719100; PVP01_0719200; PVP01_0719300(PRP24)	chaperone putative; CS domain protein putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; ADP-ribosylation factor putative; conserved Plasmodium protein unknown function; CRAL/TRIO domain-containing protein putative; inosine-5'-monophosphate dehydrogenase putative; U4/U6 snRNA-associated-splicing factor putative
7	1020000	1050000	South Asia	101	0.435	4.066	2	1.98	4.042	PVP01_0723800(aPRS); PVP01_0723900; PVP01_0724000(TLP2); PVP01_0724100; PVP01_0724200(HDAC1); PVP01_0724300; PVP01_0724400(PV5); PVP01_0724500; PVP01_0724600	proline--tRNA ligase putative; protein phosphatase-beta putative; thioredoxin-like protein 2 putative; zinc binding protein (Yippee) putative; histone deacetylase 1 putative; armadillo repeat protein putative; parasitophorous vacuolar protein 5 putative; protein kinase putative; protein kinase putative
7	1200000	1230000	South Asia	301	0.685	5.926	4	1.33	4.654	PVP01_0728600; PVP01_0728700(LRR8); PVP01_0728800(MSP1P); PVP01_0728900(MSP1); PVP01_0729000; PVP01_0729100(DGK1)	heptatricopeptide repeat-containing protein putative; leucine-rich repeat protein; merozoite surface protein 1 paralog; merozoite surface protein 1; zinc finger protein putative; diacylglycerol kinase putative
8	380000	410000	South Asia	94	0.843	6.471	3	3.19	5.111	PVP01_0808500(NOP5); PVP01_0808600; PVP01_0808700; PVP01_0808800; PVP01_0808900(AK1); PVP01_0809000(DPH5)	nucleolar protein 5 putative; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; adenylate kinase putative; diphthine methyl ester synthase putative
8	460000	490000	South Asia	84	0.521	5.618	2	2.38	5.55	PVP01_0810400(SDH4); PVP01_0810500; PVP01_0810600; PVP01_0810700(eIF2beta); PVP01_0810800; PVP01_0810900; PVP01_0811000; PVP01_0811100(ISP1); PVP01_0811200; PVP01_0811300(BET1); PVP01_0811400(ARV1)	succinate dehydrogenase subunit 4 putative; MORN repeat protein putative; ankyrin-repeat protein putative; eukaryotic translation initiation factor 2 subunit beta putative; dolichyl-phosphate-mannose--protein mannosyltransferase putative; 50S ribosomal protein L22 mitochondrial putative; conserved Plasmodium protein unknown function; inner membrane complex sub-compartment protein 1 putative; conserved Plasmodium protein unknown function; BET1-like protein putative; protein ARV1 putative
2	110000	170000	Southeast Asia	197	1.026	10.426	8	4.06	5.742	PVP01_0202200(TRAG2); PVP01_0202300; PVP01_0202400; PVP01_0202500; PVP01_0202600; PVP01_0202700; PVP01_0202800; PVP01_0202900; PVP01_0203000(MRP1); PVP01_0203100; PVP01_0203200(TatD)	tryptophan-rich protein; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein (PHIST) unknown function; 28S ribosomal RNA; 5.8S ribosomal RNA; 18S ribosomal RNA; multidrug resistance-associated protein 1 putative; conserved Plasmodium protein unknown function; TatD-like deoxyribonuclease putative
3	240000	270000	Southeast Asia	238	1.129	8.125	5	2.1	5.795	PVP01_0304600(MAg-1); PVP01_0304700(LISP2); PVP01_0304800(PRPF8); PVP01_0304900; PVP01_0305000	ag-1 blood stage membrane protein homologue; liver specific protein 2 putative; pre-mRNA-processing-splicing factor 8 putative; conserved Plasmodium protein unknown function; TMEM33 domain-containing protein putative
4	750000	780000	Southeast Asia	248	0.97	9.896	9	3.63	6.043	PVP01_0418100(IscA1); PVP01_0418200; PVP01_0418300(MSP4); PVP01_0418400(MSP5); PVP01_0418500(ADSL); PVP01_0418600; PVP01_0418700; PVP01_0418800(AROM)	iron-sulfur assembly protein putative; conserved Plasmodium protein unknown function; merozoite surface protein 4; merozoite surface protein 5; adenylosuccinate lyase; DNA-directed RNA polymerase III subunit RPC10 putative; conserved Plasmodium protein unknown function; pentafunctional AROM polypeptide putative

5	1020000	1110000	Southeast Asia	323	0.917	10.425	10	3.1	5.833	PVP01_0525200; PVP01_0525300; PVP01_0525400; PVP01_0525500(GATA); PVP01_0525600; PVP01_0525700(MCM9); PVP01_0525800(HAT1); PVP01_0525900(MAF1); PVP01_0526000(PHBL); PVP01_0526100; PVP01_0526200(SAR1); PVP01_0526300; PVP01_0526400; PVP01_0526500(PUF2); PVP01_0526600(DHFR-TS); PVP01_0526700; PVP01_0526800; PVP01_0526900; PVP01_0527000	Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; glutamyl-tRNA(Gln) amidotransferase subunit A putative; conserved protein unknown function; DNA helicase MCM9 putative; histone acetyltransferase putative; repressor of RNA polymerase III transcription MAF1 putative; prohibitin-like protein PHBL putative; CDGSH iron-sulfur domain-containing protein putative; small GTP-binding protein sar1 putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; mRNA-binding protein PUF2 putative; bifunctional dihydrofolate reductase-thymidylate synthase putative; LETM1-like protein putative; conserved Plasmodium protein unknown function; memo-like protein; conserved Plasmodium protein unknown function
6	810000	840000	Southeast Asia	320	0.6	4.798	4	1.25	4.567	PVP01_0619900; PVP01_0620000(MRScyt); PVP01_0620100; PVP01_0620200; PVP01_0620300; PVP01_0620400; PVP01_0620500; PVP01_0620600; PVP01_0620700	conserved protein unknown function; methionine--tRNA ligase putative; U2 snRNA/tRNA pseudouridine synthase putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; hypothetical protein; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; hypothetical protein
6	960000	990000	Southeast Asia	126	0.604	6.11	3	2.38	5.358	PVP01_0623500; PVP01_0623600; PVP01_0623700; PVP01_0623800(DBP); PVP01_0623900(REX4)	Plasmodium exported protein unknown function; PIR protein; Plasmodium exported protein (PHIST) unknown function; duffy binding protein; ring-exported protein 4 putative
7	610000	630000	Southeast Asia	113	0.407	5.264	2	1.77	4.792	PVP01_0712900; PVP01_0713000(ELF1); PVP01_0713100(MFR4); PVP01_0713200; PVP01_0713300; PVP01_0713400(NDH2)	conserved Plasmodium protein unknown function; transcription elongation factor 1 putative; major facilitator superfamily-related transporter putative; GINS complex subunit Psf3 putative; BSD-domain protein putative; type II NADH:ubiquinone oxidoreductase putative
7	1200000	1230000	Southeast Asia	360	0.991	8.248	16	4.44	5.673	PVP01_0728600; PVP01_0728700(LRR8); PVP01_0728800(MSP1P); PVP01_0728900(MSP1); PVP01_0729000; PVP01_0729100(DGK1)	heptatricopeptide repeat-containing protein putative; leucine-rich repeat protein; merozoite surface protein 1 paralog; merozoite surface protein 1; zinc finger protein putative; diacylglycerol kinase putative
8	120000	150000	Southeast Asia	155	0.503	4.459	2	1.29	4.455	PVP01_0802300; PVP01_0802400; PVP01_0802500; PVP01_0802600; PVP01_0802700; PVP01_0802800; PVP01_0802900; PVP01_0803000	conserved protein unknown function; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; DnaJ protein putative; conserved Plasmodium protein unknown function; conserved protein unknown function; conserved protein unknown function; ribosome-binding factor A putative
8	510000	530000	Southeast Asia	95	0.544	6.007	2	2.11	5.455	PVP01_0811900(PREBP); PVP01_0812000(HO); PVP01_0812100; PVP01_0812200; PVP01_0812300(RA); PVP01_0812400(QCR7)	PRE-binding protein putative; heme oxygenase putative; RING zinc finger protein putative; conserved protein unknown function; rhoptry associated adhesin putative; cytochrome b-c1 complex subunit 7 putative
8	650000	680000	Southeast Asia	174	0.44	5.518	2	1.15	5.518	PVP01_0814700(ADA2); PVP01_0814800(PHB2); PVP01_0814900; PVP01_0815000; PVP01_0815100; PVP01_0815200	transcriptional coactivator ADA2 putative; prohibitin 2 putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; FAD synthetase putative; CWC16 domain-containing protein putative
3	240000	270000	Southern SEA	107	0.876	9.841	3	2.8	6.065	PVP01_0304600(MAg-1); PVP01_0304700(LISP2); PVP01_0304800(PRPF8); PVP01_0304900; PVP01_0305000	ag-1 blood stage membrane protein homologue; liver specific protein 2 putative; pre-mRNA-processing-splicing factor 8 putative; conserved Plasmodium protein unknown function; TMEM33 domain-containing protein putative

S14 Table. Evidence of selection within South America at the country level, as determined by iHS

Chr	Start	End	Subpopulation	Number of markers	Mean iHS markers	Max iHS	Number of iHS extreme markers	Percentage of extreme iHS markers	Mean iHS of extreme markers	Genes in the region	Products
7	1200000	1230000	Colombia	196	1.15	8.229	11	5.61	5.816	PVP01_0728600; PVP01_0728700(LRR8); PVP01_0728800(MSP1P); PVP01_0728900(MSP1); PVP01_0729000; PVP01_0729100(DGK1)	heptatricopeptide repeat-containing protein putative; leucine-rich repeat protein; merozoite surface protein 1 paralog; merozoite surface protein 1; zinc finger protein putative; diacylglycerol kinase putative
14	3000000	3030000	Colombia	67	2.014	4.408	3	4.48	4.338	PVP01_1470100(TRAG24); PVP01_1470200; PVP01_1470300; PVP01_1470400; PVP01_1470500; PVP01_1470600; PVP01_1470700	tryptophan-rich protein; Plasmodium exported protein unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein (PHIST) unknown function; hypothetical protein
3	240000	270000	Panama	98	0.845	8.968	3	3.06	6.722	PVP01_0304600(MAg-1); PVP01_0304700(LISP2); PVP01_0304800(PRPF8); PVP01_0304900; PVP01_0305000	ag-1 blood stage membrane protein homologue; liver specific protein 2 putative; pre-mRNA-processing-splicing factor 8 putative; conserved Plasmodium protein unknown function; TMEM33 domain-containing protein putative
4	180000	210000	Panama	73	0.835	4.318	3	4.11	4.318	PVP01_0404100; PVP01_0404200(TRAG3); PVP01_0404300; PVP01_0404400; PVP01_0404500; PVP01_0404600	PIR protein; tryptophan-rich protein; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; hypothetical protein; Plasmodium exported protein unknown function
5	200000	220000	Panama	25	1.487	7.614	3	12	5.392	PVP01_0504600; PVP01_0504700; PVP01_0504800(ETRAMP)	5.8S ribosomal RNA; 18S ribosomal RNA; early transcribed membrane protein
7	1200000	1230000	Panama	107	1.552	4.644	7	6.54	4.422	PVP01_0728600; PVP01_0728700(LRR8); PVP01_0728800(MSP1P); PVP01_0728900(MSP1); PVP01_0729000; PVP01_0729100(DGK1)	heptatricopeptide repeat-containing protein putative; leucine-rich repeat protein; merozoite surface protein 1 paralog; merozoite surface protein 1; zinc finger protein putative; diacylglycerol kinase putative
14	20000	50000	Panama	27	1.414	8.668	4	14.81	6.715	PVP01_1400600; PVP01_1400700; PVP01_1400800; PVP01_1400900; PVP01_1401000; PVP01_1401100; PVP01_1401200	Plasmodium exported protein unknown function pseudogene; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function
3	240000	270000	Peru	143	0.767	4.058	2	1.4	4.045	PVP01_0304600(MAg-1); PVP01_0304700(LISP2); PVP01_0304800(PRPF8); PVP01_0304900; PVP01_0305000	ag-1 blood stage membrane protein homologue; liver specific protein 2 putative; pre-mRNA-processing-splicing factor 8 putative; conserved Plasmodium protein unknown function; TMEM33 domain-containing protein putative

5	1000000	1030000	Peru	39	1.07	13.021	2	5.13	13.021	PVP01_0524600; PVP01_0524700; PVP01_0524800; PVP01_0524900; PVP01_0525000; PVP01_0525100; PVP01_0525200; PVP01_0525300	Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function pseudogene; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein (PHIST) unknown function
6	960000	990000	Peru	75	0.525	4.862	3	4	4.862	PVP01_0623500; PVP01_0623600; PVP01_0623700; PVP01_0623800(DBP); PVP01_0623900(REX4)	Plasmodium exported protein unknown function; PIR protein; Plasmodium exported protein (PHIST) unknown function; duffy binding protein; ring-exported protein 4 putative
7	1200000	1230000	Peru	231	0.798	5.363	4	1.73	5.041	PVP01_0728600; PVP01_0728700(LRR8); PVP01_0728800(MSP1P); PVP01_0728900(MSP1); PVP01_0729000; PVP01_0729100(DGK1)	heptatricopeptide repeat-containing protein putative; leucine-rich repeat protein; merozoite surface protein 1 paralog; merozoite surface protein 1; zinc finger protein putative; diacylglycerol kinase putative
14	20000	50000	Peru	106	0.839	7.258	2	1.89	5.751	PVP01_1400600; PVP01_1400700; PVP01_1400800; PVP01_1400900; PVP01_1401000; PVP01_1401100; PVP01_1401200	Plasmodium exported protein unknown function pseudogene; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function
14	2990000	3020000	Peru	15	1.373	4.756	2	13.33	4.413	PVP01_1469900(TRAG23); PVP01_1470000; PVP01_1470100(TRAG24); PVP01_1470200; PVP01_1470300; PVP01_1470400; PVP01_1470500; PVP01_1470600	tryptophan-rich protein; Plasmodium exported protein unknown function; tryptophan-rich protein; Plasmodium exported protein unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein (PHIST) unknown function

S15 Table. Signals of positive selection within Brazil (iHS)

Chromosome	Position	IHS	LOGPVALUE	Reference	Alternative	Gene name	Brazil allele counts (n=95, monoclonal isolates)
2	159062	-4.197733753	4.569281202	G	A	NA	1(55),NA(38),0(2)
5	1006505	-4.358199592	4.882275532	T	C	NA	1(56),NA(36),0(2),0.5(1)
12	270635	-4.227020582	4.625607272	T	A	NA	1(73),NA(16),0(6)
5	1017309	4.611690406	5.398583713	T	A	PVP01_0525100	NA(75),0(18),1(2)
5	1017314	4.611690406	5.398583713	G	A	PVP01_0525100	NA(75),0(18),1(2)

S16 Table. Signals of positive selection within Brazil groupings using the iHS score

Chr	Start	End	Population	Number of markers	Mean iHS markers	Max iHS	Number of iHS extreme markers	Percentage of extreme iHS markers	Mean iHS of extreme markers	Genes in the region	Products
8	870000	900000	C6	32	0.787	5.557	2	6.25	5.338	PVP01_0819800(Cap380); PVP01_0819900(CCT5); PVP01_0820000; PVP01_0820100(SEC22); PVP01_0820200(I2); PVP01_0820300; PVP01_0820400; PVP01_0820500(ABCI3)	oocyst capsule protein Cap380 putative; T-complex protein 1 subunit epsilon putative; CPW-WPC family protein; protein transport protein SEC22 putative; protein phosphatase inhibitor 2 putative; conserved protein unknown function; conserved Plasmodium protein unknown function; ABC transporter I family member 1 putative
14	40000	70000	C6	71	1.762	7.198	6	8.45	5.159	PVP01_1401000; PVP01_1401100; PVP01_1401200; PVP01_1401300; PVP01_1401400; PVP01_1401500	Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; cytoadherence linked asexual protein CLAG putative; Plasmodium exported protein unknown function
14	860000	890000	C6	18	1.171	5.455	3	16.67	5.454	PVP01_1419900; PVP01_1420000; PVP01_1420100; PVP01_1420200(TRM1); PVP01_1420300; PVP01_1420400	conserved Plasmodium protein unknown function; ubiquitin carboxyl-terminal hydrolase MINDY putative; conserved Plasmodium protein unknown function; tRNA (guanine(26)-N(2))-dimethyltransferase putative; conserved Plasmodium protein unknown function; conserved protein unknown function
14	2990000	3030000	C6	97	1.037	7.464	8	8.25	5.577	PVP01_1469900(TRAG23); PVP01_1470000; PVP01_1470100(TRAG24); PVP01_1470200; PVP01_1470300; PVP01_1470400; PVP01_1470500; PVP01_1470600; PVP01_1470700	tryptophan-rich protein; Plasmodium exported protein unknown function; tryptophan-rich protein; Plasmodium exported protein unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein (PHIST) unknown function; hypothetical protein
3	240000	270000	C7	86	0.965	4.279	2	2.33	4.279	PVP01_0304600(MAg-1); PVP01_0304700(LISP2); PVP01_0304800(PRPF8); PVP01_0304900; PVP01_0305000	ag-1 blood stage membrane protein homologue; liver specific protein 2 putative; pre-mRNA-processing-splicing factor 8 putative; conserved Plasmodium protein unknown function; TMEM33 domain-containing protein putative
4	180000	210000	C7	108	1.131	5.578	11	10.19	4.609	PVP01_0404100; PVP01_0404200(TRAG3); PVP01_0404300; PVP01_0404400; PVP01_0404500; PVP01_0404600	PIR protein; tryptophan-rich protein; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; hypothetical protein; Plasmodium exported protein unknown function
4	750000	780000	C7	111	1.238	4.921	9	8.11	4.762	PVP01_0418100(IscA1); PVP01_0418200; PVP01_0418300(MSP4); PVP01_0418400(MSP5); PVP01_0418500(ADSL); PVP01_0418600; PVP01_0418700; PVP01_0418800(AROM)	iron-sulfur assembly protein putative; conserved Plasmodium protein unknown function; merozoite surface protein 4; merozoite surface protein 5; adenylosuccinate lyase; DNA-directed RNA polymerase III subunit RPC10 putative; conserved Plasmodium protein unknown function; pentafunctional AROM polypeptide putative

5	210000	230000	C7	17	1.27	4.175	2	11.76	4.155	PVP01_0504800(ETRAMP); PVP01_0504900	early transcribed membrane protein; conserved Plasmodium protein unknown function
6	810000	840000	C7	145	0.498	9.703	3	2.07	6.246	PVP01_0619900; PVP01_0620000(MRScyt); PVP01_0620100; PVP01_0620200; PVP01_0620300; PVP01_0620400; PVP01_0620500; PVP01_0620600; PVP01_0620700	conserved protein unknown function; methionine--tRNA ligase putative; U2 snRNA/tRNA pseudouridine synthase putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; hypothetical protein; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; hypothetical protein
6	970000	1000000	C7	32	2.338	10.64	8	25	6.342	PVP01_0623600; PVP01_0623700; PVP01_0623800(DBP); PVP01_0623900(REX4); PVP01_0624000; PVP01_0624100	PIR protein; Plasmodium exported protein (PHIST) unknown function; duffy binding protein; ring-exported protein 4 putative; Plasmodium exported protein unknown function; PIR protein
7	1200000	1230000	C7	228	1.166	6.69	12	5.26	5.378	PVP01_0728600; PVP01_0728700(LRR8); PVP01_0728800(MSP1P); PVP01_0728900(MSP1); PVP01_0729000; PVP01_0729100(DGK1)	heptatricopeptide repeat-containing protein putative; leucine-rich repeat protein; merozoite surface protein 1 paralog; merozoite surface protein 1; zinc finger protein putative; diacylglycerol kinase putative
11	500000	530000	C7	19	1.343	6.98	2	10.53	6.381	PVP01_1112100; PVP01_1112200; PVP01_1112300	ADP-ribosylation factor putative; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function
14	2160000	2190000	C7	40	1.038	5	2	5	4.675	PVP01_1449700; PVP01_1449800(PRP2); PVP01_1449900; PVP01_1450000; PVP01_1450100	mitosis protein dim1 putative; pre-mRNA-splicing factor ATP-dependent RNA helicase PRP2 putative; conserved protein unknown function; conserved Plasmodium protein unknown function; WD repeat-containing protein putative
4	750000	780000	Group B	105	1.323	5.464	11	10.48	4.767	PVP01_0418100(IscA1); PVP01_0418200; PVP01_0418300(MSP4); PVP01_0418400(MSP5); PVP01_0418500(ADSL); PVP01_0418600; PVP01_0418700; PVP01_0418800(AROM)	iron-sulfur assembly protein putative; conserved Plasmodium protein unknown function; merozoite surface protein 4; merozoite surface protein 5; adenylosuccinate lyase; DNA-directed RNA polymerase III subunit RPC10 putative; conserved Plasmodium protein unknown function; pentafunctional AROM polypeptide putative
7	1200000	1230000	Group B	174	1.569	6.942	22	12.64	4.934	PVP01_0728600; PVP01_0728700(LRR8); PVP01_0728800(MSP1P); PVP01_0728900(MSP1); PVP01_0729000; PVP01_0729100(DGK1)	heptatricopeptide repeat-containing protein putative; leucine-rich repeat protein; merozoite surface protein 1 paralog; merozoite surface protein 1; zinc finger protein putative; diacylglycerol kinase putative

S17 Table. Signals of selection (Rsb) between South America and other populations across the globe

Chr	Pos	RSB	LOGPVALUE	Ref	Alt	Gene name	South America compared to	South America allele count (n = 265 monoclonal)
5	949500	-6.05993343	8.865893371	G	A	NA	East Africa	NA(184),0(60),1(21)
5	949500	6.05993343	8.865893371	G	A	NA	East Africa	NA(184),0(60),1(21)
5	949522	-4.94344498	6.114898879	T	C	NA	East Africa	NA(193),1(63),0(9)
5	949522	4.94344498	6.114898879	T	C	NA	East Africa	NA(193),1(63),0(9)
5	951272	-4.64470467	5.46780218	G	C	PVP01_0523500	East Africa	NA(180),0(83),1(2)
5	951272	4.64470467	5.46780218	G	C	PVP01_0523500	East Africa	NA(180),0(83),1(2)
5	951303	-5.73452221	8.009716165	T	G	PVP01_0523500	East Africa	NA(187),1(42),0(36)
5	951303	5.73452221	8.009716165	T	G	PVP01_0523500	East Africa	NA(187),1(42),0(36)
5	952094	-4.46750001	5.101609398	G	T	NA	East Africa	NA(169),0(91),1(5)
5	952094	4.46750001	5.101609398	G	T	NA	East Africa	NA(169),0(91),1(5)
5	952524	-4.63704782	5.451708046	G	A	NA	East Africa	NA(173),0(89),1(3)
5	952524	4.63704782	5.451708046	G	A	NA	East Africa	NA(173),0(89),1(3)
7	1216912	-4.49820621	5.164124253	C	A	MSP1	East Africa	0(173),1(48),NA(44)
7	1216912	4.49820621	5.164124253	C	A	MSP1	East Africa	0(173),1(48),NA(44)
7	1216997	-4.75365865	5.69947362	A	T	MSP1	East Africa	0(200),NA(36),1(29)
7	1216997	4.75365865	5.69947362	A	T	MSP1	East Africa	0(200),NA(36),1(29)
7	1217012	-4.82449066	5.852752023	G	A	MSP1	East Africa	0(196),NA(42),1(27)
7	1217012	4.82449066	5.852752023	G	A	MSP1	East Africa	0(196),NA(42),1(27)
7	1217882	-5.69093195	7.898432114	T	C	MSP1	East Africa	NA(162),0(54),1(49)
7	1217882	5.69093195	7.898432114	T	C	MSP1	East Africa	NA(162),0(54),1(49)
7	1217885	-5.69106808	7.898778386	G	C	MSP1	East Africa	NA(161),0(55),1(49)
7	1217885	5.69106808	7.898778386	G	C	MSP1	East Africa	NA(161),0(55),1(49)
7	1217886	-5.6911548	7.898998987	C	G	MSP1	East Africa	NA(160),0(56),1(49)
7	1217886	5.6911548	7.898998987	C	G	MSP1	East Africa	NA(160),0(56),1(49)
7	1217936	-4.83814016	5.882530791	T	C	MSP1	East Africa	0(122),1(105),NA(38)
7	1217936	4.83814016	5.882530791	T	C	MSP1	East Africa	0(122),1(105),NA(38)
7	1217939	-4.83814016	5.882530791	T	C	MSP1	East Africa	0(122),1(105),NA(38)
7	1217939	4.83814016	5.882530791	T	C	MSP1	East Africa	0(122),1(105),NA(38)
7	1217942	-5.17645833	6.645627343	G	A	MSP1	East Africa	1(127),0(94),NA(44)
7	1217942	5.17645833	6.645627343	G	A	MSP1	East Africa	1(127),0(94),NA(44)
7	1218019	-5.8855602	8.401529081	A	C	MSP1	East Africa	0(224),NA(29),1(12)
7	1218019	5.8855602	8.401529081	A	C	MSP1	East Africa	0(224),NA(29),1(12)
7	1218053	-5.43442842	7.259855618	C	T	MSP1	East Africa	0(198),1(41),NA(26)
7	1218053	5.43442842	7.259855618	C	T	MSP1	East Africa	0(198),1(41),NA(26)
7	1218092	-4.8126002	5.826874558	C	T	MSP1	East Africa	1(151),0(78),NA(36)
7	1218092	4.8126002	5.826874558	C	T	MSP1	East Africa	1(151),0(78),NA(36)
7	1218104	-5.04414	6.341449004	G	A	MSP1	East Africa	1(150),0(76),NA(39)
7	1218104	5.04414	6.341449004	G	A	MSP1	East Africa	1(150),0(76),NA(39)
9	836594	-4.50783498	5.183808558	G	C	PVP01_0918900	East Africa	1(180),0(61),NA(24)
9	836594	4.50783498	5.183808558	G	C	PVP01_0918900	East Africa	1(180),0(61),NA(24)
11	1247014	-5.68183348	7.875305475	G	A	NA	East Africa	0(229),NA(25),1(11)
11	1247014	5.68183348	7.875305475	G	A	NA	East Africa	0(229),NA(25),1(11)
11	1247617	-5.84236827	8.288497474	T	A	NA	East Africa	0(239),1(13),NA(13)
11	1247617	5.84236827	8.288497474	T	A	NA	East Africa	0(239),1(13),NA(13)
11	1480631	-4.49034782	5.148087862	C	T	NA	East Africa	0(191),1(44),NA(30)
11	1480631	4.49034782	5.148087862	C	T	NA	East Africa	0(191),1(44),NA(30)
5	949500	5.60784865	7.688549094	G	A	NA	South Asia	NA(184),0(60),1(21)
5	949500	-5.60784865	7.688549094	G	A	NA	South Asia	NA(184),0(60),1(21)
5	949522	5.06829105	6.396419057	T	C	NA	South Asia	NA(193),1(63),0(9)
5	949522	-5.06829105	6.396419057	T	C	NA	South Asia	NA(193),1(63),0(9)
5	951303	5.36199617	7.084561716	T	G	PVP01_0523500	South Asia	NA(187),1(42),0(36)
5	951303	-5.36199617	7.084561716	T	G	PVP01_0523500	South Asia	NA(187),1(42),0(36)
5	958312	5.46150181	7.325944431	C	T	NA	South Asia	NA(175),0(79),1(11)
5	958312	-5.46150181	7.325944431	C	T	NA	South Asia	NA(175),0(79),1(11)
5	958366	5.85013088	8.308753715	A	G	NA	South Asia	NA(174),0(62),1(28),0.5(1)
5	958366	-5.85013088	8.308753715	A	G	NA	South Asia	NA(174),0(62),1(28),0.5(1)
5	958539	5.1597145	6.606728984	G	C	NA	South Asia	NA(175),0(47),1(42),0.5(1)
5	958539	-5.1597145	6.606728984	G	C	NA	South Asia	NA(175),0(47),1(42),0.5(1)
5	1014653	-5.27649394	6.880483916	T	A	NA	South Asia	NA(255),0(7),1(3)

5	1014653	5.27649394	6.880483916	T	A	NA	South Asia	NA(255),0(7),1(3)
5	1014669	-6.40978845	9.836475773	A	G	NA	South Asia	NA(256),0(6),1(3)
5	1014669	6.40978845	9.836475773	A	G	NA	South Asia	NA(256),0(6),1(3)
6	838969	5.2375683	6.788596164	A	T	PVP01_0620700	South Asia	NA(117),0(75),1(73)
6	838969	-5.2375683	6.788596164	A	T	PVP01_0620700	South Asia	NA(117),0(75),1(73)
7	1216360	5.087396	6.440077429	A	G	MSP1	South Asia	0(155),NA(60),1(49),0.5(1)
7	1216360	-5.087396	6.440077429	A	G	MSP1	South Asia	0(155),NA(60),1(49),0.5(1)
7	1216364	4.47742365	5.121769947	T	C	MSP1	South Asia	1(118),0(84),NA(62),0.5(1)
7	1216364	-4.47742365	5.121769947	T	C	MSP1	South Asia	1(118),0(84),NA(62),0.5(1)
7	1216369	5.2896543	6.911694627	T	C	MSP1	South Asia	0(157),NA(58),1(49),0.5(1)
7	1216369	-5.2896543	6.911694627	T	C	MSP1	South Asia	0(157),NA(58),1(49),0.5(1)
7	1216377	4.6113618	5.397897039	G	A	MSP1	South Asia	0(171),NA(57),1(37)
7	1216377	-4.6113618	5.397897039	G	A	MSP1	South Asia	0(171),NA(57),1(37)
7	1216403	4.97698372	6.189883467	G	T	MSP1	South Asia	0(181),NA(67),1(17)
7	1216403	-4.97698372	6.189883467	G	T	MSP1	South Asia	0(181),NA(67),1(17)
7	1216409	5.46547699	7.335674298	T	A	MSP1	South Asia	0(144),NA(81),1(40)
7	1216409	-5.46547699	7.335674298	T	A	MSP1	South Asia	0(144),NA(81),1(40)
7	1216431	5.02421625	6.296285241	A	C	MSP1	South Asia	0(155),NA(76),1(34)
7	1216431	-5.02421625	6.296285241	A	C	MSP1	South Asia	0(155),NA(76),1(34)
7	1216440	4.95407988	6.13862482	A	G	MSP1	South Asia	NA(100),1(84),0(80),0.5(1)
7	1216440	-4.95407988	6.13862482	A	G	MSP1	South Asia	NA(100),1(84),0(80),0.5(1)
7	1216444	5.00727561	6.258014886	T	A	MSP1	South Asia	NA(101),1(83),0(80),0.5(1)
7	1216444	-5.00727561	6.258014886	T	A	MSP1	South Asia	NA(101),1(83),0(80),0.5(1)
7	1216453	4.48971774	5.146803188	A	C	MSP1	South Asia	NA(122),0(80),1(62),0.5(1)
7	1216453	-4.48971774	5.146803188	A	C	MSP1	South Asia	NA(122),0(80),1(62),0.5(1)
7	1216492	4.41877265	5.003212601	A	T	MSP1	South Asia	0(152),NA(70),1(43)
7	1216492	-4.41877265	5.003212601	A	T	MSP1	South Asia	0(152),NA(70),1(43)
7	1216897	4.7852604	5.767599118	A	T	MSP1	South Asia	0(198),NA(40),1(27)
7	1216897	-4.7852604	5.767599118	A	T	MSP1	South Asia	0(198),NA(40),1(27)
7	1216907	4.787831	5.773159109	C	T	MSP1	South Asia	0(200),NA(38),1(27)
7	1216907	-4.787831	5.773159109	C	T	MSP1	South Asia	0(200),NA(38),1(27)
7	1216912	5.52397689	7.479632995	C	A	MSP1	South Asia	0(173),1(48),NA(44)
7	1216912	-5.52397689	7.479632995	C	A	MSP1	South Asia	0(173),1(48),NA(44)
7	1216997	4.43683017	5.039561265	A	T	MSP1	South Asia	0(200),NA(36),1(29)
7	1216997	-4.43683017	5.039561265	A	T	MSP1	South Asia	0(200),NA(36),1(29)
7	1217012	4.50923538	5.186674645	G	A	MSP1	South Asia	0(196),NA(42),1(27)
7	1217012	-4.50923538	5.186674645	G	A	MSP1	South Asia	0(196),NA(42),1(27)
7	1217882	6.16345938	9.147686923	T	C	MSP1	South Asia	NA(162),0(54),1(49)
7	1217882	-6.16345938	9.147686923	T	C	MSP1	South Asia	NA(162),0(54),1(49)
7	1217885	6.16359853	9.148068741	G	C	MSP1	South Asia	NA(161),0(55),1(49)
7	1217885	-6.16359853	9.148068741	G	C	MSP1	South Asia	NA(161),0(55),1(49)
7	1217886	6.16368718	9.148311987	C	G	MSP1	South Asia	NA(160),0(56),1(49)
7	1217886	-6.16368718	9.148311987	C	G	MSP1	South Asia	NA(160),0(56),1(49)
7	1217936	6.06657569	8.883836855	T	C	MSP1	South Asia	0(122),1(105),NA(38)
7	1217936	-6.06657569	8.883836855	T	C	MSP1	South Asia	0(122),1(105),NA(38)
7	1217939	6.06657569	8.883836855	T	C	MSP1	South Asia	0(122),1(105),NA(38)
7	1217939	-6.06657569	8.883836855	T	C	MSP1	South Asia	0(122),1(105),NA(38)
7	1217942	4.81993161	5.842823045	G	A	MSP1	South Asia	1(127),0(94),NA(44)
7	1217942	-4.81993161	5.842823045	G	A	MSP1	South Asia	1(127),0(94),NA(44)
7	1218019	5.92309881	8.500407924	A	C	MSP1	South Asia	0(224),NA(29),1(12)
7	1218019	-5.92309881	8.500407924	A	C	MSP1	South Asia	0(224),NA(29),1(12)
7	1218020	5.46262607	7.328695555	T	C	MSP1	South Asia	0(173),1(59),NA(33)
7	1218020	-5.46262607	7.328695555	T	C	MSP1	South Asia	0(173),1(59),NA(33)
7	1218053	6.61597398	10.43284032	C	T	MSP1	South Asia	0(198),1(41),NA(26)
7	1218053	-6.61597398	10.43284032	C	T	MSP1	South Asia	0(198),1(41),NA(26)
7	1218092	5.76090791	8.077468225	C	T	MSP1	South Asia	1(151),0(78),NA(36)
7	1218092	-5.76090791	8.077468225	C	T	MSP1	South Asia	1(151),0(78),NA(36)
7	1218100	5.33677184	7.024035513	G	A	MSP1	South Asia	1(220),NA(42),0(3)
7	1218100	-5.33677184	7.024035513	G	A	MSP1	South Asia	1(220),NA(42),0(3)
7	1218104	6.36558822	9.71098501	G	A	MSP1	South Asia	1(150),0(76),NA(39)

7	1218104	-6.36558822	9.71098501	G	A	MSP1	South Asia	1(150),0(76),NA(39)
7	1218107	5.57085907	7.596045471	G	A	MSP1	South Asia	0(210),NA(32),1(23)
7	1218107	-5.57085907	7.596045471	G	A	MSP1	South Asia	0(210),NA(32),1(23)
7	1218113	5.0739015	6.409224063	T	C	MSP1	South Asia	1(173),0(52),NA(40)
7	1218113	-5.0739015	6.409224063	T	C	MSP1	South Asia	1(173),0(52),NA(40)
7	1218126	4.93541153	6.097008121	T	C	MSP1	South Asia	1(190),NA(44),0(31)
7	1218126	-4.93541153	6.097008121	T	C	MSP1	South Asia	1(190),NA(44),0(31)
14	2042264	5.33006203	7.0079804	A	C	PVP01_1447100	South Asia	1(218),NA(34),0(13)
14	2042264	-5.33006203	7.0079804	A	C	PVP01_1447100	South Asia	1(218),NA(34),0(13)
2	148634	-4.7454501	5.681846477	A	T	NA	South East Asia	1(135),NA(79),0(51)
2	148634	4.7454501	5.681846477	A	T	NA	South East Asia	1(135),NA(79),0(51)
2	148676	-5.85086974	8.310683058	A	C	NA	South East Asia	0(182),NA(76),1(7)
2	148676	5.85086974	8.310683058	A	C	NA	South East Asia	0(182),NA(76),1(7)
2	148792	-6.1080733	8.996362671	G	A	NA	South East Asia	0(181),NA(79),1(5)
2	148792	6.1080733	8.996362671	G	A	NA	South East Asia	0(181),NA(79),1(5)
2	148876	-5.27874538	6.885818182	A	G	NA	South East Asia	0(120),NA(98),1(47)
2	148876	5.27874538	6.885818182	A	G	NA	South East Asia	0(120),NA(98),1(47)
2	148971	-7.83440013	14.3268995	G	A	NA	South East Asia	0(164),NA(99),1(2)
2	148971	7.83440013	14.3268995	G	A	NA	South East Asia	0(164),NA(99),1(2)
2	149000	-6.59399067	10.36839528	A	C	NA	South East Asia	0(166),NA(97),1(2)
2	149000	6.59399067	10.36839528	A	C	NA	South East Asia	0(166),NA(97),1(2)
2	149160	-7.35679254	12.72496113	A	T	NA	South East Asia	0(145),NA(99),1(21)
2	149160	7.35679254	12.72496113	A	T	NA	South East Asia	0(145),NA(99),1(21)
2	149172	-7.52119594	13.26538071	T	C	NA	South East Asia	0(147),NA(97),1(21)
2	149172	7.52119594	13.26538071	T	C	NA	South East Asia	0(147),NA(97),1(21)
2	149203	-7.08749284	11.86466579	G	T	NA	South East Asia	0(157),NA(105),1(3)
2	149203	7.08749284	11.86466579	G	T	NA	South East Asia	0(157),NA(105),1(3)
2	149842	-5.81737426	8.223449914	T	C	NA	South East Asia	0(191),NA(67),1(7)
2	149842	5.81737426	8.223449914	T	C	NA	South East Asia	0(191),NA(67),1(7)
2	150220	-4.69222601	5.568236968	G	A	NA	South East Asia	NA(106),1(101),0(58)
2	150220	4.69222601	5.568236968	G	A	NA	South East Asia	NA(106),1(101),0(58)
2	151589	-4.46564169	5.097838656	T	C	NA	South East Asia	1(134),NA(67),0(64)
2	151589	4.46564169	5.097838656	T	C	NA	South East Asia	1(134),NA(67),0(64)
2	151849	-5.01526968	6.27605914	G	A	NA	South East Asia	0(170),NA(62),1(33)
2	151849	5.01526968	6.27605914	G	A	NA	South East Asia	0(170),NA(62),1(33)
4	762975	4.91591324	6.053697452	G	A	MSP5	South East Asia	0(156),1(76),NA(33)
4	762975	-4.91591324	6.053697452	G	A	MSP5	South East Asia	0(156),1(76),NA(33)
5	949500	4.83103897	5.867028557	G	A	NA	South East Asia	NA(184),0(60),1(21)
5	949500	-4.83103897	5.867028557	G	A	NA	South East Asia	NA(184),0(60),1(21)
5	951303	5.01310836	6.271177956	T	G	PVP01_0523500	South East Asia	NA(187),1(42),0(36)
5	951303	-5.01310836	6.271177956	T	G	PVP01_0523500	South East Asia	NA(187),1(42),0(36)
5	957973	5.0210421	6.289105391	T	C	NA	South East Asia	NA(170),0(92),1(3)
5	957973	-5.0210421	6.289105391	T	C	NA	South East Asia	NA(170),0(92),1(3)
5	958312	4.61258404	5.400451322	C	T	NA	South East Asia	NA(175),0(79),1(11)
5	958312	-4.61258404	5.400451322	C	T	NA	South East Asia	NA(175),0(79),1(11)
5	958366	4.91369842	6.048787865	A	G	NA	South East Asia	NA(174),0(62),1(28),0.5(1)
5	958366	-4.91369842	6.048787865	A	G	NA	South East Asia	NA(174),0(62),1(28),0.5(1)
5	958539	4.66509299	5.5107765	G	C	NA	South East Asia	NA(175),0(47),1(42),0.5(1)
5	958539	-4.66509299	5.5107765	G	C	NA	South East Asia	NA(175),0(47),1(42),0.5(1)
6	838966	5.06124411	6.380354242	C	T	PVP01_0620700	South East Asia	NA(117),0(83),1(65)
6	838966	-5.06124411	6.380354242	C	T	PVP01_0620700	South East Asia	NA(117),0(83),1(65)
6	838969	5.39735622	7.169860425	A	T	PVP01_0620700	South East Asia	NA(117),0(75),1(73)
6	838969	-5.39735622	7.169860425	A	T	PVP01_0620700	South East Asia	NA(117),0(75),1(73)
6	839005	4.81870392	5.840150814	T	A	PVP01_0620700	South East Asia	NA(142),1(65),0(58)
6	839005	-4.81870392	5.840150814	T	A	PVP01_0620700	South East Asia	NA(142),1(65),0(58)
6	839035	4.78994112	5.777725177	G	C	PVP01_0620700	South East Asia	NA(151),1(58),0(56)
6	839035	-4.78994112	5.777725177	G	C	PVP01_0620700	South East Asia	NA(151),1(58),0(56)
6	839038	4.78989707	5.777629836	A	C	PVP01_0620700	South East Asia	NA(127),0(79),1(59)
6	839038	-4.78989707	5.777629836	A	C	PVP01_0620700	South East Asia	NA(127),0(79),1(59)
6	839039	4.78989707	5.777629836	C	T	PVP01_0620700	South East Asia	NA(150),1(59),0(56)

6	839039	-4.78989707	5.777629836	C	T	PVP01_0620700	South East Asia	NA(150),1(59),0(56)
6	839048	5.05803157	6.373037574	G	A	PVP01_0620700	South East Asia	NA(151),1(58),0(56)
6	839048	-5.05803157	6.373037574	G	A	PVP01_0620700	South East Asia	NA(151),1(58),0(56)
7	1216288	4.49899433	5.165733981	A	G	MSP1	South East Asia	0(207),NA(52),1(6)
7	1216288	-4.49899433	5.165733981	A	G	MSP1	South East Asia	0(207),NA(52),1(6)
7	1216318	4.94790009	6.124832233	C	A	MSP1	South East Asia	1(125),0(89),NA(51)
7	1216318	-4.94790009	6.124832233	C	A	MSP1	South East Asia	1(125),0(89),NA(51)
7	1216328	4.74795077	5.687213458	A	G	MSP1	South East Asia	1(124),0(86),NA(55)
7	1216328	-4.74795077	5.687213458	A	G	MSP1	South East Asia	1(124),0(86),NA(55)
7	1216346	4.92428003	6.072262679	A	G	MSP1	South East Asia	1(107),0(104),NA(54)
7	1216346	-4.92428003	6.072262679	A	G	MSP1	South East Asia	1(107),0(104),NA(54)
7	1216360	4.97272521	6.180336269	A	G	MSP1	South East Asia	0(155),NA(60),1(49),0.5(1)
7	1216360	-4.97272521	6.180336269	A	G	MSP1	South East Asia	0(155),NA(60),1(49),0.5(1)
7	1216364	4.53505183	5.239657878	T	C	MSP1	South East Asia	1(118),0(84),NA(62),0.5(1)
7	1216364	-4.53505183	5.239657878	T	C	MSP1	South East Asia	1(118),0(84),NA(62),0.5(1)
7	1216369	5.09834648	6.46517049	T	C	MSP1	South East Asia	0(157),NA(58),1(49),0.5(1)
7	1216369	-5.09834648	6.46517049	T	C	MSP1	South East Asia	0(157),NA(58),1(49),0.5(1)
7	1216403	4.92076648	6.06446281	G	T	MSP1	South East Asia	0(181),NA(67),1(17)
7	1216403	-4.92076648	6.06446281	G	T	MSP1	South East Asia	0(181),NA(67),1(17)
7	1216409	5.14744472	6.578299387	T	A	MSP1	South East Asia	0(144),NA(81),1(40)
7	1216409	-5.14744472	6.578299387	T	A	MSP1	South East Asia	0(144),NA(81),1(40)
7	1216411	5.31196645	6.964776169	A	G	MSP1	South East Asia	0(173),NA(71),1(21)
7	1216411	-5.31196645	6.964776169	A	G	MSP1	South East Asia	0(173),NA(71),1(21)
7	1216416	5.24657418	6.80979868	T	A	MSP1	South East Asia	1(146),NA(94),0(25)
7	1216416	-5.24657418	6.80979868	T	A	MSP1	South East Asia	1(146),NA(94),0(25)
7	1216431	4.91977569	6.062264253	A	C	MSP1	South East Asia	0(155),NA(76),1(34)
7	1216431	-4.91977569	6.062264253	A	C	MSP1	South East Asia	0(155),NA(76),1(34)
7	1216435	4.41722769	5.00010902	T	C	MSP1	South East Asia	0(137),NA(92),1(36)
7	1216435	-4.41722769	5.00010902	T	C	MSP1	South East Asia	0(137),NA(92),1(36)
7	1216440	4.99302663	6.225918578	A	G	MSP1	South East Asia	NA(100),1(84),0(80),0.5(1)
7	1216440	-4.99302663	6.225918578	A	G	MSP1	South East Asia	NA(100),1(84),0(80),0.5(1)
7	1216444	4.82217925	5.847716997	T	A	MSP1	South East Asia	NA(101),1(83),0(80),0.5(1)
7	1216444	-4.82217925	5.847716997	T	A	MSP1	South East Asia	NA(101),1(83),0(80),0.5(1)
7	1216453	4.49023472	5.147857248	A	C	MSP1	South East Asia	NA(122),0(80),1(62),0.5(1)
7	1216453	-4.49023472	5.147857248	A	C	MSP1	South East Asia	NA(122),0(80),1(62),0.5(1)
7	1216912	4.45809924	5.082549013	C	A	MSP1	South East Asia	0(173),1(48),NA(44)
7	1216912	-4.45809924	5.082549013	C	A	MSP1	South East Asia	0(173),1(48),NA(44)
7	1217882	5.35175726	7.059960937	T	C	MSP1	South East Asia	NA(162),0(54),1(49)
7	1217882	-5.35175726	7.059960937	T	C	MSP1	South East Asia	NA(162),0(54),1(49)
7	1217885	5.35187867	7.060252394	G	C	MSP1	South East Asia	NA(161),0(55),1(49)
7	1217885	-5.35187867	7.060252394	G	C	MSP1	South East Asia	NA(161),0(55),1(49)
7	1217886	5.35195602	7.060438074	C	G	MSP1	South East Asia	NA(160),0(56),1(49)
7	1217886	-5.35195602	7.060438074	C	G	MSP1	South East Asia	NA(160),0(56),1(49)
7	1217936	5.875212	8.374376286	T	C	MSP1	South East Asia	0(122),1(105),NA(38)
7	1217936	-5.875212	8.374376286	T	C	MSP1	South East Asia	0(122),1(105),NA(38)
7	1217939	5.875212	8.374376286	T	C	MSP1	South East Asia	0(122),1(105),NA(38)
7	1217939	-5.875212	8.374376286	T	C	MSP1	South East Asia	0(122),1(105),NA(38)
7	1218019	4.45050612	5.067180599	A	C	MSP1	South East Asia	0(224),NA(29),1(12)
7	1218019	-4.45050612	5.067180599	A	C	MSP1	South East Asia	0(224),NA(29),1(12)
7	1218020	5.55718193	7.56198805	T	C	MSP1	South East Asia	0(173),1(59),NA(33)
7	1218020	-5.55718193	7.56198805	T	C	MSP1	South East Asia	0(173),1(59),NA(33)
7	1218053	5.8444394	8.293899525	C	T	MSP1	South East Asia	0(198),1(41),NA(26)
7	1218053	-5.8444394	8.293899525	C	T	MSP1	South East Asia	0(198),1(41),NA(26)
7	1218092	4.96048034	6.152926823	C	T	MSP1	South East Asia	1(151),0(78),NA(36)
7	1218092	-4.96048034	6.152926823	C	T	MSP1	South East Asia	1(151),0(78),NA(36)
7	1218104	5.60249498	7.675124839	G	A	MSP1	South East Asia	1(150),0(76),NA(39)
7	1218104	-5.60249498	7.675124839	G	A	MSP1	South East Asia	1(150),0(76),NA(39)
7	1218107	5.16218189	6.612453682	G	A	MSP1	South East Asia	0(210),NA(32),1(23)
7	1218107	-5.16218189	6.612453682	G	A	MSP1	South East Asia	0(210),NA(32),1(23)
7	1218113	4.6831258	5.548930832	T	C	MSP1	South East Asia	1(173),0(52),NA(40)

7	1218113	-4.6831258	5.548930832	T	C	MSP1	South East Asia	1(173),0(52),NA(40)
11	1423431	4.85721213	5.924270469	C	G	PVP01_1133300	South East Asia	0(126),1(105),NA(34)
11	1423431	-4.85721213	5.924270469	C	G	PVP01_1133300	South East Asia	0(126),1(105),NA(34)
14	2043057	4.45011746	5.066394601	A	G	PVP01_1447100	South East Asia	NA(258),0(5),1(2)
14	2043057	-4.45011746	5.066394601	A	G	PVP01_1447100	South East Asia	NA(258),0(5),1(2)
6	838966	5.40237527	7.182010557	C	T	PVP01_0620700	Southern South East Asia	NA(117),0(83),1(65)
6	838966	-5.40237527	7.182010557	C	T	PVP01_0620700	Southern South East Asia	NA(117),0(83),1(65)
6	838969	5.29563825	6.925910095	A	T	PVP01_0620700	Southern South East Asia	NA(117),0(75),1(73)
6	838969	-5.29563825	6.925910095	A	T	PVP01_0620700	Southern South East Asia	NA(117),0(75),1(73)
7	1217936	5.39068108	7.153717672	T	C	MSP1	Southern South East Asia	0(122),1(105),NA(38)
7	1217936	-5.39068108	7.153717672	T	C	MSP1	Southern South East Asia	0(122),1(105),NA(38)
7	1217939	5.39068108	7.153717672	T	C	MSP1	Southern South East Asia	0(122),1(105),NA(38)
7	1217939	-5.39068108	7.153717672	T	C	MSP1	Southern South East Asia	0(122),1(105),NA(38)
7	1218020	5.00903053	6.261973801	T	C	MSP1	Southern South East Asia	0(173),1(59),NA(33)
7	1218020	-5.00903053	6.261973801	T	C	MSP1	Southern South East Asia	0(173),1(59),NA(33)
7	1218053	5.19161089	6.680930693	C	T	MSP1	Southern South East Asia	0(198),1(41),NA(26)
7	1218053	-5.19161089	6.680930693	C	T	MSP1	Southern South East Asia	0(198),1(41),NA(26)
7	1218092	5.48739468	7.389440888	C	T	MSP1	Southern South East Asia	1(151),0(78),NA(36)
7	1218092	-5.48739468	7.389440888	C	T	MSP1	Southern South East Asia	1(151),0(78),NA(36)
7	1218104	5.91240137	8.472169452	G	A	MSP1	Southern South East Asia	1(150),0(76),NA(39)
7	1218104	-5.91240137	8.472169452	G	A	MSP1	Southern South East Asia	1(150),0(76),NA(39)
7	1218107	5.20487193	6.711906352	G	A	MSP1	Southern South East Asia	0(210),NA(32),1(23)
7	1218107	-5.20487193	6.711906352	G	A	MSP1	Southern South East Asia	0(210),NA(32),1(23)
7	1218113	5.30031295	6.937025831	T	C	MSP1	Southern South East Asia	1(173),0(52),NA(40)
7	1218113	-5.30031295	6.937025831	T	C	MSP1	Southern South East Asia	1(173),0(52),NA(40)
7	1218126	4.81637821	5.835090297	T	C	MSP1	Southern South East Asia	1(190),NA(44),0(31)
7	1218126	-4.81637821	5.835090297	T	C	MSP1	Southern South East Asia	1(190),NA(44),0(31)
7	1218158	4.56587209	5.303274984	T	C	MSP1	Southern South East Asia	1(168),NA(50),0(47)
7	1218158	-4.56587209	5.303274984	T	C	MSP1	Southern South East Asia	1(168),NA(50),0(47)

S18 Table. Signals of selection (Rsb) between Brazil and other countries within South America

Chr	Pos	RSB	LOGPVALUE	Ref	Alt	Gene name	Brazil compared to	Brazil allele count (n = 95 monoclonal)
14	2372411	4.739332072	5.668726903	A	G	NA	Colombia	NA(41),1(29),0(25)
14	2372411	-4.739332072	5.668726903	A	G	NA	Colombia	NA(41),1(29),0(25)
14	3008899	-5.08443076	6.433291235	T	G	NA	Colombia	NA(43),0.5(33),0(14),1(5)
14	3008899	5.08443076	6.433291235	T	G	NA	Colombia	NA(43),0.5(33),0(14),1(5)
14	3008965	-4.693946743	5.571891414	T	A	NA	Colombia	0.5(35),NA(35),0(19),1(6)
14	3008965	4.693946743	5.571891414	T	A	NA	Colombia	0.5(35),NA(35),0(19),1(6)
14	3009004	-4.640396484	5.45874368	T	G	NA	Colombia	0(33),NA(29),0.5(27),1(6)
14	3009004	4.640396484	5.45874368	T	G	NA	Colombia	0(33),NA(29),0.5(27),1(6)
14	3009013	-4.482390072	5.131874988	C	T	NA	Colombia	0(35),NA(29),0.5(26),1(5)
14	3009013	4.482390072	5.131874988	C	T	NA	Colombia	0(35),NA(29),0.5(26),1(5)
14	3009060	-4.513160398	5.194712046	T	G	NA	Colombia	0(34),NA(34),0.5(24),1(3)
14	3009060	4.513160398	5.194712046	T	G	NA	Colombia	0(34),NA(34),0.5(24),1(3)
14	3009162	-4.571472482	5.314877553	T	G	NA	Colombia	NA(37),0(28),0.5(28),1(2)
14	3009162	4.571472482	5.314877553	T	G	NA	Colombia	NA(37),0(28),0.5(28),1(2)
14	3009165	-4.41957286	5.004820502	A	C	NA	Colombia	0(57),NA(36),0.5(2)
14	3009165	4.41957286	5.004820502	A	C	NA	Colombia	0(57),NA(36),0.5(2)
14	3009227	-4.441810549	5.049610387	C	A	NA	Colombia	0(49),NA(37),0.5(9)
14	3009227	4.441810549	5.049610387	C	A	NA	Colombia	0(49),NA(37),0.5(9)
14	3009263	-4.921767278	6.066684	A	C	NA	Colombia	0(60),NA(34),0.5(1)
14	3009263	4.921767278	6.066684	A	C	NA	Colombia	0(60),NA(34),0.5(1)
14	3009282	-5.264927207	6.853112875	G	T	NA	Colombia	0(50),NA(31),0.5(12),1(2)
14	3009282	5.264927207	6.853112875	G	T	NA	Colombia	0(50),NA(31),0.5(12),1(2)
14	3009309	-4.909358213	6.03917291	A	T	NA	Colombia	NA(37),0.5(29),0(23),1(6)
14	3009309	4.909358213	6.03917291	A	T	NA	Colombia	NA(37),0.5(29),0(23),1(6)
14	3011470	-4.703829814	5.592904784	T	A	NA	Colombia	NA(32),0(31),0.5(26),1(6)
14	3011470	4.703829814	5.592904784	T	A	NA	Colombia	NA(32),0(31),0.5(26),1(6)
14	3011702	-4.492306207	5.152081881	T	C	NA	Colombia	NA(35),0(26),0.5(21),1(13)
14	3011702	4.492306207	5.152081881	T	C	NA	Colombia	NA(35),0(26),0.5(21),1(13)
14	3011916	-4.520788118	5.210349991	T	C	NA	Colombia	0(63),NA(29),0.5(3)
14	3011916	4.520788118	5.210349991	T	C	NA	Colombia	0(63),NA(29),0.5(3)
14	3011931	-4.476327571	5.119541162	G	T	NA	Colombia	0(63),NA(28),0.5(3),1(1)
14	3011931	4.476327571	5.119541162	G	T	NA	Colombia	0(63),NA(28),0.5(3),1(1)
4	762854	-4.602466017	5.379325104	A	T	MSP5	Mexico	1(62),NA(20),0(13)
4	762854	4.602466017	5.379325104	A	T	MSP5	Mexico	1(62),NA(20),0(13)
4	762885	-4.536810669	5.243277689	A	C	MSP5	Mexico	0(76),1(13),NA(6)
4	762885	4.536810669	5.243277689	A	C	MSP5	Mexico	0(76),1(13),NA(6)
4	763051	-5.316655759	6.975958906	G	A	MSP5	Mexico	0(72),NA(16),1(7)
4	763051	5.316655759	6.975958906	G	A	MSP5	Mexico	0(72),NA(16),1(7)
3	251204	4.529178419	5.227579399	C	T	LISP2	Panama	1(41),0(37),NA(17)
3	251204	-4.529178419	5.227579399	C	T	LISP2	Panama	1(41),0(37),NA(17)
3	251206	4.529178419	5.227579399	A	G	LISP2	Panama	1(41),0(37),NA(17)
3	251206	-4.529178419	5.227579399	A	G	LISP2	Panama	1(41),0(37),NA(17)
3	251247	4.465564078	5.097681215	G	T	LISP2	Panama	0(42),1(32),NA(21)
3	251247	-4.465564078	5.097681215	G	T	LISP2	Panama	0(42),1(32),NA(21)
3	251256	6.279314671	9.468433347	C	T	LISP2	Panama	0(46),1(25),NA(24)
3	251256	-6.279314671	9.468433347	C	T	LISP2	Panama	0(46),1(25),NA(24)
3	251329	4.474221548	5.115260164	A	G	LISP2	Panama	0(41),1(30),NA(24)
3	251329	-4.474221548	5.115260164	A	G	LISP2	Panama	0(41),1(30),NA(24)
3	251367	4.434737268	5.035341422	C	T	LISP2	Panama	0(40),1(34),NA(21)
3	251367	-4.434737268	5.035341422	C	T	LISP2	Panama	0(40),1(34),NA(21)
3	251375	5.462408539	7.328163211	A	C	LISP2	Panama	0(41),1(34),NA(20)
3	251375	-5.462408539	7.328163211	A	C	LISP2	Panama	0(41),1(34),NA(20)
3	251377	5.365637124	7.093320405	G	T	LISP2	Panama	0(52),NA(22),1(21)
3	251377	-5.365637124	7.093320405	G	T	LISP2	Panama	0(52),NA(22),1(21)
3	251395	4.441181867	5.048341299	C	T	LISP2	Panama	0(56),NA(25),1(14)
3	251395	-4.441181867	5.048341299	C	T	LISP2	Panama	0(56),NA(25),1(14)
3	251396	4.441181867	5.048341299	A	G	LISP2	Panama	0(56),NA(25),1(14)
3	251396	-4.441181867	5.048341299	A	G	LISP2	Panama	0(56),NA(25),1(14)
4	500313	4.965327066	6.163768419	T	C	PVP01_041240(	Panama	0(46),1(25),NA(24)

4	500313	-4.965327066	6.163768419	T	C	PVP01_041240(	Panama	0(46),1(25),NA(24)
10	819187	7.029900744	11.68470047	A	C	PVP01_101840(	Panama	NA(57),0(31),1(7)
10	819187	-7.029900744	11.68470047	A	C	PVP01_101840(	Panama	NA(57),0(31),1(7)
14	2263645	4.6810523	5.544536735	C	G	AARP1	Panama	0(69),NA(15),1(11)
14	2263645	-4.6810523	5.544536735	C	G	AARP1	Panama	0(69),NA(15),1(11)
14	2271643	4.984384451	6.206493337	C	G	SENP1	Panama	0(50),1(37),NA(8)
14	2271643	-4.984384451	6.206493337	C	G	SENP1	Panama	0(50),1(37),NA(8)
14	2273947	4.72373897	5.635359746	G	C	NA	Panama	0(57),1(26),NA(11),0.5(1)
14	2273947	-4.72373897	5.635359746	G	C	NA	Panama	0(57),1(26),NA(11),0.5(1)
14	3009165	-4.45991257	5.086222726	A	C	NA	Panama	0(57),NA(36),0.5(2)
14	3009165	4.45991257	5.086222726	A	C	NA	Panama	0(57),NA(36),0.5(2)
2	599529	4.476383813	5.119655512	C	G	NA	Peru	0(61),1(29),NA(5)
2	599529	-4.476383813	5.119655512	C	G	NA	Peru	0(61),1(29),NA(5)
3	251256	4.749832939	5.691254732	C	T	LISP2	Peru	0(46),1(25),NA(24)
3	251256	-4.749832939	5.691254732	C	T	LISP2	Peru	0(46),1(25),NA(24)
3	251361	4.861508309	5.933693871	G	T	LISP2	Peru	0(54),NA(21),1(20)
3	251361	-4.861508309	5.933693871	G	T	LISP2	Peru	0(54),NA(21),1(20)
3	251368	4.905569061	6.030785181	C	T	LISP2	Peru	0(52),NA(23),1(20)
3	251368	-4.905569061	6.030785181	C	T	LISP2	Peru	0(52),NA(23),1(20)
3	251377	5.06741962	6.394431341	G	T	LISP2	Peru	0(52),NA(22),1(21)
3	251377	-5.06741962	6.394431341	G	T	LISP2	Peru	0(52),NA(22),1(21)
3	251470	4.505612343	5.179261338	T	C	LISP2	Peru	0(73),NA(15),1(7)
3	251470	-4.505612343	5.179261338	T	C	LISP2	Peru	0(73),NA(15),1(7)
3	252094	4.526615328	5.222312996	G	A	LISP2	Peru	NA(48),1(33),0(14)
3	252094	-4.526615328	5.222312996	G	A	LISP2	Peru	NA(48),1(33),0(14)
4	762854	-4.532906122	5.235243635	A	T	MSP5	Peru	1(62),NA(20),0(13)
4	762854	4.532906122	5.235243635	A	T	MSP5	Peru	1(62),NA(20),0(13)
4	763051	-6.154373081	9.122772341	G	A	MSP5	Peru	0(72),NA(16),1(7)
4	763051	6.154373081	9.122772341	G	A	MSP5	Peru	0(72),NA(16),1(7)
7	1216431	4.851698267	5.912187518	A	C	MSP1	Peru	0(51),NA(25),1(19)
7	1216431	-4.851698267	5.912187518	A	C	MSP1	Peru	0(51),NA(25),1(19)
8	854440	-4.457431788	5.081197124	G	C	NA	Peru	0(87),1(4),NA(4)
8	854440	4.457431788	5.081197124	G	C	NA	Peru	0(87),1(4),NA(4)
14	3009263	-6.723893799	10.75219506	A	C	NA	Peru	0(60),NA(34),0.5(1)
14	3009263	6.723893799	10.75219506	A	C	NA	Peru	0(60),NA(34),0.5(1)
14	3009310	-4.889755621	5.995845329	T	A	NA	Peru	0(46),NA(36),0.5(13)
14	3009310	4.889755621	5.995845329	T	A	NA	Peru	0(46),NA(36),0.5(13)
14	3009317	-4.785561339	5.768249891	A	C	NA	Peru	0(49),NA(34),0.5(12)
14	3009317	4.785561339	5.768249891	A	C	NA	Peru	0(49),NA(34),0.5(12)
14	3009484	-4.54183988	5.253635251	A	G	PVP01_147040(	Peru	0(58),NA(20),0.5(15),1(2)
14	3009484	4.54183988	5.253635251	A	G	PVP01_147040(	Peru	0(58),NA(20),0.5(15),1(2)
14	3009706	-4.591568994	5.356620212	A	G	PVP01_147040(	Peru	0(67),NA(27),0.5(1)
14	3009706	4.591568994	5.356620212	A	G	PVP01_147040(	Peru	0(67),NA(27),0.5(1)

S19 Table. Candidate regions for positive selection within South America

Chr	Start	End	Subpopulation comparison	Number of markers	Mean Rsb markers	Max Rsb	Number of Rsb extreme markers	Percentage of extreme Rsb markers	Mean Rsb of extreme markers	Genes in the region	Products
3	240000	270000	Brazil|Panama	95	1.376	9.468	10	10.53	5.969	PVP01_0304600(MAg-1); PVP01_0304700(LISP2); PVP01_0304800(PRPF8); PVP01_0304900; PVP01_0305000	ag-1 blood stage membrane protein homologue; liver specific protein 2 putative; pre-mRNA-processing-splicing factor 8 putative; conserved Plasmodium protein unknown function; TMEM33 domain-containing protein putative
14	2260000	2290000	Brazil|Panama	36	2.588	6.206	3	8.33	5.795	PVP01_1451700(AARP1); PVP01_1451800; PVP01_1451900; PVP01_1452000(SENP1); PVP01_1452100; PVP01_1452200; PVP01_1452300	asparagine and aspartate rich protein 1 putative; homocysteine S-methyltransferase putative; conserved protein unknown function; sentrin-specific protease 1 putative; conserved Plasmodium protein unknown function; bromodomain protein putative; helicase putative
3	240000	270000	Brazil|Peru	125	1.949	6.394	6	4.8	5.742	PVP01_0304600(MAg-1); PVP01_0304700(LISP2); PVP01_0304800(PRPF8); PVP01_0304900; PVP01_0305000	ag-1 blood stage membrane protein homologue; liver specific protein 2 putative; pre-mRNA-processing-splicing factor 8 putative; conserved Plasmodium protein unknown function; TMEM33 domain-containing protein putative
4	750000	780000	Brazil|Peru	124	1.462	9.123	2	1.61	7.179	PVP01_0418100(IscA1); PVP01_0418200; PVP01_0418300(MSP4); PVP01_0418400(MSP5); PVP01_0418500(ADSL); PVP01_0418600; PVP01_0418700; PVP01_0418800(AROM)	iron-sulfur assembly protein putative; conserved Plasmodium protein unknown function; merozoite surface protein 4; merozoite surface protein 5; adenylosuccinate lyase; DNA-directed RNA polymerase III subunit RPC10 putative; conserved Plasmodium protein unknown function; pentafunctional AROM polypeptide putative
14	2990000	3020000	Brazil|Peru	13	3.166	10.752	5	38.46	6.625	PVP01_1469900(TRAG23); PVP01_1470000; PVP01_1470100(TRAG24); PVP01_1470200; PVP01_1470300; PVP01_1470400; PVP01_1470500; PVP01_1470600	tryptophan-rich protein; Plasmodium exported protein unknown function; tryptophan-rich protein; Plasmodium exported protein unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein (PHIST) unknown function
3	240000	270000	Mexico|Panama	65	1.227	7.791	2	3.08	6.797	PVP01_0304600(MAg-1); PVP01_0304700(LISP2); PVP01_0304800(PRPF8); PVP01_0304900; PVP01_0305000	ag-1 blood stage membrane protein homologue; liver specific protein 2 putative; pre-mRNA-processing-splicing factor 8 putative; conserved Plasmodium protein unknown function; TMEM33 domain-containing protein putative
14	2990000	3020000	Mexico|Panama	20	1.528	5.417	2	10	5.314	PVP01_1469900(TRAG23); PVP01_1470000; PVP01_1470100(TRAG24); PVP01_1470200; PVP01_1470300; PVP01_1470400; PVP01_1470500; PVP01_1470600	tryptophan-rich protein; Plasmodium exported protein unknown function; tryptophan-rich protein; Plasmodium exported protein unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein (PHIST) unknown function

S20 Table. Candidate regions for positive selection within Brazilian groupings

Chr	Start	End	Subpopulation comparison	Number of markers	Mean Rsb markers	Max Rsb	Number of Rsb extreme markers	Percentage of extreme Rsb markers	Mean Rsb of extreme markers	Genes in the region	Products
4	750000	780000	C1|C6	80	1.012	14.844	3	3.75	13.698	PVP01_0418100(IscA1); PVP01_0418200; PVP01_0418300(MSP4); PVP01_0418400(MSP5); PVP01_0418500(ADSL); PVP01_0418600; PVP01_0418700; PVP01_0418800(AROM)	iron-sulfur assembly protein putative; conserved Plasmodium protein unknown function; merozoite surface protein 4; merozoite surface protein 5; adenylosuccinate lyase; DNA-directed RNA polymerase III subunit RPC10 putative; conserved Plasmodium protein unknown function; pentafunctional AROM polypeptide putative
14	2990000	3020000	C1|C6	119	1.308	7.391	7	5.88	5.801	PVP01_1469900(TRAG23); PVP01_1470000; PVP01_1470100(TRAG24); PVP01_1470200; PVP01_1470300; PVP01_1470400; PVP01_1470500; PVP01_1470600	tryptophan-rich protein; Plasmodium exported protein unknown function; tryptophan-rich protein; Plasmodium exported protein unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein (PHIST) unknown function
7	1200000	1230000	C1|C7	198	2.738	20.717	51	25.76	6.534	PVP01_0728600; PVP01_0728700(LRR8); PVP01_0728800(MSP1P); PVP01_0728900(MSP1); PVP01_0729000; PVP01_0729100(DGK1)	heptatricopeptide repeat-containing protein putative; leucine-rich repeat protein; merozoite surface protein 1 paralog; merozoite surface protein 1; zinc finger protein putative; diacylglycerol kinase putative
14	3000000	3030000	C1|C7	7	3.937	13.354	2	28.57	13.263	PVP01_1470100(TRAG24); PVP01_1470200; PVP01_1470300; PVP01_1470400; PVP01_1470500; PVP01_1470600; PVP01_1470700	tryptophan-rich protein; Plasmodium exported protein unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein (PHIST) unknown function; hypothetical protein
3	240000	270000	C6|C7	81	1.605	10.479	4	4.94	7.595	PVP01_0304600(MAg-1); PVP01_0304700(LISP2); PVP01_0304800(PRPF8); PVP01_0304900; PVP01_0305000	ag-1 blood stage membrane protein homologue; liver specific protein 2 putative; pre-mRNA-processing-splicing factor 8 putative; conserved Plasmodium protein unknown function; TMEM33 domain-containing protein putative
4	750000	780000	C6|C7	72	0.533	9.211	2	2.78	8.664	PVP01_0418100(IscA1); PVP01_0418200; PVP01_0418300(MSP4); PVP01_0418400(MSP5); PVP01_0418500(ADSL); PVP01_0418600; PVP01_0418700; PVP01_0418800(AROM)	iron-sulfur assembly protein putative; conserved Plasmodium protein unknown function; merozoite surface protein 4; merozoite surface protein 5; adenylosuccinate lyase; DNA-directed RNA polymerase III subunit RPC10 putative; conserved Plasmodium protein unknown function; pentafunctional AROM polypeptide putative
7	1200000	1230000	C6|C7	217	2.692	10.891	44	20.28	7.944	PVP01_0728600; PVP01_0728700(LRR8); PVP01_0728800(MSP1P); PVP01_0728900(MSP1); PVP01_0729000; PVP01_0729100(DGK1)	heptatricopeptide repeat-containing protein putative; leucine-rich repeat protein; merozoite surface protein 1 paralog; merozoite surface protein 1; zinc finger protein putative; diacylglycerol kinase putative
14	40000	80000	C6|C7	52	3.661	9.751	15	28.85	6.703	PVP01_1401000; PVP01_1401100; PVP01_1401200; PVP01_1401300; PVP01_1401400; PVP01_1401500; PVP01_1401600; PVP01_1401700	Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; cytoadherence linked asexual protein CLAG putative; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function
14	3000000	3030000	C6|C7	6	2.892	8.764	2	33.33	8.334	PVP01_1470100(TRAG24); PVP01_1470200; PVP01_1470300; PVP01_1470400; PVP01_1470500; PVP01_1470600; PVP01_1470700	tryptophan-rich protein; Plasmodium exported protein unknown function; Plasmodium exported protein (PHIST) unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein (PHIST) unknown function; hypothetical protein

3	240000	270000	Group A|Group B	98	1.73	6.071	4	4.08	5.437	PVP01_0304600(MAg-1); PVP01_0304700(LISP2); PVP01_0304800(PRPF8); PVP01_0304900; PVP01_0305000	ag-1 blood stage membrane protein homologue; liver specific protein 2 putative; pre-mRNA-processing-splicing factor 8 putative; conserved Plasmodium protein unknown function; TMEM33 domain-containing protein putative
6	820000	850000	Group A|Group B	148	2.259	7.594	5	3.38	7.471	PVP01_0620200; PVP01_0620300; PVP01_0620400; PVP01_0620500; PVP01_0620600; PVP01_0620700; PVP01_0620800; PVP01_0620900; PVP01_0621000	conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; hypothetical protein; conserved Plasmodium protein unknown function; conserved Plasmodium protein unknown function; hypothetical protein; conserved Plasmodium protein unknown function; hypothetical protein; conserved protein unknown function
6	970000	1000000	Group A|Group B	23	1.916	6.597	3	13.04	6.196	PVP01_0623600; PVP01_0623700; PVP01_0623800(DBP); PVP01_0623900(REX4); PVP01_0624000; PVP01_0624100	PIR protein; Plasmodium exported protein (PHIST) unknown function; duffy binding protein; ring-exported protein 4 putative; Plasmodium exported protein unknown function; PIR protein
7	1200000	1230000	Group A|Group B	220	1.813	6.756	21	9.55	5.959	PVP01_0728600; PVP01_0728700(LRR8); PVP01_0728800(MSP1P); PVP01_0728900(MSP1); PVP01_0729000; PVP01_0729100(DGK1)	heptatricopeptide repeat-containing protein putative; leucine-rich repeat protein; merozoite surface protein 1 paralog; merozoite surface protein 1; zinc finger protein putative; diacylglycerol kinase putative
14	40000	70000	Group A|Group B	31	3.615	12.681	12	38.71	7.371	PVP01_1401000; PVP01_1401100; PVP01_1401200; PVP01_1401300; PVP01_1401400; PVP01_1401500	Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; Plasmodium exported protein unknown function; cytoadherence linked asexual protein CLAG putative; Plasmodium exported protein unknown function

S21 Table. Tajima's D metric for monoclonal isolates within South America

Top 50 genes with highest and lowest D metric in South America presented, with reference to the same position in other regions. If gene was discarded in the regional analysis (less than 5 SNPs) is is shown in red font.

Chr	Start	End	Gene ID	Gene name	Gene product	South America	Southeast Asia	Southern SEA	South Asia	East Africa
PvP01_10_v1	1302679	1303648	PVP01_1030200	RPL31	60S ribosomal protein L31; putative	4.459	-1.854	-1.577	2.456	3.044
PvP01_10_v1	1407109	1408530	PVP01_1033100	NA	conserved protein; unknown function	4.004	-0.934	-1.982	3.595	2.931
PvP01_12_v1	2996079	2996585	PVP01_1270900	NA	hypothetical protein	3.757	-1.884	-1.577	0.394	0.584
PvP01_06_v1	998881	1000731	PVP01_0624100	NA	PIR protein	3.562	2.335	2.853	2.617	3.342
PvP01_06_v1	87703	88206	PVP01_0602100	ETRAMP	early transcribed membrane protein	2.989	-0.702	-1.474	-0.067	-1.285
PvP01_06_v1	919969	922328	PVP01_0622700	NA	conserved Plasmodium protein; unknown function	2.956	-0.632	0.19	0.89	1.382
PvP01_06_v1	838600	839634	PVP01_0620700	NA	hypothetical protein	2.924	1.772	0.765	2.036	2.146
PvP01_02_v1	391308	392067	PVP01_0209400	NA	conserved Plasmodium protein; unknown function	2.791	0.744	0.453	0.697	1.459
PvP01_12_v1	2986799	2988087	PVP01_1270700	NA	conserved Plasmodium protein; unknown function	2.778	1.558	1.426	3.357	0.141
PvP01_05_v1	1016909	1017727	PVP01_0525100	NA	Plasmodium exported protein (PHIST); unknown function	2.671	3.957	2.239	1.487	1.204
PvP01_07_v1	1215432	1220636	PVP01_0728900	MSP1	merozoite surface protein 1	2.637	2.073	1.108	2.035	1.917
PvP01_13_v1	153829	156796	PVP01_1303600	NA	NA	2.621	0.212	-1.553	2.844	1.467
PvP01_13_v1	2013814	2015366	PVP01_1346200	NA	ribosomal protein S27a; putative	2.4	1.311	-1.199	2.616	2.453
PvP01_09_v1	1322076	1322945	PVP01_0930400	NA	conserved Plasmodium protein; unknown function	2.378	-1.633	-1.976	-0.356	-0.472
PvP01_06_v1	993570	994772	PVP01_0624000	NA	Plasmodium exported protein; unknown function	2.309	-1.022	0.273	2.059	0.799
PvP01_10_v1	151961	153955	PVP01_1002800	NA	SURF1 domain-containing protein; putative	2.279	0.843	2.185	0.557	1.15
PvP01_12_v1	1880027	1882480	PVP01_1245700	NA	conserved Plasmodium protein; unknown function	2.259	-1.413	1.697	-0.316	-1.496
PvP01_07_v1	593112	594800	PVP01_0712600	NA	phospholipid or glycerol acyltransferase; putative	2.244	-0.805	1.663	0.182	0.118
PvP01_06_v1	828653	830320	PVP01_0620400	NA	hypothetical protein	2.093	0.729	1.028	2.022	1.442
PvP01_04_v1	762486	764022	PVP01_0418400	MSP5	merozoite surface protein 5	2.06	1.8	1.109	1.868	1.558
PvP01_12_v1	1912137	1914080	PVP01_1246700	NA	kelch domain-containing protein; putative	2.06	-0.937	-0.171	0.092	1.123
PvP01_04_v1	624235	626468	PVP01_0415400	NA	conserved protein; unknown function	2.032	-1.703	-0.887	1.031	0.603
PvP01_07_v1	421180	424353	PVP01_0708200	NA	FHA domain protein; putative	2.026	-2.001	0.274	-1.279	0.064
PvP01_04_v1	946292	947893	PVP01_0423200	NA	Plasmodium exported protein (PHIST); unknown function	1.984	-1.185	1.201	-0.49	0.487
PvP01_03_v1	245783	253342	PVP01_0304700	LISP2	liver specific protein 2; putative	1.968	0.821	0.117	1.423	1.038
PvP01_14_v1	51257	53698	PVP01_1401300	NA	Plasmodium exported protein; unknown function	1.944	-0.14	-1.271	1.336	0.843
PvP01_04_v1	340944	342863	PVP01_0408400	NA	MtN3-like protein	1.906	-0.334	1.333	1.034	0.37
PvP01_07_v1	878869	881193	PVP01_0720200	NA	conserved protein; unknown function	1.904	-0.749	0.999	1.148	1.165
PvP01_12_v1	788281	789447	PVP01_1219200	NA	hypothetical protein	1.895	0.335	-0.091	0.867	1.248
PvP01_05_v1	1368607	1369931	PVP01_0532400	CyRPA	cysteine-rich protective antigen; putative	1.86	0.74	0.908	1.662	2.066
PvP01_14_v1	1097783	1098889	PVP01_1425600	NA	conserved Plasmodium protein; unknown function	1.854	-0.717	0.384	-1.044	0.492
PvP01_12_v1	1357499	1358584	PVP01_1233900	NA	dihydroorotase; putative	1.826	-0.987	-0.102	-0.206	0.062
PvP01_07_v1	71450	80153	PVP01_0701200	RBP1a	reticulocyte binding protein 1a	1.808	-0.728	-0.065	0.148	0.144
PvP01_06_v1	933225	934273	PVP01_0623000	NA	Plasmodium exported protein (PHIST); unknown function	1.786	-1.452	-0.112	-0.145	0.277
PvP01_07_v1	1226520	1230924	PVP01_0729100	DGK1	diacylglycerol kinase; putative	1.777	-1.289	-1.247	-0.321	0.019
PvP01_06_v1	443798	446878	PVP01_0610100	NA	conserved Plasmodium protein; unknown function	1.749	-2.121	-1.466	-1.159	-0.331
PvP01_03_v1	118170	129734	PVP01_0302600	NA	conserved Plasmodium protein; unknown function	1.719	0.422	0.64	1.08	0.929
PvP01_09_v1	2150027	2151219	PVP01_0948800	TRAG16	tryptophan-rich protein	1.713	2.611	0.438	2.396	2.403
PvP01_11_v1	1809799	1810656	PVP01_1142500	NA	conserved Plasmodium protein; unknown function	1.696	-0.755	0.58	-0.067	0.742

PvP01_10_v1	1224465	1226132	PVP01_1028700	NA	UDP-N-acetylglucosamine transporter; putative	1.675	-2.154	-0.93	-1.67	0.228
PvP01_08_v1	507941	510898	PVP01_0811900	PREBP	PRE-binding protein; putative	1.655	-1.377	0.204	0.841	-0.336
PvP01_14_v1	702419	703273	PVP01_1415900	NA	conserved Plasmodium protein; unknown function	1.633	-0.988	0.077	-0.125	-0.279
PvP01_02_v1	744399	745643	PVP01_0217200	NA	Plasmodium exported protein; unknown function	1.627	-0.349	-0.024	0.927	1.277
PvP01_12_v1	1015349	1016629	PVP01_1226000	NA	conserved protein; unknown function	1.625	-1.953	-1.315	-0.239	-1.668
PvP01_07_v1	1459636	1463494	PVP01_0735200	NA	Plasmodium exported protein; unknown function	1.618	-0.008	1.946	1.588	0.481
PvP01_11_v1	786679	789439	PVP01_1118700	NA	TPH domain-containing protein; putative	1.618	-1.573	-1.127	0.131	-0.403
PvP01_12_v1	2099317	2105028	PVP01_1251300	NA	NA	1.618	-1.985	-1.403	-0.866	-0.532
PvP01_01_v1	604221	605585	PVP01_0113500	NA	conserved protein; unknown function	1.566	-0.307	1.315	0.09	0.737
PvP01_10_v1	188715	190337	PVP01_1003700	PPT	phosphoenolpyruvate/phosphate translocator; putative	1.528	-0.286	0.072	1.006	0.812
PvP01_03_v1	195716	197179	PVP01_0303800	P52	6-cysteine protein P52; putative	1.499	-1.6	-0.679	-0.988	-1.608
PvP01_11_v1	1132983	1135137	PVP01_1125900	NA	conserved Plasmodium protein; unknown function	-1.677	-2.321	-1.448	-1.506	-0.283
PvP01_08_v1	450779	451671	PVP01_0810100	NA	ribosomal protein L43; mitochondrial; putative	-1.678	-2.136	-1.088	-1.37	-1.536
PvP01_14_v1	544696	546045	PVP01_1412300	NA	ATP synthase subunit gamma; mitochondrial; putative	-1.678	-2.15	-1.478	-1.426	-0.81
PvP01_14_v1	1695387	1697078	PVP01_1439700	IMC1h	inner membrane complex protein 1h; putative	-1.678	-2.404	-0.937	0.093	-0.838
PvP01_01_v1	541039	543375	PVP01_0111800	SAP18	histone deacetylase complex subunit SAP18; putative	-1.683	-2.183	-1.004	-1.088	0.255
PvP01_04_v1	420492	425357	PVP01_0410300	NA	serine/threonine protein kinase; putative	-1.683	-2.152	-0.551	-1.499	-1.185
PvP01_13_v1	1431000	1434272	PVP01_1333300	NA	conserved Plasmodium protein; unknown function	-1.689	-2.165	-0.208	-1.804	-1.545
PvP01_14_v1	682467	683897	PVP01_1415200	NA	MACRO domain-containing protein; putative	-1.689	-0.461	-0.511	-0.584	-0.177
PvP01_05_v1	1327415	1328581	PVP01_0531200	EIF4A3	eukaryotic initiation factor 4A-III; putative	-1.694	-1.755	NA	-0.345	-0.896
PvP01_07_v1	398353	399995	PVP01_0707700	POFUT2	GDP-fucose protein O-fucosyltransferase 2	-1.695	-2.217	-1.681	0.356	-1.668
PvP01_09_v1	2001384	2007419	PVP01_0945900	NA	serine/threonine protein kinase; putative	-1.697	-2.42	-2.093	-1.681	-0.921
PvP01_14_v1	2579586	2581301	PVP01_1459000	NA	tetratricopeptide repeat protein; putative	-1.699	-1.458	-1.586	-0.076	-0.142
PvP01_11_v1	1640983	1642650	PVP01_1138800	RBM25	RNA-binding protein 25; putative	-1.711	-1.063	-1.786	-1.105	0.234
PvP01_12_v1	680277	682770	PVP01_1216300	NA	SprT-like domain-containing protein; putative	-1.714	-1.385	-0.061	-0.881	-0.564
PvP01_12_v1	1634644	1641396	PVP01_1240600	NA	conserved Plasmodium protein; unknown function	-1.734	-2.192	-0.872	-0.689	-0.608
PvP01_07_v1	1447577	1448712	PVP01_0734900	NA	Plasmodium exported protein (PHIST); unknown function	-1.736	-1.869	1.476	0.363	1.058
PvP01_12_v1	176234	180121	PVP01_1204300	NA	conserved Plasmodium protein; unknown function	-1.74	-1.991	-0.632	-0.962	-0.745
PvP01_12_v1	1443594	1446876	PVP01_1236300	NA	NA	-1.74	-2.107	-1.369	-0.126	0.597
PvP01_07_v1	1135807	1137261	PVP01_0726600	NA	conserved Plasmodium protein; unknown function	-1.744	-2.046	0.434	-0.092	-0.896
PvP01_11_v1	1459365	1460576	PVP01_1134000	ENR	enoyl-acyl carrier reductase; putative	-1.749	-0.6	0.858	0.537	-1.162
PvP01_08_v1	513014	514942	PVP01_0812000	HO	heme oxygenase; putative	-1.754	-1.882	-1.4	-0.549	-0.548
PvP01_03_v1	155417	158020	PVP01_0303100	NA	alpha/beta hydrolase; putative	-1.758	-1.663	-0.259	-1.346	0.404
PvP01_03_v1	135393	138083	PVP01_0302800	NA	conserved Plasmodium protein; unknown function	-1.767	-2.459	-0.298	-1.576	-0.657
PvP01_05_v1	1171930	1173693	PVP01_0528300	NA	conserved protein; unknown function	-1.768	-1.864	-1.796	-1.356	-0.253
PvP01_03_v1	679409	680839	PVP01_0315200	NA	NA	-1.77	-1.492	-0.888	-1.032	0.076
PvP01_01_v1	767750	769801	PVP01_0116900	NA	mitogen-activated protein kinase organizer 1; putative	-1.771	-1.638	-1.823	-0.658	-1.359
PvP01_02_v1	645756	646916	PVP01_0215300	NA	conserved Plasmodium protein; unknown function	-1.783	-2.387	1.623	-1.604	-0.097
PvP01_07_v1	388263	390670	PVP01_0707400	NA	conserved Plasmodium protein; unknown function	-1.79	-2.046	-1.366	-0.802	-0.7
PvP01_03_v1	451006	452275	PVP01_0309900	NA	conserved protein; unknown function	-1.793	-2.251	-1.182	-0.911	NA
PvP01_03_v1	410337	413312	PVP01_0308600	NA	methyltransferase; putative	-1.801	-2.021	-0.191	-0.601	0.092
PvP01_06_v1	141806	146626	PVP01_0603800	NA	tRNA wybutosine-synthesizing protein; putative	-1.801	-2.311	-1.524	-1.062	-0.392
PvP01_02_v1	140093	144454	PVP01_0202700	NA	28S ribosomal RNA	-1.802	-1.466	-2.069	-1.47	-1.32
PvP01_11_v1	1456103	1458269	PVP01_1133900	NA	conserved protein; unknown function	-1.808	-2.268	-0.017	-0.541	0.163
PvP01_04_v1	892242	897866	PVP01_0422100	ROPE	repetitive organellar protein; putative	-1.816	-2.162	-1.021	-1.325	-0.796
PvP01_13_v1	885864	887410	PVP01_1319400	NA	conserved protein; unknown function	-1.837	-1.56	-0.698	-1.249	-0.407
PvP01_13_v1	121737	124154	PVP01_1302600	NA	TBCC domain-containing protein; putative	-1.84	-1.963	-1.224	-1.067	0.88
PvP01_11_v1	1769863	1771244	PVP01_1141700	UROD	uroporphyrinogen III decarboxylase; putative	-1.865	-2.217	-1.376	-1.426	-1.315
PvP01_12_v1	2336848	2338350	PVP01_1256900	SRP54	signal recognition particle subunit SRP54; putative	-1.868	-2.382	-0.334	0.761	-0.675
PvP01_13_v1	166536	169378	PVP01_1304100	NA	conserved protein; unknown function	-1.89	-1.6	-0.301	0.355	-0.782

PvP01_04_v1	376027	380175	PVP01_0409300	RPB2	DNA-directed RNA polymerase II subunit RPB2; putative	-1.897	-2.447	-1.766	-1.623	-0.873
PvP01_14_v1	1438510	1439052	PVP01_1433200	EMC6	ER membrane protein complex subunit 6; putative	-1.903	-1.631	-1.448	-0.716	-1.086
PvP01_12_v1	385620	390835	PVP01_1209400	NA	conserved Plasmodium protein; unknown function	-1.923	-2.338	-0.4	-1.31	-0.676
PvP01_10_v1	176943	181553	PVP01_1003500	NA	conserved Plasmodium protein; unknown function	-1.929	-1.67	-2.236	-2.131	-1.303
PvP01_09_v1	1174659	1177421	PVP01_0926700	ThrRS	threonine--tRNA ligase; putative	-1.944	-2.08	0.606	-1.194	-0.186
PvP01_API_v1	22760	24169	PVP01_API04600	sufB	FeS cluster assembly protein SufB; putative	-1.964	-2.479	-0.318	0.339	1.462
PvP01_14_v1	2129881	2133622	PVP01_1449200	NA	conserved Plasmodium protein; unknown function	-2.096	-1.824	0.196	-0.902	-0.713
PvP01_12_v1	2618462	2624188	PVP01_1263400	NA	gamma-tubulin complex component; putative	-2.116	-2.567	-1.318	-0.041	-1.154
PvP01_API_v1	17469	19190	PVP01_API04200	rpoC	NA	-2.116	-2.382	-1.598	-0.364	-0.638
PvP01_API_v1	8576	9805	PVP01_API02800	TUFA	elongation factor Tu; putative	-2.142	-2.532	NA	0.129	1.305
PvP01_API_v1	10790	13081	PVP01_API03500	ClpM	chaperone protein ClpM; putative	-2.221	-2.529	0.095	-1.614	-2.241

S22 Table. Tajima's D metric for monoclonal isolates within Brazil

Top 50 genes with highest and lowest D metric in Brazil.

Chr	Gene ID	Start	End	Gene name	Gene product	Tajima's D	Number of SNPs
PvP01_06_v1	PVP01_0624100	998881	1000731	NA	PIR protein	3.743	35
PvP01_12_v1	PVP01_1270900	2996079	2996585	NA	hypothetical protein	3.197	7
PvP01_06_v1	PVP01_0602100	87703	88206	ETRAMP	early transcribed membrane protein	3.116	5
PvP01_12_v1	PVP01_1245700	1880027	1882480	NA	conserved Plasmodium protein; unknown function	3.062	6
PvP01_06_v1	PVP01_0620700	838600	839634	NA	hypothetical protein	2.833	52
PvP01_10_v1	PVP01_1033100	1407109	1408530	NA	conserved protein; unknown function	2.83	31
PvP01_06_v1	PVP01_0622700	919969	922328	NA	conserved Plasmodium protein; unknown function	2.759	14
PvP01_12_v1	PVP01_1270700	2986799	2988087	NA	conserved Plasmodium protein; unknown function	2.754	16
PvP01_05_v1	PVP01_0525100	1016909	1017727	NA	Plasmodium exported protein (PHIST); unknown function	2.588	18
PvP01_13_v1	PVP01_1303600	153829	156796	NA	NA	2.515	42
PvP01_02_v1	PVP01_0209400	391308	392067	NA	conserved Plasmodium protein; unknown function	2.506	8
PvP01_14_v1	PVP01_1470400	3009446	3010524	NA	Plasmodium exported protein; unknown function	2.465	37
PvP01_09_v1	PVP01_0948800	2150027	2151219	TRAG16	tryptophan-rich protein	2.394	24
PvP01_10_v1	PVP01_1030200	1302679	1303648	RPL31	60S ribosomal protein L31; putative	2.347	34
PvP01_14_v1	PVP01_1401300	51257	53698	NA	Plasmodium exported protein; unknown function	2.334	8
PvP01_08_v1	PVP01_0819500	855540	857351	SPC2	NA	2.316	6
PvP01_07_v1	PVP01_0728900	1215432	1220636	MSP1	merozoite surface protein 1	2.239	231
PvP01_04_v1	PVP01_0415400	624235	626468	NA	conserved protein; unknown function	2.15	8
PvP01_14_v1	PVP01_1428400	1211301	1215809	NA	RNA-binding protein; putative	2.143	8
PvP01_13_v1	PVP01_1345100	1969454	1970903	TRAM	translocation associated membrane protein; putative	2.117	7
PvP01_07_v1	PVP01_0712600	593112	594800	NA	phospholipid or glycerol acyltransferase; putative	2.11	5
PvP01_06_v1	PVP01_0620400	828653	830320	NA	hypothetical protein	2.101	68
PvP01_14_v1	PVP01_1450200	2192366	2194162	aFRS	phenylalanine--tRNA ligase; putative	2.075	6
PvP01_03_v1	PVP01_0304700	245783	253342	LISP2	liver specific protein 2; putative	2.063	119
PvP01_11_v1	PVP01_1142500	1809799	1810656	NA	conserved Plasmodium protein; unknown function	2.045	6
PvP01_11_v1	PVP01_1131100	1334496	1337188	NA	LMBR1 domain-containing protein; putative	2.021	7
PvP01_10_v1	PVP01_1023500	1021490	1024285	NA	topoisomerase I; putative	2.013	5
PvP01_12_v1	PVP01_1241700	1701589	1704096	NA	NA	2.008	5
PvP01_07_v1	PVP01_0729100	1226520	1230924	DGK1	diacylglycerol kinase; putative	1.995	10
PvP01_13_v1	PVP01_1328500	1207397	1209925	GlyRS	glycine--tRNA ligase; putative	1.984	7
PvP01_06_v1	PVP01_0624000	993570	994772	NA	Plasmodium exported protein; unknown function	1.982	10
PvP01_06_v1	PVP01_0605400	218286	221362	NA	conserved Plasmodium protein; unknown function	1.981	7
PvP01_07_v1	PVP01_0709800	495347	505297	CRMP1	cysteine repeat modular protein 1; putative	1.98	38
PvP01_13_v1	PVP01_1324000	1072307	1074309	NA	conserved Plasmodium protein; unknown function	1.971	5
PvP01_07_v1	PVP01_0708200	421180	424353	NA	FHA domain protein; putative	1.964	7
PvP01_06_v1	PVP01_0609900	434293	435921	NA	conserved protein; unknown function	1.945	5
PvP01_05_v1	PVP01_0532400	1368607	1369931	CyRPA	cysteine-rich protective antigen; putative	1.918	36
PvP01_03_v1	PVP01_0302600	118170	129734	NA	conserved Plasmodium protein; unknown function	1.898	258
PvP01_07_v1	PVP01_0701200	71450	80153	RBP1a	reticulocyte binding protein 1a	1.898	48

PvP01_11_v1	PVP01_1106700	266825	269071	NA	conserved Plasmodium protein; unknown function	1.892	6
PvP01_14_v1	PVP01_1430400	1310045	1316657	JmjC1	JmjC domain-containing protein; putative	1.878	15
PvP01_01_v1	PVP01_0111400	527779	531048	Cdc48	AAA family ATPase; CDC48 subfamily; putative	1.874	11
PvP01_11_v1	PVP01_1105200	208352	211133	NA	phosphatidylserine synthase; putative	1.87	9
PvP01_04_v1	PVP01_0405900	259211	260647	NA	saccharopine dehydrogenase; putative	1.858	6
PvP01_02_v1	PVP01_0217200	744399	745643	NA	Plasmodium exported protein; unknown function	1.856	13
PvP01_06_v1	PVP01_0613700	582340	583675	NA	conserved Plasmodium protein; unknown function	1.848	6
PvP01_13_v1	PVP01_1317200	805666	810872	NA	conserved Plasmodium protein; unknown function	1.844	9
PvP01_05_v1	PVP01_0511900	510175	512782	NA	conserved Plasmodium protein; unknown function	1.84	6
PvP01_06_v1	PVP01_0612200	536695	538812	MPP10	U3 small nucleolar ribonucleoprotein protein MPP10; putative	1.785	7
PvP01_06_v1	PVP01_0605600	226050	227855	NA	NA	1.778	7
PvP01_08_v1	PVP01_0830100	1311078	1313873	CDPK3	calcium-dependent protein kinase 3; putative	-1.476	5
PvP01_09_v1	PVP01_0935700	1536464	1538461	NA	conserved Plasmodium protein; unknown function	-1.478	5
PvP01_12_v1	PVP01_1235500	1412264	1414593	NA	RAP protein; putative	-1.479	5
PvP01_03_v1	PVP01_0316900	735486	737984	NA	conserved Plasmodium protein; unknown function	-1.491	10
PvP01_03_v1	PVP01_0312300	537873	540540	NA	conserved Plasmodium protein; unknown function	-1.498	5
PvP01_14_v1	PVP01_1427700	1180438	1183776	NA	RNA-binding protein; putative	-1.498	5
PvP01_12_v1	PVP01_1228000	1119462	1122303	NA	translation initiation factor eIF-2B subunit gamma; putative	-1.502	7
PvP01_07_v1	PVP01_0712700	597669	600018	NA	met-10+ like protein; putative	-1.503	6
PvP01_03_v1	PVP01_0308200	394770	397654	KAE1	tRNA N6-adenosine threonylcarbamoyltransferase; putative	-1.506	9
PvP01_02_v1	PVP01_0203900	189245	192562	NA	P-loop containing nucleoside triphosphate hydrolase; putative	-1.508	7
PvP01_05_v1	PVP01_0507600	335035	336445	RACK	NA	-1.51	8
PvP01_14_v1	PVP01_1420300	881234	886322	NA	conserved Plasmodium protein; unknown function	-1.51	15
PvP01_11_v1	PVP01_1109300	367043	371278	NA	conserved Plasmodium protein; unknown function	-1.522	11
PvP01_12_v1	PVP01_1240600	1634644	1641396	NA	conserved Plasmodium protein; unknown function	-1.522	16
PvP01_11_v1	PVP01_1109000	349163	352384	NA	conserved Plasmodium protein; unknown function	-1.543	11
PvP01_10_v1	PVP01_1007100	307222	309309	NA	conserved protein; unknown function	-1.556	6
PvP01_09_v1	PVP01_0922300	964795	967173	GCN20	protein GCN20; putative	-1.558	5
PvP01_09_v1	PVP01_0927400	1203603	1208951	ATG23	autophagy-related protein 23; putative	-1.561	19
PvP01_09_v1	PVP01_0946500	2030212	2033283	NA	conserved Plasmodium protein; unknown function	-1.563	6
PvP01_12_v1	PVP01_1204400	180938	182695	NA	ubiquitin-activating enzyme E1; putative	-1.564	5
PvP01_07_v1	PVP01_0726600	1135807	1137261	NA	conserved Plasmodium protein; unknown function	-1.566	5
PvP01_12_v1	PVP01_1216300	680277	682770	NA	SprT-like domain-containing protein; putative	-1.583	14
PvP01_11_v1	PVP01_1123500	1030329	1034924	NA	conserved Plasmodium protein; unknown function	-1.592	9
PvP01_10_v1	PVP01_1003500	176943	181553	NA	conserved Plasmodium protein; unknown function	-1.625	44
PvP01_14_v1	PVP01_1429300	1262207	1264847	CUL1	cullin-1; putative	-1.632	11
PvP01_05_v1	PVP01_0533000	1396484	1397908	TRAG12	tryptophan-rich protein	-1.633	6
PvP01_10_v1	PVP01_1032300	1380644	1381886	NA	conserved protein; unknown function	-1.633	6
PvP01_11_v1	PVP01_1143800	1857791	1860190	NA	RNA-binding protein; putative	-1.645	5
PvP01_API_v1	PVP01_API04600	22760	24169	sufB	FeS cluster assembly protein SufB; putative	-1.662	10
PvP01_12_v1	PVP01_1246600	1911090	1911950	NA	conserved protein; unknown function	-1.666	7
PvP01_13_v1	PVP01_1304100	166536	169378	NA	conserved protein; unknown function	-1.674	9
PvP01_13_v1	PVP01_1319400	885864	887410	NA	conserved protein; unknown function	-1.694	6
PvP01_12_v1	PVP01_1209400	385620	390835	NA	conserved Plasmodium protein; unknown function	-1.699	10
PvP01_12_v1	PVP01_1203800	155262	158072	NA	phosphatidylinositol transfer protein; putative	-1.706	9

PvP01_12_v1	PVP01_1263400	2618462	2624188	NA	gamma-tubulin complex component; putative	-1.72	12
PvP01_13_v1	PVP01_1330000	1271625	1274418	DBP4	ATP-dependent RNA helicase DBP4; putative	-1.735	5
PvP01_07_v1	PVP01_0703200	184756	186593	RPA1	replication protein A1; small fragment; putative	-1.737	5
PvP01_12_v1	PVP01_1251600	2115730	2117604	NA	conserved protein; unknown function	-1.737	5
PvP01_04_v1	PVP01_0405400	248395	251232	RFC1	replication factor C subunit 1; putative	-1.75	10
PvP01_01_v1	PVP01_0105300	284868	287291	NA	conserved Plasmodium protein; unknown function	-1.761	11
PvP01_01_v1	PVP01_0105600	296834	299332	MCM7	DNA replication licensing factor MCM7; putative	-1.776	10
PvP01_09_v1	PVP01_0926700	1174659	1177421	ThrRS	threonine--tRNA ligase; putative	-1.785	9
PvP01_11_v1	PVP01_1129200	1257939	1260209	NA	conserved Plasmodium protein; unknown function	-1.788	6
PvP01_14_v1	PVP01_1408700	398419	400192	NA	FHA domain-containing protein; putative	-1.808	11
PvP01_12_v1	PVP01_1256900	2336848	2338350	SRP54	signal recognition particle subunit SRP54; putative	-1.837	14
PvP01_05_v1	PVP01_0532800	1386447	1387761	TRAG10	tryptophan-rich protein	-1.93	8
PvP01_API_v1	PVP01_API04200	17469	19190	rpoC	NA	-1.973	14
PvP01_API_v1	PVP01_API02800	8576	9805	TUFA	elongation factor Tu; putative	-2.002	7
PvP01_14_v1	PVP01_1407200	341021	344257	NA	SAM dependent methyltransferase; putative	-2.031	14
PvP01_API_v1	PVP01_API03500	10790	13081	ClpM	chaperone protein ClpM; putative	-2.315	12

S23 Table. Total SNPs within genes of interest in Brazilian isolates of *P. vivax* (n = 123)

Chr	Pos	Ref	Alt	Alternate calls	Missing	Reference calls	Alternate frequency	Effect	Gene name	Gene ID	Amino acid change	Nucleotide change 1	Nucleotide change 2
PvP01_01_v1	444779	T	C	114	9	0	1	synonymous	CRT	PVP01_0109300	328D	444779T>C	
PvP01_01_v1	444987	C	T	16	3	104	0.13	synonymous	CRT	PVP01_0109300	339F	444987C>T	
PvP01_01_v1	445381	T	C	110	13	0	1	synonymous	CRT	PVP01_0109300	387S	445381T>C	
PvP01_02_v1	154107	G	T	14	55	54	0.21	missense	MRP1	PVP01_0203000	1606A>1606N	154107G>T+154108C>T	
PvP01_02_v1	154108	C	T	14	58	51	0.22	missense	MRP1	PVP01_0203000	1606A>1606N	154107G>T+154108C>T	
PvP01_02_v1	154168	G	A	2	49	72	0.03	missense	MRP1	PVP01_0203000	1586H>1586Y	154168G>A	
PvP01_02_v1	154229	G	A	1	47	75	0.01	synonymous	MRP1	PVP01_0203000	1565I	154229G>A	
PvP01_02_v1	154294	C	T	15	50	58	0.21	missense	MRP1	PVP01_0203000	1544V>1544I	154294C>T	
PvP01_02_v1	154350	G	A	2	42	79	0.02	missense	MRP1	PVP01_0203000	1525T>1525I	154350G>A	
PvP01_02_v1	154424	T	C	1	39	83	0.01	synonymous	MRP1	PVP01_0203000	1500T	154424T>C	
PvP01_02_v1	154486	T	C	2	42	79	0.02	missense	MRP1	PVP01_0203000	1480I>1480V	154486T>C	
PvP01_02_v1	154492	T	C	5	42	76	0.06	missense	MRP1	PVP01_0203000	1478I>1478V	154492T>C	
PvP01_02_v1	154668	C	G	20	47	56	0.26	missense	MRP1	PVP01_0203000	1419G>1419A	154668C>G	
PvP01_02_v1	155080	A	T	1	36	86	0.01	missense	MRP1	PVP01_0203000	1282L>1282I	155080A>T	
PvP01_02_v1	155206	A	T	2	47	74	0.03	missense	MRP1	PVP01_0203000	1240F>1240I	155206A>T	
PvP01_02_v1	155228	T	G	29	53	41	0.41	synonymous	MRP1	PVP01_0203000	1232I	155228T>G	
PvP01_02_v1	155732	A	G	22	43	58	0.28	synonymous	MRP1	PVP01_0203000	1064Y	155732A>G	
PvP01_02_v1	156089	G	T	1	47	75	0.01	missense	MRP1	PVP01_0203000	945F>945L	156089G>T	
PvP01_02_v1	156098	G	T	1	45	77	0.01	synonymous	MRP1	PVP01_0203000	942G	156098G>T	
PvP01_02_v1	156208	C	G	25	52	46	0.35	missense	MRP1	PVP01_0203000	906E>906Q	156208C>G	
PvP01_02_v1	157100	C	T	3	48	72	0.04	synonymous	MRP1	PVP01_0203000	608E	157100C>T	
PvP01_02_v1	157143	C	T	3	45	75	0.04	missense	MRP1	PVP01_0203000	594R>594K	157143C>T	
PvP01_02_v1	157220	A	G	25	54	44	0.36	synonymous	MRP1	PVP01_0203000	568N	157220A>G	
PvP01_02_v1	157457	C	T	3	46	74	0.04	synonymous	MRP1	PVP01_0203000	489V	157457C>T	
PvP01_02_v1	158272	A	C	58	65	0	1	missense	MRP1	PVP01_0203000	218Y>218D	158272A>C	
PvP01_02_v1	158273	A	G	58	65	0	1	synonymous	MRP1	PVP01_0203000	217N	158273A>G	
PvP01_02_v1	158545	C	T	60	63	0	1	missense	MRP1	PVP01_0203000	127V>127I	158545C>T	
PvP01_02_v1	158818	T	G	1	49	73	0.01	missense	MRP1	PVP01_0203000	36K>36Q	158818T>G	
PvP01_02_v1	415315	A	G	35	10	78	0.31	synonymous	UBP1	PVP01_0210400	3276A	415315A>G	
PvP01_02_v1	415407	G	T	24	9	90	0.21	missense	UBP1	PVP01_0210400	3246P>3246T	415407G>T	
PvP01_02_v1	415490	T	C	112	8	3	0.97	missense	UBP1	PVP01_0210400	3218N>3218S	415490T>C	
PvP01_02_v1	415675	G	A	3	3	117	0.03	synonymous	UBP1	PVP01_0210400	3156P	415675G>A	
PvP01_02_v1	416004	C	A	4	4	115	0.03	missense	UBP1	PVP01_0210400	3047A>3047S	416004C>A	
PvP01_02_v1	416034	T	G	21	5	97	0.18	synonymous	UBP1	PVP01_0210400	3037R	416034T>G	
PvP01_02_v1	416378	C	A	2	13	108	0.02	missense	UBP1	PVP01_0210400	2922S>2922I	416378C>A	
PvP01_02_v1	416416	G	T	1	7	115	0.01	missense	UBP1	PVP01_0210400	2909N>2909K	416416G>T	
PvP01_02_v1	416893	G	C	6	6	111	0.05	missense	UBP1	PVP01_0210400	2750F>2750L	416893G>C	
PvP01_02_v1	416896	G	A	71	8	44	0.62	synonymous	UBP1	PVP01_0210400	2749G	416896G>A	
PvP01_02_v1	417401	C	A	1	5	117	0.01	missense	UBP1	PVP01_0210400	2620D>2620Y	417401C>A	
PvP01_02_v1	417594	C	T	60	7	56	0.52	synonymous	UBP1	PVP01_0210400	2555K	417594C>T	
PvP01_02_v1	418197	C	A	6	7	110	0.05	missense	UBP1	PVP01_0210400	2354R>2354S	418197C>A	
PvP01_02_v1	418249	G	C	42	9	72	0.37	missense	UBP1	PVP01_0210400	2337P>2337R	418249G>C	
PvP01_02_v1	418478	C	A	75	12	36	0.68	missense	UBP1	PVP01_0210400	2261A>2261S	418478C>A	
PvP01_02_v1	418617	A	G	116	7	0	1	synonymous	UBP1	PVP01_0210400	2214Y	418617A>G	
PvP01_02_v1	418628	A	G	6	3	114	0.05	synonymous	UBP1	PVP01_0210400	2211L	418628A>G	
PvP01_02_v1	418894	A	G	6	4	113	0.05	missense	UBP1	PVP01_0210400	2122I>2122T	418894A>G	
PvP01_02_v1	419181	C	A	1	1	121	0.01	missense	UBP1	PVP01_0210400	2026R>2026S	419181C>A	
PvP01_02_v1	419231	G	A	2	2	119	0.02	missense	UBP1	PVP01_0210400	2010R>2010C	419231G>A	
PvP01_02_v1	419232	G	A	2	2	119	0.02	synonymous	UBP1	PVP01_0210400	2009G	419232G>A	

PvP01_02_v1	419267	A	T	1	2	120	0.01	missense	UBP1	PVP01_0210400	1998F>1998I	419267A>T
PvP01_02_v1	419360	G	T	110	13	0	1	missense	UBP1	PVP01_0210400	1967P>1967T	419360G>T
PvP01_02_v1	420021	G	A	17	3	103	0.14	synonymous	UBP1	PVP01_0210400	1746A	420021G>A
PvP01_02_v1	420279	A	G	111	12	0	1	synonymous	UBP1	PVP01_0210400	1660Y	420279A>G
PvP01_02_v1	420528	C	T	49	10	64	0.43	synonymous	UBP1	PVP01_0210400	1577E	420528C>T
PvP01_02_v1	420584	T	C	2	5	116	0.02	missense	UBP1	PVP01_0210400	1559K>1559E	420584T>C
PvP01_02_v1	420605	C	T	1	2	120	0.01	missense	UBP1	PVP01_0210400	1552V>1552I	420605C>T
PvP01_02_v1	420612	G	A	14	3	106	0.12	synonymous	UBP1	PVP01_0210400	1549F	420612G>A
PvP01_02_v1	420860	A	T	7	3	113	0.06	missense	UBP1	PVP01_0210400	1467F>1467I	420860A>T
PvP01_02_v1	421011	T	G	2	2	119	0.02	missense	UBP1	PVP01_0210400	1416K>1416N	421011T>G
PvP01_02_v1	421953	C	T	12	7	104	0.1	synonymous	UBP1	PVP01_0210400	1102G	421953C>T
PvP01_02_v1	422145	T	C	108	15	0	1	synonymous	UBP1	PVP01_0210400	1038K	422145T>C
PvP01_02_v1	422286	G	T	30	10	83	0.27	missense	UBP1	PVP01_0210400	991D>991E	422286G>T
PvP01_02_v1	422483	G	A	6	3	114	0.05	missense	UBP1	PVP01_0210400	926R>926C	422483G>A
PvP01_02_v1	422851	C	T	36	9	78	0.32	missense	UBP1	PVP01_0210400	803R>803Q	422851C>T
PvP01_02_v1	422937	G	T	10	9	104	0.09	synonymous	UBP1	PVP01_0210400	774G	422937G>T
PvP01_02_v1	422965	T	A	11	9	103	0.1	missense	UBP1	PVP01_0210400	765H>765L	422965T>A
PvP01_02_v1	423009	G	T	4	10	109	0.04	synonymous	UBP1	PVP01_0210400	750G	423009G>T
PvP01_02_v1	423218	C	T	71	11	41	0.63	missense	UBP1	PVP01_0210400	681V>681I	423218C>T
PvP01_02_v1	423253	A	C	1	6	116	0.01	missense	UBP1	PVP01_0210400	669L>669W	423253A>C
PvP01_02_v1	424182	T	G	29	28	66	0.31	synonymous	UBP1	PVP01_0210400	359G	424182T>G
PvP01_02_v1	424340	C	T	17	22	84	0.17	missense	UBP1	PVP01_0210400	307G>307R	424340C>T
PvP01_02_v1	424495	C	T	5	18	100	0.05	missense	UBP1	PVP01_0210400	255R>255H	424495C>T
PvP01_02_v1	424653	G	A	7	7	109	0.06	synonymous	UBP1	PVP01_0210400	202D	424653G>A
PvP01_02_v1	424684	T	C	5	11	107	0.04	missense	UBP1	PVP01_0210400	192D>192G	424684T>C
PvP01_02_v1	424690	C	T	2	9	112	0.02	missense	UBP1	PVP01_0210400	190R>190K	424690C>T
PvP01_02_v1	424896	G	T	2	8	113	0.02	missense	UBP1	PVP01_0210400	121D>121E	424896G>T
PvP01_02_v1	685481	G	A	1	8	114	0.01	synonymous	AP2-L	PVP01_0216000	3L	685481G>A
PvP01_02_v1	685484	G	C	10	10	103	0.09	synonymous	AP2-L	PVP01_0216000	4L	685484G>C
PvP01_02_v1	685679	T	C	46	15	62	0.43	synonymous	AP2-L	PVP01_0216000	69S	685679T>C
PvP01_02_v1	686098	C	T	20	4	99	0.17	missense	AP2-L	PVP01_0216000	209P>209L	686098C>T
PvP01_02_v1	686337	T	C	48	4	71	0.4	missense	AP2-L	PVP01_0216000	289S>289P	686337T>C
PvP01_02_v1	686372	C	T	1	2	120	0.01	synonymous	AP2-L	PVP01_0216000	300F	686372C>T
PvP01_02_v1	687004	T	C	3	5	115	0.03	missense	AP2-L	PVP01_0216000	511V>511A	687004T>C
PvP01_02_v1	687044	T	C	1	3	119	0.01	synonymous	AP2-L	PVP01_0216000	524D	687044T>C
PvP01_02_v1	687461	C	G	40	5	78	0.34	missense	AP2-L	PVP01_0216000	663D>663E	687461C>G
PvP01_02_v1	687902	G	A	1	1	121	0.01	synonymous	AP2-L	PVP01_0216000	810L	687902G>A
PvP01_02_v1	688071	G	A	118	5	0	1	missense	AP2-L	PVP01_0216000	867V>867M	688071G>A
PvP01_03_v1	192842	G	A	10	3	110	0.08	missense	P36	PVP01_0303700	325L>325F	192842G>A
PvP01_03_v1	192985	C	T	5	2	116	0.04	missense	P36	PVP01_0303700	277S>277N	192985C>T
PvP01_03_v1	192987	G	T	102	3	18	0.85	missense	P36	PVP01_0303700	276F>276L	192987G>T
PvP01_03_v1	193569	C	T	46	2	75	0.38	synonymous	P36	PVP01_0303700	82V	193569C>T
PvP01_03_v1	633963	G	A	4	53	66	0.06	missense	CRMP2	PVP01_0314000	2516A>2516V	633963G>A
PvP01_03_v1	634189	G	A	3	51	69	0.04	synonymous	CRMP2	PVP01_0314000	2441L	634189G>A
PvP01_03_v1	634195	C	T	3	51	69	0.04	missense	CRMP2	PVP01_0314000	2439G>2439R	634195C>T
PvP01_03_v1	634867	A	T	49	60	14	0.78	missense	CRMP2	PVP01_0314000	2215L>2215I	634867A>T
PvP01_03_v1	635876	G	A	1	43	79	0.01	synonymous	CRMP2	PVP01_0314000	1878I	635876G>A
PvP01_03_v1	638138	C	T	1	39	83	0.01	synonymous	CRMP2	PVP01_0314000	1124Q	638138C>T
PvP01_03_v1	638174	G	A	62	49	12	0.84	synonymous	CRMP2	PVP01_0314000	1112Y	638174G>A
PvP01_03_v1	639327	G	A	15	37	71	0.17	synonymous	CRMP2	PVP01_0314000	816P	639327G>A

PvP01_03_v1	639469	G	A	16	33	74	0.18	missense	CRMP2	PVP01_0314000	769T>769I	639469G>A
PvP01_03_v1	639675	G	A	5	34	84	0.06	synonymous	CRMP2	PVP01_0314000	700N	639675G>A
PvP01_03_v1	639881	T	C	12	36	75	0.14	missense	CRMP2	PVP01_0314000	632K>632E	639881T>C
PvP01_03_v1	640434	C	T	10	26	87	0.1	synonymous	CRMP2	PVP01_0314000	447E	640434C>T
PvP01_03_v1	641013	T	A	12	25	86	0.12	missense	CRMP2	PVP01_0314000	304I>304F	641013T>A
PvP01_03_v1	641166	C	A	14	21	88	0.14	missense	CRMP2	PVP01_0314000	253V>253L	641166C>A
PvP01_03_v1	641172	T	G	3	20	100	0.03	missense	CRMP2	PVP01_0314000	251K>251Q	641172T>G
PvP01_03_v1	642358	G	T	1	26	96	0.01	missense	CRMP2	PVP01_0314000	40P>40T	642358G>T
PvP01_03_v1	642368	C	A	1	26	96	0.01	synonymous	CRMP2	PVP01_0314000	36T	642368C>A
PvP01_04_v1	491775	C	T	3	3	117	0.03	synonymous	PVP01_0412400	PVP01_0412400	3529S	491775C>T
PvP01_04_v1	492414	C	T	2	1	120	0.02	synonymous	PVP01_0412400	PVP01_0412400	3316A	492414C>T
PvP01_04_v1	493042	T	C	94	5	24	0.8	missense	PVP01_0412400	PVP01_0412400	3107K>3107R	493042T>C
PvP01_04_v1	493044	A	G	118	5	0	1	synonymous	PVP01_0412400	PVP01_0412400	3106Y	493044A>G
PvP01_04_v1	493053	G	A	8	1	114	0.07	synonymous	PVP01_0412400	PVP01_0412400	3103F	493053G>A
PvP01_04_v1	493070	C	T	27	1	95	0.22	missense	PVP01_0412400	PVP01_0412400	3098V>3098I	493070C>T
PvP01_04_v1	493769	C	A	15	6	102	0.13	missense	PVP01_0412400	PVP01_0412400	2865A>2865S	493769C>A
PvP01_04_v1	493965	C	T	1	9	113	0.01	synonymous	PVP01_0412400	PVP01_0412400	2799G	493965C>T
PvP01_04_v1	494144	G	A	1	3	119	0.01	synonymous	PVP01_0412400	PVP01_0412400	2740L	494144G>A
PvP01_04_v1	494544	C	T	1	2	120	0.01	synonymous	PVP01_0412400	PVP01_0412400	2606E	494544C>T
PvP01_04_v1	495306	C	G	113	10	0	1	synonymous	PVP01_0412400	PVP01_0412400	2352T	495306C>G
PvP01_04_v1	495574	C	G	1	6	116	0.01	missense	PVP01_0412400	PVP01_0412400	2263R>2263T	495574C>G
PvP01_04_v1	495617	G	A	1	5	117	0.01	missense	PVP01_0412400	PVP01_0412400	2249R>2249W	495617G>A
PvP01_04_v1	495768	C	T	3	7	113	0.03	synonymous	PVP01_0412400	PVP01_0412400	2198L	495768C>T
PvP01_04_v1	495825	C	A	1	5	117	0.01	missense	PVP01_0412400	PVP01_0412400	2179M>2179I	495825C>A
PvP01_04_v1	495997	C	G	1	7	115	0.01	missense	PVP01_0412400	PVP01_0412400	2122S>2122T	495997C>G
PvP01_04_v1	496422	C	T	3	4	116	0.03	synonymous	PVP01_0412400	PVP01_0412400	1980A	496422C>T
PvP01_04_v1	496734	A	C	71	26	26	0.73	missense	PVP01_0412400	PVP01_0412400	1876H>1876Q	496734A>C
PvP01_04_v1	497127	C	G	11	13	99	0.1	missense	PVP01_0412400	PVP01_0412400	1745R>1745S	497127C>G
PvP01_04_v1	497175	A	C	57	10	56	0.5	synonymous	PVP01_0412400	PVP01_0412400	1729G	497175A>C
PvP01_04_v1	497453	C	T	5	2	116	0.04	missense	PVP01_0412400	PVP01_0412400	1637D>1637N	497453C>T
PvP01_04_v1	497463	G	A	22	11	90	0.2	synonymous	PVP01_0412400	PVP01_0412400	1633F	497463G>A
PvP01_04_v1	497738	G	A	8	7	108	0.07	synonymous	PVP01_0412400	PVP01_0412400	1542L	497738G>A
PvP01_04_v1	497979	G	T	18	5	100	0.15	synonymous	PVP01_0412400	PVP01_0412400	1461L	497979G>T
PvP01_04_v1	498072	A	T	9	7	107	0.08	synonymous	PVP01_0412400	PVP01_0412400	1430A	498072A>T
PvP01_04_v1	498429	C	A	1	3	119	0.01	missense	PVP01_0412400	PVP01_0412400	1311K>1311N	498429C>A
PvP01_04_v1	499391	T	G	6	6	111	0.05	missense	PVP01_0412400	PVP01_0412400	991S>991R	499391T>G
PvP01_04_v1	500091	C	G	106	17	0	1	synonymous	PVP01_0412400	PVP01_0412400	757S	500091C>G
PvP01_04_v1	500130	C	G	108	15	0	1	missense	PVP01_0412400	PVP01_0412400	744M>744I	500130C>G
PvP01_04_v1	500161	G	C	89	13	21	0.81	missense	PVP01_0412400	PVP01_0412400	734S>734C	500161G>C
PvP01_04_v1	500175	C	G	108	15	0	1	missense	PVP01_0412400	PVP01_0412400	729K>729N	500175C>G
PvP01_04_v1	500204	C	G	99	15	9	0.92	missense	PVP01_0412400	PVP01_0412400	720V>720L	500204C>G
PvP01_04_v1	500205	G	A	94	15	14	0.87	synonymous	PVP01_0412400	PVP01_0412400	719Y	500205G>A
PvP01_04_v1	500276	G	A	107	16	0	1	missense	PVP01_0412400	PVP01_0412400	696H>696Y	500276G>A
PvP01_04_v1	500313	T	C	31	30	62	0.33	synonymous	PVP01_0412400	PVP01_0412400	683S	500313T>C
PvP01_04_v1	500435	C	T	1	4	118	0.01	missense	PVP01_0412400	PVP01_0412400	643V>643I	500435C>T
PvP01_04_v1	500556	C	T	1	6	116	0.01	synonymous	PVP01_0412400	PVP01_0412400	602L	500556C>T
PvP01_04_v1	501162	T	G	54	15	54	0.5	synonymous	PVP01_0412400	PVP01_0412400	400G	501162T>G
PvP01_04_v1	636372	A	G	1	1	121	0.01	synonymous	P230	PVP01_0415800	2631D	636372A>G
PvP01_04_v1	637044	G	T	114	9	0	1	missense	P230	PVP01_0415800	2407N>2407K	637044G>T
PvP01_04_v1	637765	A	G	14	2	107	0.12	missense	P230	PVP01_0415800	2167V>2167A	637765A>G

PvP01_04_v1	638175	G	A	25	4	94	0.21	synonymous	P230	PVP01_0415800	2030D	638175G>A	
PvP01_04_v1	638194	T	C	1	3	119	0.01	missense	P230	PVP01_0415800	2024K>2024R	638194T>C	
PvP01_04_v1	638316	G	A	98	11	14	0.88	synonymous	P230	PVP01_0415800	1983Y	638316G>A	
PvP01_04_v1	638385	T	G	110	11	2	0.98	synonymous	P230	PVP01_0415800	1960A	638385T>G	
PvP01_04_v1	639060	T	A	91	15	17	0.84	synonymous	P230	PVP01_0415800	1735L	639060T>A	
PvP01_04_v1	639162	C	A	10	4	109	0.08	missense	P230	PVP01_0415800	1701K>1701N	639162C>A	
PvP01_04_v1	639371	C	T	8	7	108	0.07	missense	P230	PVP01_0415800	1632V>1632I	639371C>T	
PvP01_04_v1	639480	A	G	28	11	84	0.25	synonymous	P230	PVP01_0415800	1595D	639480A>G	
PvP01_04_v1	639585	C	A	2	6	115	0.02	missense	P230	PVP01_0415800	1560Q>1560H	639585C>A	
PvP01_04_v1	639591	G	A	2	5	116	0.02	synonymous	P230	PVP01_0415800	1558S	639591G>A	
PvP01_04_v1	639848	C	G	3	9	111	0.03	missense	P230	PVP01_0415800	1473A>1473P	639848C>G	
PvP01_04_v1	640577	G	C	23	4	96	0.19	missense	P230	PVP01_0415800	1230L>1230V	640577G>C	
PvP01_04_v1	640620	T	G	98	16	9	0.92	synonymous	P230	PVP01_0415800	1215G	640620T>G	
PvP01_04_v1	641279	C	T	2	7	114	0.02	missense	P230	PVP01_0415800	996E>996K	641279C>T	
PvP01_04_v1	641822	C	T	13	9	101	0.11	missense	P230	PVP01_0415800	815V>815I	641822C>T	
PvP01_04_v1	642106	T	G	44	14	65	0.4	missense	P230	PVP01_0415800	720D>720T	642106T>G+642107C>T	
PvP01_04_v1	642107	C	T	44	14	65	0.4	missense	P230	PVP01_0415800	720D>720T	642106T>G+642107C>T	
PvP01_04_v1	642453	G	A	27	18	78	0.26	synonymous	P230	PVP01_0415800	604Y	642453G>A	
PvP01_04_v1	643130	C	T	35	22	66	0.35	missense	P230	PVP01_0415800	379A>379T	643130C>T	
PvP01_04_v1	643362	G	A	10	10	103	0.09	synonymous	P230	PVP01_0415800	301F	643362G>A	
PvP01_04_v1	643404	G	A	105	18	0	1	synonymous	P230	PVP01_0415800	287P	643404G>A	
PvP01_04_v1	643474	T	G	105	18	0	1	missense	P230	PVP01_0415800	264D>264A	643474T>G	
PvP01_04_v1	643474	T	G	105	18	0	1	missense	P230	PVP01_0415800	264D>264T	643474T>G+643475C>T	762805A>C
PvP01_04_v1	643511	C	T	12	10	101	0.11	missense	P230	PVP01_0415800	252V>252M	643511C>T	
PvP01_04_v1	643628	A	C	105	18	0	1	missense	P230	PVP01_0415800	213F>213V	643628A>C	
PvP01_04_v1	643932	G	A	2	10	111	0.02	synonymous	P230	PVP01_0415800	111G	643932G>A	
PvP01_04_v1	643960	C	T	43	16	64	0.4	missense	P230	PVP01_0415800	102C>102H	643960C>T+643961A>G	
PvP01_04_v1	643960	C	T	43	16	64	0.4	missense	P230	PVP01_0415800	102C>102Y	643960C>T	
PvP01_04_v1	760567	T	G	43	12	68	0.39	missense	MSP4	PVP01_0418300	6C>6W	760567T>G	
PvP01_04_v1	761231	G	A	4	3	116	0.03	missense	MSP4	PVP01_0418300	176D>176N	761231G>A	
PvP01_04_v1	762533	C	G	5	2	116	0.04	missense	MSP5	PVP01_0418400	16C>16W	762533C>G	
PvP01_04_v1	762602	G	T	8	2	113	0.07	missense	MSP5	PVP01_0418400	39Q>39H	762602G>T	
PvP01_04_v1	762632	G	A	19	3	101	0.16	synonymous	MSP5	PVP01_0418400	49R	762632G>A	
PvP01_04_v1	762643	C	T	19	2	102	0.16	missense	MSP5	PVP01_0418400	53A>53V	762643C>T	762879A>G
PvP01_04_v1	762645	A	G	19	2	102	0.16	missense	MSP5	PVP01_0418400	54N>54D	762645A>G	
PvP01_04_v1	762751	G	A	8	2	113	0.07	missense	MSP5	PVP01_0418400	89S>89N	762751G>A	
PvP01_04_v1	762756	G	A	8	6	109	0.07	missense	MSP5	PVP01_0418400	91E>91K	762756G>A	
PvP01_04_v1	762762	G	A	8	3	112	0.07	missense	MSP5	PVP01_0418400	93D>93K	762762G>A+762764T>A	
PvP01_04_v1	762764	T	A	83	19	21	0.8	missense	MSP5	PVP01_0418400	93D>93E	762764T>A	
PvP01_04_v1	762764	T	A	83	19	21	0.8	missense	MSP5	PVP01_0418400	93D>93K	762762G>A+762764T>A	
PvP01_04_v1	762765	A	C	24	12	87	0.22	missense	MSP5	PVP01_0418400	94K>94Q	762765A>C	
PvP01_04_v1	762769	A	G	23	14	86	0.21	missense	MSP5	PVP01_0418400	95D>95G	762769A>G	
PvP01_04_v1	762771	C	T	60	14	49	0.55	missense	MSP5	PVP01_0418400	96P>96S	762771C>T	
PvP01_04_v1	762781	A	G	76	20	27	0.74	missense	MSP5	PVP01_0418400	99E>99G	762781A>G	
PvP01_04_v1	762787	A	C	5	8	110	0.04	missense	MSP5	PVP01_0418400	101D>101A	762787A>C	
PvP01_04_v1	762787	A	C	5	8	110	0.04	missense	MSP5	PVP01_0418400	101D>101P	762786G>C+762787A>C	
PvP01_04_v1	762789	A	C	5	9	109	0.04	missense	MSP5	PVP01_0418400	102T>102P	762789A>C	
PvP01_04_v1	762798	G	A	55	14	54	0.5	missense	MSP5	PVP01_0418400	105E>105K	762798G>A	
PvP01_04_v1	762805	A	C	59	20	44	0.57	missense	MSP5	PVP01_0418400	107K>107T	762805A>C+762806G>A	
PvP01_04_v1	762806	G	A	20	12	91	0.18	missense	MSP5	PVP01_0418400	107K>107T	762805A>C+762806G>A	

PvP01_04_v1	762817	C	A	3	9	111	0.03	missense	MSP5	PVP01_0418400	111A>111D	762817C>A
PvP01_04_v1	762819	G	A	21	11	91	0.19	missense	MSP5	PVP01_0418400	112A>112T	762819G>A
PvP01_04_v1	762844	A	G	52	17	54	0.49	missense	MSP5	PVP01_0418400	120E>120G	762844A>G
PvP01_04_v1	762844	A	G	52	17	54	0.49	missense	MSP5	PVP01_0418400	120E>120R	762843G>A+762844A>G
PvP01_04_v1	762846	C	A	7	9	107	0.06	missense	MSP5	PVP01_0418400	121P>121T	762846C>A
PvP01_04_v1	762847	C	A	8	11	104	0.07	missense	MSP5	PVP01_0418400	121P>121Q	762847C>A
PvP01_04_v1	762854	A	T	87	22	14	0.86	missense	MSP5	PVP01_0418400	123E>123D	762854A>T
PvP01_04_v1	762872	C	T	72	24	27	0.73	synonymous	MSP5	PVP01_0418400	129V	762872C>T
PvP01_04_v1	762875	G	T	99	24	0	1	missense	MSP5	PVP01_0418400	130E>130D	762875G>T
PvP01_04_v1	762876	C	T	71	25	27	0.72	missense	MSP5	PVP01_0418400	131P>131S	762876C>T
PvP01_04_v1	762879	A	G	32	11	80	0.29	missense	MSP5	PVP01_0418400	132T>132A	762879A>G+762881C>T
PvP01_04_v1	762881	C	T	72	24	27	0.73	missense	MSP5	PVP01_0418400	132T>132A	762879A>G+762881C>T
PvP01_04_v1	762881	C	T	72	24	27	0.73	synonymous	MSP5	PVP01_0418400	132T	762881C>T
PvP01_04_v1	762882	G	A	15	9	99	0.13	missense	MSP5	PVP01_0418400	133V>133T	762882G>A+762883T>C
PvP01_04_v1	762883	T	C	31	8	84	0.27	missense	MSP5	PVP01_0418400	133V>133A	762883T>C
PvP01_04_v1	762883	T	C	31	8	84	0.27	missense	MSP5	PVP01_0418400	133V>133T	762882G>A+762883T>C
PvP01_04_v1	762885	A	C	16	8	99	0.14	missense	MSP5	PVP01_0418400	134T>134P	762885A>C
PvP01_04_v1	762889	A	G	17	16	90	0.16	missense	MSP5	PVP01_0418400	135D>135G	762889A>G
PvP01_04_v1	762891	G	A	15	7	101	0.13	missense	MSP5	PVP01_0418400	136A>136K	762891G>A+762892C>A
PvP01_04_v1	762892	C	A	15	8	100	0.13	missense	MSP5	PVP01_0418400	136A>136E	762892C>A
PvP01_04_v1	762892	C	A	15	8	100	0.13	missense	MSP5	PVP01_0418400	136A>136K	762891G>A+762892C>A
PvP01_04_v1	762903	A	G	12	9	102	0.11	missense	MSP5	PVP01_0418400	140K>140E	762903A>G
PvP01_04_v1	762910	A	C	7	8	108	0.06	missense	MSP5	PVP01_0418400	142E>142A	762910A>C
PvP01_04_v1	762912	A	C	46	45	32	0.59	missense	MSP5	PVP01_0418400	143N>143H	762912A>C
PvP01_04_v1	762924	T	C	42	42	39	0.52	missense	MSP5	PVP01_0418400	147S>147P	762924T>C
PvP01_04_v1	762930	G	A	8	17	98	0.08	missense	MSP5	PVP01_0418400	149E>149K	762930G>A
PvP01_04_v1	762939	A	T	26	12	85	0.23	missense	MSP5	PVP01_0418400	152M>152L	762939A>T
PvP01_04_v1	762945	A	C	14	10	99	0.12	missense	MSP5	PVP01_0418400	154T>154P	762945A>C
PvP01_04_v1	762951	G	A	27	11	85	0.24	missense	MSP5	PVP01_0418400	156G>156S	762951G>A
PvP01_04_v1	762957	G	C	3	12	108	0.03	missense	MSP5	PVP01_0418400	158G>158Q	2957G>C+762958G>A+762959G>A
PvP01_04_v1	762957	G	C	3	12	108	0.03	missense	MSP5	PVP01_0418400	158G>158R	762957G>C
PvP01_04_v1	762958	G	A	30	16	77	0.28	missense	MSP5	PVP01_0418400	158G>158E	762958G>A+762959G>A
PvP01_04_v1	762958	G	A	30	16	77	0.28	missense	MSP5	PVP01_0418400	158G>158Q	2957G>C+762958G>A+762959G>A
PvP01_04_v1	762959	G	A	72	22	29	0.71	missense	MSP5	PVP01_0418400	158G>158E	762958G>A+762959G>A
PvP01_04_v1	762959	G	A	72	22	29	0.71	missense	MSP5	PVP01_0418400	158G>158Q	2957G>C+762958G>A+762959G>A
PvP01_04_v1	762959	G	A	72	22	29	0.71	synonymous	MSP5	PVP01_0418400	158G	762959G>A
PvP01_04_v1	762960	G	A	6	15	102	0.06	missense	MSP5	PVP01_0418400	159E>159K	762960G>A
PvP01_04_v1	762975	G	A	19	17	87	0.18	missense	MSP5	PVP01_0418400	164D>164N	762975G>A
PvP01_04_v1	762975	G	A	19	17	87	0.18	missense	MSP5	PVP01_0418400	164D>164S	762975G>A+762976A>G
PvP01_04_v1	762976	A	G	31	15	77	0.29	missense	MSP5	PVP01_0418400	164D>164G	762976A>G
PvP01_04_v1	762976	A	G	31	15	77	0.29	missense	MSP5	PVP01_0418400	164D>164S	762975G>A+762976A>G
PvP01_04_v1	762981	A	G	26	19	78	0.25	missense	MSP5	PVP01_0418400	166N>166D	762981A>G
PvP01_04_v1	762981	A	G	26	19	78	0.25	missense	MSP5	PVP01_0418400	166N>166E	762981A>G+762983T>G
PvP01_04_v1	762983	T	G	90	23	10	0.9	missense	MSP5	PVP01_0418400	166N>166E	762981A>G+762983T>G
PvP01_04_v1	762983	T	G	90	23	10	0.9	missense	MSP5	PVP01_0418400	166N>166K	762983T>G
PvP01_04_v1	762988	A	G	24	23	76	0.24	missense	MSP5	PVP01_0418400	168D>168G	762988A>G
PvP01_04_v1	763016	T	G	40	25	58	0.41	missense	MSP5	PVP01_0418400	177D>177E	763016T>G
PvP01_04_v1	763020	A	G	31	25	67	0.32	missense	MSP5	PVP01_0418400	179N>179D	763020A>G
PvP01_04_v1	763024	A	C	52	25	46	0.53	missense	MSP5	PVP01_0418400	180Q>180P	763024A>C
PvP01_04_v1	763030	G	A	9	24	90	0.09	missense	MSP5	PVP01_0418400	182G>182E	763030G>A

PvP01_04_v1	763032	G	A	7	24	92	0.07	missense	MSP5	PVP01_0418400	183D>183K	763032G>A+763034C>A	
PvP01_04_v1	763032	G	A	7	24	92	0.07	missense	MSP5	PVP01_0418400	183D>183N	763032G>A	
PvP01_04_v1	763034	C	A	42	28	53	0.44	missense	MSP5	PVP01_0418400	183D>183E	763034C>A	
PvP01_04_v1	763034	C	A	42	28	53	0.44	missense	MSP5	PVP01_0418400	183D>183K	763032G>A+763034C>A	
PvP01_04_v1	763036	G	A	42	27	54	0.44	missense	MSP5	PVP01_0418400	184G>184E	763036G>A	
PvP01_04_v1	763040	G	T	10	23	90	0.1	missense	MSP5	PVP01_0418400	185K>185N	763040G>T	
PvP01_04_v1	763042	A	G	85	28	10	0.89	missense	MSP5	PVP01_0418400	186E>186G	763042A>G	
PvP01_04_v1	763048	A	G	85	28	10	0.89	missense	MSP5	PVP01_0418400	188E>188G	763048A>G	
PvP01_04_v1	763051	G	A	21	22	80	0.21	missense	MSP5	PVP01_0418400	189G>189D	763051G>A	
PvP01_04_v1	763054	G	A	53	29	41	0.56	missense	MSP5	PVP01_0418400	190G>190E	763054G>A	
PvP01_04_v1	763071	G	A	31	19	73	0.3	missense	MSP5	PVP01_0418400	196D>196N	763071G>A+763073T>C	
PvP01_04_v1	763073	T	C	31	19	73	0.3	missense	MSP5	PVP01_0418400	196D>196N	763071G>A+763073T>C	
PvP01_04_v1	763074	G	A	31	19	73	0.3	missense	MSP5	PVP01_0418400	197G>197S	763074G>A	
PvP01_04_v1	763077	A	G	31	19	73	0.3	missense	MSP5	PVP01_0418400	198K>198E	763077A>G	
PvP01_04_v1	763084	A	G	8	15	100	0.07	missense	MSP5	PVP01_0418400	200D>200G	763084A>G	
PvP01_04_v1	763090	G	A	58	20	45	0.56	missense	MSP5	PVP01_0418400	202G>202E	763090G>A	
PvP01_04_v1	763111	A	C	32	17	74	0.3	missense	MSP5	PVP01_0418400	209D>209A	763111A>C+763112T>G	1368918C>T
PvP01_04_v1	763112	T	G	83	23	17	0.83	missense	MSP5	PVP01_0418400	209D>209A	763111A>C+763112T>G	
PvP01_04_v1	763112	T	G	83	23	17	0.83	missense	MSP5	PVP01_0418400	209D>209E	763112T>G	
PvP01_04_v1	763137	G	A	35	8	80	0.3	missense	MSP5	PVP01_0418400	218A>218T	763137G>A	
PvP01_04_v1	763152	G	A	10	4	109	0.08	missense	MSP5	PVP01_0418400	223G>223R	763152G>A	
PvP01_04_v1	763159	C	A	29	7	87	0.25	missense	MSP5	PVP01_0418400	225A>225D	763159C>A	
PvP01_04_v1	763164	G	A	3	5	115	0.03	missense	MSP5	PVP01_0418400	227E>227K	763164G>A	
PvP01_04_v1	763203	G	A	96	6	21	0.82	missense	MSP5	PVP01_0418400	240E>240K	763203G>A	
PvP01_04_v1	763224	G	C	23	4	96	0.19	missense	MSP5	PVP01_0418400	247D>247H	763224G>C	
PvP01_04_v1	763252	A	G	2	7	114	0.02	missense	MSP5	PVP01_0418400	256D>256G	763252A>G	
PvP01_04_v1	763289	A	T	112	8	3	0.97	missense	MSP5	PVP01_0418400	268P>268S	763287C>T+763289A>T	
PvP01_04_v1	763289	A	T	112	8	3	0.97	synonymous	MSP5	PVP01_0418400	268P	763289A>T	
PvP01_04_v1	763291	A	G	20	4	99	0.17	missense	MSP5	PVP01_0418400	269D>269G	763291A>G	
PvP01_04_v1	763297	C	A	20	4	99	0.17	missense	MSP5	PVP01_0418400	271P>271Q	763297C>A	
PvP01_04_v1	763305	G	A	51	6	66	0.44	missense	MSP5	PVP01_0418400	274E>274K	763305G>A	
PvP01_04_v1	763311	C	A	4	2	117	0.03	missense	MSP5	PVP01_0418400	276P>276T	763311C>A	
PvP01_04_v1	763930	G	T	2	1	120	0.02	missense	MSP5	PVP01_0418400	359G>359W	763930G>T	
PvP01_04_v1	763952	A	G	2	1	120	0.02	missense	MSP5	PVP01_0418400	366K>366R	763952A>G	
PvP01_04_v1	763970	C	T	10	0	113	0.08	missense	MSP5	PVP01_0418400	372S>372L	763970C>T	
PvP01_04_v1	763977	C	T	4	0	119	0.03	synonymous	MSP5	PVP01_0418400	374F	763977C>T	
PvP01_04_v1	763988	T	C	3	0	120	0.02	missense	MSP5	PVP01_0418400	378L>378S	763988T>C	
PvP01_05_v1	1077406	A	G	9	9	105	0.08	synonymous	DHFR-TS	PVP01_0526600	15A	1077406A>G	
PvP01_05_v1	1077418	C	T	4	6	113	0.03	synonymous	DHFR-TS	PVP01_0526600	19V	1077418C>T	
PvP01_05_v1	1077475	T	C	1	3	119	0.01	synonymous	DHFR-TS	PVP01_0526600	38G	1077475T>C	
PvP01_05_v1	1077534	G	A	1	4	118	0.01	missense	DHFR-TS	PVP01_0526600	58R>58K	1077534G>A	
PvP01_05_v1	1077568	T	C	90	21	12	0.88	synonymous	DHFR-TS	PVP01_0526600	69Y	1077568T>C	
PvP01_05_v1	1077707	A	G	2	10	111	0.02	missense	DHFR-TS	PVP01_0526600	116S>116G	1077707A>G	
PvP01_05_v1	1368620	G	C	13	3	107	0.11	missense	CyRPA	PVP01_0532400	359T>359R	1368620G>C	
PvP01_05_v1	1368627	A	G	11	2	110	0.09	missense	CyRPA	PVP01_0532400	357Y>357H	1368627A>G	
PvP01_05_v1	1368628	A	C	11	2	110	0.09	synonymous	CyRPA	PVP01_0532400	356L	1368628A>C	
PvP01_05_v1	1368693	G	A	10	3	110	0.08	synonymous	CyRPA	PVP01_0532400	335L	1368693G>A	
PvP01_05_v1	1368732	C	A	10	3	110	0.08	missense	CyRPA	PVP01_0532400	322D>322Y	1368732C>A	
PvP01_05_v1	1368805	C	A	10	1	112	0.08	missense	CyRPA	PVP01_0532400	297K>297N	1368805C>A	
PvP01_05_v1	1368807	T	C	2	1	120	0.02	missense	CyRPA	PVP01_0532400	297K>297E	1368807T>C	

PvP01_05_v1	1368849	C	T	14	0	109	0.11	missense	CyRPA	PVP01_0532400	283V>283I	1368849C>T
PvP01_05_v1	1368885	C	T	5	0	118	0.04	missense	CyRPA	PVP01_0532400	271V>271I	1368885C>T
PvP01_05_v1	1368916	A	G	50	7	66	0.43	missense	CyRPA	PVP01_0532400	260G>260S	1368916A>G+1368918C>T
PvP01_05_v1	1368918	C	T	50	7	66	0.43	missense	CyRPA	PVP01_0532400	260G>260S	1368916A>G+1368918C>T
PvP01_05_v1	1368927	G	T	61	8	54	0.53	missense	CyRPA	PVP01_0532400	257P>257T	1368927G>T
PvP01_05_v1	1368931	T	G	61	8	54	0.53	missense	CyRPA	PVP01_0532400	255E>255A	1368931T>G+1368932T>G
PvP01_05_v1	1368931	T	G	61	8	54	0.53	missense	CyRPA	PVP01_0532400	255E>255D	1368931T>G
PvP01_05_v1	1368934	T	A	61	9	53	0.54	synonymous	CyRPA	PVP01_0532400	254L	1368934T>A
PvP01_05_v1	1368965	G	C	1	7	115	0.01	missense	CyRPA	PVP01_0532400	244A>244G	1368965G>C
PvP01_05_v1	1368968	G	C	9	8	106	0.08	missense	CyRPA	PVP01_0532400	243T>243R	1368968G>C
PvP01_05_v1	1368969	T	C	1	8	114	0.01	missense	CyRPA	PVP01_0532400	243T>243A	1368969T>C
PvP01_05_v1	1368973	G	T	2	8	113	0.02	missense	CyRPA	PVP01_0532400	241N>241K	1368973G>T
PvP01_05_v1	1369014	C	T	45	5	73	0.38	missense	CyRPA	PVP01_0532400	228D>228N	1369014C>T
PvP01_05_v1	1369050	C	T	31	3	89	0.26	missense	CyRPA	PVP01_0532400	216V>216I	1369050C>T
PvP01_05_v1	1369384	G	A	17	9	97	0.15	missense	CyRPA	PVP01_0532400	183P>183L	1369384G>A
PvP01_05_v1	1369391	C	T	80	16	27	0.75	missense	CyRPA	PVP01_0532400	181E>181K	1369391C>T
PvP01_05_v1	1369436	T	C	25	9	89	0.22	missense	CyRPA	PVP01_0532400	166N>166D	1369436T>C
PvP01_05_v1	1369483	G	T	64	13	46	0.58	missense	CyRPA	PVP01_0532400	150A>150D	1369483G>T
PvP01_05_v1	1369495	T	C	10	12	101	0.09	missense	CyRPA	PVP01_0532400	146K>146R	1369495T>C
PvP01_05_v1	1369498	T	C	79	18	26	0.75	missense	CyRPA	PVP01_0532400	145D>145G	1369498T>C
PvP01_05_v1	1369505	G	T	2	9	112	0.02	missense	CyRPA	PVP01_0532400	143Q>143K	1369505G>T
PvP01_05_v1	1369509	G	A	1	11	111	0.01	synonymous	CyRPA	PVP01_0532400	141D	1369509G>A
PvP01_05_v1	1369510	T	C	37	11	75	0.33	missense	CyRPA	PVP01_0532400	141D>141G	1369510T>C
PvP01_05_v1	1369510	T	C	37	11	75	0.33	missense	CyRPA	PVP01_0532400	141D>141S	1369510T>C+1369511C>T
PvP01_05_v1	1369511	C	T	23	15	85	0.21	missense	CyRPA	PVP01_0532400	141D>141N	1369511C>T
PvP01_05_v1	1369511	C	T	23	15	85	0.21	missense	CyRPA	PVP01_0532400	141D>141S	1369510T>C+1369511C>T
PvP01_05_v1	1369513	T	G	1	11	111	0.01	missense	CyRPA	PVP01_0532400	140E>140T	1369513T>G+1369514C>T
PvP01_05_v1	1369514	C	T	1	11	111	0.01	missense	CyRPA	PVP01_0532400	140E>140T	1369513T>G+1369514C>T
PvP01_05_v1	1369517	T	C	2	11	110	0.02	missense	CyRPA	PVP01_0532400	139K>139E	1369517T>C
PvP01_05_v1	1369520	T	G	36	12	75	0.32	missense	CyRPA	PVP01_0532400	138K>138Q	1369520T>G
PvP01_05_v1	1369548	C	G	38	9	76	0.33	synonymous	CyRPA	PVP01_0532400	128R	1369548C>G
PvP01_05_v1	1369552	T	A	2	9	112	0.02	missense	CyRPA	PVP01_0532400	127E>127M	1369552T>A+1369553C>T
PvP01_05_v1	1369553	C	T	30	12	81	0.27	missense	CyRPA	PVP01_0532400	127E>127K	1369553C>T
PvP01_05_v1	1369553	C	T	30	12	81	0.27	missense	CyRPA	PVP01_0532400	127E>127M	1369552T>A+1369553C>T
PvP01_05_v1	1369558	A	C	39	9	75	0.34	missense	CyRPA	PVP01_0532400	125I>125S	1369558A>C
PvP01_05_v1	1369564	T	C	72	12	39	0.65	missense	CyRPA	PVP01_0532400	123E>123G	1369564T>C
PvP01_05_v1	1369568	T	C	72	12	39	0.65	missense	CyRPA	PVP01_0532400	122K>122E	1369568T>C
PvP01_05_v1	1369579	T	C	10	8	105	0.09	missense	CyRPA	PVP01_0532400	118K>118R	1369579T>C
PvP01_05_v1	1369581	T	A	2	6	115	0.02	missense	CyRPA	PVP01_0532400	117Q>117D	1369581T>A+1369583G>C
PvP01_05_v1	1369583	G	C	10	7	106	0.09	missense	CyRPA	PVP01_0532400	117Q>117D	1369581T>A+1369583G>C
PvP01_05_v1	1369583	G	C	10	7	106	0.09	missense	CyRPA	PVP01_0532400	117Q>117E	1369583G>C
PvP01_05_v1	1369657	A	C	11	4	108	0.09	missense	CyRPA	PVP01_0532400	92M>92R	1369657A>C
PvP01_05_v1	1369667	C	T	78	8	37	0.68	missense	CyRPA	PVP01_0532400	89E>89K	1369667C>T
PvP01_05_v1	1369675	T	G	62	7	54	0.53	missense	CyRPA	PVP01_0532400	86K>86T	1369675T>G
PvP01_05_v1	1369688	G	C	108	9	6	0.95	missense	CyRPA	PVP01_0532400	82Q>82E	1369688G>C
PvP01_05_v1	1369738	G	C	89	4	30	0.75	missense	CyRPA	PVP01_0532400	65A>65G	1369738G>C
PvP01_05_v1	1369897	A	G	5	16	102	0.05	missense	CyRPA	PVP01_0532400	12F>12S	1369897A>G
PvP01_06_v1	350830	A	T	26	12	85	0.23	synonymous	PVP01_0607800	PVP01_0607800	76T	350830A>T
PvP01_06_v1	351124	C	T	11	7	105	0.09	synonymous	PVP01_0607800	PVP01_0607800	174A	351124C>T
PvP01_06_v1	351403	C	G	2	7	114	0.02	missense	PVP01_0607800	PVP01_0607800	267N>267K	351403C>G

PvP01_06_v1	351597	A	G	2	5	116	0.02	missense	PVP01_0607800	PVP01_0607800	332N>332S	351597A>G	
PvP01_06_v1	352661	G	T	2	3	118	0.02	missense	PVP01_0607800	PVP01_0607800	687V>687L	352661G>T	
PvP01_06_v1	574588	G	T	6	2	115	0.05	synonymous	EBP2	PVP01_0613400	151R	574588G>T	
PvP01_06_v1	574785	T	A	1	3	119	0.01	missense	EBP2	PVP01_0613400	85E>85V	574785T>A	
PvP01_06_v1	574994	C	T	6	11	106	0.05	synonymous	EBP2	PVP01_0613400	15Q	574994C>T	
PvP01_06_v1	588169	A	G	2	5	116	0.02	missense	MTRAP	PVP01_0613800	22S>22G	588169A>G	
PvP01_06_v1	588451	G	A	16	4	103	0.13	missense	MTRAP	PVP01_0613800	116D>116N	588451G>A	
PvP01_06_v1	588969	A	C	116	7	0	1	*missense	MTRAP	PVP01_0613800	288R>288S	588969A>C	
PvP01_06_v1	589206	C	A	115	8	0	1	*missense	MTRAP	PVP01_0613800	367D>367E	589206C>A	
PvP01_06_v1	589240	T	G	2	2	119	0.02	missense	MTRAP	PVP01_0613800	379Y>379D	589240T>G	984780T>C+984781A>C
PvP01_06_v1	589259	G	C	119	4	0	1	*missense	MTRAP	PVP01_0613800	385G>385A	589259G>C	
PvP01_06_v1	670568	G	C	1	0	122	0.01	missense	PVP01_0616100	PVP01_0616100	165T>165R	670568G>C	
PvP01_06_v1	670573	A	G	1	1	121	0.01	synonymous	PVP01_0616100	PVP01_0616100	163C	670573A>G	
PvP01_06_v1	670671	T	G	119	4	0	1	missense	PVP01_0616100	PVP01_0616100	131K>131Q	670671T>G	
PvP01_06_v1	670673	G	A	107	4	12	0.9	missense	PVP01_0616100	PVP01_0616100	130T>130I	670673G>A	
PvP01_06_v1	670680	A	T	1	1	121	0.01	missense	PVP01_0616100	PVP01_0616100	128S>128T	670680A>T	
PvP01_06_v1	670727	T	A	1	1	121	0.01	missense	PVP01_0616100	PVP01_0616100	112Q>112L	670727T>A	
PvP01_06_v1	670781	C	T	12	0	111	0.1	missense	PVP01_0616100	PVP01_0616100	94G>94D	670781C>T	
PvP01_06_v1	670803	G	T	46	1	76	0.38	missense	PVP01_0616100	PVP01_0616100	87Q>87K	670803G>T	
PvP01_06_v1	982074	C	T	51	52	20	0.72	missense	DBP	PVP01_0623800	17S>17L	982074C>T	
PvP01_06_v1	982596	C	A	3	30	90	0.03	missense	DBP	PVP01_0623800	146A>146E	982596C>A	
PvP01_06_v1	982640	C	G	13	30	80	0.14	missense	DBP	PVP01_0623800	161Q>161E	982640C>G	
PvP01_06_v1	982675	T	C	1	30	92	0.01	synonymous	DBP	PVP01_0623800	172F	982675T>C	
PvP01_06_v1	982683	G	A	14	36	73	0.16	missense	DBP	PVP01_0623800	175G>175E	982683G>A	
PvP01_06_v1	982803	A	C	2	35	86	0.02	missense	DBP	PVP01_0623800	215K>215T	982803A>C	
PvP01_06_v1	982948	T	G	57	39	27	0.68	missense	DBP	PVP01_0623800	263S>263R	982948T>G	
PvP01_06_v1	983124	T	C	2	37	84	0.02	missense	DBP	PVP01_0623800	322I>322T	983124T>C	
PvP01_06_v1	983125	C	T	1	34	88	0.01	synonymous	DBP	PVP01_0623800	322I	983125C>T	
PvP01_06_v1	983158	C	T	25	38	60	0.29	synonymous	DBP	PVP01_0623800	333R	983158C>T	
PvP01_06_v1	983175	G	A	9	38	76	0.11	missense	DBP	PVP01_0623800	339G>339D	983175G>A	
PvP01_06_v1	983177	G	A	28	43	52	0.35	missense	DBP	PVP01_0623800	340E>340K	983177G>A	
PvP01_06_v1	983177	G	A	28	43	52	0.35	missense	DBP	PVP01_0623800	340E>340T	983177G>A+983178A>C	
PvP01_06_v1	983178	A	C	2	38	83	0.02	missense	DBP	PVP01_0623800	340E>340T	983177G>A+983178A>C	
PvP01_06_v1	983180	A	C	14	38	71	0.16	missense	DBP	PVP01_0623800	341K>341H	983180A>C+983182G>T	
PvP01_06_v1	983180	A	C	14	38	71	0.16	missense	DBP	PVP01_0623800	341K>341Q	983180A>C	
PvP01_06_v1	983182	G	T	28	43	52	0.35	missense	DBP	PVP01_0623800	341K>341H	983180A>C+983182G>T	
PvP01_06_v1	983182	G	T	28	43	52	0.35	missense	DBP	PVP01_0623800	341K>341N	983182G>T	
PvP01_06_v1	983193	A	G	42	42	39	0.52	missense	DBP	PVP01_0623800	345H>345R	983193A>G	
PvP01_06_v1	983216	T	A	18	39	66	0.21	missense	DBP	PVP01_0623800	353S>353T	983216T>A	
PvP01_06_v1	983235	C	G	20	37	66	0.23	missense	DBP	PVP01_0623800	359T>359R	983235C>G	
PvP01_06_v1	983275	T	A	33	45	45	0.42	missense	DBP	PVP01_0623800	372N>372K	983275T>A	
PvP01_06_v1	983281	A	G	14	41	68	0.17	missense	DBP	PVP01_0623800	374I>374M	983281A>G	
PvP01_06_v1	983294	A	T	22	43	58	0.28	missense	DBP	PVP01_0623800	379I>379L	983294A>T	
PvP01_06_v1	983333	T	C	43	42	38	0.53	missense	DBP	PVP01_0623800	392W>392R	983333T>C	
PvP01_06_v1	983363	A	T	69	54	0	1	missense	DBP	PVP01_0623800	402K>402S	983363A>T+983364A>C	
PvP01_06_v1	983364	A	C	69	54	0	1	missense	DBP	PVP01_0623800	402K>402S	983363A>T+983364A>C	
PvP01_06_v1	983416	C	A	15	45	63	0.19	synonymous	DBP	PVP01_0623800	419I	983416C>A	
PvP01_06_v1	983480	C	G	1	44	78	0.01	missense	DBP	PVP01_0623800	441Q>441E	983480C>G	
PvP01_06_v1	983532	T	A	29	46	48	0.38	missense	DBP	PVP01_0623800	458I>458K	983532T>A	
PvP01_06_v1	983588	A	T	1	34	88	0.01	missense	DBP	PVP01_0623800	477I>477L	983588A>T	

PvP01_06_v1	983609	G	A	3	40	80	0.04	missense	DBP	PVP01_0623800	484E>484K	983609G>A	
PvP01_06_v1	983617	C	T	4	43	76	0.05	synonymous	DBP	PVP01_0623800	486N	983617C>T	
PvP01_06_v1	983621	G	A	13	39	71	0.15	missense	DBP	PVP01_0623800	488V>488M	983621G>A	
PvP01_06_v1	983710	G	C	12	40	71	0.14	missense	DBP	PVP01_0623800	517Q>517H	983710G>C	
PvP01_06_v1	983885	C	A	34	36	53	0.39	missense	DBP	PVP01_0623800	576Q>576K	983885C>A	
PvP01_06_v1	983984	G	A	16	28	79	0.17	missense	DBP	PVP01_0623800	609E>609K	983984G>A	
PvP01_06_v1	984062	C	T	14	32	77	0.15	missense	DBP	PVP01_0623800	635P>635S	984062C>T	
PvP01_06_v1	984080	G	A	2	32	89	0.02	missense	DBP	PVP01_0623800	641E>641K	984080G>A	
PvP01_06_v1	984140	G	A	16	38	69	0.19	missense	DBP	PVP01_0623800	661E>661K	984140G>A	
PvP01_06_v1	984145	C	T	5	36	82	0.06	synonymous	DBP	PVP01_0623800	662N	984145C>T	
PvP01_06_v1	984173	G	A	59	43	21	0.74	missense	DBP	PVP01_0623800	672G>672R	984173G>A	
PvP01_06_v1	984357	C	G	1	37	85	0.01	missense	DBP	PVP01_0623800	733T>733R	984357C>G	
PvP01_06_v1	984438	A	G	1	37	85	0.01	missense	DBP	PVP01_0623800	760D>760G	984438A>G	495381T>C+495382T>A
PvP01_06_v1	984780	T	C	58	44	21	0.73	missense	DBP	PVP01_0623800	874V>874A	984780T>C	
PvP01_06_v1	984917	A	G	1	37	85	0.01	missense	DBP	PVP01_0623800	920I>920V	984917A>G	
PvP01_06_v1	984959	A	G	24	42	57	0.3	missense	DBP	PVP01_0623800	934N>934D	984959A>G	
PvP01_06_v1	984984	G	A	14	36	73	0.16	missense	DBP	PVP01_0623800	942R>942H	984984G>A	
PvP01_06_v1	985137	A	T	69	48	6	0.92	missense	DBP	PVP01_0623800	993H>993L	985137A>T	
PvP01_06_v1	985195	A	C	65	54	4	0.94	synonymous	DBP	PVP01_0623800	1012G	985195A>C	
PvP01_06_v1	985665	T	C	66	45	12	0.85	synonymous	DBP	PVP01_0623800	1061P	985665T>C	
PvP01_07_v1	61407	T	A	40	15	68	0.37	missense	RBP1b	PVP01_0701100	2555K>2555N	61407T>A	
PvP01_07_v1	61657	T	C	84	24	15	0.85	*missense	RBP1b	PVP01_0701100	2472K>2472R	61657T>C	
PvP01_07_v1	62184	G	T	12	12	99	0.11	synonymous	RBP1b	PVP01_0701100	2296G	62184G>T	
PvP01_07_v1	62195	A	T	12	12	99	0.11	missense	RBP1b	PVP01_0701100	2293L>2293M	62195A>T	
PvP01_07_v1	62200	G	A	12	13	98	0.11	missense	RBP1b	PVP01_0701100	2291A>2291V	62200G>A	
PvP01_07_v1	62484	A	C	91	24	8	0.92	*missense	RBP1b	PVP01_0701100	2196D>2196E	62484A>C	
PvP01_07_v1	62551	T	A	3	15	105	0.03	missense	RBP1b	PVP01_0701100	2174K>2174I	62551T>A	
PvP01_07_v1	63099	G	T	59	18	46	0.56	missense	RBP1b	PVP01_0701100	1991N>1991K	63099G>T	
PvP01_07_v1	63301	C	T	1	12	110	0.01	missense	RBP1b	PVP01_0701100	1924G>1924E	63301C>T	
PvP01_07_v1	63829	G	A	1	15	107	0.01	missense	RBP1b	PVP01_0701100	1748T>1748I	63829G>A	
PvP01_07_v1	64512	C	T	84	24	15	0.85	synonymous	RBP1b	PVP01_0701100	1520G	64512C>T	
PvP01_07_v1	64531	A	C	5	16	102	0.05	missense	RBP1b	PVP01_0701100	1514F>1514C	64531A>C	
PvP01_07_v1	64846	T	A	15	16	92	0.14	missense	RBP1b	PVP01_0701100	1409K>1409I	64846T>A	
PvP01_07_v1	65272	A	G	2	18	103	0.02	missense	RBP1b	PVP01_0701100	1267I>1267T	65272A>G	
PvP01_07_v1	65285	G	T	11	15	97	0.1	missense	RBP1b	PVP01_0701100	1263Q>1263K	65285G>T	
PvP01_07_v1	65850	G	A	12	12	99	0.11	synonymous	RBP1b	PVP01_0701100	1074F	65850G>A	
PvP01_07_v1	66173	C	T	41	17	65	0.39	missense	RBP1b	PVP01_0701100	967G>967S	66173C>T	
PvP01_07_v1	66789	G	A	13	17	93	0.12	synonymous	RBP1b	PVP01_0701100	761N	66789G>A	
PvP01_07_v1	67518	C	T	14	18	91	0.13	synonymous	RBP1b	PVP01_0701100	518R	67518C>T	
PvP01_07_v1	67586	A	G	38	32	53	0.42	missense	RBP1b	PVP01_0701100	496S>496P	67586A>G	
PvP01_07_v1	67869	A	G	18	14	91	0.17	synonymous	RBP1b	PVP01_0701100	401I	67869A>G	
PvP01_07_v1	68138	T	C	3	16	104	0.03	missense	RBP1b	PVP01_0701100	312K>312E	68138T>C	
PvP01_07_v1	68602	T	C	9	22	92	0.09	missense	RBP1b	PVP01_0701100	157D>157G	68602T>C	
PvP01_07_v1	68620	C	T	92	31	0	1	missense	RBP1b	PVP01_0701100	151C>151Y	68620C>T	
PvP01_07_v1	68805	T	G	93	30	0	1	synonymous	RBP1b	PVP01_0701100	89I	68805T>G	
PvP01_07_v1	71467	T	C	95	27	1	0.99	*missense	RBP1a	PVP01_0701200	2829I>2829V	71467T>C	
PvP01_07_v1	71500	C	T	1	20	102	0.01	*missense	RBP1a	PVP01_0701200	2818E>2818K	71500C>T	
PvP01_07_v1	71506	T	G	38	22	63	0.38	*missense	RBP1a	PVP01_0701200	2816N>2816H	71506T>G	
PvP01_07_v1	71532	C	T	40	18	65	0.38	*missense	RBP1a	PVP01_0701200	2807R>2807K	71532C>T	
PvP01_07_v1	71710	G	A	101	22	0	1	*missense	RBP1a	PVP01_0701200	2748H>2748Y	71710G>A	

PvP01_07_v1	72866	A	G	98	25	0	1	*synonymous	RBP1a	PVP01_0701200	2362D	72866A>G	
PvP01_07_v1	72973	C	T	2	16	105	0.02	missense	RBP1a	PVP01_0701200	2327A>2327T	72973C>T	
PvP01_07_v1	72991	T	G	90	30	3	0.97	*missense	RBP1a	PVP01_0701200	2321N>2321H	72991T>G	1216095G>A
PvP01_07_v1	73192	C	T	43	19	61	0.41	*missense	RBP1a	PVP01_0701200	2254V>2254M	73192C>T	
PvP01_07_v1	73790	G	A	32	21	70	0.31	*synonymous	RBP1a	PVP01_0701200	2054Y	73790G>A	
PvP01_07_v1	74247	T	C	35	15	73	0.32	*missense	RBP1a	PVP01_0701200	1902K>1902R	74247T>C	
PvP01_07_v1	74280	T	A	4	17	102	0.04	*missense	RBP1a	PVP01_0701200	1891K>1891I	74280T>A	1216263A>G
PvP01_07_v1	74661	G	C	37	24	62	0.37	*missense	RBP1a	PVP01_0701200	1764A>1764G	74661G>C	
PvP01_07_v1	74758	T	C	40	24	59	0.4	*missense	RBP1a	PVP01_0701200	1732R>1732G	74758T>C	
PvP01_07_v1	74798	C	T	46	26	51	0.47	*synonymous	RBP1a	PVP01_0701200	1718K	74798C>T	
PvP01_07_v1	75023	A	C	80	43	0	1	*missense	RBP1a	PVP01_0701200	1643N>1643K	75023A>C	
PvP01_07_v1	75023	A	C	80	43	0	1	missense	RBP1a	PVP01_0701200	1643N>1643E	75023A>C+75025T>C	1216296G>C
PvP01_07_v1	75408	T	G	92	31	0	1	*missense	RBP1a	PVP01_0701200	1515E>1515A	75408T>G	
PvP01_07_v1	75408	T	G	92	31	0	1	missense	RBP1a	PVP01_0701200	1515E>1515A	75407T>C+75408T>G	
PvP01_07_v1	75705	T	A	1	18	104	0.01	missense	RBP1a	PVP01_0701200	1416K>1416M	75705T>A	
PvP01_07_v1	75725	A	T	11	17	95	0.1	*missense	RBP1a	PVP01_0701200	1409F>1409L	75725A>T	
PvP01_07_v1	76128	T	A	39	26	58	0.4	*missense	RBP1a	PVP01_0701200	1275K>1275I	76128T>A	
PvP01_07_v1	76826	T	C	103	20	0	1	*synonymous	RBP1a	PVP01_0701200	1042K	76826T>C	
PvP01_07_v1	76852	G	A	1	12	110	0.01	synonymous	RBP1a	PVP01_0701200	1034L	76852G>A	
PvP01_07_v1	77734	C	T	1	12	110	0.01	missense	RBP1a	PVP01_0701200	740D>740N	77734C>T	
PvP01_07_v1	77815	C	T	18	17	88	0.17	*missense	RBP1a	PVP01_0701200	713E>713K	77815C>T	
PvP01_07_v1	78101	G	T	53	28	42	0.56	*missense	RBP1a	PVP01_0701200	617S>617R	78101G>T	
PvP01_07_v1	78144	A	G	25	39	59	0.3	*missense	RBP1a	PVP01_0701200	603L>603S	78144A>G	
PvP01_07_v1	78148	T	C	78	45	0	1	*missense	RBP1a	PVP01_0701200	602N>602D	78148T>C	
PvP01_07_v1	78157	C	A	48	41	34	0.59	missense	RBP1a	PVP01_0701200	599E>599Y	78155T>A+78157C>A	
PvP01_07_v1	78157	C	A	48	41	34	0.59	stop_gained	RBP1a	PVP01_0701200	599E>599*	78157C>A	
PvP01_07_v1	78223	G	T	72	51	0	1	synonymous	RBP1a	PVP01_0701200	577R	78223G>T	
PvP01_07_v1	78239	A	T	1	38	84	0.01	missense	RBP1a	PVP01_0701200	571F>571L	78239A>T	
PvP01_07_v1	78252	G	A	1	37	85	0.01	missense	RBP1a	PVP01_0701200	567S>567L	78252G>A	
PvP01_07_v1	78259	C	T	1	35	87	0.01	missense	RBP1a	PVP01_0701200	565D>565N	78259C>T	
PvP01_07_v1	78274	T	C	75	48	0	1	missense	RBP1a	PVP01_0701200	560R>560E	78273C>T+78274T>C	
PvP01_07_v1	78274	T	C	75	48	0	1	missense	RBP1a	PVP01_0701200	560R>560G	78274T>C	
PvP01_07_v1	78300	A	G	80	43	0	1	missense	RBP1a	PVP01_0701200	551M>551T	78300A>G	
PvP01_07_v1	78365	T	A	2	22	99	0.02	missense	RBP1a	PVP01_0701200	529E>529D	78365T>A	
PvP01_07_v1	78378	T	C	7	21	95	0.07	missense	RBP1a	PVP01_0701200	525E>525G	78378T>C	
PvP01_07_v1	78380	T	A	2	21	100	0.02	missense	RBP1a	PVP01_0701200	524K>524N	78380T>A	
PvP01_07_v1	78387	G	T	92	31	0	1	missense	RBP1a	PVP01_0701200	522P>522Q	78387G>T	
PvP01_07_v1	78400	C	A	10	21	92	0.1	missense	RBP1a	PVP01_0701200	518D>518Y	78400C>A	
PvP01_07_v1	78423	G	T	96	25	2	0.98	missense	RBP1a	PVP01_0701200	510A>510D	78423G>T	
PvP01_07_v1	78742	C	T	17	21	85	0.17	missense	RBP1a	PVP01_0701200	404E>404K	78742C>T	
PvP01_07_v1	78783	A	G	89	34	0	1	missense	RBP1a	PVP01_0701200	390V>390A	78783A>G	
PvP01_07_v1	78870	T	G	87	34	2	0.98	missense	RBP1a	PVP01_0701200	361N>361T	78870T>G	
PvP01_07_v1	78958	T	C	83	36	4	0.95	missense	RBP1a	PVP01_0701200	332M>332V	78958T>C	
PvP01_07_v1	79090	A	T	89	33	1	0.99	missense	RBP1a	PVP01_0701200	288L>288I	79090A>T	
PvP01_07_v1	79137	A	T	82	37	4	0.95	missense	RBP1a	PVP01_0701200	272I>272K	79137A>T	
PvP01_07_v1	79312	T	G	92	30	1	0.99	missense	RBP1a	PVP01_0701200	214K>214Q	79312T>G	
PvP01_07_v1	79327	C	T	39	21	63	0.38	missense	RBP1a	PVP01_0701200	209E>209K	79327C>T	
PvP01_07_v1	79348	G	T	19	23	81	0.19	missense	RBP1a	PVP01_0701200	202Q>202K	79348G>T	
PvP01_07_v1	79354	C	T	14	22	87	0.14	missense	RBP1a	PVP01_0701200	200D>200N	79354C>T	
PvP01_07_v1	79453	C	T	96	27	0	1	missense	RBP1a	PVP01_0701200	167E>167K	79453C>T	

PvP01_07_v1	79482	T	A	19	16	88	0.18	missense	RBP1a	PVP01_0701200	157K>157M	79482T>A	
PvP01_07_v1	79546	G	C	17	22	84	0.17	missense	RBP1a	PVP01_0701200	136Q>136E	79546G>C	
PvP01_07_v1	79582	C	T	19	25	79	0.19	missense	RBP1a	PVP01_0701200	124E>124K	79582C>T	
PvP01_07_v1	495381	T	C	36	32	55	0.4	missense	CRMP1	PVP01_0709800	12I>12T	495381T>C	
PvP01_07_v1	495637	G	A	14	12	97	0.13	synonymous	CRMP1	PVP01_0709800	97R	495637G>A	
PvP01_07_v1	495827	A	G	49	14	60	0.45	missense	CRMP1	PVP01_0709800	161K>161E	495827A>G	
PvP01_07_v1	495859	C	T	1	15	107	0.01	synonymous	CRMP1	PVP01_0709800	171D	495859C>T	
PvP01_07_v1	496172	C	G	60	22	41	0.59	missense	CRMP1	PVP01_0709800	276Q>276E	496172C>G	
PvP01_07_v1	496521	C	T	45	13	65	0.41	missense	CRMP1	PVP01_0709800	392S>392F	496521C>T	
PvP01_07_v1	496547	G	C	1	11	111	0.01	missense	CRMP1	PVP01_0709800	401V>401L	496547G>C	
PvP01_07_v1	496556	G	A	5	15	103	0.05	missense	CRMP1	PVP01_0709800	404D>404N	496556G>A	
PvP01_07_v1	496563	C	A	48	14	61	0.44	missense	CRMP1	PVP01_0709800	406P>406H	496563C>A	
PvP01_07_v1	496565	C	T	1	9	113	0.01	missense	CRMP1	PVP01_0709800	407L>407F	496565C>T	
PvP01_07_v1	496613	G	A	58	20	45	0.56	*missense	CRMP1	PVP01_0709800	423D>423N	496613G>A	
PvP01_07_v1	496659	G	A	7	16	100	0.07	missense	CRMP1	PVP01_0709800	438R>438Q	496659G>A	
PvP01_07_v1	496742	A	G	1	17	105	0.01	missense	CRMP1	PVP01_0709800	466R>466G	496742A>G	
PvP01_07_v1	496776	C	T	40	16	67	0.37	*missense	CRMP1	PVP01_0709800	477S>477L	496776C>T	
PvP01_07_v1	496847	A	G	63	14	46	0.58	*missense	CRMP1	PVP01_0709800	501I>501V	496847A>G	
PvP01_07_v1	496859	A	G	63	14	46	0.58	*missense	CRMP1	PVP01_0709800	505T>505A	496859A>G	
PvP01_07_v1	497016	G	T	4	10	109	0.04	*missense	CRMP1	PVP01_0709800	557R>557I	497016G>T	
PvP01_07_v1	497177	A	C	48	15	60	0.44	*missense	CRMP1	PVP01_0709800	611I>611L	497177A>C	
PvP01_07_v1	497365	T	A	2	7	114	0.02	missense	CRMP1	PVP01_0709800	673S>673R	497365T>A	
PvP01_07_v1	497459	A	G	41	14	68	0.38	missense	CRMP1	PVP01_0709800	705T>705A	497459A>G	
PvP01_07_v1	497489	G	C	1	12	110	0.01	*missense	CRMP1	PVP01_0709800	715A>715P	497489G>C	
PvP01_07_v1	497573	C	T	7	17	99	0.07	missense	CRMP1	PVP01_0709800	743R>743C	497573C>T	
PvP01_07_v1	497820	G	A	1	15	107	0.01	missense	CRMP1	PVP01_0709800	825G>825E	497820G>A	
PvP01_07_v1	497960	C	T	43	21	59	0.42	*missense	CRMP1	PVP01_0709800	872L>872F	497960C>T	1217238A>G+1217240T>G
PvP01_07_v1	498259	A	G	3	16	104	0.03	missense	CRMP1	PVP01_0709800	971I>971M	498259A>G	
PvP01_07_v1	498835	C	T	9	14	100	0.08	synonymous	CRMP1	PVP01_0709800	1163N	498835C>T	
PvP01_07_v1	498878	A	G	97	26	0	1	*missense	CRMP1	PVP01_0709800	1178K>1178E	498878A>G	
PvP01_07_v1	498907	G	T	18	14	91	0.17	synonymous	CRMP1	PVP01_0709800	1187V	498907G>T	
PvP01_07_v1	498997	G	T	45	19	59	0.43	*missense	CRMP1	PVP01_0709800	1217K>1217N	498997G>T	
PvP01_07_v1	499188	T	A	3	8	112	0.03	missense	CRMP1	PVP01_0709800	1281I>1281N	499188T>A	
PvP01_07_v1	499708	G	T	3	5	115	0.03	*missense	CRMP1	PVP01_0709800	1454L>1454F	499708G>T	
PvP01_07_v1	500025	A	T	47	13	63	0.43	*missense	CRMP1	PVP01_0709800	1560K>1560I	500025A>T	
PvP01_07_v1	500025	A	T	47	13	63	0.43	missense	CRMP1	PVP01_0709800	1560K>1560V	500024A>G+500025A>T	
PvP01_07_v1	500041	C	A	15	11	97	0.13	synonymous	CRMP1	PVP01_0709800	1565P	500041C>A	
PvP01_07_v1	500176	A	T	28	11	84	0.25	synonymous	CRMP1	PVP01_0709800	1610P	500176A>T	
PvP01_07_v1	500177	T	C	107	16	0	1	*missense	CRMP1	PVP01_0709800	1611S>1611P	500177T>C	
PvP01_07_v1	500313	A	C	102	21	0	1	*missense	CRMP1	PVP01_0709800	1656Y>1656S	500313A>C	
PvP01_07_v1	500492	G	A	98	25	0	1	*missense	CRMP1	PVP01_0709800	1716V>1716M	500492G>A	
PvP01_07_v1	500547	A	G	33	14	76	0.3	missense	CRMP1	PVP01_0709800	1734K>1734R	500547A>G	
PvP01_07_v1	501450	C	T	32	10	81	0.28	*missense	CRMP1	PVP01_0709800	2035T>2035I	501450C>T	
PvP01_07_v1	501625	C	T	30	6	87	0.26	synonymous	CRMP1	PVP01_0709800	2093A	501625C>T	
PvP01_07_v1	501822	T	C	26	11	86	0.23	missense	CRMP1	PVP01_0709800	2159V>2159A	501822T>C	
PvP01_07_v1	503506	C	A	1	8	114	0.01	synonymous	CRMP1	PVP01_0709800	2720P	503506C>A	
PvP01_07_v1	503558	G	A	40	19	64	0.38	*missense	CRMP1	PVP01_0709800	2738E>2738R	503558G>A+503559A>G	
PvP01_07_v1	503559	A	G	40	19	64	0.38	*missense	CRMP1	PVP01_0709800	2738E>2738R	503558G>A+503559A>G	
PvP01_07_v1	503989	G	A	14	7	102	0.12	synonymous	CRMP1	PVP01_0709800	2881A	503989G>A	1218019A>C
PvP01_07_v1	1207313	C	A	17	8	98	0.15	missense	MSP1P	PVP01_0728800	113A>113D	1207313C>A	

PvP01_07_v1	1207343	A	T	15	9	99	0.13	missense	MSP1P	PVP01_0728800	123N>123I	1207343A>T	
PvP01_07_v1	1207374	G	C	54	8	61	0.47	missense	MSP1P	PVP01_0728800	133Q>133H	1207374G>C	
PvP01_07_v1	1207377	C	T	3	5	115	0.03	synonymous	MSP1P	PVP01_0728800	134H	1207377C>T	
PvP01_07_v1	1207398	G	A	1	2	120	0.01	synonymous	MSP1P	PVP01_0728800	141E	1207398G>A	
PvP01_07_v1	1207482	C	G	2	5	116	0.02	synonymous	MSP1P	PVP01_0728800	169S	1207482C>G	
PvP01_07_v1	1207618	G	A	7	6	110	0.06	missense	MSP1P	PVP01_0728800	215A>215T	1207618G>A	
PvP01_07_v1	1207689	C	A	1	11	111	0.01	synonymous	MSP1P	PVP01_0728800	238G	1207689C>A	
PvP01_07_v1	1207905	C	T	2	7	114	0.02	synonymous	MSP1P	PVP01_0728800	310S	1207905C>T	
PvP01_07_v1	1208213	C	T	11	8	104	0.1	missense	MSP1P	PVP01_0728800	413P>413L	1208213C>T	
PvP01_07_v1	1208352	C	T	6	11	106	0.05	synonymous	MSP1P	PVP01_0728800	459H	1208352C>T	
PvP01_07_v1	1208544	C	A	1	6	116	0.01	synonymous	MSP1P	PVP01_0728800	523T	1208544C>A	
PvP01_07_v1	1209366	T	C	2	5	116	0.02	synonymous	MSP1P	PVP01_0728800	797F	1209366T>C	
PvP01_07_v1	1209600	G	A	14	7	102	0.12	synonymous	MSP1P	PVP01_0728800	875K	1209600G>A	
PvP01_07_v1	1209739	C	T	8	4	111	0.07	missense	MSP1P	PVP01_0728800	922P>922S	1209739C>T	
PvP01_07_v1	1209900	C	T	3	3	117	0.03	synonymous	MSP1P	PVP01_0728800	975L	1209900C>T	
PvP01_07_v1	1209901	C	T	3	3	117	0.03	synonymous	MSP1P	PVP01_0728800	976L	1209901C>T	
PvP01_07_v1	1209909	C	T	3	3	117	0.03	synonymous	MSP1P	PVP01_0728800	978T	1209909C>T	
PvP01_07_v1	1210887	A	G	2	4	117	0.02	synonymous	MSP1P	PVP01_0728800	1304P	1210887A>G	
PvP01_07_v1	1211116	G	A	65	11	47	0.58	missense	MSP1P	PVP01_0728800	1381A>1381T	1211116G>A	
PvP01_07_v1	1212170	A	C	7	8	108	0.06	missense	MSP1P	PVP01_0728800	1732Y>1732S	1212170A>C	
PvP01_07_v1	1215536	G	A	54	18	51	0.51	synonymous	MSP1	PVP01_0728900	35E	1215536G>A	
PvP01_07_v1	1215589	A	G	46	27	50	0.48	missense	MSP1	PVP01_0728900	53N>53S	1215589A>G	
PvP01_07_v1	1215597	A	C	45	30	48	0.48	missense	MSP1	PVP01_0728900	56K>56Q	1215597A>C	
PvP01_07_v1	1215616	G	A	43	27	53	0.45	missense	MSP1	PVP01_0728900	62S>62N	1215616G>A	
PvP01_07_v1	1215667	G	T	6	23	94	0.06	missense	MSP1	PVP01_0728900	79R>79I	1215667G>T	
PvP01_07_v1	1215678	A	G	45	29	49	0.48	missense	MSP1	PVP01_0728900	83S>83G	1215678A>G	
PvP01_07_v1	1215712	C	T	1	18	104	0.01	missense	MSP1	PVP01_0728900	94T>94I	1215712C>T	
PvP01_07_v1	1215746	G	T	85	19	19	0.82	missense	MSP1	PVP01_0728900	105M>105I	1215746G>T	
PvP01_07_v1	1215812	A	C	87	17	19	0.82	synonymous	MSP1	PVP01_0728900	127I	1215812A>C	
PvP01_07_v1	1215863	C	T	36	21	66	0.35	synonymous	MSP1	PVP01_0728900	144H	1215863C>T	
PvP01_07_v1	1215897	T	A	13	17	93	0.12	missense	MSP1	PVP01_0728900	156S>156T	1215897T>A	
PvP01_07_v1	1215902	C	G	24	17	82	0.23	missense	MSP1	PVP01_0728900	157D>157E	1215902C>G	
PvP01_07_v1	1215920	C	G	31	24	68	0.31	synonymous	MSP1	PVP01_0728900	163L	1215920C>G	
PvP01_07_v1	1215932	T	G	31	22	70	0.31	synonymous	MSP1	PVP01_0728900	167V	1215932T>G	
PvP01_07_v1	1215938	C	T	31	23	69	0.31	synonymous	MSP1	PVP01_0728900	169G	1215938C>T	
PvP01_07_v1	1215965	A	G	54	23	46	0.54	synonymous	MSP1	PVP01_0728900	178K	1215965A>G	
PvP01_07_v1	1215971	C	T	53	25	45	0.54	synonymous	MSP1	PVP01_0728900	180D	1215971C>T	
PvP01_07_v1	1215986	G	A	29	19	75	0.28	synonymous	MSP1	PVP01_0728900	185E	1215986G>A	
PvP01_07_v1	1215992	T	C	52	24	47	0.53	synonymous	MSP1	PVP01_0728900	187F	1215992T>C	
PvP01_07_v1	1216047	A	C	8	32	83	0.09	missense	MSP1	PVP01_0728900	206N>206H	1216047A>C	1219806G>A
PvP01_07_v1	1216062	G	A	23	30	70	0.25	missense	MSP1	PVP01_0728900	211G>211S	1216062G>A	
PvP01_07_v1	1216069	C	A	18	60	45	0.29	missense	MSP1	PVP01_0728900	213S>213Y	1216069C>A	
PvP01_07_v1	1216082	T	C	7	64	52	0.12	synonymous	MSP1	PVP01_0728900	217I	1216082T>C	
PvP01_07_v1	1216089	A	G	1	69	53	0.02	missense	MSP1	PVP01_0728900	220N>220D	1216089A>G	
PvP01_07_v1	1216095	G	A	6	72	45	0.12	missense	MSP1	PVP01_0728900	222A>222T	1216095G>A+1216097G>A	
PvP01_07_v1	1216097	G	A	5	73	45	0.1	missense	MSP1	PVP01_0728900	222A>222T	1216095G>A+1216097G>A	
PvP01_07_v1	1216097	G	A	5	73	45	0.1	synonymous	MSP1	PVP01_0728900	222A	1216097G>A	
PvP01_07_v1	1216127	C	A	8	78	37	0.18	synonymous	MSP1	PVP01_0728900	232A	1216127C>A	
PvP01_07_v1	1216225	A	G	8	33	82	0.09	missense	MSP1	PVP01_0728900	265H>265R	1216225A>G	
PvP01_07_v1	1216240	C	A	3	35	85	0.03	missense	MSP1	PVP01_0728900	270A>270E	1216240C>A	

PvP01_07_v1	1216253	C	T	6	18	99	0.06	synonymous	MSP1	PVP01_0728900	274I	1216253C>T
PvP01_07_v1	1216263	A	G	24	17	82	0.23	missense	MSP1	PVP01_0728900	278I>278V	1216263A>G+1216265C>G
PvP01_07_v1	1216265	C	G	27	28	68	0.28	missense	MSP1	PVP01_0728900	278I>278M	1216265C>G
PvP01_07_v1	1216265	C	G	27	28	68	0.28	missense	MSP1	PVP01_0728900	278I>278V	1216263A>G+1216265C>G
PvP01_07_v1	1216272	G	A	24	18	81	0.23	missense	MSP1	PVP01_0728900	281G>281S	1216272G>A
PvP01_07_v1	1216288	A	G	4	22	97	0.04	missense	MSP1	PVP01_0728900	286N>286S	1216288A>G
PvP01_07_v1	1216296	G	C	20	14	89	0.18	missense	MSP1	PVP01_0728900	289E>289Q	1216296G>C+1216298G>A
PvP01_07_v1	1216298	G	A	6	14	103	0.06	missense	MSP1	PVP01_0728900	289E>289Q	1216296G>C+1216298G>A
PvP01_07_v1	1216298	G	A	6	14	103	0.06	synonymous	MSP1	PVP01_0728900	289E	1216298G>A
PvP01_07_v1	1216304	G	T	20	13	90	0.18	synonymous	MSP1	PVP01_0728900	291A	1216304G>T
PvP01_07_v1	1216318	C	A	71	25	27	0.72	missense	MSP1	PVP01_0728900	296A>296E	1216318C>A
PvP01_07_v1	1216325	G	T	30	20	73	0.29	synonymous	MSP1	PVP01_0728900	298L	1216325G>T
PvP01_07_v1	1216328	A	G	70	27	26	0.73	synonymous	MSP1	PVP01_0728900	299E	1216328A>G
PvP01_07_v1	1216346	A	G	60	29	34	0.64	synonymous	MSP1	PVP01_0728900	305L	1216346A>G
PvP01_07_v1	1216357	A	C	3	22	98	0.03	missense	MSP1	PVP01_0728900	309K>309T	1216357A>C
PvP01_07_v1	1216360	A	G	32	27	64	0.33	missense	MSP1	PVP01_0728900	310D>310G	1216360A>G
PvP01_07_v1	1216364	T	C	67	31	25	0.73	synonymous	MSP1	PVP01_0728900	311I	1216364T>C
PvP01_07_v1	1216369	T	C	32	27	64	0.33	missense	MSP1	PVP01_0728900	313V>313A	1216369T>C
PvP01_07_v1	1216377	G	A	22	29	72	0.23	missense	MSP1	PVP01_0728900	316E>316K	1216377G>A
PvP01_07_v1	1216380	C	A	26	29	68	0.28	missense	MSP1	PVP01_0728900	317Q>317K	1216380C>A
PvP01_07_v1	1216383	G	A	35	30	58	0.38	missense	MSP1	PVP01_0728900	318V>318I	1216383G>A
PvP01_07_v1	1216403	G	T	9	33	81	0.1	missense	MSP1	PVP01_0728900	324K>324N	1216403G>T
PvP01_07_v1	1216404	C	T	26	32	65	0.29	missense	MSP1	PVP01_0728900	325L>325F	1216404C>T
PvP01_07_v1	1216409	T	A	26	37	60	0.3	synonymous	MSP1	PVP01_0728900	326P	1216409T>A
PvP01_07_v1	1216411	A	G	18	33	72	0.2	missense	MSP1	PVP01_0728900	327N>327R	1216411A>G+1216412T>A
PvP01_07_v1	1216411	A	G	18	33	72	0.2	missense	MSP1	PVP01_0728900	327N>327S	1216411A>G
PvP01_07_v1	1216412	T	A	39	37	47	0.45	missense	MSP1	PVP01_0728900	327N>327K	1216412T>A
PvP01_07_v1	1216412	T	A	39	37	47	0.45	missense	MSP1	PVP01_0728900	327N>327R	1216411A>G+1216412T>A
PvP01_07_v1	1216413	G	A	27	37	59	0.31	missense	MSP1	PVP01_0728900	328D>328N	1216413G>A
PvP01_07_v1	1216416	T	A	64	40	19	0.77	missense	MSP1	PVP01_0728900	329Y>329N	1216416T>A
PvP01_07_v1	1216419	C	A	39	37	47	0.45	missense	MSP1	PVP01_0728900	330P>330I	1216419C>A+1216420C>T
PvP01_07_v1	1216419	C	A	39	37	47	0.45	missense	MSP1	PVP01_0728900	330P>330T	1216419C>A
PvP01_07_v1	1216420	C	T	25	34	64	0.28	missense	MSP1	PVP01_0728900	330P>330I	1216419C>A+1216420C>T
PvP01_07_v1	1216420	C	T	25	34	64	0.28	missense	MSP1	PVP01_0728900	330P>330L	1216420C>T
PvP01_07_v1	1216423	A	C	38	36	49	0.44	missense	MSP1	PVP01_0728900	331N>331T	1216423A>C
PvP01_07_v1	1216431	A	C	24	38	61	0.28	missense	MSP1	PVP01_0728900	334N>334P	1216431A>C+1216432A>C+1216433T>C
PvP01_07_v1	1216432	A	C	35	39	49	0.42	missense	MSP1	PVP01_0728900	334N>334P	1216431A>C+1216432A>C+1216433T>C
PvP01_07_v1	1216432	A	C	35	39	49	0.42	missense	MSP1	PVP01_0728900	334N>334T	1216432A>C+1216433T>C
PvP01_07_v1	1216433	T	C	35	37	51	0.41	missense	MSP1	PVP01_0728900	334N>334P	1216431A>C+1216432A>C+1216433T>C
PvP01_07_v1	1216433	T	C	35	37	51	0.41	missense	MSP1	PVP01_0728900	334N>334T	1216432A>C+1216433T>C
PvP01_07_v1	1216435	T	C	25	41	57	0.3	missense	MSP1	PVP01_0728900	335L>335P	1216435T>C
PvP01_07_v1	1216440	A	G	59	42	22	0.73	missense	MSP1	PVP01_0728900	337N>337A	1216440A>G+1216441A>C
PvP01_07_v1	1216440	A	G	59	42	22	0.73	missense	MSP1	PVP01_0728900	337N>337D	1216440A>G
PvP01_07_v1	1216441	A	C	11	32	80	0.12	missense	MSP1	PVP01_0728900	337N>337A	1216440A>G+1216441A>C
PvP01_07_v1	1216444	T	A	55	46	22	0.71	missense	MSP1	PVP01_0728900	338V>338E	1216444T>A
PvP01_07_v1	1216452	G	A	24	56	43	0.36	missense	MSP1	PVP01_0728900	341E>341T	1216452G>A+1216453A>C
PvP01_07_v1	1216453	A	C	41	60	22	0.65	missense	MSP1	PVP01_0728900	341E>341A	1216453A>C
PvP01_07_v1	1216453	A	C	41	60	22	0.65	missense	MSP1	PVP01_0728900	341E>341T	1216452G>A+1216453A>C
PvP01_07_v1	1216466	A	C	30	28	65	0.32	missense	MSP1	PVP01_0728900	345K>345N	1216466A>C
PvP01_07_v1	1216475	G	C	46	52	25	0.65	missense	MSP1	PVP01_0728900	348E>348D	1216475G>C

PvP01_07_v1	1216481	G	A	30	39	54	0.36	synonymous	MSP1	PVP01_0728900	350E	1216481G>A
PvP01_07_v1	1216490	C	T	6	22	95	0.06	synonymous	MSP1	PVP01_0728900	353I	1216490C>T
PvP01_07_v1	1216492	A	T	22	40	61	0.27	missense	MSP1	PVP01_0728900	354E>354V	1216492A>T
PvP01_07_v1	1216495	C	A	6	16	101	0.06	missense	MSP1	PVP01_0728900	355A>355D	1216495C>A
PvP01_07_v1	1216499	C	T	6	14	103	0.06	synonymous	MSP1	PVP01_0728900	356I	1216499C>T
PvP01_07_v1	1216505	G	A	6	15	102	0.06	synonymous	MSP1	PVP01_0728900	358K	1216505G>A
PvP01_07_v1	1216508	T	G	37	31	55	0.4	synonymous	MSP1	PVP01_0728900	359T	1216508T>G
PvP01_07_v1	1216514	C	T	7	15	101	0.06	synonymous	MSP1	PVP01_0728900	361N	1216514C>T
PvP01_07_v1	1216535	T	C	36	20	67	0.35	synonymous	MSP1	PVP01_0728900	368F	1216535T>C
PvP01_07_v1	1216539	G	A	36	20	67	0.35	missense	MSP1	PVP01_0728900	370D>370N	1216539G>A
PvP01_07_v1	1216550	G	A	29	20	74	0.28	synonymous	MSP1	PVP01_0728900	373E	1216550G>A
PvP01_07_v1	1216556	G	A	1	16	106	0.01	synonymous	MSP1	PVP01_0728900	375E	1216556G>A
PvP01_07_v1	1216562	T	C	46	23	54	0.46	synonymous	MSP1	PVP01_0728900	377Y	1216562T>C
PvP01_07_v1	1216599	A	G	27	22	74	0.27	missense	MSP1	PVP01_0728900	390I>390V	1216599A>G
PvP01_07_v1	1216634	A	G	29	19	75	0.28	synonymous	MSP1	PVP01_0728900	401G	1216634A>G
PvP01_07_v1	1216670	C	T	59	19	45	0.57	synonymous	MSP1	PVP01_0728900	413H	1216670C>T
PvP01_07_v1	1216700	T	C	61	17	45	0.58	synonymous	MSP1	PVP01_0728900	423S	1216700T>C
PvP01_07_v1	1216706	T	C	60	17	46	0.57	synonymous	MSP1	PVP01_0728900	425Y	1216706T>C
PvP01_07_v1	1216778	G	A	7	15	101	0.06	synonymous	MSP1	PVP01_0728900	449P	1216778G>A
PvP01_07_v1	1216785	G	A	7	12	104	0.06	missense	MSP1	PVP01_0728900	452G>452R	1216785G>A
PvP01_07_v1	1216804	C	T	7	15	101	0.06	missense	MSP1	PVP01_0728900	458T>458I	1216804C>T
PvP01_07_v1	1216817	A	G	6	20	97	0.06	synonymous	MSP1	PVP01_0728900	462E	1216817A>G
PvP01_07_v1	1216869	A	G	7	18	98	0.07	missense	MSP1	PVP01_0728900	480N>480D	1216869A>G
PvP01_07_v1	1216886	C	T	6	17	100	0.06	synonymous	MSP1	PVP01_0728900	485Y	1216886C>T
PvP01_07_v1	1216894	A	T	1	16	106	0.01	missense	MSP1	PVP01_0728900	488K>488I	1216894A>T
PvP01_07_v1	1216897	A	T	6	15	102	0.06	missense	MSP1	PVP01_0728900	489Y>489F	1216897A>T
PvP01_07_v1	1216900	A	C	7	14	102	0.06	missense	MSP1	PVP01_0728900	490K>490T	1216900A>C
PvP01_07_v1	1216907	C	T	6	15	102	0.06	synonymous	MSP1	PVP01_0728900	492Y	1216907C>T
PvP01_07_v1	1216912	C	A	23	19	81	0.22	missense	MSP1	PVP01_0728900	494A>494E	1216912C>A
PvP01_07_v1	1216922	T	G	13	17	93	0.12	missense	MSP1	PVP01_0728900	497N>497K	1216922T>G
PvP01_07_v1	1216925	G	T	6	14	103	0.06	missense	MSP1	PVP01_0728900	498E>498D	1216925G>T
PvP01_07_v1	1216929	A	C	13	16	94	0.12	missense	MSP1	PVP01_0728900	500K>500Q	1216929A>C
PvP01_07_v1	1216936	C	A	7	16	100	0.07	missense	MSP1	PVP01_0728900	502A>502E	1216936C>A
PvP01_07_v1	1216946	C	T	6	17	100	0.06	missense	MSP1	PVP01_0728900	505H>505Y	1216944C>T+1216946C>T
PvP01_07_v1	1216946	C	T	6	17	100	0.06	synonymous	MSP1	PVP01_0728900	505H	1216946C>T
PvP01_07_v1	1216973	C	T	9	15	99	0.08	synonymous	MSP1	PVP01_0728900	514T	1216973C>T
PvP01_07_v1	1216981	A	G	9	13	101	0.08	missense	MSP1	PVP01_0728900	517E>517G	1216981A>G
PvP01_07_v1	1216983	A	G	8	14	101	0.07	missense	MSP1	PVP01_0728900	518N>518D	1216983A>G
PvP01_07_v1	1216983	A	G	8	14	101	0.07	missense	MSP1	PVP01_0728900	518N>518E	1216983A>G+1216985C>A
PvP01_07_v1	1216985	C	A	8	14	101	0.07	missense	MSP1	PVP01_0728900	518N>518E	1216983A>G+1216985C>A
PvP01_07_v1	1216993	A	G	9	15	99	0.08	missense	MSP1	PVP01_0728900	521D>521G	1216993A>G
PvP01_07_v1	1216997	A	T	19	12	92	0.17	missense	MSP1	PVP01_0728900	522E>522D	1216997A>T
PvP01_07_v1	1217012	G	A	17	15	91	0.16	synonymous	MSP1	PVP01_0728900	527R	1217012G>A
PvP01_07_v1	1217017	C	A	10	15	98	0.09	missense	MSP1	PVP01_0728900	529A>529E	1217017C>A
PvP01_07_v1	1217025	G	A	10	16	97	0.09	missense	MSP1	PVP01_0728900	532V>532K	1217025G>A+1217026T>A
PvP01_07_v1	1217026	T	A	99	24	0	1	missense	MSP1	PVP01_0728900	532V>532E	1217026T>A
PvP01_07_v1	1217026	T	A	99	24	0	1	missense	MSP1	PVP01_0728900	532V>532K	1217025G>A+1217026T>A
PvP01_07_v1	1217045	A	G	99	23	1	0.99	synonymous	MSP1	PVP01_0728900	538E	1217045A>G
PvP01_07_v1	1217068	G	C	103	20	0	1	missense	MSP1	PVP01_0728900	546S>546A	1217067A>G+1217068G>C
PvP01_07_v1	1217068	G	C	103	20	0	1	missense	MSP1	PVP01_0728900	546S>546T	1217068G>C

PvP01_07_v1	1217085	C	T	96	23	4	0.96	synonymous	MSP1	PVP01_0728900	552L	1217085C>T
PvP01_07_v1	1217102	C	A	5	14	104	0.05	synonymous	MSP1	PVP01_0728900	557T	1217102C>A
PvP01_07_v1	1217120	G	T	5	15	103	0.05	synonymous	MSP1	PVP01_0728900	563V	1217120G>T
PvP01_07_v1	1217238	A	G	89	19	15	0.86	missense	MSP1	PVP01_0728900	603I>603V	1217238A>G
PvP01_07_v1	1217252	C	T	1	9	113	0.01	synonymous	MSP1	PVP01_0728900	607F	1217252C>T
PvP01_07_v1	1217370	A	G	96	21	6	0.94	missense	MSP1	PVP01_0728900	647I>647V	1217370A>G
PvP01_07_v1	1217463	A	C	1	15	107	0.01	missense	MSP1	PVP01_0728900	678K>678Q	1217463A>C
PvP01_07_v1	1217618	A	T	57	65	1	0.98	missense	MSP1	PVP01_0728900	729Q>729H	1217618A>T
PvP01_07_v1	1217626	T	C	8	63	52	0.13	missense	MSP1	PVP01_0728900	732V>732A	1217626T>C
PvP01_07_v1	1217770	T	C	53	26	44	0.55	missense	MSP1	PVP01_0728900	780I>780T	1217770T>C
PvP01_07_v1	1217862	A	G	3	60	60	0.05	missense	MSP1	PVP01_0728900	811T>811A	1217862A>G
PvP01_07_v1	1217882	T	C	18	65	40	0.31	synonymous	MSP1	PVP01_0728900	817Y	1217882T>C
PvP01_07_v1	1217885	G	C	18	64	41	0.31	synonymous	MSP1	PVP01_0728900	818L	1217885G>C
PvP01_07_v1	1217886	C	G	18	64	41	0.31	missense	MSP1	PVP01_0728900	819Q>819E	1217886C>G
PvP01_07_v1	1217936	T	C	48	18	57	0.46	synonymous	MSP1	PVP01_0728900	835I	1217936T>C
PvP01_07_v1	1217939	T	C	48	18	57	0.46	synonymous	MSP1	PVP01_0728900	836F	1217939T>C
PvP01_07_v1	1217942	G	A	50	18	55	0.48	synonymous	MSP1	PVP01_0728900	837V	1217942G>A
PvP01_07_v1	1217958	A	G	36	18	69	0.34	missense	MSP1	PVP01_0728900	843K>843D	1217958A>G+1217960A>C
PvP01_07_v1	1217960	A	C	36	18	69	0.34	missense	MSP1	PVP01_0728900	843K>843D	1217958A>G+1217960A>C
PvP01_07_v1	1217960	A	C	36	18	69	0.34	missense	MSP1	PVP01_0728900	843K>843N	1217960A>C
PvP01_07_v1	1217964	G	A	36	19	68	0.35	missense	MSP1	PVP01_0728900	845E>845K	1217964G>A+1217966G>A
PvP01_07_v1	1217966	G	A	36	19	68	0.35	missense	MSP1	PVP01_0728900	845E>845K	1217964G>A+1217966G>A
PvP01_07_v1	1217966	G	A	36	19	68	0.35	synonymous	MSP1	PVP01_0728900	845E	1217966G>A
PvP01_07_v1	1217973	G	A	36	18	69	0.34	missense	MSP1	PVP01_0728900	848D>848K	1217973G>A+1217975T>A
PvP01_07_v1	1217973	G	A	36	18	69	0.34	missense	MSP1	PVP01_0728900	848D>848N	1217973G>A
PvP01_07_v1	1217975	T	A	37	19	67	0.36	missense	MSP1	PVP01_0728900	848D>848E	1217975T>A
PvP01_07_v1	1217975	T	A	37	19	67	0.36	missense	MSP1	PVP01_0728900	848D>848K	1217973G>A+1217975T>A
PvP01_07_v1	1217976	C	G	37	19	67	0.36	missense	MSP1	PVP01_0728900	849Q>849E	1217976C>G
PvP01_07_v1	1217982	A	G	37	19	67	0.36	missense	MSP1	PVP01_0728900	851K>851E	1217982A>G
PvP01_07_v1	1218000	C	A	37	19	67	0.36	missense	MSP1	PVP01_0728900	857Q>857K	1218000C>A
PvP01_07_v1	1218004	A	C	37	18	68	0.35	missense	MSP1	PVP01_0728900	858N>858T	1218004A>C
PvP01_07_v1	1218015	G	C	30	17	76	0.28	missense	MSP1	PVP01_0728900	862E>862Q	1218015G>C
PvP01_07_v1	1218019	A	C	9	14	100	0.08	missense	MSP1	PVP01_0728900	863N>863T	1218019A>C+1218020T>C
PvP01_07_v1	1218020	T	C	30	17	76	0.28	missense	MSP1	PVP01_0728900	863N>863T	1218019A>C+1218020T>C
PvP01_07_v1	1218020	T	C	30	17	76	0.28	synonymous	MSP1	PVP01_0728900	863N	1218020T>C
PvP01_07_v1	1218053	C	T	19	14	90	0.17	synonymous	MSP1	PVP01_0728900	874V	1218053C>T
PvP01_07_v1	1218092	C	T	49	20	54	0.48	synonymous	MSP1	PVP01_0728900	887S	1218092C>T
PvP01_07_v1	1218100	G	A	93	27	3	0.97	missense	MSP1	PVP01_0728900	890S>890N	1218100G>A
PvP01_07_v1	1218104	G	A	48	21	54	0.47	synonymous	MSP1	PVP01_0728900	891E	1218104G>A
PvP01_07_v1	1218107	G	A	17	19	87	0.16	synonymous	MSP1	PVP01_0728900	892L	1218107G>A
PvP01_07_v1	1218113	T	C	64	22	37	0.63	synonymous	MSP1	PVP01_0728900	894N	1218113T>C
PvP01_07_v1	1218126	T	C	72	28	23	0.76	synonymous	MSP1	PVP01_0728900	899L	1218126T>C
PvP01_07_v1	1218158	T	C	56	26	41	0.58	synonymous	MSP1	PVP01_0728900	909Y	1218158T>C
PvP01_07_v1	1218168	A	G	80	40	3	0.96	missense	MSP1	PVP01_0728900	913N>913D	1218168A>G
PvP01_07_v1	1218176	T	C	76	41	6	0.93	synonymous	MSP1	PVP01_0728900	915D	1218176T>C
PvP01_07_v1	1218183	C	A	75	42	6	0.93	missense	MSP1	PVP01_0728900	918L>918I	1218183C>A
PvP01_07_v1	1218237	A	C	50	55	18	0.74	missense	MSP1	PVP01_0728900	936K>936Q	1218237A>C
PvP01_07_v1	1218240	C	T	49	56	18	0.73	missense	MSP1	PVP01_0728900	937P>937S	1218240C>T
PvP01_07_v1	1218391	C	G	1	101	21	0.05	missense	MSP1	PVP01_0728900	987A>987G	1218391C>G
PvP01_07_v1	1218398	A	C	1	101	21	0.05	missense	MSP1	PVP01_0728900	989E>989D	1218398A>C

PvP01_07_v1	1218401	T	C	1	101	21	0.05	synonymous	MSP1	PVP01_0728900	990S	1218401T>C	
PvP01_07_v1	1218441	A	T	26	76	21	0.55	missense	MSP1	PVP01_0728900	1004I>1004F	1218441A>T	
PvP01_07_v1	1218449	A	G	29	73	21	0.58	synonymous	MSP1	PVP01_0728900	1006E	1218449A>G	
PvP01_07_v1	1218482	C	T	83	20	20	0.81	synonymous	MSP1	PVP01_0728900	1017D	1218482C>T	
PvP01_07_v1	1218489	C	A	83	20	20	0.81	missense	MSP1	PVP01_0728900	1020Q>1020K	1218489C>A	
PvP01_07_v1	1218494	G	A	83	20	20	0.81	synonymous	MSP1	PVP01_0728900	1021K	1218494G>A	
PvP01_07_v1	1218507	C	A	102	21	0	1	missense	MSP1	PVP01_0728900	1026Q>1026K	1218507C>A	
PvP01_07_v1	1218515	T	G	83	21	19	0.81	missense	MSP1	PVP01_0728900	1028D>1028E	1218515T>G	480551C>T+480552G>T
PvP01_07_v1	1218523	C	T	83	21	19	0.81	missense	MSP1	PVP01_0728900	1031T>1031I	1218523C>T	
PvP01_07_v1	1218536	A	T	51	22	50	0.5	missense	MSP1	PVP01_0728900	1035E>1035D	1218536A>T	
PvP01_07_v1	1218548	A	T	51	22	50	0.5	missense	MSP1	PVP01_0728900	1039K>1039N	1218548A>T	
PvP01_07_v1	1218558	G	A	51	23	49	0.51	missense	MSP1	PVP01_0728900	1043A>1043K	1218558G>A+1218559C>A	
PvP01_07_v1	1218559	C	A	51	23	49	0.51	missense	MSP1	PVP01_0728900	1043A>1043E	1218559C>A	
PvP01_07_v1	1218559	C	A	51	23	49	0.51	missense	MSP1	PVP01_0728900	1043A>1043K	1218558G>A+1218559C>A	
PvP01_07_v1	1218636	C	T	102	21	0	1	missense	MSP1	PVP01_0728900	1069L>1069F	1218636C>T	
PvP01_07_v1	1218656	A	C	2	13	108	0.02	missense	MSP1	PVP01_0728900	1075K>1075N	1218656A>C	
PvP01_07_v1	1218779	G	A	52	17	54	0.49	synonymous	MSP1	PVP01_0728900	1116K	1218779G>A	
PvP01_07_v1	1218856	G	A	1	13	109	0.01	missense	MSP1	PVP01_0728900	1142R>1142Q	1218856G>A	
PvP01_07_v1	1218872	A	G	7	13	103	0.06	missense	MSP1	PVP01_0728900	1147I>1147M	1218872A>G	
PvP01_07_v1	1218912	T	A	98	17	8	0.92	missense	MSP1	PVP01_0728900	1161L>1161M	1218912T>A	
PvP01_07_v1	1218966	C	T	35	14	74	0.32	synonymous	MSP1	PVP01_0728900	1179L	1218966C>T	
PvP01_07_v1	1218995	A	G	75	17	31	0.71	synonymous	MSP1	PVP01_0728900	1188E	1218995A>G	
PvP01_07_v1	1219036	T	A	70	20	33	0.68	missense	MSP1	PVP01_0728900	1202L>1202D	1219035C>G+1219036T>A	
PvP01_07_v1	1219036	T	A	70	20	33	0.68	missense	MSP1	PVP01_0728900	1202L>1202H	1219036T>A	
PvP01_07_v1	1219070	A	C	31	17	75	0.29	missense	MSP1	PVP01_0728900	1213K>1213N	1219070A>C	
PvP01_07_v1	1219071	G	A	24	65	34	0.41	missense	MSP1	PVP01_0728900	1214D>1214N	1219071G>A	
PvP01_07_v1	1219079	A	G	24	64	35	0.41	synonymous	MSP1	PVP01_0728900	1216K	1219079A>G	
PvP01_07_v1	1219091	A	C	23	65	35	0.4	missense	MSP1	PVP01_0728900	1220K>1220N	1219091A>C	1397780G>A+1397781A>G
PvP01_07_v1	1219098	G	A	23	65	35	0.4	missense	MSP1	PVP01_0728900	1223A>1223T	1219098G>A	
PvP01_07_v1	1219112	G	A	21	69	33	0.39	synonymous	MSP1	PVP01_0728900	1227A	1219112G>A	
PvP01_07_v1	1219114	A	C	21	69	33	0.39	missense	MSP1	PVP01_0728900	1228E>1228A	1219114A>C+1219115A>T	
PvP01_07_v1	1219115	A	T	21	69	33	0.39	missense	MSP1	PVP01_0728900	1228E>1228A	1219114A>C+1219115A>T	
PvP01_07_v1	1219122	A	G	21	70	32	0.4	missense	MSP1	PVP01_0728900	1231K>1231D	1219122A>G+1219124A>C	
PvP01_07_v1	1219122	A	G	21	70	32	0.4	missense	MSP1	PVP01_0728900	1231K>1231E	1219122A>G	
PvP01_07_v1	1219127	G	T	21	71	31	0.4	synonymous	MSP1	PVP01_0728900	1232A	1219127G>T	
PvP01_07_v1	1219133	G	A	21	71	31	0.4	synonymous	MSP1	PVP01_0728900	1234Q	1219133G>A	
PvP01_07_v1	1219157	A	G	16	75	32	0.33	synonymous	MSP1	PVP01_0728900	1242K	1219157A>G	
PvP01_07_v1	1219160	A	G	16	74	33	0.33	synonymous	MSP1	PVP01_0728900	1243V	1219160A>G	
PvP01_07_v1	1219164	T	A	16	77	30	0.35	missense	MSP1	PVP01_0728900	1245S>1245T	1219164T>A	
PvP01_07_v1	1219176	A	G	2	78	43	0.04	missense	MSP1	PVP01_0728900	1249T>1249A	1219176A>G	
PvP01_07_v1	1219341	C	A	18	63	42	0.3	missense	MSP1	PVP01_0728900	1304Q>1304K	1219341C>A	
PvP01_07_v1	1219388	G	A	43	39	41	0.51	synonymous	MSP1	PVP01_0728900	1319A	1219388G>A	
PvP01_07_v1	1219391	G	C	43	39	41	0.51	synonymous	MSP1	PVP01_0728900	1320L	1219391G>C	
PvP01_07_v1	1219394	C	T	43	39	41	0.51	synonymous	MSP1	PVP01_0728900	1321P	1219394C>T	
PvP01_07_v1	1219400	T	C	43	39	41	0.51	synonymous	MSP1	PVP01_0728900	1323F	1219400T>C	
PvP01_07_v1	1219462	C	A	53	17	53	0.5	missense	MSP1	PVP01_0728900	1344A>1344E	1219462C>A	
PvP01_07_v1	1219674	T	C	17	11	95	0.15	missense	MSP1	PVP01_0728900	1415S>1415P	1219674T>C	
PvP01_07_v1	1219740	A	C	81	23	19	0.81	missense	MSP1	PVP01_0728900	1437I>1437L	1219740A>C	
PvP01_07_v1	1219777	A	T	16	13	94	0.15	missense	MSP1	PVP01_0728900	1449K>1449M	1219777A>T	
PvP01_07_v1	1219806	G	A	35	19	69	0.34	missense	MSP1	PVP01_0728900	1459A>1459T	1219806G>A+1219808G>C	

PvP01_07_v1	1219808	G	C	35	19	69	0.34	missense	MSP1	PVP01_0728900	1459A>1459T	1219806G>A+1219808G>C
PvP01_07_v1	1219808	G	C	35	19	69	0.34	synonymous	MSP1	PVP01_0728900	1459A	1219808G>C
PvP01_07_v1	1219825	A	G	39	20	64	0.38	missense	MSP1	PVP01_0728900	1465E>1465G	1219825A>G
PvP01_07_v1	1219848	A	G	25	19	79	0.24	missense	MSP1	PVP01_0728900	1473T>1473D	1219848A>G+1219849C>A
PvP01_07_v1	1219849	C	A	25	19	79	0.24	missense	MSP1	PVP01_0728900	1473T>1473D	1219848A>G+1219849C>A
PvP01_07_v1	1219852	C	G	25	19	79	0.24	missense	MSP1	PVP01_0728900	1474A>1474G	1219852C>G
PvP01_07_v1	1219859	T	A	26	18	79	0.25	missense	MSP1	PVP01_0728900	1476N>1476K	1219859T>A
PvP01_07_v1	1219860	G	A	26	18	79	0.25	missense	MSP1	PVP01_0728900	1477A>1477K	1219860G>A+1219861C>A
PvP01_07_v1	1219860	G	A	26	18	79	0.25	missense	MSP1	PVP01_0728900	1477A>1477T	1219860G>A
PvP01_07_v1	1219861	C	A	17	21	85	0.17	missense	MSP1	PVP01_0728900	1477A>1477E	1219861C>A
PvP01_07_v1	1219861	C	A	17	21	85	0.17	missense	MSP1	PVP01_0728900	1477A>1477K	1219860G>A+1219861C>A
PvP01_07_v1	1219863	C	G	43	24	56	0.43	missense	MSP1	PVP01_0728900	1478Q>1478E	1219863C>G
PvP01_07_v1	1219866	A	G	17	22	84	0.17	missense	MSP1	PVP01_0728900	1479I>1479V	1219866A>G
PvP01_07_v1	1219882	A	C	20	24	79	0.2	missense	MSP1	PVP01_0728900	1484D>1484A	1219882A>C
PvP01_07_v1	1219892	C	A	77	25	21	0.79	missense	MSP1	PVP01_0728900	1487N>1487K	1219892C>A
PvP01_07_v1	1219893	A	G	21	23	79	0.21	missense	MSP1	PVP01_0728900	1488T>1488A	1219893A>G
PvP01_07_v1	1219893	A	G	21	23	79	0.21	missense	MSP1	PVP01_0728900	1488T>1488E	I893A>G+1219894C>A+1219895A>G
PvP01_07_v1	1219894	C	A	40	22	61	0.4	missense	MSP1	PVP01_0728900	1488T>1488E	I893A>G+1219894C>A+1219895A>G
PvP01_07_v1	1219894	C	A	40	22	61	0.4	missense	MSP1	PVP01_0728900	1488T>1488K	1219894C>A+1219895A>G
PvP01_07_v1	1219895	A	G	40	22	61	0.4	missense	MSP1	PVP01_0728900	1488T>1488E	I893A>G+1219894C>A+1219895A>G
PvP01_07_v1	1219895	A	G	40	22	61	0.4	missense	MSP1	PVP01_0728900	1488T>1488K	1219894C>A+1219895A>G
PvP01_07_v1	1219896	C	G	21	23	79	0.21	missense	MSP1	PVP01_0728900	1489Q>1489D	1219896C>G+1219898A>T
PvP01_07_v1	1219896	C	G	21	23	79	0.21	missense	MSP1	PVP01_0728900	1489Q>1489E	1219896C>G
PvP01_07_v1	1219898	A	T	16	21	86	0.16	missense	MSP1	PVP01_0728900	1489Q>1489D	1219896C>G+1219898A>T
PvP01_07_v1	1219898	A	T	16	21	86	0.16	missense	MSP1	PVP01_0728900	1489Q>1489H	1219898A>T
PvP01_07_v1	1219899	A	G	75	28	20	0.79	missense	MSP1	PVP01_0728900	1490N>1490D	1219899A>G
PvP01_07_v1	1219917	A	T	22	23	78	0.22	missense	MSP1	PVP01_0728900	1496I>1496L	1219917A>T
PvP01_07_v1	1219921	A	G	38	28	57	0.4	missense	MSP1	PVP01_0728900	1497E>1497G	1219921A>G
PvP01_07_v1	1219924	A	G	33	29	61	0.35	missense	MSP1	PVP01_0728900	1498N>1498S	1219924A>G
PvP01_07_v1	1219935	A	C	16	24	83	0.16	missense	MSP1	PVP01_0728900	1502K>1502Q	1219935A>C
PvP01_07_v1	1219946	A	G	7	21	95	0.07	synonymous	MSP1	PVP01_0728900	1505E	1219946A>G
PvP01_07_v1	1219949	G	C	7	23	93	0.07	missense	MSP1	PVP01_0728900	1506K>1506N	1219949G>C
PvP01_07_v1	1219958	T	G	34	30	59	0.37	missense	MSP1	PVP01_0728900	1509F>1509L	1219958T>G
PvP01_07_v1	1219977	C	T	44	16	63	0.41	synonymous	MSP1	PVP01_0728900	1516L	1219977C>T
PvP01_07_v1	1219994	G	A	44	15	64	0.41	synonymous	MSP1	PVP01_0728900	1521P	1219994G>A
PvP01_07_v1	1220069	A	C	1	11	111	0.01	missense	MSP1	PVP01_0728900	1546K>1546N	1220069A>C
PvP01_07_v1	1220072	C	T	36	17	70	0.34	synonymous	MSP1	PVP01_0728900	1547V	1220072C>T
PvP01_07_v1	1220077	A	G	18	17	88	0.17	missense	MSP1	PVP01_0728900	1549N>1549S	1220077A>G
PvP01_07_v1	1220089	T	A	6	12	105	0.05	missense	MSP1	PVP01_0728900	1553L>1553H	1220089T>A+1220090G>T
PvP01_07_v1	1220090	G	T	23	12	88	0.21	missense	MSP1	PVP01_0728900	1553L>1553H	1220089T>A+1220090G>T
PvP01_07_v1	1220090	G	T	23	12	88	0.21	synonymous	MSP1	PVP01_0728900	1553L	1220090G>T
PvP01_07_v1	1220137	A	C	3	17	103	0.03	missense	MSP1	PVP01_0728900	1569N>1569T	1220137A>C
PvP01_07_v1	1220215	A	G	38	14	71	0.35	missense	MSP1	PVP01_0728900	1595K>1595R	1220215A>G
PvP01_07_v1	1220379	G	A	2	11	110	0.02	missense	MSP1	PVP01_0728900	1650E>1650K	1220379G>A
PvP01_07_v1	1220505	G	A	106	17	0	1	missense	MSP1	PVP01_0728900	1692E>1692K	1220505G>A
PvP01_08_v1	34144	G	T	47	57	19	0.71	missense	RBP2b	PVP01_0800700	15L>15F	34144G>T
PvP01_08_v1	34503	C	T	21	39	63	0.25	missense	RBP2b	PVP01_0800700	39S>39L	34503C>T
PvP01_08_v1	34793	A	G	50	53	20	0.71	missense	RBP2b	PVP01_0800700	136K>136E	34793A>G
PvP01_08_v1	35037	A	G	72	51	0	1	missense	RBP2b	PVP01_0800700	217H>217C	35036C>T+35037A>G
PvP01_08_v1	35037	A	G	72	51	0	1	missense	RBP2b	PVP01_0800700	217H>217R	35037A>G

PvP01_08_v1	35080	C	A	1	42	80	0.01	missense	RBP2b	PVP01_0800700	231N>231K	35080C>A	
PvP01_08_v1	35113	C	G	34	54	35	0.49	missense	RBP2b	PVP01_0800700	242S>242R	35113C>G	
PvP01_08_v1	35113	C	G	34	54	35	0.49	missense	RBP2b	PVP01_0800700	242S>242T	35112G>C+35113C>G	
PvP01_08_v1	35249	A	G	16	49	58	0.22	missense	RBP2b	PVP01_0800700	288K>288E	35249A>G	
PvP01_08_v1	35250	A	C	4	50	69	0.05	missense	RBP2b	PVP01_0800700	288K>288T	35250A>C	
PvP01_08_v1	35264	T	G	4	48	71	0.05	missense	RBP2b	PVP01_0800700	293L>293V	35264T>G	
PvP01_08_v1	35285	A	G	2	48	73	0.03	missense	RBP2b	PVP01_0800700	300N>300D	35285A>G	
PvP01_08_v1	35287	T	A	4	49	70	0.05	missense	RBP2b	PVP01_0800700	300N>300K	35287T>A	
PvP01_08_v1	35474	G	A	27	53	43	0.39	missense	RBP2b	PVP01_0800700	363E>363K	35474G>A	
PvP01_08_v1	35484	T	A	28	53	42	0.4	missense	RBP2b	PVP01_0800700	366V>366D	35484T>A+35485T>C	
PvP01_08_v1	35484	T	A	28	53	42	0.4	missense	RBP2b	PVP01_0800700	366V>366H	35483G>C+35484T>A+35485T>C	
PvP01_08_v1	35485	T	C	28	53	42	0.4	missense	RBP2b	PVP01_0800700	366V>366D	35484T>A+35485T>C	
PvP01_08_v1	35485	T	C	28	53	42	0.4	missense	RBP2b	PVP01_0800700	366V>366H	35483G>C+35484T>A+35485T>C	
PvP01_08_v1	35532	G	A	18	40	65	0.22	missense	RBP2b	PVP01_0800700	382G>382E	35532G>A	
PvP01_08_v1	35571	T	C	2	41	80	0.02	missense	RBP2b	PVP01_0800700	395V>395A	35571T>C	
PvP01_08_v1	35623	T	A	17	45	61	0.22	missense	RBP2b	PVP01_0800700	412N>412K	35623T>A	
PvP01_08_v1	35972	G	A	1	38	84	0.01	missense	RBP2b	PVP01_0800700	529V>529M	35972G>A	
PvP01_08_v1	36061	A	C	71	50	2	0.97	missense	RBP2b	PVP01_0800700	558E>558D	36061A>C	
PvP01_08_v1	36072	G	T	1	35	87	0.01	missense	RBP2b	PVP01_0800700	562S>562I	36072G>T	
PvP01_08_v1	36078	G	A	50	46	27	0.65	missense	RBP2b	PVP01_0800700	564R>564Q	36078G>A	
PvP01_08_v1	36110	A	G	20	39	64	0.24	missense	RBP2b	PVP01_0800700	575K>575E	36110A>G	
PvP01_08_v1	36145	C	A	14	39	70	0.17	missense	RBP2b	PVP01_0800700	586S>586R	36145C>A	
PvP01_08_v1	36160	T	A	3	39	81	0.04	missense	RBP2b	PVP01_0800700	591N>591K	36160T>A	
PvP01_08_v1	36278	G	A	3	41	79	0.04	missense	RBP2b	PVP01_0800700	631E>631K	36278G>A	
PvP01_08_v1	37138	A	C	11	35	77	0.12	missense	RBP2b	PVP01_0800700	917E>917D	37138A>C	
PvP01_08_v1	37294	C	G	2	39	82	0.02	synonymous	RBP2b	PVP01_0800700	969S	37294C>G	
PvP01_08_v1	37360	G	A	1	41	81	0.01	missense	RBP2b	PVP01_0800700	991M>991I	37360G>A	
PvP01_08_v1	38103	G	A	80	37	6	0.93	missense	RBP2b	PVP01_0800700	1239G>1239E	38103G>A	
PvP01_08_v1	38942	G	A	1	26	96	0.01	missense	RBP2b	PVP01_0800700	1519E>1519K	38942G>A	
PvP01_08_v1	38945	C	T	10	28	85	0.11	missense	RBP2b	PVP01_0800700	1520H>1520Y	38945C>T	
PvP01_08_v1	38974	A	C	24	31	68	0.26	missense	RBP2b	PVP01_0800700	1529K>1529N	38974A>C	
PvP01_08_v1	39175	A	G	75	48	0	1	synonymous	RBP2b	PVP01_0800700	1596L	39175A>G	
PvP01_08_v1	39319	T	C	18	31	74	0.2	synonymous	RBP2b	PVP01_0800700	1644T	39319T>C	
PvP01_08_v1	39730	C	G	5	25	93	0.05	missense	RBP2b	PVP01_0800700	1781F>1781L	39730C>G	
PvP01_08_v1	39735	C	G	3	26	94	0.03	missense	RBP2b	PVP01_0800700	1783T>1783S	39735C>G	114697T>C+114699A>T
PvP01_08_v1	39793	C	T	78	45	0	1	synonymous	RBP2b	PVP01_0800700	1802P	39793C>T	
PvP01_08_v1	41128	G	A	23	24	76	0.23	synonymous	RBP2b	PVP01_0800700	2247Q	41128G>A	
PvP01_08_v1	41135	G	A	52	24	47	0.53	missense	RBP2b	PVP01_0800700	2250A>2250T	41135G>A	
PvP01_08_v1	41180	A	G	75	40	8	0.9	missense	RBP2b	PVP01_0800700	2265K>2265E	41180A>G	
PvP01_08_v1	41248	A	G	4	26	93	0.04	synonymous	RBP2b	PVP01_0800700	2287E	41248A>G	
PvP01_08_v1	41559	A	C	89	34	0	1	missense	RBP2b	PVP01_0800700	2391K>2391T	41559A>C	
PvP01_08_v1	41782	G	A	12	25	86	0.12	synonymous	RBP2b	PVP01_0800700	2465L	41782G>A	
PvP01_08_v1	42314	G	A	1	27	95	0.01	missense	RBP2b	PVP01_0800700	2643E>2643K	42314G>A	
PvP01_08_v1	42599	G	A	96	27	0	1	missense	RBP2b	PVP01_0800700	2738V>2738I	42599G>A	
PvP01_08_v1	42624	G	A	83	27	13	0.86	missense	RBP2b	PVP01_0800700	2746G>2746E	42624G>A	
PvP01_08_v1	42627	A	G	94	29	0	1	missense	RBP2b	PVP01_0800700	2747E>2747G	42627A>G	
PvP01_08_v1	42637	T	G	93	30	0	1	missense	RBP2b	PVP01_0800700	2750F>2750L	42637T>G	
PvP01_08_v1	42637	T	G	93	30	0	1	missense	RBP2b	PVP01_0800700	2750F>2750M	42635T>A+42637T>G	
PvP01_08_v1	42672	C	T	1	16	106	0.01	missense	RBP2b	PVP01_0800700	2762S>2762L	42672C>T	
PvP01_08_v1	734654	G	C	1	9	113	0.01	synonymous	RIPR	PVP01_0816800	11L	734654G>C	

PvP01_08_v1	734677	G	A	5	10	108	0.04	missense	RIPR	PVP01_0816800	19G>19D	734677G>A
PvP01_08_v1	734678	C	A	2	9	112	0.02	synonymous	RIPR	PVP01_0816800	19G	734678C>A
PvP01_08_v1	734777	G	C	2	8	113	0.02	missense	RIPR	PVP01_0816800	52K>52N	734777G>C
PvP01_08_v1	734784	T	G	107	16	0	1	missense	RIPR	PVP01_0816800	55L>55V	734784T>G
PvP01_08_v1	734863	C	A	8	5	110	0.07	missense	RIPR	PVP01_0816800	81A>81E	734863C>A
PvP01_08_v1	734937	A	G	2	13	108	0.02	missense	RIPR	PVP01_0816800	106K>106E	734937A>G
PvP01_08_v1	734944	C	T	107	16	0	1	missense	RIPR	PVP01_0816800	108P>108L	734944C>T
PvP01_08_v1	735135	A	C	104	10	9	0.92	missense	RIPR	PVP01_0816800	172I>172L	735135A>C
PvP01_08_v1	735636	A	G	57	6	60	0.49	missense	RIPR	PVP01_0816800	339K>339E	735636A>G
PvP01_08_v1	735639	G	C	1	4	118	0.01	missense	RIPR	PVP01_0816800	340D>340H	735639G>C
PvP01_08_v1	735690	G	A	6	6	111	0.05	missense	RIPR	PVP01_0816800	357E>357K	735690G>A
PvP01_08_v1	735692	G	C	19	5	99	0.16	missense	RIPR	PVP01_0816800	357E>357D	735692G>C
PvP01_08_v1	735708	T	C	109	14	0	1	missense	RIPR	PVP01_0816800	363Y>363H	735708T>C
PvP01_08_v1	735717	A	G	107	14	2	0.98	missense	RIPR	PVP01_0816800	366K>366E	735717A>G
PvP01_08_v1	735733	C	A	21	8	94	0.18	missense	RIPR	PVP01_0816800	371A>371E	735733C>A
PvP01_08_v1	736469	C	G	26	5	92	0.22	missense	RIPR	PVP01_0816800	616N>616K	736469C>G
PvP01_08_v1	736736	A	C	2	3	118	0.02	missense	RIPR	PVP01_0816800	705Q>705H	736736A>C
PvP01_08_v1	736797	C	G	29	7	87	0.25	missense	RIPR	PVP01_0816800	726L>726V	736797C>G
PvP01_08_v1	737001	G	A	113	9	1	0.99	missense	RIPR	PVP01_0816800	794D>794N	737001G>A
PvP01_08_v1	737332	G	T	7	8	108	0.06	missense	RIPR	PVP01_0816800	904G>904V	737332G>T
PvP01_08_v1	1509613	A	T	1	7	115	0.01	synonymous	CSP	PVP01_0835600	300P	1509613A>T
PvP01_09_v1	1459310	C	T	1	0	122	0.01	synonymous	AMA1	PVP01_0934200	5Y	1459310C>T
PvP01_09_v1	1459370	T	G	119	4	0	1	missense	AMA1	PVP01_0934200	25H>25K	1459368C>A+1459370T>G
PvP01_09_v1	1459370	T	G	119	4	0	1	missense	AMA1	PVP01_0934200	25H>25Q	1459370T>G
PvP01_09_v1	1459420	G	T	14	3	106	0.12	missense	AMA1	PVP01_0934200	42G>42V	1459420G>T
PvP01_09_v1	1459425	A	T	8	4	111	0.07	missense	AMA1	PVP01_0934200	44T>44S	1459425A>T
PvP01_09_v1	1459492	G	A	34	4	85	0.29	missense	AMA1	PVP01_0934200	66R>66K	1459492G>A
PvP01_09_v1	1459601	C	T	60	3	60	0.5	synonymous	AMA1	PVP01_0934200	102V	1459601C>T
PvP01_09_v1	1459631	C	A	58	7	58	0.5	synonymous	AMA1	PVP01_0934200	112T	1459631C>A
PvP01_09_v1	1459654	G	A	22	5	96	0.19	missense	AMA1	PVP01_0934200	120R>120K	1459654G>A
PvP01_09_v1	1459689	G	A	82	7	34	0.71	missense	AMA1	PVP01_0934200	132D>132N	1459689G>A
PvP01_09_v1	1459713	T	A	49	9	65	0.43	missense	AMA1	PVP01_0934200	140L>140I	1459713T>A
PvP01_09_v1	1459717	C	A	47	8	68	0.41	missense	AMA1	PVP01_0934200	141A>141E	1459717C>A
PvP01_09_v1	1459729	A	C	95	8	20	0.83	missense	AMA1	PVP01_0934200	145E>145A	1459729A>C
PvP01_09_v1	1459862	A	T	15	1	107	0.12	missense	AMA1	PVP01_0934200	189E>189D	1459862A>T
PvP01_09_v1	1459872	C	T	5	0	118	0.04	missense	AMA1	PVP01_0934200	193H>193Y	1459872C>T
PvP01_09_v1	1459923	T	C	111	8	4	0.97	missense	AMA1	PVP01_0934200	210S>210P	1459923T>C
PvP01_09_v1	1459975	T	A	80	8	35	0.7	missense	AMA1	PVP01_0934200	227V>227E	1459975T>A
PvP01_09_v1	1459977	G	A	87	8	28	0.76	missense	AMA1	PVP01_0934200	228D>228N	1459977G>A
PvP01_09_v1	1460036	T	C	116	7	0	1	synonymous	AMA1	PVP01_0934200	247C	1460036T>C
PvP01_09_v1	1460124	A	G	13	3	107	0.11	missense	AMA1	PVP01_0934200	277K>277E	1460124A>G
PvP01_09_v1	1460136	G	C	1	3	119	0.01	missense	AMA1	PVP01_0934200	281E>281Q	1460136G>C
PvP01_09_v1	1460158	A	G	111	10	2	0.98	missense	AMA1	PVP01_0934200	288E>288G	1460158A>G
PvP01_09_v1	1460268	C	T	9	2	112	0.07	missense	AMA1	PVP01_0934200	325L>325F	1460268C>T
PvP01_09_v1	1460363	T	C	7	1	115	0.06	synonymous	AMA1	PVP01_0934200	356I	1460363T>C
PvP01_09_v1	1460398	A	T	14	3	106	0.12	missense	AMA1	PVP01_0934200	368K>368I	1460398A>T
PvP01_09_v1	1460433	C	A	20	2	101	0.17	missense	AMA1	PVP01_0934200	380Q>380K	1460433C>A
PvP01_09_v1	1460433	C	A	20	2	101	0.17	missense	AMA1	PVP01_0934200	380Q>380R	1460433C>A+1460434A>G
PvP01_09_v1	1460440	T	A	5	4	114	0.04	missense	AMA1	PVP01_0934200	382V>382E	1460440T>A+1460441A>G
PvP01_09_v1	1460441	A	G	5	4	114	0.04	missense	AMA1	PVP01_0934200	382V>382E	1460440T>A+1460441A>G

PvP01_09_v1	1460450	T	G	117	6	0	1	missense	AMA1	PVP01_0934200	385D>385E	1460450T>G
PvP01_09_v1	1460608	A	G	99	3	21	0.82	missense	AMA1	PVP01_0934200	438H>438R	1460608A>G
PvP01_09_v1	1460628	A	G	6	2	115	0.05	missense	AMA1	PVP01_0934200	445N>445D	1460628A>G
PvP01_09_v1	1460818	A	G	119	4	0	1	missense	AMA1	PVP01_0934200	508K>508R	1460818A>G
PvP01_10_v1	254365	C	T	118	5	0	1	synonymous	PVP01_1005500	PVP01_1005500	104H	254365C>T
PvP01_10_v1	254401	T	C	105	7	11	0.91	synonymous	PVP01_1005500	PVP01_1005500	116D	254401T>C
PvP01_10_v1	254409	T	G	103	9	11	0.9	missense	PVP01_1005500	PVP01_1005500	119F>119C	254409T>G
PvP01_10_v1	254411	T	C	103	9	11	0.9	missense	PVP01_1005500	PVP01_1005500	120S>120L	254411T>C+254412C>T
PvP01_10_v1	254412	C	T	114	9	0	1	missense	PVP01_1005500	PVP01_1005500	120S>120F	254412C>T
PvP01_10_v1	254412	C	T	114	9	0	1	missense	PVP01_1005500	PVP01_1005500	120S>120L	254411T>C+254412C>T
PvP01_10_v1	254515	T	A	40	2	81	0.33	synonymous	PVP01_1005500	PVP01_1005500	154T	254515T>A
PvP01_10_v1	255415	G	A	42	2	79	0.35	synonymous	PVP01_1005500	PVP01_1005500	454Q	255415G>A
PvP01_10_v1	255640	T	C	115	8	0	1	synonymous	PVP01_1005500	PVP01_1005500	529D	255640T>C
PvP01_10_v1	314831	G	T	2	13	108	0.02	synonymous	PVP01_1007200	PVP01_1007200	2143P	314831G>T
PvP01_10_v1	314834	G	A	5	15	103	0.05	synonymous	PVP01_1007200	PVP01_1007200	2142S	314834G>A
PvP01_10_v1	314835	C	T	58	24	41	0.59	missense	PVP01_1007200	PVP01_1007200	2142S>2142N	314835C>T
PvP01_10_v1	314868	C	A	1	13	109	0.01	missense	PVP01_1007200	PVP01_1007200	2131G>2131V	314868C>A
PvP01_10_v1	314938	G	A	11	17	95	0.1	missense	PVP01_1007200	PVP01_1007200	2108L>2108F	314938G>A
PvP01_10_v1	314939	G	A	11	17	95	0.1	synonymous	PVP01_1007200	PVP01_1007200	2107D	314939G>A
PvP01_10_v1	315183	T	C	28	18	77	0.27	missense	PVP01_1007200	PVP01_1007200	2026K>2026R	315183T>C
PvP01_10_v1	315430	A	G	5	17	101	0.05	*missense	PVP01_1007200	PVP01_1007200	1944Y>1944H	315430A>G
PvP01_10_v1	315719	G	A	31	21	71	0.3	synonymous	PVP01_1007200	PVP01_1007200	1847F	315719G>A
PvP01_10_v1	315724	C	T	1	19	103	0.01	missense	PVP01_1007200	PVP01_1007200	1846G>1846R	315724C>T
PvP01_10_v1	316239	A	C	3	19	101	0.03	missense	PVP01_1007200	PVP01_1007200	1674F>1674C	316239A>C
PvP01_10_v1	316461	G	A	27	27	69	0.28	missense	PVP01_1007200	PVP01_1007200	1600A>1600V	316461G>A
PvP01_10_v1	316517	G	A	25	22	76	0.25	synonymous	PVP01_1007200	PVP01_1007200	1581D	316517G>A
PvP01_10_v1	316576	G	A	2	15	106	0.02	missense	PVP01_1007200	PVP01_1007200	1562L>1562F	316576G>A
PvP01_10_v1	316742	C	T	1	16	106	0.01	synonymous	PVP01_1007200	PVP01_1007200	1506Q	316742C>T
PvP01_10_v1	316886	C	T	9	15	99	0.08	synonymous	PVP01_1007200	PVP01_1007200	1458L	316886C>T
PvP01_10_v1	317190	T	C	1	16	106	0.01	missense	PVP01_1007200	PVP01_1007200	1357E>1357G	317190T>C
PvP01_10_v1	317420	C	T	3	16	104	0.03	synonymous	PVP01_1007200	PVP01_1007200	1280L	317420C>T
PvP01_10_v1	317628	T	G	44	28	51	0.46	missense	PVP01_1007200	PVP01_1007200	1211K>1211T	317628T>G
PvP01_10_v1	318199	C	T	23	18	82	0.22	missense	PVP01_1007200	PVP01_1007200	1021A>1021T	318199C>T
PvP01_10_v1	318227	G	A	44	27	52	0.46	synonymous	PVP01_1007200	PVP01_1007200	1011S	318227G>A
PvP01_10_v1	318235	T	C	20	25	78	0.2	missense	PVP01_1007200	PVP01_1007200	1009T>1009A	318235T>C
PvP01_10_v1	318277	A	T	69	44	10	0.87	missense	PVP01_1007200	PVP01_1007200	995C>995S	318277A>T
PvP01_10_v1	318512	C	A	1	39	83	0.01	missense	PVP01_1007200	PVP01_1007200	916K>916N	318512C>A
PvP01_10_v1	318549	T	C	16	37	70	0.19	missense	PVP01_1007200	PVP01_1007200	904N>904S	318549T>C
PvP01_10_v1	318628	C	T	39	27	57	0.41	missense	PVP01_1007200	PVP01_1007200	878G>878R	318628C>T
PvP01_10_v1	318662	C	T	24	21	78	0.24	synonymous	PVP01_1007200	PVP01_1007200	866E	318662C>T
PvP01_10_v1	318772	G	C	61	34	28	0.69	missense	PVP01_1007200	PVP01_1007200	830Q>830E	318772G>C
PvP01_10_v1	318970	C	A	9	31	83	0.1	missense	PVP01_1007200	PVP01_1007200	764D>764Y	318970C>A
PvP01_10_v1	319327	C	A	3	26	94	0.03	missense	PVP01_1007200	PVP01_1007200	645G>645C	319327C>A
PvP01_10_v1	319423	A	G	85	38	0	1	missense	PVP01_1007200	PVP01_1007200	613S>613P	319423A>G
PvP01_10_v1	319434	A	C	1	20	102	0.01	missense	PVP01_1007200	PVP01_1007200	609L>609W	319434A>C
PvP01_10_v1	320094	C	T	3	26	94	0.03	missense	PVP01_1007200	PVP01_1007200	389G>389D	320094C>T
PvP01_10_v1	320141	T	G	85	37	1	0.99	missense	PVP01_1007200	PVP01_1007200	373K>373N	320141T>G
PvP01_10_v1	320458	C	A	1	30	92	0.01	missense	PVP01_1007200	PVP01_1007200	268D>268Y	320458C>A
PvP01_10_v1	320490	G	C	24	27	72	0.25	missense	PVP01_1007200	PVP01_1007200	257S>257W	320490G>C
PvP01_10_v1	320638	T	C	79	44	0	1	missense	PVP01_1007200	PVP01_1007200	208K>208E	320638T>C

PvP01_10_v1	320966	G	A	1	32	90	0.01	synonymous	PVP01_1007200	PVP01_1007200	98S	320966G>A
PvP01_10_v1	321144	G	A	6	25	92	0.06	missense	PVP01_1007200	PVP01_1007200	39P>39L	321144G>A
PvP01_10_v1	479069	T	C	4	6	113	0.03	synonymous	MDR1	PVP01_1010900	1355K	479069T>C
PvP01_10_v1	479133	G	T	1	7	115	0.01	missense	MDR1	PVP01_1010900	1334A>1334E	479133G>T
PvP01_10_v1	480070	G	A	2	9	112	0.02	synonymous	MDR1	PVP01_1010900	1022L	480070G>A
PvP01_10_v1	480261	A	G	22	10	91	0.19	missense	MDR1	PVP01_1010900	958M>958T	480261A>G
PvP01_10_v1	480412	G	T	3	7	113	0.03	missense	MDR1	PVP01_1010900	908L>908M	480412G>T
PvP01_10_v1	480552	G	T	2	5	116	0.02	missense	MDR1	PVP01_1010900	861A>861E	480552G>T
PvP01_10_v1	481042	T	C	103	20	0	1	missense	MDR1	PVP01_1010900	698S>698G	481042T>C
PvP01_10_v1	481092	T	C	1	6	116	0.01	missense	MDR1	PVP01_1010900	681H>681R	481092T>C
PvP01_10_v1	481340	T	G	106	17	0	1	synonymous	MDR1	PVP01_1010900	598P	481340T>G
PvP01_10_v1	481547	C	T	18	7	98	0.16	synonymous	MDR1	PVP01_1010900	529T	481547C>T
PvP01_10_v1	481636	C	T	83	8	32	0.72	missense	MDR1	PVP01_1010900	500D>500N	481636C>T
PvP01_10_v1	482054	C	T	4	12	107	0.04	synonymous	MDR1	PVP01_1010900	360E	482054C>T
PvP01_10_v1	826522	A	G	5	5	113	0.04	*synonymous	PVP01_1018600	PVP01_1018600	1655S	826522A>G
PvP01_10_v1	826828	G	C	15	4	104	0.13	synonymous	PVP01_1018600	PVP01_1018600	1553A	826828G>C
PvP01_10_v1	827121	C	T	49	10	64	0.43	*missense	PVP01_1018600	PVP01_1018600	1456E>1456K	827121C>T
PvP01_10_v1	827999	C	T	3	16	104	0.03	*missense	PVP01_1018600	PVP01_1018600	1163G>1163E	827999C>T
PvP01_10_v1	828001	C	T	3	14	106	0.03	*synonymous	PVP01_1018600	PVP01_1018600	1162G	828001C>T
PvP01_10_v1	828322	T	C	111	12	0	1	*synonymous	PVP01_1018600	PVP01_1018600	1055G	828322T>C
PvP01_10_v1	828751	T	G	6	8	109	0.05	synonymous	PVP01_1018600	PVP01_1018600	912P	828751T>G
PvP01_10_v1	829459	G	T	6	11	106	0.05	stop_gained	PVP01_1018600	PVP01_1018600	676Y>676*	829459G>T
PvP01_10_v1	829517	C	T	4	15	104	0.04	missense	PVP01_1018600	PVP01_1018600	657R>657K	829517C>T
PvP01_10_v1	829671	T	C	5	11	107	0.04	missense	PVP01_1018600	PVP01_1018600	606I>606V	829671T>C
PvP01_10_v1	829733	T	C	1	15	107	0.01	missense	PVP01_1018600	PVP01_1018600	585E>585G	829733T>C
PvP01_10_v1	830191	C	T	63	8	52	0.55	synonymous	PVP01_1018600	PVP01_1018600	432Q	830191C>T
PvP01_10_v1	830383	C	A	7	11	105	0.06	missense	PVP01_1018600	PVP01_1018600	368M>368I	830383C>A
PvP01_10_v1	830910	G	A	85	19	19	0.82	missense	PVP01_1018600	PVP01_1018600	193P>193S	830910G>A
PvP01_10_v1	1053307	C	G	107	16	0	1	synonymous	PI4K	PVP01_1024200	1511S	1053307C>G
PvP01_10_v1	1054750	T	A	114	9	0	1	missense	PI4K	PVP01_1024200	1240N>1240Y	1054750T>A
PvP01_10_v1	1055342	G	T	36	7	80	0.31	synonymous	PI4K	PVP01_1024200	1042I	1055342G>T
PvP01_10_v1	1056253	T	C	118	5	0	1	missense	PI4K	PVP01_1024200	739K>739E	1056253T>C
PvP01_10_v1	1056500	G	A	61	1	61	0.5	synonymous	PI4K	PVP01_1024200	656I	1056500G>A
PvP01_10_v1	1057472	G	T	7	1	115	0.06	synonymous	PI4K	PVP01_1024200	332R	1057472G>T
PvP01_11_v1	1256157	A	G	101	22	0	1	synonymous	MSP10	PVP01_1129100	145A	1256157A>G
PvP01_11_v1	1256312	G	A	106	17	0	1	missense	MSP10	PVP01_1129100	197G>197D	1256312G>A
PvP01_11_v1	1256708	G	T	107	16	0	1	missense	MSP10	PVP01_1129100	329R>329I	1256708G>T
PvP01_11_v1	1256938	C	T	108	15	0	1	missense	MSP10	PVP01_1129100	406H>406Y	1256938C>T
PvP01_11_v1	1395023	T	C	1	1	121	0.01	synonymous	TLP	PVP01_1132600	1374G	1395023T>C
PvP01_11_v1	1395450	C	A	115	6	2	0.98	missense	TLP	PVP01_1132600	1232G>1232V	1395450C>A
PvP01_11_v1	1395492	C	A	2	2	119	0.02	missense	TLP	PVP01_1132600	1218G>1218V	1395492C>A
PvP01_11_v1	1395711	G	C	64	3	56	0.53	missense	TLP	PVP01_1132600	1145S>1145C	1395711G>C
PvP01_11_v1	1396097	C	T	4	1	118	0.03	synonymous	TLP	PVP01_1132600	1016E	1396097C>T
PvP01_11_v1	1396203	A	C	101	21	1	0.99	missense	TLP	PVP01_1132600	981I>981R	1396203A>C
PvP01_11_v1	1396312	T	C	111	12	0	1	missense	TLP	PVP01_1132600	945M>945V	1396312T>C
PvP01_11_v1	1396438	A	C	117	6	0	1	missense	TLP	PVP01_1132600	903C>903G	1396438A>C
PvP01_11_v1	1397781	A	G	108	15	0	1	missense	TLP	PVP01_1132600	455V>455A	1397781A>G
PvP01_11_v1	1397781	A	G	108	15	0	1	missense	TLP	PVP01_1132600	455V>455T	1397781A>G+1397782C>T
PvP01_11_v1	1397842	T	C	107	16	0	1	missense	TLP	PVP01_1132600	435R>435G	1397842T>C
PvP01_11_v1	1398656	G	A	108	15	0	1	synonymous	TLP	PVP01_1132600	163H	1398656G>A

PvP01_11_v1	1949796	C	T	110	11	2	0.98	missense	DHODH	PVP01_1145600	22T>22I	1949796C>T
PvP01_11_v1	1949812	T	C	113	10	0	1	synonymous	DHODH	PVP01_1145600	27R	1949812T>C
PvP01_11_v1	1950244	T	C	1	3	119	0.01	synonymous	DHODH	PVP01_1145600	171D	1950244T>C
PvP01_11_v1	1950778	T	C	100	14	9	0.92	synonymous	DHODH	PVP01_1145600	349N	1950778T>C
PvP01_11_v1	1950994	G	C	2	5	116	0.02	synonymous	DHODH	PVP01_1145600	421T	1950994G>C
PvP01_11_v1	1951178	C	T	86	7	30	0.74	synonymous	DHODH	PVP01_1145600	483L	1951178C>T
PvP01_11_v1	1951264	C	T	1	2	120	0.01	synonymous	DHODH	PVP01_1145600	511F	1951264C>T
PvP01_11_v1	1951420	C	T	1	1	121	0.01	synonymous	DHODH	PVP01_1145600	563A	1951420C>T
PvP01_12_v1	323599	A	C	105	7	11	0.91	missense	P47	PVP01_1208000	22L>22F	323599A>C
PvP01_12_v1	323603	C	T	115	8	0	1	missense	P47	PVP01_1208000	24L>24F	323603C>T
PvP01_12_v1	323611	A	G	115	8	0	1	synonymous	P47	PVP01_1208000	26T	323611A>G
PvP01_12_v1	323612	G	A	115	8	0	1	missense	P47	PVP01_1208000	27E>27K	323612G>A
PvP01_12_v1	323618	T	A	13	6	104	0.11	missense	P47	PVP01_1208000	29L>29I	323618T>A
PvP01_12_v1	323689	A	G	2	5	116	0.02	synonymous	P47	PVP01_1208000	52E	323689A>G
PvP01_12_v1	323918	G	T	11	7	105	0.09	missense	P47	PVP01_1208000	129V>129L	323918G>T
PvP01_12_v1	324114	A	G	12	6	105	0.1	missense	P47	PVP01_1208000	194D>194G	324114A>G
PvP01_12_v1	324352	G	A	114	9	0	1	missense	P47	PVP01_1208000	273M>273I	324352G>A
PvP01_12_v1	324352	G	A	114	9	0	1	missense	P47	PVP01_1208000	273M>273V	324350A>G+324352G>A
PvP01_12_v1	324650	G	A	1	2	120	0.01	missense	P47	PVP01_1208000	373V>373T	324650G>A+324651T>C
PvP01_12_v1	324651	T	C	118	5	0	1	missense	P47	PVP01_1208000	373V>373A	324651T>C
PvP01_12_v1	324651	T	C	118	5	0	1	missense	P47	PVP01_1208000	373V>373T	324650G>A+324651T>C
PvP01_12_v1	326708	A	G	116	7	0	1	missense	P48/45	PVP01_1208100	35K>35E	326708A>G
PvP01_12_v1	326794	A	G	116	7	0	1	synonymous	P48/45	PVP01_1208100	63L	326794A>G
PvP01_12_v1	327052	A	T	1	4	118	0.01	synonymous	P48/45	PVP01_1208100	149T	327052A>T
PvP01_12_v1	327073	C	T	1	3	119	0.01	synonymous	P48/45	PVP01_1208100	156C	327073C>T
PvP01_12_v1	327124	G	C	113	10	0	1	synonymous	P48/45	PVP01_1208100	173A	327124G>C
PvP01_12_v1	327236	A	C	111	8	4	0.97	missense	P48/45	PVP01_1208100	211N>211H	327236A>C
PvP01_12_v1	327349	A	G	114	8	1	0.99	synonymous	P48/45	PVP01_1208100	248K	327349A>G
PvP01_12_v1	327355	C	A	113	8	2	0.98	missense	P48/45	PVP01_1208100	250N>250K	327355C>A
PvP01_12_v1	327404	G	T	8	3	112	0.07	missense	P48/45	PVP01_1208100	267V>267L	327404G>T
PvP01_12_v1	327608	T	G	112	11	0	1	missense	P48/45	PVP01_1208100	335Y>335D	327608T>G
PvP01_12_v1	327662	G	C	4	2	117	0.03	missense	P48/45	PVP01_1208100	353E>353Q	327662G>C
PvP01_12_v1	327731	A	G	116	7	0	1	missense	P48/45	PVP01_1208100	376T>376A	327731A>G
PvP01_12_v1	327858	G	A	116	7	0	1	missense	P48/45	PVP01_1208100	418R>418K	327858G>A
PvP01_12_v1	485399	G	T	9	4	110	0.08	synonymous	K13	PVP01_1211100	205V	485399G>T
PvP01_12_v1	769132	C	T	12	1	110	0.1	missense	TRAP	PVP01_1218700	58P>58S	769132C>T
PvP01_12_v1	769318	T	A	80	3	40	0.67	missense	TRAP	PVP01_1218700	120S>120T	769318T>A
PvP01_12_v1	769324	T	A	26	2	95	0.21	missense	TRAP	PVP01_1218700	122S>122T	769324T>A
PvP01_12_v1	769336	A	T	45	2	76	0.37	missense	TRAP	PVP01_1218700	126T>126S	769336A>T
PvP01_12_v1	769438	A	G	121	2	0	1	missense	TRAP	PVP01_1218700	160I>160V	769438A>G
PvP01_12_v1	769457	C	G	117	5	1	0.99	missense	TRAP	PVP01_1218700	166T>166R	769457C>G
PvP01_12_v1	769476	A	T	80	3	40	0.67	missense	TRAP	PVP01_1218700	172K>172N	769476A>T
PvP01_12_v1	769516	A	G	78	5	40	0.66	missense	TRAP	PVP01_1218700	186I>186V	769516A>G
PvP01_12_v1	769675	A	G	2	0	121	0.02	missense	TRAP	PVP01_1218700	239N>239D	769675A>G
PvP01_12_v1	769734	C	A	44	0	79	0.36	missense	TRAP	PVP01_1218700	258H>258Q	769734C>A
PvP01_12_v1	769762	G	A	27	3	93	0.22	missense	TRAP	PVP01_1218700	268G>268K	769762G>A+769763G>A
PvP01_12_v1	769763	G	A	27	4	92	0.23	missense	TRAP	PVP01_1218700	268G>268K	769762G>A+769763G>A
PvP01_12_v1	769966	A	G	80	3	40	0.67	missense	TRAP	PVP01_1218700	336R>336G	769966A>G
PvP01_12_v1	769968	G	T	1	1	121	0.01	missense	TRAP	PVP01_1218700	336R>336S	769968G>T
PvP01_12_v1	770010	A	C	118	4	1	0.99	synonymous	TRAP	PVP01_1218700	350A	770010A>C

PvP01_12_v1	770190	A	C	111	5	7	0.94	missense	TRAP	PVP01_1218700	410K>410N	770190A>C
PvP01_12_v1	770351	C	A	122	1	0	1	missense	TRAP	PVP01_1218700	464A>464E	770351C>A
PvP01_12_v1	770517	T	C	117	4	2	0.98	synonymous	TRAP	PVP01_1218700	519G	770517T>C
PvP01_12_v1	770605	C	A	8	8	107	0.07	missense	TRAP	PVP01_1218700	549L>549M	770605C>A
PvP01_12_v1	1185755	C	T	57	24	42	0.58	synonymous	LDH	PVP01_1229700	8V	1185755C>T
PvP01_12_v1	1187665	C	T	30	13	80	0.27	synonymous	DHFS-FPGS	PVP01_1229800	42N	1187665C>T
PvP01_12_v1	1188211	T	C	111	12	0	1	synonymous	DHFS-FPGS	PVP01_1229800	224V	1188211T>C
PvP01_12_v1	1188838	C	T	10	9	104	0.09	synonymous	DHFS-FPGS	PVP01_1229800	373A	1188838C>T
PvP01_12_v1	2393419	C	A	19	1	103	0.16	missense	GEST	PVP01_1258000	16L>16M	2393419C>A
PvP01_12_v1	2393675	A	C	29	0	94	0.24	splice_region	GEST	PVP01_1258000		
PvP01_12_v1	2393734	A	G	12	0	111	0.1	missense	GEST	PVP01_1258000	53D>53G	2393734A>G
PvP01_12_v1	2394350	C	T	1	1	121	0.01	synonymous	GEST	PVP01_1258000	202D	2394350C>T
PvP01_12_v1	2394371	T	G	94	2	27	0.78	missense	GEST	PVP01_1258000	209D>209E	2394371T>G
PvP01_12_v1	2441608	G	C	109	14	0	1	missense	MDR2	PVP01_1259100	43V>43L	2441608G>C
PvP01_12_v1	2441670	A	T	3	5	115	0.03	missense	MDR2	PVP01_1259100	63R>63S	2441670A>T
PvP01_12_v1	2441682	G	A	17	5	101	0.14	synonymous	MDR2	PVP01_1259100	67S	2441682G>A
PvP01_12_v1	2442441	G	A	12	4	107	0.1	synonymous	MDR2	PVP01_1259100	320A	2442441G>A
PvP01_12_v1	2442447	T	C	52	8	63	0.45	synonymous	MDR2	PVP01_1259100	322S	2442447T>C
PvP01_12_v1	2442764	T	C	3	3	117	0.03	missense	MDR2	PVP01_1259100	428L>428S	2442764T>C
PvP01_12_v1	2442885	A	T	99	9	15	0.87	synonymous	MDR2	PVP01_1259100	468V	2442885A>T
PvP01_12_v1	2443022	A	T	26	4	93	0.22	missense	MDR2	PVP01_1259100	514Y>514F	2443022A>T
PvP01_12_v1	2443419	C	A	12	4	107	0.1	synonymous	MDR2	PVP01_1259100	646L	2443419C>A
PvP01_12_v1	2443548	A	T	29	3	91	0.24	synonymous	MDR2	PVP01_1259100	689G	2443548A>T
PvP01_12_v1	2444739	C	T	2	3	118	0.02	synonymous	MDR2	PVP01_1259100	1086N	2444739C>T
PvP01_12_v1	2444949	C	T	1	4	118	0.01	synonymous	MDR2	PVP01_1259100	1156G	2444949C>T
PvP01_12_v1	2445106	G	A	6	2	115	0.05	missense	MDR2	PVP01_1259100	1209G>1209S	2445106G>A
PvP01_12_v1	2445136	G	A	1	3	119	0.01	missense	MDR2	PVP01_1259100	1219A>1219T	2445136G>A
PvP01_12_v1	2445238	G	A	7	2	114	0.06	missense	MDR2	PVP01_1259100	1253G>1253R	2445238G>A
PvP01_12_v1	2446077	C	T	2	3	118	0.02	synonymous	MDR2	PVP01_1259100	1532P	2446077C>T
PvP01_13_v1	333465	G	T	79	44	0	1	missense	CRMP3	PVP01_1307300	290R>290I	333465G>T
PvP01_13_v1	333558	A	G	2	25	96	0.02	missense	CRMP3	PVP01_1307300	321N>321S	333558A>G
PvP01_13_v1	333652	A	G	84	39	0	1	missense	CRMP3	PVP01_1307300	352I>352M	333652A>G
PvP01_13_v1	333815	T	G	79	43	1	0.99	missense	CRMP3	PVP01_1307300	407L>407V	333815T>G
PvP01_13_v1	333824	C	G	79	44	0	1	missense	CRMP3	PVP01_1307300	410L>410V	333824C>G
PvP01_13_v1	333932	T	C	1	24	98	0.01	missense	CRMP3	PVP01_1307300	446Y>446H	333932T>C
PvP01_13_v1	334192	A	T	86	37	0	1	synonymous	CRMP3	PVP01_1307300	532P	334192A>T
PvP01_13_v1	334265	C	T	4	23	96	0.04	missense	CRMP3	PVP01_1307300	557P>557S	334265C>T
PvP01_13_v1	334388	G	A	86	35	2	0.98	missense	CRMP3	PVP01_1307300	598E>598K	334388G>A
PvP01_13_v1	334388	G	A	86	35	2	0.98	missense	CRMP3	PVP01_1307300	598E>598M	334388G>A+334389A>T
PvP01_13_v1	334389	A	T	85	32	6	0.93	missense	CRMP3	PVP01_1307300	598E>598M	334388G>A+334389A>T
PvP01_13_v1	334460	A	G	91	32	0	1	missense	CRMP3	PVP01_1307300	622K>622E	334460A>G
PvP01_13_v1	335184	C	T	87	36	0	1	missense	CRMP3	PVP01_1307300	863S>863F	335184C>T
PvP01_13_v1	335187	T	C	88	35	0	1	missense	CRMP3	PVP01_1307300	864V>864A	335187T>C
PvP01_13_v1	335351	A	G	85	38	0	1	missense	CRMP3	PVP01_1307300	919I>919V	335351A>G
PvP01_13_v1	336032	C	A	81	42	0	1	missense	CRMP3	PVP01_1307300	1146H>1146N	336032C>A
PvP01_13_v1	336738	C	T	86	33	4	0.96	missense	CRMP3	PVP01_1307300	1381P>1381L	336738C>T
PvP01_13_v1	336883	A	T	2	24	97	0.02	missense	CRMP3	PVP01_1307300	1429K>1429N	336883A>T
PvP01_13_v1	336962	G	A	82	39	2	0.98	missense	CRMP3	PVP01_1307300	1456V>1456M	336962G>A
PvP01_13_v1	337161	G	A	2	18	103	0.02	missense	CRMP3	PVP01_1307300	1522G>1522D	337161G>A
PvP01_13_v1	337428	T	A	3	25	95	0.03	missense	CRMP3	PVP01_1307300	1611F>1611Y	337428T>A+337429T>C

PvP01_13_v1	337429	T	C	3	25	95	0.03	missense	CRMP3	PVP01_1307300	1611F>1611Y	337428T>A+337429T>C
PvP01_13_v1	337753	A	C	94	29	0	1	missense	CRMP3	PVP01_1307300	1719K>1719N	337753A>C
PvP01_13_v1	338148	A	C	91	32	0	1	missense	CRMP3	PVP01_1307300	1851K>1851T	338148A>C
PvP01_13_v1	339747	C	A	21	26	76	0.22	missense	CRMP3	PVP01_1307300	2384A>2384D	339747C>A
PvP01_13_v1	341423	C	A	11	14	98	0.1	missense	CRMP3	PVP01_1307300	2943P>2943T	341423C>A
PvP01_13_v1	342463	C	A	16	14	93	0.15	missense	CRMP3	PVP01_1307300	3289H>3289Q	342463C>A
PvP01_13_v1	516685	C	T	18	7	98	0.16	synonymous	ATP4	PVP01_1311100	1237G	516685C>T
PvP01_13_v1	516804	C	A	14	9	100	0.12	missense	ATP4	PVP01_1311100	1198D>1198Y	516804C>A
PvP01_13_v1	516924	C	T	6	9	108	0.05	missense	ATP4	PVP01_1311100	1158E>1158K	516924C>T
PvP01_13_v1	517034	G	T	31	5	87	0.26	missense	ATP4	PVP01_1311100	1121A>1121D	517034G>T
PvP01_13_v1	517064	C	T	32	7	84	0.28	missense	ATP4	PVP01_1311100	1111S>1111N	517064C>T
PvP01_13_v1	517172	A	T	4	4	115	0.03	missense	ATP4	PVP01_1311100	1075F>1075Y	517172A>T
PvP01_13_v1	517790	T	A	3	12	108	0.03	missense	ATP4	PVP01_1311100	869E>869V	517790T>A
PvP01_13_v1	519436	G	T	9	4	110	0.08	synonymous	ATP4	PVP01_1311100	320P	519436G>T
PvP01_13_v1	519722	T	C	1	4	118	0.01	missense	ATP4	PVP01_1311100	225E>225G	519722T>C
PvP01_13_v1	519866	T	A	48	13	62	0.44	missense	ATP4	PVP01_1311100	177E>177V	519866T>A
PvP01_13_v1	817378	G	A	2	3	118	0.02	synonymous	PVP01_1317400	PVP01_1317400	279T	817378G>A
PvP01_13_v1	818665	T	C	67	12	44	0.6	missense	PVP01_1317400	PVP01_1317400	39K>39E	818665T>C
PvP01_13_v1	1034368	A	C	117	6	0	1	missense	ABCG2	PVP01_1322800	124M>124L	1034368A>C
PvP01_13_v1	1034368	A	C	117	6	0	1	missense	ABCG2	PVP01_1322800	124M>124Q	1034368A>C+1034369T>A
PvP01_13_v1	1034451	C	T	5	0	118	0.04	synonymous	ABCG2	PVP01_1322800	151H	1034451C>T
PvP01_13_v1	1034884	G	A	119	4	0	1	missense	ABCG2	PVP01_1322800	296V>296I	1034884G>A
PvP01_13_v1	1035259	C	T	1	2	120	0.01	missense	ABCG2	PVP01_1322800	421L>421F	1035259C>T
PvP01_13_v1	1035368	A	T	117	6	0	1	missense	ABCG2	PVP01_1322800	457K>457M	1035368A>T
PvP01_13_v1	1035572	G	C	117	6	0	1	missense	ABCG2	PVP01_1322800	525S>525T	1035572G>C
PvP01_13_v1	1802108	T	C	20	11	92	0.18	missense	PM4	PVP01_1340900	165I>165V	1802108T>C
PvP01_14_v1	113615	T	C	1	24	98	0.01	missense	RBP2a	PVP01_1402400	38L>38P	113615T>C
PvP01_14_v1	113962	G	A	1	22	100	0.01	missense	RBP2a	PVP01_1402400	154G>154K	113962G>A+113963G>A
PvP01_14_v1	113963	G	A	1	22	100	0.01	missense	RBP2a	PVP01_1402400	154G>154K	113962G>A+113963G>A
PvP01_14_v1	114058	A	G	2	34	87	0.02	missense	RBP2a	PVP01_1402400	186S>186D	114058A>G+114059G>A
PvP01_14_v1	114059	G	A	4	35	84	0.05	missense	RBP2a	PVP01_1402400	186S>186D	114058A>G+114059G>A
PvP01_14_v1	114059	G	A	4	35	84	0.05	missense	RBP2a	PVP01_1402400	186S>186N	114059G>A
PvP01_14_v1	114157	C	A	81	42	0	1	missense	RBP2a	PVP01_1402400	219R>219S	114157C>A
PvP01_14_v1	114237	G	A	12	26	85	0.12	synonymous	RBP2a	PVP01_1402400	245E	114237G>A
PvP01_14_v1	114331	A	G	5	23	95	0.05	missense	RBP2a	PVP01_1402400	277K>277E	114331A>G
PvP01_14_v1	114357	T	A	8	32	83	0.09	missense	RBP2a	PVP01_1402400	285N>285I	114356A>T+114357T>A
PvP01_14_v1	114357	T	A	8	32	83	0.09	missense	RBP2a	PVP01_1402400	285N>285K	114357T>A
PvP01_14_v1	114369	A	T	1	24	98	0.01	missense	RBP2a	PVP01_1402400	289K>289N	114369A>T
PvP01_14_v1	114412	G	A	8	23	92	0.08	missense	RBP2a	PVP01_1402400	304E>304K	114412G>A
PvP01_14_v1	114419	A	T	11	26	86	0.11	missense	RBP2a	PVP01_1402400	306D>306V	114419A>T
PvP01_14_v1	114427	G	A	1	24	98	0.01	missense	RBP2a	PVP01_1402400	309E>309K	114427G>A
PvP01_14_v1	114697	T	C	13	30	80	0.14	missense	RBP2a	PVP01_1402400	399S>399L	114697T>C+114698C>T
PvP01_14_v1	114697	T	C	13	30	80	0.14	missense	RBP2a	PVP01_1402400	399S>399P	114697T>C
PvP01_14_v1	114699	A	T	3	27	93	0.03	missense	RBP2a	PVP01_1402400	399S>399P	114697T>C+114699A>T
PvP01_14_v1	114700	C	A	2	22	99	0.02	missense	RBP2a	PVP01_1402400	400Q>400K	114700C>A
PvP01_14_v1	114736	G	T	5	24	94	0.05	missense	RBP2a	PVP01_1402400	412D>412Y	114736G>T
PvP01_14_v1	114808	G	A	2	24	97	0.02	missense	RBP2a	PVP01_1402400	436E>436K	114808G>A
PvP01_14_v1	114814	G	A	2	24	97	0.02	missense	RBP2a	PVP01_1402400	438E>438K	114814G>A
PvP01_14_v1	114814	G	A	2	24	97	0.02	missense	RBP2a	PVP01_1402400	438E>438R	114814G>A+114815A>G
PvP01_14_v1	114815	A	G	85	38	0	1	missense	RBP2a	PVP01_1402400	438E>438G	114815A>G

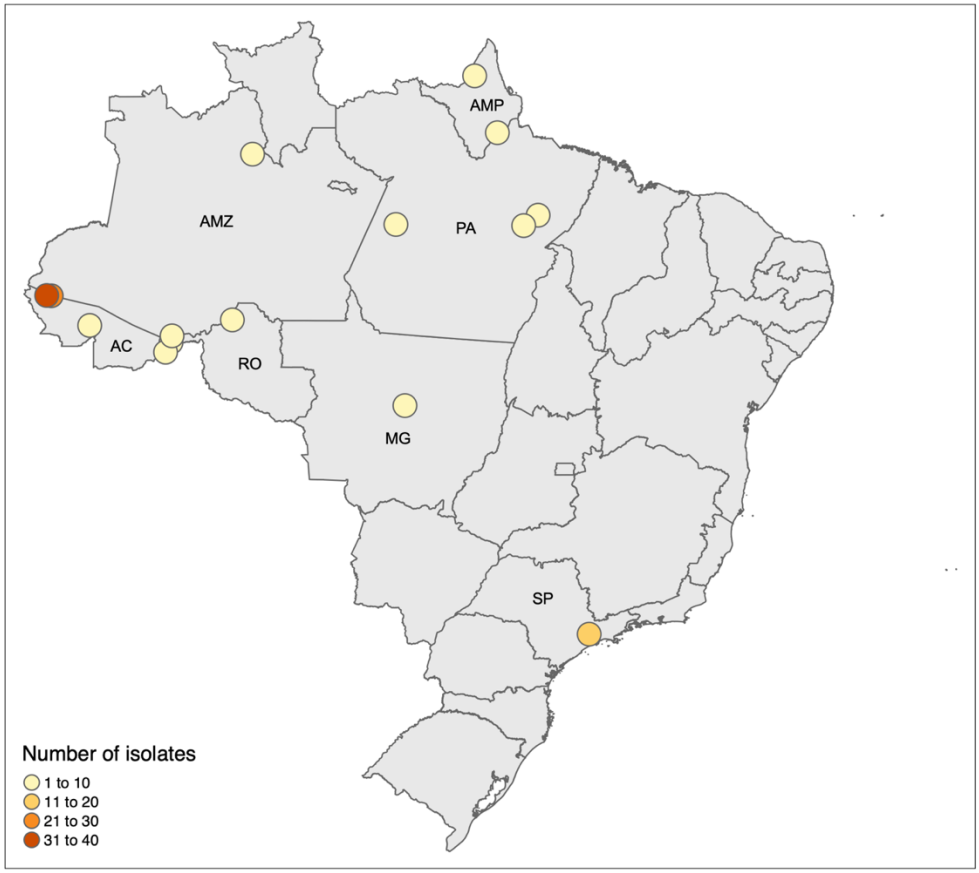
PvP01_14_v1	114815	A	G	85	38	0	1	missense	RBP2a	PVP01_1402400	438E>438R	114814G>A+114815A>G
PvP01_14_v1	114826	G	C	2	22	99	0.02	missense	RBP2a	PVP01_1402400	442A>442R	114826G>C+114827C>G
PvP01_14_v1	114827	C	G	2	24	97	0.02	missense	RBP2a	PVP01_1402400	442A>442R	114826G>C+114827C>G
PvP01_14_v1	114840	T	A	2	22	99	0.02	missense	RBP2a	PVP01_1402400	446N>446K	114840T>A
PvP01_14_v1	114884	A	G	1	27	95	0.01	missense	RBP2a	PVP01_1402400	461D>461G	114884A>G
PvP01_14_v1	114892	T	A	3	26	94	0.03	missense	RBP2a	PVP01_1402400	464Y>464N	114892T>A
PvP01_14_v1	114907	A	G	5	24	94	0.05	missense	RBP2a	PVP01_1402400	469S>469G	114907A>G
PvP01_14_v1	115034	T	A	87	36	0	1	missense	RBP2a	PVP01_1402400	511M>511K	115034T>A
PvP01_14_v1	115121	A	G	17	27	79	0.18	missense	RBP2a	PVP01_1402400	540N>540S	115121A>G
PvP01_14_v1	115171	A	G	62	30	31	0.67	missense	RBP2a	PVP01_1402400	557N>557D	115171A>G
PvP01_14_v1	115220	A	C	83	32	8	0.91	missense	RBP2a	PVP01_1402400	573K>573T	115220A>C
PvP01_14_v1	115333	A	G	9	31	83	0.1	missense	RBP2a	PVP01_1402400	611T>611A	115333A>G
PvP01_14_v1	115345	G	A	6	27	90	0.06	missense	RBP2a	PVP01_1402400	615D>615N	115345G>A
PvP01_14_v1	115397	A	G	1	26	96	0.01	missense	RBP2a	PVP01_1402400	632D>632G	115397A>G
PvP01_14_v1	115402	A	G	66	30	27	0.71	missense	RBP2a	PVP01_1402400	634K>634E	115402A>G
PvP01_14_v1	115651	G	T	3	25	95	0.03	missense	RBP2a	PVP01_1402400	717V>717F	115651G>T
PvP01_14_v1	115657	A	G	46	25	52	0.47	missense	RBP2a	PVP01_1402400	719K>719E	115657A>G
PvP01_14_v1	115993	C	G	1	21	101	0.01	missense	RBP2a	PVP01_1402400	831L>831V	115993C>G
PvP01_14_v1	116296	A	G	69	38	16	0.81	missense	RBP2a	PVP01_1402400	932R>932E	116296A>G+116297G>A
PvP01_14_v1	116297	G	A	69	38	16	0.81	missense	RBP2a	PVP01_1402400	932R>932E	116296A>G+116297G>A
PvP01_14_v1	116329	G	A	1	26	96	0.01	missense	RBP2a	PVP01_1402400	943E>943K	116329G>A
PvP01_14_v1	116433	A	G	6	26	91	0.06	synonymous	RBP2a	PVP01_1402400	977Q	116433A>G
PvP01_14_v1	116714	T	A	72	51	0	1	missense	RBP2a	PVP01_1402400	1071I>1071K	116714T>A
PvP01_14_v1	117334	C	G	3	24	96	0.03	missense	RBP2a	PVP01_1402400	1278Q>1278E	117334C>G
PvP01_14_v1	117588	C	A	5	22	96	0.05	missense	RBP2a	PVP01_1402400	1362N>1362K	117588C>A
PvP01_14_v1	117590	T	A	5	22	96	0.05	missense	RBP2a	PVP01_1402400	1363I>1363N	117590T>A
PvP01_14_v1	118007	A	T	6	31	86	0.07	missense	RBP2a	PVP01_1402400	1502D>1502V	118007A>T
PvP01_14_v1	118307	T	C	90	30	3	0.97	missense	RBP2a	PVP01_1402400	1602I>1602T	118307T>C
PvP01_14_v1	118326	A	T	10	23	90	0.1	synonymous	RBP2a	PVP01_1402400	1608G	118326A>T
PvP01_14_v1	118351	G	T	17	22	84	0.17	missense	RBP2a	PVP01_1402400	1617V>1617L	118351G>T
PvP01_14_v1	118481	A	G	6	20	97	0.06	missense	RBP2a	PVP01_1402400	1660Y>1660C	118481A>G
PvP01_14_v1	118720	A	G	1	22	100	0.01	missense	RBP2a	PVP01_1402400	1740N>1740D	118720A>G
PvP01_14_v1	118737	A	T	75	30	18	0.81	missense	RBP2a	PVP01_1402400	1745K>1745D	118735A>G+118737A>T
PvP01_14_v1	118737	A	T	75	30	18	0.81	missense	RBP2a	PVP01_1402400	1745K>1745N	118737A>T
PvP01_14_v1	119489	C	T	4	20	99	0.04	missense	RBP2a	PVP01_1402400	1996A>1996V	119489C>T
PvP01_14_v1	119712	A	G	88	30	5	0.95	missense	RBP2a	PVP01_1402400	2070I>2070M	119712A>G
PvP01_14_v1	120040	A	G	82	24	17	0.83	missense	RBP2a	PVP01_1402400	2180K>2180E	120040A>G
PvP01_14_v1	120052	G	C	89	26	8	0.92	missense	RBP2a	PVP01_1402400	2184E>2184Q	120052G>C
PvP01_14_v1	120091	T	C	1	17	105	0.01	missense	RBP2a	PVP01_1402400	2197Y>2197H	120091T>C
PvP01_14_v1	120096	T	C	90	24	9	0.91	synonymous	RBP2a	PVP01_1402400	2198N	120096T>C
PvP01_14_v1	120211	T	G	3	17	103	0.03	missense	RBP2a	PVP01_1402400	2237S>2237A	120211T>G
PvP01_14_v1	120302	T	A	95	28	0	1	missense	RBP2a	PVP01_1402400	2267M>2267K	120302T>A
PvP01_14_v1	120473	T	G	88	29	6	0.94	missense	RBP2a	PVP01_1402400	2324I>2324R	120473T>G
PvP01_14_v1	120596	G	A	73	23	27	0.73	missense	RBP2a	PVP01_1402400	2365G>2365D	120596G>A
PvP01_14_v1	120608	C	A	9	21	93	0.09	missense	RBP2a	PVP01_1402400	2369A>2369E	120608C>A
PvP01_14_v1	120644	G	A	5	18	100	0.05	missense	RBP2a	PVP01_1402400	2381G>2381E	120644G>A
PvP01_14_v1	120715	A	T	96	27	0	1	missense	RBP2a	PVP01_1402400	2405N>2405Y	120715A>T
PvP01_14_v1	120854	G	A	3	14	106	0.03	missense	RBP2a	PVP01_1402400	2451G>2451E	120854G>A
PvP01_14_v1	120959	C	T	93	24	6	0.94	missense	RBP2a	PVP01_1402400	2486A>2486V	120959C>T
PvP01_14_v1	826173	C	G	1	3	119	0.01	synonymous	PVP01_1419000	PVP01_1419000	193L	826173C>G

PvP01_14_v1	826313	A	C	116	7	0	1	missense	PVP01_1419000	PVP01_1419000	147S>147A	826313A>C
PvP01_14_v1	826496	C	T	1	4	118	0.01	missense	PVP01_1419000	PVP01_1419000	86G>86R	826496C>T
PvP01_14_v1	1070602	G	T	7	3	113	0.06	missense	PVP01_1424900	PVP01_1424900	421L>421I	1070602G>T
PvP01_14_v1	1070614	T	A	7	3	113	0.06	missense	PVP01_1424900	PVP01_1424900	417I>417F	1070614T>A
PvP01_14_v1	1071124	G	A	116	7	0	1	missense	PVP01_1424900	PVP01_1424900	247H>247Y	1071124G>A
PvP01_14_v1	1071245	T	G	114	9	0	1	synonymous	PVP01_1424900	PVP01_1424900	206L	1071245T>G
PvP01_14_v1	1071643	G	A	109	14	0	1	missense	PVP01_1424900	PVP01_1424900	74H>74Y	1071643G>A
PvP01_14_v1	1071664	G	T	108	15	0	1	missense	PVP01_1424900	PVP01_1424900	67L>67I	1071664G>T
PvP01_14_v1	1270628	G	A	4	4	115	0.03	synonymous	PPPK-DHPS	PVP01_1429500	477S	1270628G>A
PvP01_14_v1	1270743	C	T	1	2	120	0.01	missense	PPPK-DHPS	PVP01_1429500	439G>439E	1270743C>T
PvP01_14_v1	1270911	C	G	14	4	105	0.12	missense	PPPK-DHPS	PVP01_1429500	383G>383A	1270911C>G
PvP01_14_v1	1270914	G	C	40	4	79	0.34	missense	PPPK-DHPS	PVP01_1429500	382S>382C	1270914G>C
PvP01_14_v1	1271444	C	T	101	7	15	0.87	missense	PPPK-DHPS	PVP01_1429500	205M>205I	1271444C>T
PvP01_14_v1	1276331	C	T	59	3	61	0.49	missense	DBP1	PVP01_1429700	920V>920I	1276331C>T
PvP01_14_v1	1276738	G	A	2	2	119	0.02	missense	DBP1	PVP01_1429700	784S>784F	1276738G>A
PvP01_14_v1	1277142	G	A	3	5	115	0.03	synonymous	DBP1	PVP01_1429700	649N	1277142G>A
PvP01_14_v1	1277438	T	G	116	7	0	1	synonymous	DBP1	PVP01_1429700	551R	1277438T>G
PvP01_14_v1	1277664	G	A	118	5	0	1	synonymous	DBP1	PVP01_1429700	475N	1277664G>A
PvP01_14_v1	1279062	G	A	65	11	47	0.58	synonymous	DBP1	PVP01_1429700	9N	1279062G>A
PvP01_14_v1	1543436	G	T	2	2	119	0.02	missense	TRAMP	PVP01_1436800	130L>130F	1543436G>T
PvP01_14_v1	1543449	G	C	4	2	117	0.03	missense	TRAMP	PVP01_1436800	135G>135R	1543449G>C
PvP01_14_v1	1543461	A	G	1	1	121	0.01	missense	TRAMP	PVP01_1436800	139S>139G	1543461A>G
PvP01_14_v1	1543516	T	A	117	6	0	1	missense	TRAMP	PVP01_1436800	157M>157K	1543516T>A
PvP01_14_v1	1544063	C	T	107	9	7	0.94	synonymous	TRAMP	PVP01_1436800	339Y	1544063C>T
PvP01_14_v1	1560475	C	T	2	11	110	0.02	synonymous	AP2-MU	PVP01_1437100	601Y	1560475C>T
PvP01_14_v1	2020861	C	A	1	1	121	0.01	missense	MSP9	PVP01_1446800	815V>815A	2020861C>A+2020862A>G
PvP01_14_v1	2020862	A	G	1	1	121	0.01	missense	MSP9	PVP01_1446800	815V>815A	2020861C>A+2020862A>G
PvP01_14_v1	2021271	T	C	30	2	91	0.25	missense	MSP9	PVP01_1446800	679T>679A	2021271T>C
PvP01_14_v1	2021431	T	C	34	1	88	0.28	synonymous	MSP9	PVP01_1446800	625A	2021431T>C
PvP01_14_v1	2022081	T	C	120	3	0	1	missense	MSP9	PVP01_1446800	409N>409D	2022081T>C
PvP01_14_v1	2022102	C	T	120	3	0	1	missense	MSP9	PVP01_1446800	402E>402K	2022102C>T
PvP01_14_v1	2022220	G	A	2	1	120	0.02	synonymous	MSP9	PVP01_1446800	362Y	2022220G>A
PvP01_14_v1	2022309	C	G	11	2	110	0.09	missense	MSP9	PVP01_1446800	333V>333L	2022309C>G
PvP01_14_v1	2022340	C	G	120	3	0	1	synonymous	MSP9	PVP01_1446800	322V	2022340C>G
PvP01_14_v1	2022363	G	C	2	0	121	0.02	missense	MSP9	PVP01_1446800	315L>315V	2022363G>C
PvP01_14_v1	2022437	A	G	121	2	0	1	missense	MSP9	PVP01_1446800	290I>290T	2022437A>G
PvP01_14_v1	2022559	C	T	6	1	116	0.05	synonymous	MSP9	PVP01_1446800	249K	2022559C>T
PvP01_14_v1	2022656	T	C	82	6	35	0.7	missense	MSP9	PVP01_1446800	217H>217R	2022656T>C
PvP01_14_v1	2022663	G	C	117	6	0	1	missense	MSP9	PVP01_1446800	215Q>215E	2022663G>C
PvP01_14_v1	2022941	C	T	118	5	0	1	missense	MSP9	PVP01_1446800	122G>122D	2022941C>T
PvP01_14_v1	2022941	C	T	118	5	0	1	missense	MSP9	PVP01_1446800	122G>122N	2022941C>T+2022942C>T
PvP01_14_v1	2022942	C	T	5	3	115	0.04	missense	MSP9	PVP01_1446800	122G>122N	2022941C>T+2022942C>T
PvP01_14_v1	2052257	C	G	112	11	0	1	missense	MRP2	PVP01_1447300	1956D>1956H	2052257C>G
PvP01_14_v1	2052294	C	T	5	2	116	0.04	synonymous	MRP2	PVP01_1447300	1943V	2052294C>T
PvP01_14_v1	2052369	T	C	116	5	2	0.98	synonymous	MRP2	PVP01_1447300	1918V	2052369T>C
PvP01_14_v1	2052588	A	G	117	6	0	1	synonymous	MRP2	PVP01_1447300	1845N	2052588A>G
PvP01_14_v1	2053134	G	A	4	14	105	0.04	synonymous	MRP2	PVP01_1447300	1663P	2053134G>A
PvP01_14_v1	2053883	A	G	118	5	0	1	missense	MRP2	PVP01_1447300	1414Y>1414H	2053883A>G
PvP01_14_v1	2053904	C	G	38	4	81	0.32	missense	MRP2	PVP01_1447300	1407E>1407Q	2053904C>G
PvP01_14_v1	2053938	G	A	4	4	115	0.03	synonymous	MRP2	PVP01_1447300	1395H	2053938G>A

PvP01_14_v1	2054372	T	A	11	1	111	0.09	missense	MRP2	PVP01_1447300	1251N>1251Y	2054372T>A
PvP01_14_v1	2054573	G	A	63	5	55	0.53	missense	MRP2	PVP01_1447300	1184P>1184S	2054573G>A
PvP01_14_v1	2054843	C	A	59	9	55	0.52	missense	MRP2	PVP01_1447300	1094A>1094S	2054843C>A
PvP01_14_v1	2055068	G	A	1	2	120	0.01	missense	MRP2	PVP01_1447300	1019L>1019F	2055068G>A
PvP01_14_v1	2055069	G	A	1	2	120	0.01	synonymous	MRP2	PVP01_1447300	1018D	2055069G>A
PvP01_14_v1	2055095	T	C	4	2	117	0.03	missense	MRP2	PVP01_1447300	1010M>1010V	2055095T>C
PvP01_14_v1	2055185	T	C	10	2	111	0.08	missense	MRP2	PVP01_1447300	980I>980V	2055185T>C
PvP01_14_v1	2055753	A	G	112	11	0	1	synonymous	MRP2	PVP01_1447300	790Y	2055753A>G
PvP01_14_v1	2055810	T	G	109	9	5	0.96	synonymous	MRP2	PVP01_1447300	771I	2055810T>G
PvP01_14_v1	2056117	C	A	3	11	109	0.03	missense	MRP2	PVP01_1447300	669S>669I	2056117C>A
PvP01_14_v1	2057278	A	C	3	6	114	0.03	missense	MRP2	PVP01_1447300	282M>282R	2057278A>C
PvP01_14_v1	2057525	G	T	10	2	111	0.08	missense	MRP2	PVP01_1447300	200L>200I	2057525G>T
PvP01_14_v1	2057652	C	G	1	4	118	0.01	missense	MRP2	PVP01_1447300	157Q>157H	2057652C>G
PvP01_14_v1	2057664	G	C	109	7	7	0.94	synonymous	MRP2	PVP01_1447300	153L	2057664G>C
PvP01_14_v1	2057712	G	T	17	4	102	0.14	synonymous	MRP2	PVP01_1447300	137T	2057712G>T
PvP01_14_v1	2057755	C	T	7	4	112	0.06	missense	MRP2	PVP01_1447300	123R>123K	2057755C>T
PvP01_14_v1	2057861	C	G	13	2	108	0.11	missense	MRP2	PVP01_1447300	88E>88Q	2057861C>G
PvP01_14_v1	2057943	A	G	116	7	0	1	synonymous	MRP2	PVP01_1447300	60Y	2057943A>G
PvP01_14_v1	2058037	G	C	3	2	118	0.02	missense	MRP2	PVP01_1447300	29T>29S	2058037G>C
PvP01_14_v1	2206594	A	G	14	7	102	0.12	synonymous	PVP01_1450700	PVP01_1450700	830Y	2206594A>G
PvP01_14_v1	2206666	T	C	102	9	12	0.89	synonymous	PVP01_1450700	PVP01_1450700	806V	2206666T>C
PvP01_14_v1	2206669	G	A	113	10	0	1	synonymous	PVP01_1450700	PVP01_1450700	805F	2206669G>A
PvP01_14_v1	2206719	G	T	12	14	97	0.11	missense	PVP01_1450700	PVP01_1450700	789L>789I	2206719G>T
PvP01_14_v1	2206735	G	C	104	19	0	1	synonymous	PVP01_1450700	PVP01_1450700	783P	2206735G>C
PvP01_14_v1	2206774	T	G	110	13	0	1	synonymous	PVP01_1450700	PVP01_1450700	770T	2206774T>G
PvP01_14_v1	2206809	C	G	4	6	113	0.03	missense	PVP01_1450700	PVP01_1450700	759E>759Q	2206809C>G
PvP01_14_v1	2206819	G	A	110	10	3	0.97	synonymous	PVP01_1450700	PVP01_1450700	755R	2206819G>A
PvP01_14_v1	2207633	C	G	33	8	82	0.29	missense	PVP01_1450700	PVP01_1450700	484G>484A	2207633C>G
PvP01_14_v1	2208534	G	T	2	3	118	0.02	synonymous	PVP01_1450700	PVP01_1450700	184R	2208534G>T
PvP01_14_v1	2208948	A	G	111	10	2	0.98	missense	PVP01_1450700	PVP01_1450700	91L>91P	2208948A>G
PvP01_14_v1	2750044	C	T	1	3	119	0.01	missense	PKB	PVP01_1464000	384S>384N	2750044C>T
PvP01_14_v1	2750469	C	T	17	3	103	0.14	missense	PKB	PVP01_1464000	287A>287T	2750469C>T
PvP01_14_v1	2750512	A	C	20	4	99	0.17	missense	PKB	PVP01_1464000	272S>272R	2750512A>C
PvP01_14_v1	2750517	C	T	14	2	107	0.12	missense	PKB	PVP01_1464000	271E>271K	2750517C>T
PvP01_14_v1	2750737	C	T	20	0	103	0.16	synonymous	PKB	PVP01_1464000	197K	2750737C>T
PvP01_14_v1	2905414	G	T	120	3	0	1	missense	PVP01_1468200	PVP01_1468200	7A>7S	2905414G>T
PvP01_14_v1	2906043	C	T	13	6	104	0.11	missense	PVP01_1468200	PVP01_1468200	77T>77I	2906043C>T
PvP01_14_v1	2906054	T	A	113	9	1	0.99	missense	PVP01_1468200	PVP01_1468200	81L>81I	2906054T>A
PvP01_14_v1	2906268	G	A	105	18	0	1	missense	PVP01_1468200	PVP01_1468200	97G>97R	2906268G>A
PvP01_14_v1	2906323	C	T	76	9	38	0.67	missense	PVP01_1468200	PVP01_1468200	115A>115V	2906323C>T
PvP01_14_v1	2909573	G	T	1	5	117	0.01	missense	PVP01_1468300	PVP01_1468300	492A>492S	2909573G>T
PvP01_14_v1	2909751	T	G	21	9	93	0.18	missense	PVP01_1468300	PVP01_1468300	551F>551C	2909751T>G

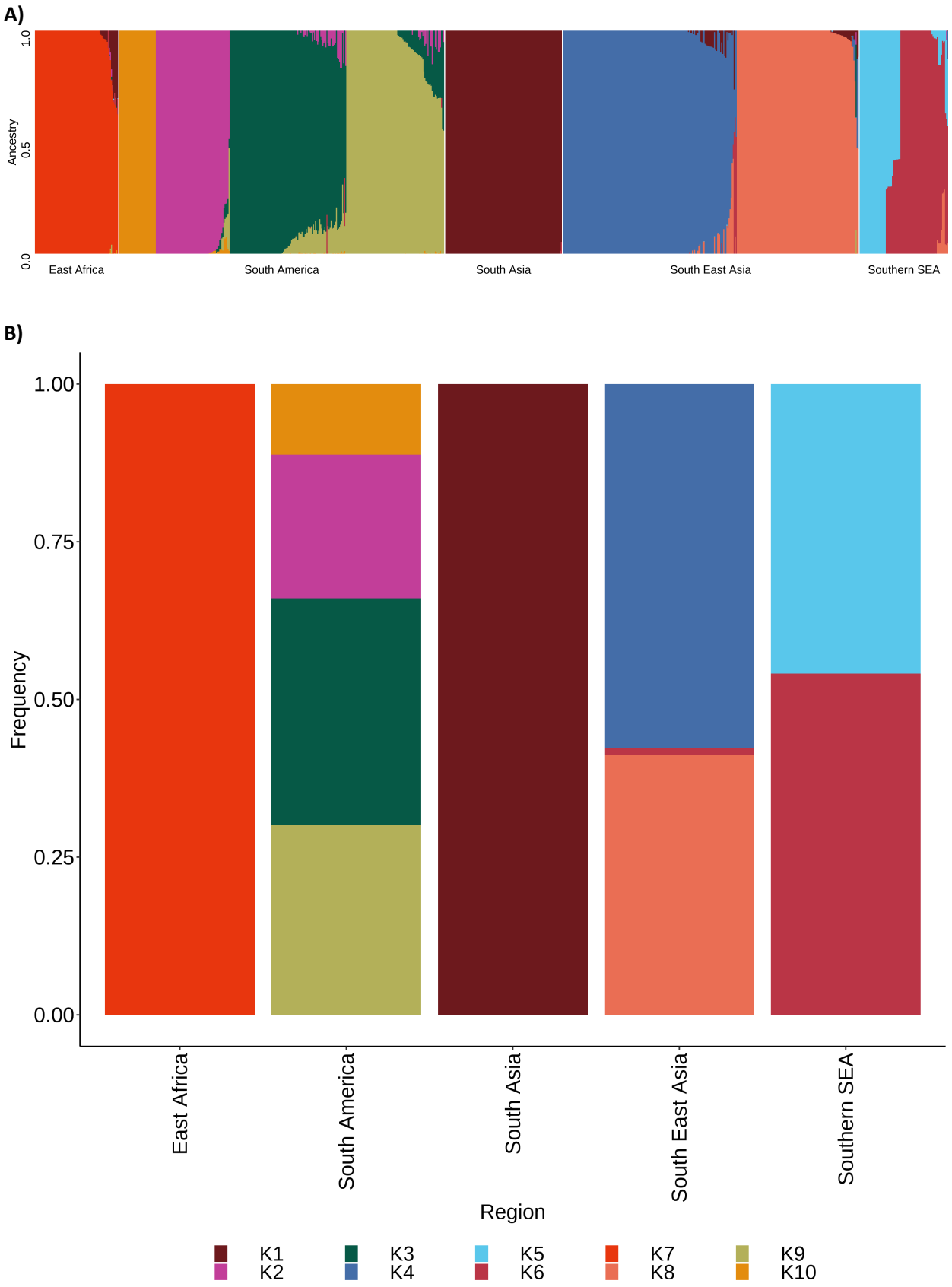
SUPPLEMENTARY FIGURES

SUPPLEMENTARY FIGURES

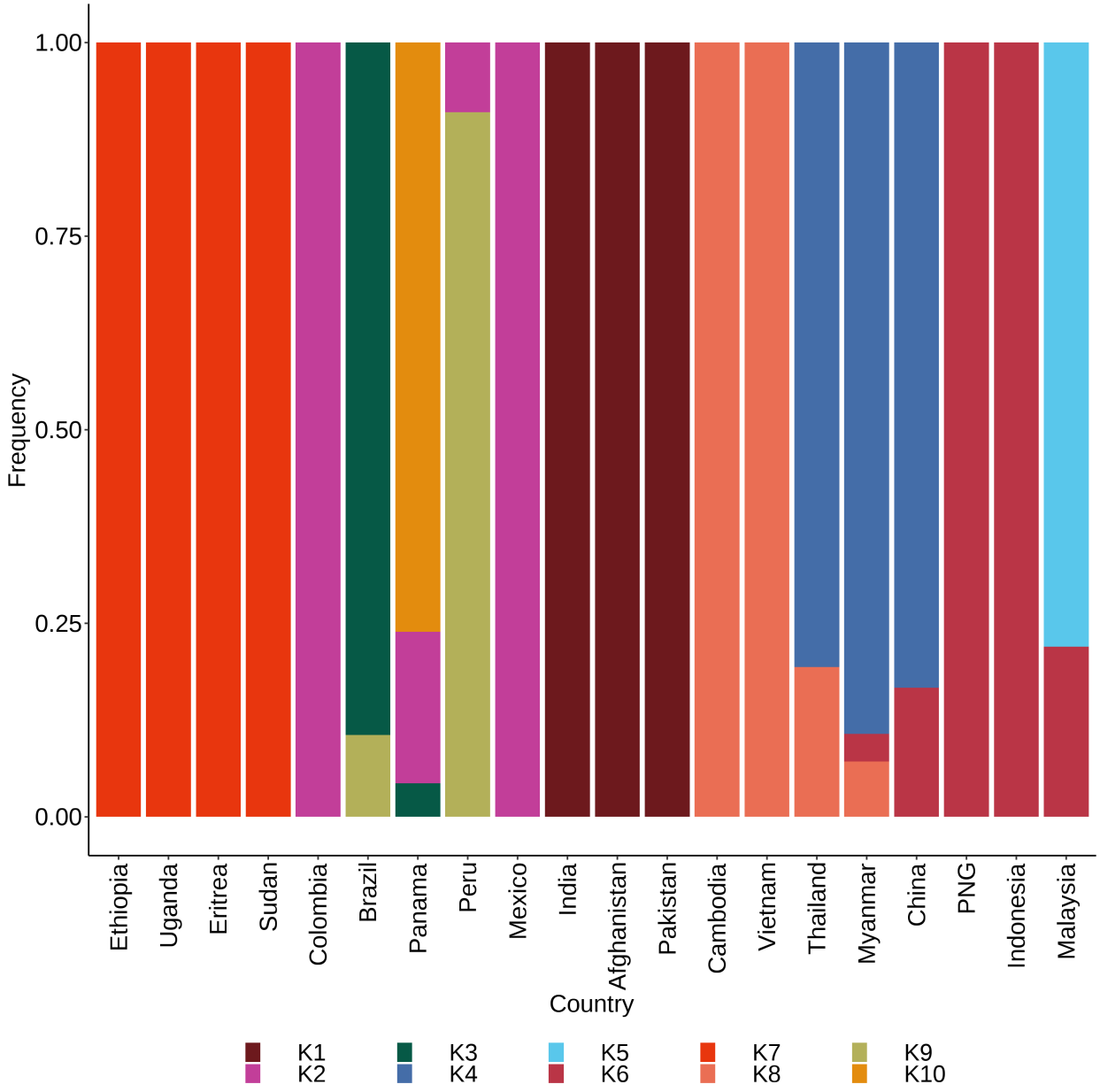


S1 Figure. Location of Brazilian *Plasmodium vivax* isolates.

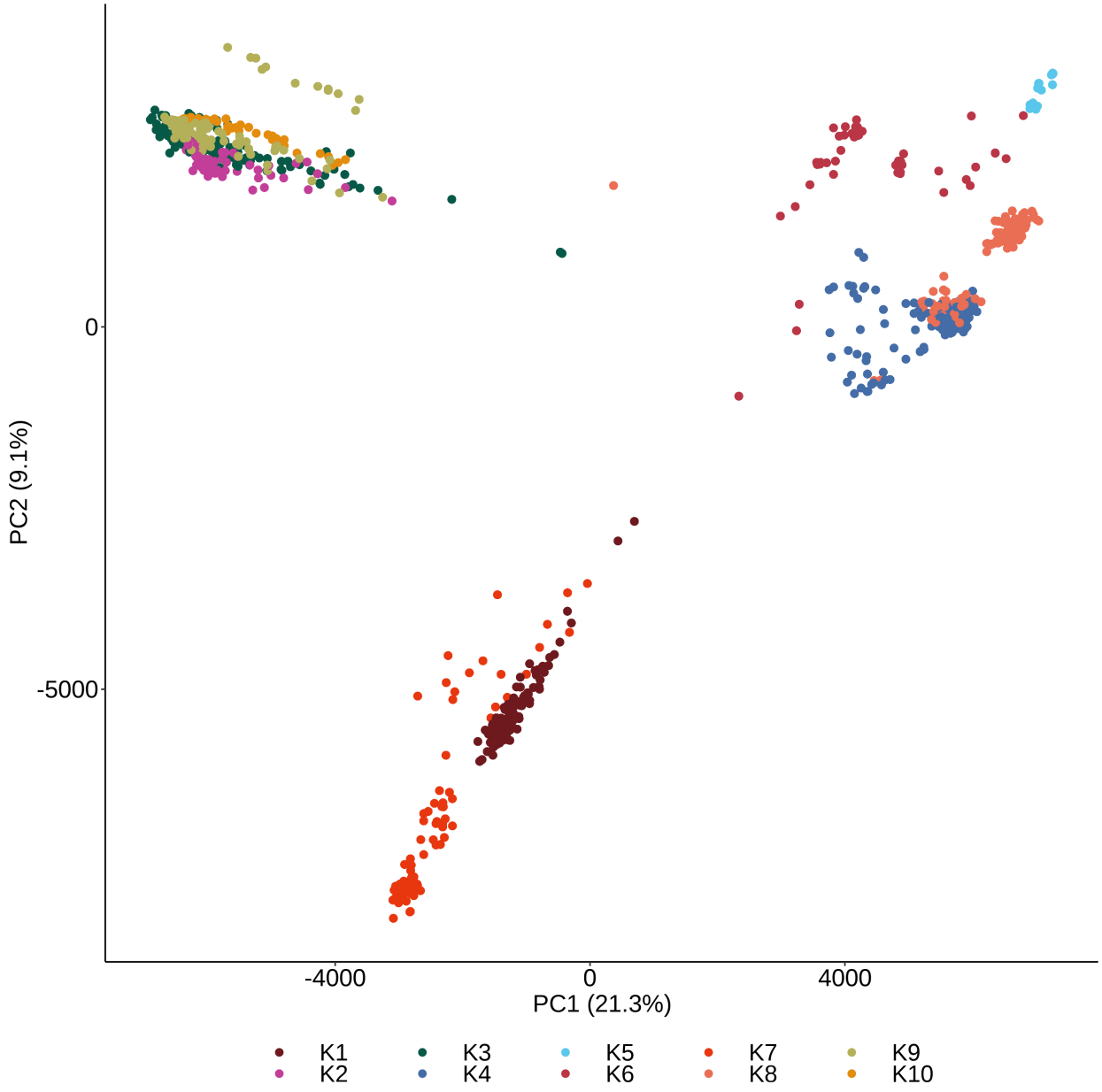
The locations of the 123 Brazilian *P. vivax* isolates included within the final filtered database are plotted on a map of Brazil with the states where isolates were obtained labelled (AC; Acre, AMZ; Amazonas, RO; Rondônia, AMP; Amapá, PA; *Pará*, SP; São Paulo). Location points are coloured according to the scale, which represents the number of isolates from each location point. The location with the most isolates was Mancio Lima in Acre state (N = 38), and the lowest were Macapá within Amapá (N = 1) and Remansinho in Amazonas state (N = 1). For specific isolate location, date of collection and the corresponding accession numbers, please refer to **S2 Table**.


A)
Ancestry
1.0
0.5
0.0
East Africa
South America
South Asia
South East Asia
Southern SEA
B)
Frequency
1.00
0.75
0.50
0.25
0.00
East Africa
South America
South Asia
South East Asia
Southern SEA
Region
K1
K2
K3
K4
K5
K6
K7
K8
K9
K10

C)

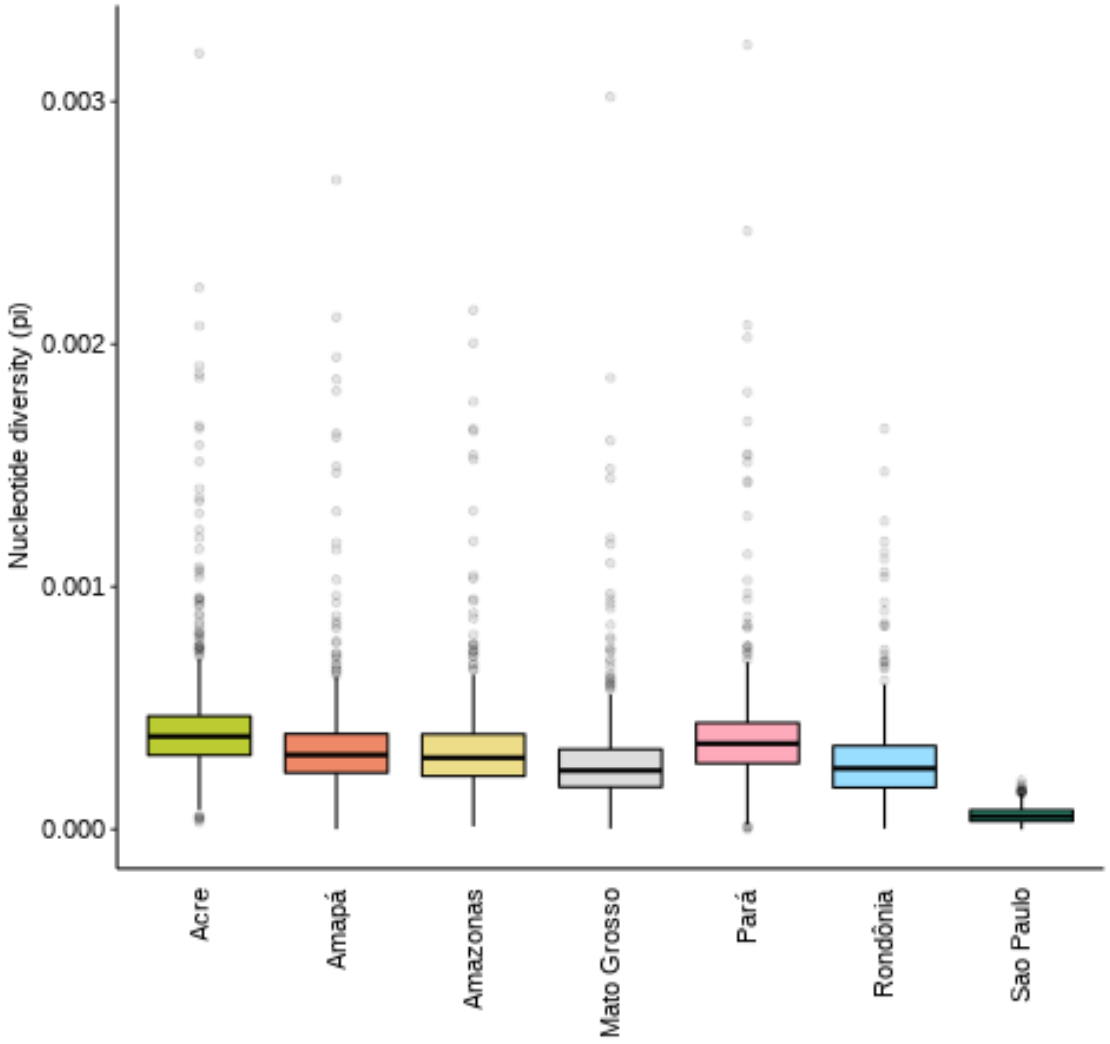

1.00
0.75
0.50
0.25
0.00
Frequency
Ethiopia
Uganda
Eritrea
Sudan
Colombia
Brazil
Panama
Peru
Mexico
India
Afghanistan
Pakistan
Cambodia
Vietnam
Thailand
Myanmar
China
PNG
Indonesia
Malaysia
Country
K1
K2
K3
K4
K5
K6
K7
K8
K9
K10

D)



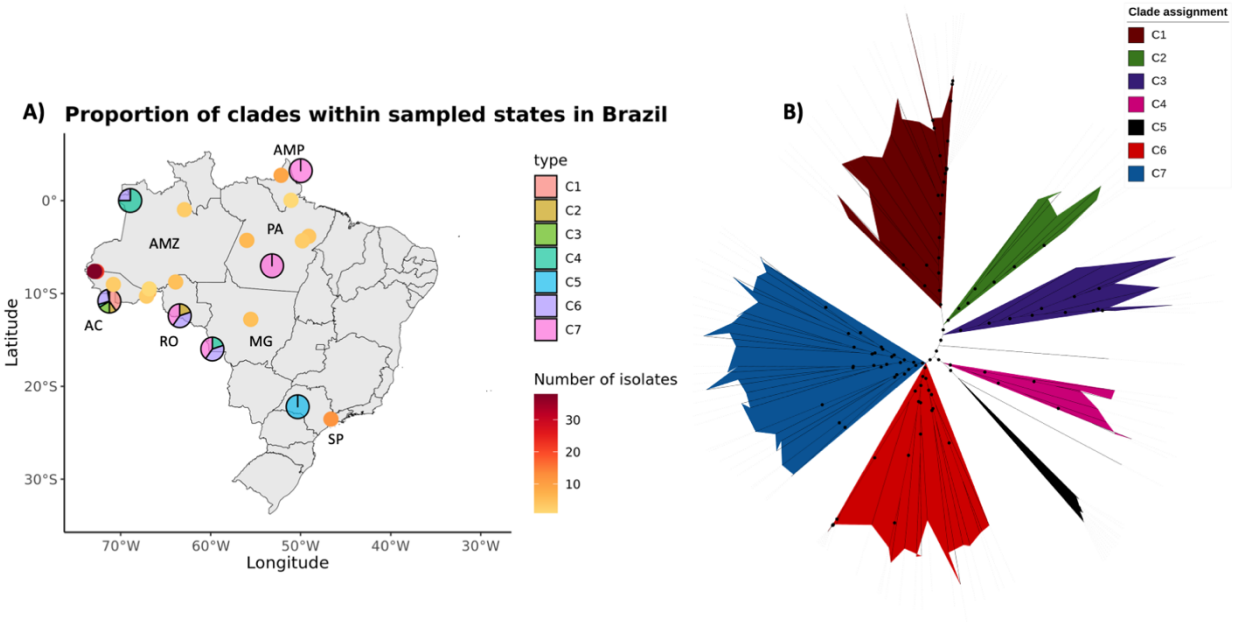
S2 Figure. Admixture analysis

All global isolates (n = 855) were assessed for population structure using admixture. Ten ancestral populations were assigned to the global dataset, the distribution of each ancestral population amongst each region **(A, B)** and country **(C)** is highlighted in the boxplots below. The PCA generated using the global SNP matrix is coloured by ancestral populations **(D).**

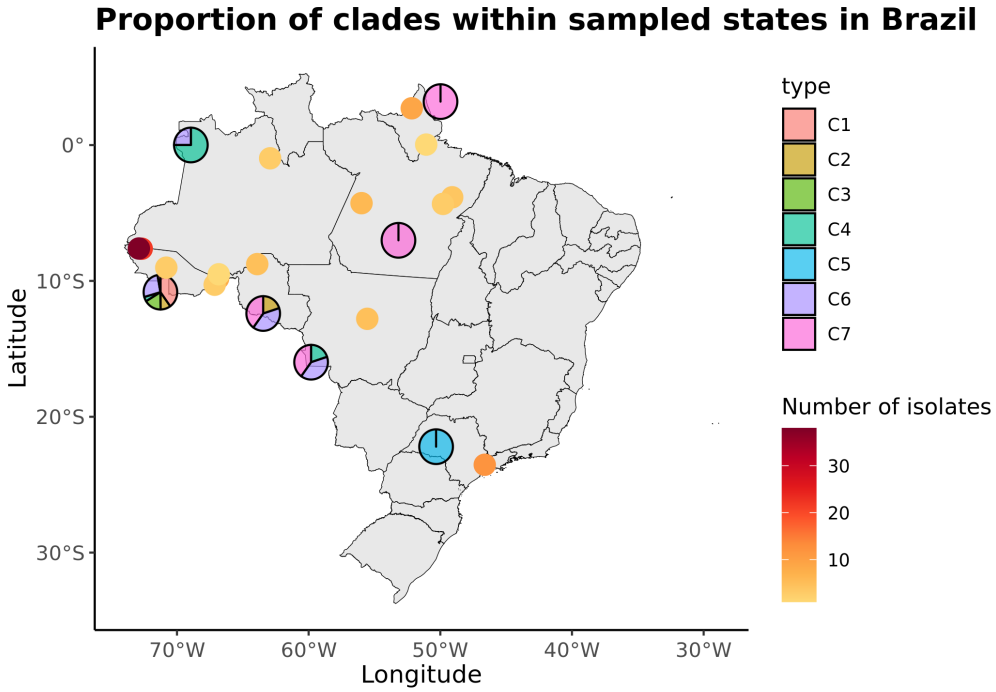


S3 Figure. Nucleotide diversity within each state in Brazil

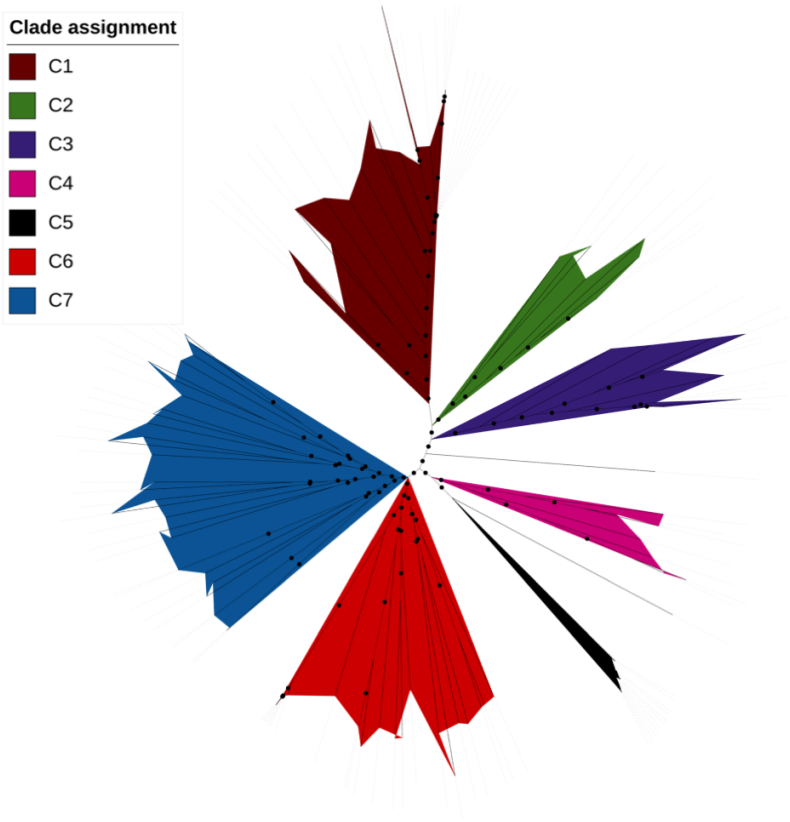
Boxplot highlighting nucleotide diversity (π), calculated in sliding windows of 25 kbp across the genome using VCFtools for isolates within each state (*Pará*, n = 13; Amapá, n = 10; Mato Grosso, n = 5; Rondônia, n = 5; Acre, n = 74; Amazonas, n = 4; São Paulo, n = 12). All boxplot boxes consist of the median and interquartile range of the data for each country, with whiskers extending to extreme data points within 1.5 times the interquartile range from the box, data points beyond this range are outliers and plotted as points.


A) Proportion of clades within sampled states in Brazil
AMP
PA
AMZ
AC
RO
MG
SP
0°
10°S
20°S
30°S
70°W
60°W
50°W
40°W
30°W
Latitude
Longitude
type
C1
C2
C3
C4
C5
C6
C7
Number of isolates
30
20
10
B)
Clade assignment
C1
C2
C3
C4
C5
C6
C7

A)

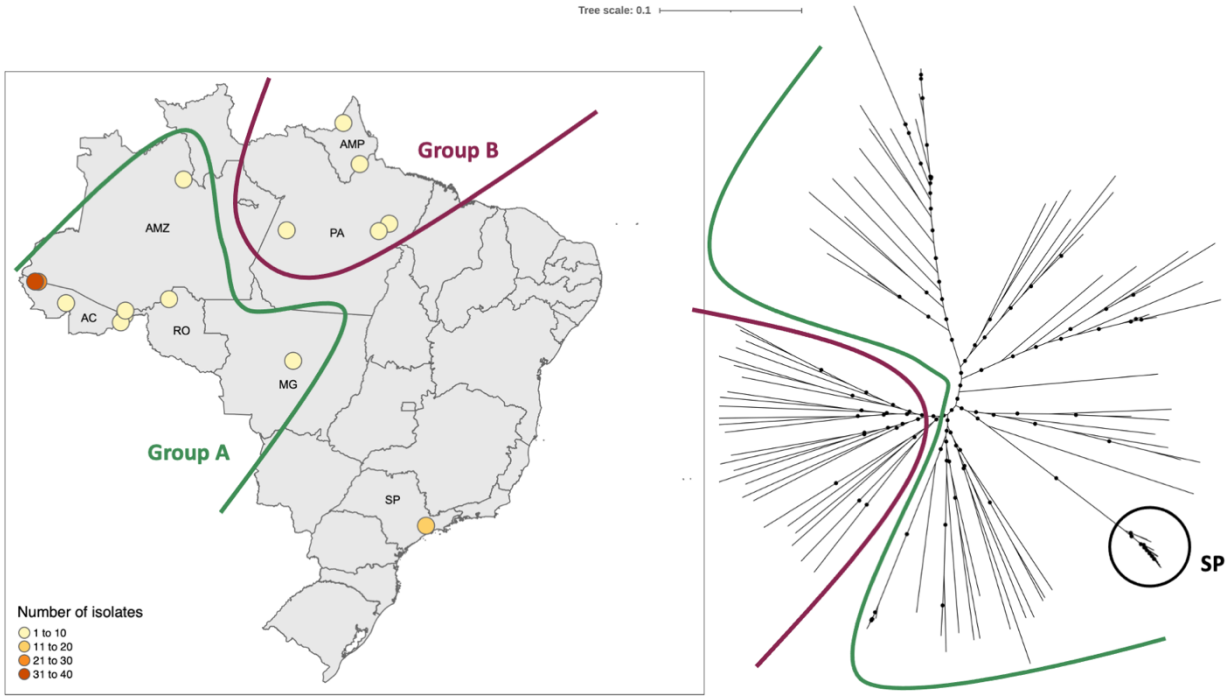

Proportion of clades within sampled states in Brazil
type
C1
C2
C3
C4
C5
C6
C7
Number of isolates
30
20
10
0°
10°S
20°S
30°S
Latitude
70°W
60°W
50°W
40°W
30°W
Longitude

B)


Clade assignment
C1
C2
C3
C4
C5
C6
C7

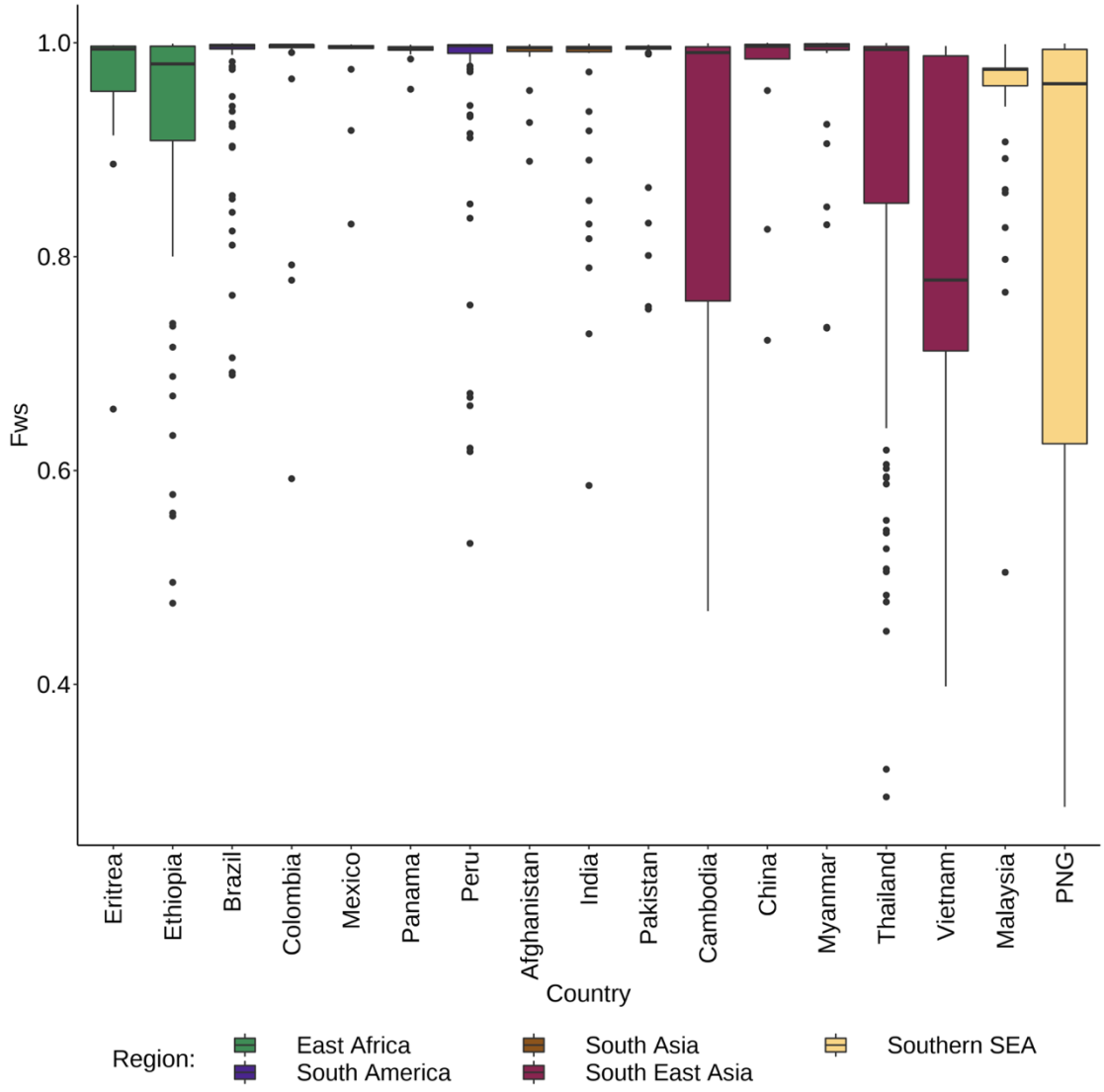
S4 Figure. Proportion of assigned clades across sampled states in Brazil

A) Map of Brazil with isolate location plotted as points, coloured according to the number of isolates sequenced at this location. States where isolates were obtained are labelled on the map (AC; Acre, AMZ; Amazonas, RO; Rondônia, MG; Mato Grosso, PA; *Pará*, AMP; Amapá, SP; São Paulo). Pie charts indicate the proportion of each clade found within each state. **B)** The Maximum Likelihood phylogenetic tree of 123 isolates from Brazil, comprising 70,757 SNPs, demonstrating the 7 clades assigned within the Brazilian dataset.



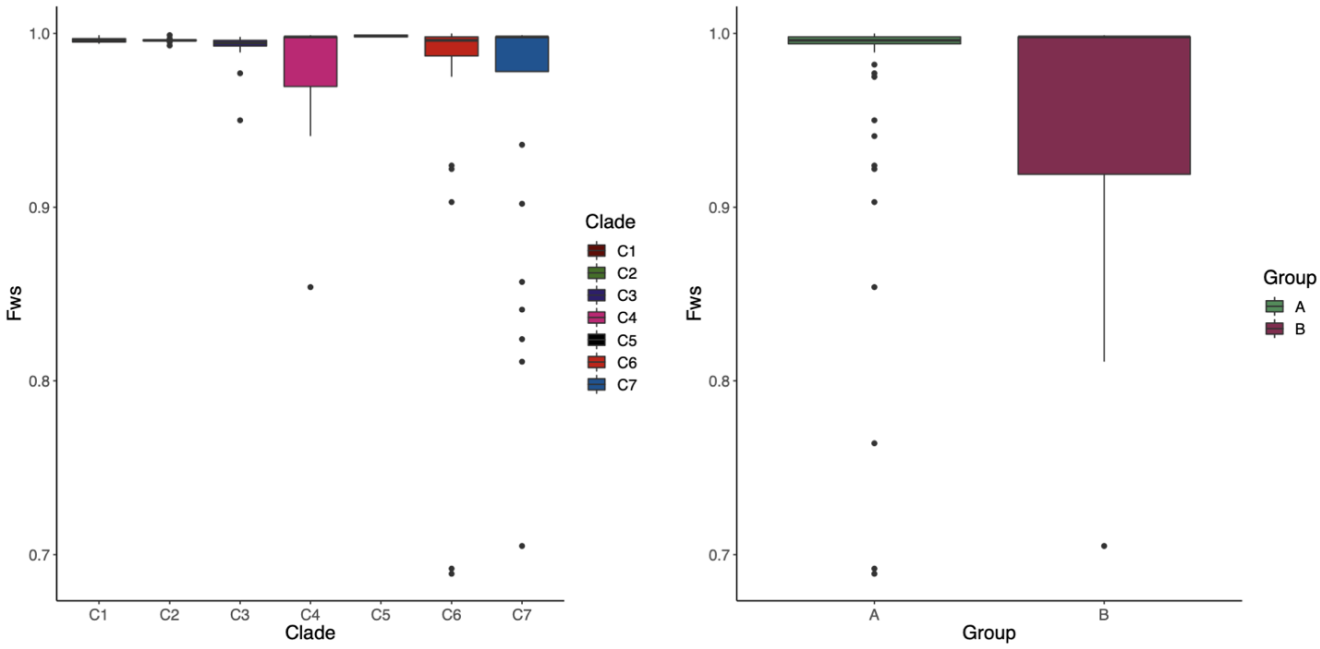
S5 Figure. Brazilian isolates split into two population groups: Group A and Group B.

A) The location of all Brazilian isolates within the filtered genomics database are shown on a map of Brazil with the states where isolates were obtained labelled (AC; Acre, AMZ; Amazonas, RO; Rondônia, MG; Mato Grosso, PA; *Pará*, AMP; Amapá, SP; São Paulo). A green border separates the Southwestern portion of the map where isolates are assigned into Group A. A purple border highlights the northern portion of Brazil where isolates are assigned to Group B. B) The Maximum Likelihood phylogenetic tree of 123 isolates from Brazil, comprising 70,757 SNPs with the corresponding borders to highlight isolates from group A (green border, N = 88 isolates) and group B (purple border, N = 23 isolates). Isolates from São Paulo are highlighted in a black circle labelled 'SP' and are excluded from these population groupings.

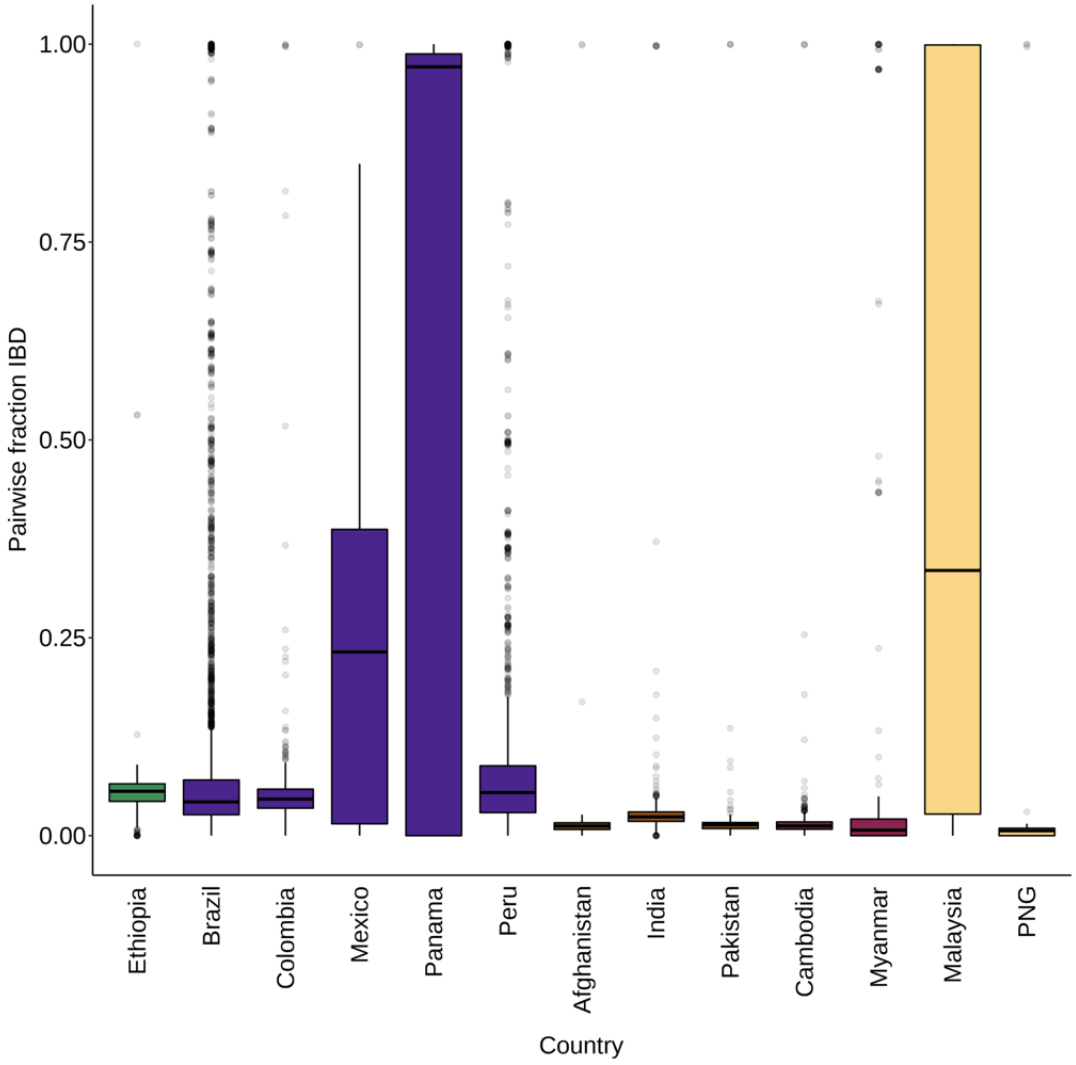


S6 Figure. Multiplicity of infection within global isolates by country.

Boxplot representing F_{WS} for *P. vivax* isolates across all countries with more than 10 isolates. Isolates are grouped by country and coloured according to region. All boxplot boxes consist of the median and interquartile range of the data for each country, with whiskers extending to extreme data points within 1.5 times the interquartile range from the box, data points beyond this range are outliers and plotted as points.

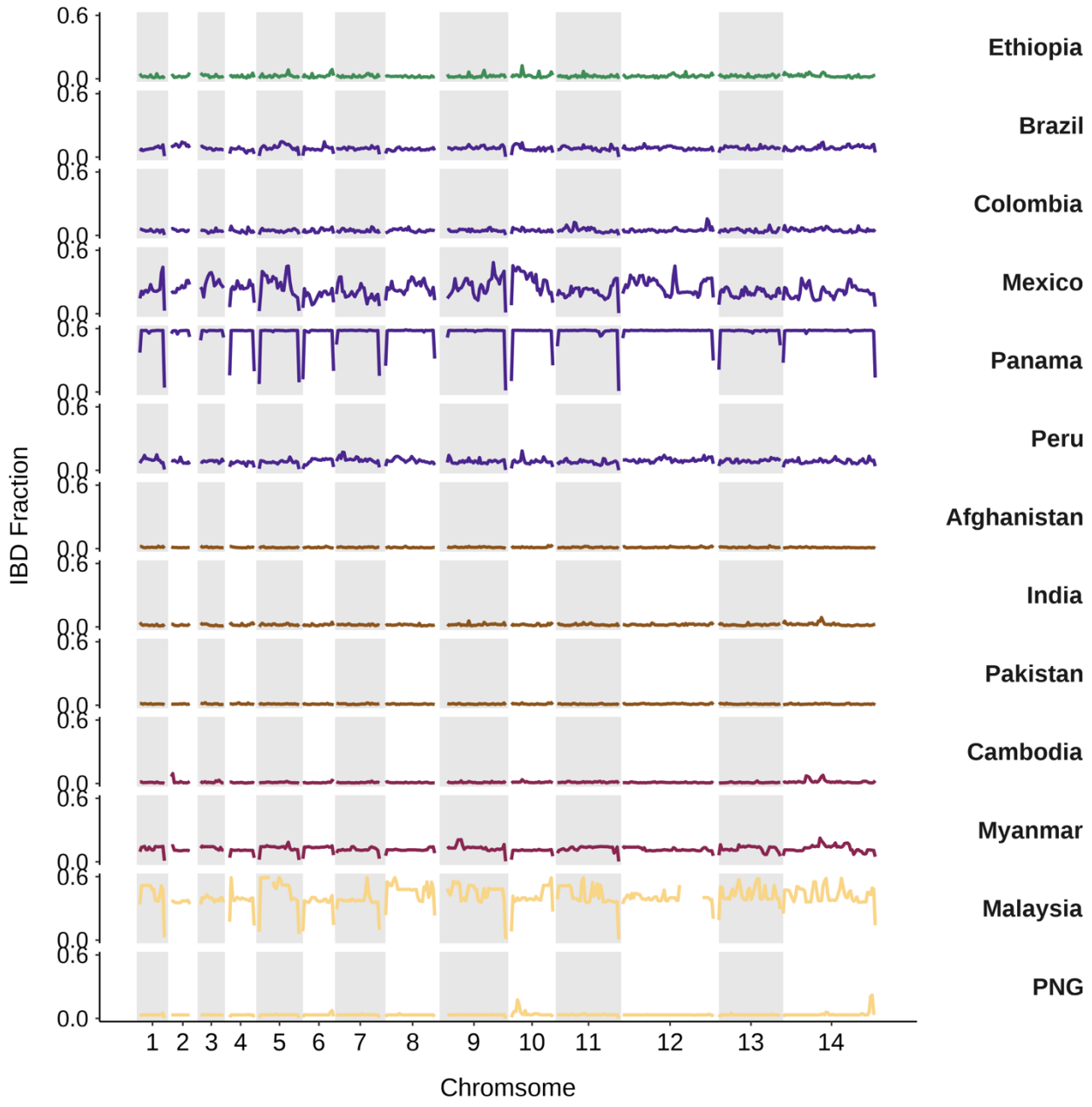


S7 Figure Multiplicity of infection in isolates within Brazil groupings.
Boxplot representing F_{WS} for Brazilian *P. vivax* isolates split into **A)** Clade groupings 1-7 and **B)** Geographical groupings A and B. All boxplot boxes consist of the median and interquartile range of the data for each country, with whiskers extending to extreme data points within 1.5 times the interquartile range from the box, data points beyond this range are outliers and plotted as points.



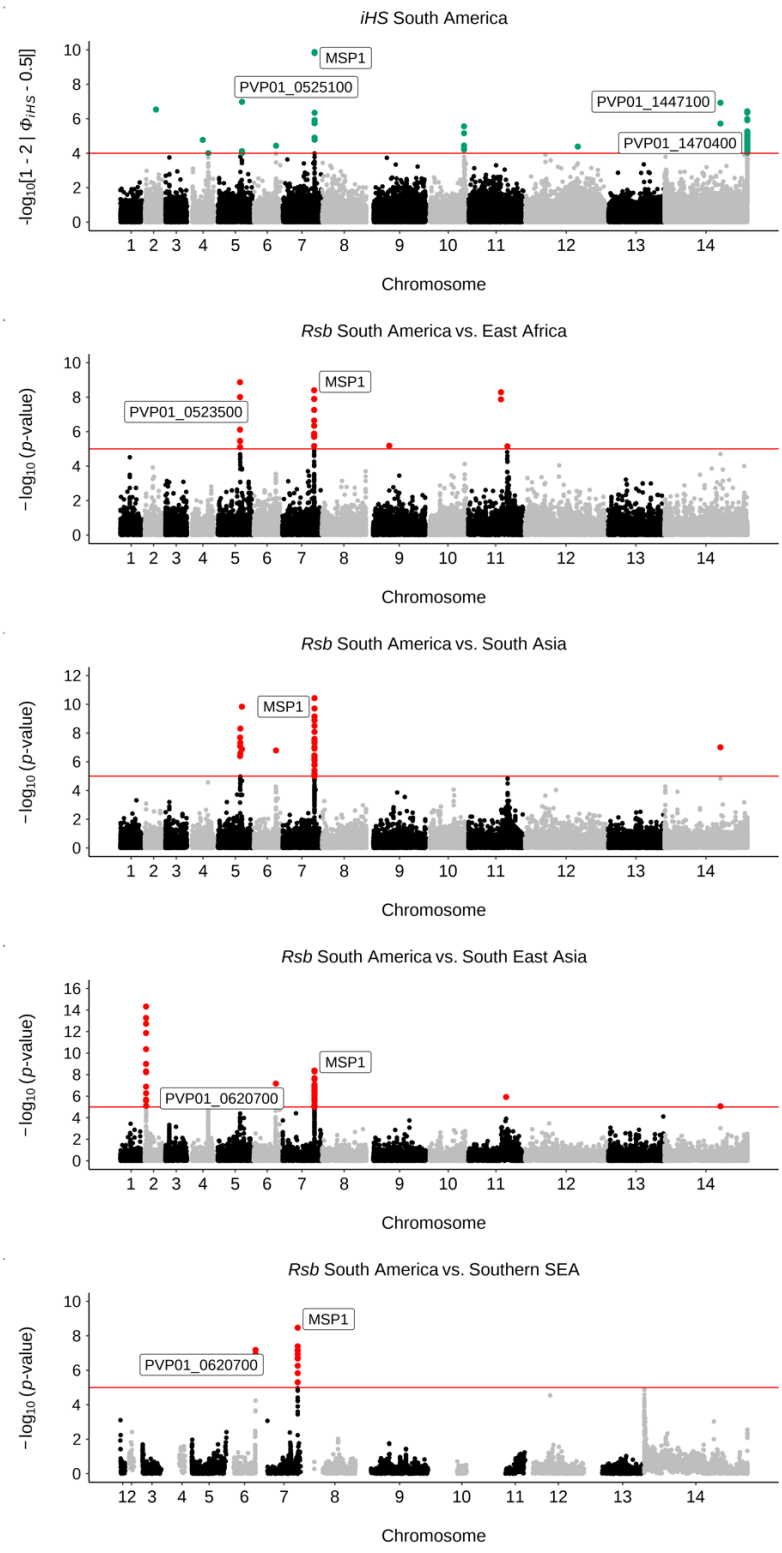
S8 Figure. Pairwise IBD fractions within each country's *P. vivax* population.

Pairwise comparisons of IBD fractions amongst all isolates within each country (where there are >10 isolates per country). All boxplot boxes consist of the median and interquartile range of the data for each country, with whiskers extending to extreme data points within 1.5 times the interquartile range from the box, data points beyond this range are outliers and plotted as points. Median IBD value for each country, and the total number of isolates for each country included in IBD analysis is summarised in **S9 Table**. Boxplot boxes are coloured according to the assigned geographical region (East Africa = green, South America = purple, South Asia = brown, Southeast Asia = pink, Southern Southeast Asia = yellow).

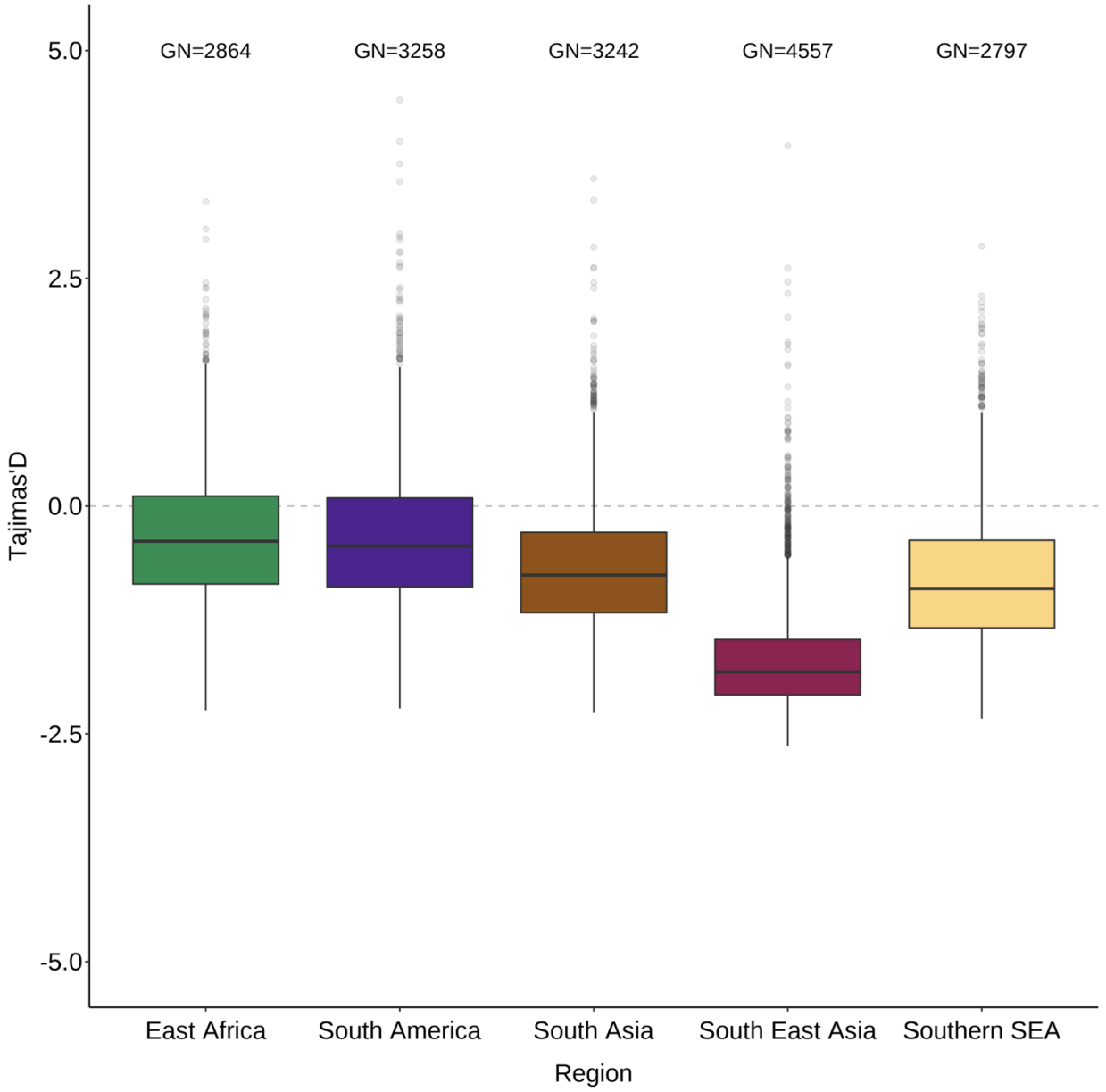


S9 Figure. Genome-wide analysis of IBD fractions across each country.

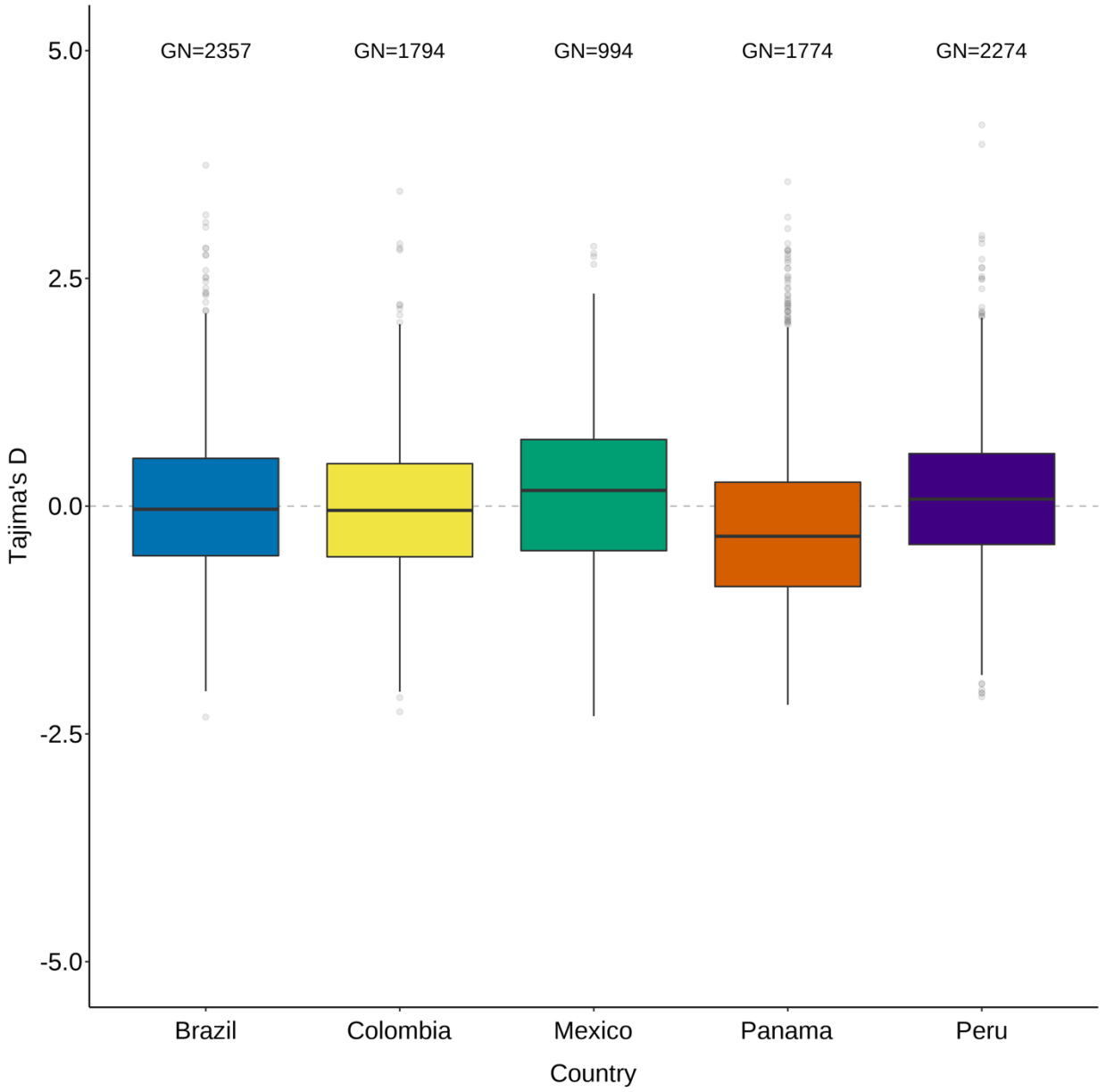
Using a sliding window of 50 kb, the genome wide IBD fractions are summarised for each country with >10 isolates. Line graphs are coloured according to the assigned geographical region (East Africa = green, South America = purple, South Asia = brown, Southeast Asia = pink, Southern Southeast Asia = yellow). Genome regions containing the top 1% of IBD fractions for each country are highlighted in **S8 Table**.



S9 Figure. Selection signals (iHS) across the global dataset at the regional level.



S10 Figure. Distribution of Tajima's D metric across all global regions.
Boxplot demonstrating the distribution of Tajima's D metric in each global region. Boxes are colour coded according to the geographical region (green = East Africa, purple = South America, brown = South Asia, pink = SEA, cream = SSEA) with the number of genes used in the Tajima D calculation (number of genes with > 5 SNPs in monoclonal isolates) is annotated at the top.



S11 Figure. Distribution of Tajima's D metric across all countries within South America. Boxplot demonstrating the distribution of Tajima's D metric in each global region. Boxes are colour coded according to the country within South America (blue = Brazil, yellow = Colombia, green = Mexico, orange = Panama, purple = Peru) with the number of genes used in the Tajima D calculation (number of genes with > 5 SNPs in monoclonal isolates) is annotated at the top.

CHAPTER 6

6 Discussion and conclusions

6. 1 Discussion

Genomic analysis of Plasmodium parasite species has been pivotal to support malaria control [1], with large-scale studies of *P. falciparum* demonstrating high genetic variability between populations [2–4] and highlighting loci associated with drug resistance [5–7]. The genomes of the human infective *Plasmodium spp.* vary in length and are continually evolving to evade disease control programmes through mutations and recombination.

Widespread disease control methods to reduce the burden of malaria have been in place for decades, with the main approaches aiming to: 1) prevent malaria infection through preventing the bite of an infected mosquito, through either vector control methods to reduce the mosquito population or physical barriers such as bed nets to hinder its ability to take a blood meal; 2) prevent malaria through chemoprevention, either at the level of large-scale drug treatments in populations at risk (such as intermittent preventative treatment in children, IPTi, or pregnant women, IPTp), or chemotherapy that individuals can take before travelling to a malaria endemic region to reduce the risk of infection; 3) prevent malaria and severe disease through rapid diagnosis and effective treatment, thereby reducing malaria transmission in a community. All these approaches can be thwarted by genetic variation in the Plasmodium parasite, including single nucleotide polymorphisms or duplications in drug resistance genes [5,8–10], or variants in the mosquito vector genomes, such as mutations and amplifications linked to insecticide resistance [11,12]. Regular monitoring of parasite genomes through sequencing is essential to infer important changes in genetic diversity, linked to parasite phenotypes.

Molecular screening of Plasmodium parasites has been a widely used research tool for diagnostics and surveillance, with capillary sequencing of specific genes of interest allowing for initial drug resistance screening in populations [13–15]. These molecular surveillance studies provide crucial information on the drug susceptibility profile of parasites in specific populations, and enable changes in drug treatment policies, in addition to aiding the planning of large-scale control programmes. Whilst capillary sequencing is a vital tool for initial genetic investigations, the advent in next generation sequencing technologies over the past decade has dramatically facilitated and reduced the cost of whole genome sequencing (WGS), which enables us to investigate genetic variation genome-wide, rather than choosing specific known loci. This increase in resolution is

particularly important when we are not sure which gene or locus underlies a particular parasite phenotype.

Whilst genomic research has predominantly focused on the *P. falciparum* parasite species, investigations of other human-infective Plasmodium is lacking. Prior to this thesis, there were > 7,000 publicly available WGS for *P. falciparum* [2], compared to > 1,000 for *P. vivax* [16], > 150 for *P. knowlesi* [17,18], 5 for *P. malariae* [19], 1 for *P. ovale wallikeri* and 1 for *P. ovale curtisi* [19]. The initial aim of my thesis was to increase the amount of genetic data available for understudied populations of Plasmodium parasites, which includes a global investigation into *P. malariae* isolates (novel sequence data; n = 235) and a specific analysis of Brazilian *P. vivax* isolates (novel sequence data; n = 89) across different endemic areas including the generation of novel sequence data for previously unstudied transmission regions in Brazil.

A fundamental barrier to the WGS of malaria parasites is the difficulty in generating sufficient high-quality parasite DNA, particularly as clinical blood samples contain leukocytes and large amounts of human DNA, which is detrimental to pathogen sequencing. In the case of *P. falciparum* infections, this process can be facilitated by the ability to culture parasites *in vitro* [20], allowing for an increase in the quantities of parasite DNA. With the exception of *P. knowlesi* [21], other human-infective *Plasmodium spp.* do not have a stable *in vitro* culture system. Other approaches, including leukocyte depletion [22] and selective whole genome amplification (SWGA) have been utilised to increase the relative amounts of parasite DNA obtained from clinical samples [23,24].

In **Chapter 2,** I described the development of a method for SWGA of *P. malariae* isolates, in particular, designing and testing primer sets that preferentially bind within the *P. malariae* genome in comparison to the human genome. I designed a set consisting of 5 primers which was tested using two isolates and demonstrated an 18.6-fold increase in the percentage of the genome covered by > sequencing 5 reads after WGS. The efficiency of amplification was affected by the initial parasitaemia of the isolate, consistent with findings in *P. falciparum* and *P. vivax* [23,24]. Using a range of parasitaemias, I determined a limit of 0.01% parasitaemia (400 parasites /μl) for effective SWGA. Using this validated primer set, I amplified and sequenced additional isolates, which after filtering, created a dataset of 18 high-quality isolates from Africa (n = 9, Kenya, Liberia, Sierra

Leone, Sudan, Uganda) and Asia (n = 9, Thailand). Using this initial dataset, I demonstrated isolation by distance, where African isolates clustered together, and independently from those from Asia. The *P. malariae* genome is large and complex to work with due to large repetitive subtelomeric regions with expansions of gene families including the *pmfam* genes [25], further complicating the analysis of short read sequence data. Using the sequence data generated from the 18 high-quality isolates, I determined the approximate genomic locations of the hypervariable subtelomeric regions to define the position of the core genome, further simplifying WGS analysis of short-read sequence data for future work on the *P. malariae* genome. I investigated both the presence of SNPs and the coverage after WGS for multiple genes associated with drug susceptibility and demonstrated the presence of SNPs within *pmdhfr* (3 nonsynonymous (NS)), *pmdhps* (1 NS) and *pmmdr1* (2 NS). Due to low coverage at the *pmcrt* locus after SWGA and WGS, it was not possible to explore potential SNPs within this gene across all samples, suggesting that further development of primers may be required to specifically increase the amplification at this locus [26].

Building on the SWGA work in **Chapter 2**, in **Chapter 3**, I generated a genomic dataset of 235 globally sourced *P. malariae* isolates. Sequence data was filtered to leave a final high-quality database of 155 isolates spanning 4 continents and 25 countries. Through population genomic analysis, I further verified that African isolates are distinct from those found in Asia. *P. malariae* isolates from Brazil demonstrated the highest levels of identity by decent, followed by Asia and Africa, with signals of homology in all three continents within chromosome 10 where a gene, *pmact*, whose ortholog is associated with multi-drug resistance in *P. falciparum,* is located [27]. When comparing populations, highly differentiating SNPs were found in genes encoding surface proteins (*pmcsp* and *pmmsp*9), in addition to loci associated with potential resistance pathways including *pmmrp2* [28]. I investigated the presence of SNPs within *P. malariae* orthologs of genes either predicted or confirmed to be associated with drug resistance and identified 286 SNPs leading to amino acid substitutions across 15 genes. Of note were 15 SNPs within *pmdhfr*, implicated in pyrimethamine susceptibility, which were taken forward for functional analysis in **Chapter 4**. Whilst SWGA has been a revolutionary tool for the generation of sufficient parasite DNA for WGS, there are limitations. For instance, 80 samples did not pass stringent quality filtering requirements, which may be addressed through additional optimisation of the SWGA protocol.

For *P. vivax*, an improved primer set was used to capture GC-rich regions of the genome [23], or through vacuum filtration of samples before amplification [29].

Using the findings within **Chapter 3**, I aimed to further investigate the phenotypic effects of SNPs found within *pmdhfr* with regards to pyrimethamine resistance, which is outlined in **Chapter 4**. Whilst there is no routine *in vitro* culture method for *P. malariae* parasites, recent studies using *P. knowlesi* have demonstrated the use of this parasite species to express orthologs of genes associated with drug resistance in an *in vitro* context [30]. Through ortholog replacement using CRISPR-Cas9 genome editing, I targeted the *pkdhfr* locus and replaced the DHFR domain with the orthologous *P. malariae* and *P. falciparum* DHFR version, alongside creating genotypes based on the mutated positions seen in the global dataset in **Chapter 3.** I demonstrate that ortholog replacement can be used to generate drug susceptibilities akin to those seen in *in vitro* studies using *P. knowlesi* and *P. falciparum* culture adapted parasite lines, with some variation likely due to multiple genes being involved in susceptibility (*P. knowlesi* 5.9-fold more susceptible to pyrimethamine than *P. falciparum* through ortholog replacement compared to 10.6-fold difference in *in vitro* culture studies [31]). I validate pyrimethamine sensitive lines using the wild-type *P. knowlesi* A1H1 culture, and this line after ortholog replacement with a recodonised version of the *P. knowlesi* DHFR domain, in addition to a version including the *P. falciparum* 3D7 DHFR sequence. Additionally, I validated the assay for determining pyrimethamine resistant lines, where the *P. falciparum* line with the IRN haplotype at positions 51, 59 and 108, shows pyrimethamine resistance as expected. In addition, I demonstrated that the PmUG01 reference strain displayed a pyrimethamine resistant phenotype, which is understandable when looking at the amino acid sequence, where the reference genome contains an asparagine at position 114 (N114) which aligns with the N108 position within *P.* falciparum, the key driver of pyrimethamine resistance [32]. The PmUG01 line without this substitution was sensitive to pyrimethamine.

Finally, within **Chapter 5,** I focus on the understudied parasite population of *P. vivax* within Brazil, where most previous sequence data has been generated from isolates sourced in Acre state. I collated the largest current genomic database for *P. vivax* isolates (n = 855 across 26 countries) and used this to confirm that South American isolates are a distinct global population. I generated novel WGS data for 89 isolates spanning 7 states within Brazil, and the analysis demonstrated a

complex population structure within Brazil, with multiple clades of transmission in the Amazonian regions (high transmission areas). Further, I identified South American signals of selection within genes associated with drug susceptibility including *pvmdr1* and *pvdhfr-ts*, and specifically signals of selection within *pvabci3*, an ABC transporter, potentially highlighting novel candidates for future functional work.

Overall, through the implementation of SWGA and WGS, I have increased the availability of genomic data for understudied groups of Plasmodium parasites. This is particularly seen with *P. malariae*, which, prior to my thesis, was greatly understudied with only 5 WGS publicly available [19,25], now increased 47-fold to 235 genome sequences, with 155 isolates after filtering for high quality genomic data. Genomic investigations of parasites have enabled fundamental discoveries regarding basic parasite biology, and by increasing the amount of data available, I develop a database which other researchers can utilise to investigate questions on *P. malariae* biology. *P. malariae* parasites demonstrate many unique features including the longer life cycle, the ability to remain in human hosts for decades and associations with nephropathologies, which using genomic data, we may uncover the genomic signatures that lead to these traits. My thesis uses this data specifically to investigate the potential for drug resistance in this parasite species, a concern due to reported treatment failures with both chloroquine [33] and ACTs [34,35]. After identifying potentially relevant variants within *pmdhfr-ts*, I developed an ortholog replacement system to determine the phenotypic effects of these variants. There is no routine *in vitro* culture method for *P. malariae* parasites, which complicates the ability to assay phenotypes associated with variants found through the genomics database. I wanted to not only identify variants, but to also take genomics through to functional verification, and to do this we developed a platform using ortholog replacement of *P. knowlesi* parasites. *P. knowlesi* is routinely grown *in vitro* [21] and ortholog replacement has been previously used to investigate invasion properties of *P. vivax* parasites, which also cannot be cultured *in vitro* [36]. Through orthologue replacement of *pkdhfr-ts*, I demonstrate a pyrimethamine resistant genotype of *pmdhfr-ts*, which is seen in the global genomic database and is therefore relevant for the treatment of *P. malariae* infections. Not only is this finding relevant, but the validation of this system has greater impact for potential future studies, whereby any interesting variant found within the *P. malariae* database could potentially be assayed for its phenotype using

cultures of *P. knowlesi*, whether in regard to drug resistance phenotypes, invasion of red blood cells, potential vaccine candidates, or many other avenues of research.

6.2 Conclusions

This thesis presents a genomic analysis of two understudied species of malaria parasites: *P. malariae* and *P. vivax*. I validate a method of SWGA to enable WGS of *P. malariae* isolates, which has previously been difficult, in part due to low parasite density in clinical infections. Through developing this methodology, previous questions about the *P. malariae* genome can now be investigated, and I have performed the first global population genomics investigation into this parasite.

WGS is a significant tool for disease control, and can be utilised to investigate transmission intensity, drug resistance, potential vaccine efficacy and the impact of disease interventions amongst other things, and these approaches have been utilised in the case of *P. falciparum* infections, where vast amounts of genomic data are available [37]. By increasing the WGS data available for understudied parasite populations, I enable further investigations into their unique biology, which is increasingly important in understudied populations of malaria parasites. Previous control methods targeting malaria have been designed against *P. falciparum* and are likely to be less effective on non-falciparum malaria species, where further knowledge is needed to ascertain how best to target and control these parasite species.

In summary, the malaria genomics field is rapidly advancing due to technological change and cost reductions, leading to the possibility of performing large-scale WGS to gain a baseline understanding of each species' unique biology, as well as the molecular evolution and genomic landscape of each parasite population within different regions.

6.3 Future of malaria genomics and future work

Genomics is a useful tool for infectious disease research that is becoming increasingly accessible due to advances in sequencing technologies, which have reduced the cost and created more

accessible methodologies for WGS. Genomic data has provided significant insights into Plasmodium parasite biology, and has been particularly instrumental in determining the mechanisms of drug resistance [5,10,38], a major challenge in malaria eradication.

WGS data can be utilised in many ways, including large-scale genomic studies that are paired with clinical data to identify genotypes associated with disease phenotypes, as in the case of genome wide association studies (GWAS) [39], large scale population genomics studies to identify distinct populations of parasites and signals of selection that differ between regions [16,17], and assessment of clinical isolates in the context of an outbreak to determine the origin of infection [40]. The development of portable sequencing devices such as the Oxford Nanopore Technologies MinION platform, and its validation for use in *P. falciparum* isolates [41] provides exciting opportunities for malaria genomics and control. Long read sequence data can be generated in real time, allowing for rapid identification of parasite species, and investigation into specific SNPs and gene amplifications of interest. This may be particularly useful in disease outbreak scenarios and in areas monitoring for malaria elimination, where genome data could potentially be used to identify whether infections originated from local transmission or importation in real-time.

The COVID-19 global pandemic has had a detrimental impact on healthcare services and the control of other infectious diseases, which has been demonstrated by an increase in malaria incidence shown in the 2021 World Malaria Report [42]. Prior to the COVID-19 pandemic, there were reductions in funding for malaria control, and the subsequent pandemic has strained healthcare systems, and led to delays in delivering vital anti-malarials and bed nets for routine control programmes [42]. These increased difficulties in malaria control are accompanied by the threat of artemisinin resistance emergence in Africa [43]. Whilst novel antimalarial drugs are in development, none have been accepted for medical use, and continued surveillance of Plasmodium parasites is essential to ascertain the extent of artemisinin resistance in differing regions.

Future concerns for malaria control include non-falciparum parasite species, which are likely to be more difficult to eradicate due to higher chances of asymptomatic silent infections as observed for *P. malariae*, and in the case of *P. vivax* and *P. ovale* species; the development of dormant hypnozoite stages which can cause disease relapses (if not treated with radical cure therapy). Of

particular note are the reported treatment failures with non-falciparum malaria parasites, coupled with mutations found within genes associated with drug susceptibility in *P. vivax* [44] and in *P. malariae* as described in **Chapter 3**. This thesis demonstrates not only the importance of using genomic screening to identify novel mutations in genes associated with drug susceptibility but goes further to develop an *in vitro* system for testing candidate genes and SNPs found within genomic studies for non-falciparum malaria parasites that have no routine *in vitro* culture method (**Chapter 4**). Future work will continue to use this system of culture-adapted parasites coupled with genomic screens of WGS data to identify novel mutations in candidate genes and create resulting parasite lines to investigate the effect of polymorphisms on important phenotypes such as parasite drug susceptibility and invasion of red blood cells. Such work and its insights will assist the development of tools for malaria contrsol, such as diagnostics, treatments, and vaccines.

6.4 References

1. Grad, Y. H. & Lipsitch, M. Epidemiologic data and pathogen genome sequences: a powerful synergy for public health. *Genome Biol.* **15**, (2014).
2. Gen, M. *et al.* An open dataset of *Plasmodium falciparum* genome variation in 7,000 worldwide samples. *Wellcome Open Res. 2021 642* **6**, 42 (2021).
3. Manske, M. *et al.* Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. *Nature* **487**, 375–379 (2012).
4. C. de Oliveira, T. *et al.* Genome-wide diversity and differentiation in New World populations of the human malaria parasite Plasmodium vivax. *PLoS Negl. Trop. Dis.* **11**, (2017).
5. Ariey, F. *et al.* A molecular marker of artemisinin-resistant Plasmodium falciparum. *Nature* **505**, 50–55 (2016).
6. Takala-Harrison, S. *et al.* Genetic loci associated with delayed clearance of plasmodium falciparum following artemisinin. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 240–245 (2013).
7. Miotto, O. *et al.* Multiple populations of artemisinin-resistant Plasmodium falciparum in Cambodia. *Nat. Genet.* **45**, 648–55 (2013).
8. Sidhu, A., Verdier-Pinard, D. & Fidock, D. A. Chloroquine resistance in Plasmodium falciparum malaria parasites conferred by pfcrt mutations. *Science (80-.).* **298**, 210–213 (2002).
9. Le Bras, J. & Durand, R. The mechanisms of resistance to antimalarial drugs in Plasmodium falciparum. *Fundam. Clin. Pharmacol.* **17**, 147–153 (2003).
10. Heinberg, A. & Kirkman, L. The molecular basis of antifolate resistance in Plasmodium falciparum: looking beyond point mutations. doi:10.1111/nyas.12662
11. Clarkson, C. S., Temple, H. J. & Miles, A. The genomics of insecticide resistance: insights from recent studies in African malaria vectors. *Curr. Opin. Insect Sci.* **27**, 111 (2018).
12. Adolfi, A. *et al.* Functional genetic validation of key genes conferring insecticide resistance in the major African malaria vector, Anopheles gambiae. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 25764–25772 (2019).
13. Straimer, J., Gandhi, P., Renner, K. C. & Schmitt, E. K. High Prevalence of Plasmodium falciparum K13 Mutations in Rwanda Is Associated With Slow Parasite Clearance After Treatment With Artemether-Lumefantrine. *J. Infect. Dis.* (2021). doi:10.1093/INFDIS/JIAB352
14. Muwanguzi, J. *et al.* Lack of K13 mutations in Plasmodium falciparum persisting after artemisinin combination therapy treatment of Kenyan children. *Malar. J.* **15**, (2016).
15. Henriques, G. *et al.* Directional selection at the pfmdr1, pfcrt, pfubp1, and pfap2mu loci of Plasmodium falciparum in Kenyan children treated with ACT. *J. Infect. Dis.* **210**, 2001–2008 (2014).
16. Diez Benavente, E. *et al.* Distinctive genetic structure and selection patterns in Plasmodium vivax from South Asia and East Africa - under review - to update reference. *Nat. Commun.* **XX**, (2021).
17. Benavente, E. D. *et al.* Whole genome sequencing of amplified Plasmodium knowlesi DNA from unprocessed blood reveals genetic exchange events between Malaysian Peninsular and Borneo subpopulations. *Sci. Rep.* **9**, 9873 (2019).
18. 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 (2020).
19. Rutledge, G. G. *et al.* Plasmodium malariae and P. ovale genomes provide insights into malaria parasite evolution. *Nature* **542**, 101–104 (2017).
20. Trager, W. & Jensen, J. B. Human malaria parasites in continuous culture. *Science* **193**, 673–675 (1976).
21. 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).
22. Venkatesan, M. *et al.* Using CF11 cellulose columns to inexpensively and effectively remove human DNA from Plasmodium falciparum-infected whole blood samples. *Malar. J.* (2012). doi:10.1186/1475-2875-11-41
23. 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**, 1–15 (2017).
24. Oyola, S. O. *et al.* Whole genome sequencing of Plasmodium falciparum from dried blood spots using selective whole genome amplification. *Malar. J.* **15**, 597 (2016).
25. Ansari, H. R. *et al.* Genome-scale comparison of expanded gene families in Plasmodium ovale wallikeri and Plasmodium ovale curtisi with Plasmodium malariae and with other Plasmodium species. *Int. J. Parasitol.* **46**, 685–696 (2016).
26. Ibrahim, A. *et al.* Selective whole genome amplification of Plasmodium malariae DNA from clinical samples reveals insights into population structure. *Sci. Rep.* **10**, 1–11 (2020).
27. Lim, M. Y. X. *et al.* UDP-galactose and acetyl-CoA transporters as Plasmodium multidrug resistance genes. *Nat. Microbiol.* **1**, (2016).
28. Gil, J. P. & Fançony, C. Plasmodium falciparum Multidrug Resistance Proteins (pfMRPs). *Front. Pharmacol.* **12**, 2896 (2021).
29. Shah, Z. *et al.* Optimization of parasite DNA enrichment approaches to generate whole genome sequencing data for Plasmodium falciparum from low parasitaemia samples. *Malar. J.* **19**, 1–10 (2020).
30. Verzier, L. H., Coyle, R., Singh Id, S., Sanderson, T. & Id, J. C. R. Plasmodium knowlesi as a model system for characterising Plasmodium vivax drug resistance candidate genes. (2019). doi:10.1371/journal.pntd.0007470
31. 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).
32. Cowman, A. F., Morry, M. J., Biggs, B. A., Cross, G. A. M. & Foote, S. J. Amino acid changes linked to pyrimethamine resistance in the dihydrofolate reductase-thymidylate synthase gene of Plasmodium falciparum. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 9109–9113 (1988).
33. Maguire, J. D. *et al.* Chloroquine-resistant Plasmodium malariae in south Sumatra, Indonesia. *Lancet* (2002). doi:10.1016/S0140-6736(02)09336-4
34. Betson, M. *et al.* Detection of persistent Plasmodium spp. infections in Ugandan children after artemether-lumefantrine treatment. *Parasitology* **141**, 1880–1890 (2014).
35. Betson, M., Clifford, S., Stanton, M., Kabatereine, N. B. & Stothard, J. R. Emergence of Nonfalciparum Plasmodium Infection Despite Regular Artemisinin Combination Therapy in an 18-Month Longitudinal Study of Ugandan Children and Their Mothers. *J. Infect.*

Dis. **217**, 1099–1109 (2018).
36. 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).
37. Neafsey, D. E. & Volkman, S. K. Malaria Genomics in the Era of Eradication. *Cold Spring Harb. Perspect. Med.* **7**, (2017).
38. Djimde, A. A. *et al.* A molecular marker for chloroquine-resistant falciparum malaria. *N. Engl. J. Med.* **344**, 257–263 (2001).
39. Miotto, O. *et al.* Genetic architecture of artemisinin-resistant Plasmodium falciparum. *Nat. Genet.* **47**, 226–34 (2015).
40. Da Veiga Leal, S. *et al.* Drug resistance profile and clonality of Plasmodium falciparum parasites in Cape Verde: the 2017 malaria outbreak. *Malar. J.* **20**, (2021).
41. Razook, Z. *et al.* Real time, field-deployable whole genome sequencing of malaria parasites using nanopore technology. *bioRxiv* 2020.12.17.423341 (2020). doi:10.1101/2020.12.17.423341
42. World Health Organization. *World Malaria Report 2021*. (2021).
43. Balikagala, B. *et al.* Evidence of Artemisinin-Resistant Malaria in Africa. *N. Engl. J. Med.* **385**, 1163–1171 (2021).
44. Auliff, A. *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).