

***PLASMODIUM* INFECTIONS OF WILD APES AND HUMAN ZONOTIC RISK**

Dorothy Elizabeth Loy

A DISSERTATION

in

Cell and Molecular Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2018

Supervisor of Dissertation

Beatrice H. Hahn

Professor of Medicine

Graduate Group Chairperson

Daniel Kessler

Associate Professor of Cell and Developmental Biology

Dissertation Committee

Phillip Scott, Professor of Microbiology and Immunology

Frederic D. Bushman, William Maul Measey Professor of Microbiology

Sara Cherry, Professor of Microbiology

Michael Povelones, Assistant Professor of Pathobiology

David S. Roos, E. Otis Kendall Professor of Biology

PLASMODIUM INFECTIONS OF WILD APES AND HUMAN ZOONOTIC RISK

COPYRIGHT

2018

Dorothy Elizabeth Loy

This work is licensed under the
Creative Commons Attribution-
NonCommercial-ShareAlike 3.0
License

To view a copy of this license, visit

<https://creativecommons.org/licenses/by-nc-sa/3.0/us/>

ACKNOWLEDGMENT

I thank my mentor, Beatrice Hahn, for all the support she provided during my PhD training. Leading by example, she taught me to ask a lot of questions, to think critically, to communicate clearly, and to strive for perfection in my work. It has been an honor to work with such a curious and motivated scientist.

I also thank the students and staff in the Hahn lab for all their advice and moral support. In particular, I thank Yingying Li who taught me many experimental techniques and importance of a clean workspace, Weimin Liu who generated sequences from fecal samples when no one else could, Sesh Sundararaman who taught me genome assembly starting with opening my first terminal session, Jerry Learn who made and discussed countless phylogenetic trees, Alexa Avitto and Meagan Rubel who dove into our *Plasmodium* screen with enthusiasm, Fred Bibollet-Ruche who helped obtain and process ape blood samples, Scott Sherrill-Mix and Jesse Connell who answered all my statistics and programming questions, and Shivani Sethi who taught me how to present data in an attractive way. I am especially grateful to Paul Sharp and Lindsey Plenderleith, who analyzed and interpreted much of the ape *P. vivax* data with me. I can't thank Lindsey enough for her detailed answers to my many questions.

I thank the Penn faculty members who have mentored me along the way, including the members of my thesis committee: David Roos, Michael Povelones, Phillip Scott, Sara Cherry, and Rick Bushman. The scientific and professional guidance you provided was invaluable. In addition, I am grateful to Skip Brass, Maggie Krall, and Matt Deardorff for giving advice and encouragement during the transition from medical school to graduate school, and to Sarah Tishkoff for her willingness to collaborate and to share her precious samples.

I am incredibly grateful to the community of family and friends that surrounds me. Without their encouragement I might not have written any of the words that follow. And many thanks to Tom Gemmer for believing in me, for unwavering support while I worked through weekends and holidays, and for suggesting that I take a break to get out into the mountains from time to time.

ABSTRACT

PLASMODIUM INFECTIONS OF WILD APES AND HUMAN ZOONOTIC RISK

Dorothy Elizabeth Loy

Beatrice H. Hahn, MD

Plasmodium falciparum and *Plasmodium vivax* cause over 95% of all human malaria infections. To control and potentially eliminate these pathogens, it is important to understand their origins and evolutionary history. The recent discovery of a multitude of *Plasmodium* species in apes revealed that *P. vivax* and *P. falciparum* evolved from parasites infecting African apes, but the zoonotic threat posed by ape parasites and the precise circumstances surrounding the emergence of *Plasmodium* in humans remain unknown. Thus, in this thesis, I asked two questions: Are humans exposed to ape parasites and what can be learned about the history of human *P. vivax* through analyses of related ape parasites. To address the first question, I asked whether humans living near *Plasmodium*-infected apes develop pre-erythrocytic infections in the absence of blood stage infections. Screening 504 Cameroonian fecal samples for ape *Plasmodium* species, I found no evidence of abortive liver infection. Next, to facilitate genome sequencing of ape *P. vivax*, I adapted selective whole genome amplification (SWGA) to *P. vivax*, achieving a dramatic increase in the proportion of *P. vivax* DNA in human samples without introducing systemic sequence errors. I then generated partial *P. vivax* genome sequences from six chimpanzees and one gorilla, which revealed that human strains of *P. vivax* exhibit ~10-fold less diversity and have a unique excess of nonsynonymous nucleotide polymorphisms. This suggests a recent bottleneck and greatly relaxed purifying selection in the human parasite lineage. Investigating potential host specificity determinants, I found that ape *P. vivax* parasites encode three reticulocyte binding protein

genes (*rbp2d*, *rbp2e*, and *rbp3*) whose orthologs are pseudogenes in human *P. vivax* strains. However, recombinant RBP2e and RBP3 proteins bound human, chimpanzee, and gorilla erythrocytes with similar efficiency. These results suggest that the *P. vivax* ancestor infected humans and apes in Africa, and that modern human *P. vivax* is derived from parasites that escaped Africa. Although many questions remain concerning the biology and zoonotic potential of ape malaria parasites, my studies show that comparative genomics, coupled with functional parasite studies, can yield new insights that are relevant to the prevention and eradication of human malaria.

TABLE OF CONTENTS

ACKNOWLEDGMENT	III
ABSTRACT	IV
LIST OF TABLES	VIII
LIST OF ILLUSTRATIONS	IX
CHAPTER 1: OUT OF AFRICA: ORIGINS AND EVOLUTION OF THE HUMAN MALARIA PARASITES <i>PLASMODIUM FALCIPARUM</i> AND <i>PLASMODIUM VIVAX</i> ...	1
1.1 Introduction	2
1.2 Early studies of ape <i>Plasmodium</i> infections	3
1.3 Seven <i>Laverania</i> spp. in wild-living apes.....	5
1.4 Origin of <i>P. falciparum</i> in western gorillas and emergence in humans.....	14
1.5 A sylvatic resevoir of <i>P. vivax</i> and African origin of human <i>P. vivax</i>	16
1.6 Zoonotic potential of ape parasites	21
1.7 Emergence of <i>P. vivax</i> in human populations	22
1.8 Chapter references	23
CHAPTER 2 - INVESTIGATING ZOOONOTIC INFECTION BARRIERS TO APE <i>PLASMODIUM</i> PARASIES USING FECAL DNA ANALYSIS.....	29
2.1 Abstract.....	30
2.2 Introduction	31
2.3 Materials and methods.....	34
2.4 Results	40
2.5 Discussion	56
2.6 Chapter acknowledgments	61
2.7 Chapter references	62
CHAPTER 3 - SELECTIVE WHOLE-GENOME AMPLIFICATION IS A ROBUST METHOD THAT ENABLES SCALABE WHOLE-GENOME SEQUENCING OF <i>PLASMODIUM VIVAX</i> FROM UNPROCESSED CLINICAL SAMPLE.....	69
3.1 Abstract.....	70

3.2 Introduction	70
3.3 Results	73
3.4 Discussion	87
3.5 Materials and methods.....	89
3.6 Chapters acknowledgments	95
3.7 Supplemental figures and tables	97
3.8 Chapter references	105
CHAPTER 4 - EVOLUTIONARY HISTORY OF HUMAN <i>PLASMODIUM VIVAX</i> REVEALED BY GENOME-WIDE ANALYSIS OF RELATED APE PARASITES.....	111
4.1 Abstract.....	112
4.2 Introduction	112
4.3 Results	115
4.4 Discussion	132
4.5 Materials and methods.....	137
4.6 Chapter acknowledgments	142
4.7 Supplemental Information Appendix	143
4.8 Chapter References	182
CHAPTER 5 – SUMMARY AND FUTURE DIRECTIONS.....	190
5.1 Exposure of humans to ape <i>Laverania</i> and <i>P. vivax</i> parasites	191
5.2 <i>P. malariae</i> and <i>P. ovale</i> -like parasite species in African great apes.....	194
5.3 Evolutionary history of <i>P. vivax</i> in humans	196
5.4 No evidence of positive selection unique to the human <i>P. vivax</i> lineage	197
5.5 Loss of three invasion genes in human <i>P. vivax</i>	198
5.6 <i>P. vivax</i> parasites in gorillas and bonobos	200
5.7 <i>P. vivax</i> transmission in modern African humans	201
5.8 Summary.....	202
5.9 Chapter References	203

LIST OF TABLES

Table 1-1 Feces-based prevalence estimates of <i>Laverania</i> and <i>Plasmodium vivax</i> infections in wild-living African apes.....	12
Table 2-1 Detection of <i>Plasmodium</i> DNA in fecal samples of subjects from rural Cameroon	43
Table 2-2 Relative parasite burden in <i>Plasmodium</i> -positive human fecal samples	45
Table 2-3 <i>Plasmodium</i> spp. in the blood of fecal <i>Plasmodium</i> -negative individuals	53
Table 2-4. Prevalence of gastrointestinal parasites in fecal <i>Plasmodium</i> -positive and -negative subjects	54
Table 3-1. Sequencing statistics for <i>P. vivax</i> sequences from clinical samples that underwent selective whole-genome amplification (SWGA).....	80
Table 3-2 SNPs in SWGA versus leukocyte filtered whole genome sequences.....	83
Table 3-3 Nonsynonymous SNPs in known drug resistance genes.	84
Table 3-3 Clonality estimates post-SWGA.	85
Table 4-1 Features of assembled ape <i>P. vivax</i> genomes.....	116
Table 4-2. Nucleotide polymorphism in ape and human <i>P. vivax</i>	118

LIST OF ILLUSTRATIONS

Figure 1-1 Evolutionary relationships of <i>Plasmodium</i> spp.	7
Figure 1-2. Geographic distribution of <i>Laverania</i> and <i>Plasmodium vivax</i> infections in wild-living apes.	10
Figure 1-3. Evolutionary relationships of ape and human <i>Laverania</i> parasites.....	13
Figure 1-4. Evolutionary relationships of ape and human <i>Plasmodium vivax</i> parasites.	18
Figure 2-1. Human sampling sites in Cameroon	42
Figure 2-2. <i>Plasmodium</i> spp. detected in the stool and blood of rural Cameroonians.	45
Figure 2-3. Single nucleotide variants distinguishing human and ape <i>Plasmodium malariae</i> parasites.	48
Figure 2-4. Detection of a porcupine parasite in human stool.	50
Figure 2-5. Fecal <i>Plasmodium</i> -positive individuals have higher blood parasitemia than fecal <i>Plasmodium</i> -negative individuals.....	54
Figure 3-1. Selective whole genome amplification (SWGA) of <i>Plasmodium vivax</i> genomic DNA (gDNA) from human blood samples.....	72
Figure 3-2. <i>Plasmodium vivax</i> chromosomal coverage following SWGA using primer set pvset1 or pvset1920	74
Figure 3-3. Testing of SWGA primer sets on DNA from an unprocessed, <i>P. vivax</i> -infected blood sample.....	75
Figure 3-4. Comparison of <i>P. vivax</i> genome coverage from samples treated with SWGA and with leukocyte filtration	80
Figure 3-5. GC bias plots for <i>P. vivax</i> genomes generated following leukocyte filtration or SWGA.	81
Figure 3-6. Neighbor-joining tree of <i>P. vivax</i> clinical samples from different regions of the world.	85
Figure 4-1. Nucleotide diversity of ape and human <i>P. vivax</i>	118
Figure 4-2. Relationships among <i>P. vivax</i> strains from apes and humans	123
Figure 4-3. The <i>rbp</i> gene family in ape and human <i>P. vivax</i>	127
Figure 4-4. Binding of RBPs to ape and human <i>P. vivax</i>	130

CHAPTER 1 – Out of Africa: Origins and evolution of the human malaria parasites

Plasmodium falciparum* and *Plasmodium vivax

Dorothy E. Loy^{a,b}, Weimin Liu^a, Yingying Li^a, Gerald H. Learn^a, Lindsey J. Plenderleith^c,
Sesh A. Sundararaman^{a,b}, Paul M. Sharp^c, Beatrice H. Hahn^{a,b}

Departments of Medicine^a and Microbiology^b, Perelman School of Medicine, University
of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; ^cInstitute of Evolutionary
Biology, and Centre for Immunity, Infection and Evolution, University of Edinburgh,
Edinburgh EH9 3FL, UK;

Adapted from a review published in 2017 in *Int J Para* 47(2-3):87-97.

1.1 Introduction

Of the *Plasmodium* species known to commonly infect humans, *P. falciparum* and *P. vivax* cause the vast majority of malaria morbidity and mortality, and are the principal targets of malaria prevention and eradication efforts. *P. falciparum* is highly prevalent in sub-Saharan Africa where it is responsible for an estimated 216 million clinical cases and over 400,000 malaria-related deaths annually, predominantly in children under 5 years of age (World Health Organization, 2017). *P. vivax* is rare in sub-Saharan Africa, but endemic in many parts of Asia, Oceania, as well as Central and South America where it causes an estimated 8.6 million cases of clinical malaria, which represent about half of all malaria cases outside Africa (World Health Organization, 2017).

Given the devastating effects of malaria, the origins of the human *Plasmodium* parasites have long been a subject of interest. Descriptions of malaria-like illness can be found in ancient texts from China, India, the Middle East, Africa, and Europe, indicating that humans have been combatting *Plasmodium* infections through much of our recorded history (Carter and Mendis, 2002). Indeed, variants in the human genome that are associated with resistance to *Plasmodium* infection and malaria-associated disease are estimated to be thousands of years old (Hedrick, 2011). One such variant is the sickle cell trait, which is common in African populations and protects against fatal *P. falciparum* malaria (Taylor et al., 2012). Similarly, a mutation that abolishes the expression of the Duffy antigen receptor of chemokines on the surface of red blood cells (the so-called “Duffy negative phenotype”) approaches fixation in west and central Africa and confers almost complete protection from *P. vivax* parasitemia (Miller et al., 1976; Howes et al., 2011). Together, these findings indicate that *Plasmodium* infections have impacted human health for millennia, but the prevailing view has been that this history goes back much further.

One long-standing hypothesis suggested that humans and chimpanzees each inherited *P. falciparum*-like infections from their common ancestor, and that these parasites co-evolved with their respective host species for millions of years (Escalante and Ayala, 1994). In contrast, *P. vivax* was believed to have arisen several hundred thousand years ago, following the cross-species transmission of a macaque parasite in Southeast Asia (Escalante et al., 2005; Jongwutiwes et al., 2005; Mu et al., 2005; Neafsey et al., 2012). However, both of these theories have recently been refuted following the characterization of a large number of additional *Plasmodium* parasites from African apes. Specifically, it is now clear that *P. falciparum* is a relatively new infection of humans, which arose after the acquisition of a gorilla parasite, likely within the past 10,000 years (Liu et al., 2010a; Sundararaman et al., 2016). Similarly, *P. vivax* did not emerge in Asia, but represents a bottlenecked lineage that escaped out of Africa before the spread of Duffy negativity rendered African humans resistant to *P. vivax* (Liu et al., 2014). In this chapter, I describe the findings that led to this new understanding and summarize what is known about the epidemiology, vector tropism, zoonotic potential, and pathogenicity of the ape precursors of the human parasites.

1.2 Early studies of ape *Plasmodium* infections

The first indication that African apes harbor *Plasmodium* infections was the finding of three morphologically distinct forms of parasites in the blood of wild-caught chimpanzees (*Pan troglodytes*) and western gorillas (*Gorilla gorilla*) in Cameroon (Reichenow, 1920). Microscopic characterization identified ape parasites that resembled human *P. falciparum*, *P. malariae*, and either *P. ovale* or (the similar) *P. vivax*, suggesting the existence of distinct *Plasmodium* species, which were classified as *P. reichenowi*, *P. rhodaini*, and *P. schwetzi*, respectively (Sluiter et al., 1922; Brumpt, 1939). Moreover, *P. falciparum* and

P. reichenowi were found to differ substantially from the other *Plasmodium* species in both life cycle and gametocyte morphology, prompting their placement into a separate subgenus, termed *Laverania* (Bray, 1958; Coatney et al., 1971). The existence of two divergent clades of malaria parasites infecting primates was subsequently confirmed when the various *Plasmodium* species were first molecularly characterized (Figure 1-1). Comparing rRNA small subunit gene sequences, Escalante and Ayala (1994) showed that among the known species *P. falciparum* and *P. reichenowi* were each other's closest relatives, and that both were only distantly related to other *Plasmodium* species. Assuming that rRNA gene sequences in *Plasmodium* species evolved at the same rate as had been estimated for some bacteria, it was inferred that *P. falciparum* and *P. reichenowi* had diverged ~10 million years ago, close to the time of the human-chimpanzee common ancestor. This led to the conclusion that human and chimpanzee parasites had co-diverged with their respective hosts (Escalante and Ayala, 1994). Due to a lack of preserved material, gene sequences from *P. schwetzi* and *P. rhodaini* were never determined, and so their relationship to other malaria parasites remains unknown.

Interest in ape *Laverania* infections was rekindled in 2009 when Ollomo and colleagues found parasites morphologically similar to *P. reichenowi* in the blood of two pet chimpanzees from Gabon (Ollomo et al., 2009). Analysis of mitochondrial DNA (mtDNA) sequences revealed that these parasites were related to, but divergent from, *P. falciparum* and *P. reichenowi*, suggesting the existence of a third *Laverania* species which they named *P. gaboni* (Ollomo et al., 2009). Follow-up studies of additional captive and wild apes confirmed a greater diversity of *Laverania* parasites, but interpretations differed as to the number of species and their host associations. Amplifying mtDNA and nuclear gene sequences of parasites from members of two chimpanzee subspecies, Rich and colleagues identified several distinct *Laverania* lineages, but chose to consider all of them

as “*P. reichenowi*”, even though one of these new lineages corresponded to *P. gaboni* (Rich et al., 2009). In contrast, Krief and colleagues classified a similar diversity of chimpanzee parasites into three species, termed *P. reichenowi*, *P. billcollinsi*, and *P. billbrayi*, where the latter corresponded to *P. gaboni* (Krief et al., 2010). These investigators also amplified *P. falciparum* mtDNA from the blood of captive bonobos (*Pan paniscus*), concluding that this ape species represents the likely source of the human infection (Krief et al., 2010). Finally, Prugnolle and colleagues developed non-invasive methods that permitted parasite detection in ape fecal samples, which identified diverse *Laverania* lineages not only in chimpanzees but also in western gorillas (Prugnolle et al., 2010). However, these investigators classified all chimpanzee parasites as either *P. reichenowi* or *P. gaboni*, and concluded that *P. falciparum*-like sequences found in fecal samples of wild-living western gorillas indicated ongoing transmission from humans to gorillas (Prugnolle et al., 2010). The consensus of these studies was that wild-living apes harbor a much greater diversity of *Laverania* parasites than previously recognized. However, there was disagreement concerning the number of ape *Laverania* species as well as the origin of *P. falciparum*, with some investigators implicating chimpanzees (Rich et al., 2009; Duval et al., 2010; Prugnolle et al., 2010) and others bonobos (Krief et al., 2010) as the likely original source of the parasites now infecting humans.

1.3 Seven *Laverania* species in wild-living apes

The seemingly discrepant results from these early studies were reconciled by comprehensive studies of *Laverania* infections in wild-living apes, which employed improved fecal-based detection methods and targeted different regions of both organelle and nuclear parasite genomes (Liu et al., 2010a). One technical advance was the use of limiting dilution PCR (termed single genome amplification or SGA), which in contrast to

standard (bulk) PCR precludes the generation of *in vitro* recombinants that confound phylogenetic analyses (Liu et al., 2010b). Using this approach to characterize the molecular epidemiology of ape malaria, *Plasmodium* infections were found to be widespread in both chimpanzees and western gorillas, including parasites that were closely related to human *P. malariae*, *P. ovale* and *P. vivax* (Liu et al., 2010a). However, the great majority of parasite sequences grouped within one of three chimpanzee-specific or three gorilla-specific parasite lineages, with each clade being well supported and quite distinct from the others, pointing to the existence of six *Laverania* species in chimpanzees and gorillas (Figure 1-1). Subsequent surveys of wild-living apes in Gabon (Boundenga et al., 2015) and Cote d'Ivoire (Kaiser et al., 2010) confirmed these findings, demonstrating that chimpanzees and western gorillas represent a substantial *Laverania* reservoir. Recently, a comprehensive survey of wild-living bonobos revealed the existence of a seventh *Laverania* species, termed *P. lomaminensis* (Figure 1-1; Liu et al., 2017).

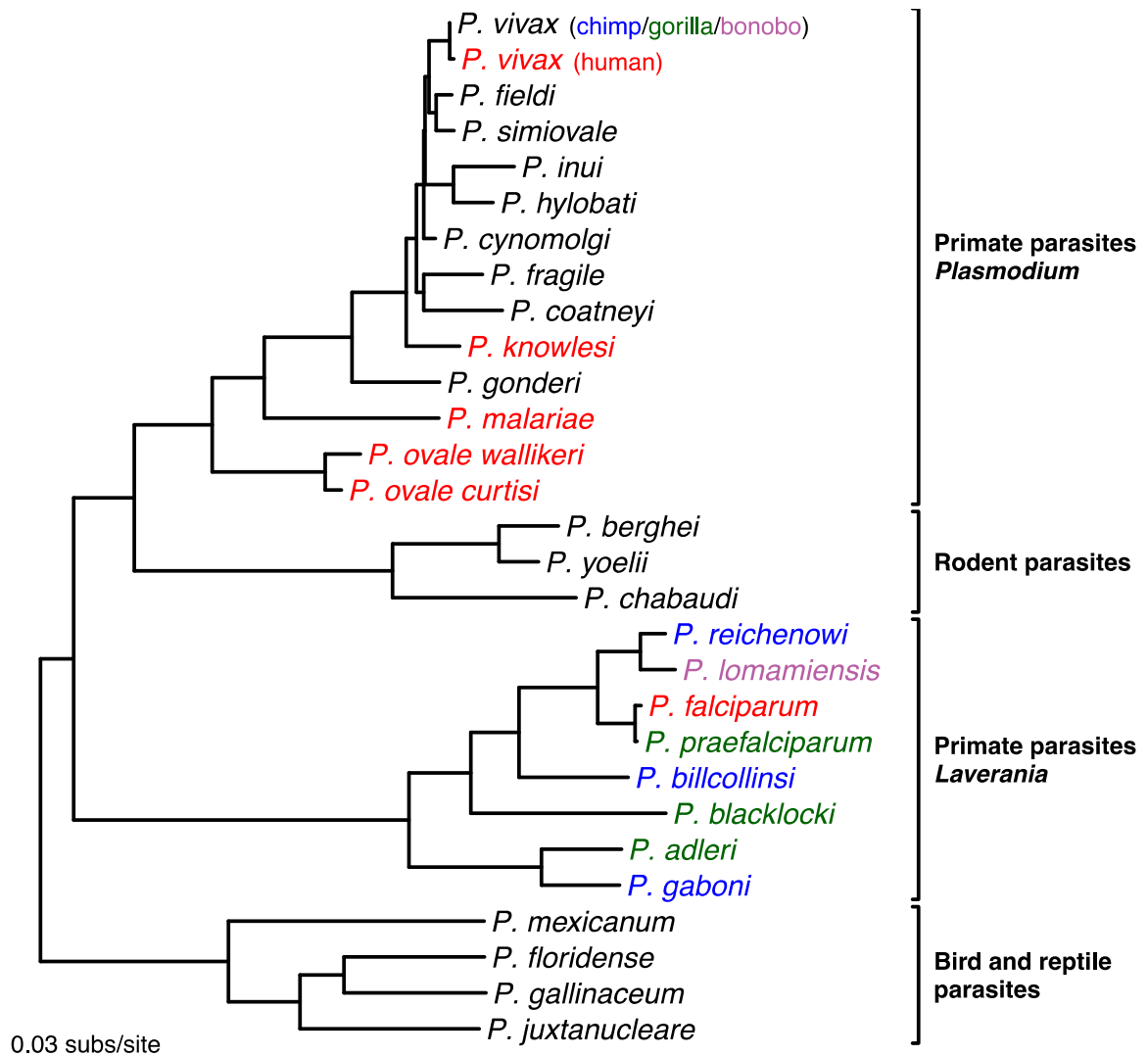


Figure 1-1. Evolutionary relationships of *Plasmodium* species. Colors highlight *Plasmodium* species that infect humans (red), chimpanzees (blue), bonobos (purple), and gorillas (green). Four groups of *Plasmodium* species are shown, with subgenus designations indicated for primate parasites. The phylogeny was estimated by maximum likelihood analysis of 2.4 kb of the mitochondrial genome; the scale bar indicates 0.03 substitutions per site.

Figure 1-2A summarizes current knowledge concerning the geographic distribution and host species association of ape *Laverania* infections at over 100 field sites across sub-Saharan Africa (Kaiser et al., 2010; Liu et al., 2010a; De Nys et al., 2013; Boundenga et al., 2015; Liu et al., 2016, Liu et al., 2017). All chimpanzee subspecies, including western (*P. t. verus*), Nigeria-Cameroon (*P. t. ellioti*), central (*P. t. troglodytes*) and eastern (*P. t. schweinfurthii*) chimpanzees, as well as western lowland gorillas (*G. g. gorilla*) are endemically infected with *Laverania* parasites, with fecal detection rates ranging from 24% to 40% (Table 1-1). The true prevalence rates are likely to be considerably higher, since the amount of *Laverania* DNA that is shed into fecal samples is substantially less than that from replicating parasites in the blood (Liu et al., 2010a; Sundararaman et al., 2016). Although Cross River gorillas (*G. g. diehli*) and eastern lowland gorillas (*G. beringei graueri*) have appeared to be free of *Laverania* infections, the numbers of individuals tested from these potential hosts are still too small to draw definitive conclusions (Liu et al., 2010a). Analyses of nearly 3,500 SGA-derived mitochondrial, apicoplast and nuclear DNA sequences from ape fecal and blood samples have confirmed the existence of seven *Laverania* species (Figures 1-1 and 1-3). Of these, *P. reichenowi*, *P. gaboni*, and *P. billcollinsi* are found in wild-living chimpanzees, while *P. praefalciparum*, *P. blacklocki*, and *P. adleri* are found in western gorillas and *P. lomaminensis* is found in bonobos. All seven *Laverania* species have been classified based on numerous SGA-derived organelle and nuclear gene sequences from many different field isolates (Liu et al., 2010a; Liu et al., 2016, Liu et al., 2017). Whole genome sequencing of *P. reichenowi* and *P. gaboni* parasites confirmed that they represent distinct species, with no evidence of interspecific hybridization (Otto et al., 2014; Sundararaman et al., 2016). While it has been argued that detection of parasite DNA in either feces or blood, in itself, is not proof of productive *Plasmodium* infection (Valkiunas et al., 2011), the high prevalence rates of *Laverania*

infections (Table 1-1) and their widespread distribution (Figure 1-2A) provide compelling evidence for significant ongoing transmission.

Of note, five of the six *Laverania* parasites that infect chimpanzees and gorillas exhibit strict host specificity when infecting wild-living apes, including at field sites where *Laverania* species are co-circulating in sympatric chimpanzee and gorilla populations (octagons in Figure 1-2A). Single genome amplification, which yields a proportional representation of all parasites present in a sample, failed to detect even minor fractions of *P. reichenowi*, *P. billcollinsi*, *P. adleri*, *P. praefalciparum*, or *P. blacklocki* from the “wrong” host species in over a hundred *Laverania* infected chimpanzee and gorillas (Liu et al., 2016). In contrast, *P. gaboni* is able to infect multiple host species, as it has been amplified from numerous chimpanzee and bonobo fecal samples (Liu et al., 2017, see Figure 1-3A) though it has not been detected in gorilla fecal samples. In addition, the bonobo parasite *P. lomaminesis* has been amplified from a single chimpanzee sample, suggesting that its restriction to bonobos is not absolute (Liu et al., 2017, see Figure 1-3A). However, our failure to amplify *P. lomaminesis* from additional chimpanzee samples despite extensive screening of chimpanzees living near bonobos suggests that *P. lomaminesis* is not endemic in chimpanzees. Interestingly, this host species specificity can be broken when chimpanzees and gorillas are kept together in captivity (Duval et al., 2010; Pacheco et al., 2013), and so it will be of great interest to decipher the aspects of host and/or vector biology that contribute to host species restriction in the wild.

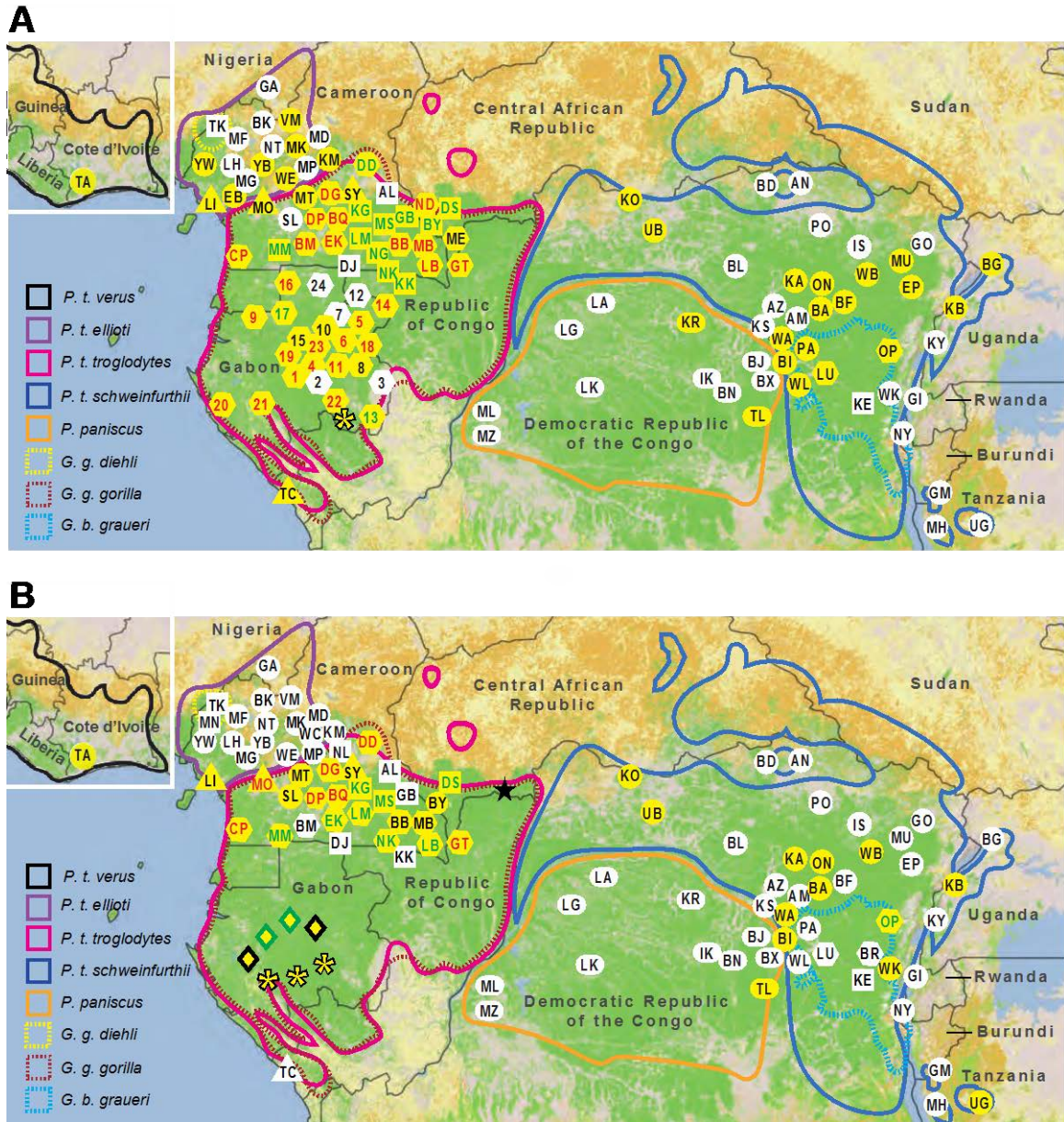


Figure 1-2. Geographic distribution of (A) *Laverania* and (B) *P. vivax* infections in wild-living apes. Field sites are shown in relation to the ranges of the four subspecies of the common chimpanzee (*P. t. verus*, black, upper left inset, *P. t. ellioti*, purple; *P. t. troglodytes*, magenta; *P. t. schweinfurthii*, blue), the Cross River (*G. g. diehli*, yellow stripe), western lowland (*G. g. gorilla*, red stripe), and eastern lowland (*G. b. graueri*, cyan stripe) gorilla, as well as the bonobo (*P. paniscus*, orange) in sub-Saharan Africa (Caldecott and Miles, 2005). Field sites are labeled by a two-letter code (Liu et al., 2010a; Liu et al., 2014, Liu et al., 2017) or numbers (Boundenga et al., 2015), and those where ape malaria was detected are highlighted in yellow. Because chimpanzee and gorilla ranges overlap, we have colored the two letter code to indicate whether chimpanzees (black), gorillas (green), or both (red) were infected at a given field site. Triangles denote ape rescue centers and asterisks mosquito collection sites. Circles, squares, and

hexagons identify locations where fecal samples were collected from chimpanzees, gorillas, or both species, respectively. Ovals indicate bonobo sites. At the TA and KB sites, blood and tissue samples were obtained from injured or deceased habituated chimpanzees (Kaiser et al., 2010; Krief et al., 2010; De Nys et al., 2013). Diamonds in panel B indicate the capture sites of ape *P. vivax* infected sanctuary chimpanzees (black border) and gorillas (green border), respectively, and a star denotes the location where a European forester became infected with ape *P. vivax* (Prugnolle et al., 2013). Data were compiled from published (Kaiser et al., 2010; Liu et al., 2010a; De Nys et al., 2013; Paupy et al., 2013; Prugnolle et al., 2013; Liu et al., 2014; Boundenga et al., 2015; Liu et al., 2017) and unpublished studies (Table 1-1).

Table 1-1. Fecal-based prevalence estimates of *Laverania* and *P. vivax* infections in wild-living African apes

Species/subspecies	<i>Laverania</i> ^a					<i>P. vivax</i> ^b				
	Field sites tested ^c	Field sites positive	Fecal samples tested	Fecal samples positive	% Infection rate (CI) ^d	Field sites tested ^e	Field sites positive	Fecal samples tested	Fecal samples positive	% Infection rate (CI) ^d
Western chimpanzee (<i>P. t. verus</i>) ^g	1	1	171	34	40 (31-50)	1	1	171	2	4 (2-11)
Nigeria-Cameroon chimpanzee (<i>P. t. ellioti</i>)	15	7	148	21	29 (20-39)	15	0	149	0	0 (0-4)
Central chimpanzee (<i>P. t. troglodytes</i>)	47	31	1412	271	39 (36-42)	25	11	1130	25	8 (6-10)
Eastern chimpanzee (<i>P. t. schweinfurthii</i>)	33	17	1676	199	24 (20-28)	34	10	1784	20	4 (3-6)
Cross River gorilla (<i>G. g. diehli</i>)	1	0	9	0	0 (0-53)	2	0	80	0	0 (0-8)
Western lowland gorilla (<i>G. g. gorilla</i>)	49	38	1584	256	33 (30-36)	22	13	1575	30	7 (5-9)
Eastern lowland gorilla (<i>G. b. graueri</i>)	3	0	148	0	0 (0-4)	4	1	189	2	4 (1-9)
Bonobo (<i>P. paniscus</i>) ^f	11	1	1418	16	2 (1-3)	4	1	407	2	2 (0-4)

^a*Laverania* infection results were compiled from six studies (Kaiser et al., 2010; Liu et al., 2010; De Nys et al., 2013; Boundenga et al., 2015; Liu et al., 2016; Liu et al., 2017) as well as recently obtained unpublished data from additional field sites (BJ, BK, DJ, GA, GI, GM, GO, IK, KB, KY, MD, MG, MH, MK, MP, NY, SL, TK, UG).

^bApe *P. vivax* infection results were compiled from five studies (Kaiser et al., 2010; Liu et al., 2010; De Nys et al., 2013; Liu et al., 2014; Liu et al., 2017) as well as recently obtained unpublished data from additional field sites (BG, GA, GI, KB, KY, MH, NY).

^cThe location of field sites is shown in Fig. 2.

^dInfection rates were estimated for each ape species or subspecies based on the combined numbers of PCR-positive fecal samples per total number of fecal samples screened, but assuming similar levels of specimen degradation, redundant sampling and diagnostic test sensitivities across all studies (Liu et al., 2010; Liu et al., 2014). Since there is less *Plasmodium* DNA shed into fecal samples than can be detected in the blood, the values represent minimum estimates. Brackets indicate 95% confidence intervals (CI). Results from chimpanzee blood samples at the TA and KB sites are not included (Kaiser et al., 2010; Krief et al., 2010).

^eFecal samples from *P. t. verus* were screened using pan-*Plasmodium* *cylB* primers, not *Laverania* or *P. vivax* specific PCR primers (Kaiser et al., 2010; De Nys et al., 2013).

^fScreened using Intensive PCR, multiple reactions per sample tested.

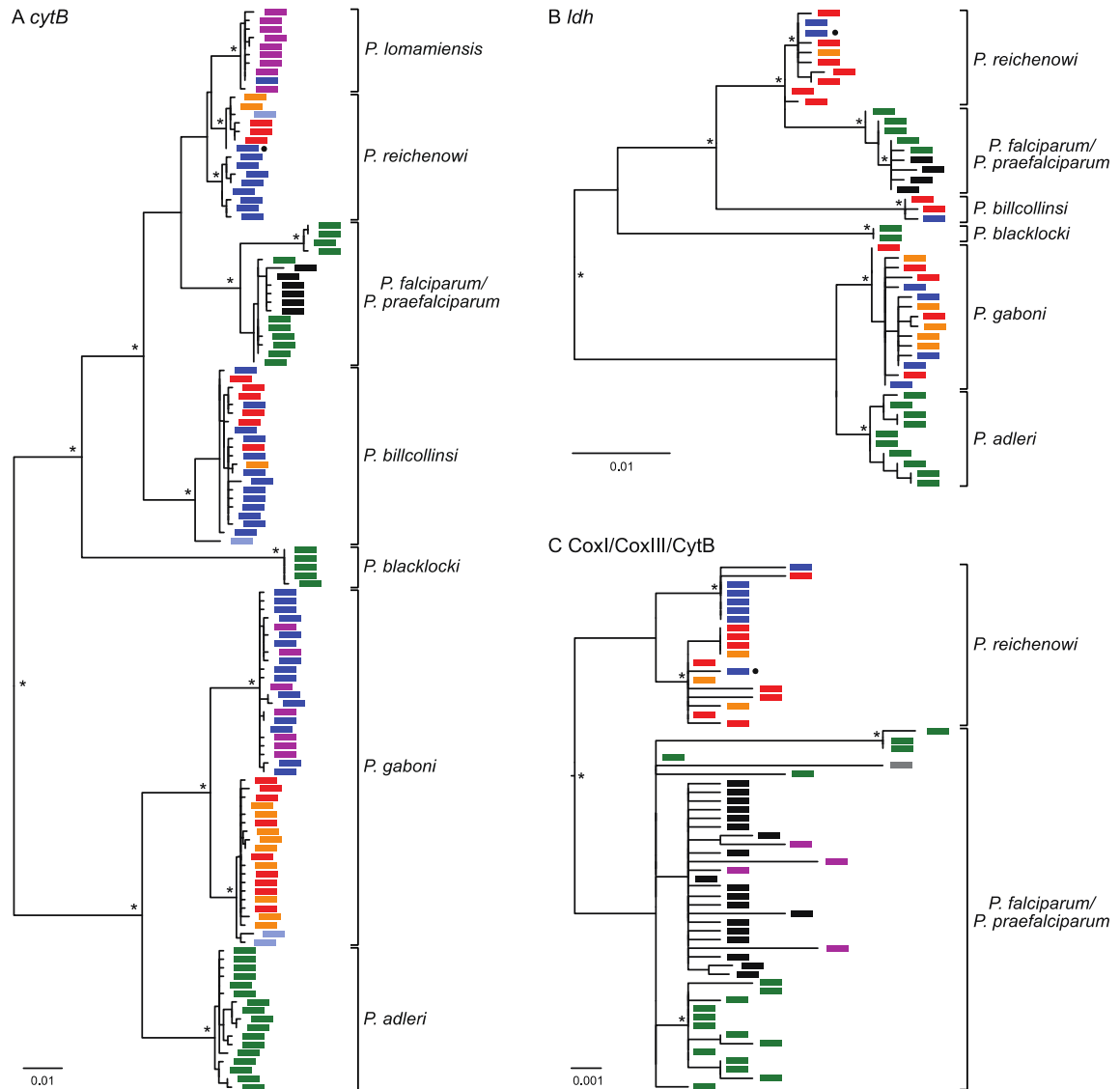


Figure 1-3. Evolutionary relationships of ape and human *Laverania* parasites. The phylogenetic relationships of (A) mitochondrial cytochrome B (*cytB*; 956 bp) and (B) nuclear lactate dehydrogenase (*ldh*; 772 bp) gene sequences, as well as (C) concatenated mitochondrial protein (CoxI/CoxIII/CytB; 981 amino acids) sequences are shown. Ape parasite sequences are colored according to their host species (*Pan troglodytes verus*, light blue; *Pan troglodytes troglodytes*, red; *Pan troglodytes schweinfurthii*, dark blue; *Pan troglodytes ellioti*, orange; *Gorilla gorilla gorilla*, green; *Pan paniscus*, purple), and human parasite reference sequences are shown in black. A black circle denotes the *Plasmodium reichenowi* PrCDC reference sequence (Otto et al., 2014) derived from a chimpanzee captured in the Belgian Congo (now the Democratic Republic of the Congo) (*Pan troglodytes schweinfurthii*) (Coatney et al., 1971). (C) Four *Plasmodium falciparum* sequences from captive bonobos (Krief et al., 2010) and one *Plasmodium praefalciparum* sequence from a captive greater spot-nosed monkey (Prugnolle et al., 2011) are shown in purple and grey, respectively. Parenteses indicate *Laverania* spp. Phylogenies were generated using maximum likelihood methods. Asterisks at major nodes indicate

bootstrap values $\geq 65\%$, and the scale bars represent (A, B) 0.01 nucleotide substitutions per site, or (C) 0.001 amino acid replacements per site, respectively. Sequences were combined from multiple studies (Kaiser et al., 2010; Krief et al., 2010; Liu et al., 2010a, 2016, 2017; Prugnolle et al., 2011).

1.4 Origin of *P. falciparum* in western gorillas and emergence in humans

Characterization of the various ape *Laverania* species identified one lineage in western gorillas that was comprised of parasites that were nearly identical to *P. falciparum* (Liu et al., 2010a; Prugnolle et al., 2010). This was initially interpreted as indicating that human parasites can infect gorillas (Prugnolle et al., 2010). However, with the characterization of mtDNA sequences from large numbers of additional wild-living gorillas, it became apparent that all extant *P. falciparum* strains from humans fall within the radiation of these gorilla parasites (Liu et al., 2010a). Analyses of both mitochondrial (Figure 1-3A) and nuclear (Figure 1-3B) sequences confirmed these relationships, indicating that human *P. falciparum* resulted from the cross-species transmission of a parasite that had previously diversified in gorillas. This gorilla parasite lineage has been named *P. praefalciparum* to indicate its role in the origin of *P. falciparum*. To investigate how often *P. praefalciparum* crossed the species barrier to humans, we constructed a phylogenetic tree from concatenated mitochondrial protein sequences of these and closely related *P. reichenowi* parasites, which yielded evidence for only a single transmission event (Figure 1-3C). These findings are consistent with results from epidemiological surveys in Cameroon and Gabon, which demonstrated that humans living in the immediate vicinity of wild-living chimpanzees and gorillas do not harbor ape *Laverania* parasites (Sundararaman et al., 2013; Delicat-Loembet et al., 2015). Thus, *P. praefalciparum* parasites appear incapable of infecting humans, suggesting that the particular gorilla parasite strain that was able to cross the host species barrier must have carried one or more highly unusual mutations that conferred an ability to colonize humans.

Although alternative hypotheses concerning the origin of *P. falciparum* have been proposed, none has stood the test of time. For example, the finding of a *P. praefalciparum* infection in a greater spot-nosed monkey (*Cercopithecus nictitans*) (Figure 1-3C) was taken to indicate that *P. falciparum* could have originated in monkeys (Prugnolle et al., 2011). However, this theory ignored the fact that *P. praefalciparum* sequences had been amplified from numerous wild-living gorillas at 11 different field sites up to 750 km apart, whereas only a single captive infected monkey was reported (Sharp et al., 2011). Indeed, subsequent testing of nearly 300 wild-caught greater spot-nosed monkeys failed to identify a single *P. praefalciparum* infection, indicating that this monkey species is not a natural reservoir for this parasite (Ayouba et al., 2012). Similarly, amplification of *P. falciparum* sequences from captive bonobos was taken to indicate that the human malaria parasite originated in this ape species (Krief et al., 2010). However, phylogenetic analysis of these sequences revealed that they were completely interspersed with human *P. falciparum* (Figure 1-3C), which together with the finding of drug resistance mutations in the bonobo parasites (Krief et al., 2010), indicated that these apes had acquired parasites from the local human population. This is not without precedent, since human *P. falciparum* has on occasion been found to infect chimpanzees in captivity (Duval et al., 2010; Pacheco et al., 2013).

P. falciparum has long been suspected to exhibit unusually low levels of genetic diversity (Rich et al., 1998), but the underlying causes have remained unclear. Recent genome-wide comparisons of the chimpanzee parasites *P. gaboni* and *P. reichenowi* have shown that their within-species genetic diversity is about 10-fold higher than that seen in *P. falciparum* (Sundararaman et al., 2016). Thus, the extremely low diversity among extant *P. falciparum* strains is not a general characteristic of *Laverania* parasites. Recent selective sweeps of drug resistance mutations have reduced levels of polymorphism in

P. falciparum, but because resistant and sensitive strains continue to recombine in mosquitoes, diversity has only been reduced in the immediate vicinity of the selected loci (Nair et al., 2003; Volkman et al., 2007). Instead, the greatly reduced level of diversity across the entire *P. falciparum* genome most likely resulted from a recent severe population bottleneck, which is most plausibly explained by the gorilla-to-human cross-species transmission event though subsequent bottlenecks secondary to human and mosquito adaptation may also have contributed to the observed genetic diversity.

1.5 A sylvatic reservoir of *P. vivax* and African origin of *P. vivax*

Although early studies indicated that chimpanzees and gorillas harbor *P. vivax*-like parasites, the number of sequences recovered was too limited to draw definitive conclusions (Kaiser et al., 2010; Krief et al., 2010; Liu et al., 2010a). As for the *Laverania* species, elucidation of the molecular epidemiology of *P. vivax* in apes required a comprehensive analysis of wild-living populations across central Africa (Liu et al., 2014). Table 1-1 and Figure 1-2B summarize available data from published studies, showing that *P. vivax* is relatively common among central and eastern chimpanzees as well as western lowland gorillas, which together represent a considerable sylvatic *P. vivax* reservoir (Kaiser et al., 2010; Liu et al., 2010a; De Nys et al., 2013; Liu et al., 2014). In addition, *P. vivax* sequences have also been amplified from wild-living bonobos (Liu et al., 2017). However, amplification of *P. vivax* DNA sequences from fecal samples was considerably less efficient than from blood samples, most likely reflecting much lower parasite loads in fecal samples compared to blood (Liu et al., 2014). Thus, the observed fecal-based infection rates, which ranged from 2% to 8% for the various ape species and subspecies (Table 1-1), are expected to greatly underestimate the actual prevalence rates, perhaps by as much as an order of magnitude. The low sensitivity of fecal parasite detection may

also explain why *P. vivax* has not yet been detected in wild-living Nigeria-Cameroon chimpanzees or in Cross River gorillas. Indeed, *P. vivax*-like sequences were readily amplified from the blood of captive Nigeria-Cameroon chimpanzees, indicating that this subspecies is susceptible to *P. vivax* infection (Figure 1-4).

Phylogenetic analyses of SGA-derived sequences showed that ape and human *P. vivax* were very closely related. In phylogenetic trees of mitochondrial (Figure 1-4A), nuclear (Figure 1-4B), and apicoplast (Figure 1-4C) sequences, human *P. vivax* sequences formed a single lineage within the radiation of the ape parasites. In contrast, parasite sequences derived from ape samples were interspersed, suggesting that *P. vivax* strains circulate freely between ape species. Of note, analysis of nearly 1,000 bushmeat samples failed to identify related sequences in samples from any of 16 different monkey species, strongly suggesting that *P. vivax* is restricted to apes (Liu et al., 2014).

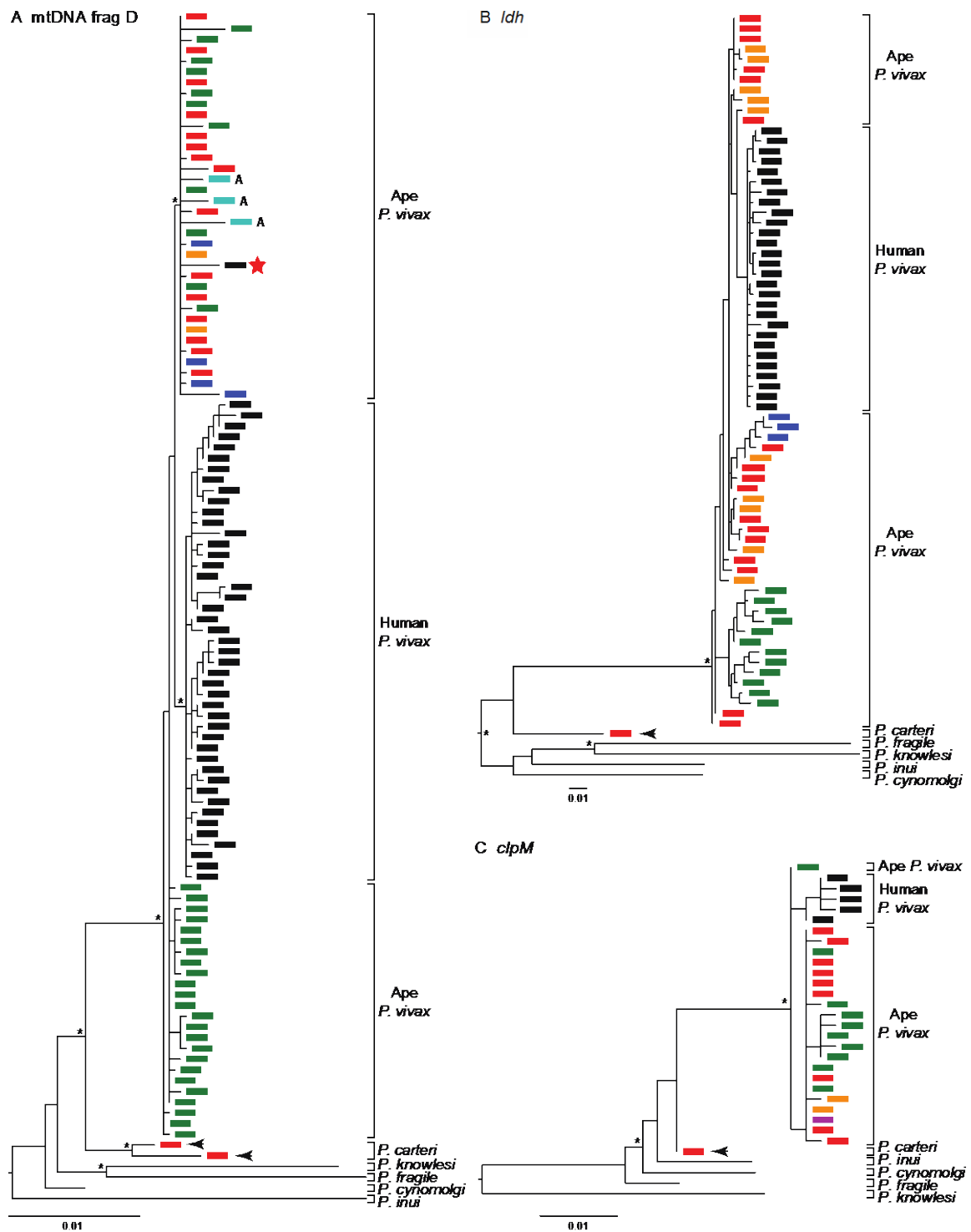


Figure 1-4: Evolutionary relationships of ape and human *P. vivax* parasites. Phylogenies were derived from (A) mtDNA fragment D (2,539 bp), (B) nuclear DNA (*ldh* gene; 711 bp), and (C) apicoplast DNA (*clpM* gene; 574 bp). Parasite sequences are colored according to their host species (*P. t. troglodytes*, red; *P. t. schweinfurthii*, dark

blue; *P. t. ellioti*, orange; *G. g. gorilla*, green; *Pan paniscus*, purple; human, black); the red star denotes a parasite from a European who worked in an African forest. Mosquito (*A. moucheti*) derived sequences are shown in cyan. Reference sequences for *P. cynomolgi*, *P. inui*, *P. fragile*, and *P. knowlesi* are indicated. A lineage of parasite sequences from wild chimpanzees, which is related to ape and human *P. vivax*, likely represents a new *Plasmodium* species, which has been designated *P. carteri* (black arrows). Phylogenies were generated using maximum likelihood methods. Asterisks at major nodes indicate bootstrap values $\geq 65\%$, and the scale bars represent 0.01 nucleotide substitutions per site. Sequences were combined from multiple studies (Krief et al., 2010; Paupy et al., 2013; Prugnolle et al., 2013; Liu et al., 2014, 2017).

Until recently, the closest known relative of *P. vivax* was a parasite, *P. cynomolgi*, which infects macaques in Asia (Tachibana et al., 2012). In phylogenetic trees, *P. vivax* and *P. cynomolgi* fall within a clade of parasites that includes at least eight other *Plasmodium* species infecting Asian macaques (Figure 1-1). The consensus view has thus been that *P. vivax* emerged in Southeast Asia following the cross-species transmission of a macaque parasite (Escalante et al., 2005; Jongwutiwes et al., 2005; Mu et al., 2005; Neafsey et al., 2012). However, this hypothesis has always been at odds with two other observations. First, the high prevalence of the Duffy negative phenotype in sub-Saharan Africans, which suggested that this mutation arose in response to prolonged selection pressure from *P. vivax* (Carter, 2003) rather than another unidentified pathogen (Livingstone, 1984). Second, modern humans did not arrive in Asia until about 60,000 years ago (Mellars, 2006); yet, *P. vivax* has likely diverged from macaque parasites much earlier than this (Escalante et al., 2005; Jongwutiwes et al., 2005; Mu et al., 2005; Neafsey et al., 2012). Thus, *P. vivax* would have had a rather convoluted history, requiring transmission from macaques to an early hominin, such as *Homo erectus*, followed by its diversification in that host before numerous lineages were transmitted to modern humans after they emerged from Africa. The discovery of *P. vivax* in large numbers of African great apes now resolves these inconsistencies, providing compelling evidence for an African, rather than an Asian, origin of human *P. vivax*.

The phylogenetic relationships of *P. vivax* strains suggest that all extant human *P. vivax* strains form a monophyletic clade within the radiation of ape parasites (Figure 1-4). This could be interpreted to mean that *P. vivax* originated in humans following a single transmission event. However, the lack of host specificity of ape *P. vivax* in natural settings (Liu et al., 2014, Liu et al., 2017), along with the finding of a zoonotic ape *P. vivax* infection (Prugnolle et al., 2013), argues against this theory. Instead, it seems more likely, based on these short sequences, that extant human *P. vivax* represents a lineage that survived after spreading out of Africa. Human *P. vivax* strains that are currently found in Madagascar and parts of Africa are likely the result of a reintroduction of this parasite from Asia (Culleton and Carter, 2012).

While it could be argued that the ape *P. vivax* was brought to Africa by humans who migrated from Asia (Prugnolle et al., 2013), this hypothesis has been refuted by sequences indicating the existence of a related, but distinct, *Plasmodium* species that also infects African apes. This *Plasmodium* species, which is apparent in trees of mitochondrial, nuclear and apicoplast sequences (Figure 1-4), has been found in chimpanzees from two different locations in Cameroon (the BQ and DG field sites in Figure 1-2) and represents the closest known relative of *P. vivax*. The most parsimonious interpretation of this finding is that the common ancestor of these two species was in Africa, indicating that the lineage existed there for a long time before *P. vivax* arose as a distinct species (Figure 1-4). We propose to designate this newly described species *Plasmodium carteri*, in honor of Richard Carter, who has long championed the hypothesis that *P. vivax* originated in Africa (Carter, 2003; Culleton and Carter, 2012).

1.6 Zoonotic potential of ape parasites

Despite the identification of suitable bridge vectors (Paupy et al., 2013; Makanga et al., 2016), both experimental transmission and molecular epidemiological studies indicate that

ape *Laverania* parasites do not normally cause blood stage infections in humans. Attempts to inoculate humans with a parasite identified as "*P. reichenowi*" over 100 years ago did not result in parasitemia (Blacklock and Adler, 1922). More importantly, two recent field studies conducted in rural Cameroon and Gabon failed to identify ape *Laverania* infections in humans living in close proximity to infected chimpanzees and gorillas (Sundararaman et al., 2013; Delicat-Loembet et al., 2015). In contrast, experimental inoculation with *P. schwetzi* did result in patent parasitemia and clinical malaria in Duffy positive, but not Duffy negative, humans suggesting that the inoculum was likely ape *P. vivax* (Contacos 1970). Furthermore, natural infections of Duffy positive individuals can occur, as exemplified by the case of a Caucasian male who acquired this infection after working for 18 days in a forest in the Central African Republic (Figure 1-2B). Parasite sequences amplified from this individual's blood did not fall within the human *P. vivax* lineage, but instead clustered with parasites obtained from wild-living apes (Figure 1-4A), confirming acquisition by cross-species transmission from an ape (Prugnolle et al., 2013). These data indicate that ape *Laverania* parasites do not switch between host species, except under highly unusual circumstances, while ape *P. vivax* is much less host-specific and has the potential to infect Duffy positive humans, suggesting that human and ape *P. vivax* parasites represent a single species.

Importantly, however, none of these studies address the possibility that humans are exposed to ape parasites, which may cause a liver stage infection in the absence of a blood stage infection. Thus, it is currently unclear whether absence of blood stage infections in humans is a result of lack of exposure to mosquitoes carrying ape *Plasmodium* sporozoites, parasite-host incompatibility at the liver stage of infection, and/or parasite-host incompatibility at the blood stage of infection. In Chapter 2 of this dissertation, I use a fecal screen to search for ape *Plasmodium* infections in humans living

near infected apes, including hunter-gatherers who spend extensive time in or near ape habitat. Although studies in mice show that pre-erythrocytic *Plasmodium* DNA can be detected in stool, I did not detect evidence of abortive ape *Plasmodium* liver infections in the 504 Cameroonian humans screened. Instead, I identified a link between parasitemia and fecal *Plasmodium* DNA detection, suggesting that parasite DNA in fecal samples is derived from blood stage infections.

1.7 Emergence of *P. vivax* in human populations

While available organellar and nuclear sequences suggest an African origin of human *P. vivax* (Liu et al., 2014), nothing is known about the circumstances and mechanistic processes that led to the parasite's emergence in humans. Yet, such information is critical to understand how ape parasites crossed the species barrier and whether such events are likely to occur again. The lack of *in vitro* culture systems poses a significant challenge to the functional analysis of ape *Plasmodium* parasites, but whole genome sequencing represents a critical first step towards understanding their biology (Otto et al., 2014; Sundararaman et al., 2016). In this dissertation, I present the results of comparative genomic analysis of ape and human *P. vivax*. In Chapter 3 I describe the adaptation of selective whole genome amplification (SWGA) to *P. vivax*-infected samples, and subsequently use this method in Chapter 4 to generate *P. vivax* genome sequences from six chimpanzees and one gorilla. Differences in the magnitude and pattern of polymorphism observed in ape and human *P. vivax* indicate that the human *P. vivax* lineage has undergone a recent bottleneck followed by rapid population expansion. Furthermore, both genome-wide analysis and functional protein studies suggest that the human *P. vivax* lineage has acquired remarkably few human-specific adaptations. Thus,

these studies provide insight into the zoonotic ape reservoir of malaria parasites and reveal the evolutionary history of *P. vivax* in humans.

1.8 Chapter references

- Ayoub, A., Mouacha, F., Learn, G.H., Mpoudi-Ngole, E., Rayner, J.C., Sharp, P.M., Hahn, B.H., Delaporte, E., and Peeters, M. 2012. Ubiquitous *Hepatocystis* infections, but no evidence of *Plasmodium falciparum*-like malaria parasites in wild greater spot-nosed monkeys (*Cercopithecus nictitans*). *Int. J. Parasitol.* 42, 709-713.
- Bhatt, S., Weiss, D.J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., Battle, K.E., Moyes, C.L., Henry, A., Eckhoff, P.A., Wenger, E.A., Briet, O., Penny, M.A., Smith, T.A., Bennett, A., Yukich, J., Eisele, T.P., Griffin, J.T., Fergus, C.A., Lynch, M., Lindgren, F., Cohen, J.M., Murray, C.L., Smith, D.L., Hay, S.I., Cibulskis, R.E., and Gething, P.W. 2015. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature* 526, 207-211.
- Blacklock, B., and Adler, S. 1922. A parasite resembling *Plasmodium falciparum* in a chimpanzee. *Ann. Trop. Med. Parasitol.* XVI, 99-106.
- Boundenga, L., Ollomo, B., Rougeron, V., Mouele, L.Y., Mve-Ondo, B., Delicat-Loembet, L.M., Moukodoum, N.D., Okouga, A.P., Arnathau, C., Elguero, E., Durand, P., Liegeois, F., Boue, V., Motsch, P., Le Flohic, G., Ndoungouet, A., Paupy, C., Ba, C.T., Renaud, F., and Prugnolle, F. 2015. Diversity of malaria parasites in great apes in Gabon. *Malar. J.* 14, 111.
- Bray, R.S. 1958. Studies on malaria in chimpanzees VI. *Laverania falciparum*. *Am. J. Trop. Med. Hyg.* 7, 20-24.
- Brumpt, E. 1939. Les parasites du paludisme des chimpanzes. *C. R. Soc. Bio.* 130, 837-840.
- Caldecott, J.O., and Miles, L. 2005. *World Atlas of Great Apes and their Conservation*. University of California Press, Berkley, California, USA.
- Carter, R. 2003. Speculations on the origins of *Plasmodium vivax* malaria. *Trends Parasitol.* 19, 214-219.
- Carter, R., and Mendis, K.N. 2002. Evolutionary and historical aspects of the burden of malaria. *Clin. Microbiol. Rev.* 15, 564-594.
- Coatney G.R., Collins W.E., Warren, M., Contacos P.G. 1971. *The Primate Malaras*. Washington, DC: US Government Printing Office.
http://phsource.us/PH/PARA/Book/primate_24.pdf
- Conway, D.J., Fanello, C., Lloyd, J.M., Al-Joubori, B.M., Baloch, A.H., Somanath, S.D., Roper, C., Oduola, A.M., Mulder, B., Pova, M.M., Singh, B., and Thomas, A.W.

2000. Origin of *Plasmodium falciparum* malaria is traced by mitochondrial DNA. *Mol. Biochem. Parasitol.* 111, 163-171.
- Culleton, R., and Carter, R. 2012. African *Plasmodium vivax*: distribution and origins. *Int. J. Parasitol.* 42, 1091-1097.
- De Nys, H.M., Calvignac-Spencer, S., Thiesen, U., Boesch, C., Wittig, R.M., Mundry, R., and Leendertz, F.H. 2013. Age-related effects on malaria parasite infection in wild chimpanzees. *Biol. Lett.* 9, 20121160.
- Delicat-Loembet, L., Rougeron, V., Ollomo, B., Arnathau, C., Roche, B., Elguero, E., Moukodoum, N.D., Okougha, A.P., Mve Ondo, B., Boundenga, L., Houze, S., Galan, M., Nkoghe, D., Leroy, E.M., Durand, P., Paupy, C., Renaud, F., and Prugnolle, F. 2015. No evidence for ape *Plasmodium* infections in humans in Gabon. *PLoS One* 10, e0126933.
- Duval, L., Fourment, M., Nerrienet, E., Rousset, D., Sadeuh, S.A., Goodman, S.M., Andriaholinirina, N.V., Randrianarivelosia, M., Paul, R.E., Robert, V., Ayala, F.J., and Ariey, F. 2010. African apes as reservoirs of *Plasmodium falciparum* and the origin and diversification of the *Laverania* subgenus. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10561-10566.
- Escalante, A.A., and Ayala, F.J. 1994. Phylogeny of the malarial genus *Plasmodium*, derived from rRNA gene sequences. *Proc. Natl. Acad. Sci. U.S.A.* 91, 11373-11377.
- Escalante, A.A., Cornejo, O.E., Freeland, D.E., Poe, A.C., Durrego, E., Collins, W.E., and Lal, A.A. 2005. A monkey's tale: The origin of *Plasmodium vivax* as a human malaria parasite. *Proc. Natl. Acad. Sci. U.S.A.* 102, 1980-1985.
- Hedrick, P.W. 2011. Population genetics of malaria resistance in humans. *Heredity (Edinb)* 107, 283-304.
- Howes, R.E., Patil, A.P., Piel, F.B., Nyangiri, O.A., Kabaria, C.W., Gething, P.W., Zimmerman, P.A., Barnadas, C., Beall, C.M., Gebremedhin, A., Menard, D., Williams, T.N., Weatherall, D.J., and Hay, S.I. 2011. The global distribution of the Duffy blood group. *Nat. Commun.* 2, 266.
- Jongwutiwes, S., Putaporntip, C., Iwasaki, T., Ferreira, M.U., Kanbara, H., and Hughes, A.L. 2005. Mitochondrial genome sequences support ancient population expansion in *Plasmodium vivax*. *Mol. Biol. Evol.* 22, 1733-1739.
- Kaiser, M., Lowa, A., Ulrich, M., Ellerbrok, H., Goffe, A.S., Blasse, A., Zommers, Z., Couacy-Hymann, E., Babweteera, F., Zuberbuhler, K., Metzger, S., Geidel, S., Boesch, C., Gillespie, T.R., and Leendertz, F.H. 2010. Wild chimpanzees infected with 5 *Plasmodium* species. *Emerg. Infect. Dis.* 16, 1956-1959.
- Krief, S., Escalante, A.A., Pacheco, M.A., Mugisha, L., Andre, C., Halbwax, M., Fischer, A., Krief, J.M., Kasenene, J.M., Crandfield, M., Cornejo, O.E., Chavatte, J.M., Lin, C., Letourneur, F., Gruner, A.C., McCutchan, T.F., Renia, L., and Snounou,

- G. 2010. On the diversity of malaria parasites in African apes and the origin of *Plasmodium falciparum* from bonobos. *PLoS Pathog.* 6, e1000765.
- Liu, W., Li, Y., Learn, G.H., Rudicell, R.S., Robertson, J.D., Keele, B.F., Ndjango, J.B., Sanz, C.M., Morgan, D.B., Locatelli, S., Gonder, M.K., Kranzusch, P.J., Walsh, P.D., Delaporte, E., Mpoudi-Ngole, E., Georgiev, A.V., Muller, M.N., Shaw, G.M., Peeters, M., Sharp, P.M., Rayner, J.C., and Hahn, B.H. 2010a. Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. *Nature* 467, 420-425.
- Liu, W., Li, Y., Peeters, M., Rayner, J.C., Sharp, P.M., Shaw, G.M., and Hahn, B.H. 2010b. Single genome amplification and direct amplicon sequencing of *Plasmodium* spp. DNA from ape fecal specimens. *Protocol Exchange*. doi:10.1038/nprot.2010.156.
- Liu, W., Li, Y., Shaw, K.S., Learn, G.H., Plenderleith, L.J., Malenke, J.A., Sundararaman, S.A., Ramirez, M.A., Crystal, P.A., Smith, A.G., Bibollet-Ruche, F., Ayoub, A., Locatelli, S., Esteban, A., Mouacha, F., Guichet, E., Butel, C., Ahuka-Mundeki, S., Inogwabini, B.I., Ndjango, J.B., Speede, S., Sanz, C.M., Morgan, D.B., Gonder, M.K., Kranzusch, P.J., Walsh, P.D., Georgiev, A.V., Muller, M.N., Piel, A.K., Stewart, F.A., Wilson, M.L., Pusey, A.E., Cui, L., Wang, Z., Farnert, A., Sutherland, C.J., Nolder, D., Hart, J.A., Hart, T.B., Bertolani, P., Gillis, A., LeBreton, M., Tafon, B., Kiyang, J., Djoko, C.F., Schneider, B.S., Wolfe, N.D., Mpoudi-Ngole, E., Delaporte, E., Carter, R., Culleton, R.L., Shaw, G.M., Rayner, J.C., Peeters, M., Hahn, B.H., and Sharp, P.M. 2014. African origin of the malaria parasite *Plasmodium vivax*. *Nat. Commun.* 5, 3346.
- Liu, W., Sundararaman, S.A., Loy, D.E., Learn, G.H., Li, Y., Plenderleith, L.J., Ndjango, J.-B.N., Speede, S., Atencia, R., Cox, D., Shaw, G.M., Ayoub, A., Peeters, M., Rayner, J.C., Hahn, B.H., and Sharp, P.M. 2016. Multigenomic delineation of *Plasmodium* species of *Laverania* subgenus infecting wild-living chimpanzees and gorillas. *Genome Biol. Evol.* 8, 1929-1939.
- Liu, W., Sherrill-Mix S., Learn G.H., Scully E.J., Li Y., Avitto A.N., Loy D.E., Lauder A.P., Sundararaman S.A., Plenderleith L.J., Ndjango J.N., Georgiev A.V., Ahuka-Mundeki S., Peeters M., Bertolani P., Dupain J., Garai C., Hart J.A., Hart T.B., Shaw G.M., Sharp P.M., Hahn B.H. 2017. Wild bonobos host geographically restricted malaria parasites including a putative new *Laverania* species. *Nat. Commun.* 8, 1635.
- Livingstone, F.B. 1984. The Duffy blood groups, vivax malaria, and malaria selection in human populations: a review. *Hum. Biol.* 56, 413-425.
- Makanga, B., Yangari, P., Rahola, N., Rougeron, V., Elguero, E., Boundenga, L., Moukoudou, N.D., Okouga, A.P., Arnathau, C., Durand, P., Willaume, E., Ayala, D., Fontenille, D., Ayala, F.J., Renaud, F., Ollomo, B., Prugnolle, F., and Paupy, C. 2016. Ape malaria transmission and potential for ape-to-human transfers in Africa. *Proc. Natl. Acad. Sci. U.S.A.*, 201603008.
- Mellars, P. 2006. Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. *Science* 313, 796-800.

- Miller, L.H., Mason, S.J., Clyde, D.F., and McGinniss, M.H. 1976. The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, *FyFy*. N. Engl. J. Med. 295, 302-304.
- Molina-Cruz, A., and Barillas-Mury, C. 2016. Mosquito vectors of ape malarias: another piece of the puzzle. Proc. Natl. Acad. Sci. U.S.A. 113, 5153-5154.
- Mu, J., Joy, D.A., Duan, J., Huang, Y., Carlton, J., Walker, J., Barnwell, J., Beerli, P., Charleston, M.A., Pybus, O.G., and Su, X.Z. 2005. Host switch leads to emergence of *Plasmodium vivax* malaria in humans. Mol. Biol. Evol. 22, 1686-1693.
- Nair, S., Williams, J.T., Brockman, A., Paiphun, L., Mayxay, M., Newton, P.N., Guthmann, J.P., Smithuis, F.M., Hien, T.T., White, N.J., Nosten, F., and Anderson, T.J. 2003. A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. Mol. Biol. Evol. 20, 1526-1536.
- Neafsey, D.E., Galinsky, K., Jiang, R.H., Young, L., Sykes, S.M., Saif, S., Gujja, S., Goldberg, J.M., Young, S., Zeng, Q., Chapman, S.B., Dash, A.P., Anvikar, A.R., Sutton, P.L., Birren, B.W., Escalante, A.A., Barnwell, J.W., and Carlton, J.M. 2012. The malaria parasite *Plasmodium vivax* exhibits greater genetic diversity than *Plasmodium falciparum*. Nat. Genet. 44, 1046-1050.
- Ollomo, B., Durand, P., Prugnolle, F., Douzery, E., Arnathau, C., Nkoghe, D., Leroy, E., and Renaud, F. 2009. A new malaria agent in African hominids. PLoS Pathog. 5, e1000446.
- Otto, T.D., Rayner, J.C., Bohme, U., Pain, A., Spottiswoode, N., Sanders, M., Quail, M., Ollomo, B., Renaud, F., Thomas, A.W., Prugnolle, F., Conway, D.J., Newbold, C., and Berriman, M. 2014. Genome sequencing of chimpanzee malaria parasites reveals possible pathways of adaptation to human hosts. Nat. Commun. 5, 4754.
- Pacheco, M.A., Cranfield, M., Cameron, K., and Escalante, A.A. 2013. Malarial parasite diversity in chimpanzees: the value of comparative approaches to ascertain the evolution of *Plasmodium falciparum* antigens. Malar. J. 12, 238.
- Paupy, C., Makanga, B., Ollomo, B., Rahola, N., Durand, P., Magnus, J., Willaume, E., Renaud, F., Fontenille, D., and Prugnolle, F. 2013. *Anopheles moucheti* and *Anopheles vinckei* are candidate vectors of ape *Plasmodium* parasites, including *Plasmodium praefalciparum* in Gabon. PLoS One 8, e57294.
- Prugnolle, F., Durand, P., Neel, C., Ollomo, B., Ayala, F.J., Arnathau, C., Etienne, L., Mpoudi-Ngole, E., Nkoghe, D., Leroy, E., Delaporte, E., Peeters, M., and Renaud, F. 2010. African great apes are natural hosts of multiple related malaria species, including *Plasmodium falciparum*. Proc. Natl. Acad. Sci. U.S.A. 107, 1458-1463.
- Prugnolle, F., Ollomo, B., Durand, P., Yalcindag, E., Arnathau, C., Elguero, E., Berry, A., Pourrut, X., Gonzalez, J.P., Nkoghe, D., Akiana, J., Verrier, D., Leroy, E., Ayala, F.J., and Renaud, F. 2011. African monkeys are infected by *Plasmodium*

- falciparum* nonhuman primate-specific strains. Proc. Natl. Acad. Sci. U.S.A. 108, 11948-11953.
- Prugnotte, F., Rougeron, V., Becquart, P., Berry, A., Makanga, B., Rahola, N., Arnathau, C., Ngoubangoye, B., Menard, S., Willaume, E., Ayala, F.J., Fontenille, D., Ollomo, B., Durand, P., Paupy, C., and Renaud, F. 2013. Diversity, host switching and evolution of *Plasmodium vivax* infecting African great apes. Proc. Natl. Acad. Sci. U.S.A. 110, 8123-8128.
- Ramachandran, S., Deshpande, O., Roseman, C.C., Rosenberg, N.A., Feldman, M.W., and Cavalli-Sforza, L.L. 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc. Natl. Acad. Sci. U.S.A. 102, 15942-15947.
- Reichenow, E. 1920. Über das Vorkommen der Malariaparasiten des Menschen bei den Afrikanischen Menschenaffen. Centralbl. f. Bakt. I. Abt. Orig. 85, 207-216.
- Rich, S.M., Leendertz, F.H., Xu, G., LeBreton, M., Djoko, C.F., Aminake, M.N., Takang, E.E., Dikko, J.L.D., Pike, B.L., Rosenthal, B.M., Formenty, P., Boesch, C., Ayala, F.J., and Wolfe, F.D. 2009. The origin of malignant malaria. Proc. Natl. Acad. Sci. U.S.A. 106, 14902-14907.
- Rich, S.M., Licht, M.C., Hudson, R.R., and Ayala, F.J. 1998. Malaria's Eve: evidence of a recent population bottleneck throughout the world populations of *Plasmodium falciparum*. Proc. Natl. Acad. Sci. U.S.A. 95, 4425-4430.
- Sharp, P.M., Liu, W., Learn, G.H., Rayner, J.C., Peeters, M., and Hahn, B.H. 2011. Source of the human malaria parasite *Plasmodium falciparum*. Proc. Natl. Acad. Sci. U.S.A. 108, E744-745.
- Sluiter, C.P., Swellengrebel, N.H., and Ihle, J.E. 1922. De dierlijke parasieten van den mensch en van onze huisdieren. 3rd ed. Scheltema and Holkema's Boekhandel, Amsterdam.
- Sundararaman, S.A., Liu, W., Keele, B.F., Learn, G.H., Bittinger, K., Mouacha, F., Ahuka-Mundike, S., Manske, M., Sherrill-Mix, S., Li, Y., Malenke, J.A., Delaporte, E., Laurent, C., Mpoudi Ngole, E., Kwiatkowski, D.P., Shaw, G.M., Rayner, J.C., Peeters, M., Sharp, P.M., Bushman, F.D., and Hahn, B.H. 2013. *Plasmodium falciparum*-like parasites infecting wild apes in southern Cameroon do not represent a recurrent source of human malaria. Proc. Natl. Acad. Sci. U.S.A. 110, 7020-7025.
- Sundararaman, S.A., Plenderleith, L.J., Liu, W., Loy, D.E., Learn, G.H., Li, Y., Shaw, K.S., Ayoub, A., Peeters, M., Speede, S., Shaw, G.M., Bushman, F.D., Brisson, D., Rayner, J.C., Sharp, P.M., and Hahn, B.H. 2016. Genomes of cryptic chimpanzee *Plasmodium* species reveal key evolutionary events leading to human malaria. Nat. Commun. 7, 11078.
- Tachibana, S., Sullivan, S.A., Kawai, S., Nakamura, S., Kim, H.R., Goto, N., Arisue, N., Palacpac, N.M., Honma, H., Yagi, M., Tougan, T., Katakai, Y., Kaneko, O., Mita, T., Kita, K., Yasutomi, Y., Sutton, P.L., Shakhbatyan, R., Horii, T., Yasunaga, T.,

- Barnwell, J.W., Escalante, A.A., Carlton, J.M., and Tanabe, K. 2012. *Plasmodium cynomolgi* genome sequences provide insight into *Plasmodium vivax* and the monkey malaria clade. *Nat. Genet.* 44, 1051-1055.
- Tanabe, K., Mita, T., Jombart, T., Eriksson, A., Horibe, S., Palacpac, N., Ranford-Cartwright, L., Sawai, H., Sakihama, N., Ohmae, H., Nakamura, M., Ferreira, M.U., Escalante, A.A., Prugnolle, F., Bjorkman, A., Farnert, A., Kaneko, A., Horii, T., Manica, A., Kishino, H., and Balloux, F. 2010. *Plasmodium falciparum* accompanied the human expansion out of Africa. *Curr. Biol.* 20, 1283-1289.
- Taylor, S.M., Parobek, C.M., and Fairhurst, R.M. 2012. Haemoglobinopathies and the clinical epidemiology of malaria: a systematic review and meta-analysis. *Lancet Infect. Dis.* 12, 457-468.
- Valkiunas, G., Ashford, R.W., Bensch, S., Killick-Kendrick, R., and Perkins, S. 2011. A cautionary note concerning *Plasmodium* in apes. *Trends Parasitol.* 27, 231-232.
- Volkman, S.K., Sabeti, P.C., DeCaprio, D., Neafsey, D.E., Schaffner, S.F., Milner, D.A., Jr., Daily, J.P., Sarr, O., Ndiaye, D., Ndir, O., Mboup, S., Duraisingh, M.T., Lukens, A., Derr, A., Stange-Thomann, N., Waggoner, S., Onofrio, R., Ziaugra, L., Mauceli, E., Gnerre, S., Jaffe, D.B., Zainoun, J., Wiegand, R.C., Birren, B.W., Hartl, D.L., Galagan, J.E., Lander, E.S., and Wirth, D.F. 2007. A genome-wide map of diversity in *Plasmodium falciparum*. *Nat. Genet.* 39, 113-119.
- World Health Organization, 2017. World Malaria Report 2017.
<http://www.who.int/malaria/publications/world-malaria-report-2017/report/en/>

CHAPTER 2 - Investigating zoonotic infection barriers to ape *Plasmodium* parasites using fecal DNA analysis

Dorothy E. Loy^{a,b}, Meagan A. Rubel^{c,d}, Alexa N. Avitto^a, Weimin Liu^a, Yingying Li^a,
Gerald H. Learn^a, Alessia Ranciaro^d, Eric Mbunwe^d, Charles Fokunang^e, Alfred K.
Njamnshi^f, Paul M. Sharp^g, Sarah A. Tishkoff^{d,h}, Beatrice H. Hahn^{a,b,*}

^aDepartment of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA, ^bDepartment of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA, ^cDepartment of Anthropology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA, ^dDepartment of Genetics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA, ^eDepartment of Pharmacotoxicology and Pharmacokinetics, Faculty of Medicine and Biomedical Sciences, University of Yaoundé I, Yaoundé, Cameroon, ^fDepartment of Neurology, Faculty of Medicine and Biomedical Sciences, Central Hospital Yaoundé, University of Yaoundé I, Yaoundé, Cameroon, ^gInstitute of Evolutionary Biology and Centre for Immunity, Infection and Evolution, University of Edinburgh, Edinburgh EH9 3FL, United Kingdom, ^hDepartment of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

Originally published in *Int J Para* 48(7):531-542

2.1 Abstract

African apes are endemically infected with numerous *Plasmodium* spp. including close relatives of human *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale*, and *Plasmodium malariae*. Although these ape parasites are not believed to pose a zoonotic threat, their ability to colonize humans has not been fully explored. In particular, it remains unknown whether ape parasites are able to initiate exo-erythrocytic replication in human hepatocytes following the bite of an infective mosquito. Since animal studies have shown that liver stage infection can result in the excretion of parasite nucleic acids into the bile, we screened fecal samples from 504 rural Cameroonians for *Plasmodium* DNA. Using pan-*Laverania* as well as *P. malariae*- and *P. vivax*-specific primer sets, we amplified human *P. falciparum* ($n=14$), *P. malariae* ($n=1$), and *P. ovale wallikeri* ($n=1$) mitochondrial sequences from fecal DNA of 15 individuals. However, despite using an intensified PCR screening approach we failed to detect ape *Laverania*, ape *P. vivax* or ape *P. malariae* parasites in these same subjects. One fecal sample from a hunter-gatherer contained a sequence closely related to the porcupine parasite *Plasmodium atheruri*. Since this same fecal sample also contained porcupine mitochondrial DNA, but a matching blood sample was *Plasmodium*-negative, it is likely that this hunter-gatherer consumed *Plasmodium*-infected bushmeat. Fecal *Plasmodium* detection was not secondary to intestinal bleeding and/or infection with gastrointestinal parasites, but indicative of blood parasitemia. Quantitative PCR identified 26-fold more parasite DNA in the blood of fecal *Plasmodium*-positive than fecal *Plasmodium*-negative individuals ($P=0.01$). However, among blood-positive individuals only 10% - 20% had detectable *Plasmodium* sequences in their stool. Thus, fecal screening of rural Cameroonians failed to uncover abortive ape *Plasmodium* infections, but detected infection with human parasites, albeit with reduced sensitivity compared with blood analysis.

2.2 Introduction

Wild-living African apes are naturally infected with a wide variety of *Plasmodium* spp., including seven members of the *Laverania* subgenus as well as close relatives of human *Plasmodium vivax*, *Plasmodium malariae* and *Plasmodium ovale* (Ollomo et al., 2009; Rich et al., 2009; Krief et al., 2010; Liu et al., 2010a, 2014, 2016, 2017; Prugnolle et al., 2010). While the impact of these infections on wild ape populations remains to be explored, their genetic characterization has provided important new insight into the origins and evolution of human malaria. For example, *Plasmodium praefalciparum*, a parasite that naturally infects gorillas, was found to have crossed the species barrier to humans several thousand years ago, giving rise to *Plasmodium falciparum* (Liu et al., 2010a). In contrast, human *P. vivax* emerged from an ancestral stock of parasites that infected chimpanzees, gorillas and humans in Africa, until the spread of the protective Duffy-negative mutation eliminated *P. vivax* in humans there (Liu et al., 2014). African apes also harbor parasites that are closely related to human *P. malariae*, *Plasmodium ovale wallikeri* and *Plasmodium ovale curtisi* as well as *Plasmodium* spp. that do not have a known human counterpart (Duval et al., 2009, 2010; Hayakawa et al., 2009; Kaiser et al., 2010; Liu et al., 2010a, 2014). However, the precise relationships between these ape and human parasites, including their host species association and prevalence, have yet to be determined. Given the magnitude of the sylvatic *Plasmodium* reservoir and the fact that ape parasite-transmitting *Anopheles* vectors also feed on humans (Paupy et al., 2013; Makanga et al., 2016), the possibility of spillovers into the human population has to be considered.

Laverania infections are common and widespread in chimpanzees (*Pan troglodytes*) and western gorillas (*Gorilla gorilla*), with estimated prevalence rates ranging from 29% to 40% (Loy et al., 2017). Similarly, ape *P. vivax* infections are widely distributed,

although fecal detection rates are generally lower than for *Laverania* parasites, likely reflecting lower blood titers (Loy et al., 2017). Several studies have addressed whether *Plasmodium*-infected apes serve as a recurrent source of human infection. PCR screening of more than 5,000 blood samples from individuals in rural Cameroon and Gabon failed to identify human infection with ape *Laverania* parasites (Sundararaman et al., 2013; Delicat-Loembet et al., 2015; Ngoubangoye et al., 2016). However, one case of a zoonotic ape *P. vivax* infection was documented in a (Duffy-positive) European forest worker (Prugnolle et al., 2013). Consistent with experimental transmission studies conducted nearly 100 years ago (Blacklock and Adler, 1922), these results indicate substantial barriers to the cross-species transmission of ape *Laverania* spp., while non-*Laverania* parasites such as ape *P. vivax* appear to exhibit a more promiscuous host tropism (Prugnolle et al., 2010, 2013; Rayner et al., 2011; Loy et al., 2017; Liu et al., 2017).

Laverania host specificity is determined, at least in part, by interactions of parasite ligands with receptors on the surface of host red blood cells (Martin et al., 2005; Wanaguru et al., 2013). However, barriers to erythrocyte invasion do not necessarily preclude infection of host hepatocytes. Among non-*Laverania* spp., an African monkey parasite, *Plasmodium gonderi*, can invade and replicate in hepatocytes of several New World monkey species, but fails to establish a productive blood stage infection in these same hosts (Sullivan et al., 2002). Similarly, the macaque parasite *Plasmodium fieldi* and human *P. ovale* can progress through the exo-erythrocytic stage of their life cycle without resulting in subsequent parasitemia in owl monkeys (*Aotus vociferans*) and squirrel monkeys (*Saimiri boliviensis*), respectively (Millet et al., 1994; Sullivan et al., 1998). Although these examples represent experimental infections of non-natural host species, they raise the possibility that similar non-blood stage initiating (abortive) liver infections could also occur under natural conditions. For example, it has been reported that exo-erythrocytic

replication of *P. vivax* may occur in Duffy-negative humans who are protected from blood stage infection, since some individuals were found to have antibodies to the *P. vivax* MSP1 protein (Herrera et al., 2005; Culleton et al., 2009), which is expressed on liver schizonts (Szarfman et al., 1988; Suhrbier et al., 1989) and merozoites (Holder and Freeman, 1984). Thus, in addition to blood analyses, studies of zoonotic *Plasmodium* infections should include approaches that can detect exo-erythrocytic stages of parasite development.

Following the bite of an infective mosquito, sporozoites of mammalian *Plasmodium* parasites migrate to the liver where they initiate asexual replication in hepatocytes that mature into schizonts, which then release merozoites into the blood stream (Prudencio et al., 2006; Vaughan and Kappe, 2017). Since direct analysis of liver stage infection requires invasive methods, we wondered whether there was a non-invasive alternative. Previous studies of *Plasmodium yoelli*-infected mice demonstrated the presence of parasite DNA in the liver, gall bladder and stool of sporozoite-inoculated animals several days before the onset of parasitemia, indicating that pre-erythrocytic parasite forms are excreted from the infected liver into the bile (Abkallo et al., 2014). In addition, *Plasmodium knowlesi* DNA was found in the stool of a macaque, which was inoculated with infected erythrocytes, 12 days after parasite DNA was last detected in the blood, indicating protracted clearance possibly through accumulation of parasite nucleic acids in the gall bladder (Kawai et al., 2014). We thus reasoned that fecal analyses might reveal human *Plasmodium* infections even if they did not progress to a productive blood stage and/or caused only transient parasitemia.

Here, we used a sensitive PCR screening approach to test fecal samples from 504 rural Cameroonians for aye *Plasmodium* parasites, including subsistence agriculturalists and hunter-gatherers at risk of exposure to transmitting forest mosquito vectors. Targeting both *Laverania* and non-*Laverania* spp., we amplified *P. falciparum*, *P. malariae* and *P.*

ovale wallikeri sequences from a subset of individuals. However, none of the fecal samples contained ape *Plasmodium* parasites. Moreover, all were negative for *P. vivax*, which has recently been reported to infect both Duffy-positive and negative individuals in Cameroon (Fru-Cho et al., 2014; Mbenda and Das, 2014; Russo et al., 2017; Zimmerman, 2017). Thus, fecal analyses of humans living in close proximity to wild chimpanzees and gorillas failed to yield evidence of abortive ape *Plasmodium* infections.

2.3 Materials and methods

Study sites and sample collection. Study participants were recruited from nine villages in the Northwest (Ntambang, NT; Sabga, SA), South (Bidou I, BI; Ndtoua, ND) and East (Nkolbikong, NK; Missoume, MI; Njibot, NJ; Aviation, AV; Bosquet, BO) Administrative Regions of Cameroon (Figure 2-1A), all of which represented rural communities. Moreover, the villages in the South and East Administrative Regions were located in densely forested areas in close proximity to the habitat of wild chimpanzees and gorillas previously shown to be *Plasmodium*-infected (Figure 2-1B). Fecal samples were obtained from a total of 504 individuals including 37 Fulani pastoralists (individuals who raise livestock), 142 rainforest hunter-gatherers (individuals who forage for meat and plant materials), and 325 Bantu-speaking agro-pastoralists (individuals who grow crops and raise livestock). For 80 of these individuals, matching blood samples collected on the same day as the fecal samples were also available. All samples were obtained from asymptomatic subjects with no signs of clinical illness. Fecal samples (~5 g) were placed in sterile plastic containers without preservatives, frozen at -80 °C, and shipped to the US on dry ice. Blood samples (collected in EDTA tubes) were obtained from adults (~10 ml) and juveniles (~6 ml), but not from children 10 years of age or younger. For long-term storage, whole blood from a subset of individuals was also blotted onto Whatman filter

cards (GE Healthcare, USA). All samples were coded with an alphanumeric identifier to protect participant confidentiality. The study was approved by the Institutional Review Board of the University of Pennsylvania, the Cameroonian National Ethics Committee and the Cameroonian Ministry of Public Health. All subjects provided written informed consent for the collection and analysis of samples.

Extraction of fecal and blood DNA. Fecal DNA was extracted from ~220 mg aliquots using a modified “bead beating” method (Salonen et al., 2010) and the PSP Spin Stool DNA Plus Kit (Stratec Biomedical, Germany). Blood was processed in the field to isolate leukocytes (Miller et al., 1988), and DNA was extracted using the Puregene kit (Quigen, Germany). This was done to maximize DNA utilization for studies other than malaria. For quantitative PCR (qPCR) analysis of *P. falciparum* in blood samples, DNA from whole blood was extracted from three filter paper punches (3 mm in diameter) using the QIAamp DNA Mini Kit (Qiagen, Germany). DNA quality was determined using a Nanodrop, and the concentrations were determined using Qubit fluorometric quantitation.

Intensified PCR for *Plasmodium* detection. Fecal DNA was screened using highly cross-reactive primers previously shown to amplify a 956 bp mitochondrial cytochrome B (*cytB*) fragment of both *Laverania* and non-*Laverania* parasites (Liu et al., 2010a, 2017). Studies of wild apes have shown that fecal samples contain only very limited quantities of parasite DNA (Liu et al., 2010a). To increase the likelihood of parasite detection, we intensified our PCR screening by testing multiple aliquots of the same fecal DNA using the same primers and cycling conditions (Liu et al., 2017). Depending on the amount of starting material, we performed between 10 and 20 independent PCRs per sample. This intensified approach was also used to screen fecal DNA for *P. vivax* and *P. malariae*

sequences, although in each case only four DNA aliquots were tested due to limited sample availability. For *P. vivax*, lineage-specific primers were used to amplify a 295 bp fragment of the cytochrome oxidase 1 (*cox1*) gene (Liu et al., 2014), while for *P. malariae* a 600 bp *cytB* fragment was targeted for amplification (Liu et al., 2017). DNA from matched, leukocyte-enriched blood samples, which were available for nine fecal *Plasmodium*-positive and 70 fecal *Plasmodium*-negative subjects from five villages (AV, BI, BO, MI, ND), were screened by regular PCR for *Laverania*, *P. vivax* and *P. malariae* infections. This was also done for a paired dried blood spot DNA from one additional fecal-positive individual. All amplicons were sequenced using Sanger technologies, and amplicons with ambiguous bases were subsequently Illumina MiSeq sequenced using Nextera library preparation followed by read mapping to the appropriate *Plasmodium* reference sequence for base calling and variant identification.

Differentiation of ape and human *P. malariae*. To increase the number of *P. malariae* reference sequences, we selected stored blood and fecal samples of infected apes (Liu et al., 2010a, 2014, 2016) and humans (Sundararaman et al., 2013) from our specimen bank previously shown to harbor this parasite. Using single template PCR and published primer sets, we amplified 3.4 kb (Liu et al., 2010a), 2.5 kb (Liu et al., 2014), or 956 bp (Liu et al., 2010a) fragments of the *P. malariae* mitochondrial genome. Two single nucleotide variants (SNVs) were used to assess whether *P. malariae* parasites identified in the blood ($n=18$) and fecal ($n=1$) samples of rural Cameroonians were of likely human or ape origin. SNV1 was amplified using *P. malariae*-specific *cox1* primers, which generated a 104 bp fragment. Fecal DNA (2.5 μ l) was amplified in a 25 μ l reaction volume using 40 μ M dNTPs, 20 pmol of forward (ApePm_SNP1_oF 5'-ATTTTATCTACAGCTGCTGAATTT-3') and reverse (ApePm_SNP1_oR 5'-TGTAATTAATAATGACCATGTTGATA-3') primers, 1x

PCR buffer, and 0.25 µl of Taq from the Expand Long Template enzyme kit (Roche, USA). First round cycling conditions included an initial hot start of 2 min at 94 °C, followed by 15 cycles of denaturation (94 °C, 10 s), annealing (45 °C, 30 s) and elongation (68 °C, 2 min), followed by 35 cycles of denaturation (94 °C, 10 s), annealing (48 °C, 30 s) and elongation (68 °C, 2 min, with 15 s increments for each successive cycle), followed by a final elongation step of 10 min at 68 °C. A 1 µl aliquot of the first round product was amplified using forward (ApePm_SNP1_iF 5'-TATCCACCATTAAGTACTTCTCTTAT-3') and reverse (ApePm_SNP1_iR 5'-ACCTAATGTTAATCCTTTTGATCTTA-3') primers in the second round PCR. Second round cycling conditions included an initial denaturation step of 2 min at 94 °C, followed by 60 cycles of denaturation (94 °C, 10 s), annealing (52 °C, 30 s) and elongation (68 °C 1 min), followed by a final elongation step of 10 min at 68 °C. SNV2 was amplified using pan-*Plasmodium* or *P. malariae*-specific *cytB* primer pairs as previously described (Liu et al., 2010a; Liu et al., 2017), or by targeting a 110 bp *cytB* fragment using primers ApePm_SNP2_oF (5'-TTTTACCATTTTATGCAATGTTAAAA-3') and ApePm_SNP2_oR (5'-AAATGAAAATTTCTTGTGGTAATTGA-3') in the first round of PCR, and ApePm_SNP2_iF (5'-AATACCTAGTAAACAGCAGGTT-3') and ApePm_SNP2_iR (5'-AATGAACACATAAACCATATAATTGG-3') in the second round. Amplification conditions were the same as for SNV1.

Mammalian species determination A subset of human fecal samples ($n=27$) was subjected to mitochondrial DNA (D-loop) analysis to confirm their host species origin as previously described (Gao et al., 1999). To test for bushmeat consumption, the same samples were also screened for mitochondrial DNA from the African brush-tailed porcupine (*Atherurus africanus*), targeting a 219 bp *cytB* fragment using primers Aafricanuscytb_1F (5'-CTCCTYAAAATCATTAACCACTCATT-3') and Aafricanuscytb_1R

(5'-GTTGCTATCACYGTAAGTAGTAATA-3') in the first, and Aafricanuscytb_2F (5'-AATATCTCAGRATGATGAAACTTC-3') and Aafricanuscytb_2R (5'-ATATTTCAAGTTTCYGTGAATGTGTA-3') in the second round of PCR. Amplification conditions were the same as for *P. malariae* SNV1.

Fecal testing for intestinal parasites and occult blood. Fecal samples were analyzed by wet-mount fecal microscopy with and without iodine staining in the field to identify visible gastrointestinal parasites or parasite ova, including human whipworm (*Trichuris trichiura*), hookworm (*Ancylostoma duodenale*), giant roundworm (*Ascaris lumbricoides*), and amebiasis (*Entamoeba histolytica*). A subset of stored, frozen fecal samples ($n=84$) were later thawed on ice and screened for occult blood using the Hemosure test (Hemosure, USA), a qualitative immunochemical test that detects human haemoglobin in stool samples, according to the manufacturer's recommendations. Positive and negative control samples were processed in parallel.

Quantitative PCR to determine *P. falciparum* DNA levels in blood samples. To quantify *P. falciparum* DNA in blood samples, we developed a quantitative Taqman-based PCR assay targeting a 141 bp fragment of the parasite mitochondrial *cytB* region. DNA from dried blood spots, which were available from eight fecal *Plasmodium*-positive and 20 fecal *Plasmodium*-negative individuals, was incubated in a 25 μ l reaction volume containing 1x Taqman Fast Universal MasterMix (Invitrogen), 0.4 μ M of both forward (cytb_qpcr1_F 5'-GAGAATTATGGAGTGGATGGTGT-3') and reverse (cytb_qpcr1_R 5'-AGACATAACCAACGAAAGCAGT-3') primers, and a FAM-labeled Taqman probe (6FAM-ACATGCACGCAACAGGTGCT-TAMRA). PCR conditions included incubation for 2 min

at 50 °C, followed by a hot start of 10 min at 95 °C, followed by 45 cycles of 15 s at 95 °C and 45 s at 58 °C. A qPCR standard curve was generated by diluting known quantities of *P. falciparum* 3D7 DNA in human genomic DNA. All samples were run in triplicate alongside positive and negative controls. *Plasmodium falciparum* copies were averaged and normalized by the amount of input DNA (ng).

Phylogenetic analyses. Parasite and host mitochondrial sequences were aligned using CLUSTAL W (Thompson et al., 1994) implemented in Geneious 9.1.2 (Kearse et al., 2012) with regions that could not be unambiguously aligned removed from subsequent analyses. Maximum likelihood phylogenetic trees were estimated using PhyML (Guindon et al., 2010) with evolutionary models selected by jModelTest (Darriba et al., 2012).

Statistical analyses. Differences in gastrointestinal parasite detection between fecal *Plasmodium*-positive and -negative individuals were assessed using a one-sided Fisher's exact test after correcting for multiple tests using a Bonferroni correction. Differences in hemoglobin detection between fecal *Plasmodium*-positive and fecal *Plasmodium*-negative individuals were also assessed using a one-sided Fisher's exact test. Differences in fecal *Plasmodium* detection between children and adults were assessed using a one-sided Fisher's exact test. Blood parasite loads (determined by qPCR) of fecal *Plasmodium*-positive and -negative individuals were compared using a one-sided Mann-Whitney test.

Accession numbers. All newly derived *Plasmodium* and host mitochondrial sequences have been deposited in GenBank under accession numbers MF693406 - MF693457 (also see Tables S3-1).

2.4 Results

Intensified PCR identifies *P. falciparum*, but not ape *Laverania* parasites, in fecal samples from rural Cameroonians Previous analyses of humans from rural Gabon and Cameroon failed to identify evidence of zoonotic ape *Plasmodium* infections (Sundararaman et al., 2013; Delicat-Loembet et al., 2015; Ngoubangoye et al., 2016). However, since only blood samples were tested, the possibility of abortive (non-blood stage initiating) liver infection was not explored. Here, we analyzed fecal samples from 504 Fulani pastoralists, rainforest hunter-gatherers and Bantu-speaking agro-pastoralists from nine rural communities in Cameroon (Table 2-1). Most of these communities were located in forested areas in the immediate vicinity of the habitat of wild apes (Figure 2-1A), which were previously shown to harbor both *Laverania* and non-*Laverania* infections (Liu et al., 2010a, 2014). Although the intensity of *Plasmodium* transmission in the nine villages was not determined, this close proximity, in addition to the fact that most villagers from southern field sites spent many hours every day in or adjacent to the forest (Figure 2-1B), suggested that these communities were at risk of exposure to *Anopheles* spp. that carry ape parasites.

Fecal DNAs were first screened with highly cross-reactive primers previously shown to amplify a 956 bp mitochondrial *cytB* fragment from both *Laverania* and non-*Laverania* parasites (Liu et al. 2010a, 2017). Because fecal samples frequently contain only limited amounts of *Plasmodium* DNA, we intensified our PCR screening to maximise parasite detection (Liu et al., 2017). Performing 10 to 20 independent PCRs for each DNA sample (depending on sample availability), we amplified human *Plasmodium* sequences from the stool of 15 individuals (Table 2-1). For three of these subjects, more than half of 20 PCR replicates were positive, suggesting a relatively high fecal parasite burden (Table

2-2). However, for most other subjects fecal *Plasmodium* PCR was at or near the limits of detection, with five samples yielding only a single positive reaction (Table 2-2).

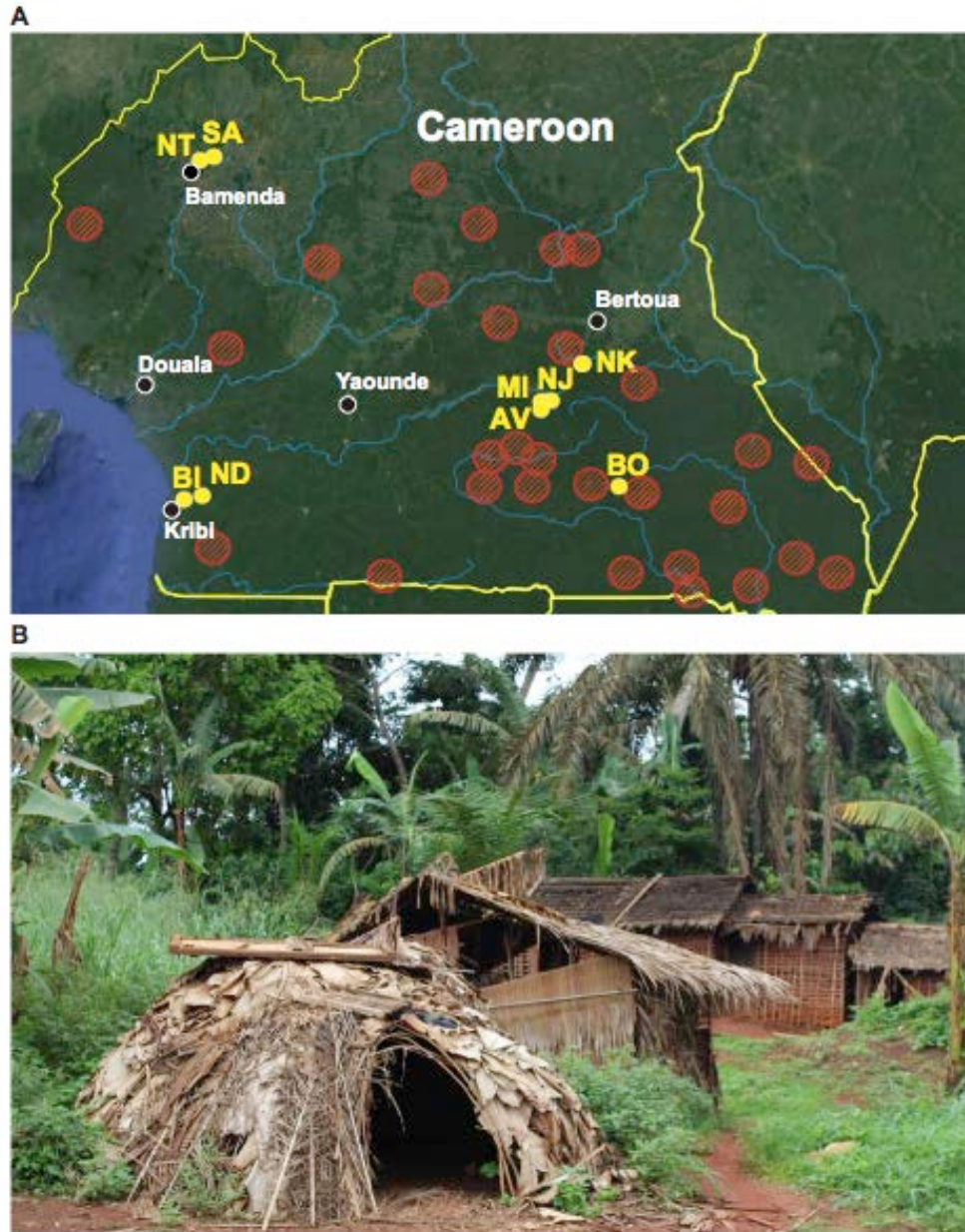


Figure 2-1. Human sampling sites in Cameroon. (A) The locations of nine rural villages (yellow circles) are shown in relation to field sites (red hatched circles) where wild-living chimpanzees and gorillas have previously been shown to harbor *Plasmodium* parasites (Liu et al., 2010a, 2014; Loy et al., 2017). Villages are denoted by a two-letter code (NT, Ntambang; SA, Sabga; BI, Bidou I; ND, Ndtoua; NK, Nkolbikong; MI, Missoume; NJ, Njibot; AV, Aviation; BO, Bosquet). Major cities, national boundaries and rivers are shown

in black, yellow and blue, respectively. (B) Image of a representative village with a large population of hunter-gatherers (photograph by Meagan A. Rubel).

Table 2-1. Detection of *Plasmodium* DNA in fecal samples of subjects from rural Cameroon

Village (Code)	Fecal samples screened	<i>Plasmodium</i> species ^a						
		<i>P. falciparum</i>	<i>P. malariae</i>	<i>P. ovale wallikeri</i>	Ape <i>Laverania</i>	Ape <i>P. vivax</i>	Ape <i>P. malariae</i>	Other spp.
Aviation (AV)	19	2	0	0	0	0	0	0
Bidou I (BI)	68	3	0	0	0	0	0	0
Bosquet (BO)	51	6	1 ^b	1	0	0	0	0
Missoume (MI)	31	1	0	0	0	0	0	0
Ndtoua (ND)	65	2	0	0	0	0	0	1 ^c
Njibot (NJ)	20	0	0	0	0	0	0	0
Nkolbikong (NK)	24	0	0	0	0	0	0	0
Ntambang (NT)	174	0	0	0	0	0	0	0
Sabga (SA)	52	0	0	0	0	0	0	0
Total	504	14	1	1	0	0	0	1

^a*Plasmodium* species were identified by direct amplicon sequencing and phylogenetic analysis.

^bSample BO-6 contained both *P. falciparum* (956 bp *cytB* fragment) and *P. malariae* (956 bp and 600 *cytB* fragments) sequences.

^cSample ND-1 yielded a rodent *Plasmodium* sequence (295 bp *cox1* fragment; see Figure 2-4).

All amplicons were sequenced directly and subjected to phylogenetic analysis. Fecal samples from 14 individuals contained *P. falciparum*, one of which (BO-6) also contained *P. malariae*, while the remaining sample (BO-3) contained *P. ovale wallikeri* (Figure 2-2). Replicate PCRs from the same samples yielded identical sequences, except for BI-2 and BO-2, both of which yielded one amplicon that contained two *P. falciparum* variants in approximately equal proportions, differing by one and two nucleotides, respectively (Figure 2-2). Analysis of available blood samples confirmed the fecal results, identifying additional *P. malariae* co-infections in three individuals that were not detected by fecal analysis (Table 2-2). None of the 504 fecal samples yielded sequences from ape *Laverania* parasites.

Table 2-2. Relative parasite burden in *Plasmodium* positive human fecal samples

Subject	Age	Ethnicity	Fraction of PCR positives using pan- <i>Plasmodium</i> primers ^a	Fraction of PCR positives using <i>P. malariae</i> primers ^a	Fraction of PCR positives using <i>P. vivax</i> primers ^a	<i>Plasmodium</i> species in fecal sample ^b	<i>Plasmodium</i> species in matching blood ^b
AV-1	32	Baka	1/20	0/4	0/4	<i>Pf</i>	<i>Pf</i>
AV-2	7	Maka	3/20	0/4	0/4	<i>Pf</i>	n/a
BI-1	44	Fang	1/20	0/4	0/4	<i>Pf</i>	<i>Pf</i>
BI-2	50	Fang	4/20	0/4	0/4	<i>Pf</i>	<i>Pf, Pm</i>
BI-3	11	Bagyeli	8/20	0/4	0/4	<i>Pf</i>	<i>Pf, Pm</i>
BO-1	20	Baka	1/20	0/4	0/4	<i>Pf</i>	<i>Pf</i>
BO-2	15	Baka	1/20	0/4	0/4	<i>Pf</i>	<i>Pf</i>
BO-3	7	Baka	4/20 ^c	0/4	0/4	<i>Pow</i> ^c	n/a
BO-4	5	Baka	16/20	0/4	0/4	<i>Pf</i>	n/a
BO-5	28	Baka	7/20	0/4	0/4	<i>Pf</i>	<i>Pf</i>
BO-6	7	Baka	3/20 ^d	3/4	0/4	<i>Pf, Pm</i>	n/a
BO-7	8	Baka	20/20	0/4	0/4	<i>Pf</i>	n/a
MI-1	17	Maka	1/20	0/4	0/4	<i>Pf</i>	<i>Pf, Pm</i>
ND-1	30	Bagyeli	0/20	0/4	1/4 ^e	rodent spp.	neg
ND-2	6	Bagyeli	14/20	0/4	0/4	<i>Pf</i>	n/a
ND-3	60	Fang	4/20	0/4	0/4	<i>Pf</i>	<i>Pf</i>

^aNumber of positive reactions per total number of PCR replicates of the same DNA sample.

^b*Pf*, *P. falciparum*, *Pow*, *P. ovale wallikeri*, *Pm*, *P. malariae*, n/a, blood sample not available; neg, negative for *Plasmodium* sequences.

^cAll four positive PCR reactions yielded *P. ovale wallikeri* sequences due to primer cross-reactivity.

^dTwo of three positive PCR reactions yielded *P. falciparum* sequences, while the third yielded *P. malariae* sequences.

^eOne positive PCR reaction yielded a *P. atheruri*-like sequence (see Figure 2-4A).

sample, only distinct haplotypes are shown (two haplotypes from the same fecal sample are indicated as v1 and v2, respectively). Fecal samples for which matching blood samples were not available are indicated by stars. Brackets indicate *Plasmodium* spp. (*Pf*, *Plasmodium falciparum*; *Po*-related, *Plasmodium ovale*-related, *Poc*, *Plasmodium ovale curtisi*, *Pow*, *Plasmodium ovale wallikeri*, *Pm*-related, *Plasmodium malariae*-related, *Pm*, *P. malariae*; *Pv*, *Plasmodium vivax*). The tree was constructed using PhyML (Guindon et al., 2010) with TIM2+I as the evolutionary model. Bootstrap values $\geq 80\%$ are shown (the scale bar represents 0.01 nucleotide substitutions per site). GenBank accession numbers for individual sequences are listed in Table S3-1.

Fecal samples from rural Cameroonians contain *P. malariae* but not *P. vivax* sequences. To specifically search for non-*Laverania* parasites, all 504 human fecal samples were rescreened using *P. vivax* and *P. malariae*-specific primer pairs. Again, each fecal sample was subjected to repeat PCR testing, this time using four DNA aliquots per primer set. This analysis confirmed the presence of *P. malariae* in the fecal sample of subject BO-6, but failed to identify additional positive samples (Table 2-2).

Plasmodium vivax-specific primers identified one positive sample, but sequence analysis of the respective amplicon identified a porcupine rather than a human parasite (see Section 3.3 below). Thus, despite the wide distribution of *P. vivax* among chimpanzees and gorillas in southern Cameroon (Liu et al., 2010a), and the recent finding of *P. vivax* infection in both Duffy-positive and -negative humans in the same region (Frucho et al., 2014; Mbenda and Das, 2014; Russo et al., 2017; Zimmerman, 2017), *P. vivax* sequences were not amplified from the fecal samples of any of these individuals.

To investigate the possible host origin of the newly identified *P. malariae* and *P. ovale wallikeri* parasites, we constructed a phylogenetic tree from available *cytB* sequences (Figure 2-2). This analysis showed that both *P. malariae* and *P. ovale wallikeri* sequences failed to form host-specific lineages, most likely because existing genetic information is too limited to differentiate ape and human parasites (Figure 2-2). We thus performed additional PCR amplifications to increase the number (and in some cases the

length) of ape- and human-derived *P. malariae* sequences (this was not possible for *P. ovale wallikeri* due to a lack of positive samples). Using limiting dilution PCR, we amplified mitochondrial fragments (3.4kb, 2.5kb, 956 bp) from stored ape ($n=9$) and human ($n=16$) samples previously shown to contain this parasite (Liu et al., 2010a, 2014; Sundararaman et al., 2013). An alignment of this expanded sequence set revealed two SNVs spaced 1,906 bp apart that distinguished human from most ape-derived *P. malariae* strains (Figure 2-3). All human-derived *P. malariae* parasites contained a T at position 276 in the *cox1* gene (SNV1), and a T or A at position 690 in the *cytB* gene (SNV2), while nearly all *P. malariae* parasites from wild apes contained a C at these two positions (Figure 2-3); the single exception was a parasite from a habituated chimpanzee (Leo) sampled in the Tai Forest, which contained an A in *cytB* (SNV1 was not sequenced).

Since two bonobos and three chimpanzees living in sanctuaries also harbored *P. malariae* strains with human-specific SNVs, it appears that captive apes are susceptible to infection with human *P. malariae* parasites, a finding consistent with experimental infections of chimpanzees (Rodhain, 1948). Whether humans can become infected with ape *P. malariae* under natural conditions remains unknown; however, the fact that the *P. malariae* strain in the BO-6 fecal sample was identical to all other human parasites, including *P. malariae* sequences amplified from matching blood samples, suggests that it is of human origin (Figure 2-3).

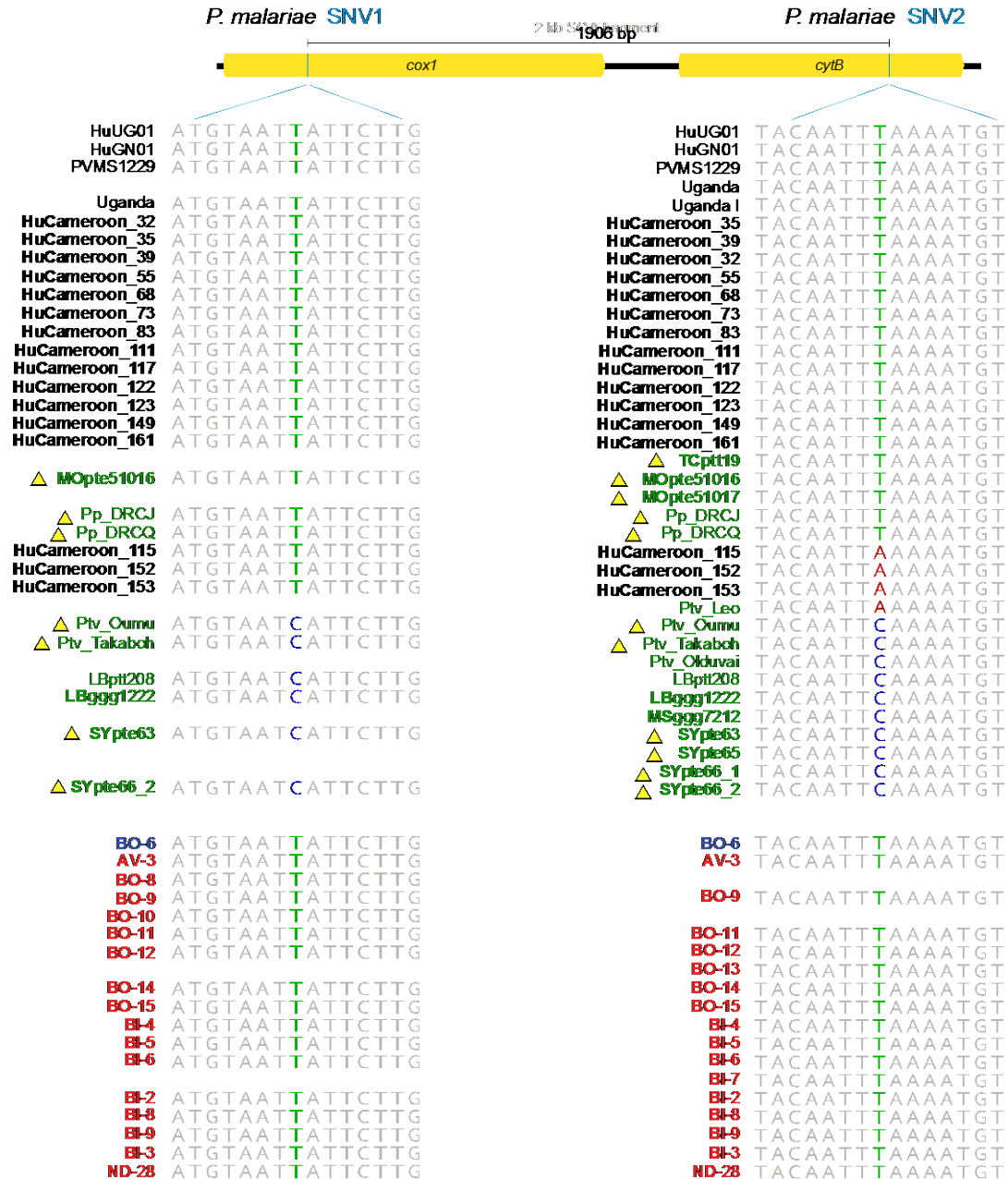


Figure 2-3. Single nucleotide variants distinguishing human and ape *Plasmodium malariae* parasites. A schematic representation of a portion of the *P. malariae* mitochondrial genome is shown (black line), with yellow boxes denoting *cox1* and *cytB* coding regions. Blue lines indicate the locations of two single nucleotide variants (SNV1 and SNV2) spaced 1,906 bp apart, which differentiate human (black) from most ape-derived (green) *P. malariae* sequences. Newly derived *P. malariae* reference sequences from humans (Hu), bonobos (Pp), chimpanzees (ptv; *Pan troglodytes verus*; ptt; *Pan troglodytes troglodytes*; pte, *Pan troglodytes ellioti*) and gorillas (ggg, *Gorilla gorilla gorilla*) are shown, with newly derived sequences in bold and sequences from captive apes

indicated with a yellow triangle (see Table S3-1 for a more detailed description of each sequence). *P. malariae* sequences amplified from the fecal (blue) and blood (red) samples of rural Cameroonians are shown at the bottom.

Detection of a porcupine *Plasmodium* parasite in the fecal sample of a hunter-gatherer. Although the *P. vivax*-specific primers failed to uncover *P. vivax* infection, they amplified a 295 bp *cox1* sequence (Figure 2-4A) which was very similar to the corresponding sequence of *Plasmodium atheruri*, a parasite that infects African brush-tailed porcupines (*Atherurus africanus*) (Van Den Berghe et al., 1958; Pacheco et al., 2011). Since this parasite was detected in the fecal sample of a hunter-gatherer (ND-1), we investigated whether the subject was productively infected with the rodent parasite (Table 2-2). Screening a blood sample collected on the same day, we failed to amplify *P. atheruri* or any other *Plasmodium* sequences from this individual. To exclude sample mix-up, we subjected the ND-1 fecal sample to host mitochondrial analysis (Gao et al., 1999). Amplifying a hypervariable D-loop fragment, we confirmed that this and 26 other fecal samples from hunter-gatherers of the same ethnic group sampled at the same village contained human mitochondrial sequences (Figure 2-4B). Reasoning that the detected parasite may have infected consumed bushmeat, we designed porcupine-specific mitochondrial primers and used these to probe the same fecal samples. Interestingly, this analysis revealed porcupine mitochondrial DNA in the feces of eight hunter-gatherers, including subject ND-1 (Figure 2-4C). Since two different mitochondrial DNA haplotypes were identified, it is likely that these individuals consumed at least two porcupines in the days prior to fecal testing. This is not surprising given that porcupines are frequently hunted for food in this and other regions in western central Africa (Jori et al., 1998; Gaubert et al., 2015). Thus, the *P. atheruri*-like parasite, which was detected in the stool, but not in

the blood, of subject ND-1, most likely infected a prey animal whose sequences survived passage through the human digestive tract (De Nys et al., 2015). Since great apes are also hunted for bushmeat, a similar scenario would have to be excluded if ape parasite sequences were detected in human fecal samples.

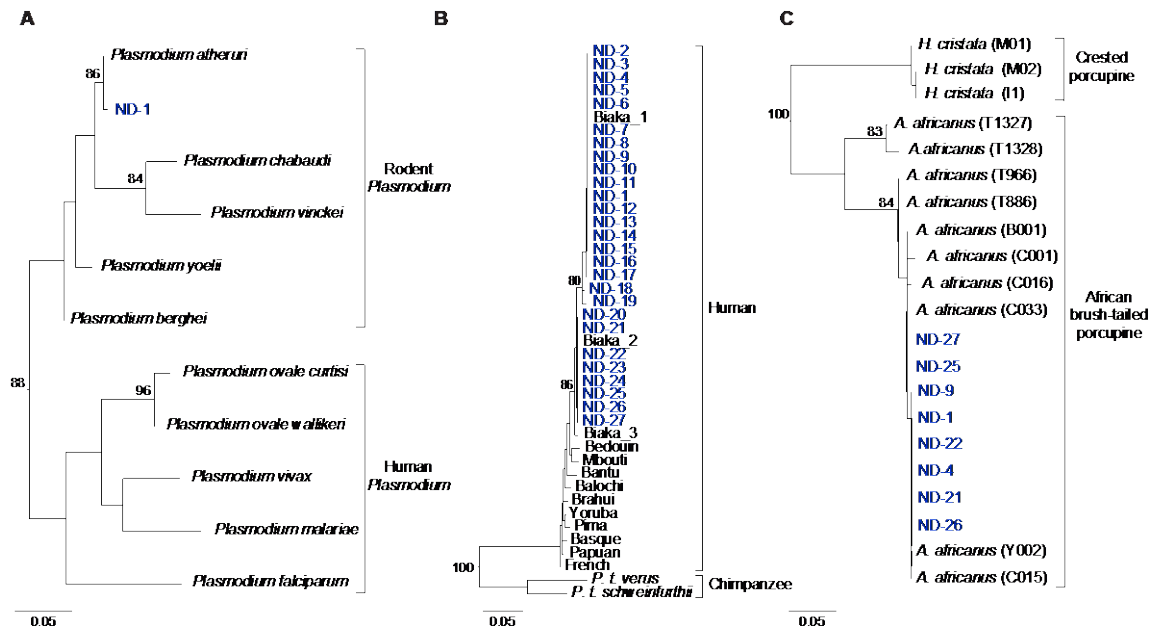


Figure 2-4. Detection of a porcupine parasite in human stool. (A) The phylogenetic position of a *Plasmodium* sequence (295 bp *cox1* fragment) amplified from the fecal sample of a human hunter-gatherer (blue) is shown in relation to rodent and human *Plasmodium* reference sequences (black). (B) Confirmation that fecal samples from 27 hunter-gatherers (blue) are of human origin. Newly derived mitochondrial sequences (500 bp D-loop fragment) are shown in relation to published human and chimpanzee reference sequences (black). (C) Fecal samples of nine hunter-gatherers (blue) contained porcupine mitochondrial DNA. Amplified sequences (219 bp *cytB* fragment) are shown in relation to reference sequences for African brush-tailed (*Atherurus africanus*) and crested (*Hystrix cristata*) porcupines (black). All trees were constructed using PhyML (Guindon et al., 2010) with TIM1+I (A), TPM3uf+I+G (B) and HKY+G (C) as the evolutionary models. Bootstrap values $\geq 80\%$ are shown (the scale bars represents 0.05 nucleotides substitutions per site). GenBank accession numbers for individual sequences are listed in Table S3-1.

Fecal detection of *Plasmodium* DNA is indicative of concurrent parasitemia. To determine whether any of the other individuals who were fecal *Plasmodium*-positive lacked a corresponding blood infection, we screened all subjects for whom blood samples were available (blood was not collected from six fecal-positive children). Using pan-

Laverania-, *P. malariae*- and *P. vivax*-specific primers, we identified *P. falciparum* in all paired blood samples (Figure 2-2), three of which (MI-1, BI-2, and BI-3) also contained *P. malariae* parasites (Table 2-2). None of the blood samples was *P. vivax*-positive. All blood-derived *P. falciparum* sequences were identical to the corresponding fecal parasite sequences (Figure 2-2). Thus, for all subjects for whom matching fecal and blood samples could be tested, we found that the detection of *Plasmodium* DNA in the fecal sample was indicative of a concurrent blood stage infection.

Fecal *Plasmodium* testing underestimates blood stage infections. The fact that all fecal PCR-positive individuals (except for subject ND-1) were also blood stage-infected suggested that fecal *Plasmodium* DNA derived from blood parasites. However, it remained unclear to what extent fecal-negative individuals had blood stage infections. To address this question, we screened blood samples from a representative subset of fecal-negative individuals ($n=70$) from the five villages where fecal-positive individuals were identified (AV, BI, BO, MI, ND). Using all three *Plasmodium* primer sets, we identified parasite sequences in a large fraction (53%) of these individuals. The most common infection was *P. falciparum*, which was detected in the blood of 28 of the 70 fecal-negative individuals (Table 2-3). *Plasmodium malariae*, *P. ovale wallikeri* and *P. ovale curtisi* were also detected, many as coinfections with *P. falciparum* (Table 2-3). Of note, all blood-derived *P. malariae* sequences contained the human-specific SNVs, and none of the 70 blood samples was *P. vivax*-positive. These results show that the great majority of blood stage-infected individuals do not have detectable parasite DNA in their stool, suggesting that PCR-based fecal screening is of limited utility for malaria epidemiological studies.

Table 2-3. *Plasmodium* species in the blood of fecal *Plasmodium* negative individuals

Village	Samples screened	<i>Plasmodium</i> species ^a							
		<i>Pf</i>	<i>Pm</i>	<i>Poc</i>	<i>Pow</i>	<i>Pv</i>	<i>Pf/Pm</i> ^b	<i>Pf/Poc</i> ^b	<i>Pf/Pow</i> ^b
AV	10	3	0	0	0	0	1	0	0
BI	25	6	3	1	0	0	4	2	1
BO	25	3	5	0	0	0	4	0	0
MI	5	0	0	0	0	0	0	0	0
ND	5	3	0	0	0	0	1	0	0
Total	70	15	8	1	0	0	10	2	1

^a*Plasmodium* species were identified by direct amplicon sequencing and phylogenetic analysis. *Pf*, *P. falciparum*; *Pm*, *P. malariae*; *Pow*, *P. ovale wallikeri*; *Poc*, *P. ovale curtisi*; *Pv*, *P. vivax*.

^bDual infections

Fecal *Plasmodium* detection is not secondary to gastrointestinal bleeding. The finding that most blood-positive individuals were fecal *Plasmodium*-negative raised the possibility that fecal parasite DNA was secondary to intestinal bleeding. Gastrointestinal parasites are known to cause bloody stools and are frequently detected in humans from Cameroon (Tchuem-Tchuente et al., 2012; Karagiannis-Voules et al., 2015). To examine whether bleeding caused by intestinal parasites was a primary reason for fecal *Plasmodium* positivity, we compared the percentage of *Plasmodium*-positive and -negative stool samples that were also positive for gastrointestinal parasites for individuals from five villages (AV, BI, BO, MI, ND). Although fecal *Plasmodium*-positive individuals tended to have higher rates of whipworm (*Trichuris trichiura*), hookworm (*Ancylostoma duodenale* or *Necator americanus*), and giant roundworm (*Ascaris lumbricoides*) infections than fecal *Plasmodium*-negative individuals, these differences were not statistically significant (Table 2-4). Separate analyses of adults and children yielded the same result (not shown). Thus, the presence of *Plasmodium* DNA in fecal samples was not secondary to infection with gastrointestinal parasites.

Table 2-4. Prevalence of gastrointestinal parasites in *Plasmodium* fecal positive and negative subjects

Intestinal Parasite	Fecal <i>Plasmodium</i> positive			Fecal <i>Plasmodium</i> negative			p value ^a
	Number of samples	Number positive	% positive	Number of samples	Number positive	% positive	
Whipworm	15	6	40	218	56	26	0.90
Giant roundworm	15	8	53	218	59	27	0.17
Amebiasis	15	0	0	218	9	4	1.0
Hookworm	15	3	20	218	13	6	0.36
Any parasite	15	10	67	218	88	40	0.22

^ap values were calculated using a one-tailed Fisher's exact test with Bonferroni correction.

To examine whether intestinal bleeding due to any cause was associated with fecal *Plasmodium* detection, we examined the 15 *Plasmodium*-positive stool samples as well as 69 *Plasmodium*-negative stool samples from 37 blood-positive and 32 blood-negative individuals for occult blood. Two of the *Plasmodium*-positive and two of the *Plasmodium*-negative fecal samples were Hemosure positive; however, these differences were not statistically significant regardless of whether all (2/15 versus 2/69; Fisher's exact test, $P = 0.14$) or only the parasitemic individuals (2/15 versus 0/37; Fisher's exact test, $P = 0.08$) were included in the analysis. Thus, while gastrointestinal bleeding may in some instances increase the likelihood of fecal *Plasmodium* detection, occult blood does not explain the presence of parasite sequences in most stool samples.

Fecal *Plasmodium* detection is associated with parasitemia. Comparing the age distribution of fecal PCR-positive individuals, we noted that parasite detection was disproportionately higher in fecal samples of children (6/15, 40%) than in the stool of adults (9/204, 4%) (Fisher's exact test, $P < 0.0001$). Since even among asymptomatic individuals parasitemia tends to decrease with age (Baird, 1998; Bousema et al., 2004; Balaraine et

al., 2009; Zhou et al., 2016), we hypothesized that fecal *Plasmodium* positivity may be associated with higher parasite titers in the blood. To examine this possibility, we developed a qPCR assay and used it to determine the copy number of *P. falciparum* in dried blood spot samples available for eight fecal-positive and 20 fecal-negative individuals. Testing all samples in triplicate, we found 26-fold more parasite DNA in the blood of fecal-positive than fecal-negative subjects (Figure 2-5). Thus, individuals with high *Plasmodium* blood titers were more likely to have detectable parasite DNA in their stool than individuals with low parasitemia.

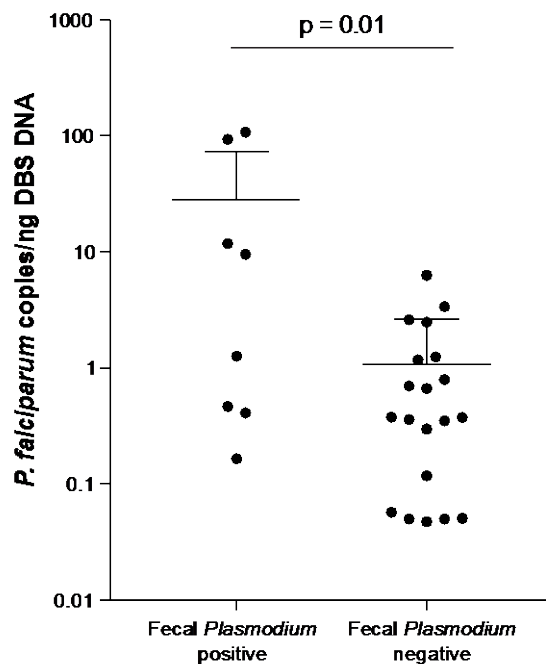


Figure 2-5. Fecal *Plasmodium*-positive individuals have higher blood parasitemia than fecal *Plasmodium*-negative individuals. Blood parasite loads were determined for fecal *Plasmodium*-positive ($n = 8$) and -negative ($n = 20$) individuals by Taqman quantitative PCR. Values are expressed as copies of *P. falciparum* per ng of whole blood DNA derived from dried blood spots (DBS). Each sample was tested in triplicate. Two samples that were below the limit of detection were set to background DNA levels (0.05). Error bars shows standard deviation from the mean and the P value was calculated using a one sided Mann-Whitney test.

2.5 Discussion

To examine the zoonotic potential of ape *Plasmodium* parasites, we screened fecal samples from humans living in close proximity of wild chimpanzees and gorillas for evidence of abortive liver infection. We reasoned that while receptor/ligand incompatibilities at the erythrocyte invasion stage may protect humans from blood stage infection, similar barriers may not exist to prevent invasion of hepatocytes. Despite using a highly sensitive PCR approach, we found no evidence for such a scenario. Although we amplified *P. falciparum*, *P. malariae*, and *P. ovale wallikeri* from human stool, none of the 504 samples tested was positive for ape *Plasmodium* DNA. This finding is unlikely the result of insensitive detection methods. Assuming that ape *Plasmodium* sporozoites delivered by a mosquito bite would infect at least five hepatocytes, liver stage schizonts should produce at least 150,000 merozoites (Shortt et al., 1951). Assuming further that at least 5% of these merozoites are cleared into the biliary system, this would yield ~75 parasites per gram of fecal material (the average adult excretes ~250 g of feces per day; Rose et al., 2015). Since we extracted ~0.22 g of fecal material and screened over 45 □ of the resulting DNA, ~45 mitochondrial genomes (assuming 20 copies per parasite) should have been present and readily amplified. Indeed, we detected a porcupine parasite after passage through the human gastrointestinal tract, indicating sufficient PCR sensitivity to find even rare sequences. However, the likelihood of amplifying exo-erythrocytic parasite DNA from human fecal samples depends primarily on the local transmission intensity, especially since abortive liver infections would only be detectable for a couple of days following an infective mosquito bite. In the absence of such information, we cannot exclude that our failure to detect ape parasites reflects a lack of human exposure to infective mosquitoes. Alternatively, ape parasites may be unable to replicate in the human liver and/or pre-erythrocytic parasite DNA may not be shed into the human stool. Until

entomological inoculation rates (EIRs) for *Anopheles* spp. that transmit ape *Plasmodium* parasites are determined for these communities, it is impossible to differentiate between these possibilities.

Previous studies of human communities in rural Cameroon reported *P. falciparum* EIRs to range between 17 and >500 infected bites per person per year (Njan Nloga et al., 1993; Meunier et al., 1999; Antonio-Nikidjio et al., 2005; Atangana et al., 2010). We thus expected a sizable fraction of our study population to have been exposed to mosquitoes carrying this parasite in the days prior to the fecal collection. Since only 15 of 504 individuals contained human *Plasmodium* sequences in their stool, we examined parasite titers in matching blood samples. This analysis revealed that all individuals with parasite DNA in their fecal samples were also blood stage-infected. While this alone does not argue against the shedding of pre-erythrocytic parasite DNA, the fact that fecal-positive individuals also had higher parasite titers in their blood strongly suggests that most fecal parasite DNA derives from infected erythrocytes. This is consistent with the observations that fecal and blood *Plasmodium* copy numbers rose concurrently in experimentally infected mice (Abkallo et al., 2014) and that a monkey solely infected with parasitized erythrocytes also had detectable *Plasmodium* DNA in its stool (Kawai et al., 2014). Moreover, *Laverania* parasites, which replicate to high blood titers, are more readily amplified from fecal material than non-*Laverania* parasites such as *P. vivax*, which exhibit only low levels of parasitemia (Liu et al., 2014). Although *P. yoelli* DNA was detected in the liver, gall bladder and stool of experimentally-infected mice several days before the onset of parasitemia, the sporozoite dose used in these studies (1,500 - 25,000 per mouse; Abkallo et al., 2014) was much higher than what would be delivered by a typical infective mosquito, which injects on average fewer than 100 sporozoites per bite (Medica

and Sinnis, 2005). Thus, it is possible that most, if not all, of the parasite DNA detected in fecal samples of naturally infected primates is derived from blood stage infection.

The routes by which *Plasmodium* DNA is shed into the human stool are not understood. One possibility is gastrointestinal bleeding, but the detection of occult blood in the stool of fecal *Plasmodium*-positive individuals was not significantly higher than that of fecal *Plasmodium*-negative individuals. Thus, while potentially contributing to parasite detection in some individuals, gastrointestinal bleeding is unlikely to be the primary cause for fecal *Plasmodium* DNA positivity. A more plausible scenario is excretion into the biliary system, as has been proposed for *P. yoelli*-infected mice (Abkallo et al., 2014). Given the central role of the liver in removing damaged erythrocytes from the circulation (Theurl et al., 2016), it is likely that parasitized red blood cells are cleared into the bile proportional to their numbers in the blood. This scenario is consistent with the fact that higher parasitemia levels are associated with higher *Plasmodium* detection rates in fecal samples.

Most indigenous people in central Africa are protected from *P. vivax* malaria because their erythrocytes lack the Duffy antigen receptor for chemokines (DARC), which is used by *P. vivax* merozoites to invade red blood cells (Miller et al., 1976). However, since ape *P. vivax* has been identified in a Duffy-positive European forest worker (Prugnolle et al., 2013), we reasoned that rural Cameroonians might acquire ape *P. vivax* liver infections, even if their Duffy-negative phenotype protected them from blood stage infection. This hypothesis was further supported by the recent finding of *P. vivax* in some Duffy-negative individuals in Cameroon, which could have, at least theoretically, been derived from the introduction of an ape parasite (Fru-Cho et al., 2014; Mbenda and Das, 2014; Russo et al., 2017; Zimmerman, 2017). To examine this possibility, we used species-specific primers to screen all 504 human fecal samples for *P. vivax* sequences.

Despite the wide distribution of ape *P. vivax* among chimpanzees and gorillas in Cameroon (Liu et al., 2014) as well as previous reports of *P. vivax*-specific seroreactivity in Duffy-negative individuals from the Republic of Congo (Culleton et al., 2009), we failed to amplify ape or human *P. vivax* sequences from any of the specimens tested, including 80 matched blood samples. Since the Duffy-negative phenotype is unlikely to protect humans against liver stage infection (Herrera et al., 2005; Culleton et al., 2009), our findings indicate that Cameroonian forest dwellers are either not exposed to ape *P. vivax*, which seems unlikely, or that abortive liver infection does occur, but is not detectable by fecal analyses. The resolution of this question may require development of new methods such as serological assays capable of detecting immune responses that are specific for ape *P. vivax* antigens.

We also examined the host species origin of *P. malariae* parasites that were detected in one fecal and 18 blood samples from our study cohort. Amplifying a 956 bp *cytB* fragment, we realized that this sequence was not sufficiently diverse to differentiate ape- and human-specific *P. malariae* lineages (Figure 2-2). We thus generated longer mitochondrial fragments from additional *P. malariae*-containing samples, identifying two single nucleotide variants that differentiated human from most ape *P. malariae* parasites. Both fecal- and blood-derived *P. malariae* sequences from humans contained a T at position 276 in the *cox1* gene (SNV1), and a T or A at position 690 in the *cytB* gene (SNV2), while *P. malariae* sequences from all but one wild ape contained Cs at these two positions (Figure 2-3). These data suggest that most wild apes harbor *P. malariae* strains that are distinct from those infecting humans and that the rural Cameroonians were infected with human, and not ape, parasites. However, it remains unknown whether human and ape *P. malariae* parasites represent two distinct *Plasmodium* spp., and if so, whether ape *P. malariae* strains would infect humans in natural settings. The same is true

for ape- and human-derived *P. ovale* parasites, which are even less well characterized due to the paucity of naturally occurring infections.

Non-invasive fecal-based methods have been instrumental in characterizing the molecular epidemiology of ape malaria parasites, identifying numerous new *Plasmodium* spp. that were not previously known to exist (Kaiser et al., 2010; Liu et al., 2010a, 2014, 2017; Prugnolle et al., 2010). Ape *Laverania* parasites, in particular, are readily detected in endemic areas such as Cameroon, with 18% to 25% of fecal samples from both chimpanzees and gorillas yielding parasite sequences when screened by diagnostic PCR (Liu et al., 2010a; Prugnolle et al., 2010). In light of these data, the fact that we detected *Plasmodium* DNA in only 3% of human fecal samples was surprising. Although we extracted DNA from a smaller fecal aliquot than is usually used for wild apes, this difference cannot account for the nearly 10-fold difference in detection frequency. The reason(s) for the much higher fecal parasite detection rates in wild apes is not known, but may be due to higher overall prevalence rates similar to what has been observed in wild macaques (Zhang et al., 2016), higher blood titers, more frequent gastrointestinal bleeding, or a combination of these factors. It is also possible that apes are much more frequently bitten by infective mosquitoes, which might result in the fecal shedding of larger quantities of pre-erythrocytic parasite DNA.

In summary, we report here the absence of ape *Plasmodium* sequences in fecal samples of rural Cameroonians at risk of exposure to such parasites. While our results failed to provide evidence for abortive liver infections, they do not rule out that such infections are, in fact, occurring. This is particularly true for ape *P. vivax*, which is prevalent in wild apes and has been shown to productively infect humans. One fundamental question, therefore, is whether liver-derived merozoites, which lack the ability to parasitize erythrocytes, are cleared through excretion into the biliary tract or through other pathways

such as phagocytic cells in the liver or other organs. A second fundamental question is how often and under what circumstance humans are bitten by mosquitoes that carry ape *Plasmodium* sporozoites. Given the critical role of non-invasive testing in identifying ape malaria parasites and its potential to probe barriers of cross-species infection, further experimentation, including longitudinal studies in macaques, will be required to determine the sources, kinetics and mechanisms of fecal parasite shedding. Ultimately, the utility of fecal testing in detecting exo-erythrocytic parasite DNA will have to be addressed in humans, possibly in the context of *P. falciparum* and *P. vivax* vaccination studies that use mosquito bites to challenge immunised individuals (Roestenberg et al., 2009; Herrera et al., 2011; Bijker et al., 2013; Spring et al., 2014; Dunachie et al., 2015; Arevalo-Herrera et al., 2016). Such studies, in combination with investigations of the distribution, ecology, and biting behavior of the mosquito vectors that transmit ape *Plasmodium* parasites, will greatly aid in determining the zoonotic potential of these parasites and the associated human infection risk.

2.6 Chapter acknowledgments

We thank André Essiane, Julius Y. Fonsah, Valentine N. Ndze, Eric N. Ngwang, Grace N. Tenjei and Peter Kfu for field work in Cameroon; William Beggs for sample coordination; Scott Sherrill-Mix for helpful discussions; the Cameroonian villagers for their willingness to participate in this research; and the Cameroonian Ministry of Public Health and the Cameroon National Ethics Committee for permission to collect samples in Cameroon. We also thank community outreach groups such as Mbororo Social and Cultural Association for their support, as well as the partnership collaboration with the Faculty of Medicine and Biomedical Sciences of the University of Yaounde 1, Cameroon. This work was supported by grants from the National Institutes of Health, USA (R01 AI 091595, R01 AI 058715,

R01 AI 120810, R37 AI 050529, P30 AI 045008, and R01 GM 113657), the Lewis and Clark Fund, USA, the Leakey Foundation, USA and the National Science Foundation USA (BCS-1540432 and BCS-1317217). D.E.L. (T32 AI 007532) and M.A.R. (T32 AI007532) were supported by training grants. The authors declare no competing financial interests.

Supplemental information available online:

<https://proxy.library.upenn.edu:2100/content/image/1-s2.0-S0020751918300249-mmc1.docx>

2.7 Chapter references

- Ayoubu, A., Mouacha, F., Learn, G.H., Mpoudi-Ngole, E., Rayner, J.C., Sharp, P.M., Hahn, B.H., Delaporte, E., and Peeters, M. 2012. Ubiquitous *Hepatocystis* infections, but no evidence of *Plasmodium falciparum*-like malaria parasites in wild greater spot-nosed monkeys (*Cercopithecus nictitans*). *Int. J. Parasitol.* 42, 709-713.
- Bhatt, S., Weiss, D.J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., Battle, K.E., Moyes, C.L., Henry, A., Eckhoff, P.A., Wenger, E.A., Briet, O., Penny, M.A., Smith, T.A., Bennett, A., Yukich, J., Eisele, T.P., Griffin, J.T., Fergus, C.A., Lynch, M., Lindgren, F., Cohen, J.M., Murray, C.L., Smith, D.L., Hay, S.I., Cibulskis, R.E., and Gething, P.W. 2015. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature* 526, 207-211.
- Blacklock, B., and Adler, S. 1922. A parasite resembling *Plasmodium falciparum* in a chimpanzee. *Ann. Trop. Med. Parasitol.* XVI, 99-106.
- Bopp, S.E.R., Manary, M.J., Bright, A.T., Johnston, G.L., Dharia, N.V., Luna, F.L., McCormack, S., Plouffe, D., McNamara, C.W., Walker, J.R., Fidock, D.A., Denchi, E.L., and Winzeler, E.A. 2013. Mitotic evolution of *Plasmodium falciparum* shows a stable core genome but recombination in antigen families. *PLoS Genet.* 9, e1003293.
- Boundenga, L., Ollomo, B., Rougeron, V., Mouele, L.Y., Mve-Ondo, B., Delicat-Loembet, L.M., Moukodoum, N.D., Okouga, A.P., Arnathau, C., Elguero, E., Durand, P., Liegeois, F., Boue, V., Motsch, P., Le Flohic, G., Ndougouet, A., Paupy, C., Ba, C.T., Renaud, F., and Prugnolle, F. 2015. Diversity of malaria parasites in great apes in Gabon. *Malar. J.* 14, 111.
- Bray, R.S. 1958. Studies on malaria in chimpanzees VI. *Laverania falciparum*. *Am. J. Trop. Med. Hyg.* 7, 20-24.
- Brumpt, E. 1939. Les parasites du paludisme des chimpanzes. *C. R. Soc. Bio.* 130, 837-840.

- Caldecott, J.O., and Miles, L. 2005. World Atlas of Great Apes and their Conservation. University of California Press, Berkley, California, USA.
- Carter, R. 2003. Speculations on the origins of *Plasmodium vivax* malaria. Trends Parasitol. 19, 214-219.
- Carter, R., and Mendis, K.N. 2002. Evolutionary and historical aspects of the burden of malaria. Clin. Microbiol. Rev. 15, 564-594.
- Coatney G.R., Collins W.E., Warren, M., Contacos P.G. 1971. The Primate Malaria. Washington, DC: US Government Printing Office.
http://phsource.us/PH/PARA/Book/primate_24.pdf
- Conway, D.J., Fanello, C., Lloyd, J.M., Al-Joubori, B.M., Baloch, A.H., Somanath, S.D., Roper, C., Oduola, A.M., Mulder, B., Pova, M.M., Singh, B., and Thomas, A.W. 2000. Origin of *Plasmodium falciparum* malaria is traced by mitochondrial DNA. Mol. Biochem. Parasitol. 111, 163-171.
- Culleton, R., and Carter, R. 2012. African *Plasmodium vivax*: distribution and origins. Int. J. Parasitol. 42, 1091-1097.
- De Nys, H.M., Calvignac-Spencer, S., Boesch, C., Dorny, P., Wittig, R.M., Mundry, R., and Leendertz, F.H. 2014. Malaria parasite detection increases during pregnancy in wild chimpanzees. Malar. J. 13, 413.
- De Nys, H.M., Calvignac-Spencer, S., Thiesen, U., Boesch, C., Wittig, R.M., Mundry, R., and Leendertz, F.H. 2013. Age-related effects on malaria parasite infection in wild chimpanzees. Biol. Lett. 9, 20121160.
- Delicat-Loembet, L., Rougeron, V., Ollomo, B., Arnathau, C., Roche, B., Elguero, E., Moukodoum, N.D., Okougha, A.P., Mve Ondo, B., Boundenga, L., Houze, S., Galan, M., Nkoghe, D., Leroy, E.M., Durand, P., Paupy, C., Renaud, F., and Prugnolle, F. 2015. No evidence for ape *Plasmodium* infections in humans in Gabon. PLoS One 10, e0126933.
- Duval, L., Fourment, M., Nerrienet, E., Rousset, D., Sadeuh, S.A., Goodman, S.M., Andriaholinirina, N.V., Randrianarivelosia, M., Paul, R.E., Robert, V., Ayala, F.J., and Arie, F. 2010. African apes as reservoirs of *Plasmodium falciparum* and the origin and diversification of the *Laverania* subgenus. Proc. Natl. Acad. Sci. U.S.A. 107, 10561-10566.
- Escalante, A.A., and Ayala, F.J. 1994. Phylogeny of the malarial genus *Plasmodium*, derived from rRNA gene sequences. Proc. Natl. Acad. Sci. U.S.A. 91, 11373-11377.
- Escalante, A.A., Cornejo, O.E., Freeland, D.E., Poe, A.C., Durrego, E., Collins, W.E., and Lal, A.A. 2005. A monkey's tale: The origin of *Plasmodium vivax* as a human malaria parasite. Proc. Natl. Acad. Sci. U.S.A. 102, 1980-1985.
- Hedrick, P.W. 2011. Population genetics of malaria resistance in humans. Heredity (Edinb) 107, 283-304.

- Herbert, A., Boundenga, L., Meyer, A., Moukodoum, D.N., Okouga, A.P., Arnathau, C., Durand, P., Magnus, J., Ngoubangoye, B., Willaume, E., Ba, C.T., Rougeron, V., Renaud, F., Ollomo, B., and Prugnolle, F. 2015. Malaria-like symptoms associated with a natural *Plasmodium reichenowi* infection in a chimpanzee. *Malar. J.* 14, 220.
- Howes, R.E., Patil, A.P., Piel, F.B., Nyangiri, O.A., Kabaria, C.W., Gething, P.W., Zimmerman, P.A., Barnadas, C., Beall, C.M., Gebremedhin, A., Menard, D., Williams, T.N., Weatherall, D.J., and Hay, S.I. 2011. The global distribution of the Duffy blood group. *Nat. Commun.* 2, 266.
- Hughes, A.L., and Verra, F. 2001. Very large long-term effective population size in the virulent human malaria parasite *Plasmodium falciparum*. *Proc. Biol. Sci.* 268, 1855-1860.
- Ioerger, T.R., Clark, A.G., and Kao, T.H. 1990. Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. *Proc. Natl. Acad. Sci. U.S.A.* 87, 9732-9735.
- Jongwutiwes, S., Putaporntip, C., Iwasaki, T., Ferreira, M.U., Kanbara, H., and Hughes, A.L. 2005. Mitochondrial genome sequences support ancient population expansion in *Plasmodium vivax*. *Mol. Biol. Evol.* 22, 1733-1739.
- Kaiser, M., Lova, A., Ulrich, M., Ellerbrok, H., Goffe, A.S., Blasse, A., Zommers, Z., Couacy-Hymann, E., Babweteera, F., Zuberbuhler, K., Metzger, S., Geidel, S., Boesch, C., Gillespie, T.R., and Leendertz, F.H. 2010. Wild chimpanzees infected with 5 *Plasmodium* species. *Emerg. Infect. Dis.* 16, 1956-1959.
- Krief, S., Escalante, A.A., Pacheco, M.A., Mugisha, L., Andre, C., Halbwax, M., Fischer, A., Krief, J.M., Kasenene, J.M., Crandfield, M., Cornejo, O.E., Chavatte, J.M., Lin, C., Letourneur, F., Gruner, A.C., McCutchan, T.F., Renia, L., and Snounou, G. 2010. On the diversity of malaria parasites in African apes and the origin of *Plasmodium falciparum* from bonobos. *PLoS Pathog.* 6, e1000765.
- Liu, W., Li, Y., Learn, G.H., Rudicell, R.S., Robertson, J.D., Keele, B.F., Ndjanga, J.B., Sanz, C.M., Morgan, D.B., Locatelli, S., Gonder, M.K., Kranzusch, P.J., Walsh, P.D., Delaporte, E., Mpoudi-Ngole, E., Georgiev, A.V., Muller, M.N., Shaw, G.M., Peeters, M., Sharp, P.M., Rayner, J.C., and Hahn, B.H. 2010a. Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. *Nature* 467, 420-425.
- Liu, W., Li, Y., Peeters, M., Rayner, J.C., Sharp, P.M., Shaw, G.M., and Hahn, B.H. 2010b. Single genome amplification and direct amplicon sequencing of *Plasmodium* spp. DNA from ape fecal specimens. *Protocol Exchange*. doi:10.1038/nprot.2010.156.
- Liu, W., Li, Y., Shaw, K.S., Learn, G.H., Plenderleith, L.J., Malenke, J.A., Sundararaman, S.A., Ramirez, M.A., Crystal, P.A., Smith, A.G., Bibollet-Ruche, F., Ayoub, A., Locatelli, S., Esteban, A., Mouacha, F., Guichet, E., Butel, C., Ahuka-Mundeki, S., Inogwabini, B.I., Ndjanga, J.B., Speede, S., Sanz, C.M., Morgan, D.B., Gonder, M.K., Kranzusch, P.J., Walsh, P.D., Georgiev, A.V., Muller, M.N., Piel,

- A.K., Stewart, F.A., Wilson, M.L., Pusey, A.E., Cui, L., Wang, Z., Farnert, A., Sutherland, C.J., Nolder, D., Hart, J.A., Hart, T.B., Bertolani, P., Gillis, A., LeBreton, M., Tafon, B., Kiyang, J., Djoko, C.F., Schneider, B.S., Wolfe, N.D., Mpoudi-Ngole, E., Delaporte, E., Carter, R., Culleton, R.L., Shaw, G.M., Rayner, J.C., Peeters, M., Hahn, B.H., and Sharp, P.M. 2014. African origin of the malaria parasite *Plasmodium vivax*. Nat. Commun. 5, 3346.
- Liu, W., Sundararaman, S.A., Loy, D.E., Learn, G.H., Li, Y., Plenderleith, L.J., Ndjango, J.-B.N., Speede, S., Atencia, R., Cox, D., Shaw, G.M., Ayoub, A., Peeters, M., Rayner, J.C., Hahn, B.H., and Sharp, P.M. submitted. Multigenomic delineation of *Plasmodium* species of *Laverania* subgenus infecting wild-living chimpanzees and gorillas.
- Liu, W., Sundararaman, S.A., Loy, D.E., Learn, G.H., Li, Y., Plenderleith, L.J., Ndjango, J.B., Speede, S., Rayner, J.C., Peeters, M., Hahn, B.H., and Sharp, P.M. (2015). On the origins of allelic dimorphism of the *Plasmodium falciparum* msp-1 and msp-6 genes. Paper presented at: Am. Soc. Trop. Med. Hyg. (ASTMH), 64th Annual Meeting (Philadelphia, PA).
- Livingstone, F.B. 1958. Anthropological implications of sickle cell gene distribution in west Africa Am. Anthropol. 60, 533-562.
- Livingstone, F.B. 1984. The Duffy blood groups, vivax malaria, and malaria selection in human populations: a review. Hum. Biol. 56, 413-425.
- Lynch, M. 2010. Evolution of the mutation rate. Trends Genet. 26, 345-352.
- Makanga, B., Yangari, P., Rahola, N., Rougeron, V., Elguero, E., Boundenga, L., Moukodoum, N.D., Okouga, A.P., Arnathau, C., Durand, P., Willaume, E., Ayala, D., Fontenille, D., Ayala, F.J., Renaud, F., Ollomo, B., Prugnolle, F., and Paupy, C. 2016. Ape malaria transmission and potential for ape-to-human transfers in Africa. Proc. Natl. Acad. Sci. U.S.A., 201603008.
- Mellars, P. 2006. Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. Science 313, 796-800.
- Miller, L.H., Mason, S.J., Clyde, D.F., and McGinniss, M.H. 1976. The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, *FyFy*. N. Engl. J. Med. 295, 302-304.
- Molina-Cruz, A., and Barillas-Mury, C. 2016. Mosquito vectors of ape malarias: another piece of the puzzle. Proc. Natl. Acad. Sci. U.S.A. In press.
- Mu, J., Joy, D.A., Duan, J., Huang, Y., Carlton, J., Walker, J., Barnwell, J., Beerli, P., Charleston, M.A., Pybus, O.G., and Su, X.Z. 2005. Host switch leads to emergence of *Plasmodium vivax* malaria in humans. Mol. Biol. Evol. 22, 1686-1693.
- Nair, S., Williams, J.T., Brockman, A., Paiphun, L., Mayxay, M., Newton, P.N., Guthmann, J.P., Smithuis, F.M., Hien, T.T., White, N.J., Nosten, F., and

- Anderson, T.J. 2003. A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. *Mol. Biol. Evol.* 20, 1526-1536.
- Neafsey, D.E., Galinsky, K., Jiang, R.H., Young, L., Sykes, S.M., Saif, S., Gujja, S., Goldberg, J.M., Young, S., Zeng, Q., Chapman, S.B., Dash, A.P., Anvikar, A.R., Sutton, P.L., Birren, B.W., Escalante, A.A., Barnwell, J.W., and Carlton, J.M. 2012. The malaria parasite *Plasmodium vivax* exhibits greater genetic diversity than *Plasmodium falciparum*. *Nat. Genet.* 44, 1046-1050.
- Ollomo, B., Durand, P., Prugnolle, F., Douzery, E., Arnathau, C., Nkoghe, D., Leroy, E., and Renaud, F. 2009. A new malaria agent in African hominids. *PLoS Pathog.* 5, e1000446.
- Otto, T.D., Rayner, J.C., Bohme, U., Pain, A., Spottiswoode, N., Sanders, M., Quail, M., Ollomo, B., Renaud, F., Thomas, A.W., Prugnolle, F., Conway, D.J., Newbold, C., and Berriman, M. 2014. Genome sequencing of chimpanzee malaria parasites reveals possible pathways of adaptation to human hosts. *Nat. Commun.* 5, 4754.
- Pacheco, M.A., Battistuzzi, F.U., Junge, R.E., Cornejo, O.E., Williams, C.V., Landau, I., Rabetafika, L., Snounou, G., Jones-Engel, L., and Escalante, A.A. 2011. Timing the origin of human malarias: the lemur puzzle. *BMC Evol. Biol.* 11, 299.
- Pacheco, M.A., Cranfield, M., Cameron, K., and Escalante, A.A. 2013. Malarial parasite diversity in chimpanzees: the value of comparative approaches to ascertain the evolution of *Plasmodium falciparum* antigens. *Malar. J.* 12, 238.
- Paget-McNicol, S., and Saul, A. 2001. Mutation rates in the dihydrofolate reductase gene of *Plasmodium falciparum*. *Parasitology* 122, 497-505.
- Paupy, C., Makanga, B., Ollomo, B., Rahola, N., Durand, P., Magnus, J., Willaume, E., Renaud, F., Fontenille, D., and Prugnolle, F. 2013. *Anopheles moucheti* and *Anopheles vinckei* are candidate vectors of ape *Plasmodium* parasites, including *Plasmodium praefalciparum* in Gabon. *PLoS One* 8, e57294.
- Prugnolle, F., Durand, P., Neel, C., Ollomo, B., Ayala, F.J., Arnathau, C., Etienne, L., Mpoudi-Ngole, E., Nkoghe, D., Leroy, E., Delaporte, E., Peeters, M., and Renaud, F. 2010. African great apes are natural hosts of multiple related malaria species, including *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 1458-1463.
- Prugnolle, F., Ollomo, B., Durand, P., Yalcindag, E., Arnathau, C., Elguero, E., Berry, A., Pourrut, X., Gonzalez, J.P., Nkoghe, D., Akiana, J., Verrier, D., Leroy, E., Ayala, F.J., and Renaud, F. 2011. African monkeys are infected by *Plasmodium falciparum* nonhuman primate-specific strains. *Proc. Natl. Acad. Sci. U.S.A.* 108, 11948-11953.
- Prugnolle, F., Rougeron, V., Becquart, P., Berry, A., Makanga, B., Rahola, N., Arnathau, C., Ngoubangoye, B., Menard, S., Willaume, E., Ayala, F.J., Fontenille, D., Ollomo, B., Durand, P., Paupy, C., and Renaud, F. 2013. Diversity, host

- switching and evolution of *Plasmodium vivax* infecting African great apes. Proc. Natl. Acad. Sci. U.S.A. 110, 8123-8128.
- Ramachandran, S., Deshpande, O., Roseman, C.C., Rosenberg, N.A., Feldman, M.W., and Cavalli-Sforza, L.L. 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc. Natl. Acad. Sci. U.S.A. 102, 15942-15947.
- Reichenow, E. 1920. Über das Vorkommen der Malariaparasiten des Menschen bei den Afrikanischen Menschenaffen. Centralbl. f. Bakt. I. Abt. Orig. 85, 207-216.
- Rich, S.M., Leendertz, F.H., Xu, G., LeBreton, M., Djoko, C.F., Aminake, M.N., Takang, E.E., Dikko, J.L.D., Pike, B.L., Rosenthal, B.M., Formenty, P., Boesch, C., Ayala, F.J., and Wolfe, F.D. 2009. The origin of malignant malaria. Proc. Natl. Acad. Sci. U.S.A. 106, 14902-14907.
- Rich, S.M., Licht, M.C., Hudson, R.R., and Ayala, F.J. 1998. Malaria's Eve: evidence of a recent population bottleneck throughout the world populations of *Plasmodium falciparum*. Proc. Natl. Acad. Sci. U.S.A. 95, 4425-4430.
- Roy, S.W. 2015. The *Plasmodium gaboni* genome illuminates allelic dimorphism of immunologically important surface antigens in *P. falciparum*. Infect. Genet. Evol. 36, 441-449.
- Roy, S.W., and Ferreira, M.U. 2015. A new model for the origins of allelic dimorphism in *Plasmodium falciparum*. Parasitol. Int. 64, 229-237.
- Roy, S.W., Ferreira, M.U., and Hartl, D.L. 2008. Evolution of allelic dimorphism in malarial surface antigens. Heredity (Edinb) 100, 103-110.
- Sharp, P.M., Liu, W., Learn, G.H., Rayner, J.C., Peeters, M., and Hahn, B.H. 2011. Source of the human malaria parasite *Plasmodium falciparum*. Proc. Natl. Acad. Sci. U.S.A. 108, E744-745.
- Sluiter, C.P., Swellengrebel, N.H., and Ihle, J.E. 1922. De dierlijke parasieten van den mensch en van onze huisdieren. 3rd ed. Scheltema and Holkema's Boekhandel, Amsterdam.
- Sundararaman, S.A., Liu, W., Keele, B.F., Learn, G.H., Bittinger, K., Mouacha, F., Ahuka-Mundeke, S., Manske, M., Sherrill-Mix, S., Li, Y., Malenke, J.A., Delaporte, E., Laurent, C., Mpoudi Ngole, E., Kwiatkowski, D.P., Shaw, G.M., Rayner, J.C., Peeters, M., Sharp, P.M., Bushman, F.D., and Hahn, B.H. 2013. *Plasmodium falciparum*-like parasites infecting wild apes in southern Cameroon do not represent a recurrent source of human malaria. Proc. Natl. Acad. Sci. U.S.A. 110, 7020-7025.
- Sundararaman, S.A., Plenderleith, L.J., Liu, W., Loy, D.E., Learn, G.H., Li, Y., Shaw, K.S., Ayoub, A., Peeters, M., Speede, S., Shaw, G.M., Bushman, F.D., Brisson, D., Rayner, J.C., Sharp, P.M., and Hahn, B.H. 2016. Genomes of cryptic chimpanzee *Plasmodium* species reveal key evolutionary events leading to human malaria. Nat. Commun. 7, 11078.

- Tachibana, S., Sullivan, S.A., Kawai, S., Nakamura, S., Kim, H.R., Goto, N., Arisue, N., Palacpac, N.M., Honma, H., Yagi, M., Tougan, T., Katakai, Y., Kaneko, O., Mita, T., Kita, K., Yasutomi, Y., Sutton, P.L., Shakhbatyan, R., Horii, T., Yasunaga, T., Barnwell, J.W., Escalante, A.A., Carlton, J.M., and Tanabe, K. 2012. *Plasmodium cynomolgi* genome sequences provide insight into *Plasmodium vivax* and the monkey malaria clade. *Nat. Genet.* 44, 1051-1055.
- Tanabe, K., Mackay, M., Goman, M., and Scaife, J.G. 1987. Allelic dimorphism in a surface antigen gene of the malaria parasite *Plasmodium falciparum*. *J. Mol. Biol.* 195, 273-287.
- Tanabe, K., Mita, T., Jombart, T., Eriksson, A., Horibe, S., Palacpac, N., Ranford-Cartwright, L., Sawai, H., Sakihama, N., Ohmae, H., Nakamura, M., Ferreira, M.U., Escalante, A.A., Prugnolle, F., Bjorkman, A., Farnert, A., Kaneko, A., Horii, T., Manica, A., Kishino, H., and Balloux, F. 2010. *Plasmodium falciparum* accompanied the human expansion out of Africa. *Curr. Biol.* 20, 1283-1289.
- Taylor, S.M., Parobek, C.M., and Fairhurst, R.M. 2012. Haemoglobinopathies and the clinical epidemiology of malaria: a systematic review and meta-analysis. *Lancet Infect. Dis.* 12, 457-468.
- Valkiunas, G., Ashford, R.W., Bensch, S., Killick-Kendrick, R., and Perkins, S. 2011. A cautionary note concerning *Plasmodium* in apes. *Trends Parasitol.* 27, 231-232.
- Volkman, S.K., Sabeti, P.C., DeCaprio, D., Neafsey, D.E., Schaffner, S.F., Milner, D.A., Jr., Daily, J.P., Sarr, O., Ndiaye, D., Ndir, O., Mboup, S., Duraisingh, M.T., Lukens, A., Derr, A., Stange-Thomann, N., Waggoner, S., Onofrio, R., Ziaugra, L., Mauceli, E., Gnerre, S., Jaffe, D.B., Zainoun, J., Wiegand, R.C., Birren, B.W., Hartl, D.L., Galagan, J.E., Lander, E.S., and Wirth, D.F. 2007. A genome-wide map of diversity in *Plasmodium falciparum*. *Nat. Genet.* 39, 113-119.
- World Health Organization, 2015. World Malaria Report 2015.
http://apps.who.int/iris/bitstream/10665/200018/1/9789241565158_eng.pdf?ua=1

CHAPTER 3 - Selective whole genome amplification: a robust method that enables scalable whole genome sequencing of *Plasmodium vivax* from unprocessed clinical samples

Annie N. Cowell^{*1}, Dorothy E. Loy^{*2}, Sesh A. Sundararaman², Hugo Valdivia^{3, 6}, Kathleen Fisch⁴, Andres G. Lescano^{5,6}, G. Christian Baldeviano⁵, Salomon Durand⁶, Vince Gerbasi⁶, Colin J. Sutherland⁷, Debbie Nolder⁷, Joseph M. Vinetz¹, Beatrice H. Hahn², Elizabeth A. Winzeler⁸

¹*Division of Infectious Diseases, Department of Medicine, University of California, San Diego, CA,* ²*Departments of Medicine and Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA,* ³*Universidade Federal de Minas Gerais, Brazil,* ⁴*Center for Computational Biology & Bioinformatics, UC San Diego, CA,* ⁵*Universidad Peruana Cayetano Heredia, Lima, Peru,* ⁶*U.S. Naval Medical Research No. 6, Callao, Peru,* ⁷*Public Health England Malaria Reference Laboratory, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, UK,* ⁸*Division of Host-Microbe Systems & Therapeutics, Department of Pediatrics, UC San Diego, CA*

*Annie N. Cowell and Dorothy E. Loy contributed equally to this work.

Originally published in 2017 in *MBio* 8(1):e02257

3.1 Abstract

Whole genome sequencing (WGS) of microbial pathogens from clinical samples is a highly sensitive tool used to gain a deeper understanding of the biology, epidemiology, and drug resistance mechanisms of many infections. However, WGS of organisms which exhibit low densities in their host is challenging due to high levels host genomic DNA (gDNA), which leads to very low coverage of the microbial genome. WGS of *Plasmodium vivax*, the most widely distributed form of malaria, is especially difficult because of low parasite densities and the lack of an *ex-vivo* culture system. Current techniques used to enrich *P. vivax* DNA from clinical samples require significant resources or are not consistently effective. Here, we demonstrate that selective whole genome amplification (SWGA) can enrich *P. vivax* gDNA from unprocessed human blood samples and dried blood spots for high quality WGS, allowing genetic characterization of isolates that would have otherwise been prohibitively expensive or impossible to sequence. We achieved an average genome coverage of 24x with up to 95% of the *P. vivax* core genome covered by ≥ 5 reads. Single nucleotide polymorphism (SNP) characteristics and drug resistance mutations seen were consistent with other *P. vivax* sequences from a similar region in Peru, demonstrating that SWGA produces high quality sequences for downstream analysis. SWGA is a robust tool that will enable efficient, cost-effective WGS of *P. vivax* isolates from clinical samples that can be applied to other neglected microbial pathogens.

3.2 Introduction

Malaria is a mosquito-borne infection caused by protozoan parasites of the *Plasmodium* genus. Of the six *Plasmodium* species known to infect humans (Calderaro et al., 2013; Singh et al., 2004; Sutherland et al., 2010), *P. vivax* is the most widely distributed, causing approximately half of all clinical cases of malaria outside of Africa (World Health

Organisation, 2017). Whole genome sequencing (WGS) of *Plasmodium* parasites from clinical samples has revealed important insights into the biology, epidemiology, and mechanisms of drug resistance of malaria (Borrmann et al., 2013; Chan et al., 2012; Flannery et al., 2016; Hester et al., 2013; Manske et al., 2012; Miotto et al., 2013; Pearson et al., 2016; Winter et al., 2015). For *P. vivax*, WGS of clinical isolates has the potential to uncover mechanisms underlying some of the unique aspects of this parasite's biology, such as distinguishing between reinfection and relapse due to activation of dormant liver parasites. However, WGS of *P. vivax* from clinical samples is challenging, mainly due to low parasite densities in clinical samples compared to *P. falciparum* and the lack of a robust *ex-vivo* culture system.

Multiple techniques have been developed to enrich *Plasmodium* genomic DNA (gDNA) from clinical samples, including leukocyte depletion (Auburn et al., 2012; Pearson et al., 2016; Venkatesan et al., 2012), hybrid selection with RNA baits (Bright et al., 2012; Hupalo et al., 2016) short-term *ex-vivo* culture (Auburn et al., 2013), adaptation to growth in splenectomized monkeys (Carlton et al., 2008; Neafsey et al., 2012) and single cell sequencing (Nair et al., 2014). The majority of these techniques require significant labor and resources. While leukocyte depletion is the most cost-effective, it requires sample processing within 6 hours of sample collection, which is not feasible at many field sites, and is not always effective. Two recent studies that performed WGS on over 400 clinical isolates of *P. vivax* (Hupalo et al., 2016; Pearson et al., 2016) employed hybrid selection and leukocyte depletion to enrich *P. vivax* gDNA from clinical samples. Pearson et al. (2016) used leukocyte depletion on 292 clinical samples and had to eliminate 144 (49%) of their samples from further population genetics analysis due to low quality, which often occurs due to contaminating human DNA. Hupalo et al. (2016) used hybrid selection to enrich their samples, with 31 out of 170 sequences (21%) removed from further analysis

due to low quality. Although more frequently successful, the hybrid selection technique requires either expensive synthetic RNA baits or a large amount of pure *P. vivax* DNA to create the RNA baits, which is difficult to obtain. In addition, hybrid selection can introduce bias, since is approximately half as efficient at capturing regions with GC content >50% (Bright et al., 2012).

An alternative method is selective whole genome amplification (SWGA). SWGA has been used to enrich DNA of the ape *Plasmodium* parasites *P. reichenowi* and *P. gaboni*, from whole blood samples with submicroscopic DNA levels (Larremore et al., 2015; Sundararaman et al., 2016) as well as *P. falciparum* genomes from dried blood spots (Guggisberg et al., 2016) for WGS. SWGA preferentially amplifies the genomes of pathogens from complex mixtures of target and host DNA (Leichty and Brisson, 2014) (Figure 3-1). SWGA does not require separation of target DNA from background DNA, making it an attractive option for pathogens that cannot be amplified in culture. DNA amplification is carried out by the highly processive, strand displacing phi29 DNA polymerase and a set of pathogen specific primers that target short (6-12 nucleotide) motifs that are common in the pathogen genome and uncommon in the host genome. The strand displacement function of phi29 results in the amplification of genomic regions where primers bind frequently, leading to the preferential amplification of genomes with frequent primer-binding sites. Here, we show that SWGA efficiently enriches *P. vivax* gDNA from unprocessed human blood samples and dried blood spots for cost-effective, high-quality whole genome sequencing.

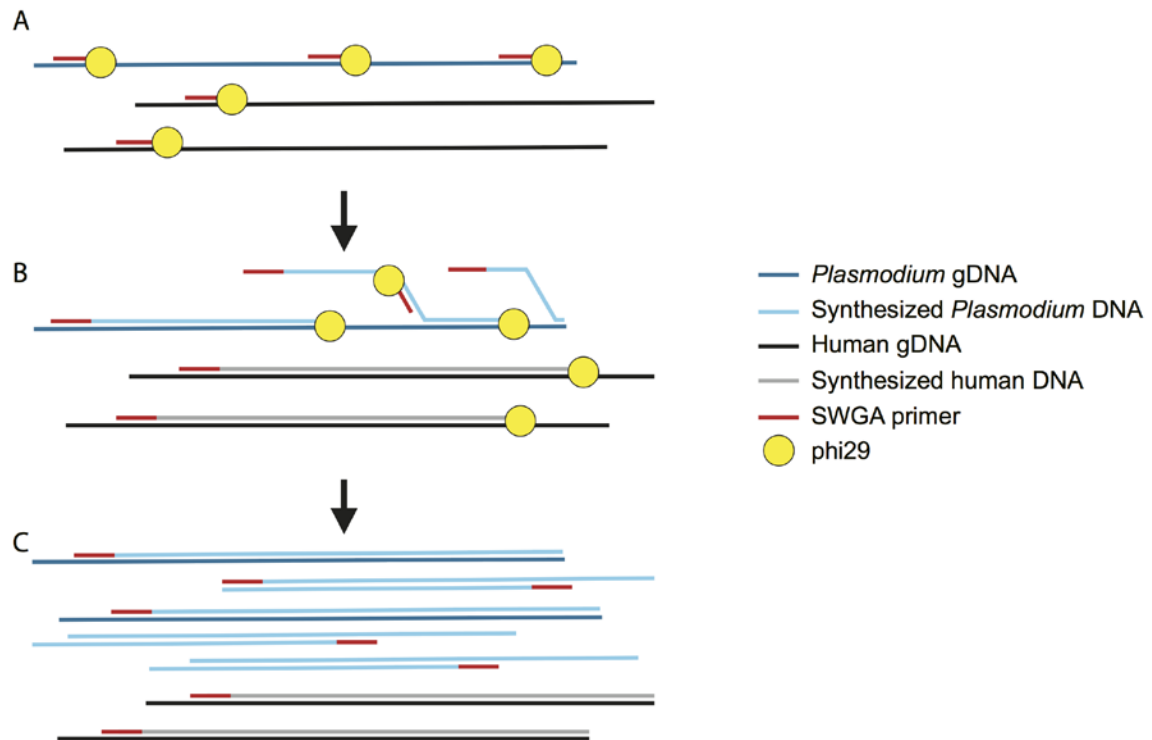


Figure 3-1. Selective whole genome amplification (SWGA) of *Plasmodium vivax* genomic DNA (gDNA) from human blood samples. (A) SWGA primers bind frequently to *Plasmodium vivax* gDNA and infrequently to human gDNA. (B) When phi29 encounters double stranded gDNA, it displaces the newly synthesized strand, opening new primer binding sites on the synthesized gDNA, leading to selective amplification of templates with frequent primer binding sites. (C) Post-SWGA, the percentage of *P. vivax* DNA has increased relative to host DNA.

3.3 Results

Primer design and optimization using *P. vivax* infected whole blood samples. To

perform SWGA on *P. vivax* gDNA from unprocessed human blood samples, we designed primers that specifically amplified this parasite's DNA using a previously published approach for *P. falciparum* (Sundararaman et al., 2016). Briefly, we identified the most frequently occurring motifs of 6-12 nucleotides in length in the *P. vivax* Salvador-1 (Sal-1) reference genome. We selected the top 10,000 primers of each length, yielding a total of 70,000 primers for further analysis. We filtered these primers based on characteristics

such as melting temperature (18–32°C), ability to homodimerize (no greater than 3 consecutive matches), binding frequency on the human genome and Sal-1 genome (less frequent than once every 500,000 bp and more frequent than once every 50,000 bp, respectively), and infrequent binding of the human mitochondrial genome (less than 4 binding sites). Next, we removed primers predicted to bind the Sal-1 subtelomeres. These filters resulted in a pool of 222 primers. We separated these primers into 6 sets that were predicted not to form heterodimers, and identified the top set (pvset1) of 10 primers using a selection algorithm described previously (Sundararaman et al., 2016). gDNA from an unprocessed *P. vivax* infected whole blood sample (MRL2) was subjected to SWGA with primer set pvset1 prior to shotgun sequencing (Figure S3-1). SWGA significantly increased the percent of reads that mapped to the *P. vivax* Sal-1 reference genome from 0.7% to 73.5% (Figure S3-1A) and improved the genome coverage obtained from ~80 million base pairs of sequencing from 1.5% to 58% (Figure S3-1B).

We observed that genome coverage obtained per base pair sequenced was lower than that achieved with SWGA of *P. falciparum* gDNA using the same primer set design methods (Sundararaman et al., 2016). Visual inspection of the *P. vivax* genome coverage from samples subjected to SWGA revealed that coverage gaps were typically in regions with comparatively higher GC content (Figure 3-2A). The *P. falciparum* genome is extremely AT-rich with only 19.4% of bases consisting of Gs or Cs, while the GC content for the *P. vivax* genome is 42.3% (Carlton et al., 2008). The *P. vivax* genome also has an isochores structure: internal chromosomal areas have a high GC-content and subtelomeres and centromeres have a lower GC-content (McCutchan et al., 1984). Since phi29 DNA polymerase pauses more frequently during strand displacement and primer extension in regions with high GC content of DNA (Morin et al., 2012), we hypothesized that differences

in base composition could explain the more uneven amplification of the *P. vivax* genome compared to that of *P. falciparum*.

Figure 3-2. *Plasmodium vivax* chromosomal coverage following SWGA using primer set pvset1 (A) or pvset1920 (B). The base compositions of chromosomes 2 and 6 were visualized in Geneious (version 9.1) using the *P. vivax* Sal-1 reference genome; green and blue lines represent percent AT and GC content, respectively, plotted for 25 bp windows across the chromosome (scale shown above the graph). Shown in blue and red below is the corresponding MiSeq read coverage depth using primer sets pvset1 and pvset1920, respectively. Coverage plots were generated using IGVTools (version 2.3.40) and are shown on a log scale with maximum read depth indicated in the upper left corner of the plot.

We thus designed primer sets specifically targeting regions of the *P. vivax* Sal-1 reference genome with high GC content and poor coverage using the *swga* program (Clarke et al., 2017), a program that identifies and scores SWGA primer sets (Figure S3-2). Primers were designed to bind regions of the *P. vivax* Sal-1 genome that had even AT/GC composition, were longer than 195,000 bp, and had low sequence coverage when amplified with pvset1. We identified 1,939 primer sets (consisting of up to 15 primers) with minimal human genome binding and maximal *P. vivax* genome binding and scored them based on evenness of binding as well as mean distance between primer binding sites in

the foreground and background genomes. The primer set with the best score, pvset1920, was chosen for subsequent testing. SWGA of an unprocessed human blood sample with pvset1920 yielded an overall superior *P. vivax* genome coverage compared to SWGA with pvset1 (Figure 3-3). Visual inspection of these post-SWGA *P. vivax* sequences revealed that pvset1920 achieved improved coverage particularly in regions with high GC content (Figure 3-2B), with troughs in coverage in genomic regions of lower GC content, which include the centromeres and subtelomeres.

Figure 3-3. Testing of SWGA primer sets on DNA from an unprocessed, *P. vivax*-infected blood sample. (A) Unamplified DNA (black) and DNA amplified with SWGA primer set pvset1 (blue) or pvset1920 (red) was sequenced on a MiSeq (Illumina). The percent of MiSeq reads that mapped to the *P. vivax* Sal-1 reference genome in Geneious (39) (version 9.1) was plotted for both unamplified and SWGA-amplified samples. (B) The 1x *P. vivax* genome coverage is shown relative to the total sequencing depth (in millions of base pairs sequenced) for samples subjected to SWGA with pvset1920 or pvset1, and for unamplified DNA.

Having developed a method that worked well for SWGA of *P. vivax* gDNA from whole blood, we tested whether the method could also be applied to gDNA extracted from dried blood spot samples. Dried blood spots are a common method of storing patient and parasite DNA that utilizes a smaller volume of blood and does not require immediate cold storage. DNA extracted from dried blood spots can have variable quality depending on the method of collection and storage (Schwartz et al., 2015). SWGA has been used to enrich *P. falciparum* DNA from dried blood spots for WGS (Guggisberg et al., 2016), with an average of 48.1% $\pm$ 3.5% of the genome covered at $\geq 5x$ for samples with an average parasite density of 73,601 parasites/ μ l $\pm$ 19,399 (1.5% parasitemia). Since *P. vivax* clinical samples generally have lower parasite densities, we wondered if it would be feasible to obtain significant genome coverage on *P. vivax* from dried blood spots with SWGA. We extracted DNA from blood spots obtained from symptomatic patients in Peru and performed SWGA with pvset1920, achieving 73% and 42% of genome coverage at 1x on initial testing for samples with high (Sample C, 56,790 parasites/ μ l; 1.1%) and low (Sample L, 2572 parasites/ μ l; 0.05%) parasitemia, respectively (Figure S3-3).

We finally tested whether an enzymatic digest to remove contaminating human DNA could further improve *P. vivax* genome coverage. Modification-dependent restriction endonucleases (MDREs), such as MspJI and FspEI, which specifically recognize cytosine C5 methylation or hydroxymethylation (Cohen-Karni et al., 2011), have been used to selectively degrade human DNA in *P. falciparum* clinical samples. Enzyme digest of DNA extracted from clinical samples with $>80\%$ human contamination has previously been shown to enrich *P. falciparum* DNA ~ 9 -fold for more efficient WGS (Oyola et al., 2013). However, when we performed a digest with MspJI and FspEI enzymes on gDNA extracted from whole blood obtained from patients with *P. vivax* infection, we observed either no

change, or markedly decreased genome coverage in the 5 enzyme-digested samples (Figure S3-4).

SWGA and WGS of *P. vivax* from patient samples. To test the utility of SWGA for variant calling and population genetics analysis of *P. vivax* from unprocessed clinical blood samples, we used primer set pvset1920 to perform SWGA on *P. vivax* gDNA from 18 whole blood and 4 dried blood spot samples collected from symptomatic patients with *P. vivax* infection in Peru. Since the whole blood samples had not been leukocyte-filtered, they had significant contamination with human DNA, with less than 1.5% of reads mapping to the Sal-1 reference genome in unamplified samples (not shown). For all samples, SWGA significantly increased the proportion of reads that mapped to the *P. vivax* Sal-1 reference genome, resulting in a higher genome coverage and a higher percent of callable total and core genome regions (covered by ≥ 5 reads) (Table 3-1). Comparison of the SWGA-amplified samples to 10 leukocyte-filtered samples from a field study in Peru which were sequenced to a similar depth (1.5 billion bp $\pm$ 0.2 sequenced for SWGA samples vs. 1.5 $\pm$ 0.5 billion bp for leukocyte-filtered samples) showed that SWGA yields a 2-fold increase in the percent of sequencing reads that map to the *P. vivax* genome and an average 5x *P. vivax* core genome coverage of 60.1% $\pm$ 26.0 compared to 43.7% $\pm$ 41.4 for leukocyte-filtered samples (Flannery et al., 2016). For the 4 dried blood spot samples, we achieved an average 5x core genome coverage of 54.0 $\pm$ 34.6%.

There was a trend towards improved mean coverage and percent of the genome callable in samples with higher parasite densities (Figure S3-5). This is consistent with previous SWGA results for *P. falciparum* (Sundararaman et al., 2016), and results from other *P. vivax* enriching methods, such as hybrid selection (Bright et al., 2012). Samples

subjected to SWGA yielded similar $\geq 5\times$ genome coverage per sequenced base pair when compared to direct sequencing of a leukocyte-filtered patient sample (Figure 3-4A). The percent of the 14 large chromosomes of *P. vivax* considered callable for samples that underwent SWGA fell within the range of that obtained by direct sequencing of leukocyte-filtered samples (Figure 3-4B). Additionally, post-SWGA sequences yielded similar mean base quality when normalized across 100 base pair windows of varying percent GC-content in the reference genome compared to leukocyte-filtered samples (Figure 3-5).

Table 3-1. Sequencing statistics for *P. vivax* sequences from clinical samples that underwent selective whole genome amplification (SWGA).

18 blood samples (SWGA)						
Sample	Parasite density^a	Total bp sequenced (billions)	Percent aligned reads^b	Mean coverage (x)^b	Percent genome callable^{b,c}	Percent core genome callable^{b,d}
1	45680 (0.9%)	1.42	89.5	37.1	71.7	83.6
2	34268 (0.7%)	1.68	80.9	37.9	72.9	85.1
3	28474 (0.6%)	1.41	88.3	30.4	66.6	76.3
4	25680 (0.5%)	1.50	65.1	13.8	40.2	42.1
5	19241 (0.4%)	2.00	71.9	44.1	79.9	95.4
6	13064 (0.3%)	1.84	83.7	43.8	76.1	90.4
7	11438 (0.2%)	1.52	79.5	20.3	32.7	34.5
8	9961 (0.2%)	1.36	75.4	31.0	74.6	87.3
9	8842 (0.2%)	1.84	74.2	29.6	72.9	84.7
10	7382 (0.1%)	1.23	72.6	23.7	65.6	73.6
11	6258 (0.1%)	1.41	82.8	25.1	59.5	67.7
12	5135 (0.1%)	1.82	54.9	18.0	34.8	37.0
13	2942 (0.06%)	1.16	40.7	12.9	52.1	57.9
14	1873 (0.04%)	1.44	17.0	5.7	13.7	14.1
15	1652 (0.03%)	1.52	53.2	23.5	52.0	57.3
16	1471 (0.03%)	1.26	44.0	12.9	21.5	22.9
17	537 (0.01%)	1.92	22.5	11.1	38.1	40.7
18	495 (0.01%)	1.44	28.6	10.0	30.0	30.5
Average^e	12466 +/- 13107.2	1.5 +/- 0.2	62.7 +/- 23.2	23.9 +/- 11.8	53.1 +/- 20.9	60.1 +/- 26.0
4 dried blood spot samples (SWGA)						
DBS-4	50330 (1.0%)	1.97	79.4	34.8	74.4	87.4
DBS-3	35730 (0.7%)	1.91	46.8	20.9	69.5	80.3
DBS-2	5932 (0.1%)	0.93	11.0	3.0	25.8	24.2
DBS-1	3885 (0.08%)	1.57	17.7	4.2	23.2	24.1
Average^e	23962 +/- 22826.4	1.6 +/- 0.5	38.7 +/- 27.1	15.7 +/- 13.1	48.2 +/- 27.5	54.0 +/- 34.6
10 blood samples (Leukocyte filtration)^f						
Mdio01	N/A	2.43	25.7	6.35	8.2	5.5
Mdio02	N/A	0.80	16.2	1.65	1.3	0.6
Mdio03	N/A	2.04	19.9	8.23	57.2	61.2
Mdio04	N/A	1.52	14.2	1.97	18.7	9.6
Mdio05	N/A	1.56	49	22.3	78.1	91.7
Mdio06	N/A	1.62	55.4	28.1	79.6	93.4
Mdio07	N/A	1.46	20.7	4.23	15.8	11.3
Mdio08	N/A	1.50	30.3	10.6	67.3	74.9
Mdio09	N/A	0.74	18.3	1.74	1.6	0.8
Mdio10	N/A	1.26	43.5	16.1	76	88.1
Average^e	N/A	1.5 +/- 0.5	31.4 +/- 14.8	10.1 +/- 9.2	40.4 +/- 34.0	43.7 +/- 41.4

^a Parasites per microliter (parasitemia) as determined by microscopy.

^b Sequencing statistics were determined using the Genome Analysis Toolkit's (GATK) DepthOfCoverage tool.

^c Covered by ≥ 5 reads

^d The core genome was defined by coordinates determined in the large scale *P. vivax* sequencing study by Pearson et al. (2016)

^e The standard deviation from the mean is shown.

^f Leukocyte-filtered sequencing statistics presented here were from *P. vivax* clinical samples obtained from a previously published study in Peru (Flannery et al, 2015).

Figure 3-4. Comparison of *P. vivax* genome coverage generated from samples treated with SWGA versus leukocyte filtration. (A) 5x *P. vivax* genome coverage is shown relative to total sequencing depth (in millions of base pairs sequenced) for a sample amplified with pvset1920 (Sample 5) and for a leukocyte-filtered sample (Mdio6). Both samples were sequenced on an Illumina HiSeq. (B) The percent of the 14 chromosomes callable (covered by ≥ 5 reads) for samples that underwent SWGA (colored lines) or leukocyte filtration (black dashed lines) was compared between multiple samples.

Figure 3-5. GC-bias plots for *P. vivax* genomes generated following leukocyte filtration (A) or SWGA (B).

Variant analysis. To examine the utility of post-SWGA sequences for variant analysis, we called 45,821 single nucleotide polymorphisms (SNPs) from the whole blood and dried blood spot samples that were subjected to SWGA. For the whole blood samples, an average of 14,463 SNPs was identified per sample, which is consistent with prior studies of *P. vivax* field isolates (Flannery et al., 2016). Compared to leukocyte-filtered samples, SNP characteristics such as SNP rate, transition to transversion (Ti/Tv) ratio, and nonsynonymous to synonymous ratio were near identical in the samples that underwent SWGA (Table 3-2).

Table 3-2. SNPs in SWGA versus leukocyte filtered whole genome sequences

Sample preparation method	SNP characteristics				SNP effects	
	Transition/transversion ratio	Exonic	Intronic	Intergenic	SNP per base	Nonsynonymous to Synonymous Ratio
SWGA ^a	1.41	20,865 (40%)	3,714 (7%)	25,178 (49%)	0.002	1.63
Leukocyte filtration	1.36	18,365 (40%)	3,029 (7%)	23,157 (50%)	0.002	1.74

^aSamples for SWGA were obtained from Iquitos, Peru and samples for leukocyte filtration were from a previously published study in Madre de dios, Peru (10).

In addition, the proportion of SNPs that were exonic, intronic, intergenic, or at 5' and 3' untranslated regions were similar between the two methods of *P. vivax* enrichment. We also detected SNPs in several known drug resistance genes previously detected in samples from Peru (Flannery et al., 2016) and Colombia (Winter et al., 2015) in the whole blood and dried blood spot samples (Table 3-3, Table S3-1), further validating the utility of sequences derived from SWGA for variant calling. This includes several intronic mutations around a putative chloroquine resistance transporter (*pvcrt*), in addition to coding mutations in dihydrofolate reductase (*pvdhfr*), multidrug resistance protein 1

(*pvm-dr1*), multidrug resistance protein 2 (*pvm-rp2*), and dihydropteroate synthetase (*dhps*).

Table 3-3. Nonsynonymous SNPs in known drug resistance genes

Locus	Chr	Position	Ref	Alt	Amino Acid	Samples ^a
<i>pvcrt-0</i> (PVX_087980)	1	331151	T	C	Intron	17 (18)
		331819	G	A	Intron	11 (18)
		332453	T	C	Intron	18 (18)
		332874	A	C	Intron	18 (18)
<i>pvdhfr</i> (PVX_089950)	5	964763	C	G,A	Ser58Arg	12 (12)
		964939	G	A	Ser117Asn	16 (16)
<i>pvm-dr1</i> (PVX_080100)	10	363223	A	G	Thr958Met	12 (12)
		363374	T	G	Met908Leu	11 (11)
		365435	C	A	Val221Leu	1 (12)
<i>pvm-rp2</i> (PVX_124085)	14	2043859	G	C	Gln1407Glu	11 (13)
		2045050	C	T	Val1010Met	14 (15)
		2047090	G	A	Pro330Ser	1(15)
		2047233	C	A	Arg282Met	13 (14)
		2047816	C	G	Glu88Gln	3 (18)
<i>dhps</i> (PVX_123230)	14	1257856	G	C	Ala383Gly	9 (15)
		1258389	C	T	Met205Ile	10 (12)

^aSamples shows the number of samples confidently genotyped (in parentheses) and the number bearing the indicated allele.

We also compared sample clonality estimates of post-SWGA sequences to microsatellite analyses on the same unamplified samples. We estimated the clonality of the 6 post-SWGA sequences with the highest coverage using the F_{ws} statistic, a measure of within-host diversity previously used to characterize multiplicity of infection in *Plasmodium falciparum* patient samples (Auburn et al., 2012; Manske et al., 2012) (Table 3-4). An F_{ws} score of ≥ 0.95 indicates low within-host diversity and infection with a single parasite, while $F_{ws} \leq 0.70$ is suggestive of a multiclonal infection. Microsatellite analysis on these same 6 unamplified samples indicated that all were clonal, except for sample 9, where the

presence of 2 microsatellite markers at more than one position suggested that it could be a multiclonal sample. However, for all 6 post-SWGA sequences, $F_{ws} \geq 0.95$ suggesting that all were clonal infections. Thus, while SWGA does not introduce errors that lead to a falsely low F_{ws} , it may lead to underestimations of clonality in multiclonal samples.

Table 3-4. Clonality estimates post-SWGA

Sample	F_{ws}	Heterozygous SNPs	<u>Microsatellite markers before SWGA</u>					
			v23	v24	v25	v28	v30	v33
1	0.98	377	266	170	102	85	149	153
2	0.97	446	266	170	102	85	149	163
5	0.97	372	266	171	102	86	150	153
6	0.98	326	266	170	102	86	149	153
8	0.98	300	282	167	102	94	136	134
9	0.97	354	282	163	99,94	86,118	143,150	151

Finally, we constructed a neighbor-joining tree using core genome SNPs to determine the relatedness of our samples to one another and to other *P. vivax* isolates from Peru and around the world (Figure 3-6). In this tree, our samples, which were from the Iquitos region of Peru, clustered with one another, and were most closely related to leukocyte-filtered samples from another region of Peru (Flannery et al., 2016). They also exhibited the expected degree of relatedness to previously published *P. vivax* sequences derived from other leukocyte-filtered (Pearson et al., 2016), hybrid-selected (Hupalo et al., 2016), and monkey-adapted (Neafsey et al., 2012) clinical samples from diverse areas of the world, further validating the use of post-SWGA sequences from downstream analyses.

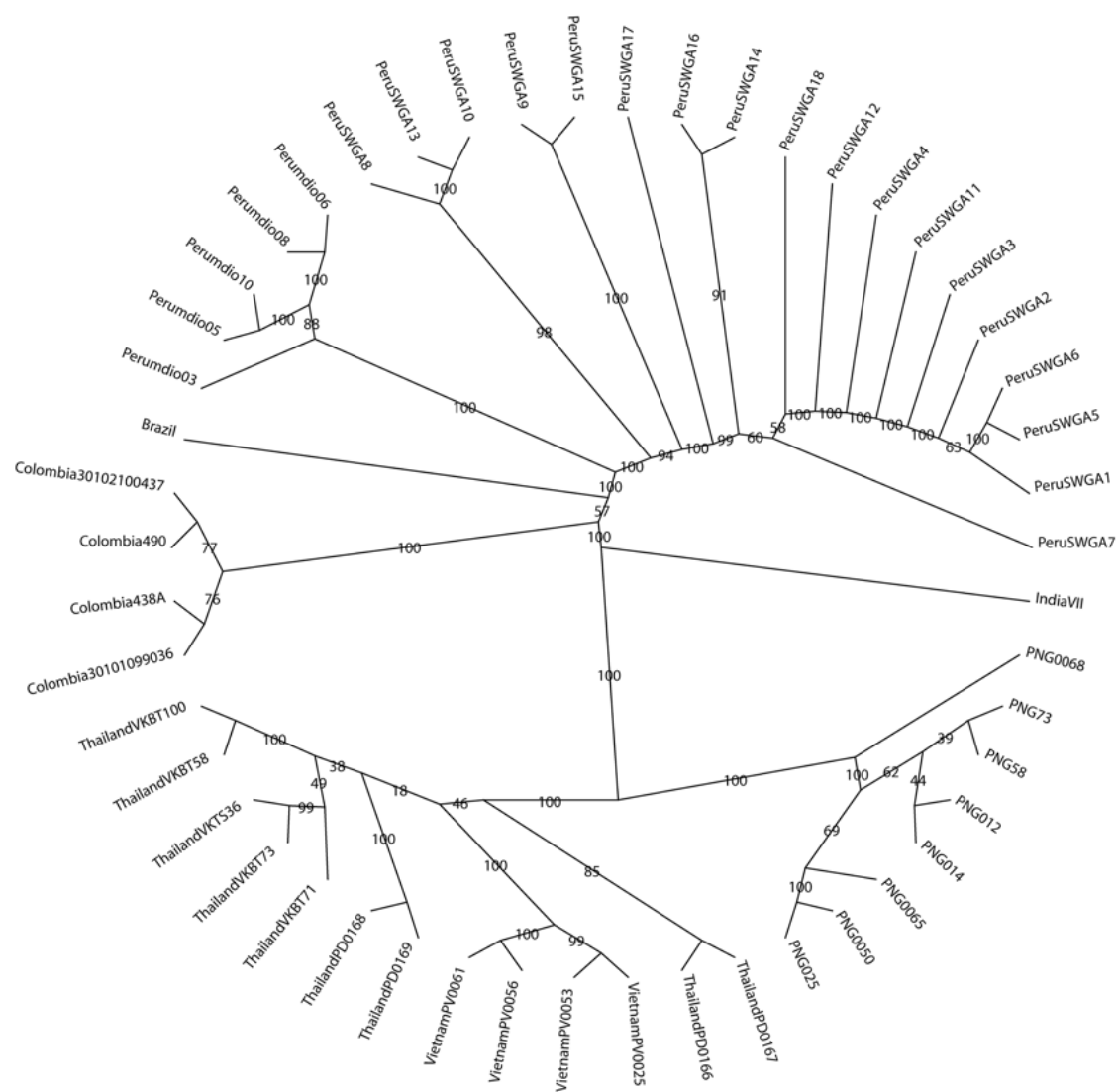


Figure 3-6. Neighbor joining tree of *Plasmodium vivax* clinical samples from different regions of the world. The tree was constructed with core genome-wide SNP data from *P. vivax* samples that underwent selective whole genome amplification (Peru-SWGA-1 through Peru-SWGA-18) and *P. vivax* clinical samples from previously published studies that underwent leukocyte filtration (Pearson et al., 2016) (Thailand-PD0169, Thailand-PD0168, Thailand-PD0166, Thailand-PD-0167, PNG-0050, PNG-0065, PNG-0068, Vietnam-PV0025, Vietnam-PV0053, Vietnam-PV0056, Vietnam-PV0061), hybrid selection (Hupalo et al., 2016) (Colombia-30102100437, Colombia-490, Colombia-438-A, Colombia-30101099036, Thailand-VKTS-36, Thailand-VKBT-73, Thailand-VKBT-58, Thailand VKBT-71, Thailand-VKBT-100, PNG-73, PNG-58, PNG-012, PNG-014, PNG-025), or adaption to growth in splenectomized monkeys (Neafsey et al., 2012) (Brazil, IndiaVII) prior to sequencing. Bootstrap values are shown on each corresponding branch.

3.4 Discussion

In this study, we validate SWGA as a cost-effective, robust method to enrich *P. vivax* gDNA from unprocessed whole blood and dried blood spot clinical samples to improve the efficiency and decrease the cost of subsequent WGS. This is a method that can be applied to clinical samples infected with other malaria species, such as *P. malariae*, *P. ovale curtisi*, and *P. ovale wallikeri*, where parasite densities are low, and where there is no routine *ex-vivo* culture (Ansari et al., 2016), though species-specific primer sets would likely be required. SWGA utilizes readily available reagents, does not require processing at the time of sample collection, and can be performed in a simple, overnight reaction. While several methods have been used successfully to enrich *P. vivax* gDNA for WGS from infected whole blood samples, most are resource and labor intensive. Short-term *ex-vivo* culture of *P. vivax* isolates or adaptation to growth in monkeys produce a large amount of *P. vivax* DNA, but require significant resources. Single cell sequencing allows for highly sensitive dissection of multiclonal samples, however this approach requires cryopreserved samples and specialized laboratory equipment (Nair et al., 2014). While leukocyte filtration is cost-effective and efficient, it is not always possible to perform at field sites with limited infrastructure, because samples require refrigeration within 6 hours to minimize white blood cell lysis and reduce irreversible contamination from human DNA. Hybrid selection is less labor intensive, but the production of the RNA baits used for capture requires either large amounts of *P. vivax* Sal-1 DNA or costly commercially-synthesized RNA bait.

Using SWGA, we achieved a higher than average callable *P. vivax* genome compared to leukocyte-depleted clinical samples sequenced at a similar depth. SWGA generally yielded the highest genome coverage for clinical samples with the highest parasite densities, consistent with our experience with *P. falciparum* (Sundararaman et

al., 2016). For the 12 samples with parasite density >5,000 parasite/μl (0.1% parasitemia), we were able to call on average 71.5% of the core genome, compared to 37% for the 6 samples with parasite densities <5,000 parasite/μl. Increased sequencing effort is needed to obtain maximal genome coverage for samples with lower parasite densities (Figure S3-8). In these cases, the low genome coverage is likely the result of stochastic amplification of a small number of starting *P. vivax* genomes, which leads to very deep coverage of some genomic regions and little or no coverage of others (Sundararaman et al., 2016). If maximal genome coverage is desired, sequential SWGA reactions with pvset1920 followed by pvset1 increases coverage slightly (3-5% 1x genome coverage) (Figure S3-6). Additionally, performing multiple independent SWGA reactions on a sample and combining the sequencing reads can improve genome coverage (4-12% 1x genome coverage) (Figure S3-7). Since multiple rounds of SWGA or pooling the products from multiple reactions increases workload and expenses for a small improvement in genome coverage, we opted for a single amplification reaction with pvset1920 for our samples. However, these protocol modifications may be useful if high genome coverage is needed from samples with low parasitemia.

One potential limitation of SWGA of *P. vivax* from clinical samples is the ability to amplify all clones in a multiclonal sample. One of our samples appeared to be multiclonal by microsatellite analysis of the unamplified gDNA; yet, F_{ws} analysis of the post-SWGA sequence suggested that the sample was comprised of a single clone. It is possible that in this case a majority clone was amplified preferentially over minority clones. Another possibility is that the microsatellite markers assessment may overestimate clonality. Analyses of additional multiclonal samples will be necessary to address this question. Another important limitation of SWGA is that copy number variant (CNV) detection is not possible on post-SWGA sequences. The uneven distribution of primer-targeted motifs in

the target genome results in peaks and troughs in mean genome coverage that can confound CNV detection methods. Finally, SWGA requires long strands of gDNA for efficient amplification of the target genome and is unlikely to work well on degraded or ancient DNA samples.

Whole genome analysis has the potential to reveal much about the biology and epidemiology of *P. vivax* infections. For example, comparison of recurrent infections using WGS can help distinguish relapse due to reactivation of hypnozoites from reinfection or drug resistance, an epidemiological distinction of public health importance. SWGA enables high quality and cost-effective WGS of *P. vivax* from unprocessed blood samples that would otherwise be impossible or prohibitively expensive to sequence. Advanced technologies such as SWGA will greatly facilitate future *P. vivax* whole genome sequencing projects, thereby improving our ability to understand and combat the most widespread form of malaria.

3.5 Materials and methods

Patient Sample Collection and Preparation. The *P. vivax* infected DNA sample used for initial testing of selective amplification primer sets (MRL2) was provided by the Malaria Reference Laboratory of the London School of Hygiene and Tropical Medicine, London, UK. The six dried blood spot samples used in this study were collected from patients with symptomatic *P. vivax* infection in Peru. Parasitemia was quantified with real time PCR and gDNA was extracted using a QIAamp DNA Blood Mini Kit (Qiagen).

Eighteen whole blood samples used for additional testing and further sequencing analysis were derived from whole blood samples collected from patients with symptomatic *P. vivax* infections from two sites around Iquitos, Peru during a study conducted by the US Naval Medical Research Unit No. 6 (Durand et al., 2014). Thick blood smears were

examined to identify the parasite species and to determine the level of parasitemia. Parasite density was calculated by counting the number of asexual parasites per 200 white blood cells in the thick smear (Assuming an average of 6,000 white blood cells per μl). Each blood smear was examined by two microscopists independently and was examined by a third microscopist in the event of a discrepancy. The final parasite density was calculated as the average of density readings from the two concordant microscopists. Whole blood samples were collected in the field using EDTA-containing vacutainer tubes. Samples were frozen and transported to the central lab until further processing. DNA was isolated from thawed whole blood using QIAamp DNA Blood Mini Kit (Qiagen) following the manufacturer's recommendation and as described elsewhere (Baldeviano et al., 2015). Samples were subsequently resuspended in TE buffer and gDNA was quantified using a Qubit 2.0 fluorometer.

Primer design. The initial set of pvset1 primers was designed as described (Sundararaman et al., 2016). Primer set pvset1920 was designed using the *swga* program, which scores primer sets based on selectivity and evenness of binding (measured using the Gini index), and thus automates and improves primer selection. The source code of *swga*, along with download links and documentation are available at <https://www.github.com/eclarke/swga>. pvset1920 was designed to specifically amplify longer regions (>195,000 bp) of the *P. vivax* reference genome (Sal-1) that were GC rich (48.5-50.6%) and yielded low genome coverage following SWGA with pvset1. A total of 1,939 primer sets were identified that exhibited a minimum background binding distance of 25,000 bp and a maximum foreground binding distance of 37,000 bp. These were scored using *swga*'s composite primer scoring algorithm (Gini score*foreground

mean/background mean) and the set with the lowest score (pvset1920) was chosen for testing.

pvset1 consists of 10 primers: 5'- CGTTG*C*G-3' 5'-TTTTTTC*G*C-3', 5'-TCGTG*C*G-3', 5'-CGTTTTTT*T*T-3', 5'-TTTTTTC*G*T-3', 5'-CCGTT*C*G-3', CGTTTC*G*T-3', 5'-CGTTTC*G*C-3', 5'-CGTTTT*C*G-3', 5'-TCGTTC*G*T-3', with asterisks indicating phosphorothioate bonds that are necessary to prevent degradation by phi29. pvset1920 set consists of 12 primers: 5'-AACGAAGC*G*A-3', 5'-ACGAAGCG*A*A-3', 5'-ACGACGA*A*G-3', 5'-ACGCGCA*A*C-3', 5'-CAACGCG*G*T-3', 5'-GACGAAA*C*G-3', 5'-GCGAAAAA*G*G-3', 5'-GCGAAGC*G*A-3', 5'-GCGGAAC*G*A-3', 5'-GCGTCGA*A*G-3', 5'-GGTTAGCG*G*C-3', and 5'-AACGAAT*C*G-3'.

Selective Whole Genome Amplification. 30-70 ng of input DNA was added to a 50 µl reaction containing 3.5 uM SWGA primers, 30 U phi29 DNA polymerase enzyme (New England Biolabs), phi29 DNA buffer (New England Biolabs), 1% bovine serum albumin, and water. The reaction was carried out on a thermocycler consisting of a ramp down from 35°C to 30°C (10 minutes per degree), 16 hours at 30°C, 10 minutes at 65°C, and hold at 4°C. The samples were diluted 1:1 with DNase-free, RNase-free water and purified with Ampure XP beads (Beckman-Coulter) at a 1:1 ratio per the manufacturer's protocol. When performed, a second round of selective amplification contained 100-200 ng of the Ampure XP purified product from the first reaction.

Methylation digest. 125-500 ng of gDNA extracted from *P. vivax* infected whole blood samples was digested with 5 units of FspEI (New England Biolabs) and 5 units of MspJI (New England Biolabs) enzymes in a 30 µl reaction. A mock digest with an identical amount of gDNA and no enzymes was run in parallel. Samples were digested for 2 hours

at 37°C then were heat inactivated at 80°C for 15 minutes. For coverage analysis, rarefaction analysis was performed on Illumina MiSeq sequencing reads derived from samples digested with MspJI or FspEI (digest and SWGA) or mock digested with no enzyme prior to SWGA with primer set pvset1920. Coverage obtained by mapping 200,000 MiSeq sequencing reads (~25 million base pairs of sequencing depth) was compared between digested and mock digested samples. P value was obtained using a two tailed t-test.

Whole genome sequencing. SWGA products and unamplified DNA used for primer set testing were sequenced on an Illumina MiSeq using a modified Nextera library preparation method. For Hiseq runs, next generation sequencing libraries of SWGA products were prepared using the Nextera XT DNA preparation kit (Illumina) per the manufacturer's protocol. These samples were pooled and clustered on a Hiseq 2500 (Illumina) in Rapid Run mode with 100 base pair paired end reads. For the coverage analysis of leukocyte-filtered samples, fastq files from a prior study of *P. vivax* field samples (Flannery et al., 2016) were used.

Raw fastq files were aligned to the Sal-1 reference genome (PlasmoDB version 13, <http://plasmodb.org/common/downloads/release-13.0/PvivaxSal1/fasta/data/>) using the Burroughs-Wheeler Aligner (version 0.7.8) (Li, 2013.) and samtools (version 0.1.19) (Li, 2011; Li et al., 2009) as previously described in the Platypus pipeline (Manary et al., 2014). Picard (version 2.0.1) was used to remove unmapped reads and the Genome Analysis Toolkit (GATK) (McKenna et al., 2010) was used to realign the sequences around the indels. Picard's CollectGcBiasMetrics tool was used to generate the GC bias plots. GATK's DepthOfCoverage tool was used to determine the percent of the total and core genome covered by ≥ 5 reads, mean coverage, and coverage over the core genome. The

coordinates of the *P. vivax* core genome, which excludes subtelomeric and hypervariable regions with significantly higher read mapping errors, was obtained from a recent analysis of hundreds *P. vivax* sequences from clinical isolates (Pearson et al., 2016).

For rarefaction analyses, sequences were aligned with smalt (Wellcome Trust Sanger Institute) and mapped with custom scripts in R (<https://www.r-project.org/>). To visualize base composition across chromosomes, plots were created in Geneious (version 9.1) (Kearse et al., 2012) using the Sal-1 *P. vivax* reference sequences and 25 bp windows. Plots of chromosome coverage were created with IGVTools (version 2.3.40) (Robinson et al., 2011; Thorvaldsdóttir et al., 2013).

Variant Calling and Analysis. We followed the GATK's best practices to call variants (Auwera et al., 2014; DePristo et al., 2011). The aligned sequences were run through GATK's HaplotypeCaller in "reference confidence" mode to create genomic GVCF files for each sample. This reference confidence model highlights areas of the genome that are likely to have variation and produces a comprehensive record of genotype likelihoods and annotations for each site. The samples were joint genotyped using the GenotypeGVCFs tool. Variants were further filtered based on quality scores and sequencing bias statistics based on default parameters from GATK. SNPs were filtered out if they met any of the following criteria: Quality Depth (QD) < 2.0, Mapping Quality (MQ) < 50.0, Phred-scaled p-value using Fisher's exact test to detect strand bias (FS) >60.0, Symmetric Odds Ratio (SOR) >4.0, Z-score from Wilcoxon rank sum test of Alternative vs. Reference read mapping qualities (MQRankSum) < -12.5, ReadPosRankSum (RPRS) < -8.0. Variants were annotated using snpeff (version 4.2) (Cingolani et al., 2012).

F_{ws} of samples with the highest genome coverage was estimated using *moimix* (<http://dx.doi.org/10.5281/zenodo.58257>) a package available through R. The package

calculates F_{ws} statistic using the equation $F_{ws} = 1 - (Hw/Hs)$, where Hw is the within-host heterozygosity and Hs is the population-level heterozygosity (Auburn et al., 2012; Manske et al., 2012). The core *P. vivax* genome, as defined by Pearson et al. (2016) was used for core genome analysis. For microsatellite genotyping, five neutral microsatellite loci of significant variability in the Peruvian Amazon were typed in a previous study (Durand et al., 2014). If there was more than one marker at any given locus, the sample was considered multiclonal, per prior genotyping studies (De Souza et al., 2015; Imwong et al., 2006; Karunaweera et al., 2008).

A neighbor joining tree was constructed using SNPs from the core *P. vivax* genome from sequences obtained in this study, along with sequences from previously published studies that are available in the NCBI Short Read Archive. The mdio samples were from a previous study conducted by our lab in Peru (Flannery et al., 2016) and the rest of the sequences were obtained from two recent large scale *P. vivax* sequencing studies (Hupalo et al., 2016; Pearson et al., 2016). In order to assess the phylogenetic relationships of sequenced isolates we constructed a multiple sequence alignment from filtered SNPs called in GATK using an in-house perl script. This alignment was used as input for Maximum Likelihood phylogenetic analysis in the Randomized Accelerated Maximum Likelihood program (Stamatakis, 2014) (RAxML) with 500 pseudoreplicates using the generalized time reversible model and the resulting tree was visualized in dendroscope (Huson and Scornavacca, 2012)

Ethics approval and consent to participate. The sample from Malaria Research Laboratories (MRL) was an anonymized DNA sample previously collected by the MRL and provided under the MRL's remit to undertake epidemiological surveillance relevant to imported malaria in the UK. The protocol for the collection of field samples was approved

by the Institutional Review Board of the US Naval Medical Research Center (Protocol NMRC.D.2005.0005) and the National Institutes of Health of Peru (Protocol 009-2004) in compliance with all applicable Federal regulation governing the protection of human subjects. All adult subjects provided written informed consent and all children 8-17 years old provided verbal assent to participate in the study.

Availability of data and materials. The *P. vivax* genome Illumina sequencing reads of the 22 samples used for variant analysis in this study will be available on the National Center for Biotechnology Information's Short Read Archive with the accession number: PRJNA344889.

3.6 Chapter acknowledgments

EAW was supported a National Institutes of Health (NIH) Grant R01 AI 103058. Support for sample preparation and sequencing at University of California, San Diego was provided by NIH Grant P50 GM 085764. ANC was supported through by a NIH T32 AI 007036 grant. Work done by BHH, DEL, and SAS was supported by grants from the NIH (R01 AI 091595, T32 AI 007532, P30 AI 045008). KF was supported by a UC San Diego Clinical and Translational Research Institute Grant UL1TR001442. JMV received support for the collection of dried blood spot samples from grants from the NIH/NIAID (U19 AI 089681, and D43 TW 007120). AGL received support from the training grant 2D43 TW007393 awarded by the Fogarty International Center of the NIH. We wish to thank the Malaria Research Laboratory UK, and all the patients who provided the samples for this study.

3.7 Supplemental figures and tables

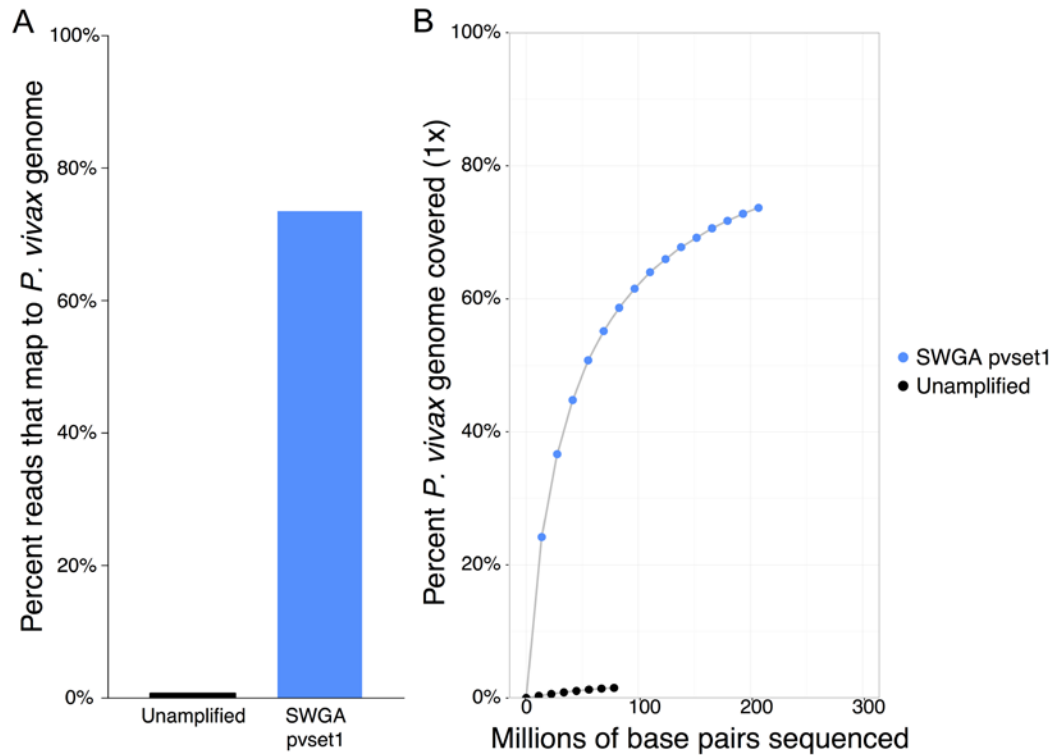


Figure S3-1. Testing of SWGA primer sets on an unprocessed, *P. vivax*-infected blood sample. (A) Unamplified DNA (black) and DNA amplified with SWGA primer set pvset1 (blue) was sequenced on a MiSeq (Illumina). The percentage of MiSeq reads that mapped to the *P. vivax* Sal-1 reference genome in Geneious (version 9.1) was plotted for both the unamplified and SWGA-amplified sample. (B) Rarefaction analysis compares the $\geq 1\times$ *P. vivax* genome coverage relative to total sequencing depth (in millions of base pairs sequenced) with and without SWGA.

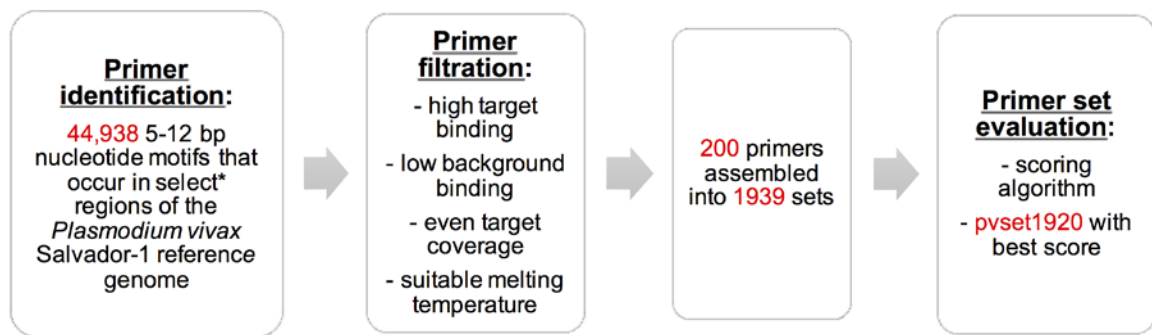


Figure S3-2. Flowchart describing the process of SWGA primer set selection for *P. vivax*. *, selected regions are >195,000 bp in length with an average 48.5 to 50.6% GC composition.

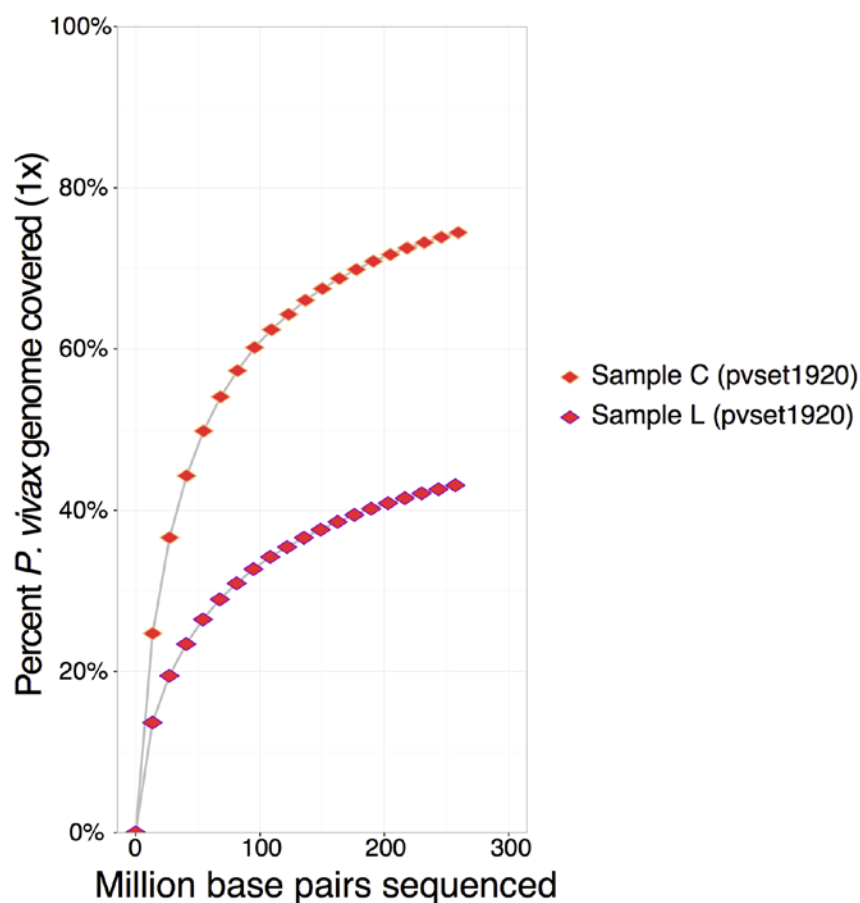


Figure S3-3. Testing of SWGA primer set pvset1920 on DNA from *P. vivax* patient-derived dried blood spots. Genomic DNA from two dried blood spots with either a relatively high (sample C, 56,790 parasites/ μ l) or low (sample L, 2,572 parasites/ μ l) parasite density was subjected to SWGA with pvset1920 and sequenced on a MiSeq (Illumina). The percentages of the *P. vivax* genome with $\geq 1\times$ coverage are shown relative to total sequencing depth (in millions of base pairs sequenced).

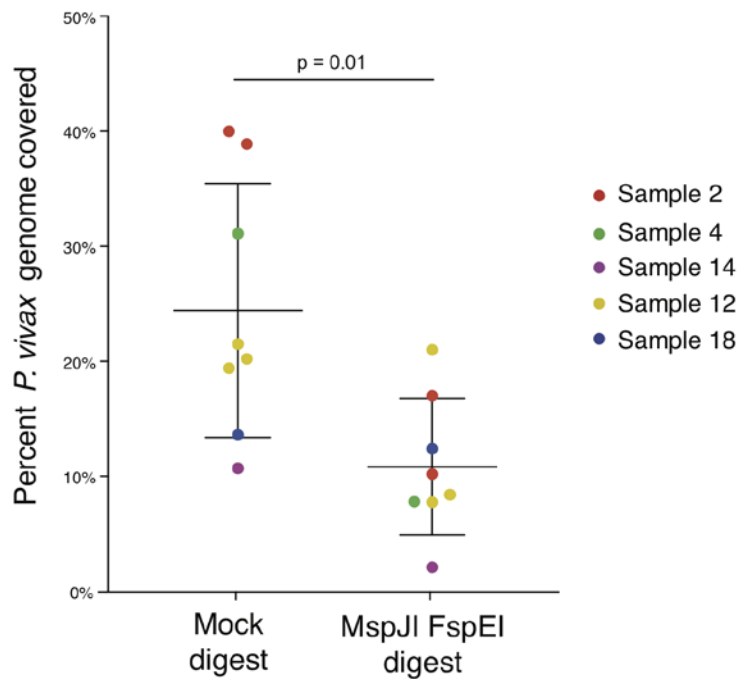


Figure S3-4. Comparison of SWGA following either a mock digestion or methylation-dependent restriction digest. Genomic DNA extracted from five *P. vivax*-infected patient samples was subjected to a mock digest or digest with the methylation-dependent restriction enzymes MspJI and FspEI. The percentage of the *P. vivax* reference genome covered (1×) by mapping 200,100 MiSeq (Illumina) reads (approximately 25 million bp of sequencing depth) from each sample with or without digest is shown. Statistical significance was calculated using a two-tailed t test.

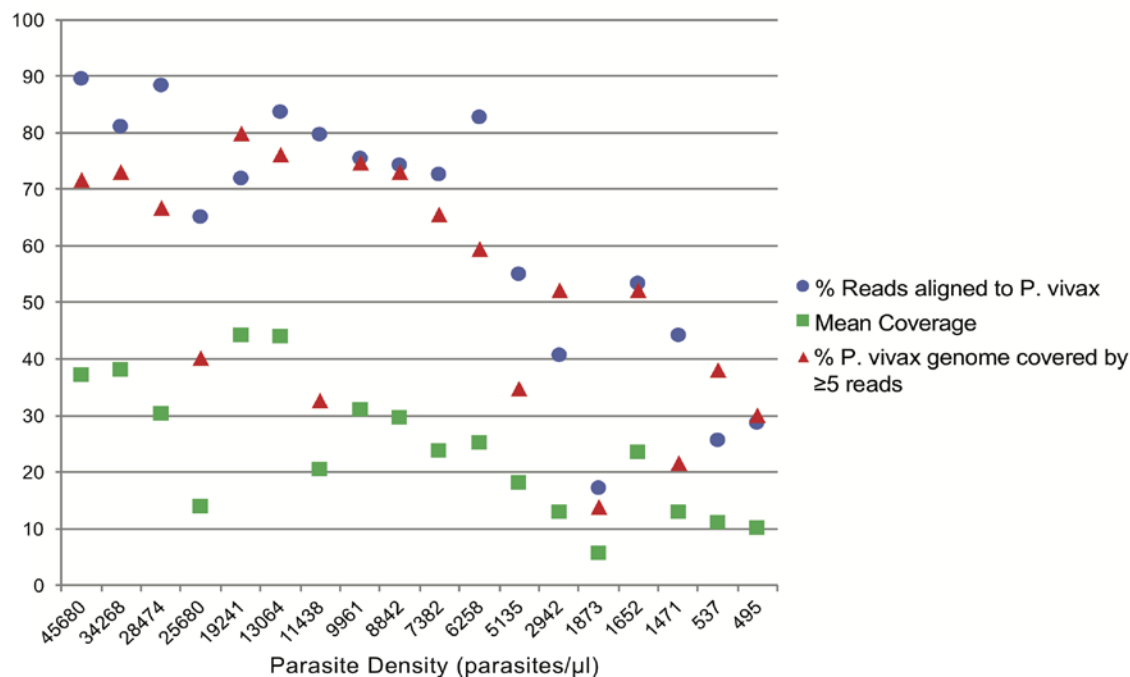


Figure S3-5. Sequencing statistics for *P. vivax* clinical samples with various parasite densities. The percentage of reads that map to *P. vivax* Sal-1 (blue circles), the mean *P. vivax* genome coverage (green squares), and the percentage of the *P. vivax* genome covered by ≥5 reads (red triangles) are shown for all 18 samples amplified by SWGA, with sample parasitemia indicated on the x axis. Sequencing statistics were determined using the Genome Analysis Toolkit (GATK) DepthOfCoverage tool.

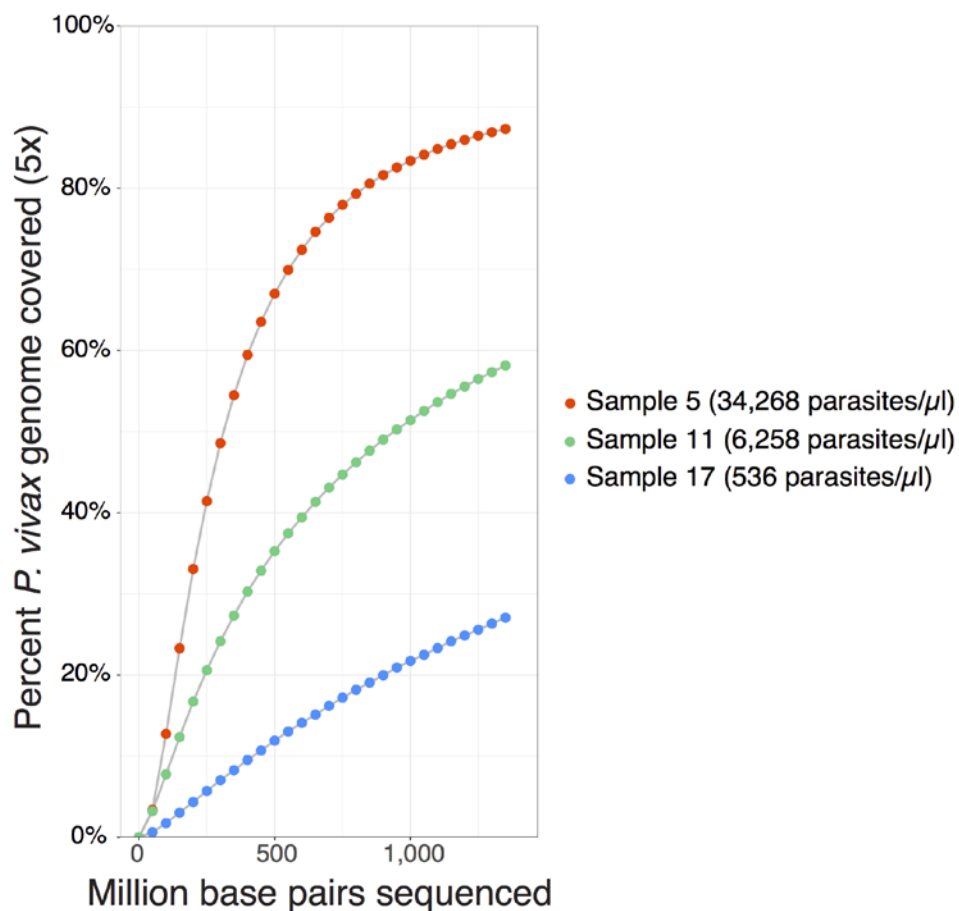


Figure S3-6. *P. vivax* genome coverage for samples with high, medium, and low parasite densities. The percentages of the *P. vivax* reference genome covered at $\geq 1\times$ are shown relative to the total sequencing depth generated on an Illumina HiSeq (in millions of base pairs sequenced) for samples with high (red), medium (green), and low (blue) parasite densities.

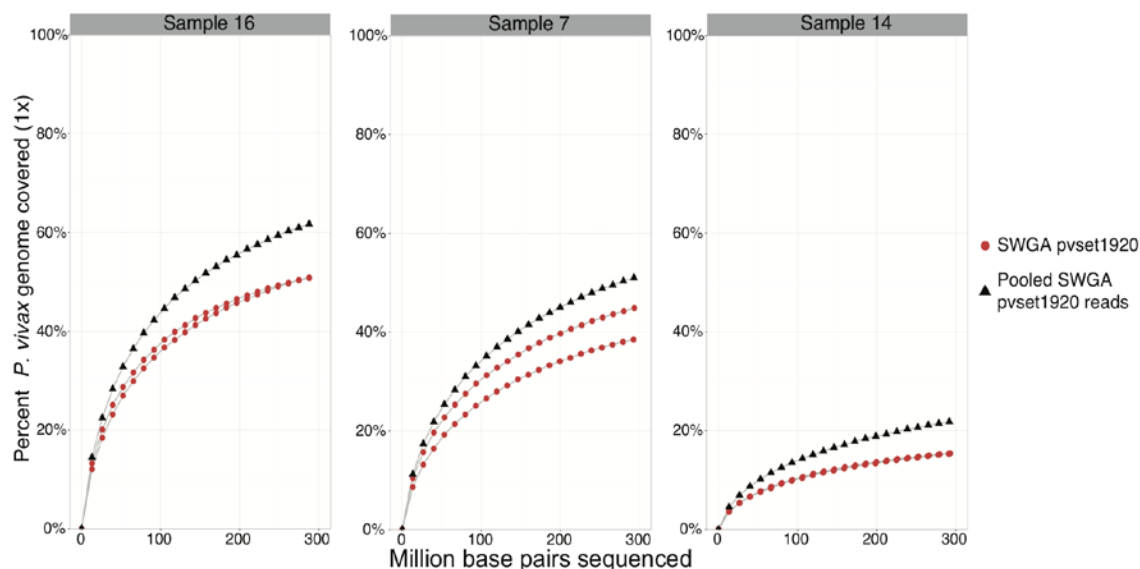


Figure S3-7. Impact of multiple independent SWGA reactions per sample on *P. vivax* genome coverage. Each panel represents the results obtained for a single patient sample (samples 16, 7, and 14). The percentage of the *P. vivax* reference genome covered at $\geq 1\times$ is shown relative to the total sequencing depth generated on an Illumina MiSeq (in millions of base pairs sequenced) for each individual SWGA reaction (red) or pooled reads from the two independent SWGA reactions (black).

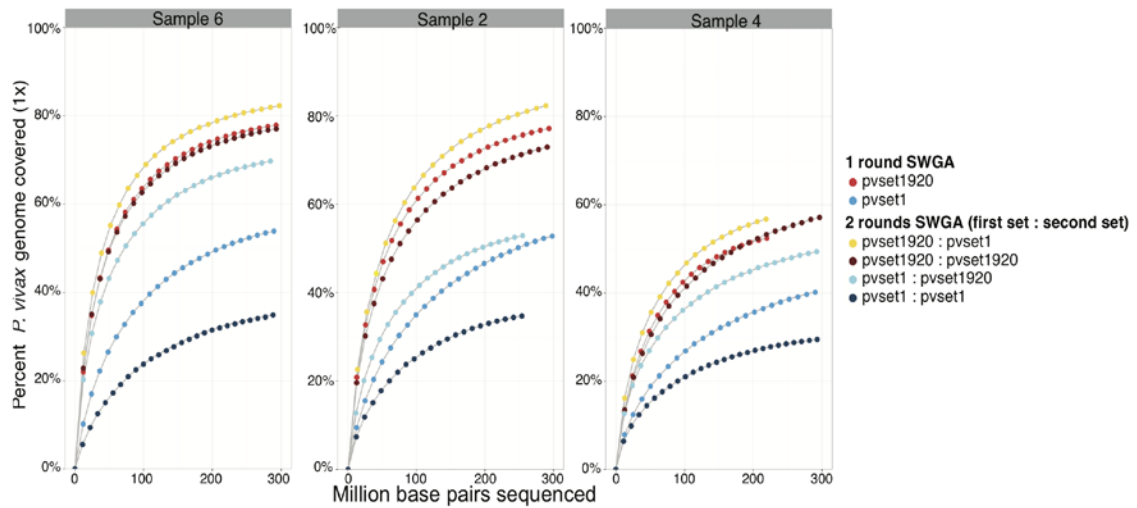


Figure S3-8. Combinatorial testing of SWGA primer sets on three *P. vivax*-infected patient samples. Each panel represents the results obtained for one of the three individual patient samples (samples 6, 2, and 4) after one round of SWGA with primer set pvset1920 (red) or pvset1 (blue), after two rounds of SWGA with the same primer set (dark red for pvset1920 and dark blue for pvset1), or after two rounds of SWGA with different primer sets (yellow for pvset1920 followed by pvset1 and light blue for pvset1 followed by pvset1920). For all primer set combinations, the percentage of the *P. vivax* reference genome covered at $\geq 1\times$ is shown relative to the total sequencing depth (in millions of base pairs sequenced) generated on a MiSeq (Illumina).

Table S3-1. Drug resistance mutations in *Plasmodium vivax* sequences from dried blood spot DNA subjected to SWGA.

Locus	Chr	Position	Ref	Alt	Amino Acid	Samples*
<i>pvcrt-0</i> (PVX_087980)	1	331151	T	C	Intron	3 (4)
		331819	G	A	Intron	1 (4)
		332453	T	C	Intron	4 (4)
		332874	A	C	Intron	4 (4)
<i>pvdhfr</i> (PVX_089950)	5	964762	G	A	Ser58Asn	1 (4)
		964763	C	G,A	Ser58Arg	3 (3)
		964939	G	A	Ser117Asn	3 (3)
<i>pvm-dr1</i> (PVX_080100)	10	363223	A	G	Thr958Met	4 (4)
		363374	T	G	Met908Leu	2 (3)
<i>pvmrp2</i> (PVX_124085)	14	2043859	G	C	Gln1407Glu	1 (3)
		2044327	T	A	Asn1251Tyr	1 (3)
		2045050	C	T	Val1010Met	4 (4)
		2047233	C	A	Arg282Met	2 (2)
		2047816	C	G	Glu88Gln	2 (4)
		2047893	C	T	Cys62Tyr	1 (4)
<i>dhps</i> (PVX_123230)	14	1257856	G	C	Ala383Gly	1 (4)
		1258389	C	T	Met205Ile	2 (3)

*The number of samples confidently genotyped (in parentheses) and the number bearing the indicated allele.

3.8 Chapter references

- Ansari, H.R., Templeton, T.J., Subudhi, A.K., Ramaprasad, A., Tang, J., Lu, F., Naeem, R., Hashish, Y., Oguike, M.C., Benavente, E.D., Clark, T.G., Sutherland, C.J., Barnwell, J.W., Culleton, R., Cao, J., Pain, A., 2016. 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.
- Auburn, S., Campino, S., Miotto, O., Djimde, A.A., Zongo, I., Manske, M., Maslen, G., Mangano, V., Alcock, D., MacInnis, B., Rockett, K.A., Clark, T.G., Doumbo, O.K., Ouédraogo, J.B., Kwiatkowski, D.P., 2012. Characterization of within-host *Plasmodium falciparum* diversity using next-generation sequence data. *PLoS One* 7, e32891.
- Auburn, S., Marfurt, J., Maslen, G., Campino, S., Ruano Rubio, V., Manske, M., MacHunter, B., Kenangalem, E., Noviyanti, R., Trianty, L., Sebayang, B., Wirjanata, G., Sriprawat, K., Alcock, D., MacInnis, B., Miotto, O., Clark, T.G., Russell, B., Anstey, N.M., Nosten, F., Kwiatkowski, D.P., Price, R.N., 2013. Effective preparation of *Plasmodium vivax* field isolates for high-throughput whole genome sequencing. *PLoS One* 8, e53160.
- Auwer, G.A. Van Der, Carneiro, M.O., Hartl, C., Poplin, R., Levy-moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D.,

- Gabriel, S., Depristo, M.A., 2014. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinforma.* 43, 11.10.1-11.10.33.
- Baldeviano, G.C., Akinyi Okoth, S., Arrospide, N., Gonzalez, R. V., Sánchez, J.F., Macedo, S., Conde, S., Tapia, L.L., Salas, C., Gamboa, D., Herrera, Y., Edgel, K.A., Udhayakumar, V., Lescano, A.G., 2015. Molecular epidemiology of *Plasmodium falciparum* malaria outbreak, Tumbes, Peru, 2010–2012. *Emerg. Infect. Dis.* 21, 797–803. =
- Borrmann, S., Straimer, J., Mwai, L., Abdi, A., Rippert, A., Okombo, J., Muriithi, S., Sasi, P., Kortok, M.M., Lowe, B., Campino, S., Assefa, S., Auburn, S., Manske, M., Maslen, G., Peshu, N., Kwiatkowski, D.P., Marsh, K., Nzila, A., Clark, T.G., 2013. Genome-wide screen identifies new candidate genes associated with artemisinin susceptibility in *Plasmodium falciparum* in Kenya. *Sci. Rep.* 3, 3318.
- Bright, A.T., Tewhey, R., Abeles, S., Chuquiyauri, R., Llanos-Cuentas, A., Ferreira, M.U., Schork, N.J., Vinetz, J.M., Winzeler, E.A., 2012. Whole genome sequencing analysis of *Plasmodium vivax* using whole genome capture. *BMC Genomics* 13, 262.
- Calderaro, A., Piccolo, G., Gorrini, C., Rossi, S., Montecchini, S., Dell’Anna, M.L., De Conto, F., Medici, M.C., Chezzi, C., Arcangeletti, M.C., 2013. Accurate identification of the six human *Plasmodium* spp. causing imported malaria, including *Plasmodium ovale wallikeri* and *Plasmodium knowlesi*. *Malar. J.* 12, 321.
- Carlton, J.M., Adams, J.H., Silva, J.C., Bidwell, S.L., Lorenzi, H., Caler, E., Crabtree, J., Angiuoli, S. V., Merino, E.F., Amedeo, P., Cheng, Q., Coulson, R.M.R., Crabb, B.S., Del Portillo, H.A., Essien, K., Feldblyum, T. V., Fernandez-Becerra, C., Gilson, P.R., Gueye, A.H., Guo, X., Kang’A, S., Kooij, T.W.A., Korsinczky, M., Meyer, E.V.S., Nene, V., Paulsen, I., White, O., Ralph, S.A., Ren, Q., Sargeant, T.J., Salzberg, S.L., Stoeckert, C.J., Sullivan, S.A., Yamamoto, M.M., Hoffman, S.L., Wortman, J.R., Gardner, M.J., Galinski, M.R., Barnwell, J.W., Fraser-Liggett, C.M., 2008. Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature* 455, 757–763.
- Chan, E.R., Menard, D., David, P.H., Ratsimbaoa, A., Kim, S., Chim, P., Do, C., Witkowski, B., Mercereau-Puijalon, O., Zimmerman, P.A., Serre, D., 2012. Whole genome sequencing of field isolates provides robust characterization of genetic diversity in *Plasmodium vivax*. *PLoS Negl. Trop. Dis.* 6, e1811.
- Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 6, 80–92.
- Clarke, E.L., Sundararaman, S.A., Seifert, S.N., Bushman, F.D., Hahn, B.H., Brisson, D., 2017. *swga*: a primer design toolkit for selective whole genome amplification. *Bioinformatics* 33, 2071–2077.

- Cohen-Karni, D., Xu, D., Apone, L., Fomenkov, A., Sun, Z., Davis, P.J., Morey Kinney, S.R., Yamada-Mabuchi, M., Xu, S. -y., Davis, T., Pradhan, S., Roberts, R.J., Zheng, Y., 2011. The MspJI family of modification-dependent restriction endonucleases for epigenetic studies. *Proc. Natl. Acad. Sci.* 108, 11040–11045.
- De Souza, A.M., De Araújo, F.C.F., Fontes, C.J.F., Carvalho, L.H., De Brito, C.F.A., De Sousa, T.N., 2015. Multiple-clone infections of *Plasmodium vivax*: Definition of a panel of markers for molecular epidemiology. *Malar. J.* 14, 330.
- DePristo, M. a., Banks, E., Poplin, R.E., Garimella, K. V., Maguire, J.R., Hartl, C., Philippakis, a. a., del Angel, G., Rivas, M. a, Hanna, M., McKenna, A., Fennell, T.J., Kernysky, a. M., Sivachenko, a. Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly, M.J., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43, 491–498.
- Durand, S., Cabezas, C., Lescano, A.G., Galvez, M., Gutierrez, S., Arrospide, N., Alvarez, C., Santolalla, M.L., Bacon, D.J., Graf, P.C.F., 2014. Efficacy of three different regimens of primaquine for the prevention of relapses of *Plasmodium vivax* malaria in the Amazon Basin of Peru. *Am. J. Trop. Med. Hyg.* 91, 18–26.
- Flannery, E.L., Wang, T., Akbari, A., Corey, V.C., Gunawan, F., Bright, A.T., Abraham, M., Sanchez, J.F., Santolalla, M.L., Baldeviano, G.C., Edgel, K.A., Rosales, L.A., Lescano, A.G., Bafna, V., Vinetz, J.M., Winzeler, E.A., 2016. Next-Generation sequencing of *Plasmodium vivax* tatient samples shows evidence of direct evolution in drug-resistance Genes. *ACS Infect. Dis.* 1, 367–379.
- Guggisberg, A.M., Sundararaman, S.A., Lanaspá, M., Moraleda, C., González, R., Mayor, A., Cisteró, P., Hutchinson, D., Kremsner, P.G., Hahn, B.H., Bassat, Q., Odom, A.R., 2016. Whole-genome sequencing to evaluate the resistance landscape following antimalarial treatment failure with fosmidomycin-clindamycin. *J. Infect. Dis.* 214, 1085–1091.
- Hester, J., Chan, E.R., Menard, D., Mercereau-Puijalon, O., Barnwell, J., Zimmerman, P.A., Serre, D., 2013. *De novo* assembly of a field isolate genome reveals novel *Plasmodium vivax* erythrocyte invasion genes. *PLoS Negl. Trop. Dis.* 7, e2569.
- Hupalo, D.N., Luo, Z., Melnikov, A., Sutton, P.L., Rogov, P., Escalante, A., Vallejo, A.F., Herrera, S., Arévalo-Herrera, M., Fan, Q., Wang, Y., Cui, L., Lucas, C.M., Durand, S., Sanchez, J.F., Christian Baldeviano, G., Lescano, A.G., Laman, M., Barnadas, C., Barry, A., Mueller, I., Kazura, J.W., Eapen, A., Kanagaraj, D., Valecha, N., Ferreira, M.U., Roobsoong, W., Nguitragool, W., Sattabonkot, J., Gamboa, D., Kosek, M., Vinetz, J.M., González-Cerón, L., Birren, B.W., Neafsey, D.E., Carlton, J.M., 2016. Population genomics studies identify signatures of global dispersal and drug resistance in *Plasmodium vivax*. *Nat. Genet.* 48, 953–958.
- Huson, D.H., Scornavacca, C., 2012. Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* 61, 1061–1067.
- Imwong, M., Sudimack, D., Pukrittayakamee, S., Osorio, L., Carlton, J.M., Day, N.P.J., White, N.J., Anderson, T.J.C., 2006. Microsatellite variation, repeat array length,

- and population history of *Plasmodium vivax*. *Mol. Biol. Evol.* 23, 1016–1018.
- Karunaweera, N.D., Ferreira, M.U., Munasinghe, A., Barnwell, J.W., Collins, W.E., King, C.L., Kawamoto, F., Hartl, D.L., Wirth, D.F., 2008. Extensive microsatellite diversity in the human malaria parasite *Plasmodium vivax*. *Gene* 410, 105–112.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649.
- Larremore, D.B., Sundararaman, S.A., Liu, W., Proto, W.R., Clauset, A., Loy, D.E., Speede, S., Plenderleith, L.J., Sharp, P.M., Hahn, B.H., Rayner, J.C., Buckee, C.O., 2015. Ape parasite origins of human malaria virulence genes. *Nat. Commun.* 6, 8368.
- Leichty, A.R., Brisson, D., 2014. Selective whole genome amplification for resequencing target microbial species from complex natural samples. *Genetics* 198, 473–481.
- Li, H., 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993.
- Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997*.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Manary, M.J., Singhakul, S.S., Flannery, E.L., Bopp, S.E.R., Corey, V.C., Bright, A.T., McNamara, C.W., Walker, J.R., Winzeler, E.A., 2014. Identification of pathogen genomic variants through an integrated pipeline. *BMC Bioinformatics* 15, 63.
- Manske, M., Miotto, O., Campino, S., Auburn, S., Almagro-Garcia, J., Maslen, G., O'Brien, J., Djimde, A., Doumbo, O., Zongo, I., Ouedraogo, J.-B., Michon, P., Mueller, I., Siba, P., Nzila, A., Borrmann, S., Kiara, S.M., Marsh, K., Jiang, H., Su, X.-Z., Amaratunga, C., Fairhurst, R., Socheat, D., Nosten, F., Imwong, M., White, N.J., Sanders, M., Anastasi, E., Alcock, D., Drury, E., Oyola, S., Quail, M.A., Turner, D.J., Ruano-Rubio, V., Jyothi, D., Amenga-Etego, L., Hubbard, C., Jeffreys, A., Rowlands, K., Sutherland, C., Roper, C., Mangano, V., Modiano, D., Tan, J.C., Ferdig, M.T., Amambua-Ngwa, A., Conway, D.J., Takala-Harrison, S., Plowe, C. V., Rayner, J.C., Rockett, K.A., Clark, T.G., Newbold, C.I., Berriman, M., MacInnis, B., Kwiatkowski, D.P., 2012. Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature* 487, 375–379.
- McCutchan, T., Dame, J., Miller, L., Barnwell, J., 1984. Evolutionary relatedness of *Plasmodium* species as determined by the structure of DNA. *Science*. 225, 808–811.

- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K. V, Altshuler, D., Gabriel, S., Daly, M., Depristo, M.A., 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303.
- Miotto, O., Almagro-Garcia, J., Manske, M., MacInnis, B., Campino, S., Rockett, K. a, Amaratunga, C., Lim, P., Suon, S., Sreng, S., Anderson, J.M., Duong, S., Kwiatkowski, D.P., 2013. Multiple populations of artemisinin-resistant *Plasmodium falciparum* in Cambodia *Nat. Genet.* 45, 648–655.
- Morin, J.A., Cao, F.J., Lazaro, J.M., Arias-Gonzalez, J.R., Valpuesta, J.M., Carrascosa, J.L., Salas, M., Ibarra, B., 2012. Active DNA unwinding dynamics during processive DNA replication. *Proc. Natl. Acad. Sci.* 109, 8115–8120.
- Nair, S., Nkhoma, S.C., Serre, D., Zimmerman, P.A., Gorena, K., Daniel, B.J., Nosten, F., Anderson, T.J.C., Cheeseman, I.H., 2014. Single-cell genomics for dissection of complex malaria infections. *Genome Res.* 24, 1028–1038.
- Neafsey, D.E., Galinsky, K., Jiang, R.H.Y., Young, L., Sykes, S.M., Saif, S., Gujja, S., Goldberg, J.M., Young, S., Zeng, Q., Chapman, S.B., Dash, A.P., Anvikar, A.R., Sutton, P.L., Birren, B.W., Escalante, A.A., Barnwell, J.W., Carlton, J.M., 2012. The malaria parasite *Plasmodium vivax* exhibits greater genetic diversity than *Plasmodium falciparum*. *Nat. Genet.* 44, 1046–1050.
- Oyola, S.O., Gu, Y., Manske, M., Otto, T.D., O'Brien, J., Alcock, D., MacInnis, B., Berriman, M., Newbold, C.I., Kwiatkowski, D.P., Swerdlow, H.P., Quail, M.A., 2013. Efficient depletion of host DNA contamination in malaria clinical sequencing. *J. Clin. Microbiol.* 51, 745–751.
- Pearson, R., R, A., S, A., O, M., J, A.-G., Amaratunga, et al., 2016. Genomic analysis of local variation and recent evolution in *Plasmodium vivax*. *Nat Genet* 48, 959–964.
- Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative Genome Viewer. *Nat. Biotechnol.* 29, 24–6.
- Schwartz, A., Baidjoe, A., Rosenthal, P.J., Dorsey, G., Bousema, T., Greenhouse, B., 2015. The effect of storage and extraction methods on amplification of *Plasmodium falciparum* DNA from dried blood spots. *Am. J. Trop. Med. Hyg.* 92, 922–925.
- Singh, B., Sung, L.K., Matusop, A., Radhakrishnan, A., Shamsul, S.S.G., Cox-Singh, J., Thomas, A., Conway, D.J., 2004. A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. *Lancet* 363, 1017–1024.
- Stamatakis, A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
- Sundararaman, S.A., Plenderleith, L.J., Liu, W., Loy, D.E., Learn, G.H., Li, Y., Shaw, K.S., Ayoub, A., Peeters, M., Speede, S., Shaw, G.M., Bushman, F.D., Brisson, D., Rayner, J.C., Sharp, P.M., Hahn, B.H., 2016. Genomes of cryptic chimpanzee *Plasmodium* species reveal key evolutionary events leading to human malaria. *Nat. Commun.* 7, 11078.

- Sutherland, C.J., Tanomsing, N., Nolder, D., Oguike, M., Jennison, C., Pukrittayakamee, S., Dolecek, C., Hien, T.T., do Rosário, V.E., Arez, A.P., Pinto, J., Michon, P., Escalante, A.A., Nosten, F., Burke, M., Lee, R., Blaze, M., Otto, T.D., Barnwell, J.W., Pain, A., Williams, J., White, N.J., Day, N.P.J., Snounou, G., Lockhart, P.J., Chiodini, P.L., Imwong, M., Polley, S.D., 2010. Two Nonrecombining Sympatric Forms of the Human Malaria Parasite *Plasmodium ovale* Occur Globally. *J. Infect. Dis.* 201, 1544–1550.
- Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* 14, 178–192.
- Venkatesan, M., Amaratunga, C., Campino, S., Auburn, S., Koch, O., Lim, P., Uk, S., Socheat, D., Kwiatkowski, D.P., Fairhurst, R.M., Plowe, C. V., 2012. Using CF11 cellulose columns to inexpensively and effectively remove human DNA from *Plasmodium falciparum*-infected whole blood samples. *Malar. J.* 11, 41.
- Winter, D.J., Pacheco, M.A., Vallejo, A.F., Schwartz, R.S., Arevalo-Herrera, M., Herrera, S., Cartwright, R.A., Escalante, A.A., 2015. Whole genome sequencing of field Isolates Reveals Extensive Genetic Diversity in *Plasmodium vivax* from Colombia. *PLoS Negl. Trop. Dis.* 9, e0004252.
- World Health Organisation, 2017. World Malaria Report 2017.
<https://doi.org/http://www.who.int/malaria/publications/world-malaria-report-2017/report/en/>

**CHAPTER 4: Evolutionary history of human *Plasmodium vivax*
revealed by genome-wide analyses of related ape parasites**

Dorothy E. Loy^{1#}, Lindsey J. Plenderleith^{2#}, Sesh A. Sundararaman¹, Weimin Liu¹,
Jakub Gruszczyk³, Yi-Jun Chen^{3,4}, Stephanie Trimboli¹, Gerald H. Learn¹, Oscar A.
MacLean², Alex L.K. Morgan², Yingying Li¹, Alexa N. Avitto¹, Jasmin Giles¹, Sébastien
Calvignac-Spencer⁵, Andreas Sachse⁵, Fabian H. Leendertz⁵, Sheri Speede⁶, Ahidjo
Ayoub⁷, Martine Peeters⁷, Julian C. Rayner⁸, Wai-Hong Tham^{3,4}, Paul M. Sharp^{2*} and
Beatrice H. Hahn^{1*}

¹Department of Medicine and Microbiology, University of Pennsylvania, Philadelphia, United States; ²Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, United Kingdom; ³Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia, ⁴Department of Medical Biology, The University of Melbourne, Melbourne, Australia, ⁵Robert Koch Institute, Berlin, Germany; ⁶Sanaga-Yong Chimpanzee Rescue Center, IDA-Africa, Portland, United States; ⁷TransVIHMI, Institut de Recherche pour le Développement, University of Montpellier, INSERM, Montpellier, France; ⁸Wellcome Trust Sanger Institute, Cambridge, United Kingdom

#contributed equally

*co-senior authors

Accepted for publication in *PNAS*, with modifications

4.1 Abstract

Wild-living African apes are endemically infected with parasites that are closely related to human *Plasmodium vivax*, a leading cause of malaria outside of Africa. This finding suggests that the origin of *P. vivax* was in Africa, even though the parasite is now rare in humans there. To elucidate the emergence of human *P. vivax* and its relationship to the ape parasites, we analyzed genome sequence data of *P. vivax* strains infecting six chimpanzees and one gorilla from Cameroon, Gabon and Côte d'Ivoire. We found that ape and human parasites share near-identical core genomes, differing by only 2% of coding sequences. However, compared with the ape parasites, human strains of *P. vivax* exhibit about 10-fold less diversity, and have a relative excess of nonsynonymous nucleotide polymorphisms, with site frequency spectra suggesting they are subject to greatly relaxed purifying selection. These data suggest that human *P. vivax* has undergone an extreme bottleneck, followed by rapid population expansion. Investigating potential host specificity determinants, we found that ape *P. vivax* parasites encode intact orthologs of three reticulocyte binding protein genes (*rbp2d*, *rbp2e*, *rbp3*), which are pseudogenes in all human *P. vivax* strains. However, binding studies of recombinant RBP2e and RBP3 proteins to human, chimpanzee and gorilla erythrocytes revealed no evidence of host-specific barriers of red blood cell invasion. These data suggest that, from an ancient stock of *P. vivax* parasites capable of infecting both humans and apes, a severely bottlenecked lineage emerged out of Africa and underwent rapid population growth as it spread globally.

4.2 Introduction

Plasmodium vivax causes over 8 million cases of human malaria per year, with the vast majority occurring in Southeast Asia and South America (World Health Organisation,

2017). *P. vivax* is rare in humans in Africa due to the high prevalence of the Duffy negative mutation (Howes et al., 2011), which abrogates expression of the Duffy antigen receptor for chemokines (DARC) on erythrocytes. Since DARC serves as a receptor for *P. vivax*, its absence protects Duffy negative humans from *P. vivax* infection (Miller et al., 1976), although this protection is not absolute (Zimmerman, 2017). Until recently, *P. vivax* was thought to have emerged in Asia following the cross-species transmission of a macaque parasite (Cornejo and Escalante, 2006; Mu et al., 2005). However, the finding of closely related parasites in wild-living chimpanzees and gorillas suggested an African origin of *P. vivax* (Liu et al., 2014). Indeed, parasite sequences closely resembling *P. vivax* have been detected in western (*Pan troglodytes verus*), central (*P. t. troglodytes*) and eastern (*P. t. schweinfurthii*) chimpanzees, eastern (*Gorilla beringei graueri*) and western (*Gorilla gorilla gorilla*) lowland gorillas, and most recently in bonobos (*Pan paniscus*) (Kaiser et al., 2010; Liu et al., 2017, 2014, 2010; Prugnolle et al., 2013). Phylogenetic analyses of available sequences revealed that ape and human parasites were near-identical, with human *P. vivax* sequences forming a monophyletic lineage that usually fell within the radiation of the ape parasites (Liu et al., 2014). These findings suggested that *P. vivax* infected apes including humans in Africa, until the spread of the Duffy negative mutation largely eliminated the parasite in humans there. However, definitive conclusions could not be drawn since all analyses of ape *P. vivax* genomes to date have rested on a small number of gene fragments amplified almost exclusively from parasite mitochondrial DNA present in ape fecal samples.

Understanding the origin of human *P. vivax* and its relationship to the ape parasites is important for several reasons. First, only six *Plasmodium* species, out of several hundred so far described to infect vertebrate hosts (Faust and Dobson, 2015), have successfully colonized humans (Scully et al., 2017). Thus, the circumstances that

surround the emergence of each of these human pathogens is of interest, especially since most, if not all, have non-human primate parasites as their closest relatives (Liu et al., 2014, 2010; Rutledge et al., 2017). Second, it is currently unclear whether the ape parasites represent a separate species, distinct from *P. vivax*. Although sequences from human *P. vivax* parasites form a monophyletic clade in phylogenetic trees, this may be a reflection of their geographic separation and not the existence of host-specific infection barriers. Indeed, ape *P. vivax* has been shown to cause malaria in a Duffy positive European traveler (Prugnolle et al., 2013), and human *P. vivax* has been transmitted to wild-living monkeys in South America, generating what has been called *Plasmodium simium* (Buery et al., 2017). If cross-species infection and recombination of ape and human *P. vivax* were possible, as appears to be the case for *P. simium* and *P. vivax* in South America (Brasil et al., 2017; Buery et al., 2017), this could have implications for malaria eradication efforts. Finally, ape and human *P. vivax* strains may have acquired adaptations that limit parasite transmission between different host species. Such findings could explain why the macaque parasites *Plasmodium knowlesi* and *Plasmodium cynomolgi* can infect and cause malaria in humans, but do not appear to be commonly transmitted between different human hosts (Brock et al., 2016).

To elucidate the events that led to the emergence of human *P. vivax*, we sought to obtain genome sequences of parasites infecting chimpanzees and gorillas. A similar approach has recently uncovered processes that may have allowed the gorilla precursor of *P. falciparum* to cross the species barrier to infect humans (Otto et al., 2018; Sundararaman et al., 2016). However, obtaining blood samples from *Plasmodium* infected apes is challenging due to the endangered species status of these hosts. Moreover, ape *P. vivax*, like its human-infecting counterpart, exhibits only low levels of parasitemia (Liu et al., 2014) and has not been cultured. Although removal of host leukocytes from whole

blood samples (Pearson et al., 2016) and parasite nucleic acid capture (Hupalo et al., 2016) have improved the recovery of human *P. vivax* genomes, these approaches are not readily applicable to ape parasites. We thus adapted a previously developed selective whole genome amplification (SWGA) method (Sundararaman et al., 2016) to generate *P. vivax* genome sequences from unprocessed chimpanzee and gorilla blood samples obtained in different parts of Africa. Analysis of these genomes revealed that the ape parasites are about 10 times more diverse than global representatives of human *P. vivax* (Hupalo et al., 2016), indicating that the human parasite has undergone a severe genetic bottleneck. Ape *P. vivax* genomes were found to have intact orthologs of three reticulocyte binding protein genes that are pseudogenized in all human *P. vivax* strains, but functional studies of two of the encoded proteins revealed no evidence of species-specific receptor interactions. The *P. vivax* ancestor therefore likely infected both humans and apes in Africa, before being eliminated in humans there by the spread of the Duffy negative mutation. Thus, the current human-infecting parasites represent a lineage that had escaped out of Africa.

4.3 Results

Genome assemblies of chimpanzee *P. vivax*. Leftover blood samples from routine health examinations of chimpanzees cared for at the Sanaga Yong (SY) Chimpanzee Rescue Center in Cameroon were screened for *Plasmodium* infection using nested PCR with pan-*Plasmodium* and *P. vivax*-specific primers. Two samples, SY56 and SY43, were positive for ape *P. vivax*, with limiting dilution PCR detecting one strain in SY56 and up to five variants in SY43, with two strains predominating. Since these two samples lacked other *Plasmodium* species, they were suitable for selective whole genome amplification (SWGA) without the risk of generating inter-species recombinants (Leichty and Brisson,

2014; Sundararaman et al., 2016). SWGA uses the highly processive phi29 DNA polymerase and specific primers to preferentially amplify pathogen sequences from complex mixtures of target and host DNA, and has been used successfully in the past to generate *Plasmodium* sequences from blood smear negative, unprocessed blood samples (Sundararaman et al., 2016). Since SWGA can result in stochastic amplification when target templates are rare (Cowell et al., 2017; Sundararaman et al., 2016), each sample was amplified on more than one occasion using different primer sets, with and without digestion of sample DNA with methylation-dependent restriction enzymes to degrade host DNA (see SI Appendix, Table S4-1). Individual SWGA reactions were pooled and sequenced on Illumina and PacBio platforms (see SI Appendix).

Draft genomes of the chimpanzee *P. vivax* strains PvSY56 and PvSY43 were generated using iterative reference-guided assembly to the human PvP01 reference genome (Auburn et al., 2016), followed by gap filling steps. In addition, PvSY56 reads that did not map to the PvP01 core genome were *de novo* assembled to obtain subtelomeric contigs. The resulting assemblies yielded 21.9 Mbp and 21.2 Mbp of sequence for PvSY56 and PvSY43, respectively (Table 4-1). Because sample SY43 contained at least five *P. vivax* strains (Liu et al., 2014), the PvSY43 genome represents a consensus of these variants. Annotations were transferred from PvP01, with additional genes predicted in the *de novo* contigs. Since a large number of genes contained frameshifts in homopolymer tracts, we manually corrected annotations spanning these presumed sequencing errors to maintain an open reading frame. Overall, PvSY56 and PvSY43 shared a highly conserved core genome with human *P. vivax*. More than 98% of PvP01 core genes (as defined in 19) were identified in each chimpanzee *P. vivax* assembly (Table 4-1), with 88% present as full-length genes. Although 10 human *P. vivax* core genes were absent from both PvSY56 and PvSY43, ape *P. vivax* reads at least partially covered these coding regions,

implicating assembly difficulties rather than differences between ape and human *P. vivax* genomes for their absence. Assembly issues likely also account for the small number of genes in subtelomeric and internal hypervariable regions that could be annotated for PvSY56 and PvSY43, respectively (Table 4-1).

Table 4-1. Genome features of ape *P. vivax*

	PvSY56	PvSY43*	PvP01 [†]
Host species	Chimpanzee	Chimpanzee	Human
Country	Cameroon	Cameroon	Indonesia
Chromosomal assembly (bp) [§]	21,928,114	21,224,756	24,177,188
Mean depth of coverage [¶]	319	240	N/A
Chromosomal contigs	7,112	6,604	14
G + C content (%)	44.1	44.6	43.3
Core [#] protein-coding genes (% of PvP01)	4,883 (98.8)	4,908 (99.3)	4,941 (100)
Full length [#] (% of PvP01)	4,391 (88.9)	4,350 (88.0)	N/A
Partial (% of PvP01)	492 (10.0)	558 (11.3)	N/A
Genes in hypervariable regions	415	276	1,702

*The genome assembly of PvSY43 represents a consensus sequence of at least two major and three minor chimpanzee *P. vivax* variants (Fig. 2C and D; see SI Appendix, Fig. S5).

[†]Chimpanzee *P. vivax* genomes were compared to the human *P. vivax* reference PvP01 (Auburn et al., 2016).

[§]Number of unambiguous bases.

[¶]Calculated by dividing the number of nucleotides in reads mapped to the assemblies by the expected genome size in PvP01.

[#]Genes classified as full-length comprised at least 90% of the length of the corresponding PvP01 ortholog.

^{||}Protein-coding genes in subtelomeric and internal hypervariable regions were defined as described (Pearson et al., 2016).

Polymorphism in ape and human *P. vivax*. Comparison of coding sequences between the PvSY56 and PvSY43 assemblies revealed that they differ at 0.61% of sites, which contrasts with a difference of only 0.11% between the two human *P. vivax* reference genomes, PvSall and PvP01 (see SI Appendix, Figure S4-1A). Since PvSY56 and PvSY43 were both derived from chimpanzees housed at the same sanctuary, we reasoned that they might not represent the full extent of ape *P. vivax* diversity (the two human reference strains were sampled on two different continents in Latin America and

Southeast Asia). This prompted us to obtain *P. vivax* genome sequences from additional infected apes. Using SWGA followed by Illumina sequencing, we amplified ape *P. vivax* from blood samples of two additional SY chimpanzees (SY81 and SY90), from a wild-living western chimpanzee (Sagu) from Cote d'Ivoire (Kaiser et al., 2010), and from a western lowland gorilla (Gor3157) sampled in Cameroon (see SI Appendix, Table S4-2). We also mined the read database from a blood sample of a *P. malariae*-infected sanctuary chimpanzee from Gabon (Rutledge et al., 2017), which we had noted contained a substantial number of ape *P. vivax* reads. Reads from each sample were mapped to the PvSY56 assembly and single nucleotide polymorphisms (SNPs) were identified. The extent of genome coverage varied considerably among the six chimpanzee samples; however, we were able to recover between 695 and 3,006 core genes (see SI Appendix, Table S4-2), with 65% of genes analyzed being covered in four or more parasite genomes (see SI Appendix, Figure S4-2). The gorilla *P. vivax* strain, which was derived from a partially degraded bushmeat sample, yielded only 10 genes despite repeated amplification, and was thus not included in the diversity analysis. For comparison, we included sequences from seven additional human *P. vivax* strains (Hupalo et al., 2016), each from a different country (India, Myanmar, Papua New Guinea, Thailand, Colombia, Mexico, Peru), and identified SNPs using the same methods. Diversity values were then calculated across a common set of 4,262 core genes for which we obtained sequences from two or more strains for both ape and human *P. vivax* (Table 4-2). The results revealed a mean pairwise nucleotide sequence diversity (π) among the six chimpanzee *P. vivax* strains of 0.694%, about eight times higher than the value (0.085%) for the global sample of human strains (Table 4-2, Figure 4-1A; see SI Appendix, Figure S4-1B). Removal of the PvSY43 sequence did not decrease this difference (see SI Appendix, Figure S4-1C), indicating that inclusion of this multiply infected sample did not artificially inflate the

diversity of the chimpanzee *P. vivax* parasites. Similarly, there was no evidence that SWGA increased chimpanzee *P. vivax* diversity. Analysis of transition/transversion ratios at 4-fold degenerate sites yielded near identical results for chimpanzee (1.08) and human (1.06) *P. vivax*, excluding a technical artifact (Cowell et al., 2017; Sundararaman et al., 2016). Thus, the *P. vivax* strains currently circulating in humans appear to have undergone a stringent population bottleneck in the recent past.

Table 4-2. Nucleotide polymorphism in ape and human *P. vivax*

	n*	$\pi_{\text{all}}^{\dagger}$	$\pi_0^{\ddagger}$	$\pi_4^{\S}$	NS polymorphisms [¶]	S polymorphisms [¶]	NS/S
Ape <i>P. vivax</i>	6	0.00694	0.00354	0.01599	31,880	47,020	0.68
Human <i>P. vivax</i>	9	0.00085	0.00061	0.00143	10,479	7,657	1.37

*n = number of strains included in the analysis (see SI Appendix, Fig. S2 for gene coverage among the different strains).

[†]Mean pairwise diversity at coding sites from 4,262 genes (6.5 million sites)

[‡]Mean pairwise diversity at 0-fold degenerate sites from 4,262 genes (4.0 million sites)

[§]Mean pairwise diversity at 4-fold degenerate sites from 4,262 genes (0.7 million sites)

[¶]Number of nonsynonymous (NS) and synonymous (S) polymorphisms were calculated by counting the number of SNPs that changed (NS) or did not change (S) the protein sequence of the respective reference.

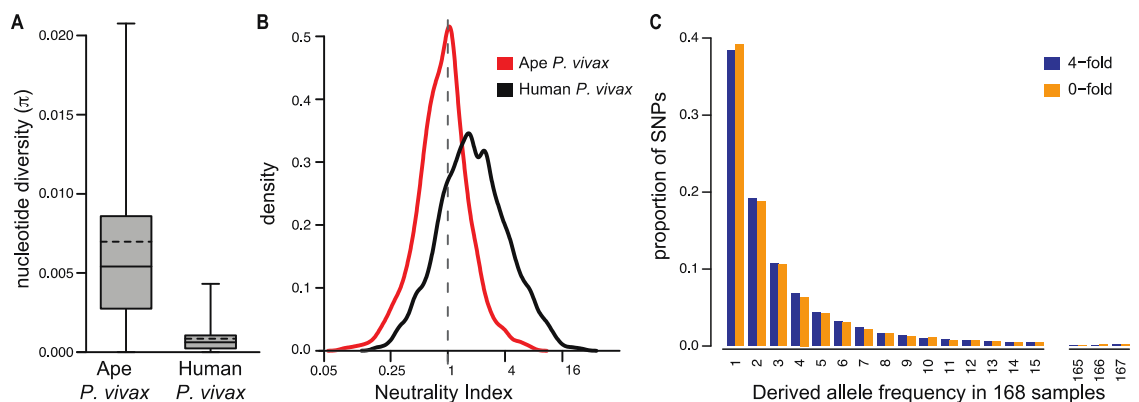


Figure 4-1. Nucleotide sequence diversity in ape and human *P. vivax*. (A) Pairwise nucleotide sequence diversity (π) calculated across a common set of 4,259 core genes for six chimpanzee and nine human *P. vivax* strains (as in Table 4-2, but three genes with fewer than 35 aligned sites were excluded). Median and mean (weighted by gene length) π values are indicated by solid and dashed lines, with box and whiskers indicating the interquartile range and 99th percentiles, respectively. Plots including outliers are shown in the SI Appendix, Fig. S1B. (B) Density plots of neutrality index (NI) values shown on a log₂ scale for ape (red) and human (black) *P. vivax* genes. Values are shown for 1,582 genes with non-zero values of NI in both populations. (C) Site frequency spectra of

polymorphisms at 4-fold degenerate (blue) and 0-fold degenerate (orange) sites extracted from SNP data of human *P. vivax* samples from Southeast Asia (20).

The nature of the nucleotide polymorphisms also differed substantially between chimpanzee and human *P. vivax* strains. Among the chimpanzee parasites the majority of polymorphisms were synonymous, with the ratio of nonsynonymous to synonymous SNPs (NS/S) being 0.68. In contrast, the majority of polymorphisms among the human strains were nonsynonymous, with a NS/S ratio of 1.37 (Table 4-2). The ratio of NS/S among polymorphisms can be compared to (i.e., divided by) the NS/S ratio among interspecies differences to yield the neutrality index (Rand and Kann, 1996). This neutrality index, or NI, assumes that synonymous changes, both within and between species, are selectively neutral, and has an expected value of one when nonsynonymous changes are also neutral. We thus compiled a set of 3,912 genes that were comparable among ape and human *P. vivax* strains as well as between these parasites and the macaque parasite *P. cynomolgi* (their closest relative with an available genome sequence). The overall NI value for ape *P. vivax* strains was close to the neutral expectation (NI = 0.95; see SI Appendix, Table S4-3). In contrast, the overall value for human *P. vivax* strains was much larger (NI = 1.91), indicating a large excess of nonsynonymous polymorphisms among human strains, relative to the expectation derived from patterns of divergence between species.

Comparison of the ratios of nonsynonymous and synonymous changes, between within-species polymorphisms and between-species fixed differences, forms the basis of the McDonald-Kreitman (MK) test for adaptive evolution of individual genes (McDonald and Kreitman, 1991). The discrepancy in overall NI values for ape versus human *P. vivax* strains indicates that MK tests are likely to produce different results depending on which *P. vivax* strains are used. For the chimpanzee *P. vivax* strains, only two genes yielded significant results after correction for multiple testing, both with NI<1 indicating a significant

excess of fixed, potentially adaptive, nonsynonymous differences: one was found to be orthologous to PVP01_1201800, which is an immunogenic member of the tryptophan-rich antigen family of *P. vivax* (Chuquiyauri et al., 2015; Wang et al., 2015), while the other (orthologous to PVP01_1406200) encodes a conserved *Plasmodium* protein of unknown function. For the human *P. vivax* strains, five genes yielded significant MK test results, but all with $NI > 1$, indicating an excess of nonsynonymous polymorphisms (see SI Appendix, SI Dataset). Such results are usually interpreted as evidence of balancing selection maintaining polymorphism, but the large overall NI value for the human strains suggests that there is a more pervasive factor, such as past demography, that is influencing these genes. Figure 4-1B shows the distributions of NI values for 1,582 individual genes with non-zero values in both ape and human *P. vivax*. These results indicate that the difference between ape and human *P. vivax* is not due to a subset of unusual genes, but rather that the entire distribution is shifted from being centered around 1.0 in ape *P. vivax*, to being centered around 2.0 in human *P. vivax* (Figure 4-1B; see SI Appendix, SI Dataset).

To investigate this in greater detail, we examined the frequencies of nonsynonymous and synonymous polymorphisms per site. We obtained different results depending on the methodology used, primarily because different methods use different approaches to estimate the numbers of sites available for synonymous changes. To avoid this problem, we included only 4-fold and 0-fold degenerate sites where, due to the structure of the genetic code, either all or none of the possible changes are synonymous. Among ape *P. vivax* strains, the nucleotide diversity at 0-fold degenerate sites (0.00354) was 22% of that at 4-fold degenerate sites (0.01599), whereas among human *P. vivax* strains, the value at 0-fold degenerate sites (0.0061) was 43% of that at 4-fold degenerate sites (0.00143) (Table 4-2; see SI Appendix, Figure S4-1D and E). Thus, while chimpanzee parasites were 11 times more diverse than human parasites at 4-fold

degenerate sites, they were only 6 times more diverse at 0-fold degenerate sites, indicating that nonsynonymous polymorphisms in human *P. vivax* strains are almost twice as numerous as expected.

A large number of human *P. vivax* genome sequences have been characterized (Hupalo et al., 2016; Pearson et al., 2016), and so it is possible to investigate the frequencies at which SNPs are segregating within the population. Synonymous polymorphisms are expected to be neutral, and so their site frequency spectrum (SFS) should reflect past demography. By comparison, many nonsynonymous polymorphisms are expected to be slightly deleterious, and their SFS should thus be more skewed towards lower frequencies. To examine a large number of parasite sequences from a single geographic region, we focused on SNP data from MalariaGen samples from Southeast Asia (Pearson et al., 2016). Ancestral and derived alleles at each site were identified by comparison with two outgroups: the chimpanzee *P. vivax* strain PvSY56 and a *P. cynomolgi* reference strain. The unfolded site frequency spectra obtained for SNPs at 0-fold and 4-fold degenerate sites are almost identical (Figure 4-1C). Thus, the unusually large fraction of nonsynonymous polymorphisms among human *P. vivax* sequences appears to reflect mutations that are segregating as effectively neutral alleles.

Relationship of *P. vivax* strains from humans and apes. In previous analyses of a small number of partial gene sequences, we found that sequences from human *P. vivax* parasites always formed a monophyletic clade, which usually fell within the radiation of sequences from chimpanzee and gorilla samples (Liu et al., 2014). This was observed for mitochondrial and apicoplast sequences as well as for three nuclear genes, while for a fourth nuclear gene ape and human *P. vivax* sequences formed sister clades (Liu et al., 2014). Here we found that, across their genome, chimpanzee parasites were much more

divergent from the human parasites than they were from each other. For example, across 3,957 core genes, the chimpanzee parasite genomes (PvSY56 and PvSY43) differed from one another at 0.6% of sites, but from the human *P. vivax* reference genomes (PvSall and PvP01) on average at 2.2% of sites. This relationship is summarized in a neighbor-joining tree constructed from a matrix of pairwise genetic distances from an alignment of 232 nuclear genes available for all six chimpanzee parasites (Figure 4-2A). Although this tree may not reflect the true evolutionary history of any one particular gene (due to recombination), the overall relationships were confirmed by phylogenetic network analysis (see SI Appendix, Figure S4-3A), which showed that the four chimpanzee parasites sampled at the same location (SY) in Cameroon are on average a little more closely related to each other than they are to the strains identified in Gabon (GA02) and Cote d'Ivoire (Sagu). Inclusion of *P. vivax* sequences from the gorilla sample restricted the analysis to six genes and only five chimpanzee parasites. For these genes the gorilla *P. vivax* strain was quite divergent (Figure 4-2B, see SI Appendix Figure S4-3B), differing almost as much from the chimpanzee strains (on average 1.8%) as from the human strains (on average 2.4%). Whether the human *P. vivax* lineage falls within the radiation of the ape strains, or groups as a sister clade, depends on the position of the root of these trees. The closest available outgroup is *P. cynomolgi*, which is much more distant from the *P. vivax* sequences than they are from each other, and may not root the tree reliably.

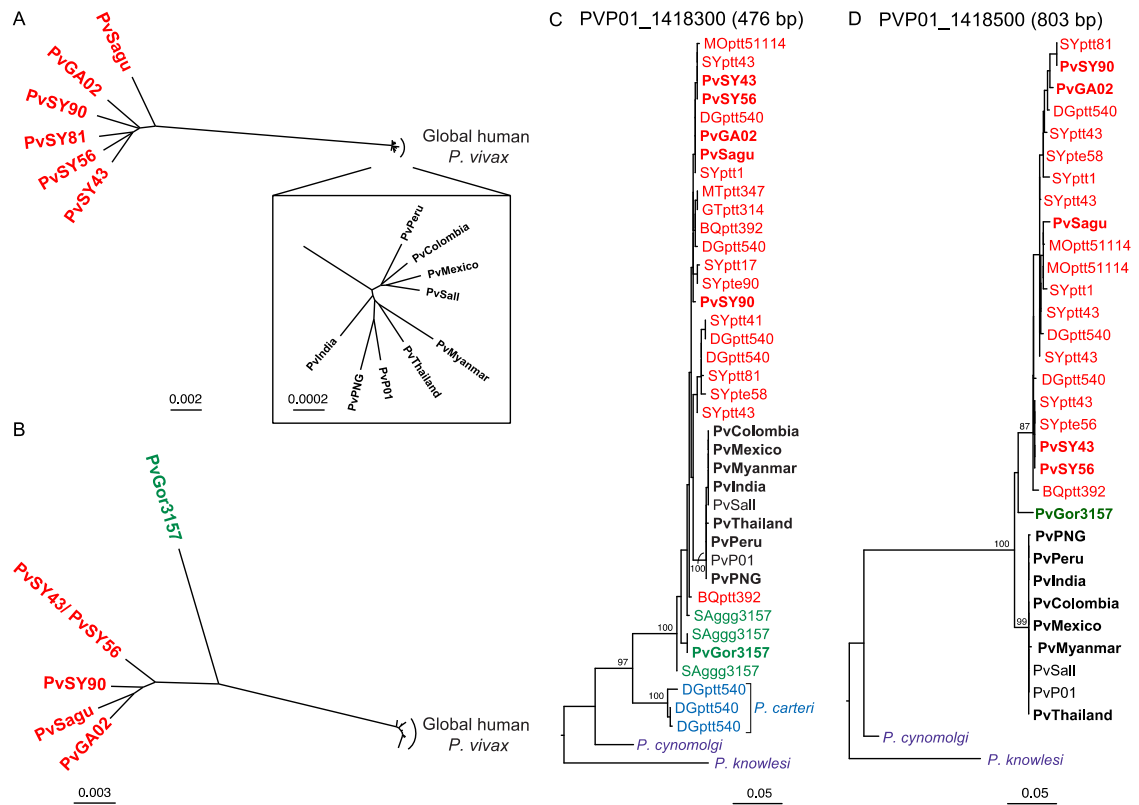


Figure 4-2. Evolutionary relationships of ape and human *P. vivax* strains (A) An unrooted neighbor-joining tree constructed from a matrix of pairwise genetic distances from an alignment of 232 nuclear genes is shown for nine human (black) and six chimpanzee (red) *P. vivax* strains (the inset shows the human *P. vivax* strains in greater detail). (B) As in (A), but based on 6 nuclear genes with coverage in one gorilla *P. vivax* strain (green). The same human and chimpanzee *P. vivax* strains were included, except for PvSY81, which did not cover these genes. (C and D) Maximum likelihood trees for fragments of nuclear genes PVP01_1418300 (C) and PVP01_1418500 (D) with *P. cynomolgi* and *P. knowlesi* included as outgroups. Sequences of *P. carteri* parasites are shown in blue. ‘Pv’ denotes sequences from genome-wide analyses, shown in bold if generated by SWGA or derived from published data (see SI Appendix, Table S4-2); all other sequences except for *P. cynomolgi* and *P. knowlesi* were generated by SGA and include a code identifying their geographic origin (see SI Appendix, Fig. S5), ape subspecies (*P.t.t.*, *P. t. troglodytes*; *P.t.e.*, *P. t. ellioti*; *G.g.g.*, *Gorilla gorilla gorilla*), and sample number. Bootstrap values ≥ 70 are shown for clades with two or more non-identical tips. Fragment lengths in PVP01 are indicated above the trees. The scale bars indicate substitutions per site (also see SI Appendix, Figure S4-3 for phylogenetic network analysis).

To investigate further the relationships among ape and human *P. vivax* strains, we focused on the 10 genes that could be recovered from the single gorilla sample (see SI

Appendix, Table S2). Including both *P. cynomolgi* and *P. knowlesi* as outgroups, we found that four genes yielded a tree topology in which the human strains fell within the radiation of ape strains, while the six other genes yielded tree topologies in which human and ape parasites represented sister clades (see SI Appendix, Figure S4-4). To increase the number of geographically diverse ape *P. vivax* sequences, we used single genome amplification (SGA) to screen an existing collection of *P. vivax* positive ape blood and fecal samples for five of these genes (see SI Appendix, Figure S4-5). Each DNA preparation was diluted such that only single DNA templates were amplified, which precludes *in vitro* recombination. For three of these genes, this analysis also yielded sequences from *Plasmodium carteri*, a rare parasite species thus far only found in two wild chimpanzees, which is distinct from, but most closely related to, the *P. vivax* clade (Liu et al., 2014; Loy et al., 2017). When these additional sequences were included in the phylogenetic analyses, four of the five trees showed the human strains within the radiation of ape strains, including two where the previous topology depicted ape and human strains as sister clades, with three of these four trees including gene sequences for *P. carteri* (Figure 4-2C; see SI Appendix, Figure S4-6). Thus, inclusion of gorilla *P. vivax* and/or *P. carteri* changed the topology from sister clades to a nested relationship. For the remaining gene, human and ape parasite sequences remained as sister clades, but only a single gorilla parasite sequence was available for analysis and *P. carteri* sequences could not be amplified (Figure 4-2D). These results indicate that the inferred relationships among *P. vivax* strains from apes and humans depend in large part on the number of available sequences, especially from gorilla parasites, as well as on the presence of a closely related outgroup.

Ape *P. vivax* strains maintain open reading frames for reticulocyte binding protein genes that are pseudogenized in human *P. vivax*. Adaptation of *Plasmodium* parasites to new host species has been associated with gains and losses of genes encoding proteins involved in red blood cell invasion (Scully et al., 2017). We therefore compared the repertoire of *P. vivax* invasion genes in the genomes of human and chimpanzee parasites. Like human *P. vivax*, the chimpanzee parasite genomes contained genes encoding the Duffy binding protein (DBP) and a related erythrocyte binding protein (DBP2, or EBP) (Hester et al., 2013), but no additional DBP-like genes were identified.

Variation in the complement of reticulocyte binding proteins (RBPs) is thought to influence the ability of *Plasmodium* parasites to invade erythrocytes (Cowman et al., 2017; Scully et al., 2017). Human *P. vivax*, which exclusively invades reticulocytes, has full-length open reading frames for five *rbp* genes (*rbp1a*, *rbp1b*, *rbp2a*, *rbp2b*, *rbp2c*), all of which were conserved in the two chimpanzee *P. vivax* genomes PvSY56 and PvSY43. Two shorter *rbp* genes annotated in human *P. vivax* (*rbp2p1* and *rbp2p2*) are believed to encode proteins that lack a C-terminal transmembrane domain; one of these (*rbp2p1*) appears to be present in all human *P. vivax* strains and also in *P. cynomolgi*, while *rbp2p2* has been found only in a subset of human *P. vivax* strains (Hietanen et al., 2016). We identified orthologs of both of these partial genes in the chimpanzee *P. vivax* genomes (Figure 4-3A), indicating that variation in the presence of *rbp2p2* among human *P. vivax* strains is the result of a deletion after the divergence of human and ape parasites, rather than a recent gene duplication. The finding of partial *rbp2p1* and *rbp2p2* genes in both ape and human *P. vivax*, and of *rbp2p1* in *P. cynomolgi*, suggests that their encoded proteins have a conserved function. However, looking for positive selection (Yang, 2007) in invasion genes (*dbp*, *ebp*, *rbp1a*, *rbp1b*, *rbp2a*, *rbp2b*) on the branch leading to human *P. vivax* failed to yield evidence of human specific adaptation.

The human *P. vivax* genome also contains three *rbp* pseudogenes, termed *rbp2d*, *rbp2e* and *rbp3*. Seemingly functional orthologs of *rbp2e* and *rbp3* are present in the genomes of the monkey parasites *P. cynomolgi*, *P. knowlesi* and *P. inui*, while *P. fragile* has an intact *rbp2e* gene but a *rbp3* pseudogene (Figure 4-3A). So far, *rbp2d* has only been identified in *P. vivax*. The two chimpanzee *P. vivax* genomes PvSY56 and PvSY43 contained full length intact open reading frames corresponding to each of these three human *P. vivax* pseudogenes, indicating that the loss of function occurred after the divergence of human and ape parasites (Figure 4-3A).

Since the pseudogenization of *rbp2d*, *rbp2e*, and *rbp3* seems to be unique to the human lineage of *P. vivax*, we considered the possibility that these genes may be intact in some human strains. We mapped sequencing reads from 374 published human *P. vivax* strains (Hupalo et al., 2016; Pearson et al., 2016) to the *rbp2d*, *rbp2e* and *rbp3* reference genes and analyzed those that yielded greater than a three-fold read coverage of the entire coding sequences. In each gene, we found at least one inactivating mutation that was present in all human parasite samples, as well as numerous additional mutations that likely occurred subsequent to the initial pseudogenization event (Figure 4-3B). The accumulation of additional frameshifts and stop mutations, some of which occur very close to the 5' end of the coding sequence, suggests that these genes do not encode truncated proteins.

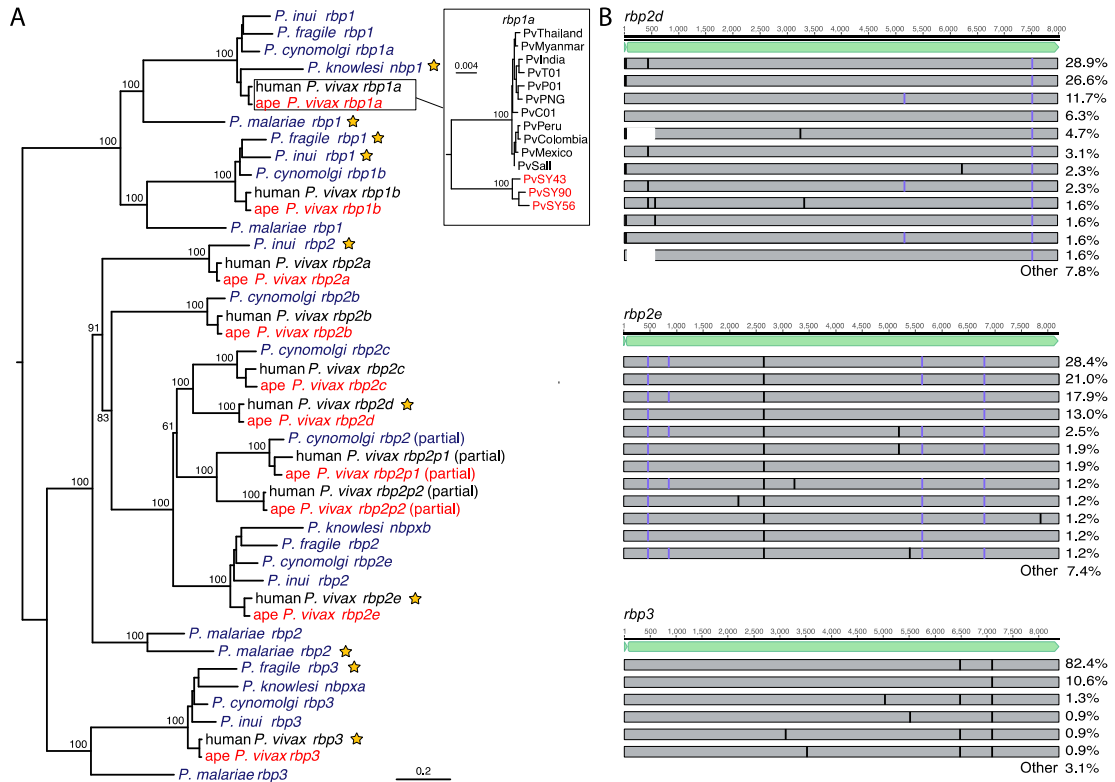


Figure 4-3. The *rbp* gene family in ape and human *P. vivax*. (A) A midpoint-rooted maximum likelihood phylogenetic tree is shown depicting the relationships of human (black) and chimpanzee (PvSY56 and PvSY43, red) *P. vivax rbp* genes with their orthologs in *P. knowlesi*, *P. cynomolgi*, *P. inui*, *P. fragile* and human *P. malariae* (purple). *P. vivax*, *P. cynomolgi* and *P. knowlesi* genes are labeled according to their published names; genes from *P. inui*, *P. fragile* and *P. malariae* are labeled according to the clade in which they are placed. Pseudogenes are indicated with yellow stars. The inset shows the relationship of *rbp1a* sequences among representative human and three sequenced chimpanzee *P. vivax* strains, rooted using *P. cynomolgi* (see SI Appendix for details). (B) Locations of frameshift (purple) and premature stop (black) mutations in *rbp2d*, *rbp2e*, and *rbp3* sequences assembled from published human *P. vivax* strains (20, 21), relative to the full length coding sequence from chimpanzee *P. vivax* (light green). Each bar represents a set of mutations that occurred in two or more human *P. vivax* strains for which a full length sequence was assembled (128, 162, and 227 sequences for *rbp2d*, *rbp2e*, *rbp3*, respectively); the percentage of sequences containing the respective mutations is shown on the right, with "other" summarizing all mutations that occurred only once.

To examine whether the other chimpanzee *P. vivax* strains (PvSY81, PvSY90, PvSagu and PvGA02) contained any of the *rbp2d*, *rbp2e*, and *rbp3* inactivating mutations, we mapped available sequencing reads to their respective genes. We also used SGA to

amplify the same regions from *P. vivax* positive gorilla samples. Although in most instances the coverage of the *rbp2d*, *rbp2e*, and *rbp3* genes was incomplete, none of the recovered sequences contained the frameshift and stop codon mutations found in all human *P. vivax* strains (see SI Appendix, Figure S4-7). This was also true for SGA-derived gorilla parasite sequences from the multiply infected bushmeat sample SAgg3157, which contained a number of polymorphism, none of which disrupted the respective reading frame. Thus, both chimpanzee and gorilla *P. vivax* parasites appear to maintain three genes encoding RBPs that have been lost in all human *P. vivax* strains, which could influence their ability to infect different host species.

Recombinant RBP2e and RBP3 do not exhibit species-specific red blood cell binding. The pseudogenization of *rbp2d*, *rbp2e* and *rbp3* in all human *P. vivax* strains raised the possibility that these proteins bind gorilla- and/or chimpanzee-specific erythrocyte receptors that are no longer used by the human parasite. Recombinant proteins comprising the amino-terminal domain of RBPs encoded by human *P. vivax* (RBP2a₁₆₀₋₁₀₀₀ and RBP2b₁₆₁₋₉₆₉) have been used to characterize their erythrocyte binding properties (Gruszczyk et al., 2018b, 2016). These studies showed that RBP2b binds the reticulocyte-specific Transferrin Receptor 1 (TfR1), also termed CD71 (Gruszczyk et al., 2018b), while RBP2a binds an unknown receptor present on both normocytes and reticulocytes (Gruszczyk et al., 2016). To examine the function of chimpanzee *P. vivax* RBP2d, RBP2e, and RBP3 proteins, we expressed their amino-terminal domains in bacteria for subsequent erythrocyte binding studies. Although RBP2d₁₆₅₋₉₆₇ could not be purified due to protein aggregation, RBP2e₁₅₆₋₉₅₇ and RBP3₁₄₉₋₉₆₈ were efficiently expressed and exhibited an alpha-helical and beta-sheet content similar to human *P. vivax* RBP2b (see SI Appendix, Figure S4-8). Because some RBPs bind only reticulocytes, we

s to enrich these cells from blood samples obtained from four humans, four chimpanzees and one gorilla using a Percoll density gradient as previously described (Gruszczyk et al., 2018b, 2016). Despite repeated attempts, this approach yielded only partial reticulocyte enrichment for the ape blood samples, possibly due to differences in erythrocyte density between the different species (see SI Appendix, SI Methods). Nonetheless, some enrichment of ape reticulocytes (up to 1.8%) was achieved as determined by thiazole orange (TO) staining.

To examine binding to ape red blood cells, we first tested the two previously characterized human *P. vivax* RBP proteins, RBP2a₁₆₀₋₁₀₀₀ and RBP2b₁₆₁₋₉₆₉ (Gruszczyk et al., 2018b, 2016). Ape and human red blood cells were incubated with each recombinant protein and binding was assessed using protein-specific polyclonal rabbit antibodies, followed by a fluorophore-labeled anti-rabbit antibody (Gruszczyk et al., 2018b, 2016). Reticulocytes were stained with thiazole orange (TO) prior to flow cytometry (Figure 4-4A). Consistent with previous results, we observed robust binding of RBP2a to both human normocytes (10.3% of TO negative cells; Figure 4-4B) and reticulocytes (32.4% of TO positive cells; Figure 4-4C). Interestingly, a similar binding profile was also observed for gorilla and chimpanzee red blood cells (Figure 4-4; see SI Appendix, Figure S4-9). As expected, RBP2b exhibited a strong preference for reticulocytes, binding 19.9% of human reticulocytes (Figure 4-4C), but only a minor fraction (0.9%) of human normocytes (Figure 4-4B), likely reflecting incomplete reticulocyte staining and/or non-specific binding. RBP2b also bound chimpanzee and gorilla reticulocytes, albeit at a reduced level (Figure 4-4C). Although the TfR1 proteins of chimpanzees and gorillas differ from their human counterpart by a few amino acids (see SI Appendix, Figure S4-10), none of these residues were identified to represent RBP2b contact sites (Gruszczyk et al., 2018a). Thus, the decreased binding of RBP2b to ape reticulocytes is unlikely the result

of sequence differences between chimpanzee, gorilla and human TfR1 proteins and may instead reflect differences in TfR1 expression levels, post-translational modifications, or other factors.

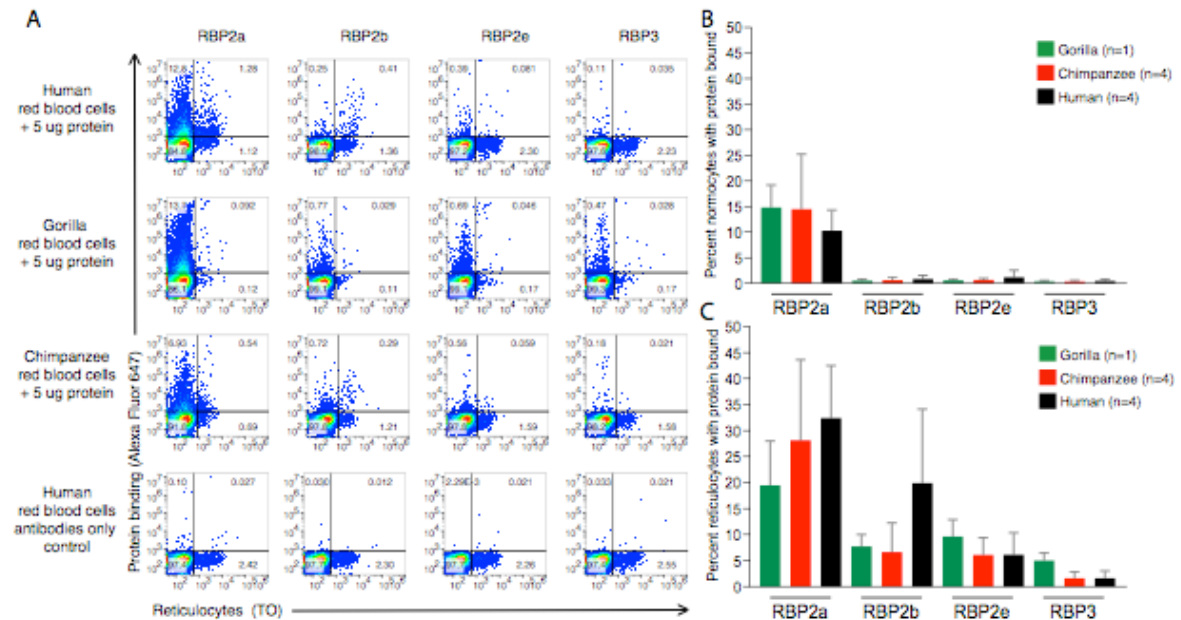


Figure 4-4. Binding of RBPs to ape and human red blood cells. (A) Dot plots are shown that depict the binding of human *P. vivax* RBP2a and RBP2b proteins and chimpanzee *P. vivax* RBP2e and RBP3 proteins to human (first row), gorilla (second row), and chimpanzee (third row) red blood cells, respectively, along with antibody-only controls of human red blood cells (fourth row). RBP binding was detected using an RBP-specific polyclonal rabbit antibody and an anti-rabbit (Alexa Fluor 647 labelled) secondary antibody (y-axis), and reticulocytes were identified by staining with thiazole orange (TO) (x-axis). Flow cytometry gates separating normocytes from reticulocytes, and protein binding from no protein binding, are shown by vertical and horizontal lines, respectively. Numbers indicate the percentage of total cells within the respective gate. (B) Percentage of gorilla (green), chimpanzee (red) and human (black) normocytes bound by the respective RBP. (C) Percentage of gorilla, chimpanzee and human reticulocytes bound by the respective RBP. Experiments were performed as three technical replicates with the background signal from the antibody-only control subtracted from each binding result.

Having validated the experimental system, we next tested the binding of chimpanzee *P. vivax* RBP2e₁₅₆₋₉₅₇ and RBP3₁₄₉₋₉₆₈ to ape and human red blood cells. We

found that neither of these two proteins bound particularly well to ape red blood cells, although RBP2e consistently yielded a higher signal than RBP3 (Figure 4-4B and C). Like the human *P. vivax* RBP2a and RBP2b proteins, RBP2e and RBP3 appeared to bind reticulocytes more efficiently than normocytes (Figure 4-4; see SI Appendix, Figure S4-9). However, there was no clear evidence for host specificity. Although RBP3 bound gorilla reticulocytes slightly more efficiently than chimpanzee and human reticulocytes, this result has to be interpreted with caution since only a single gorilla sample containing very few reticulocytes was available for testing (Figure 4-4A). Indeed, when red blood cells from a macaque were tested, RBP2e and RBP3 were found to also bind to reticulocytes from this host species (see SI Appendix, Figure S4-9). To determine whether the low level of RBP2e and RBP3 binding was due to inefficient reticulocyte enrichment, we tested an additional chimpanzee blood sample with a particularly high reticulocyte count. Although Percoll gradient enrichment of this sample yielded twice as many reticulocytes (4%), this higher fraction did not improve RBP2e and RBP3 binding (see SI Appendix, Figure S4-11). Thus, the maintenance of the *rbp2e* and *rbp3* genes in chimpanzee *P. vivax* cannot be readily explained by invoking interaction with a host-specific erythrocyte receptor.

4.4 Discussion

It has recently become apparent that wild-living African apes, including western and eastern gorillas as well as chimpanzees and bonobos, harbor malaria parasites that appear to be very closely related to *P. vivax* strains infecting humans in Asia and Central/South America (Liu et al., 2017, 2014; Prugnolle et al., 2013). Since these results were based on a small number of mostly mitochondrial DNA fragments, we have now generated two near complete and several partial genome sequences from ape-infecting parasites. Analyses of these sequences show that ape and human parasites are indeed

very closely related, with the human *P. vivax* sequences forming a monophyletic lineage either within or as a sister group to the ape parasites. The new data also reveal that ape and human *P. vivax* strains exhibit distinct patterns of genetic diversity, reflecting very different demographic histories. The parasites infecting humans and apes are largely allopatric, but there is as yet no evidence that they represent distinct *Plasmodium* species. Thus, *P. vivax* in African apes represents a substantial, and genetically diverse, parasite reservoir from which future human infections could arise, even if eradication of current human strains were successful.

Recent papers have emphasized that, across the genome, the global genetic diversity of human *P. vivax* is somewhat greater than that of *P. falciparum* (Hupalo et al., 2016; Neafsey et al., 2012). Here we find that the level of neutral genetic diversity among *P. vivax* parasites from chimpanzees is about 10 times higher than that of human *P. vivax* strains. *P. falciparum* isolates also exhibit about 10 times less genetic diversity than is seen in closely related *Laverania* species (Otto et al., 2018; Sundararaman et al., 2016). This is consistent with both human parasites having undergone severe genetic bottlenecks. In the case of *P. falciparum*, this most likely occurred at the point when a gorilla parasite made the cross-species jump into humans (Liu et al., 2010; Loy et al., 2017). For *P. vivax*, it is also possible that the ancestral parasites infected only non-human apes, and that one of these crossed the species barrier and gave rise to the population of parasites currently infecting humans. Alternatively, ancient *P. vivax* may have infected all African ape species, including humans (Liu et al., 2014). In the latter case, the bottleneck would likely have occurred when *P. vivax* migrated with human out of Africa, prior to the spread of the Duffy negative mutation that eliminated *P. vivax* from humans in most (or at that time, perhaps all) of sub-Saharan Africa. While it is difficult to distinguish between these two scenarios on the basis of genetic data, we believe that the second scenario is

the more likely. Ape *P. vivax* seems to circulate freely between chimpanzees and gorillas in the wild (Liu et al., 2014) and has caused disease in a Duffy-positive human (Prugnolle et al., 2013). Consistent with this, the very high frequency of the Duffy-negative mutation suggests a long history of *P. vivax* exerting selective pressure on humans in Africa (Carter, 2003). It is likely that the geographic areas in which this mutation is most frequent (Howes et al., 2011) were influenced over a long timescale by the distribution of ape *P. vivax* to which humans were exposed.

The nature of the genetic diversity also differs markedly between the populations of *P. vivax* parasites that infect apes and humans. Among human strains, there is an unusually high fraction of nonsynonymous polymorphism. Moreover, these nonsynonymous polymorphisms exhibit site frequency spectra that are almost identical to those seen for synonymous mutations, implying that they are segregating as if effectively neutral. Similar observations have been made for polymorphism in *P. falciparum* (Chang et al., 2012), where this has been attributed, at least in part, to the fact that the parasite undergoes repeated bottleneck events in every life cycle at the obligate transmission events between host and vector, with rapid expansion in the human host (Chang et al., 2013). Here, we find that the ratio of nonsynonymous to synonymous mutations among the much more numerous polymorphisms among ape *P. vivax* strains is similar to the ratio at sites of divergence between *P. vivax* and its macaque relative *P. cynomolgi*. Thus, the unusual pattern of polymorphism in human *P. vivax* cannot reflect its life cycle, but is more likely the consequence of the population having undergone a rapid expansion subsequent to the spread out of Africa.

Looking for possible human specific adaptive changes, we performed McDonald-Kreitman tests on approximately 4,000 core genes, comparing polymorphisms within ape *P. vivax* with fixed differences between ape *P. vivax* and human *P. vivax*. After correcting

for multiple tests, we found no gene that exhibited a significant excess of non-synonymous fixed differences. We also looked for positive selection in invasion genes on the branch leading to human *P. vivax*, but found no evidence of human specific adaptation. Although this does not exclude that a small number of human adaptive changes have occurred, the time when human *P. vivax* became specialized to humans and no longer infected apes seems too short to expect significant differences.

The pseudogenization of three reticulocyte binding proteins (*rbp2d*, *rbp2e*, and *rbp3*) in all human *P. vivax* strains could be taken to indicate human-specific adaptation after the parasite migrated out of Africa and no longer encountered chimpanzees or gorillas. However, our erythrocyte binding results are not consistent with this scenario. We observed equivalent binding of RBP2e to ape and human red blood cells, and only a modest increase in RBP3 binding to gorilla reticulocytes compared to human reticulocytes. Overall, the chimpanzee *P. vivax*-derived RBP2e and RBP3 proteins bound red blood cells much less well than the human *P. vivax*-derived RBP2a and RBP2b proteins, which could be due to low affinity interactions, improper folding of the recombinant proteins, or the absence of other parasite proteins required for receptor engagement. We also considered the possibility that RBP2e and RBP3 are not involved in erythrocyte binding, or that the core receptor binding domains were not included in the expressed proteins. However, others have shown that the *P. knowlesi* orthologs of RBP3 and RBP2e, termed PkNPBXa and PkNPBXb, both bind red blood cells (Semenya et al., 2012), and that the binding domain of PkNPBXb maps to a region included in our RBP2e construct (Semenya et al., 2012). Moreover, deletion of PkNPBXa severely reduced the ability of *P. knowlesi* merozoites to invade human red blood cells *in vitro* (Moon et al., 2016), despite the fact that all human-infecting *P. vivax* strains have lost RBP3. It could be argued that the large number of RBP genes in *P. vivax* (eight full-length and two partial genes) compared with

P. knowlesi (two full-length genes and one pseudogene) provides functional redundancy that compensates for the loss of *rbp* genes in human *P. vivax*. However, this would also apply to ape *P. vivax*, where pseudogenes have not been found. Instead, the loss of *rbp* genes may be slightly deleterious and incur a fitness cost. In ape *P. vivax*, such an inactivating mutation would be expected to be selected against and thus not spread in the population. However, in human *P. vivax* the same mutation could be effectively neutral, as many nonsynonymous mutations appear to be, and so it could survive and drift to fixation. Whether this explains the loss of *rbp2d*, *rbp2e*, and *rbp3* in human *P. vivax* remains unknown. However, we found no evidence that *rbp2e* and *rbp3* genes are maintained in ape *P. vivax* parasites because they interact with chimpanzee- and/or gorilla-specific erythrocyte receptors that are absent from human red blood cells.

Despite their common origin in Africa, ape and human *P. vivax* populations have likely had little or no geographic overlap subsequent to the escape of the human-infecting lineage out of Africa. Under the bottleneck scenario, the most recent common ancestor of human *P. vivax* was in the lineage that emerged from Africa. This may have been coincident with the emergence of modern humans from Africa, perhaps around 75,000 years ago (Pagani et al., 2016). Molecular clock estimates have placed the *P. vivax* most recent common ancestor (MRCA) at least 50,000-70,000 years ago (McManus et al., 2017; Mu et al., 2005), but these relied on rate assumptions that may not be accurate. We have argued that the MRCA of *P. falciparum* strains may have existed within the last 10,000 years (Sundararaman et al., 2016), despite molecular clock estimates that place the origin of that species much earlier. Similarly, the MRCA of human *P. vivax* may have left Africa in a more recent wave of human migration, although its higher levels of genetic diversity (Hupalo et al., 2016) suggest that human *P. vivax* is older than *P. falciparum*. Once out of Africa, *P. vivax* spread through Asia and Europe, and probably from Europe

into the Americas (Carter, 2003; Gelabert et al., 2016). Strains of *P. vivax* now present in Madagascar and East Africa, in areas where non-human apes are absent, likely reflect reintroductions from Asia (Culleton et al., 2011). Given this historical isolation of ape and human *P. vivax* strains, substantial gene flow between the two populations is unlikely. The mixture of topologies found for different genes, where some show separation of the ape and human parasite lineages, while others have the human parasites nested within the radiation of the ape parasites, likely reflects an ongoing process of lineage sorting in the absence of introgression. However, this does not mean that the two populations have become isolated species. Both ape and human *P. vivax* exhibit broad natural tropism (Buery et al., 2017; Liu et al., 2014; Loy et al., 2017; Prugnolle et al., 2013), and it therefore seems very likely that ape and human strains could infect the same hosts, and undergo genetic exchange, if their geography overlapped. In recent years, reports of *P. vivax* infection of African humans have been increasing, including instances of infection of Duffy-negative individuals (Zimmerman, 2017). It will be important to monitor these cases to determine whether any reflect zoonotic transmissions from apes, and whether there is any sign of introgression between ape- and human-infecting strains.

4.5 Materials and methods

Sample collection, DNA extraction, and *Plasmodium* screening. Blood samples were obtained from chimpanzees at the Sanaga-Yong Chimpanzee Rescue Center following routine veterinary examination. Blood was also collected from a wild-living habituated chimpanzee from the Tai Forest in Cote d'Ivoire during emergency surgery (Köndgen et al., 2011). The gorilla blood sample (Gor3157, also referred to as SAggg3157) was obtained from confiscated bushmeat in Cameroon. All samples were shipped in compliance with Convention on International Trade in Endangered Species of Wild Fauna

and Flora regulations and country-specific import and export permits. DNA was extracted from whole blood samples using the QIAmp Blood DNA Mini Kit or the Puregene Core Blood Kit (Qiagen), and screened for *Plasmodium* using both pan-*Plasmodium* primers and *P. vivax*-specific primers, as described (Liu et al., 2014, 2010). Amplicons were sequenced using Sanger or MiSeq technologies (see SI Appendix, Materials and Methods).

Selective whole genome amplification. Near full-length and partial *P. vivax* genomes were amplified from unprocessed chimpanzee and gorilla blood as described (Cowell et al., 2017; Sundararaman et al., 2016) using multiple rounds of SWGA with different primer sets, with and without prior digestions with MspJI and FspEI to selectively degrade host DNA (see SI Appendix, Table S4-1).

Illumina and PacBio sequencing. Short insert libraries were prepared from pooled SWGA products of samples SY43 and SY56 using a KAPA HyperPlus Kit and MiSeq sequenced. Pooled SWGA products were also linearized with S1 nuclease (Promega), needle sheared to reduce fragment size, and subjected to PacBio SMRT Cell sequencing (University of Delaware Sequencing Core). SWGA products from samples SY81, SY90, Sagu, and Gor3157 were Illumina sequenced to obtain partial *P. vivax* genomes.

Assembly of chimpanzee *P. vivax* draft genomes. Illumina sequencing reads were error-corrected using SPAdes (Bankevich et al., 2012), mapped to the chimpanzee reference genome, and unmapped reads were mapped to the *P. vivax* P01 reference genome (Auburn et al., 2016). Regions with low coverage or poor paired-read support were removed and gaps closed using FGAP (Piro et al., 2014) with proofread-corrected

PacBio reads (Hackl et al., 2014), followed by iterations of GapFiller (Nadalin et al., 2012) and IMAGE (Tsai et al., 2010). The PvSY56 draft genome was further improved by *de novo* assembly of subtelomeric reads using SPAdes (Bankevich et al., 2012). Annotations were transferred to the final assembly using RATT (Otto et al., 2011), and predicted using Companion (Steinbiss et al., 2016) in the *de novo* assembled subtelomeres, followed by manual correction. Sets of orthologous genes from PvSY43, PvSY56, PvP01 and PvSall were masked using segmasker (<http://nebc.nox.ac.uk/bioinformatics/docs/blast+.html>) to exclude low-complexity regions, and aligned using TranslatorX/MUSCLE (Abascal et al., 2010). Outgroup sequences, where available, were added to these alignments using MUSCLE. Divergence between sequences was calculated in R using ape (Paradis et al., 2004) with no correction for multiple substitutions ('raw' model). 4-fold and 0-fold degenerate sites were extracted if the classification was true for all sequences in the alignment.

Single nucleotide polymorphism (SNP) calling. In addition to sequencing SWGA products, we also mined publicly available databases for human (Hupalo et al., 2016) and chimpanzee (Rutledge et al., 2017) *P. vivax* reads. SNPs were called following best practices for the Genome Analysis Toolkit with hard filtering (Auwera et al., 2014), using PvSY56 and PvP01 as ape and human *P. vivax* reference sequences, respectively. Regions classified as low-complexity by dustmasker, genes in subtelomeric and internal hypervariable regions (Pearson et al., 2016), and all *vir* and *phist* genes were excluded. Sites with a coverage of at least five reads were considered callable, with genes containing fewer than 60% of sites of the reference excluded. Only sites callable for all strains were analyzed. Site frequency spectra were generated from high-quality SNP calls for all high-coverage South East Asian *P. vivax* strains in the MalariaGen Genome Variation project

(Pearson et al., 2016). 4-fold and 0-fold degenerate sites were identified in the PvSalI reference sequence using custom R scripts, and derived alleles at polymorphic sites were identified using the est-sfs unfoldr (Keightley and Jackson, 2018) with PvSY56 and PcyM as outgroups.

Single genome amplification (SGA) of geographically diverse *P. vivax* strains. Ape *P. vivax* and *P. carteri* sequences were amplified as described from stored ape blood and fecal samples previously shown to be positive for these parasites (Liu et al., 2014). Fragments of five genes were targeted using newly designed primers.

Identification of reticulocyte binding protein (*rbp*) genes. Chimpanzee *P. vivax* *rbp* genes were identified during annotation of the PVS56 and PVS43 genomes. *rbp* orthologs from *P. malariae* strain PmUG01 (Rutledge et al., 2017), *P. knowlesi* strain H ((Pain et al., 2008), 2015 update), *P. cynomolgi* strains B and Berok (Pasini et al., 2017; Tachibana et al., 2012), *P. inui* strain San Antonio 1 and *P. fragile* strain Nilgiri (PlasmoDB) were identified from annotations and using *P. vivax* *rbp* genes as query sequences. Human *P. vivax* sequencing reads (Hupalo et al., 2016; Pearson et al., 2016) were downloaded and mapped to *rbp2e* (PVP01_0700500), *rbp2p1* (PVP01_0534400), *rbp2p2* (PVP01_101590), *rbp3* (PVP01_1469400), and *rbp2d* (PVP01_1471400 and PVX_101585), with a consensus sequence called for all positions with ≥ 3 -fold coverage. The position of frameshifts and stop codons was identified relative to the ape *P. vivax* sequence (see SI Appendix, Materials and Methods).

Reticulocyte binding protein (RBP) expression and red cell binding assays.

Chimpanzee *P. vivax* *rbp2d*, *rbp2e*, and *rbp3* sequences were codon optimized, synthesized, and cloned into pET-32a(+) (Novagen), yielding constructs RBP2d₁₆₅₋₉₆₇, RBP2e₁₅₆₋₉₅₇ and RBP3₁₄₉₋₉₆₈, respectively. Proteins were expressed using *E. coli* SHuffle T7 (New England Biolabs) and purified using a HisTrap (GE Healthcare) column. After hexa-histidine tag cleavage and purification through a Q-Sepharose HiTrap column (GE Healthcare), fractions containing the respective proteins (determined by SDS PAGE) were concentrated and further purified by size exclusion chromatography. RBP2d₁₆₅₋₉₆₇ could not be purified due to protein aggregation. Expression and purification of two human *P. vivax* RBP proteins, RBP2a₁₆₀₋₁₀₀₀ and RBP2b₁₆₁₋₉₆₉, were performed as described (Gruszczyk et al., 2018b, 2016).

Whole blood from five chimpanzees (New Iberia Research Center, Lafayette, Louisiana), one gorilla (Lincoln Park Zoo, Chicago, Illinois), and one rhesus macaque (BioIVT, Westbury, New York) was collected in ACD collection tubes (BD Biosciences). All ape blood samples represented left over specimens obtained during routine health screenings (the macaque blood was purchased). Blood was also obtained from healthy human volunteers at the University of Pennsylvania under IRB protocol #813699. White blood cells were first removed by leukocyte filtration and reticulocyte were subsequently enriched by spinning red blood cells (50% hematocrit) through a 65-75% (v/v) isotonic Percoll cushion and collecting the cell band at the interface. To assess RBP binding, recombinant protein was incubated with red blood cells for one hour, followed by detection with an RBP-specific polyclonal rabbit antibody and an anti-rabbit (Alexa Fluor 647 labelled) secondary antibody (ThermoFisher). Between each incubation step, cells were washed in 180 µl PBS containing 1% BSA (Sigma). Cells were stained in the dark with BD Retic-Count reagent for 30 mins at room temperature, spun, and resuspended in 1.2 ml

of PBS prior to analysis on an Accuri flow cytometer. Experiments were performed as three technical replicates with the background signal from the antibody-only control subtracted from each binding result.

4.6 Chapter acknowledgments.

We thank Andrew Smith, Catherine Bahari and Alex Shazad for performing Illumina sequencing; the University of Delaware Sequencing Core for PacBio sequencing; Dana Bellissimo for assistance with flow cytometry; the staff of the Sanaga Yong Rescue Center as well as Jane Fontenot, Melany Musso and Francois Villinger at the New Iberia Research Center for providing left-over blood samples from captive chimpanzees; Marisa Shender from the Lincoln Park Zoo for providing left over gorilla blood; the Taï Chimpanzee Project assistants for field work; the Cameroonian Ministries of Health, Forestry and Wildlife, and Scientific Research and Innovation for permission to perform studies in Cameroon; the Ministries of Environment and Forests and Research, Office Ivoirien des Parcs et Réserves, Taï National Park, and the Swiss Centre for Scientific Research for permission to perform studies in Côte d'Ivoire. This work was supported in part by grants from the National Institutes of Health to B.H.H. (R01 AI097137, R01 AI091595, R37 AI050529, P30 AI045008); the Agence Nationale de Recherche (Programme Blanc, Sciences de la Vie, de la Santé et des Ecosystèmes and ANR 11 BSV3 021 01, Projet PRIMAL) to M.P.; and an Australian Research Council Future Fellowship and a Howard Hughes Medical Institute - Wellcome Trust International Research Scholar Award (208693/Z/17/Z) to W.-H.T. D.E.L. was supported by a training grant (T32 AI 007532), and O.A.M. by Biotechnology and Biological Sciences Research Council (BBSRC) grant BB/M010996/1 (EASTBIO).

Accession codes. Chimpanzee *P. vivax* whole genome sequences and all short-read data have been deposited in the GenBank BioProject database under project number PRJNA474492, with genome annotations for PvSY43 and PvSY56 available upon request. Limiting dilution PCR-derived sequences have been deposited in GenBank under accession codes MH443154-MH443228.

4.7 Supplemental Information Appendix

Ape samples. Whole blood samples (5-10 ml) were obtained from sanctuary chimpanzees (*Pan troglodytes*) cared for at the Sanaga Yong (SY) Chimpanzee Rescue Center in Cameroon as previously described (Liu et al., 2010; Sundararaman et al., 2016), including from members of the central (*P. t. troglodytes*) and Nigeria-Cameroonian (*P. t. ellioti*) subspecies. These samples were obtained for veterinary purposes only or represented leftover specimens from yearly health examinations. None of the chimpanzees exhibited symptoms of malaria at the time of sampling. Blood was preserved in RNAlater (1:1 vol/vol) without further processing at room temperature until shipment to the United States and long term storage at -80°C. Blood from a wild-living habituated chimpanzee (“Sagu”) in the Tai Forest (*P. t. verus*) was obtained during an emergency field immobilization for treatment of a respiratory condition as described (Köndgen et al., 2011) and immediately frozen in liquid nitrogen prior to shipment to Germany and storage at -80°C. A small amount of blood (frozen directly without preservation) was also available from a western gorilla (*Gorilla gorilla*) of unknown geographic origin (Gor3157), which was killed by hunters and confiscated by the anti-poaching program of the Cameroonian Ministry of Environment and Forestry (Liu et al., 2016; Sundararaman et al., 2016). Ape fecal samples were selected from an existing bank of chimpanzee, bonobo, and gorilla specimens previously shown to contain *P. vivax* parasite DNA (Liu et al., 2017, 2014,

2010). These specimens were collected non-invasively from non-habituated apes living in remote forest areas, with a two-letter-code indicating their field site of origin (Figure S4-5). Additionally, dried blood spots were available from two chimpanzees and one gorilla housed at the Mfou National Park Wildlife Rescue Center in Cameroon, which were previously shown to be *P. vivax* positive (Liu et al., 2014). DNA was extracted from whole blood and dried blood spots using the QIAmp Blood DNA Mini Kit or the Puregene Core Blood Kit (Qiagen). Fecal DNA was extracted using the QIAamp Stool DNA mini kit (Qiagen). All specimens were subjected to host mitochondrial DNA analysis to confirm their species and subspecies origin (Liu et al., 2017, 2014, 2010). Sample collection was approved by the Ministry of Environment and Forestry in Cameroon and by the Ministry of the Environments and Forests in Cote d'Ivoire. All samples were shipped in compliance with Convention on International Trade in Endangered Species of Wild Fauna and Flora regulations and country-specific import and export permits.

PCR screening for *Plasmodium* infection. Ape blood and fecal samples were screened for *Plasmodium* sequences by diagnostic PCR using both pan-*Plasmodium* and *P. vivax*-specific primer sets as previously described (Liu et al., 2017, 2014, 2010). Briefly, nested PCR was used to amplify a 956 bp fragment of the *cytochrome b* gene (pan-*Plasmodium* primers) and a 295 bp fragment of the *cytochrome oxidase I* gene (*P. vivax*-specific primers) of the parasite mitochondrial genome. Amplicons were subjected to Sanger or Illumina sequencing, and phylogenetic analysis was performed to determine the species of the amplified *Plasmodium* sequences. Because the blood sample from the chimpanzee Sagu was previously reported to be positive for *P. vivax* (Kaiser et al., 2010), it was not rescreened for this study.

Selective amplification of ape *P. vivax* genomes. To generate *P. vivax* genome sequences from unprocessed ape blood, we used selective whole genome amplification (SWGA), which utilizes the highly processive phi29 DNA polymerase to preferentially amplify pathogen sequences from complex mixtures of target and host DNA (Leichty and Brisson, 2014; Sundararaman et al., 2016). Six sets of *P. vivax* specific primers were used (Table S4-1), of which pvset1 and pvset6 (also termed pvset1920) have previously been reported (Cowell et al., 2017). The remaining four sets (pvset2-5) were newly designed to increase overall *P. vivax* genome coverage. Using custom scripts (Sundararaman et al., 2016), we initially identified sequence motifs (6-12 bp in length) that occurred frequently in the *P. vivax* Sall reference genome (Carlton et al., 2008), but only infrequently in the human genome. These were filtered to remove primers that exhibited extreme melting temperatures, were predicted to form homodimers, and/or bound the Sall mitochondrial genome or its subtelomeric regions more than three times. The resulting primer sets, including pvset1 (5'-CGTTG*C*G-3', 5'-TTTTTTC*G*C-3', 5'-TCGTG*C*G-3', 5'-CGTTTTTT*T*T-3', 5'-TTTTTTC*G*T-3', 5'-CCGTT*C*G-3', 5'-CGTTTC*G*T-3', 5'-CGTTTC*G*C-3', 5'-CGTTTT*C*G-3', and 5'-TCGTTC*G*T-3') and pvset2 (5'-CGAAAAA*A*A-3', 5'-CGCAA*C*G-3', 5'-GCGAAA*T*G-3', 5'-CGCAC*G*A-3', 5'-GCGAAAA*A*A-3', 5'-AACGAAAA*A*A-3', 5'-AACGAA*C*G-3', 5'-ACGAAA*C*G-3', 5'-CGAACG*A*A-3', and 5'-CGAAAC*G*G-3'), exhibited high coverage of the *P. vivax* genome and low coverage of the human genome (asterisks indicate the location of phosphorothioate bonds necessary to prevent degradation by the phi29 polymerase). Two additional sets, including pvset3 (5'-CTTCGAA*C*G-3', 5'-GCGAAAC*G*T-3', 5'-GGCGAAAAA*A*A-3', 5'-TCGCGAA*A*A-3', 5'-TTTCGCG*T*A-3', and 5'-TTTCGTG*C*G-3') and pvset4 (5'-CGAAGCGG*A*G-3', 5'-CTTCGAA*C*G-3', 5'-

GCGAAAC*G*T-3', 5'-GGCGAAAAA*A*A-3', 5'-TCGCGAA*A*A-3', 5'-TTTCGCG*T*A-3', and 5'-TTTCGTG*C*G-3'), were generated using the program *swga* (Clarke et al., 2017), which is designed to select primer sets that bind evenly across the reference genome. After noticing preferential amplification of AT-rich subtelomere regions in SWGA products, we designed two final primer sets using only GC-rich regions of the *Sall* reference sequence as the foreground genome (Cowell et al., 2017), resulting in pvset5 (5'-AGCGAAAAA*A*A-3', 5'-AGCGAAC*G*T-3', 5'-CAAACGGG*T*G-3', 5'-CGAACGA*A*T-3', 5'-CGAAGCGG*A*G-3', 5'-CGAATGGG*G*G-3', 5'-CGAGCGA*A*C-3', 5'-CGTTTGG*C*G-3', 5'-GCGGGAAAA*A*A-3', 5'-GCGTGTA*C*G-3', 5'-TACGACG*A*G-3', and 5'-TTCAGCG*C*G-3') and pvset6 (5'-AACGAAGC*G*A-3', 5'-ACGAAGCG*A*A-3', 5'-ACGACGA*A*G-3', 5'-ACGCGCA*A*C-3', 5'-CAACGCG*G*T-3', 5'-GACGAAA*C*G-3', 5'-GCGAAAAA*G*G-3', 5'-GCGAAGC*G*A-3', 5'-GCGGAAC*G*A-3', 5'-GCGTCGA*A*G-3', 5'-GGTTAGCG*G*C-3', and 5'-AACGAAT*C*G-3').

SWGA was performed as described (Cowell et al., 2017; Sundararaman et al., 2016) by amplifying whole blood DNA (100-750 ng) in a 50 µl reaction with 1x phi29 Buffer (New England Biolabs), 1 mM dNTPs (Roche), 3.5 µM of SWGA primers (an equimolar mix of primers in the set), 1% BSA and 30 units of phi29 polymerase (New England Biolabs). SWGA conditions included a 1 h ramp-down step (35°C to 30°C), followed by an amplification step for 16 h at 30°C, followed by a phi29 denaturation step for 10 min at 65°C. SWGA products were diluted 1:1 in water, purified using AMPure Beads (Beckman Coulter), and stored at 4°C. To mitigate the stochastic nature of SWGA at low template concentrations (Sundararaman et al., 2016), genomic DNA from each ape-derived sample was amplified on multiple independent occasions with different primer sets (Table S4-1). Because pretreatment with restriction enzymes that selectively degrade host DNA can

improve SWGA efficiency (Sundararaman et al., 2016), some DNA aliquots were digested with the methylation sensitive enzymes MspJI and FspEI (5 units each) for 2 hours at 37°C prior to SWGA amplification (Table S4-1). To obtain sufficient quantities of parasite genomic DNA for sequencing, ape-derived DNA samples were subjected to multiple rounds (up to 4) of successive SWGA amplification, some of which were performed with alternating primer sets to improve genome coverage (Cowell et al., 2017).

Illumina and PacBio sequencing. To generate chimpanzee *P. vivax* draft genomes, we used SWGA amplicons from samples SY43 and SY56 for Illumina and PacBio sequencing. These chimpanzee blood samples were selected because they were PCR positive for *P. vivax cytb* and *cox1* regions, but lacked *Laverania* sequences. For Illumina sequencing, we pooled second and fourth round SWGA products to prepare short insert libraries using the KAPA HyperPlus kit (Roche) with an enzyme fragmentation time of 3 minutes. Fragmentation products were purified with AMPure Beads (Beckman Coulter), followed by dual-sided solid phase reversible immobilization (SPRI) to select for fragments 550 bp in length. The resulting libraries were sequenced on an Illumina MiSeq platform using V2 chemistry. For PacBio sequencing, second and third round SWGA products were pooled for library preparation. Briefly, amplification products (7.5-40 µg) were incubated with S1 nuclease (15 units) for 30 min at 37°C for DNA linearization, purified using AMPure beads, and passed through a 26 gauge blunt end needle to reduce DNA fragment size from >60,000 bp to ~15,000 bp. DNA was purified, eluted in 40 µl of water, and sent to the University of Delaware Sequencing Core where fragments of 7,000-18,000 bp length were size-selected using BluePippin (Sage Biosciences) prior to SMRT Bell library preparation and PacBio SMRT Cell sequencing.

Assembly of PvSY43 and PvSY56 draft genomes. Illumina MiSeq reads from chimpanzee blood samples SY43 and SY56 were error corrected using SPAdes (Bankevich et al., 2012) and then mapped to the chimpanzee reference genome using smalt (<http://www.sanger.ac.uk/science/tools/smalt-0>). Unmapped reads were extracted and converted to fastq files using SAMtools (Li et al., 2009) and BEDtools (Quinlan and Hall, 2010), respectively. PacBio reads from samples SY43 and SY56 that were longer than 1,500 bp and 1,000 bp, respectively, were filtered to remove chimpanzee reads using BLASR (Chaisson and Tesler, 2012). The resulting non-chimpanzee reads were corrected using proovread (Hackl et al., 2014).

The non-chimpanzee MiSeq reads were iteratively mapped (10 times) with correction to the human *P. vivax* PvP01 reference sequence in Geneious (version 9) to generate a preliminary chimpanzee *P. vivax* consensus sequence (Kearse et al., 2012). Errors were identified based on low read support as described (Sundararaman et al., 2016), the assembly was separated into contigs by splitting at assembly gaps, and contigs <100 bp were removed using Geneious (Kearse et al., 2012), SAMtools (Li et al., 2009), BEDtools (Quinlan and Hall, 2010) and custom scripts. The resulting contigs were then mapped to the PvP01 reference genome using ABACAS (Assefa et al., 2009). Gaps were closed using FGAP (Piro et al., 2014) with proovread-corrected, non-chimpanzee PacBio reads. After initial gap closure, regions with low read support were again removed. Finally, gaps were filled using IMAGE (Tsai et al., 2010) and GapFiller (Nadalin et al., 2012), followed by removal of likely duplications and inversions at the edges of gaps and a final error correction with iCORN (Otto et al., 2010).

The PvSY56 genome assembly was improved by *de novo* assembly of subtelomere and internal hypervariable regions. Orthologs of genes that bounded the subtelomeres and internal hypervariable regions in PvSall (Pearson et al., 2016) were

identified in PvP01 and PvSY56, and the subtelomeres and internal hypervariable regions of these genomes were then defined as the sequence between the boundary gene and the nearest chromosome end, or the sequence between and including the two internal boundary genes, respectively. For *de novo* assembly, corrected Illumina reads from sample SY56 were mapped using smalt (<http://www.sanger.ac.uk/science/tools/smalt-0>) to a version of PvP01 in which the subtelomeres and internal hypervariable regions were masked. Reads that did not map to the masked genome were extracted using SAMtools (Li et al., 2009) and BEDtools (Quinlan and Hall, 2010). Since these contained a large number of *Pseudomonas* reads, we removed these by mapping to the *Pseudomonas yamanorum* genome (GenBank accession number: LT629793). The resulting reads were assembled using SPAdes (Bankevich et al., 2012). *De novo* contigs were retained if they were at least 2 kb in length and either could be mapped to PvP01 subtelomeres by ABACAS or had a blastn hit (e-value $<10^{-4}$) to any of four human *P. vivax* genome assemblies (PvSall, PvP01, PvC01, PvT01) (Auburn et al., 2016). These were ordered to the PvP01 genome using Companion (Steinbiss et al., 2016). Ordered *de novo* contigs that provided a better assembly of the subtelomeres or internal hypervariable regions than the initial PvSY56 assembly were exchanged.

Genome annotation. Annotations were transferred from the human *P. vivax* PvP01 reference sequence to the PvSY56 and PvSY43 genome assemblies using RATT (Otto et al., 2011), with additional genes in PvSY56 subtelomeres predicted using Companion (Steinbiss et al., 2016). All annotations on the chromosomes were visually inspected, and manually corrected where necessary. In roughly 10% of genes, we noted one or two base pair insertions or deletions in homopolymer tracts. Since these small indels were restricted to homopolymer regions but found throughout the genome, we reasoned that they

represented sequencing errors. We thus manually corrected the annotation to maintain the reading frame. Additional *ad hoc* manual corrections were performed when alignments indicated an error and mapped reads supported either sequence correction or removal.

Partial ape *P. vivax* genome sequences. To generate genome sequences from additional ape *P. vivax* strains, we subjected diagnostic PCR positive chimpanzee (SY81, SY90, Sagu) and gorilla (Gor3157) blood DNA to SWGA analysis (Table S4-1). The resulting amplification products were prepared for sequencing using Nextera protocols and sequenced using the Illumina MiSeq and/or MiniSeq platforms. Reads were mapped to the respective host reference genome (chimpanzee or gorilla) using smalt (<http://www.sanger.ac.uk/science/tools/smalt-0>) and unmapped reads were extracted using SAMtools (Li et al., 2009) and BEDtools (Quinlan and Hall, 2010).

We also mined a publicly available read database from a blood sample of a *P. malariae*-infected sanctuary chimpanzee from Gabon (GA02), which we noted contained a substantial number of ape *P. vivax* reads (Rutledge et al., 2017). Sequencing reads (ERS434565) were mapped to concatenated reference sequences of the chimpanzee *P. reichenowi* strain CDC (Otto et al., 2014), the human *P. vivax* strain P01 (Auburn et al., 2016), and the human *P. malariae* strain UG01 (Rutledge et al., 2017) using smalt (<http://www.sanger.ac.uk/science/tools/smalt-0>). Reads that mapped to *P. vivax* were used for this study.

Variant calling. For human *P. vivax*, we included genome data from previously sequenced field isolates from different geographic regions (SRA accession numbers SRP046091, SRS805942, SRP046182, SRP046094, SRP045997, SRP046126, SRP046031) that were not monkey adapted and classified as high-quality single infections

by the authors (Hupalo et al., 2016). For ape *P. vivax*, we analyzed genome data from four chimpanzee-derived samples (PvSY81, PvSY90, PvSagu, PvGA02) and one gorilla-derived sample (PvGor3157), with uncorrected Illumina sequencing reads from PvSY43 and PvSY56 also included to aid validation of variants. Variants were called using the Genome Analysis Toolkit (GATK) version 4.0 (Auwera et al., 2014). Sequencing reads were mapped to either PvSY56 (ape *P. vivax* samples) or PVP01 (human *P. vivax* samples) using bwa (<http://bio-bwa.sourceforge.net>), with duplicate reads removed and base quality scores recalibrated. 'Known variants' for base quality score recalibration (BQSR) were generated for each reference genome by a bootstrap procedure of variant calling, hard filtering with GATK suggested generic filters, and BQSR using the filtered variants, which was repeated until variants called in subsequent iterations showed little difference. The de-duplicated, recalibrated bam files were used to generate genomic variant call format files for each sample using HaplotypeCaller, with the OverClippedReadFilter applied. Variants were called jointly in human *P. vivax* and ape *P. vivax* samples, excluding insertion-deletion polymorphisms. Single nucleotide polymorphisms (SNPs) were hard-filtered using a set of annotations and values selected to minimize the number of SNPs removed at 4-fold degenerate sites, while minimizing the number of subtelomeric SNPs retained (QualByDepth <2.0, FisherStrand >50.0, RMSMappingQuality <45.0, StrandOddsRatio >2.5, GenotypeQuality_Mean <35.0, Quality <30.0, MappingQualityRankSumTest <-2.0 or >2.0, ReadPosRankSumTest <-6.0 or >10.0, BaseQualityRankSumTest <-6.0 or >10.0). Dustmasker (BLAST+ suite) (Camacho et al., 2009) was used with default settings to identify low-complexity regions within each reference, which were then excluded in all samples. Since the reference genome for ape *P. vivax* was PvSY56, sites that were called as SNPs with PvSY56 reads were assumed to be errors in the PvSY56 assembly and excluded. The ability to call sites

was assessed using CallableLoci (minimum mapping quality 20, minimum depth 5, OverClippedReadFilter applied). Uncallable sites and sites with heterozygous SNP calls in a given sample were excluded from that sample.

Orthologous gene alignments. Gene coding sequences were extracted from the chimpanzee PvSY43 and PvSY56 genome assemblies, the human *P. vivax* reference strains PvP01 and PvSall, the *P. cynomolgi* reference strain PcyM (Pasini et al., 2017) and the *P. knowlesi* reference strain PknH (Pain et al., 2008). Low-complexity regions were masked using segmasker (BLAST+ suite) (Camacho et al., 2009). Groups of 1:1 orthologs were identified from the RATT annotation transfer for PvP01, PvSY43 and PvSY56, and from PlasmoDB ortholog groups. Genes in subtelomeric or internal hypervariable regions, genes that were annotated as pseudogenes or had internal stop codons, and all *vir* and *phist* genes were excluded from the analysis. Orthologous gene sequences were aligned from the *P. vivax* genome assemblies using TranslatorX/MUSCLE (Abascal et al., 2010) and genes with unusually high divergence were inspected and manually corrected when necessary. Orthologs from *P. cynomolgi* and *P. knowlesi* were added to these alignments where available, excluding the 2% and 3% of genes with the highest divergence, respectively. To add additional ape and human *P. vivax* strains to the alignment, gene sequences were generated from SNP calls by changing the reference sequence to the alternative allele at variant sites. Sites that were excluded from SNP calling were masked. Sequences of genes with >60% of sites callable and no internal stop codons were added to the alignments, following the alignment of the appropriate reference sequence. Ambiguous and masked sites as well as assembly gaps in any sequence were masked in all sequences.

Polymorphism analysis. Nucleotide diversity and divergence were calculated from gene alignments using the `dist.dna` function with the 'raw' model in the `ape` R package (Paradis et al., 2004) and custom scripts. Synonymous (P_s) and nonsynonymous (P_n) polymorphisms were counted by determining the effect of each variant allele on the codon in the reference genome (PvSY56 or PvP01). If a site had multiple alternative alleles, each of these was counted as a separate SNP. Synonymous and nonsynonymous fixed differences were counted by comparing codons between *P. cynomolgi* strain M and PvSY56 or PvP01, after exclusion of sites that were polymorphic in ape *P. vivax* or human *P. vivax*, respectively, and assuming the mutation path with the smallest number of nonsynonymous changes. The neutrality index [$NI = (P_n/P_s)/(D_n/D_s)$] of each gene was calculated from the number of polymorphisms within ape and human *P. vivax* and the number of fixed differences (D_s and D_n) from the *P. cynomolgi* strain PcyM. Density distributions of $\log_2(NI)$ were generated in R for genes that had a defined $\log_2(NI)$ value in both ape and human *P. vivax*. McDonald-Kreitman tests were performed in R, using a two-tailed Fisher exact test (`fisher.test`), followed by correction of the p -values for multiple testing (`p.adjust, method=fdr`).

Site frequency spectra were generated from SNP calls from the MalariaGen *P. vivax* Genome Variation project, using a subset of Southeast Asian field isolates (Cambodia, Indonesia, Laos, Myanmar, Malaysia, Papua New Guinea, Thailand and Vietnam) for which $\geq 80\%$ of variant loci had been called. 173 samples met this inclusion criterion. Heterozygous calls were counted as one occurrence of the reference and one of the alternative allele. The number of allele calls varied between sites because of missing data and heterozygous SNPs. To standardize the number of calls per site without losing large amounts of data, we down-sampled to 168 calls per site, excluding sites with fewer

calls (this threshold was chosen such that 95% of sites were retained). 4-fold and 0-fold degenerate sites were identified in the Sall reference sequence, as this was the reference utilized in the MalariaGen SNP data set (Pearson et al., 2016). Alignments of orthologous genes from PvSall, PvSY56 and PcyM were generated with TranslatorX/MUSCLE and used with the est-sfs (Keightley and Jackson, 2018) unfoldr to identify the derived allele at each site and calculate the frequency spectrum.

Single genome amplification. To increase the number of geographically diverse ape parasite sequences, we subjected *P. vivax* positive ape blood and fecal samples to single genome amplification (SGA), which utilizes end-point-diluted template DNA and thus generates *Plasmodium* sequences devoid of PCR-induced sequence artifacts (Liu et al., 2014, 2010). Five nuclear gene regions were targeted to complement existing genome data, including PVP01_1216000 (610bp), PVP01_1418300 (476 or 491bp), PVP01_1418500 (809 or 815bp), PVP01_1418600 (351bp), and PVP01_1418800 (576bp). These regions were amplified using first round primers Pv6000F1 (5'-ATGGAAAGGCAGGGCGACGC-3' and Pv6000R1 (5'-GCTGCACAGGTAGGAGATGTACT-3'), Pv8300F1 (5'-AACGTGGAGATGTAATTCCTGCC-3' and Pv8300R1 (5'-TTGTGTGCATTTTCGAGCAGGCTG-3'), Pv8500F1p (5'-ATGGAGGACGAGACGGAGAAC-3') and Pv8500R1 (5'-CTGAAATAGATGTAGTTGTAGAAGG -3'), Pv8600F1 (5'-AAGAMAAACATTTTGGAAAACGCAG-3') and Pv8600R1 (5'-TCAAACTCC ATGGGGATGTTCTGC-3'), as well as Pv8800F1 (5'-TGTACGACTCGATGAGTTACTTCC-3') and Pv8800R1 (5'-TCACAGGAAGACCGTCGAAAACG-3'), respectively. Samples were multiplexed using

2.5 µl of end-point diluted DNA in a 25 µl reaction volume containing 0.5 µl dNTPs (10mM of each dNTP), 2.5 µmol of each first round primer, 2.5 µl PCR buffer, 0.1 µl BSA solution (50µg/ml), and 0.25 µl expand long template enzyme mix (Expand Long Template PCR System, Sigma). Cycling conditions included an initial denaturing step of 2 min at 94°C, followed by 15 cycles of denaturation (94°C, 10 sec), annealing (45°C, 30 sec), and elongation (68°C, 1 min), followed by 35 cycles of denaturation (94°C, 10 sec), annealing (48°C, 30 sec), and elongation (68 °C, 1 min; with 10-sec increments for each successive cycle), followed by a final elongation step of 10 minutes at 68°C. For second round PCR, 2 µl of the first round product were amplified using second round primers Pv6000F2 (5'-TAGAGGAGCAAGAGCGAGTGC-3') and Pv6000R2 (5'-TTCGACTCCTGCATTTGCCACTTG-3'), Pv8300F2 (5'-TTAACACGGAGGAGGCCACAGAATG-3') and Pv8300R2 (5'-CTCTCTCGTTTGTCTGCCTTCTTCC-3'), Pv8500F2p (5'-CAGAACTTGAAATGTCCAGGGAG-3') and Pv8500R2 (5'-AGCTGCCAGTTGTGCTTGTCTGCG-3'), Pv8600F2 (5'-GCAAAGGACATGACGCAAAGTG-3') and Pv8600R2 (5'-TTTCATCAAACGTGCATCTCTTGG-3'), as well as Pv8800F2 (5'-CTTATTTTGCTACGAAGATTG-3') and Pv8800R2 (5'-GCAATATATCCGCCTCTCTCCTC-3'), respectively. Cycling conditions included an initial denaturation step of 2 min at 94°C, followed by 60 cycles of denaturation (94°C, 10 sec), annealing (52°C, 30 sec), and elongation (68°C, 1 min), followed by a final elongation step of 10 minutes at 68°C. Amplification products were sequenced directly without interim cloning.

We also used SGA to amplify regions of ape *P. vivax* *rbp* genes, which in human *P. vivax* contained inactivating mutations. Four fragments were targeted, including *rbp2d*_

frameshift (81bp), *rbp2e_frameshift* (120bp), *rbp2e_stop* (61bp), and *rbp3_stop* (52bp). These regions were amplified using first round primers *anfsRBP2d_1F* (5'-AATGATGCAAAGAATTTTATTTTCGGAT-3' and *anfsRBP2d_1R* (5'-ACGCTTTCCTTTTCACTATCAATT-3'), *anfsRBP2e_1F* (5'-TGCAAGAAAACCATCTCGCT-3' and *anfsRBP2e_1R* (5'-TGCTCTCTTCATTTCTTCGTCA-3'), *anstopRBP2e_1F* (5'-ACAAAGCAAAAGGGCGAAGT-3') and *anstopRBP2e_1R* (5'-AGCGGATTCTTTGTGACTCCT-3'), as well as *anstopRBP3_1F* (5'-AATGAAGGGGAAGTGAAGGT-3') and *anstopRBP3_1R* (5'-TTTCTTTTCGCCGCACTATGG-3'), respectively. PCRs were multiplexed using 2.5 µl of end-point diluted sample DNA in a 25 µl reaction volume, containing 0.5 µl dNTPs (10mM of each dNTP), 2.5 µmol of each first round primer (4 pairs), 2.5µl PCR buffer, 0.1 µl BSA solution (50µg/ml), and 0.25 µl expand long template enzyme mix (Expand Long Template PCR System, Sigma). Cycling conditions included an initial denaturing step of 2 minutes at 94°C, followed by 50 cycles of denaturation (94°C, 10 sec), annealing (48°C, 30 sec), and elongation (68°C, 30 sec), followed by a final elongation step of 10 minutes at 68°C. For second round PCR, 2 µl of the first round product were amplified using second round primers *anfsRBP2d_2F* (5'-AGATGATCTGAATAAACGTTTCACA-3') and *anfsRBP2d_2R* (5'-ACAAATTCGTCAACGTAAAGTGT-3'), *anfsRBP2e_2F* (5'-AGGACAACACATATGCAGTTACT-3') and *anfsRBP2e_2R* (5'-ACTTTTATGGTCACCGTAGATACA-3'), *anstopRBP2e_2F* (5'-ACACATGATATTGATGCACTCAAAGA-3') and *anstopRBP2e_2R* (5'-TCTTGATTTGTCTCACTATTCTCTGT-3'), as well as *anstopRBP3_2F* (5'-ACAATGTGTGTAAGAATATTGAGACCA-3') and *anstopRBP3_2R* (5'-TGGGACACATTTTCTATACAGGCT-3'), respectively. Cycling conditions included an

initial denaturation step of 2 minutes at 94°C, followed by 50 cycles of denaturation (94°C, 10 sec), annealing (52°C, 30 sec), and elongation (68°C, 30 sec), followed by a final elongation step of 10 minutes at 68°C. Amplification products were sequenced directly without interim cloning.

Phylogenetic analyses and *rbp* gene comparisons. To examine the evolutionary relationships of ape and human *P. vivax* strains, phylogenetic trees were constructed from (i) nuclear gene sequences generated by SGA from infected ape blood or fecal DNA, (ii) PvP01, PvSall, PvSY56, PvSY43, *P. cynomolgi* and *P. knowlesi* genome assemblies, and (iii) SNP data by changing the reference sequence to the alternative allele at variant sites. Sequences were aligned with TranslatorX/MUSCLE and the alignments manually corrected, including truncation to the SGA amplicon where appropriate and removal of ambiguously aligned regions. Trees were generated using PhyML with a GTR+G model of nucleotide evolution, 10 random starts and best of NNI/SPR trees, and bootstrap values calculated from 100 replicates. For neighbor-joining trees, matrices of pairwise genetic distances were calculated from alignments of genes that were covered in all strains, and unrooted trees were generated from these matrices in R with ape 'nj'. Phylogenetic networks were generated from the same alignments in SplitsTree4 (Huson and Bryant, 2006) using SplitDecomposition with uncorrected p-distances.

Ape and human *P. vivax* *rbp* genes were identified from genome annotations. For *rbp1a*, sequences generated from SNP data were also included. For other *Plasmodium* species, *rbp* genes were identified from annotations in the *P. malariae* strain PmUG01 (PmUG01_07014300, PmUG01_07014200, PmUG01_08058500, PmUG01_12081700, PmUG01_14085600) (Rutledge et al., 2017), the *P. knowlesi* strain H (PKNH_0700200, PKNH_1472300) (Pain et al., 2008), the *P. cynomolgi* strains B and Berok

(PCYB_071060, PCYB_081060, PCYB_053840, PCYB_071010, PCYB_053850, PCYB_147650, JQ422038) (Tachibana et al., 2012), the *P. inui* strain San Antonio 1 (C922_04999, C922_01465) (PlasmoDB) and the *P. fragile* strain Nilgiri (AK88_00929, AK88_00936) (PlasmoDB). Blastn search using *P. vivax* *rbp* genes also identified two non-annotated pseudogenes (an ortholog of *rbp2a* in the *P. cynomolgi* strain B, and *nbp1* in *P. knowlesi*) as well as pseudogenes annotated as multiple gene fragments (*rbp1a*, *rbp2a* and *rbp2e* orthologs in *P. inui*; *rbp1b* and *rbp3* orthologs in *P. fragile*). Sequences were aligned with TranslatorX/MUSCLE, and alignments manually corrected. A tree was constructed for the most conserved region of this alignment (corresponding to nucleotides 478-7938 in PVP01 *rbp1a*) using PhyML with a GTR+G model of nucleotide evolution, 10 random starts and best of NNI/SPR trees, and bootstrap values calculated from 100 replicates. For tests of selection, branch-site models were fitted to the data using codeml in the PAML package (Yang, 2007), using alignments excluding *P. ovale* spp. and *P. malariae*, which were considered too divergent. The fit of the null model (no selection) was compared with the fit of the model with selection along the branch of the PhyML maximum-likelihood phylogenetic tree leading to human *P. vivax* (foreground branch), by comparing twice the difference in log-likelihood ($2\Delta\ln L$) with a X^2 distribution with one degree of freedom.

Published human *P. vivax* reads (Hupalo et al., 2016; Pearson et al., 2016) were downloaded from the SRA database and mapped to *rbp2d* (PVP01_1471400 and PVX_101585), *rbp2e* (PVP01_0700500), *rbp3* (PVP01_1469400), *rbp2p1* (PVP01_0534400), and *rbp2p2* (PVX_101590) sequences using smalt (<http://www.sanger.ac.uk/science/tools/smalt-0>). Mapped reads were extracted, converted to fastq files using SAMtools (Li et al., 2009) and BEDtools (Quinlan and Hall, 2010), and then imported into Geneious (version 9) (Kearse et al., 2012) to again be

mapped to the references. Because a truncation of *rbp2d* has been identified in some human *P. vivax* field isolates (Hietanen et al., 2016), including PvP01, reads were mapped to both the Sall and PvP01 alleles of *rbp2d*, and the correct allele for each human *P. vivax* sample was chosen by visual inspection of read mapping. A majority consensus sequence was called for all positions with ≥ 3 -fold coverage. Alignments of genes with complete coding sequences were made to chimpanzee *P. vivax* *rbp* genes and the positions of frameshifts and pseudogenes were noted. After identification of the ancestral stop or frameshift mutation, consensus sequences from samples with fewer than 3-fold coverage in some regions of the genes were inspected to verify that the ancestral mutation was indeed in all sequenced human *P. vivax* strains.

Recombinant RBP expression and polyclonal rabbit antibody production.

Chimpanzee *P. vivax* *rbp2d*, *rbp2e* and *rbp3* gene sequences were generated by aligning PvSY43 and PvSY56 sequencing reads to human *P. vivax* reference sequences. Deduced amino acid sequences spanning positions 100 to 1,000 of RBP2d (from PvSY43), RBP2e (from PvSY56), and RBP3 (from PvSY56) were codon optimized for expression in *Escherichia coli*, and the synthesized genes were purchased from Eurofins Genomics (RBP2d and RBP3) and GenScript (RBP2e). These were then used to generate shorter gene fragments (RBP2d₁₆₅₋₉₆₇, RBP2e₁₅₆₋₉₅₇, RBP3₁₄₉₋₉₆₈) predicted to express stable recombinant proteins, which included the respective binding domain based on homology to human *P. vivax* RBP2a and RBP2b (van den Ent and Löwe, 2006). RBP2d₁₆₅₋₉₆₇, RBP2e₁₅₆₋₉₅₇ and RBP3₁₄₉₋₉₆₈ were cloned into the pET-32a(+) vector (Novagen), which expresses proteins with a hexa-histidine tag (C-terminus) but also contains a tobacco etch

virus (TEV) protease cleavage site to allow removal of this tag. All clones were sequence confirmed.

Proteins were expressed using *E. coli* strain SHuffle T7 (New England Biolabs) and Terrific Broth (TB) supplemented with 100 µg/ml of carbenicillin. Flasks containing 1 liter of medium were incubated in a Multitron shaker (Infors HT) at 37°C at 200 rpm. At OD₆₀₀ of ~1.0, isopropyl-(β)-D-thiogalactopyranoside (IPTG, Astral) was added to a final concentration of 1.0 mM and protein expression was allowed to continue for 20 hours at 16°C. Cells were harvested by centrifugation at 6,000 x g, resuspended in freezing buffer containing 50 mM TrisHCl pH 7.5, 500 mM NaCl and 10% (v/v) glycerol supplemented with cOmplete EDTA-free protease inhibitor cocktail (Roche), flash-frozen in liquid nitrogen and stored at -80°C until further processing.

For protein purification, bacterial cell pellets were thawed on ice and resuspended in freezing buffer supplemented with 0.5 mg/ml of DNase and 1.0 mg/ml of lysozyme (Sigma-Aldrich). Cells were sonicated and centrifuged at 30,000 x g for 45 minutes at 4°C, and the resulting supernatant was added to a 5 ml HisTrap column (GE Healthcare). All unbound material was washed off using at least 10 column volumes of washing buffer (50 mM TrisHCl pH 7.5, 500 mM NaCl, 10% (v/v) glycerol, 20 mM imidazole). Bound protein was eluted from the column using the same buffer, but with the imidazole concentration increased to 300 mM. The eluted fractions were pooled and dialyzed in the presence of TEV protease in a buffer containing (i) 20 mM CAPS pH 10.0 and 100 mM NaCl for RBP2d₁₆₅₋₉₆₇, (ii) 20 mM TrisHCl pH 8.5 and 100 mM NaCl for RBP2e₁₅₆₋₉₅₇, and (iii) 20 mM TrisHCl pH 8.0 and 100 mM NaCl for RBP3₁₄₉₋₉₆₈. The resulting sample was added to a 5 ml Q-Sepharose HiTrap column (GE Healthcare), with unbound material washed off using at least 10 column volumes of the buffer. The protein was eluted using a gradient

(0-50%) of (i) 20 mM CAPS pH 10.0 and 1.0 M NaCl for RBP2d₁₆₅₋₉₆₇, (ii) 20 mM TrisHCl pH 8.5 and 1.0 M NaCl for RBP2e₁₅₆₋₉₅₇, and (iii) 20 mM TrisHCl pH 8.0 and 1.0 M NaCl for RBP3₁₄₉₋₉₆₈. Collected fractions (2.5 ml) were analyzed on SDS PAGE and those containing protein were concentrated using Vivaspin 15 Turbo centrifugal concentrators with a molecular weight cut-off 5 kDa (Sartorius) and injected onto S200 Superdex 16/600 size exclusion column (GE Healthcare) preequilibrated with (i) 20 mM NaHEPES pH 7.5, 500 mM NaCl, 5 mM β -mercaptoethanol and 5% (v/v) glycerol for RBP2d₁₆₅₋₉₆₇, (ii) 20 mM TrisHCl pH 8.5, 300 mM NaCl and 10% (v/v) glycerol for RBP2e₁₅₆₋₉₅₇, and (iii) 20 mM TrisHCl pH 8.5, 300 mM NaCl and 10% (v/v) glycerol for RBP3₁₄₉₋₉₆₈. Fractions containing pure proteins (2 ml) were pooled, concentrated and flash-frozen in liquid nitrogen. RBP2e₁₅₆₋₉₅₇ and RBP3₁₄₉₋₉₆₈ yielded monodisperse peaks; however, RBP2d₁₆₅₋₉₆₇ was heavily aggregated despite using high salt and a reducing agent in the buffer. To examine protein folding, circular dichroism (CD) data (Figure S4-8) were collected using a CD spectrometer Model 410 (Aviv Biomedical) and analyzed as described (Gruszczyk et al., 2016). Expression and purification of RBP2a₁₆₀₋₁₀₀₀ and RBP2b₁₆₁₋₉₆₉ has been described previously (Gruszczyk et al., 2018b, 2016).

Polyclonal rabbit antibodies were generated by the Walter and Eliza Hall Institute (WEHI) Antibody Facility. Rabbits were immunized 5 times with 150 μ g of the respective recombinant protein. The first immunization was administered in Complete Freund's adjuvant and the remainder in Incomplete Freund's Adjuvant. Rabbit IgG fractions were purified from serum using Protein G sepharose. These studies were approved by the WEHI Institutional Animal Care and Use Committee.

RBP binding assays. 30-100 ml of whole blood was collected in 10 ml ACD blood collection tubes (BD Biosciences) from five chimpanzees (New Iberia Research Center,

Lafayette, Louisiana), one gorilla (Lincoln Park Zoo, Chicago, Illinois), and one rhesus macaque (BioIVT, Westbury, New York). All ape blood samples represented left over specimens obtained during routine health screenings and were approved by the respective Institutional Animal Care and Use Committees. The macaque blood was purchased. Blood was also obtained from healthy human volunteers at the University of Pennsylvania under IRB protocol #813699 with informed consent (kindly provided by R. Collman). Whole blood was centrifuged at $1,500 \times g$ for 20 minutes (maximum acceleration and low brake speed). Buffy coats containing leukocytes were removed and red blood cells were resuspended in their respective plasma or in PBS. Resuspended red blood cells were passed through a SepaCell R-500 II filter (Fenwal) or an Acrodisc filter with leukosorb media (PALL) to remove remaining leukocytes. Reticulocytes were enriched by carefully layering 5.5 ml of red blood cells diluted in plasma (50% hematocrit) over 6 ml of a 65-75% (v/v) isotonic Percoll cushion. The percent Percoll was varied to achieve maximum reticulocyte enrichment for each sample after testing a small aliquot on a 70% (v/v) isotonic Percoll cushion. Blood layered over Percoll was centrifuged at $1,650 \times g$ for 20 min (maximum acceleration and low brake speed) and the resulting cell band at the Percoll interface was removed, pooled, washed three times in PBS, and stored in RPMI (Gibco). Although this protocol achieved reticulocyte enrichment for human blood samples (up to 60%), this was not the case for ape and monkey blood samples. Although different centrifugation speeds ($1,200 - 2,100 \times g$), Percoll densities (65-75%), and PBS formulations (Gibco 1x and 10x DBPS, Ambion 10x PBS pH 7.4, and human tonicity (HT)PBS) were tested, we were unable to achieve reticulocyte enrichment of more than 2% for the great majority of ape and monkey samples, as measured by thiazole orange staining by flow cytometry (BD Retic-Count; BD Biosciences). The exception was a blood sample from an anemic chimpanzee, which yielded 4.0% reticulocyte enrichment. For

each chimpanzee and macaque, we selected the fraction with the highest percentage of reticulocytes for subsequent binding studies. The single gorilla blood sample was of limited quantity and thus only subjected to leukocyte filtration with no reticulocyte enrichment performed. Human control samples were diluted until the percentage of enriched reticulocytes matched those of the ape and monkey samples (0.5 – 4% by flow cytometry).

For binding assays, 800,000 red blood cells were incubated with 5 µg of recombinant protein in 100 µl of PBS with 1% (w/v) BSA at room temperature for one hour after gentle mixing. Cells were then washed with 180 µl PBS containing 1% BSA, spun at 4,000 x g for 1 minute, resuspended in 180 µl PBS containing 1% BSA, and incubated with a polyclonal rabbit antibody (1 µg) raised against the respective protein. Following incubation at room temperature for one hour, cells were washed again, and then incubated with a secondary (Alexa Fluor 647 labeled) chicken anti-rabbit antibody (1 µg). After a final wash, 100 µl of room temperature thiazole orange (TO) (Retic-Count Reagent; BD Biosciences) was added and incubated for 30 min. Supernatant was removed and cells were resuspended in 1.2 ml PBS immediately prior to analysis on an Acurri flow cytometer. Antibody only controls were run in parallel, omitting the protein in the first incubation but otherwise following the same protocol. Approximately 200,000 events were captured for each sample. Data were analyzed using FlowJo software, gating on erythrocytes, then single cells, then TO-positive (reticulocyte) or TO-negative (normocyte) populations. Binding signal for each sample was obtained for both reticulocyte and normocyte populations by subtracting the fluorescence value of the corresponding antibody-only control from the sample value. RBP binding experiments were performed three times per sample on different days, with the average of these values reported.

SI Figures

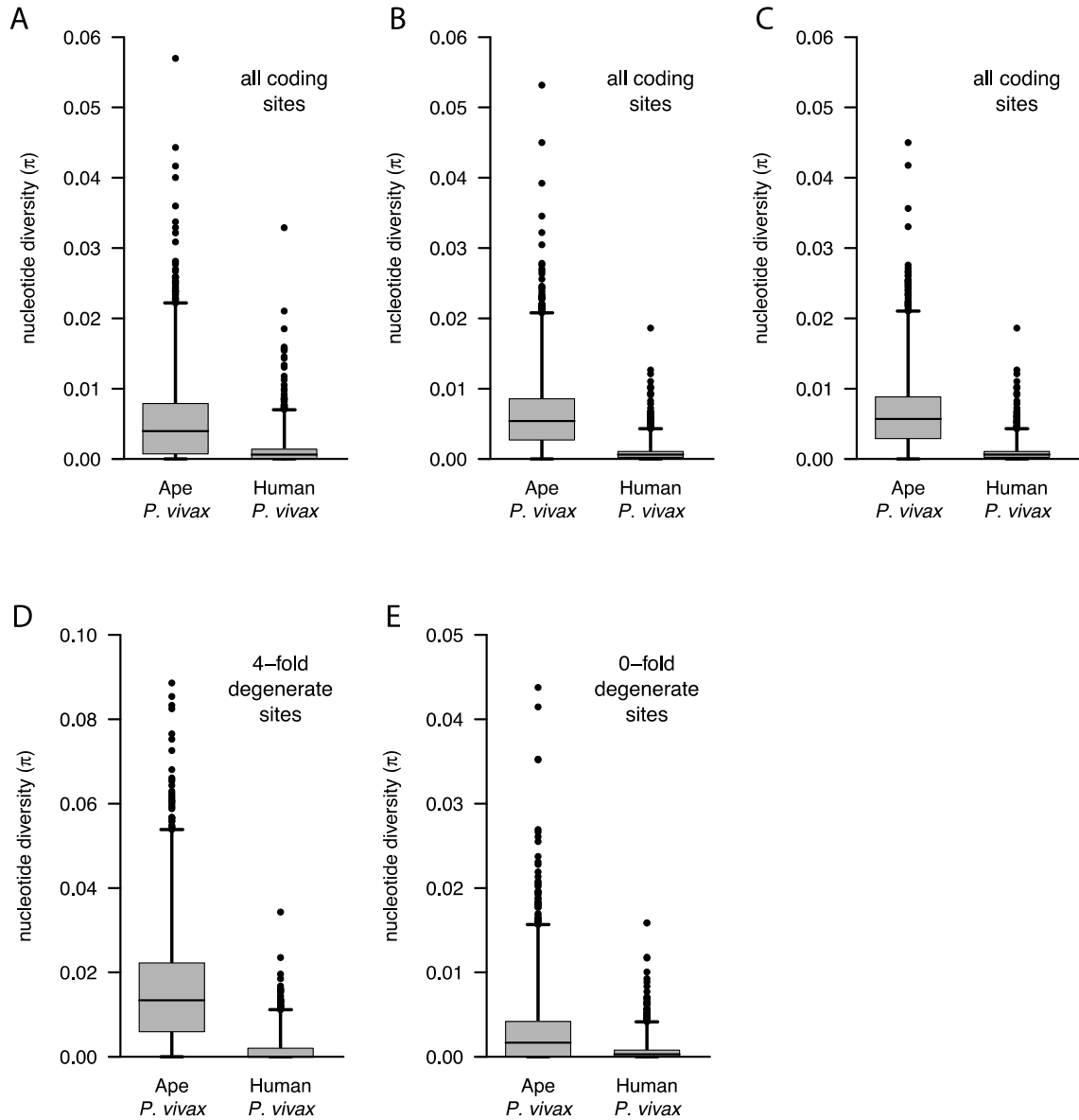


Figure S4-1. Sequence diversity between ape and human *P. vivax*. (A) Nucleotide sequence diversity (π) calculated for all coding sites in 3,955 genes of two chimpanzee (PvSY43 and PvSY56) and two human (PvSall and PvP01) *P. vivax* strains. (B) Nucleotide sequence diversity (π) for all coding sites in 4,259 genes of six chimpanzee and nine human *P. vivax* strains (Table S4-2). (C) Nucleotide sequence diversity (π) for all coding sites in 3,904 genes of five chimpanzee and nine human *P. vivax* strains, after removal of the multiply infected PvSY43 sample. (D, E) Nucleotide sequence diversity calculated for chimpanzee and human *P. vivax* strains as in (B), but considering only 4-fold degenerate sites (D, 3,837 genes) or 0-fold degenerate sites (E, 4,259 genes). For all plots, the interquartile range is shown as a box, with the upper and lower 99th percentiles indicated by whiskers (outliers are shown as black dots). For each alignment set, genes with fewer

than 35 aligned sites were excluded (2 genes in A, 3 genes in B and E, 3 genes in C, 425 genes in D) to avoid plotting spurious extreme values from very short sequences (these genes were retained in Table 4-2).

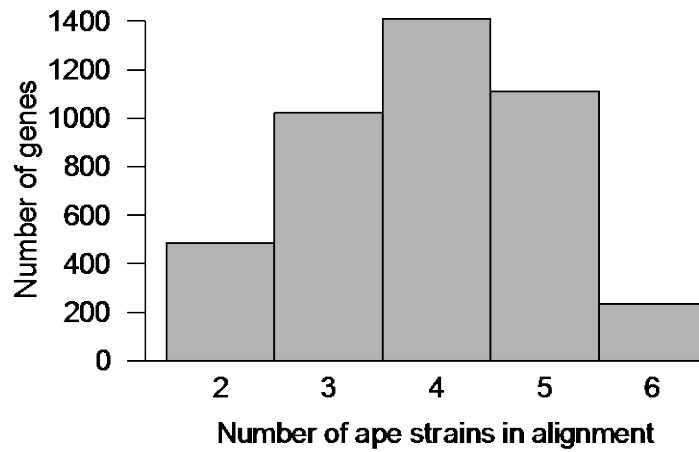


Figure S4-2. Chimpanzee *P. vivax* core genes used for diversity analyses. The number of genes (y-axis) that could be compared for two or more chimpanzee *P. vivax* strains (x-axis) is indicated. PvSY56 and PvSY43 genes were included if they covered 90% or more of the length of the corresponding ortholog in the human PvP01 reference. Genes from the other chimpanzee parasites were included if the number of callable sites covered at least 60% of sites in the corresponding PvSY56 gene (with ≥ 5 reads mapped).

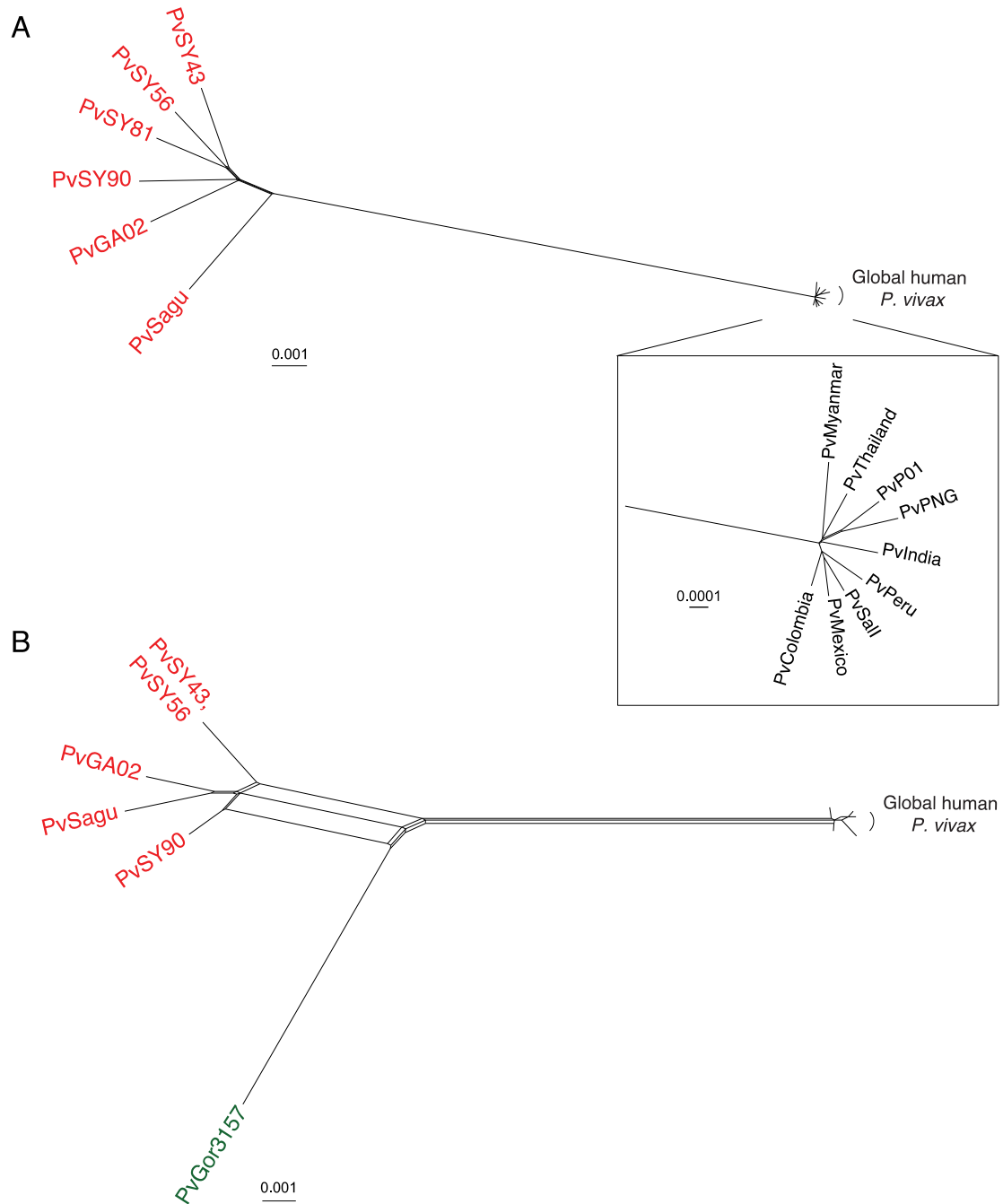


Figure S4-3. Phylogenetic network analysis of ape and human *P. vivax* strains. (A) A phylogenetic network was constructed using split decomposition from pairwise distances in an alignment of 232 nuclear genes. Nine human (black) and six chimpanzee (red) *P. vivax* strains are shown (the inset shows the human *P. vivax* strains in greater detail). The topology confirms the overall relationships of human and chimpanzee *P. vivax* depicted in Figure 4-2A. (B) As in (A), but based on 6 nuclear genes with coverage in a gorilla *P. vivax* strain (green). The same human and chimpanzee *P. vivax* strains were

included except for PvSY81, which did not cover these genes. The network suggests that some recombination has occurred between gorilla and ape parasites in these genes, but supports the overall topology shown in Figure 4-2B.

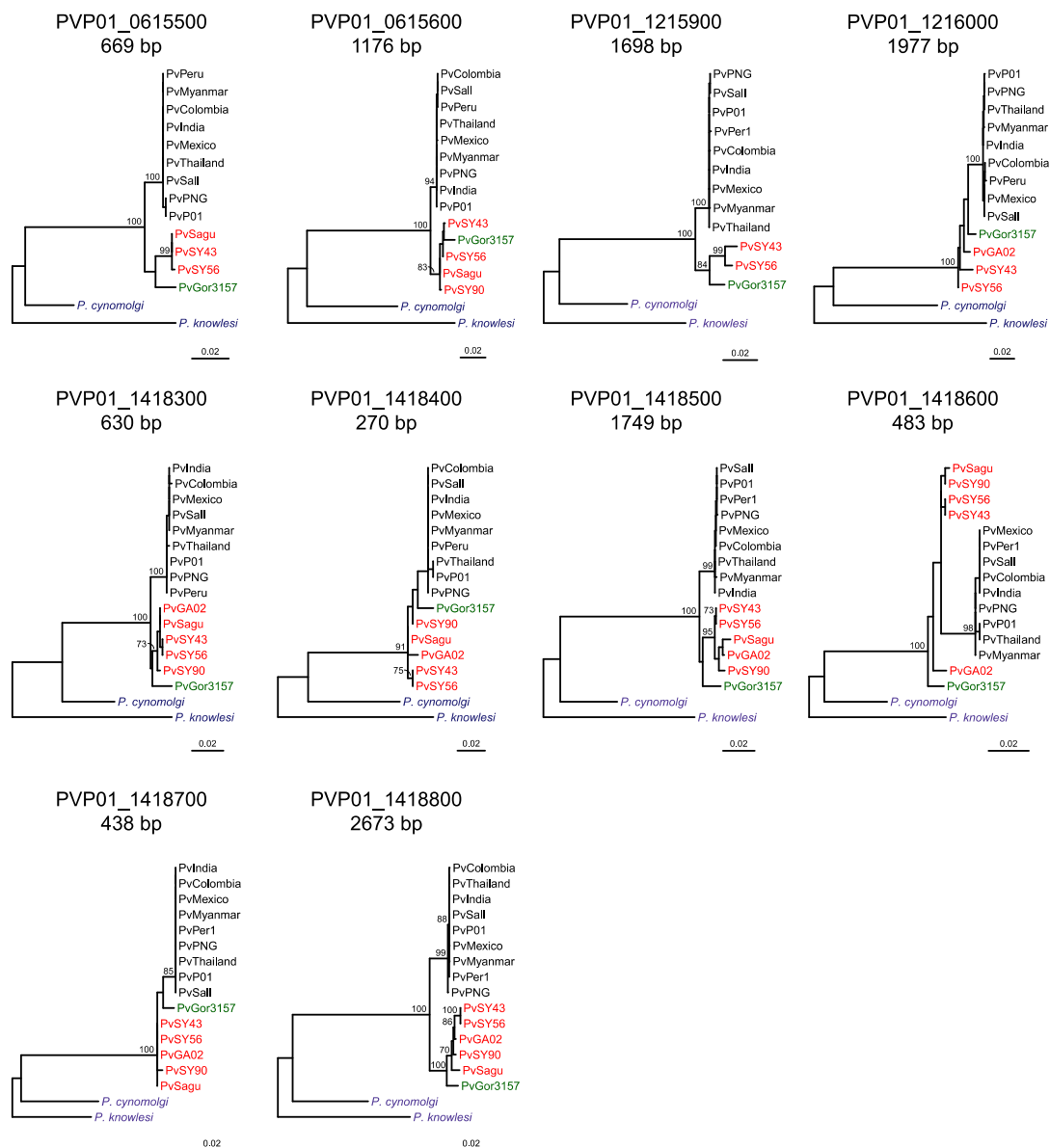


Figure S4-4. Phylogenetic relationships of ape and human *P. vivax* using whole genome and SNP data. Maximum likelihood trees, rooted with *P. knowlesi*, are shown for 10 nuclear genes (gene names and lengths for the PvP01 reference are indicated). *P. vivax* sequences from humans, chimpanzees and gorillas are shown in black, red and green, respectively. The monkey parasite species *P. cynomolgi* (strain M) and *P. knowlesi* (strain H) were included as outgroups (purple). Bootstrap values ≥ 70 are shown. The scale bar represents 0.02 substitutions per site.

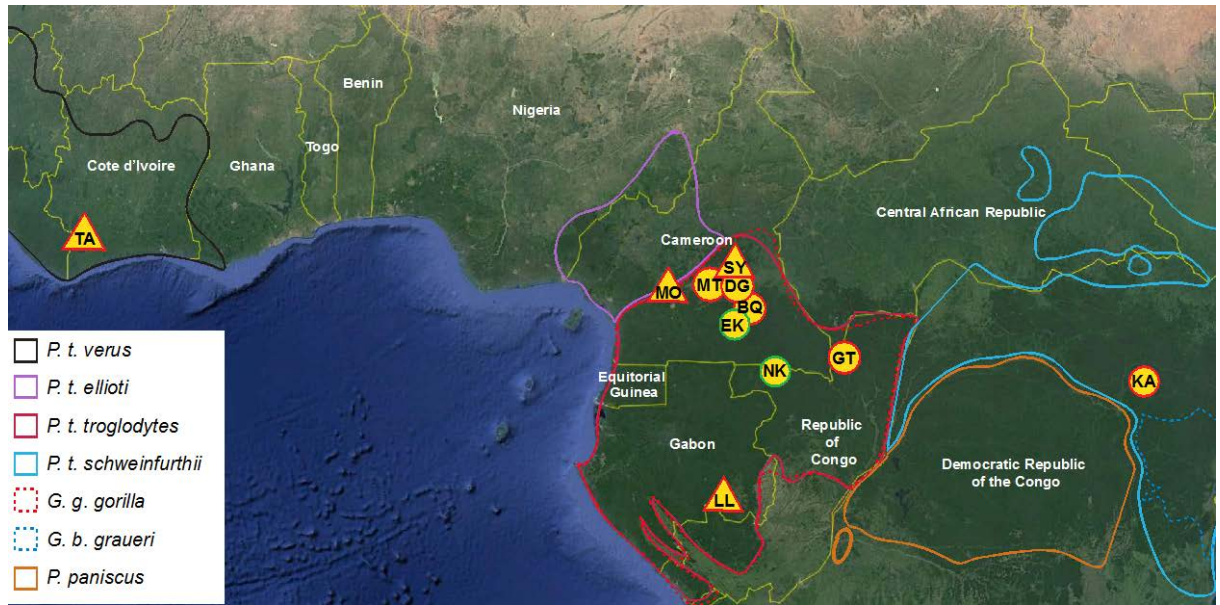


Figure S4-5. Geographical origin of *P. vivax*-positive ape samples. Study sites are shown in relation to the natural ranges of chimpanzees (*Pan troglodytes verus*, black; *P. t. ellioti*, purple; *P. t. troglodytes*, red; *P. t. schweinfurthii*, blue), bonobos (*Pan paniscus*, orange), western lowland gorillas (*Gorilla gorilla gorilla*, dashed red) and eastern lowland gorillas (*Gorilla beringei graueri*, dashed blue) (Caldecott and Miles, 2005). Circles denote forest sites, while triangles indicate the location of sanctuaries, with colors denoting whether chimpanzee (red) or gorilla (green) *P. vivax* sequences were obtained (BQ, Belgique; DG, Diang; EK, E'kom; GT, Goualougo Triangle; KA, Kabuka; MO, Mfou National Park Wildlife Rescue Center; MT, Minta; NK, Ndongo; SY, Sanaga Yong Chimpanzee Rescue Center; TA, Tai Forest). The chimpanzee infected with PvGA02 was previously reported (Rutledge et al., 2017) to have been sampled in Park of La Lékédi (LL).

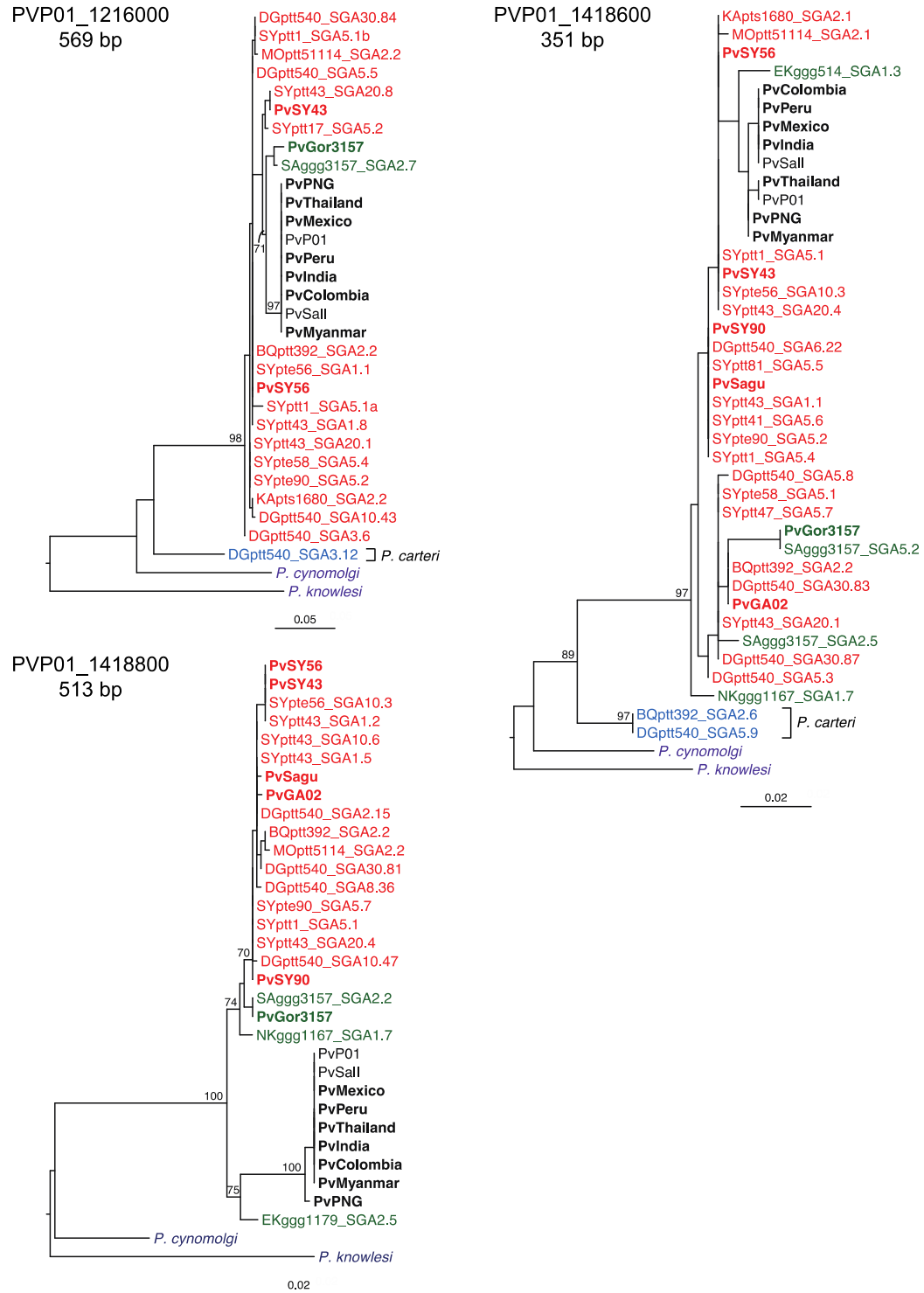
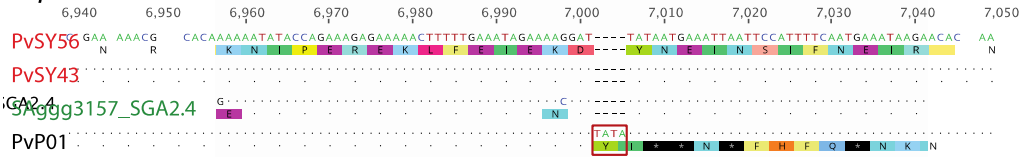


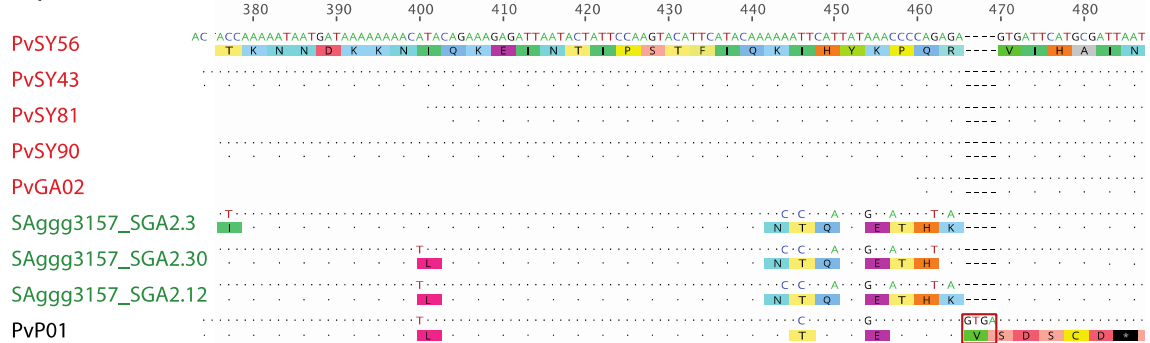
Figure S4-6. Phylogenetic relationships of ape and human *P. vivax*. Maximum likelihood trees rooted with *P. knowlesi* are shown for fragments of three nuclear genes (PVP01_1216000, PVP01_1418600, PVP01_1418800; the fragment size in PvP01 is

indicated). *P. vivax* sequences from humans, chimpanzees and gorillas are shown in black, red and green, respectively. Sequences generated by SWGA or derived from published data (see Table S4-2) are shown in bold. SGA-amplified sequences from ape fecal and blood samples include a two-letter code to denote the field site (Figure S4-5), lower case letters to indicate their species origin (ptt: *P. t. troglodytes*, red; pte: *P. t. ellioti*, red; pts: *P. t. schweinfurthii*, red; ggg: *G. g. gorilla*, green), a the sample number, as well as SGA dilution and well position (e.g., EKggg1179_SGA2.5 represents an SGA derived sequence amplified from a 1:2 dilution of fecal DNA from a western lowland gorilla sample 1179 and identified at position 5 of a plate of multiple PCR reactions). Sequences of *P. carteri*, a parasite species closely related to *P. vivax* that has thus far only been identified in wild-living chimpanzees, are shown in blue. GenBank accession numbers for SGA sequences are listed in Table S4-4. The monkey parasite species *P. cynomolgi* (strain M) and *P. knowlesi* (strain H) were included as outgroups (purple). Bootstrap values ≥ 70 are shown for clades containing at least two non-identical tips. The scale bar represents 0.02 substitutions per site.

rbp2d: ancestral frameshift mutation



rbp2e: ancestral frameshift mutation



rbp2e: ancestral stop mutation



rbp3: ancestral stop mutation

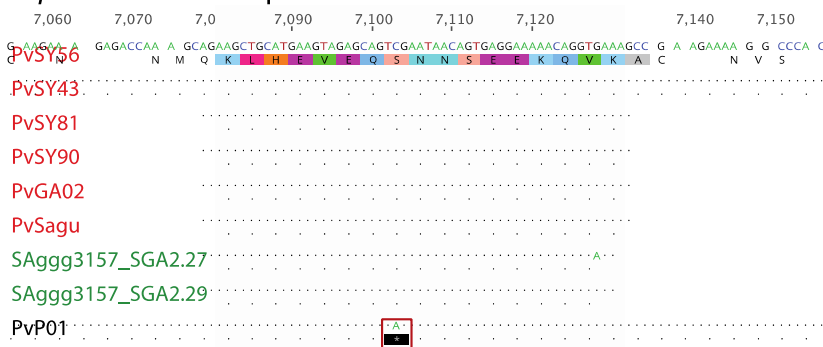


Figure S4-7. Ape *P. vivax* encode three intact *rbp* genes that are pseudogenized in human *P. vivax*. Alignments are shown for chimpanzee (red), gorilla (green) and human (black) *P. vivax* *rbp2d*, *rbp2e* and *rbp3* gene sequences (numbers indicate the nucleotide position within the respective gene). 'Pv' denotes sequences from genome-wide analyses (Table S4-2) which were generated by mapping sequencing reads to the PvSY56 reference and calling bases covered by ≥ 3 reads. All gorilla *P. vivax* sequences were derived by single genome amplification (SGA) from the same multiply infected sample

(SAGgg3157) and represent individual parasites. Nucleotides that differ from the PvSY56 reference are highlighted, with dots indicating sequence identity. Inactivating mutations that cause frameshift and premature stop codons in all published human *P. vivax* strains are boxed.

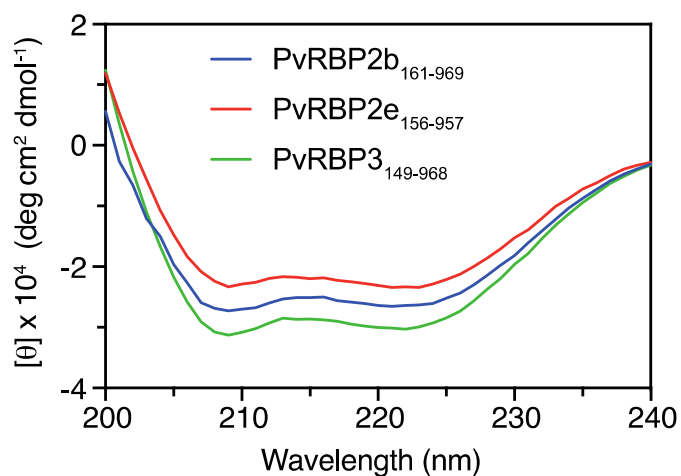


Figure S4-8. Circular dichroism spectra of recombinant RBP proteins from chimpanzee *P. vivax* strains. Far UV circular dichroism spectra of two newly expressed chimpanzee *P. vivax* proteins PvRBP2e₁₅₆₋₉₅₇ (red) and PvRBP3₁₄₉₋₉₆₈ (green) are compared to the CD spectrum of the human *P. vivax* PvRBP2b₁₆₁₋₉₆₉ protein (blue) as previously reported (Gruszczuk et al., 2018b). The mean residue molar ellipticity (θ , y-axis) is plotted relative to the wavelength (nm; x-axis). The three spectra superimpose, suggesting proper folding of the newly expressed proteins.

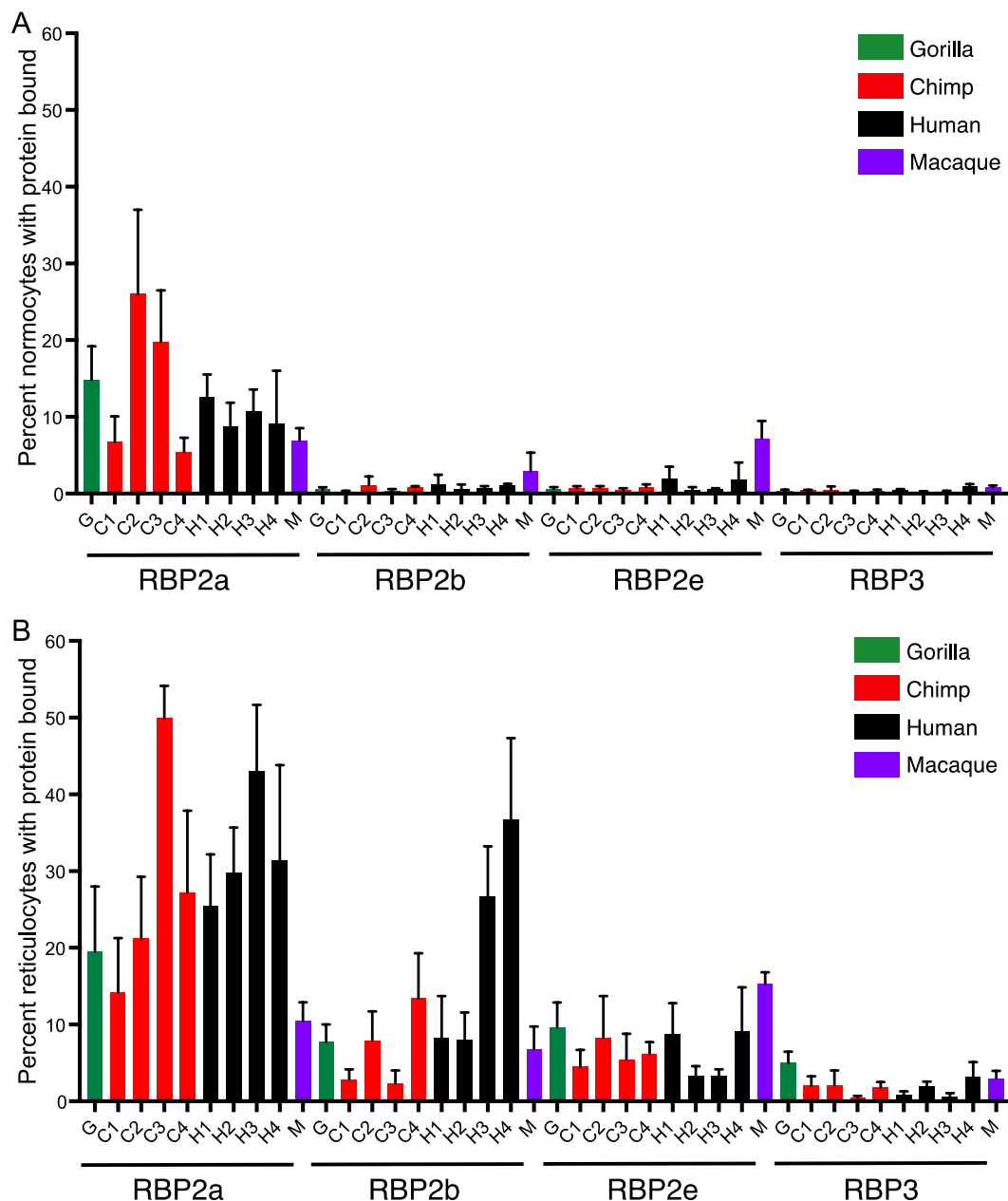


Figure S4-9. Binding of chimpanzee and human *P. vivax* RBP proteins to red blood cells from different host species. The binding of human *P. vivax* RBP2a and RBP2b and chimpanzee *P. vivax* RBP2e and RBP3 proteins to normocytes (A) and reticulocytes (B) from one gorilla (G), four chimpanzees (C1-C4), four humans (H1-H4), and one macaque (M) is shown. Columns indicate the percentage of red blood cells (RBCs) that bound the respective RBP proteins after subtracting background binding in the absence of protein. Error bars represent one standard deviation. All red blood cell preparations were tested three times in independent technical replicates.

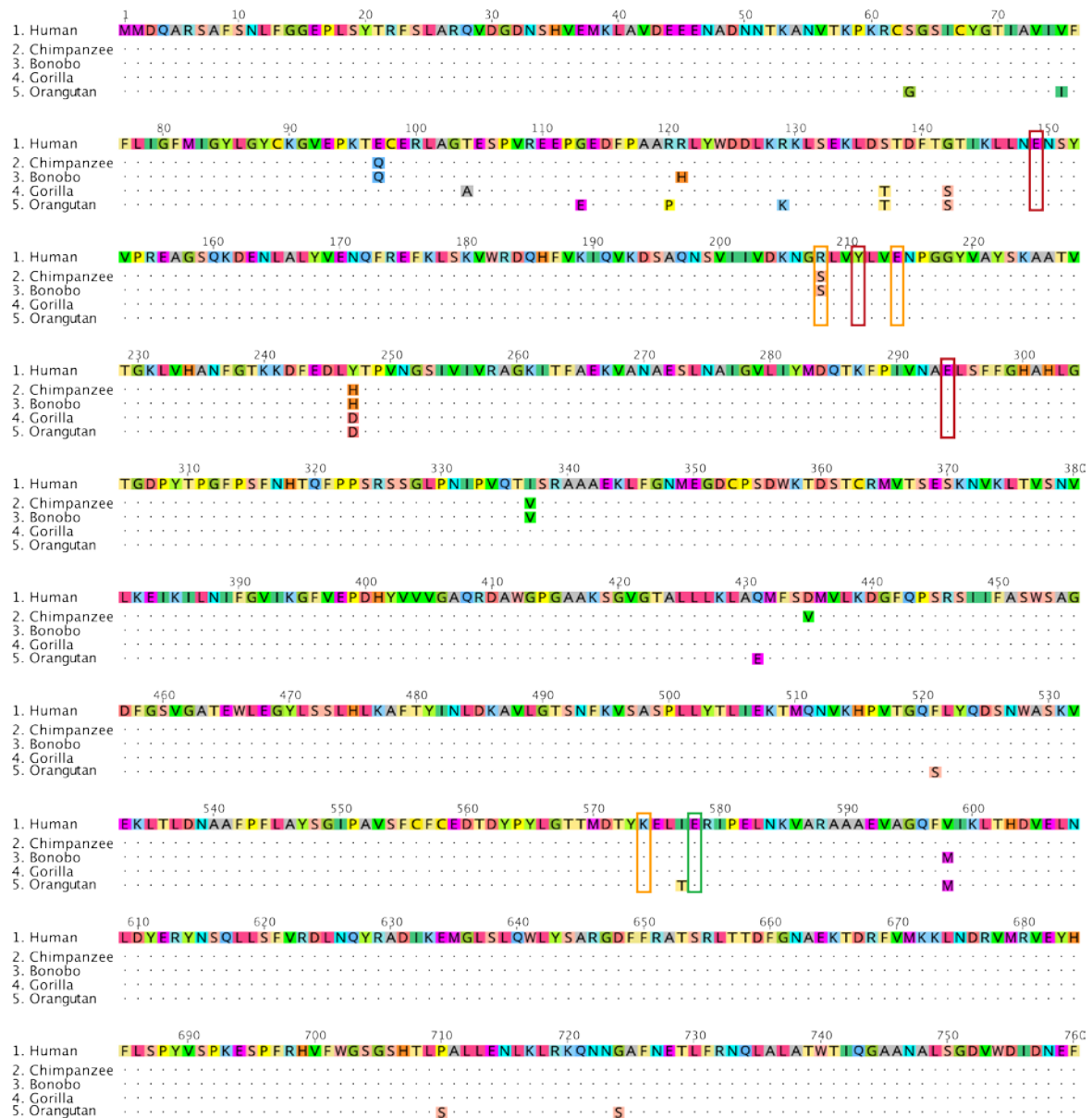


Figure S4-10. Alignment of human and ape transferrin receptor 1 (TfR1) protein sequences. An alignment of the deduced amino acid sequences of human (GRCh38, NC_000003.12), chimpanzee (*Pan troglodytes verus*, XM_003310191.3), bonobo (*Pan paniscus*, XM_003806407.2), gorilla (*Gorilla gorilla gorilla*, XM_004038250.2), and orangutan (*Pongo abelii*, NM_001131591.1) TfR1 sequences is shown. Residues that differ from the human sequence are highlighted, with dots indicating sequence identity. Alanine mutagenesis of TfR1 residues that were identified to form stacking interactions or salt bridges with RBP2b are boxed (Gruszczyk et al., 2018b), with border color indicating the degree to which the mutation abrogated invasion complex formation (no effect, green; moderate effect, orange; severe effect, red).

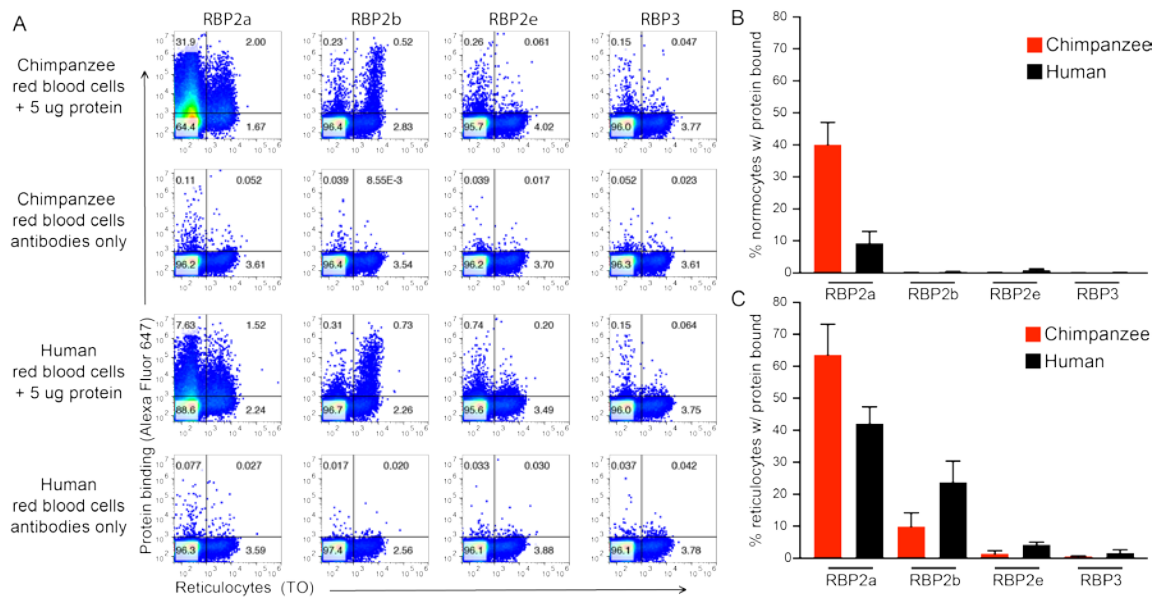


Figure S4-11. Binding of chimpanzee and human *P. vivax* RBP proteins to reticulocyte-enriched chimpanzee and human red blood cells. (A) Dot plots are shown that depict the binding of human *P. vivax* RBP2a and RBP2b and chimpanzee *P. vivax* RBP2e and RBP3 proteins to chimpanzee (first row) and human (third row) red blood cells, respectively. Both the chimpanzee and the human blood sample contained a large fraction of reticulocytes (the human sample was diluted so that equivalent numbers of reticulocytes were tested in the binding assays). The x-axis depicts thiazole orange (TO) staining of reticulocytes; the y-axis indicates protein binding as detected using an RBP-specific polyclonal rabbit antibody, followed by a secondary chicken (Alexa 647 labeled) anti-rabbit antibody. The position of gates denoting normocytes versus reticulocytes as well as protein binding versus no protein binding are shown by vertical and horizontal lines, respectively. Numbers indicate the percentage of total cells within the respective gate. Antibody-only negative controls in which no protein was added are shown in the second (chimpanzee cells) and fourth (human cells) rows, respectively. (B, C) The percentage of chimpanzee (red) and human (black) normocytes (B) and reticulocytes (C) that bound the respective RBP proteins are shown. Three independent replicates were performed and background signal from the antibody-only control was subtracted (see SI Appendix, Materials and Methods).

Table S4-1. Select whole genome amplification and sequencing of ape *P. vivax* genomes

Blood Sample*	Digest†	pvset1#	pvset2#	pvset3#	pvset4#	pvset5#	pvset6#	Illumina‡	PacBio§
SY43	+	+				+		+	+
SY56	+	+	+	+		+	+	+	+
SY81	+	+				+	+	+	
SY90	+	+				+	+	+	
Sagu		+		+		+	+	+	
Gor3157	+	+			+	+	+	+	

**P. vivax* genomes were amplified by SWGA from whole blood DNA of four sanctuary chimpanzees (SY43, SY56, SY81, SY90) housed at the Sanaga Yong (SY) Chimpanzee Rescue Center, one wild-living habituated chimpanzee (Sagu) from the Tai forest, and one western lowland gorilla bushmeat sample (Gor3157) confiscated by the Cameroonian Ministry of Environment and Forestry.

#pvset, set of *P. vivax*-specific SWGA primers (see text for primer sequences).

†Aliquots of whole blood DNA were digested with the methylation dependent restriction enzymes MspJI and FspEI prior to SWGA to selectively degrade host DNA (Sundararaman et al., 2016).

‡Illumina sequencing was performed on MiSeq or MiniSeq platforms.

§PacBio sequencing was performed on a PacBio RS II platform.

Table S4-2. Host species and geographic origin of ape and human *P. vivax* isolates

<i>P. vivax</i> strain	Host species*	Country	Genes analyzed	Reference
PvSY43	<i>P.t.t.</i>	Cameroon	3,973	This study
PvSY56	<i>P.t.e.</i>	Cameroon	4,262	This study
PvSY81	<i>P.t.t.</i>	Cameroon	695	This study
PvSY90	<i>P.t.e.</i>	Cameroon	2,159	This study
PvSagu	<i>P.t.v.</i>	Cote d'Ivoire	2,542	This study
PvGor3157	<i>G.g.g.</i>	Cameroon	10	This study
PvGA02 [#]	<i>P.t.t.</i>	Gabon	3,005	Rutledge et al., 2017
PvColombia	human	Colombia	4,257	Hupalo et al., 2016
PvIndia	human	India	4,258	Hupalo et al., 2016
PvMexico	human	Mexico	4,258	Hupalo et al., 2016
PvMyanmar	human	Myanmar	4,258	Hupalo et al., 2016
PvPeru	human	Peru	4,259	Hupalo et al., 2016
PvPNG	human	Papua New Guinea	4,260	Hupalo et al., 2016
PvThailand	human	Thailand	4,254	Hupalo et al., 2016
PvSall	human	El Salvador	4,241	Carlton et al., 2008
PvP01	human	Indonesia	4,262	Auburn et al., 2016

**P.t.t.*, *Pan troglodytes troglodytes*; *P.t.e.*, *P. t. ellioti*; *P.t.v.*, *P. t. verus*; *G.g.g.*, *Gorilla gorilla gorilla*;

[#]PvGA02 sequences were derived from a read database generated from the blood of a sanctuary chimpanzee (GA02) that was coinfectd with ape *P. malariae* and *P. vivax* (Rutledge et al., 2017).

Table S4-3. Polymorphisms between chimpanzee and human *P. vivax* and *P. cynomolgi*

	Polymorphisms in <i>P. vivax</i>			Fixed differences from <i>P. cynomolgi</i>			
	NS*	S*	NS/S	NS†	S†	NS/S	NI‡
Chimpanzee <i>P. vivax</i> [#]	27,209	41,177	0.66	266,011	383,274	0.69	0.95
Human <i>P. vivax</i> [#]	8,858	6,715	1.32	276,851	400,376	0.69	1.91

*The number of nonsynonymous (NS) and synonymous (S) polymorphisms was calculated by counting the number of SNPs that changed (NS) or did not change (S) the amino acid sequence of PvSY56 (ape *P. vivax*) or PvP01 (human *P. vivax*).

†The number of nonsynonymous (NS) and synonymous (S) differences was calculated between *P. cynomolgi* strain M and PvSY56 (ape *P. vivax*) or PvP01 (human *P. vivax*); sites that were polymorphic in ape or human *P. vivax* were excluded.

‡NI, neutrality index.

#A common set of 3,912 genes was compared among six chimpanzee and nine human *P. vivax* strains as well as between these parasites and the *P. cynomolgi* strain M.

4.8 Chapter references

- Abascal, F., Zardoya, R., Telford, M.J., 2010. TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* 38, W7–W13.
- Assefa, S., Keane, T.M., Otto, T.D., Newbold, C., Berriman, M., 2009. ABACAS: Algorithm-based automatic contiguation of assembled sequences. *Bioinformatics* 25, 1968–1969.
- Auburn, S., Böhme, U., Steinbiss, S., Trimarsanto, H., Hostetler, J., Sanders, M., Gao, Q., Nosten, F., Newbold, C.I., Berriman, M., Price, R.N., Otto, T.D., 2016. A new *Plasmodium vivax* reference sequence with improved assembly of the subtelomeres reveals an abundance of *pir* genes. *Wellcome Open Res.* 1:4.
- Auweru, G.A. Van Der, Carneiro, M.O., Hartl, C., Poplin, R., Levy-moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., Depristo, M.A., 2014. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 43, 11.10.1-11.10.33.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477.
- Brasil, P., Zalis, M.G., de Pina-Costa, A., Siqueira, A.M., Júnior, C.B., Silva, S., Areas, A.L.L., Pelajo-Machado, M., de Alvarenga, D.A.M., da Silva Santelli, A.C.F., Albuquerque, H.G., Cravo, P., Santos de Abreu, F.V., Peterka, C.L., Zanini, G.M., Suárez Mutis, M.C., Pissinatti, A., Lourenço-de-Oliveira, R., de Brito,

- C.F.A., de Fátima Ferreira-da-Cruz, M., Culleton, R., Daniel-Ribeiro, C.T., 2017. 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.
- Brock, P.M., Fornace, K.M., Parmiter, M., Cox, J., Drakeley, C.J., Ferguson, H.M., Kao, R.R., 2016. *Plasmodium knowlesi* transmission: Integrating quantitative approaches from epidemiology and ecology to understand malaria as a zoonosis. *Parasitology* 143, 389–400.
- Buery, J.C., Rodrigues, P.T., Natal, L., Salla, L.C., Loss, A.C., Vicente, C.R., Rezende, H.R., Duarte, A.M.R.D.C., Fux, B., Malafronte, R.D.S., Falqueto, A., Cerutti, C., 2017. Mitochondrial genome of *Plasmodium vivax/simium* detected in an endemic region for malaria in the Atlantic Forest of Espírito Santo state, Brazil: Do mosquitoes, simians and humans harbour the same parasite? *Malar. J.* 16, 437.
- Caldecott, J., Miles, L., 2005. *World Atlas of Great Apes and their Conservation*. Univ of California Press.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: Architecture and applications. *BMC Bioinformatics* 10, 1–9.
- Carlton, J.M., Adams, J.H., Silva, J.C., Bidwell, S.L., Lorenzi, H., Caler, E., Crabtree, J., Angiuoli, S. V., Merino, E.F., Amedeo, P., Cheng, Q., Coulson, R.M.R., Crabb, B.S., Del Portillo, H.A., Essien, K., Feldblyum, T. V., Fernandez-Becerra, C., Gilson, P.R., Gueye, A.H., Guo, X., Kang'A, S., Kooij, T.W.A., Korsinczky, M., Meyer, E.V.S., Nene, V., Paulsen, I., White, O., Ralph, S.A., Ren, Q., Sargeant, T.J., Salzberg, S.L., Stoeckert, C.J., Sullivan, S.A., Yamamoto, M.M., Hoffman, S.L., Wortman, J.R., Gardner, M.J., Galinski, M.R., Barnwell, J.W., Fraser-Liggett, C.M., 2008. Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature* 455, 757–763.
- Carter, R., 2003. Speculations on the origins of *Plasmodium vivax* malaria. *Trends Parasitol.* 19, 214–219.
- Chaisson, M.J., Tesler, G., 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): Application and theory. *BMC Bioinformatics* 13, 238.
- Chang, H.-H., Moss, E.L., Park, D.J., Ndiaye, D., Mboup, S., Volkman, S.K., Sabeti, P.C., Wirth, D.F., Neafsey, D.E., Hartl, D.L., 2013. Malaria life cycle intensifies both natural selection and random genetic drift. *Proc. Natl. Acad. Sci.* 110, 20129–20134.
- Chang, H.-H., Park, D.J., Galinsky, K.J., Schaffner, S.F., Ndiaye, D., Ndir, O., Mboup, S., Wiegand, R.C., Volkman, S.K., Sabeti, P.C., Wirth, D.F., Neafsey, D.E., Hartl, D.L., 2012. Genomic sequencing of *Plasmodium falciparum* malaria parasites from Senegal reveals the demographic history of the population. *Mol. Biol. Evol.*

- Chuquiyauri, R., Molina, D.M., Moss, E.L., Wang, R., Gardner, M.J., Brouwer, K.C., Torres, S., Gilman, R.H., Llanos-Cuentas, A., Neafsey, D.E., Felgner, P., Liang, X., Vinetz, J.M., 2015. Genome-scale protein microarray comparison of human antibody responses in *Plasmodium vivax* relapse and reinfection. *Am. J. Trop. Med. Hyg.* 93, 801–809.
- Clarke, E.L., Sundararaman, S.A., Seifert, S.N., Bushman, F.D., Hahn, B.H., Brisson, D., 2017. *swga*: a primer design toolkit for selective whole genome amplification. *Bioinformatics* 33, 2071–2077.
- Cornejo, O.E., Escalante, A.A., 2006. The origin and age of *Plasmodium vivax*. *Trends Parasitol.* 22, 558–563.
- Cowell, A.N., Loy, D.E., Sundararaman, S.A., Valdivia, H., Fisch, K., Lescano, A.G., Baldeviano, G.C., Durand, S., Gerbasi, V., Sutherland, C.J., Nolder, D., Vinetz, J.M., Hahn, B.H., Winzeler, E.A., 2017. Selective whole-genome amplification is a robust method that enables scalable whole-genome sequencing of *Plasmodium vivax* from unprocessed clinical samples. *MBio* 8, e02257.
- Cowman, A.F., Tonkin, C.J., Tham, W.H., Duraisingh, M.T., 2017. The molecular basis of erythrocyte invasion by malaria parasites. *Cell Host Microbe* 22, 232–245.
- Culleton, R., Coban, C., Zeyrek, F.Y., Cravo, P., Kaneko, A., Randrianarivelojosia, M., Andrianaranjaka, V., Kano, S., Farnert, A., Arez, A.P., Sharp, P.M., Carter, R., Tanabe, K., 2011. The origins of African *Plasmodium vivax*; insights from mitochondrial genome sequencing. *PLoS One* 6, e29137.
- Faust, C., Dobson, A.P., 2015. Primate malarias: Diversity, distribution and insights for zoonotic *Plasmodium*. *One Health* 1, 66–75.
- Gelabert, P., Sandoval-Velasco, M., Olalde, I., Fregel, R., Rieux, A., Escosa, R., Aranda, C., Paaijmans, K., Mueller, I., Gilbert, M.T.P., Lalueza-Fox, C., 2016. Mitochondrial DNA from the eradicated European *Plasmodium vivax* and *P. falciparum* from 70-year-old slides from the Ebro Delta in Spain. *Proc. Natl. Acad. Sci.* 113, 11495–11500.
- Gruszczyk, J., Huang, R.K., Chan, L.-J., Menant, S., Hong, C., Murphy, J.M., Mok, Y.-F., Griffin, M.D.W., Pearson, R.D., Wong, W., Cowman, A.F., Yu, Z., Tham, W.-H., 2018a. Cryo-EM structure of an essential *Plasmodium vivax* invasion complex. *Nature* in press.
- Gruszczyk, J., Kanjee, U., Chan, L.-J., Menant, S., Malleret, B., Lim, N.T.Y., Schmidt, C.Q., Mok, Y.-F., Lin, K.-M., Pearson, R.D., Rangel, G., Smith, B.J., Call, M.J., Weekes, M.P., Griffin, M.D.W., Murphy, J.M., Abraham, J., Sriprawat, K., Menezes, M.J., Ferreira, M.U., Russell, B., Renia, L., Duraisingh, M.T., Tham, W.-H., 2018b. Transferrin receptor 1 is a reticulocyte-specific receptor for *Plasmodium vivax*. *Science*. 359, 48–55.
- Gruszczyk, J., Lim, N.T.Y., Arnott, A., He, W.-Q., Nguitragool, W., Roobsoong, W., Mok,

- Y.-F., Murphy, J.M., Smith, K.R., Lee, S., Bahlo, M., Mueller, I., Barry, A.E., Tham, W.-H., 2016. Structurally conserved erythrocyte-binding domain in *Plasmodium* provides a versatile scaffold for alternate receptor engagement. *Proc. Natl. Acad. Sci.* 113, E191–E200.
- Hackl, T., Hedrich, R., Schultz, J., Förster, F., 2014. Proovread: Large-scale high-accuracy PacBio correction through iterative short read consensus. *Bioinformatics* 30, 3004–3011.
- Hester, J., Chan, E.R., Menard, D., Mercereau-Puijalon, O., Barnwell, J., Zimmerman, P.A., Serre, D., 2013. *De novo* assembly of a field isolate genome reveals novel *Plasmodium vivax* erythrocyte invasion genes. *PLoS Negl. Trop. Dis.* 7, e2569.
- Hietanen, J., Chim-Ong, A., Chiramanewong, T., Gruszczyk, J., Roobsoong, W., Tham, W.H., Sattabongkot, J., Nguitragool, W., 2016. Gene models, expression repertoire, and immune response of *Plasmodium vivax* reticulocyte binding proteins. *Infect. Immun.* 84, 677–685.
- Howes, R.E., Patil, A.P., Piel, F.B., Nyangiri, O.A., Kabaria, C.W., Gething, P.W., Zimmerman, P.A., Barnadas, C., Beall, C.M., Gebremedhin, A., Ménard, D., Williams, T.N., Weatherall, D.J., Hay, S.I., 2011. The global distribution of the Duffy blood group. *Nat. Commun.* 2, 226.
- Hupalo, D.N., Luo, Z., Melnikov, A., Sutton, P.L., Rogov, P., Escalante, A., Vallejo, A.F., Herrera, S., Arévalo-Herrera, M., Fan, Q., Wang, Y., Cui, L., Lucas, C.M., Durand, S., Sanchez, J.F., Christian Baldeviano, G., Lescano, A.G., Laman, M., Barnadas, C., Barry, A., Mueller, I., Kazura, J.W., Eapen, A., Kanagaraj, D., Valecha, N., Ferreira, M.U., Roobsoong, W., Nguitragool, W., Sattabonkot, J., Gamboa, D., Kosek, M., Vinetz, J.M., González-Cerón, L., Birren, B.W., Neafsey, D.E., Carlton, J.M., 2016. Population genomics studies identify signatures of global dispersal and drug resistance in *Plasmodium vivax*. *Nat. Genet.* 48, 953–958.
- Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267.
- Kaiser, M., Löwa, A., Ulrich, M., Ellerbrok, H., Goffe, A.S., Blasse, A., Zommers, Z., Couacy-Hymann, E., Babweteera, F., Zuberbühler, K., Metzger, S., Geidel, S., Boesch, C., Gillespie, T.R., Leendertz, F.H., 2010. Wild chimpanzees infected with 5 *Plasmodium* species. *Emerg. Infect. Dis.* 16, 1956–1959.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649.
- Keightley, P.D., Jackson, B.C., 2018. Inferring the probability of the derived versus the ancestral allelic state at a polymorphic site. *Genetics* early online doi: 10.1534/genetics.118.301120

- Köndgen, S., Leider, M., Lankester, F., Bethe, A., Lübke-Becker, A., Leendertz, F.H., Ewers, C., 2011. *Pasteurella multocida* involved in respiratory disease of wild chimpanzees. PLoS One 6, e24236.
- Leichty, A.R., Brisson, D., 2014. Selective whole genome amplification for resequencing target microbial species from complex natural samples. Genetics 198, 473–481.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.
- Liu, W., Li, Y., Learn, G.H., Rudicell, R.S., Robertson, J.D., Keele, B.F., Ndjango, J.B.N., Sanz, C.M., Morgan, D.B., Locatelli, S., Gonder, M.K., Kranzusch, P.J., Walsh, P.D., Delaporte, E., Mpoudi-Ngole, E., Georgiev, A. V., Muller, M.N., Shaw, G.M., Peeters, M., Sharp, P.M., Rayner, J.C., Hahn, B.H., 2010. Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. Nature 467, 420–425.
- Liu, W., Li, Y., Shaw, K.S., Learn, G.H., Plenderleith, L.J., Malenke, J.A., Sundararaman, S.A., Ramirez, M.A., Crystal, P.A., Smith, A.G., Bibollet-Ruche, F., Ayoub, A., Locatelli, S., Esteban, A., Mouacha, F., Guichet, E., Butel, C., Ahuka-Mundeki, S., Inogwabini, B.I., Ndjango, J.B.N., Speede, S., Sanz, C.M., Morgan, D.B., Gonder, M.K., Kranzusch, P.J., Walsh, P.D., Georgiev, A. V., Muller, M.N., Piel, A.K., Stewart, F.A., Wilson, M.L., Pusey, A.E., Cui, L., Wang, Z., Färnert, A., Sutherland, C.J., Nolder, D., Hart, J.A., Hart, T.B., Bertolani, P., Gillis, A., LeBreton, M., Tafon, B., Kiyang, J., Djoko, C.F., Schneider, B.S., Wolfe, N.D., Mpoudi-Ngole, E., Delaporte, E., Carter, R., Culleton, R.L., Shaw, G.M., Rayner, J.C., Peeters, M., Hahn, B.H., Sharp, P.M., 2014. African origin of the malaria parasite *Plasmodium vivax*. Nat. Commun. 5, 3346.
- Liu, W., Sherrill-Mix, S., Learn, G.H., Scully, E.J., Li, Y., Avitto, A.N., Loy, D.E., Lauder, A.P., Sundararaman, S.A., Plenderleith, L.J., Ndjango, J.B.N., Georgiev, A. V., Ahuka-Mundeki, S., Peeters, M., Bertolani, P., Dupain, J., Garai, C., Hart, J.A., Hart, T.B., Shaw, G.M., Sharp, P.M., Hahn, B.H., 2017. Wild bonobos host geographically restricted malaria parasites including a putative new *Laverania* species. Nat. Commun. 8, 1685.
- Liu, W., Sundararaman, S.A., Loy, D.E., Learn, G.H., Li, Y., Plenderleith, L.J., Ndjango, J.B.N., Speede, S., Atencia, R., Cox, D., Shaw, G.M., Ayoub, A., Peeters, M., Rayner, J.C., Hahn, B.H., Sharp, P.M., 2016. Multigenomic delineation of *Plasmodium* species of the *Laverania* subgenus infecting wild-living chimpanzees and gorillas. Genome Biol. Evol. 8, 1929–1939.
- Loy, D.E., Liu, W., Li, Y., Learn, G.H., Plenderleith, L.J., Sundararaman, S.A., Sharp, P.M., Hahn, B.H., 2017. Out of Africa: origins and evolution of the human malaria parasites *Plasmodium falciparum* and *Plasmodium vivax*. Int. J. Parasitol. 47, 87–97.
- McDonald, J.H., Kreitman, M., 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature 351, 652–654.

- McManus, K.F., Taravella, A.M., Henn, B.M., Bustamante, C.D., Sikora, M., Cornejo, O.E., 2017. Population genetic analysis of the DARC locus (Duffy) reveals adaptation from standing variation associated with malaria resistance in humans. *PLoS Genet.* 13, 48–65.
- Miller, L.H., Mason, S.J., Clyde, D.F., McGinniss, M.H., 1976. The resistance factor to *Plasmodium vivax* in Blacks: The Duffy-Blood-Group Genotype, FyFy. *N. Engl. J. Med.* 295, 302–304.
- Moon, R.W., Sharaf, H., Hastings, C.H., Ho, Y.S., Nair, M.B., Rchiad, Z., Knuepfer, E., Ramaprasad, A., Mohring, F., Amir, A., Yusuf, N.A., Hall, J., Almond, N., Lau, Y.L., Pain, A., Blackman, M.J., Holder, A.A., 2016. Normocyte-binding protein required for human erythrocyte invasion by the zoonotic malaria parasite *Plasmodium knowlesi*. *Proc. Natl. Acad. Sci.* 113, 7231–7236.
- Mu, J., Joy, D.A., Duan, J., Huang, Y., Carlton, J., Walker, J., Barnwell, J., Beerli, P., Charleston, M.A., Pybus, O.G., Su, X.Z., 2005. Host switch leads to emergence of *Plasmodium vivax* malaria in humans. *Mol. Biol. Evol.* 22, 1686–1693.
- Nadalín, F., Vezzi, F., Policriti, A., 2012. GapFiller: A *de novo* assembly approach to fill the gap within paired reads. *BMC Bioinformatics* 13, S8.
- Neafsey, D.E., Galinsky, K., Jiang, R.H.Y., Young, L., Sykes, S.M., Saif, S., Gujja, S., Goldberg, J.M., Young, S., Zeng, Q., Chapman, S.B., Dash, A.P., Anvikar, A.R., Sutton, P.L., Birren, B.W., Escalante, A.A., Barnwell, J.W., Carlton, J.M., 2012. The malaria parasite *Plasmodium vivax* exhibits greater genetic diversity than *Plasmodium falciparum*. *Nat. Genet.* 44, 1046–1050.
- Otto, T.D., Dillon, G.P., Degraeve, W.S., Berriman, M., 2011. RATT: Rapid Annotation Transfer Tool. *Nucleic Acids Res.* 39, e57.
- Otto, T.D., Rayner, J.C., Böhme, U., Pain, A., Spottiswoode, N., Sanders, M., Quail, M., Ollomo, B., Renaud, F., Thomas, A.W., Prugnolle, F., Conway, D.J., Newbold, C., Berriman, M., 2014. Genome sequencing of chimpanzee malaria parasites reveals possible pathways of adaptation to human hosts. *Nat. Commun.* 5, 4754.
- Otto, T.D., Gilabert, A., Crellen, T., Böhme, U., Arnathau, C., Sanders, M., Oyola, S., Okauga, A.P., Boundenga, L., Guillaume, E., Ngoubangoye, B., Moukoudoum, N., Paupy, C., Durand, P., Rougeron, V., Ollomo, B., Renaud, F., Newbold, C., Berriman, M., Prugnolle, F., 2018. Genomes of an entire *Plasmodium* subgenus reveal paths to virulent human malaria. *Nat. Microbiol.* 3, 687–697.
- Pagani, L., Lawson, D.J., Jagoda, E., Mörseburg, A., Eriksson, A., Mitt, M., Clemente, F., Hudjashov, G., Degiorgio, M., Saag, L., Wall, J.D., Cardona, A., Mägi, R., Sayres, M.A.W., Kaewert, S., Inchley, C., Scheib, C.L., Järve, M., Karmin, M., Jacobs, G.S., Antao, T., Iliescu, F.M., Kushniarevich, A., Ayub, Q., Tyler-Smith, C., Xue, Y., Yunusbayev, B., Tambets, K., Mallick, C.B., Pocheshkhova, E., Andriadze, G., Muller, C., Westaway, M.C., Lambert, D.M., Zoraqi, G., Turdikulova, S., Dalimova, D., Sabitov, Z., Sultana, G.N.N., Lachance, J.,

- Tishkoff, S., Momyaliev, K., Isakova, J., Damba, L.D., Gubina, M., Nymadawa, P., Evseeva, I., Atramentova, L., Utevska, O., Ricaut, F.X., Brucato, N., Sudoyo, H., Letellier, T., Cox, M.P., Barashkov, N.A., Škaro, V., Mulahasanović, L., Primorac, D., Sahakyan, H., Mormina, M., Eichstaedt, C.A., Lichman, D. V., Abdullah, S., Chaubey, G., Wee, J.T.S., Mihailov, E., Karunas, A., Litvinov, S., Khusainova, R., Ekomasova, N., Akhmetova, V., Khidiyatova, I., Marjanović, D., Yepiskoposyan, L., Behar, D.M., Balanovska, E., Metspalu, A., Derenko, M., Malyarchuk, B., Voevoda, M., Voevoda, M., Osipova, L.P., Lahr, M.M., Gerbault, P., Leavesley, M., Migliano, A.B., Petraglia, M., Balanovsky, O., Khusnutdinova, E.K., Metspalu, E., Thomas, M.G., Manica, A., Nielsen, R., Villems, R., Willerslev, E., Kivisild, T., Metspalu, M., 2016. Genomic analyses inform on migration events during the peopling of Eurasia. *Nature* 538, 238–242.
- Pain, A., Böhme, U., Berry, A.E., Mungall, K., Finn, R.D., Jackson, A.P., Mourier, T., Mistry, J., Pasini, E.M., Aslett, M.A., Balasubramaniam, S., Borgwardt, K., Brooks, K., Carret, C., Carver, T.J., Cherevach, I., Chillingworth, T., Clark, T.G., Galinski, M.R., Hall, N., Harper, D., Harris, D., Hauser, H., Ivens, A., Janssen, C.S., Keane, T., Larke, N., Lapp, S., Marti, M., Moule, S., Meyer, I.M., Ormond, D., Peters, N., Sanders, M., Sanders, S., Sargeant, T.J., Simmonds, M., Smith, F., Squares, R., Thurston, S., Tivey, A.R., Walker, D., White, B., Zuiderwijk, E., Churcher, C., Quail, M.A., Cowman, A.F., Turner, C.M.R., Rajandream, M.A., Kocken, C.H.M., Thomas, A.W., Newbold, C.I., Barrell, B.G., Berriman, M., 2008. The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature* 455, 799–803.
- Paradis, E., Claude, J., Strimmer, K., 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290.
- Pasini, E.M., Böhme, U., Rutledge, G.G., Voorberg-Van der Wel, A., Sanders, M., Berriman, M., Kocken, C.H., Otto, T.D., 2017. An improved *Plasmodium cynomolgi* genome assembly reveals an unexpected methyltransferase gene expansion. *Wellcome Open Res.* 2, 371.
- Pearson, R.D., Amato, R., Auburn, S., Miotto, O., Almagro-Garcia, J., Amaratunga, C., Suon, S., Mao, S., Noviyanti, R., Trimarsanto, H., Marfurt, J., Anstey, N.M., William, T., Boni, M.F., Dolecek, C. Hien, T.T., White, N.J., Michon, P., Siba, P., Tavul, L., Harrison, G., Barry, A., Mueller, I., Ferreira, M.U., Karunaweera, N., Randrianarivelojosa, M., Gao, Q., Hubbart, C., Hart, L., Jeffery, B., Drury, E., Mead, D., Kekre, M., Campino, S., Manske, M., Cornelius, V.J., MacInnis, B., Rockett, K.A., Miles, A., Rayner, J.C., Fairhurst, R.M., Nosten, F., Price, R.N., Kwiatkowski, D.P., 2016. Genomic analysis of local variation and recent evolution in *Plasmodium vivax*. *Nat Genet* 48, 959–964.
- Piro, V.C., Faoro, H., Weiss, V.A., Steffens, M.B.R., Pedrosa, F.O., Souza, E.M., Raittz, R.T., 2014. FGAP: An automated gap closing tool. *BMC Res. Notes* 7, 1–5.
- Prugnolle, F., Rougeron, V., Becquart, P., Berry, A., Makanga, B., Rahola, N., Arnathau, C., Ngoubangoye, B., Menard, S., Willaume, E., Ayala, F.J., Fontenille, D., Ollomo, B., Durand, P., Paupy, C., Renaud, F., 2013. Diversity, host switching and evolution of *Plasmodium vivax* infecting African great apes. *Proc. Natl. Acad.*

Sci. 110, 8123–8128.

- Quinlan, A.R., Hall, I.M., 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Rand, D., Kann, L., 1996. Excess amino acid polymorphism in mitochondrial DNA: Contrasts among genes from *Drosophila*, mice, and humans. *Mol. Biol. Evol.* 13, 735–748.
- Rutledge, G.G., Böhme, U., Sanders, M., Reid, A.J., Cotton, J.A., Maiga-Ascofare, O., Djimdé, A.A., Apinjoh, T.O., Amenga-Etego, L., Manske, M., Barnwell, J.W., Renaud, F., Ollomo, B., Prugnolle, F., Anstey, N.M., Auburn, S., Price, R.N., McCarthy, J.S., Kwiatkowski, D.P., Newbold, C.I., Berriman, M., Otto, T.D., 2017. *Plasmodium malariae* and *P. ovale* genomes provide insights into malaria parasite evolution. *Nature* 542, 101–104.
- Scully, E.J., Kanjee, U., Duraisingh, M.T., 2017. Molecular interactions governing host-specificity of blood stage malaria parasites. *Curr. Opin. Microbiol.* 40, 21–31.
- Semenya, A.A., Tran, T.M., Meyer, E.V., Barnwell, J.W., Galinski, M.R., 2012. Two functional reticulocyte binding-like (RBL) invasion ligands of zoonotic *Plasmodium knowlesi* exhibit differential adhesion to monkey and human erythrocytes. *Malar. J.* 11, 228.
- Steinbiss, S., Silva-Franco, F., Brunk, B., Foth, B., Hertz-Fowler, C., Berriman, M., Otto, T.D., 2016. Companion: a web server for annotation and analysis of parasite genomes. *Nucleic Acids Res.* 44, W29–W34.
- Sundararaman, S.A., Plenderleith, L.J., Liu, W., Loy, D.E., Learn, G.H., Li, Y., Shaw, K.S., Ayoub, A., Peeters, M., Speede, S., Shaw, G.M., Bushman, F.D., Brisson, D., Rayner, J.C., Sharp, P.M., Hahn, B.H., 2016. Genomes of cryptic chimpanzee *Plasmodium* species reveal key evolutionary events leading to human malaria. *Nat. Commun.* 7, 11078.
- Tachibana, S.I., Sullivan, S.A., Kawai, S., Nakamura, S., Kim, H.R., Goto, N., Arisue, N., Palacpac, N.M.Q., Honma, H., Yagi, M., Tougan, T., Katakai, Y., Kaneko, O., Mita, T., Kita, K., Yasutomi, Y., Sutton, P.L., Shakhbatyan, R., Horii, T., Yasunaga, T., Barnwell, J.W., Escalante, A.A., Carlton, J.M., Tanabe, K., 2012. *Plasmodium cynomolgi* genome sequences provide insight into *Plasmodium vivax* and the monkey malaria clade. *Nat. Genet.* 44, 1051–1055.
- Tsai, I.J., Otto, T.D., Berriman, M., 2010. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. *Genome Biol.* 11, R41.
- van den Ent, F., Löwe, J., 2006. RF cloning: a restriction-free method for inserting target genes into plasmids. *J. Biochem. Biophys. Methods* 67, 67–74.
- Wang, B., Lu, F., Cheng, Y., Chen, J.H., Jeon, H.Y., Ha, K.S., Cao, J., Nyunt, M.H., Han, J.H., Lee, S.K., Kyaw, M.P., Sattabongkot, J., Takashima, E., Tsuboi, T., Han, E.T., 2015. Immunoprofiling of the Tryptophan-Rich Antigen Family in *Plasmodium vivax*. *Infect. Immun.* 83, 3083–3095.

- World Health Organisation, 2017. World Malaria Report 2017.
<https://doi.org/http://www.who.int/malaria/publications/world-malaria-report-2017/report/en/>
- Yang, Z., 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591.
- Zimmerman, P.A., 2017. *Plasmodium vivax* infection in Duffy-negative people in Africa. *Am. J. Trop. Med. Hyg.* 97, 636–638.

CHAPTER 5 – Conclusions and future directions

Dorothy Elizabeth Loy

Department of Medicine and Microbiology, Perelman School of Medicine at the
University of Pennsylvania

5.1 Exposure of humans to ape *Laverania* and *P. vivax* parasites

In the past decade, characterization of ape *Plasmodium* species has led to the discovery that the two most prevalent human malaria parasites, *P. falciparum* and *P. vivax*, both have an African ape origin (Liu et al., 2014, 2010). Yet despite extremely high prevalence rates in ape communities, multiple studies have failed to detect ape *Plasmodium* strains in human blood samples obtained from individuals living or working near infected apes (Délicat-Loembet et al., 2015; Ngoubangoye et al., 2016; Sundararaman et al., 2013). However, screening blood samples for parasite DNA does not address the possibility that individuals may be exposed to ape parasites through bites of infected mosquitoes and may develop pre-erythrocytic (i.e. liver stage) infection that fails to produce a blood stage infection. In Chapter 2, I used a PCR-based screen of human fecal samples to attempt to detect pre-erythrocytic ape *Plasmodium* infections. Despite using an intensified PCR approach, I did not detect any ape *Laverania* species or ape *P. vivax*. While one interpretation of this result is that humans are not exposed to ape parasites or do not develop liver-stage infection, it is also possible that failure to amplify ape parasite sequences from human fecal samples is due to an inability to detect pre-erythrocytic parasite DNA in human stool (in contrast to results of experimental infections in mice (Abkallo et al., 2014)). Indeed, I found several links between blood stage parasitemia and fecal *Plasmodium* detection, suggesting that fecal parasite DNA is derived from blood stage infections. Going forward, it will be necessary to determine whether liver stage parasite DNA is detectable in fecal samples of humans, perhaps by collecting and analyzing fecal samples during controlled human malaria exposure in the context of clinical trials.

As no screen of human fecal or blood samples has detected infection with ape *Plasmodium* species, such infections are likely extremely rare events. However, it is not

clear whether this is due to biological differences that impact susceptibility to ape *Plasmodium* infection or whether humans are rarely (if ever) exposed to ape parasites. Thus, characterization of the transmission dynamics of ape *Plasmodium* species is critical for assessing the zoonotic risk posed by these parasites. Specifically, it is necessary to characterize the species of *Anopheles* mosquitoes that transmit the ape parasites, their host species preferences, and the extent to which humans are exposed to infectious mosquitoes (Molina-Cruz and Barillas-Mury, 2016). To date, studies of ape *Plasmodium* vectors have focused on sampling mosquitoes near ape sanctuaries or in national parks and have identified three *Anopheles* species that transmit ape parasites (Makanga et al., 2016; Paupy et al., 2013). However, a critical missing piece of data is the frequency at which humans are exposed to these mosquitoes. Thus, future work should characterize the biting times, host species preferences, and geographic distributions of the vectors of ape *Plasmodium*. Additionally, entomological surveys of villages that are geographically close to infected apes could determine whether humans are routinely exposed to ape parasites. Finally, it will be critical to assess whether the most common vectors of human malaria (i.e. *Anopheles gambiae*, *Anopheles dirus*, and *Anopheles albanus* (Molina-Cruz et al., 2016)) can transmit ape *Plasmodium* species, as this will impact the ability of the parasites to spread in densely populated cities should they emerge in human populations.

Human exposure to ape *Plasmodium* species could also be assessed by identifying antibody responses that are specific to the ape parasites. Interestingly, Duffy-negative individuals have detectable antibody response to blood stage *P. vivax* antigens, despite being protected from blood stage infection (Culleton et al., 2009; Herrera et al., 2005). Thus, humans exposed to ape *Plasmodium* species might have antibody responses to blood stage antigens in the absence of blood stage infections. Whole genome

sequencing of ape *Laverania* parasites and ape *P. vivax* have revealed several erythrocyte invasion genes that are unique to ape parasites, such as *rbp2d* in ape *P. vivax* (see Chapter 4) and *eba165* in *Laverania* species (Otto et al., 2014; Sundararaman et al., 2016). Thus, one avenue to test for exposure to ape parasites would be to ascertain whether antibody response to any of these proteins is sufficiently sensitive and specific to enable screening of humans for past exposure to ape parasites. Alternatively, it may be possible to utilize whole genome sequencing data to identify molecules that are both immunogenic in *P. falciparum* and *P. vivax* infections and also highly divergent in ape parasites. If a panel of such molecules could be identified and confirmed to be specific for ape parasite exposure, it would be possible to assess the magnitude of human exposure to ape *Plasmodium* species even if the parasites fail to result in productive blood stage infections.

While studies of wild apes suggest that the host species restriction of most *Laverania* parasites is strict (Liu et al., 2016; Sundararaman et al., 2016), there is evidence that these parasites can cross species barriers under some circumstances. In ape sanctuaries housing both chimpanzees and gorillas, the gorilla parasites *P. praefalciparum* and *P. adleri* were detected in chimpanzees and the chimpanzee parasite *P. reichenowi* was detected in a gorilla (Ngoubangoye et al., 2016). Yet these parasites have never been identified in the “wrong” host species in the wild, despite screening hundreds of fecal samples (Liu et al., 2016). This suggests that the biological barriers to cross-species transmission of ape parasites can be overcome in certain environmental conditions. Thus, changing climate and land use patterns could increase the risk of spillover of ape parasites into humans. In Southeast Asia, natural transmission of nonhuman primate malarias have only recently been appreciated (Singh et al., 2004; Ta et al., 2014), and fragmentation of forests is thought to be partially responsible for the

rising incidence of *P. knowlesi* in humans (Brock et al., 2016; Fornace et al., 2016). As such, even though there is no evidence that ape *Plasmodium* parasites routinely cause disease in humans at the present time, it is important to continue to monitor for cross-species transmission, especially in regions where forest fragmentation or other social, economic, or political factors lead to increased exposure of humans to vectors of ape *Plasmodium*.

5.2 *P. malariae* and *P. ovale*-like parasite species in African great apes

While the geographical distribution and host species association of *P. vivax* and the *Laverania* parasites have been well characterized (Liu et al., 2017, 2016, 2014, 2010), less is known about the *P. malariae*- and *P. ovale*-like parasites of African great apes. Partial genome sequences of chimpanzee *P. malariae* strains indicate that chimpanzees are infected with genetically diverse parasites that are distinct from those infecting humans (Rutledge et al., 2017). In Chapter 2 of this dissertation, I generated *P. malariae* mitochondrial sequences from apes and humans and identified two SNVs that differ in parasites isolated from humans versus wild apes. None of the *P. malariae* strains from human samples contained the *P. malariae* SNVs found in wild chimpanzees or gorillas. However, both sets of SNVs were found in sanctuary chimpanzees. This suggests that human *P. malariae* parasites can infect chimpanzees (at least in captivity), and is consistent with the relatively relaxed host tropism of this parasite species, which infects several species of South American monkeys in the wild (Alvarenga et al., 2017; Collins and Jeffery, 2007). The *P. ovale*-like parasites in apes are less well characterized. Short, mitochondrial *P. ovale wallikeri* sequences have been obtained from one chimpanzee (Duval et al., 2009) and one gorilla (Mapua et al., 2018). These short sequences are identical to human parasites, but given this limited data it is impossible to know whether

the *P. ovale wallikeri* parasites in apes represent a distinct population or whether there is ongoing zoonotic transmission (Mapua et al., 2018). In contrast, the sequences most closely related to *P. ovale curtisi* in apes differ from human sequences at multiple nucleotide positions over a short sequence (Duval et al., 2009; Kaiser et al., 2010), suggesting the existence of a distinct population in apes. Additional data is needed to determine whether *P. malariae* and *P. ovale* strains circulate freely between ape and human hosts, or whether there are distinct populations in different host species. This will lead to a better understanding of the origin of these malaria parasites in humans and their zoonotic reservoir in African great apes.

In addition to the *P. ovale*-like and *P. malariae*-like parasites in apes, there are other ape parasite species in the *Plasmodium* subgenus about which almost nothing is known. This includes *P. carteri*, a parasite related to *P. vivax* that has thus far only been detected in two chimpanzees (Loy et al., 2017), and a species termed *P. malariae*-related (distinct from ape *P. malariae*) which has been amplified from a chimpanzee and a bonobo (Liu et al., 2017). Both species were discovered through off-target effects of primers designed to amplify *Laverania* or *P. vivax* parasites from ape fecal samples. Given the relatively relaxed host tropism of this clade of parasites, better characterization of all members of the *Plasmodium* subgenus in apes is warranted. However, such studies are severely limited by the very small number of infected samples available for analysis. Specifically screening existing banks of ape fecal samples and nonhuman primate blood samples for these parasites could reveal their distribution and host-species association. However, high quality DNA samples are required to generate sequences longer than a few hundred basepairs. Thus, whole genome sequencing of these parasites will likely be impossible unless infected blood samples can be obtained from sanctuary animals or through collection of sanguinivorous insects that have fed on infected apes.

5.3 Evolutionary history of *P. vivax* in humans

P. vivax-like parasites have been identified in chimpanzees, gorillas, and in bonobos (Kaiser et al., 2010; Liu et al., 2017, 2014). Phylogenetic analysis of short sequences derived from ape and human parasites suggested an African origin of *P. vivax* (Liu et al., 2014), yet the precise relationship of ape and human parasites and the circumstances surrounding the emergence of *P. vivax* in humans remained unknown. Reasoning that comparative genomics of ape and human *P. vivax* could lead to insights into the origins of *P. vivax* in humans, I optimized selective whole genome amplification (SWGA) of *P. vivax* to facilitate sequencing of parasite genomes from asymptomatic apes. In Chapter 3, I showed that SWGA can dramatically increase the proportion of *P. vivax* DNA in unprocessed human whole blood and dried blood spot samples. While primers targeting the entire *P. vivax* genome biased amplification towards the AT-rich subtelomeres (likely a result of phi29's increased replication efficiency in AT-rich genomic regions), I was able to overcome this issue by designing primers to target GC-rich regions of the genome. Application of this method to *P. vivax*-infected ape samples resulted in amplification of partial to near-complete genomes from samples with low parasitemia, without introduction of systemic sequencing errors. Thus, SWGA is a flexible method that is suited to a variety of eukaryotic genomes including those with mixed AT/GC content.

Performing genome wide comparison of the magnitude and character of diversity in ape and human *P. vivax* strains, I found that ape parasites are approximately 10-fold more diverse than human parasites and that human, but not ape, *P. vivax* has a relative excess of nonsynonymous polymorphism. The latter observation suggests this unusual pattern of polymorphism in human *P. vivax* is not a byproduct of the parasite life cycle (which is characterized by bottlenecks during transmission between hosts followed by

rapid multiplication within hosts), as has been suggested (Chang et al., 2013, 2012). Instead, this pattern must reflect evolutionary forces unique to the human parasite population. The most likely explanation is that the human parasite lineage has undergone rapid population expansion, which has resulted in relaxed selection pressure and maintenance of slightly deleterious mutations. If this is indeed the explanation, it suggests that all human *Plasmodium* species studied to date have undergone the same expansion, as an excess of nonsynonymous polymorphism is observed in human *P. falciparum*, *P. vivax*, and *P. malariae* populations (Chang et al., 2013, 2012; Rutledge et al., 2017), yet not in any of the ape parasite species analyzed to date (data from ape *P. vivax* presented here; unpublished data from *P. gaboni* and *P. reichenowi* from O. A. MacLean). Future studies should seek to identify the timing of and circumstances surrounding the dramatic population expansions of human malaria parasites.

5.4 No evidence of positive selection unique to the human *P. vivax* lineage

Unlike ape *P. vivax*, which is transmitted between chimpanzees, bonobos, and gorillas, human *P. vivax* has been exclusively infecting a single species (humans) for thousands of years. Reasoning that the human lineage of *P. vivax* may therefore have accumulated human-specific adaptations, I tested for evidence of selection in the human lineage using McDonald-Kreitman tests on over 4,000 genes. However, I did not detect a signal of positive selection (a relative excess of fixed, nonsynonymous differences) in any of the genes analyzed. This result does not rule out the possibility that some genes in human *P. vivax* may have acquired mutations that improve their ability to infect human hosts, as a small number of nonsynonymous changes can be adaptive yet fail to meet the threshold required for significance in the McDonald-Kreitman test. However, it is possible there is a true paucity of host-specific adaptations due to the relatively short time in which human *P.*

vivax strains have been exclusively circulating in humans. Indeed, this is consistent with the fact that natural and experimental transmission studies indicate that *P. vivax* is able to infect multiple host species: some chimpanzees develop parasitemia upon injection with human *P. vivax* parasites (Rodhain, 1956), splenectomized chimpanzees are readily infected with human *P. vivax* (Sullivan et al., 1996) and ape *P. vivax* can cause malaria in Duffy-positive humans (Prugnolle et al., 2013). Host species adaptation can be further explored *in vitro* if human *P. vivax* is adapted to long-term culture, as this would facilitate side-by-side comparisons of invasion efficiency and growth dynamics within ape and human erythrocytes.

5.5 Loss of three invasion genes in human *P. vivax*

While no genes exhibited signals of positive selection unique to the human *P. vivax* lineage, I did find that human *P. vivax* has lost three reticulocyte binding protein (RBP) genes (*rbp2d*, *rbp2e*, and *rbp3*) that are maintained in ape *P. vivax* parasites. Hypothesizing that these proteins were lost in human *P. vivax* because they bind poorly or not at all to human erythrocyte receptors, I expressed recombinant RBP2e and RBP3 proteins and assessed their binding to red blood cells from humans and chimpanzees, as well a single gorilla and a macaque. Surprisingly, neither protein bound ape cells better than human cells. This suggests that they were not lost in human parasites because they bind poorly to their human receptor orthologs or because they bind receptors unique to chimpanzee and/or gorilla erythrocytes.

I was also surprised to observe that the ape *P. vivax*-derived RBP2e and RBP3 proteins both bound a much smaller fraction of erythrocytes compared to human *P. vivax*-derived RBP2a and RBP2b. While it's possible that this suggests expression of a domain insufficient for receptor binding and/or improper protein folding, RBP2e bound very well to

macaque red blood cells and both proteins adopted secondary structure similar to that of RBP2a and RBP2b. One possible explanation is that the interaction between the proteins and their ape receptors is a low affinity interaction or that the receptors are expressed at low levels on a small number of cells. If erythrocyte receptors are identified for RBP2e and RBP3, the first hypothesis could be tested using surface plasmon resonance and/or more sensitive binding assays which utilize recombinant host protein and pentamerized recombinant RBPs to quantify protein interactions (Bushell et al., 2008). Identification of the receptors would also allow characterization of their abundance and distribution on ape and human erythrocytes. Alternatively, RBP2e and RBP3 may require additional proteins to bind host receptors, as is the case for *P. falciparum* Rh5 which binds only weakly to human erythrocytes in the absence of its essential interaction partners (Reddy et al., 2015; Volz et al., 2016; Wanaguru et al., 2013; W. H. Tham unpublished observation). This possibility could be explored by harvesting native RBP2e and RBP3 proteins from *P. vivax* (or, more feasibly, *P. knowlesi* or *P. cynomolgi* parasites) and using the RBPs to pull down other parasite proteins for identification through mass spectrometry.

While analysis of recombinant RBP binding to ape and human erythrocytes suggests that these proteins do not impact host species specificity, *in vitro* adaptation of *Plasmodium* species to non-native host cells is often accompanied by gains or losses of invasion genes (Scully et al., 2017), including loss of *rbp2e* during adaptation of *P. knowlesi* to cynomolgus erythrocytes (Moon et al., 2016). Additional studies will be required to resolve this apparent discrepancy. The recent adaptation of *P. cynomolgi* to *in vitro* culture (J. Straimer, unpublished data) opens up new avenues for experimentation with this protein family, as *P. cynomolgi* has a large complement of RBPs orthologous to those found in *P. vivax*. For example, it is now possible to use gene-editing technologies such as CRISPR/Cas9 to delete *rbp2e* and *rbp3* in cultured *P. cynomolgi* parasites to

determine whether this impacts invasion efficiency of monkey and/or human red blood cells. This would help clarify whether the loss of these RBPs in human *P. vivax* was adaptive or whether their loss was slightly deleterious but maintained due to the relaxed purifying selection within the human *P. vivax* population.

5.6 *P. vivax* parasites in gorillas and bonobos

One limitation of my analysis of ape *P. vivax* is the fact that, while ape *P. vivax* has been detected in chimpanzees, bonobos, and gorillas, most (6 out of 7) of the samples I analyzed were from chimpanzees. While I was able to obtain one *P. vivax*-infected gorilla sample (Gor3157), amplification of *P. vivax* was highly stochastic and yielded coverage of only 10 genes. While phylogenetic trees of short sequences from chimpanzees and gorillas fail to show host-species specific clades (Liu et al., 2014), in many cases I found that the most divergent *P. vivax* sequences were obtained from gorilla samples. Indeed, in the 10 genes analyzed, the Gor3157 strain was nearly as divergent from chimpanzee parasites as it was from human parasites. Additional sequences from gorilla parasites are required to determine whether gorillas are infected with a *P. vivax* population that is distinct from chimpanzee-infecting strains. At the present time, all *P. vivax* sequences from bonobos are very short (Liu et al., 2017), so it is not possible to determine how bonobo parasites compare to the strains in chimpanzees and gorillas. In addition, it will be important to investigate whether the *rbp* genes that contain open reading frames in chimpanzee but not human parasites are fully intact in gorilla- and bonobo-infecting parasites. Thus, characterization of gorilla and bonobo *P. vivax* strains will allow better estimation of the diversity of ape parasites may reveal unappreciated differences in transmission dynamics and host-species susceptibility of ape *P. vivax*.

5.7 *P. vivax* transmission in modern African humans

P. vivax was long thought to be completely absent in African humans due to the high prevalence of the protective Duffy-negative phenotype, yet there are now numerous reports of *P. vivax* malaria Africa (Howes et al., 2015; Zimmerman, 2017). Sequences obtained from a small subset of *P. vivax* strains from Duffy-positive individuals suggest that African *P. vivax* represents a reintroduction of human parasite strains from outside Africa (Koepfli et al., 2015; Rodrigues et al., 2014). However, most of the analyzed samples were collected in Madagascar or from the horn of Africa, neither of which are populated with apes. While *P. vivax* has been reported in humans in countries with ape populations, the sequencing data available for these strains is either nonexistent or limited to extremely short DNA fragments. As such, it is impossible to determine what proportion, if any, of these cases represent ape-derived parasites. Future research should therefore focus on obtaining informative sequences from *P. vivax* strains infecting African humans, especially in countries with large ape populations. Diagnostic fragments need not be long: mitochondrial genome sequences are easily obtained from infected blood samples and are adequate for distinguishing ape and human *P. vivax* strains (Liu et al., 2014).

Importantly, *P. vivax* in Africa is not confined to Duffy-positive humans. Recently it has become clear that human *P. vivax* is able to infect Duffy-negative individuals (Gunalan et al., 2018; Zimmerman, 2017). It is not known, however, whether the sudden appearance of *P. vivax* in Duffy-negative humans is a result of parasite adaptation to Duffy-negative hosts, or whether this represents better surveillance and/or more sensitive screening methods for *P. vivax*. One way to address this question would be to test the ability of *P. vivax* strains from Duffy-negative individuals to infect Duffy-negative cells *in vitro*, compared to strains obtained from Duffy-positive individuals. However, this would require a large number of parasites and ability to perform short term *in vitro* culture near collection

sites, and is likely not feasible at this time. On the other hand, whole genome sequencing of *P. vivax* from Duffy-negative individuals is relatively straightforward, especially as SWGA is an effective tool for generating genome sequences from samples with low parasitemia (as is often the case in these infections). Genome sequencing could reveal whether the *P. vivax* strains isolated from Duffy-negative individuals are closely related to each other and/or whether there are genetic features (such as genes gained or lost, shared polymorphisms, or locally reduced genetic diversity suggestive of recent selective sweep) that are found at significantly higher frequency in this *P. vivax* population.

5.8 Conclusions

The work presented in this dissertation clarifies the history of *P. vivax* in human populations and provides further evidence that African humans rarely (if ever) acquire zoonotic malaria infections from great apes. Additionally, in light of the fact that these parasites have not been adapted to *in vitro* culture, the ape *P. vivax* genomes presented here represent a vital resource for exploring *P. vivax* biology. Thus, this work provides a foundation for exploration of the mechanisms that underlie cross species transmission and host species adaptation of malaria parasites.

5.9 Chapter references

- Abkallo, H.M., Liu, W., Hokama, S., Ferreira, P.E., Nakazawa, S., Maeno, Y., Quang, N.T., Kobayashi, N., Kaneko, O., Huffman, M.A., Kawai, S., Marchand, R.P., Carter, R., Hahn, B.H., Culleton, R., 2014. DNA from pre-erythrocytic stage malaria parasites is detectable by PCR in the faeces and blood of hosts. *Int. J. Parasitol.* 44, 467–473. <https://doi.org/10.1016/j.ijpara.2014.03.002>
- Alvarenga, D.A.M., Pina-Costa, A., Bianco, C., Moreira, S.B., Brasil, P., Pissinatti, A., Daniel-Ribeiro, C.T., Brito, C.F.A., 2017. New potential *Plasmodium brasilianum* hosts: Tamarin and marmoset monkeys (family *Callitrichidae*). *Malar. J.* 16, 1–7. <https://doi.org/10.1186/s12936-017-1724-0>

- Brock, P.M., Fornace, K.M., Parmiter, M., Cox, J., Drakeley, C.J., Ferguson, H.M., Kao, R.R., 2016. *Plasmodium knowlesi* transmission: Integrating quantitative approaches from epidemiology and ecology to understand malaria as a zoonosis. *Parasitology* 143, 389–400. <https://doi.org/10.1017/S0031182015001821>
- Bushell, K.M., Söllner, C., Schuster-Boeckler, B., Bateman, A., Wright, G.J., 2008. Large-scale screening for novel low-affinity extracellular protein interactions. *Genome Res.* 18, 622–630. <https://doi.org/10.1101/gr.7187808>
- Chang, H.-H., Moss, E.L., Park, D.J., Ndiaye, D., Mboup, S., Volkman, S.K., Sabeti, P.C., Wirth, D.F., Neafsey, D.E., Hartl, D.L., 2013. Malaria life cycle intensifies both natural selection and random genetic drift. *Proc. Natl. Acad. Sci.* 110, 20129–20134. <https://doi.org/10.1073/pnas.1319857110>
- Chang, H.-H., Park, D.J., Galinsky, K.J., Schaffner, S.F., Ndiaye, D., Ndir, O., Mboup, S., Wiegand, R.C., Volkman, S.K., Sabeti, P.C., Wirth, D.F., Neafsey, D.E., Hartl, D.L., 2012. Genomic sequencing of *Plasmodium falciparum* malaria parasites from Senegal reveals the demographic history of the population. *Mol. Biol. Evol.* 29, 3427–3439. <https://doi.org/10.1093/molbev/mss161>
- Collins, W.E., Jeffery, G.M., 2007. *Plasmodium malariae*: Parasite and disease. *Clin. Microbiol. Rev.* 20, 579–592. <https://doi.org/10.1128/CMR.00027-07>
- Culleton, R., Ndounga, M., Zeyrek, F.Y., Coban, C., Casimiro, P.N., Takeo, S., Tsuboi, T., Yadava, A., Carter, R., Tanabe, K., 2009. Evidence for the Transmission of *Plasmodium vivax* in the Republic of the Congo, West Central Africa. *J. Infect. Dis.* 200, 1465–1469. <https://doi.org/10.1086/644510>
- Délicat-Loembet, L., Rougeron, V., Ollomo, B., Arnathau, C., Roche, B., Elguero, E., Moukodoum, N.D., Okougha, A.P., Ondo, B.M., Boundenga, L., Houzé, S., Galan, M., Nkoghé, D., Leroy, E.M., Durand, P., Paupy, C., Renaud, F., Prugnolle, F., 2015. No evidence for ape *Plasmodium* infections in humans in Gabon. *PLoS One* 10, 1–15. <https://doi.org/10.1371/journal.pone.0126933>
- Duval, L., Nerrienet, E., Rousset, D., Sadeuh Mba, S.A., Houze, S., Fourment, M., Le Bras, J., Robert, V., Arie, F., 2009. Chimpanzee malaria parasites related to *Plasmodium ovale* in Africa. *PLoS One* 4, 3–9. <https://doi.org/10.1371/journal.pone.0005520>
- Fornace, K.M., Abidin, T.R., Alexander, N., Brock, P., Grigg, M.J., Murphy, A., William, T., Menon, J., Drakeley, C.J., Cox, J., 2016. Association between landscape factors and spatial patterns of *Plasmodium knowlesi* infections in Sabah, Malaysia. *Emerg. Infect. Dis.* 22, 3–10. <https://doi.org/10.3201/eid2202.150656>
- Gruszczyk, J., Kanjee, U., Chan, L.-J., Menant, S., Malleret, B., Lim, N.T.Y., Schmidt, C.Q., Mok, Y.-F., Lin, K.-M., Pearson, R.D., Rangel, G., Smith, B.J., Call, M.J., Weekes, M.P., Griffin, M.D.W., Murphy, J.M., Abraham, J., Sriprawat, K., Menezes, M.J., Ferreira, M.U., Russell, B., Renia, L., Duraisingh, M.T., Tham, W.-H., 2018. Transferrin receptor 1 is a reticulocyte-specific receptor for *Plasmodium vivax*. *Science (80-.)*. 359, 48–55.

- Gunalan, K., Niangaly, A., Thera, M.A., Doumbo, O.K., Miller, L.H., 2018. *Plasmodium vivax* infections of Duffy-negative erythrocytes: Historically undetected or a recent adaptation? Trends Parasitol. 34, 420–429.
<https://doi.org/10.1016/j.pt.2018.02.006>
- Herrera, S., Gómez, A., Vera, O., Vergara, J., Valderrama-Aguirre, A., Maestre, A., Méndez, F., Wang, R., Chitnis, C.E., Yazdani, S.S., Arévalo-Herrera, M., 2005. Antibody response to *Plasmodium vivax* antigens in FY-negative individuals from the Colombian pacific coast. Am. J. Trop. Med. Hyg. 73, 44–49.
<https://doi.org/10.1016/j.vaccine.2010.04.030>
- Howes, R.E., Reiner, R.C., Battle, K.E., Longbottom, J., Mappin, B., Ordanovich, D., Tatem, A.J., Drakeley, C., Gething, P.W., Zimmerman, P.A., Smith, D.L., Hay, S.I., 2015. *Plasmodium vivax* transmission in Africa. PLoS Negl. Trop. Dis. 9, 1–27. <https://doi.org/10.1371/journal.pntd.0004222>
- Kaiser, M., Löwa, A., Ulrich, M., Ellerbrok, H., Goffe, A.S., Blasse, A., Zommers, Z., Couacy-Hymann, E., Babweteera, F., Zuberbühler, K., Metzger, S., Geidel, S., Boesch, C., Gillespie, T.R., Leendertz, F.H., 2010. Wild chimpanzees infected with 5 *Plasmodium* species. Emerg. Infect. Dis. 16, 1956–1959.
<https://doi.org/10.3201/eid1612.100424>
- Koepfli, C., Rodrigues, P.T., Antao, T., Orjuela-sánchez, P., Eede, P. Van Den, Gamboa, D., Hong, N. Van, Bendezu, J., 2015. *Plasmodium vivax* diversity and population structure across four continents 9, e0003872.
<https://doi.org/10.5061/dryad.72kd2.Funding>
- Liu, W., Li, Y., Learn, G.H., Rudicell, R.S., Robertson, J.D., Keele, B.F., Ndjango, J.B.N., Sanz, C.M., Morgan, D.B., Locatelli, S., Gonder, M.K., Kranzusch, P.J., Walsh, P.D., Delaporte, E., Mpoudi-Ngole, E., Georgiev, A. V., Muller, M.N., Shaw, G.M., Peeters, M., Sharp, P.M., Rayner, J.C., Hahn, B.H., 2010. Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. Nature 467, 420–425.
<https://doi.org/10.1038/nature09442>
- Liu, W., Li, Y., Shaw, K.S., Learn, G.H., Plenderleith, L.J., Malenke, J.A., Sundararaman, S.A., Ramirez, M.A., Crystal, P.A., Smith, A.G., Bibollet-Ruche, F., Ayoub, A., Locatelli, S., Esteban, A., Mouacha, F., Guichet, E., Butel, C., Ahuka-Mundeki, S., Inogwabini, B.I., Ndjango, J.B.N., Speede, S., Sanz, C.M., Morgan, D.B., Gonder, M.K., Kranzusch, P.J., Walsh, P.D., Georgiev, A. V., Muller, M.N., Piel, A.K., Stewart, F.A., Wilson, M.L., Pusey, A.E., Cui, L., Wang, Z., Färnert, A., Sutherland, C.J., Nolder, D., Hart, J.A., Hart, T.B., Bertolani, P., Gillis, A., LeBreton, M., Tafon, B., Kiyang, J., Djoko, C.F., Schneider, B.S., Wolfe, N.D., Mpoudi-Ngole, E., Delaporte, E., Carter, R., Culleton, R.L., Shaw, G.M., Rayner, J.C., Peeters, M., Hahn, B.H., Sharp, P.M., 2014. African origin of the malaria parasite *Plasmodium vivax*. Nat. Commun. 5, 3346.
- Liu, W., Sherrill-Mix, S., Learn, G.H., Scully, E.J., Li, Y., Avitto, A.N., Loy, D.E., Lauder, A.P., Sundararaman, S.A., Plenderleith, L.J., Ndjango, J.B.N., Georgiev, A. V., Ahuka-Mundeki, S., Peeters, M., Bertolani, P., Dupain, J., Garai, C., Hart, J.A., Hart, T.B., Shaw, G.M., Sharp, P.M., Hahn, B.H., 2017. Wild bonobos host

- geographically restricted malaria parasites including a putative new *Laverania* species. Nat. Commun. 8, 1685.
- Liu, W., Sundararaman, S.A., Loy, D.E., Learn, G.H., Li, Y., Plenderleith, L.J., Ndjango, J.B.N., Speede, S., Atencia, R., Cox, D., Shaw, G.M., Ayoub, A., Peeters, M., Rayner, J.C., Hahn, B.H., Sharp, P.M., 2016. Multigenomic delineation of *Plasmodium* species of the *Laverania* subgenus infecting wild-living chimpanzees and gorillas. Genome Biol. Evol. 8, 1929–1939. <https://doi.org/10.1093/gbe/evw128>
- Loy, D.E., Liu, W., Li, Y., Learn, G.H., Plenderleith, L.J., Sundararaman, S.A., Sharp, P.M., Hahn, B.H., 2017. Out of Africa: origins and evolution of the human malaria parasites *Plasmodium falciparum* and *Plasmodium vivax*. Int. J. Parasitol. 47, 87–97. <https://doi.org/10.1016/j.ijpara.2016.05.008>
- Makanga, B., Yangari, P., Rahola, N., Rougeron, V., Elguero, E., Boundenga, L., Moukodoum, N.D., Okouga, A.P., Arnathau, C., Durand, P., Willaume, E., Ayala, D., Fontenille, D., Ayala, F.J., Renaud, F., Ollomo, B., Prugnolle, F., Paupy, C., 2016. Ape malaria transmission and potential for ape-to-human transfers in Africa. Proc. Natl. Acad. Sci. 113, 5329–5334. <https://doi.org/10.1073/pnas.1603008113>
- Mapua, M.I., Fuehrer, H.-P., Petrželková, K.J., Todd, A., Noedl, H., Qablan, M.A., Modrý, D., 2018. Plasmodium ovale wallikeri in Western Lowland Gorillas and Humans, Central African Republic. Emerg. Infect. Dis. J. <https://doi.org/10.3201/eid2408.180010>
- Molina-Cruz, A., Barillas-Mury, C., 2016. Mosquito vectors of ape malarias: Another piece of the puzzle. Proc. Natl. Acad. Sci. 113, 5153–5154. <https://doi.org/10.1073/pnas.1604913113>
- Molina-Cruz, A., Zilversmit, M.M., Neafsey, D.E., Hartl, D.L., Barillas-Mury, C., 2016. Mosquito Vectors and the Globalization of *Plasmodium falciparum* Malaria. Annu. Rev. Genet. 50, 447–465. <https://doi.org/10.1146/annurev-genet-120215-035211>
- Moon, R.W., Sharaf, H., Hastings, C.H., Ho, Y.S., Nair, M.B., Rchiad, Z., Knuepfer, E., Ramaprasad, A., Mohring, F., Amir, A., Yusuf, N.A., Hall, J., Almond, N., Lau, Y.L., Pain, A., Blackman, M.J., Holder, A.A., 2016. Normocyte-binding protein required for human erythrocyte invasion by the zoonotic malaria parasite *Plasmodium knowlesi*. Proc. Natl. Acad. Sci. 113, 7231–7236. <https://doi.org/10.1073/pnas.1522469113>
- Ngoubangoye, B., Boundenga, L., Arnathau, C., Mombo, I.M., Durand, P., Tsoumbou, T.A., Otoro, B.V., Sana, R., Okouga, A.P., Moukodoum, N., Willaume, E., Herbert, A., Fouchet, D., Rougeron, V., Bâ, C.T., Ollomo, B., Paupy, C., Leroy, E.M., Renaud, F., Pontier, D., Prugnolle, F., 2016. The host specificity of ape malaria parasites can be broken in confined environments. Int. J. Parasitol. 46, 737–744. <https://doi.org/10.1016/j.ijpara.2016.06.004>
- Otto, T.D., Rayner, J.C., Böhme, U., Pain, A., Spottiswoode, N., Sanders, M., Quail,

- M., Ollomo, B., Renaud, F., Thomas, A.W., Prugnolle, F., Conway, D.J., Newbold, C., Berriman, M., 2014. Genome sequencing of chimpanzee malaria parasites reveals possible pathways of adaptation to human hosts. *Nat. Commun.* 5, 4754. <https://doi.org/10.1038/ncomms5754>
- Paupy, C., Makanga, B., Ollomo, B., Rahola, N., Durand, P., Magnus, J., Willaume, E., Renaud, F., Fontenille, D., Prugnolle, F., 2013. *Anopheles moucheti* and *Anopheles vinckei* Are Candidate Vectors of Ape *Plasmodium* Parasites, Including *Plasmodium praefalciparum* in Gabon. *PLoS One* 8, e57294. <https://doi.org/10.1371/journal.pone.0057294>
- Prugnolle, F., Rougeron, V., Becquart, P., Berry, A., Makanga, B., Rahola, N., Arnathau, C., Ngoubangoye, B., Menard, S., Willaume, E., Ayala, F.J., Fontenille, D., Ollomo, B., Durand, P., Paupy, C., Renaud, F., 2013. Diversity, host switching and evolution of *Plasmodium vivax* infecting African great apes. *Proc. Natl. Acad. Sci.* 110, 8123–8128. <https://doi.org/10.1073/pnas.1306004110>
- Reddy, K.S., Amlabu, E., Pandey, A.K., Mitra, P., Chauhan, V.S., Gaur, D., 2015. Multiprotein complex between the GPI-anchored CyRPA with PfRH5 and PfRipr is crucial for *Plasmodium falciparum* erythrocyte invasion. *Proc. Natl. Acad. Sci.* 112, 1179–1184. <https://doi.org/10.1073/pnas.1415466112>
- Rodhain, J., 1956. Paradoxical behaviour of *Plasmodium vivax* in the chimpanzee 50, 287–293.
- Rodrigues, P.T., Alves, J.M.P., Santamaria, A.M., Calzada, J.E., Xayavong, M., Parise, M., Da Silva, A.J., Ferreira, M.U., 2014. Using mitochondrial genome sequences to track the origin of imported *plasmodium vivax* infections diagnosed in the United States. *Am. J. Trop. Med. Hyg.* 90, 1102–1108. <https://doi.org/10.4269/ajtmh.13-0588>
- Rutledge, G.G., Böhme, U., Sanders, M., Reid, A.J., Cotton, J.A., Maiga-Ascofare, O., Djimdé, A.A., Apinjoh, T.O., Amenga-Etego, L., Manske, M., Barnwell, J.W., Renaud, F., Ollomo, B., Prugnolle, F., Anstey, N.M., Auburn, S., Price, R.N., McCarthy, J.S., Kwiatkowski, D.P., Newbold, C.I., Berriman, M., Otto, T.D., 2017. *Plasmodium malariae* and *P. ovale* genomes provide insights into malaria parasite evolution. *Nature* 542, 101–104. <https://doi.org/10.1038/nature21038>
- Scully, E.J., Kanjee, U., Duraisingh, M.T., 2017. Molecular interactions governing host-specificity of blood stage malaria parasites. *Curr. Opin. Microbiol.* 40, 21–31. <https://doi.org/10.1016/j.mib.2017.10.006>
- Singh, B., Sung, L.K., Matusop, A., Radhakrishnan, A., Shamsul, S.S.G., Cox-Singh, J., Thomas, A., Conway, D.J., 2004. A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. *Lancet* 363, 1017–1024. [https://doi.org/10.1016/S0140-6736\(04\)15836-4](https://doi.org/10.1016/S0140-6736(04)15836-4)
- Sullivan, J.S., Morris, C.L., McClure, H.M., Strobert, E., Richardson, B.B., Gale Galland, G., Goldman, I.F., Collins, W.E., 1996. *Plasmodium vivax* infections in chimpanzees for sporozoite challenge studies in monkeys. *Am. J. Trop. Med.*

Hyg. 55, 344–349. <https://doi.org/10.4269/ajtmh.1996.55.344>

- Sundararaman, S.A., Liu, W., Keele, B.F., Learn, G.H., Bittinger, K., Mouacha, F., Ahuka-Mundake, S., Manske, M., Sherrill-Mix, S., Li, Y., Malenke, J.A., Delaporte, E., Laurent, C., Mpoudi Ngole, E., Kwiatkowski, D.P., Shaw, G.M., Rayner, J.C., Peeters, M., Sharp, P.M., Bushman, F.D., Hahn, B.H., 2013. *Plasmodium falciparum*-like parasites infecting wild apes in southern Cameroon do not represent a recurrent source of human malaria. *Proc. Natl. Acad. Sci.* 110, 7020–7025.
- Sundararaman, S.A., Plenderleith, L.J., Liu, W., Loy, D.E., Learn, G.H., Li, Y., Shaw, K.S., Ayoub, A., Peeters, M., Speede, S., Shaw, G.M., Bushman, F.D., Brisson, D., Rayner, J.C., Sharp, P.M., Hahn, B.H., 2016. Genomes of cryptic chimpanzee *Plasmodium* species reveal key evolutionary events leading to human malaria. *Nat. Commun.* 7, 11078.
- Ta, T.H., Hisam, S., Lanza, M., Jiram, A.I., Ismail, N., Rubio, J.M., 2014. First case of a naturally acquired human infection with *Plasmodium cynomolgi*. *Malar. J.* 13, 68. <https://doi.org/10.1186/1475-2875-13-68>
- Volz, J.C., Yap, A., Sisquella, X., Thompson, J.K., Lim, N.T.Y., Whitehead, L.W., Chen, L., Lampe, M., Tham, W.H., Wilson, D., Nebl, T., Marapana, D., Triglia, T., Wong, W., Rogers, K.L., Cowman, A.F., 2016. Essential Role of the PfRh5/PfRipr/CyRPA Complex during *Plasmodium falciparum* Invasion of Erythrocytes. *Cell Host Microbe* 20, 60–71. <https://doi.org/10.1016/j.chom.2016.06.004>
- Wanaguru, M., Liu, W., Hahn, B.H., Rayner, J.C., Wright, G.J., 2013. RH5-Basigin interaction plays a major role in the host tropism of *Plasmodium falciparum*. *Proc. Natl. Acad. Sci.* 110, 20735–20740. <https://doi.org/10.1073/pnas.1320771110>
- Zimmerman, P.A., 2017. *Plasmodium vivax* infection in Duffy-negative people in Africa. *Am. J. Trop. Med. Hyg.* 97, 636–638. <https://doi.org/10.4269/ajtmh.17-0461>