Published in final edited form as:

Ann N Y Acad Sci. 2015 April ; 1342(1): 53–61. doi:10.1111/nyas.12708.

The biology of *Plasmodium vivax* explored through genomics

Zunping Luo, Steven A. Sullivan, and Jane M. Carlton

Center for Genomics and Systems Biology, New York University, New York, New York

Abstract

Malaria is a mosquito-borne disease caused by the *Plasmodium* parasite. Of the four *Plasmodium* species that routinely cause human malaria, *P. vivax* is the most widespread species outside Africa, causing ~ 18.9 million cases in 2012. *P. vivax* cannot be cultured continuously *in vitro*, which severely hampers research in non-endemic and endemic countries alike. Consequently, whole-genome sequencing has become an effective means to interrogate the biology of the *P. vivax* parasite. Our comparative genomic analysis of five *P. vivax* reference genomes and several whole-genome sequences of the closely related monkey malaria species *Plasmodium cynomolgi* has revealed an extraordinary level of genetic diversity and enabled characterization of novel multi-gene families and important single-copy genes. The generation of whole-genome sequences from multiple clinical isolates is also driving forward knowledge concerning the biology and evolution of the species. Understanding the biology of *P. vivax* is crucial to develop potential antimalarial drugs and vaccines and to achieve the goal of eliminating malaria.

Keywords

malaria; genomics; *Plasmodium vivax*; *Plasmodium cynomolgi*

Introduction

Malaria is an infectious disease caused by the *Plasmodium* parasite that is transmitted by the bite of infected female *Anopheles* mosquitoes. Over 200 species of *Plasmodium* infect mammals, birds, and reptiles, and of the 53 mammalian *Plasmodium* species, 30 infect primates.¹ Among the four species that routinely cause malaria in humans (*P. falciparum*, *P. vivax*, *P. malariae* and *P. ovale*), *P. falciparum* receives the most attention because of its high level of mortality, mostly of children under the age of 5 in Africa. However, *Plasmodium vivax*, while not endemic in Africa because of its restriction to infecting Duffy⁺ red blood cells (RBCs), also has a significant impact on the health and longevity of human populations in tropical and subtropical parts of the world. It is the most prevalent human malaria parasite outside sub-Saharan Africa, causing an estimated 18.9 million cases in 2012² and contributing to 50% of the total global malaria cases there.² Indeed, high endemicity and large population densities make Central Asia and Southeast Asia globally

Address for correspondence: Center for Genomics and Systems Biology, Department of Biology, New York University, 12 Waverly Place, New York, NY 10003. jane.carlton@nyu.edu.

Conflicts of Interest

The authors declare no conflicts of interest.

significant for *P. vivax*. For example, of the 2.48 billion people at risk, India contributes to ~46% of the burden.³ Recent studies have described the emergence of severe vivax infections (reviewed in Ref. 4), underscoring the importance of research into this global health burden.

P. vivax is closely related to a large clade of malaria parasites that infect cercopithecoid (Old World) monkeys and lesser apes of Southeast Asia (Fig. 1), including *Plasmodium cynomolgi*, one of its closest sister taxa, which infects Asian macaque monkeys. Because of this close relatedness, and the apparent lack of *P. vivax* in Africa, it was assumed that *P. vivax* had arisen from a host switch from monkeys to humans some time in its evolutionary past.⁵ Recently, however, *P. vivax*-like parasites have been found in wild apes (gorillas and chimpanzees) in parts of West Africa.⁶ Interestingly, the ape parasites and human *P. vivax* in this study appear to represent a single species, although the human *P. vivax* from various non-African locations form a distinct lineage within the ape parasites, clustering according to geographical origin.

In addition to a distinctive evolution, *P. vivax* has evolved a group of unique biological characteristics (Table 1). First, unlike *P. falciparum*, *P. vivax* has a hypnozoite stage, during which parasites can remain dormant for long periods in the liver, causing relapses weeks or months later. Also, *P. vivax* displays a preference for invading reticulocytes, which represent 1–2% of circulating human RBCs, whereas *P. falciparum* is capable of invading both reticulocytes and normocytes. Furthermore, after merozoites are released from the liver and invade RBCs, the sexual stages are produced early in an infection compared to *P. falciparum*. In addition to these biological characteristics that render research on the species challenging, *P. vivax* cannot be cultured continuously *in vitro*, resulting in a limited capability of researchers to perform functional assays.⁷ The ability to interrogate the biology of *P. vivax* and to undertake research on this species therefore would appear bleak, were it not for the tremendous insights that whole-genome sequencing of monkey-adapted strains and patient isolates has recently started to provide. In this brief review, we show how the generation and analysis of *P. vivax* whole-genome sequences by ourselves and others has proven highly effective at probing the biology of the species and has provided insights into the parasite's population genetics, genetic diversity, and evolution that were previously unknown.

The genomes of *P. vivax* and *P. cynomolgi*

There are currently five *P. vivax* reference genome sequences, generated from isolates adapted to growth in monkey hosts and taken from patients in El Salvador, Brazil, North Korea, India, and Mauritania (Table 2). The Carlton lab led the project to sequence the first reference genome, Salvador I, from a patient in El Salvador, a multi-institution and multi-investigator project that was published in 2008⁸ in conjunction with the first genome of *P. knowlesi*, a malaria parasite of monkeys known to cause zoonotic infections in humans.⁹ All *Plasmodium* species have three genomes, a nuclear genome, a mitochondrial genome, and an apicoplast genome. The *P. vivax* nuclear genome was revealed to be ~27 megabases (Mb) and ~42% average G–C composition, containing ~5400 genes distributed among 14 chromosomes. The distinct isochores structure¹⁰ of high G–C content in chromosome internal regions but high A–T content in subtelomeric regions was confirmed. Despite the unknown

significance of the isochore structure in *P. vivax*, 11 genes located in G–C–rich isochores were found to evolve faster than genes in A–T–rich regions.⁸ A large proportion (77%) of *P. vivax* genes were found to be orthologous with the genes identified from three other mammalian *Plasmodium* species (*P. falciparum*, *P. knowlesi*, and a species of rodent malaria, *P. yoelii yoelii*), although ~ 50% of these code for hypothetical proteins with uncharacterized functions.⁸

Recognizing the limitations of a single genome sequence from one geographical area, we subsequently published whole-genome sequences of four more *P. vivax* strains from Brazil, India, Mauritania, and North Korea in 2012,¹² more than doubling the number of reference genomes for this species and providing further resources for the malaria community. All strains were originally taken from infected patients, but the parasites were adapted to growth in monkeys, so that biological material can be generated and is available in the malaria repository MR4 (<http://www.mr4.org>). Our major finding from the analysis of these four additional *P. vivax* genomes with Salvador I and a comparator set of *P. falciparum* isolates taken from similar regions of the world is that *P. vivax* displays almost twice as much genetic diversity as *P. falciparum* in terms of single-nucleotide polymorphism (SNP) diversity and microsatellite and gene family variability.¹² This implies that the global population of *P. vivax* may have a capacity for greater functional variation. Indeed, the most sequence diversity was observed in gene families associated with immune evasion and erythrocyte invasion, suggesting that vaccines targeting polymorphic antigens may encounter an even greater hurdle in eliciting an effective immune response than they do in *P. falciparum*, where strain-specific immunity has been shown to limit vaccine efficacy.¹³ The observation of extremely high sequence diversity in *P. vivax*, combined with its dormant stage in the human host, which provides a hidden reservoir of infections, thus reinforces the long-held belief that vivax malaria will be the more difficult malaria species to eliminate.

P. cynomolgi is a simian parasite that causes malaria in Asian Old World monkeys. It is a sister taxon to *P. vivax* (Fig. 1), and because it shares many biological and genetic characteristics, it is often referred to as a model system for the study of the human species (Table 1). Until recently, *P. cynomolgi* had been thought to only infect Asian macaque monkeys naturally, with a few reported cases of laboratory-induced transmission to humans. However, a patient from Malaysia initially diagnosed as infected with *P. malariae*/*P. knowlesi* was later confirmed as infected by *P. cynomolgi*¹⁴ in 2011. We published the first genomes of three strains of *P. cynomolgi* from Malaysia and Cambodia in 2012, and performed comparative genomic analysis with the reference genome of *P. vivax* and *P. knowlesi*.¹⁵ At ~ 26.3 Mb, the genome is very similar in size to that of *P. vivax*, and, interestingly, we found that the difference in gene number (~ 5700 genes in *P. cynomolgi* and ~ 5400 in *P. vivax*, although the number in the *P. vivax* genome may be an underestimate due to misassembly) is explainable by copy-number variation in multigene families involved in human immune system evasion and host erythrocyte invasion. Moreover, copy numbers of the genes composing multigene families were generally greater in the *P. cynomolgi*–*P. vivax* lineage than in *P. knowlesi*, suggesting repeated gene duplication in the ancestral lineage of *P. cynomolgi* and *P. vivax* (or repeated gene deletion

in the *P. knowlesi* lineage). Thus, it appears that a main difference between the genomes of *P. cynomolgi*, *P. vivax*, and *P. knowlesi* is not due to the presence of species-specific genes but rather variations in the copy number of multigene family members shared across the species.

Whole-genome analysis of the aforementioned *P. vivax* and *P. cynomolgi* genomes has allowed us and others to explore the biology of the vivax parasite further, including identification and analysis of important gene families and development of synteny and population genetic maps. A brief description of some of our work in these areas and in the context of research by others is given below.

vir* gene family: the largest multigene family of *P. vivax

The largest gene family in *P. vivax* was first identified in a study that sequenced a 155-kb chromosomal region of a *P. vivax* field isolate.¹⁶ Thirty-two members of the *vir* (variable interspersed repeat) family were identified that could be grouped into six different subfamilies, and *in silico* analysis of protein structures showed that some *vir* gene subfamilies shared structural characteristics with *P. falciparum* SURFIN proteins found on the surface of infected RBCs as well as *Pfmc-2tm* localized at Maurer's clefts (single membrane structures in the cytoplasm of RBCs infected by *P. falciparum*).¹⁷ SURFINS may have a role in the invasion of erythrocytes, and *Pfmc-2tm* may be involved in trafficking of parasite proteins.¹⁸ However, the question remained as to how many copies might exist in the whole genome. With the completion of the *P. vivax* Salvador I reference genome,⁸ a more complete picture emerged. First, ~350 total *vir* genes were identified, and clustering analysis revealed several more subfamilies—but many more genes that could not be clustered because of their extensive diversity, including a variable number of 1–5 exons and size range of 156–2316 bp in length. Motif analysis of the genes revealed that only half possess transmembrane domains and only ~160 contain a PEXEL-like (*Plasmodium* export element) motif,⁸ suggesting that there may be alternative pathways for export of VIR proteins to the erythrocyte surface, and that VIR proteins might have different subcellular localizations and functions.

More recently, we studied the *vir* gene family in all five *P. vivax* reference genomes, and, as expected, it was revealed to be highly divergent, with limited overlapping repertoires between strains.¹² However, we identified 15 ultra-conserved *vir* genes that displayed low SNP diversity in all *P. vivax* genomes, with one locus, PVX_113230, absolutely invariant, with high similarity to the *kir* gene family in *P. knowlesi*, and exhibiting conserved synteny in rodent malaria species. These data indicate that PVX_113230 might be the founder gene in *P. vivax* and may perform its ancestral role, although that role is currently unknown.¹² A recent publication used novel computational methods to cluster 295 VIR proteins into 10 subfamilies, and an additional 39 hypothetical proteins were predicted to be new VIR proteins by applying this new computational approach.¹⁹

Expansion of RBP gene family in *P. vivax*

One family of *Plasmodium* parasite ligands that facilitates RBC invasion is the reticulocyte binding-like (RBL) superfamily, which functions by binding to reticulocyte receptors with

high affinity.²⁰ RBL proteins are large, ranging in size from 230–350 kDa,²¹ and localize to merozoites (the form of the parasite that invades RBCs).^{21,22} Members of the RBL superfamily have a conserved two-exon gene structure, with the first exon encoding a signal peptide and a large second exon coding the remainder of the protein sequence.²¹ Regardless of these similarities, RBLs have different preferences for RBCs during invasion in different *Plasmodium* species. Reticulocyte-binding proteins (RBPs), members of the RBL superfamily in *P. cynomolgi* and *P. vivax*, are proposed to specifically recognize the reticulocyte subpopulation of RBCs during the invasion process.²²

RBPs in *P. vivax* are promising candidates for development of vaccines blocking receptor–ligand interactions.²³ However, the sequence polymorphism of *rbp* genes may hinder the effectiveness of such vaccines;²⁴ thus, unraveling genetic diversity in *rbp* genes in *P. vivax* is a crucial step for successful development of erythrocytic-stage vaccines. In 1992, two genes coding for RBPs, *rbp1* and *rbp2*, were identified in *P. vivax*,²² giving rise to the assumption that *P. vivax* has an uncomplicated erythrocyte-invasion pathway compared to *P. falciparum*. However, analysis of the first *P. vivax* genome sequence identified ten *rbp* copies, including three fragmented genes and two pseudogenes, suggesting that *P. vivax* may have alternative pathways of RBC invasion.⁸ Although functional genomic studies have yet to be performed, a more recent validation of the multi-copy nature of this gene family has come in the form of phylogenetic analysis of the genome sequences generated from three strains of *P. cynomolgi*.¹⁵ This showed that *rbl* genes from *P. cynomolgi*, *P. vivax*, and *P. knowlesi* can be classified into three distinct groups, RBP1, RBP2, and RBP3, on the basis of sequence similarity. All three groups of RBPs are represented in the closely related monkey species *P. cynomolgi*, whereas *P. vivax* and *P. knowlesi* lack functional genes from the RBP3 and RBP1 groups, respectively. Thus, *rbl* gene family expansion seems to have occurred after speciation, indicating that the three species may have multiple species-specific erythrocyte-invasion mechanisms.

One final interesting analysis facilitated by the availability of these several *P. vivax* and *P. cynomolgi* genomes was identification of copy-number variants (CNVs) of an *rbp* gene between strains of a single *Plasmodium* species. An ortholog of the *P. vivax rbp1b* gene is present in one *P. cynomolgi* strain, Berok, but absent in two strains, B and Cambodian.¹⁵ As noted by the authors, the repeated birth and death of *rbp* genes, a signature of adaptive evolution, may allow malaria parasites to switch between monkey and human hosts.

The Duffy binding protein gene family

Another family of invasion ligands that bind RBCs during invasion is the erythrocyte binding-like (EBL) superfamily. Members of the EBL superfamily share similar features, including an N-terminal signal sequence, a large extracellular segment followed by a transmembrane domain, and a C-terminal cytoplasmic domain. Members of the EBL superfamily in *P. cynomolgi* and *P. vivax* are called Duffy binding proteins (DBPs), and are exported to the merozoite surface where they bind to the host receptor Duffy antigen receptor for chemokines (DARC).^{25–27} Erythrocyte invasion by *P. vivax* is restricted by the interaction between DBPs and DARC receptors, because Duffy[−] people were shown to be resistant to infection by *P. vivax*.^{28,29} However, recently *P. vivax* infections in Duffy[−]

individuals have been found in several countries in Africa^{30–34} and Brazil,³⁵ although the highest prevalence of such cases remains in Madagascar.³⁶ Sequencing the *P. vivax* and *P. cynomolgi* reference genomes confirmed the presence of one DBP gene in *P. vivax* and two in *P. cynomolgi*;^{8,12,15} unlike the RBPs, diversity is relatively conserved in the DBP genes. However, a recent study identified a tandem duplication of the *dbp* gene in whole-genome data from *P. vivax* field isolates from Madagascar.³⁷ The duplication was extremely prevalent in the patients, suggesting that *P. vivax* may somehow overcome the barrier to infection presented by Duffy negativity through the duplication of the *dbp* gene. Moreover, the highly conserved nature of the *dbp* duplication in these patients indicated that it was a recent event. More studies are required to determine how much of a global phenomenon vivax infection of Duffy[−] individuals is, and how much of a threat it may be to control of the species.

Whole-genome alignment and synteny maps

With the availability of several whole-genome sequences of *Plasmodium*, we have been able to generate genome-wide alignment and synteny maps for comparative studies of the species. Synteny is the conservation of the order and orientation of genes between genomes of different species, and synteny maps have various uses including (1) identifying orthologs through position when sequence similarity is low; (2) plotting regions of the genome that have evolved differently; (3) identifying regions of a genome that may be co-regulated; (4) inferring the evolution of the genus through tracking chromosomal rearrangements; and (5) identification of conserved non-coding regions in the *Plasmodium* genome. The first genome-wide synteny map between *P. falciparum* and *P. y. yoelii* was published as part of the *P. falciparum* genome paper, and revealed long tracts of genes conserved in synteny between the two distantly related species.³⁸ After generation of more reference genomes, we developed a synteny map of six *Plasmodium* species (*P. vivax*, *P. falciparum*, *P. y. yoelii*, *P. knowlesi*, and two species of rodent malaria parasites, *Plasmodium chabaudi* and *Plasmodium berghei*).⁸ Of the ~ 3300 orthologs on the map between all six species, 99% were found to be positionally conserved. In addition, it was possible to reconstruct the ancestral karyotype of the six species, which turned out to correspond to the karyotypes of *P. vivax* and *P. knowlesi*; the karyotypes of *P. falciparum* and the rodent malaria parasites can be reconstructed from this form through nine and six chromosomal rearrangements, respectively. No synteny breakage hot-spots were identified, indicating that intersyntenic breakpoints were unique during the divergence of the species. More recently, the synteny map was expanded to include *P. cynomolgi*.¹⁵ Not unexpectedly for species from the same monkey malaria clade, gene synteny along the 14 chromosomes of *P. knowlesi*, *P. vivax*, and *P. cynomolgi* is highly conserved, although numerous microsyntenic breaks are present in regions containing multigene families. The map also identified two apparent errors in existing public sequence databases: an inversion in chromosome 3 of *P. knowlesi* and an inversion in chromosome 6 of *P. vivax*. This underlies the importance of such maps for assembly and annotation checking (Fig. 2).

Importantly, alignment of whole *P. vivax*, *P. knowlesi*, and *P. cynomolgi* genomes has enabled calculation of key evolutionary parameters, enabling genes to be classified based on their mode of evolution. For example, analysis of ~ 3300 high-quality *P. vivax*/*P. knowlesi*

orthologs enabled determination of maximum likelihood estimates of the rate of substitution at synonymous (dS) and non-synonymous (dN) sites, as well as ω (dN/dS).⁸ *P. vivax* chromosomes were found to differ significantly in their average values for both dS and dN, but the two variables are strongly correlated within and between chromosomes, and the existence of heterogeneous mutation rates across the genome is also suggested. The degree of selective constraint has also shown to vary across classes of *Plasmodium* genes. Genes encoding glycosylphosphatidylinositol (GPI)-anchored proteins, cell adhesion proteins, proteins predicted to be exported, and proteins with transmembrane or signal peptide motifs (i.e., extracellular proteins) have been found to evolve significantly faster than genes involved in housekeeping functions. Thus, the host immune system appears to have strongly influenced evolutionary rate variation between gene classes in *Plasmodium* by targeting extracellular proteins.

Similarly, selective constraint analysis within *P. cynomolgi* and between *P. vivax* and *P. cynomolgi* revealed some interesting findings.¹⁵ Approximately 60 genes with dN > dS, and ~ 3200 genes with dS > dN, from ~ 4000 pairs of orthologs between two *P. cynomolgi* strains were identified, providing clues as to the types of genes subject to different selective pressures in the species. Similarly, whether *P. cynomolgi* is a good model system for *P. vivax* was investigated by exploring the degree to which evolution of orthologs between the two species had been constrained over evolutionary time. Of ~ 4600 pairs of orthologs analyzed between the two species, less than 2% were found to be under positive selection (i.e., evolving quickly), and at least 81% to be under strong functional constraint (i.e., evolving slowly), indicating that the genome of *P. cynomolgi* is highly conserved with *P. vivax* and emphasizing the value of the monkey malaria species as a biomedical and evolutionary model for studying *P. vivax*.

Future directions

So what is next for whole-genome sequencing studies of *P. vivax*? We are currently leading a project to sequence ~ 180 *P. vivax* clinical samples collected from diverse geographic regions, many of which have been collected as part of the National Institutes of Health-funded International Centers of Excellence in Malaria Research initiative³⁹ (Table 2). The genomes are being sequenced at New York University and the Broad Institute using a combination of hybrid selection⁴⁰ and next-generation sequencing. The inability to easily obtain large amount of host-free DNA has hindered the genome-sequencing projects for *P. vivax*, and so by using a host-depletion method of hybrid selection, parasite DNA in clinical samples can be enriched as much as 40-fold, reducing the cost associated with generating sufficient parasite coverage for variant discovery or genome assembly. Analyzing the sequences of these global isolates will greatly advance our knowledge of the global genetic diversity and population genetics of *P. vivax* and the species' phylogeography, and will enable examination of the evolution of its genome in contrasting regions of the world. Genomic studies in other labs are also pushing the boundaries of our knowledge of *P. vivax*. For example, whole-genome sequencing studies of successive relapses in a single patient have shown that genetically related parasites can cause both short and long latency relapses,⁴¹ and may provide a method of differentiating between recrudescence, relapse, and re-infection in a patient. Several labs are developing genomic technologies and

computational approaches to determine within-host diversity of *P. vivax* infections, and thus permit rapid diagnosis of, for example, drug resistance phenotypes.⁴² Finally, whole-genome sequencing of additional species of non-human malaria species,⁴³ including those in the monkey malaria clade, will provide evolutionary context and comparative data for deeper studies into the fascinating biology of *P. vivax*.

Conclusion

P. vivax is particularly suited to interrogation by genomic technologies because the lack of a long-term *in vitro* culture system makes it difficult to study in a laboratory setting. The generation of five *P. vivax* reference assemblies and genome data from several patient isolates has dramatically increased our understanding of the genetic diversity and population genetics of this species. Whole-genome sequencing of several *P. cynomolgi* (a sister taxon to *P. vivax*) strains has also significantly advanced our knowledge of the biology and evolutionary history of the monkey malaria clade. We look forward to the release of hundreds of *P. vivax* whole-genome sequences isolated from patients throughout the world in the coming years, both from our own studies as well as from others, enabling real-time epidemiology and population genomic studies of extant parasites and leading to the development of better methods of malaria control.

Acknowledgments

ZL is supported by the MacCracken Program in the Graduate School of Arts and Science at New York University (New York, NY). This work was supported by National Institute of Allergy and Infectious Diseases, National Institutes of Health (NIH) International Centers of Excellence in Malaria Research Grant U19AI089676. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Martinsen, ES.; Perkins, SL. The Diversity of *Plasmodium* and other Haemosporidians: The Intersection of Taxonomy, Phylogenetics and Genomics. Caister Academic Press; Norfolk, UK: 2013.
2. WHO. World Malaria Report 2013. 2013
3. Gething PW, et al. A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010. PLoS Negl Trop Dis. 2012; 6:e1814. [PubMed: 22970336]
4. Rahimi BA, et al. Severe vivax malaria: a systematic review and meta-analysis of clinical studies since 1900. Malaria journal. 2014; 13:481. [PubMed: 25486908]
5. Carlton JM, Das A, Escalante AA. Genomics, population genetics and evolutionary history of *Plasmodium vivax*. Adv Parasitol. 2013; 81:203–222. [PubMed: 23384624]
6. Liu W, et al. African origin of the malaria parasite *Plasmodium vivax*. Nature communications. 2014; 5:3346.
7. Noulin F, et al. 1912–2012: a century of research on *Plasmodium vivax* in vitro culture. Trends in parasitology. 2013; 29:286–294. [PubMed: 23623759]
8. Carlton JM, et al. Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. Nature. 2008; 455:757–763. [PubMed: 18843361]
9. Pain A, et al. The genome of the simian and human malaria parasite *Plasmodium knowlesi*. Nature. 2008; 455:799–803. [PubMed: 18843368]
10. McCutchan TF, et al. Evolutionary Relatedness of *Plasmodium* Species as Determined by the Structure of DNA. Science. 1984; 225:808–811. [PubMed: 6382604]

11. Carlton J. The *Plasmodium vivax* genome sequencing project. Trends in parasitology. 2003; 19:227–231. [PubMed: 12763429]
12. Neafsey DE, et al. The malaria parasite *Plasmodium vivax* exhibits greater genetic diversity than *Plasmodium falciparum*. Nat Genet. 2012; 44:1046–1050. [PubMed: 22863733]
13. Ellis RD, et al. Blood stage vaccines for *Plasmodium falciparum*. Human Vaccines. 2014; 6:627–634. [PubMed: 20519960]
14. Ta TH, et al. First case of a naturally acquired human infection with *Plasmodium cynomolgi*. Malaria journal. 2014; 13:68. [PubMed: 24564912]
15. Tachibana S, et al. *Plasmodium cynomolgi* genome sequences provide insight into *Plasmodium vivax* and the monkey malaria clade. Nat Genet. 2012; 44:1051–1055. [PubMed: 22863735]
16. del Portillo HA, et al. A superfamily of variant genes encoded in the subtelomeric region of *Plasmodium vivax*. Nature. 2001; 410:839–842. [PubMed: 11298455]
17. Merino EF, et al. Multi-character population study of the vir subtelomeric multigene superfamily of *Plasmodium vivax*, a major human malaria parasite. Mol Biochem Parasitol. 2006; 149:10–16. [PubMed: 16730808]
18. Frech C, Chen N. Variant surface antigens of malaria parasites: functional and evolutionary insights from comparative gene family classification and analysis. BMC Genomics. 2013; 14:427. [PubMed: 23805789]
19. Lopez FJ, et al. A new computational approach redefines the subtelomeric *vir* superfamily of *Plasmodium vivax*. BMC Genomics. 2013; 14:8. [PubMed: 23324551]
20. Okenu DM, et al. The reticulocyte binding proteins of *Plasmodium cynomolgi*: a model system for studies of *P. vivax*. Mol Biochem Parasitol. 2005; 143:116–120. [PubMed: 15990180]
21. Rayner JC, et al. Rapid evolution of an erythrocyte invasion gene family: the *Plasmodium reichenowi* Reticulocyte Binding Like (RBL) genes. Molecular and Biochemical Parasitology. 2004; 133:287–296. [PubMed: 14698440]
22. Galinski MR, et al. A Reticulocyte-Binding Protein Complex of *Plasmodium vivax* Merozoites. Cell. 1992; 69:1213–1226. [PubMed: 1617731]
23. Galinski MR, Barnwell JW. *Plasmodium vivax* Merozoites, Invasion of Reticulocytes and Considerations for Malaria Vaccine Development. Parasitology Today. 1996; 12:20–29. [PubMed: 15275304]
24. Rayner JC, et al. Dramatic difference in diversity between *Plasmodium falciparum* and *Plasmodium vivax* reticulocyte binding-like genes. The American Journal of Tropical Medicine and Hygiene. 2005; 72:666–674. [PubMed: 15964948]
25. Adams JH, et al. The Duffy Receptor Family of *Plasmodium knowlesi* is located within the micronemes of invasive malaria merozoites. Cell. 1990; 63:141–153. [PubMed: 2170017]
26. Adams JH, et al. A family of erythrocyte binding proteins of malaria parasites. Proc Natl Acad Sci. 1992; 89:7085–7089. [PubMed: 1496004]
27. Batchelor JD, Zahm JA, Tolia NH. Dimerization of *Plasmodium vivax* DBP is induced upon receptor binding and drives recognition of DARC. Nat Struct Mol Biol. 2011; 18:908–914. [PubMed: 21743458]
28. Miller LH, et al. The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, FyFy. The New England Journal of Medicine. 1976; 295:302–304. [PubMed: 778616]
29. Miller LH, et al. Erythrocyte receptors for (*Plasmodium knowlesi*) malaria: Duffy blood group determinants. Science. 1975; 189:561–563. [PubMed: 1145213]
30. Culleton R, et al. Evidence for the transmission of *Plasmodium vivax* in the Republic of the Congo, West Central Africa. J Infect Dis. 2009; 200:1465–1469. [PubMed: 19803728]
31. Mendes C, et al. Duffy negative antigen is no longer a barrier to *Plasmodium vivax*—molecular evidences from the African West Coast (Angola and Equatorial Guinea). PLoS Negl Trop Dis. 2011; 5:e1192. [PubMed: 21713024]
32. Ryan JR, et al. Evidence for transmission of *Plasmodium vivax* among a Duffy antigen negative population in western Kenya. The American Journal of Tropical Medicine and Hygiene. 2006; 75:575–581. [PubMed: 17038676]

33. Woldearegai TG, et al. Plasmodium vivax malaria in Duffy-negative individuals from Ethiopia. *Trans R Soc Trop Med Hyg.* 2013; 107:328–331. [PubMed: 23584375]
34. Wurtz N, et al. Vivax malaria in Mauritania includes infection of a Duffy-negative individual. *Malaria journal.* 2011; 10:336–343. [PubMed: 22050867]
35. Cavasini CE, et al. Duffy blood group gene polymorphisms among malaria vivax patients in four areas of the Brazilian Amazon region. *Malaria journal.* 2007; 6:167. [PubMed: 18093292]
36. Menard D, et al. Plasmodium vivax clinical malaria is commonly observed in Duffy-negative Malagasy people. *Proc Natl Acad Sci U S A.* 2010; 107:5967–5971. [PubMed: 20231434]
37. Menard D, et al. Whole genome sequencing of field isolates reveals a common duplication of the Duffy binding protein gene in Malagasy Plasmodium vivax strains. *PLoS Negl Trop Dis.* 2013; 7:e2489. [PubMed: 24278487]
38. Carlton JM, et al. Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii. *Nature.* 2002; 419:512–519. [PubMed: 12368865]
39. Rao M. Foreword. The international centers of excellence for malaria research. *Acta tropica.* 2012; 121:157. [PubMed: 21819955]
40. Melnikov A, et al. Hybrid selection for sequencing pathogen genomes from clinical samples. *Genome Biol.* 2011; 12:R73. [PubMed: 21835008]
41. Bright AT, et al. Whole genome sequencing analysis of Plasmodium vivax using whole genome capture. *BMC Genomics.* 2012; 13:262. [PubMed: 22721170]
42. Mideo N, et al. Ahead of the curve: next generation estimators of drug resistance in malaria infections. *Trends in parasitology.* 2013; 29:321–328. [PubMed: 23746748]
43. Otto TD, et al. Genome sequencing of chimpanzee malaria parasites reveals possible pathways of adaptation to human hosts. *Nature communications.* 2014; 5:4754.
44. Duval L, et al. African apes as reservoirs of Plasmodium falciparum and the origin and diversification of the Laverania subgenus. *Proceedings of the National Academy of Sciences of the United States of America.* 2010; 107:10561–10566. [PubMed: 20498054]
45. Dharia NV, et al. Whole-genome sequencing and microarray analysis of ex vivo Plasmodium vivax reveal selective pressure on putative drug resistance genes. *Proc Natl Acad Sci.* 2010; 107:20045–20050. [PubMed: 21037109]
46. Chan ER, et al. Whole genome sequencing of field isolates provides robust characterization of genetic diversity in Plasmodium vivax. *PLoS Negl Trop Dis.* 2012; 6:e1811. [PubMed: 22970335]

Figure 1.

A *Plasmodium* cartoon phylogeny. The *Plasmodium* species that routinely cause malaria in humans are highly evolutionarily diverse, and only *P. falciparum* and *P. vivax* are shown here. The tree was generated using mitochondrial fragments of cytochrome B and cytochrome C oxidase subunit I coding regions, using maximum likelihood methods, and is adapted from Refs. ⁶ and ⁴⁴.

Figure 2.

Genome synteny between representative species of three lineages of the *Plasmodium* parasite. Protein coding genes of *P. knowlesi*, representative of species of the monkey malaria clade, are shown aligned with those of *P. falciparum*, and *P. berghei* as a representative of the rodent malaria clade. Chromosomes 1 through 14 are shown at the top, and conserved protein coding genes between the genomes are colored in order from red (5'-end of chromosome 1) to blue (3'-end of chromosome 14). An approximate scale in megabases is shown in the middle panel. Such synteny maps can reveal how chromosomes evolved from the first *Plasmodium* ancestor into the current karyotypes of extant malaria parasite species. Adapted from Ref. 15.

Table 1

Summary of some biological features between the human malaria species *P. vivax* and *P. falciparum*, and the monkey model *P. cynomolgi*.

	<i>P. vivax</i>	<i>P. cynomolgi</i>	<i>P. falciparum</i>
RBC preference	Reticulocytes	Reticulocytes	Reticulocytes, normocytes
Invasion receptor	DARC*	DARC	Multiple
Dormant liver phase	Yes	Yes	No
Gametocyte production	Early	Early	Late
Chromosome isochores	Yes	Yes	No
Caveola-vesicle complexes	Yes	Yes	No
Sequestration	No?	No?	Yes
Gametocyte shape	Round	Round	Crescent

NOTE: DARC*: *P. vivax* can invade both Duffy⁺ and Duffy[−] erythrocytes in certain areas of the world.^{27–33} An isochore is a region of a genome that displays a nucleotide bias, for example extreme A–T composition. Caveola-vesicle complexes are flask-shaped indentations with vesicles in the parasite-infected erythrocyte membrane. Their function may be related to nutrient transport.

Table 2

Summary of all *P. vivax* genome-sequencing projects from 2008 to the present.

Strain name	Country (no. of isolates)	Published date; reference	Comments
Five reference genomes:			
Salvador I	El Salvador	2008 ⁸	Salvador I was the first reference genome to be sequenced from a patient sample adapted to growth in a monkey.
Brazil I India VII Mauritania I North Korean	Brazil India Mauritania North Korea	2012 ¹²	Four more reference genomes of monkey-adapted strains were subsequently assembled and annotated.
12 patient genomes:			
IQ07	Peru	2010 ⁴⁵	First patient isolate sequenced.
M08, M19 C08, C15, C127 SA-94, SA-95, SA-96, SA-97, SA-98	Madagascar Cambodia Peru	2012 ⁴⁶ 2012 ⁴¹	Several more patient isolates sequenced for SNP detection only (i.e., no assembly or annotation)
M15	Madagascar	2013 ³⁷	
178 global <i>P. vivax</i> patient samples	Brazil (20), Columbia (31), PNG (23), Peru (47), India (9), China (8), Mexico (20), Thailand (20)	Unpublished	For generation of a global genetic map of <i>P. vivax</i> , including population structure and phylogeography.