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

Dorothy E. Loya,b,1, Lindsey J. Plenderleithc,d,1, Sesh A. Sundaramanan,b,1, Weimin Lia, Jakub Gruszczyk1, Yi-Jun Chend,f, Stephanie Trimbiolb, Gerald H. Learna, Oscar A. MacLeanc,d, Alex L. K. Morganc,d, Yangying Li, Alex A. N. Avittoa, graueri

*Department of Medicine, University of Pennsylvania, Philadelphia, PA 19104; bDepartment of Microbiology, University of Pennsylvania, Philadelphia, PA 19104; cInstitute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3FL, United Kingdom; dCentre for Immunity, Infection and Evolution, University of Edinburgh, Edinburgh EH9 3FL, United Kingdom; eWalter and Eliza Hall Institute of Medical Research, Parkville VIC 3052, Australia; fDepartment of Medical Biology, The University of Melbourne, Parkville VIC 3010, Australia; gRobert Koch Institute, 13353 Berlin, Germany; hSanaga-Yong Chimpanzee Rescue Center, International Development Association-Africa, Portland, OR 97208; iRecherche Translationnelle Appliquée au VIH et aux Maladies Infectieuses, Institut de Recherche pour le Développement, University of Montpellier, INSERM, 34090 Montpellier, France; and jMalaria Programme, Wellcome Trust Sanger Institute, Genome Campus, Hinxton Cambridgeshire CB10 1SA, United Kingdom

Contributed by Beatrice H. Hahn, July 13, 2018 (sent for review June 12, 2018; reviewed by David Serre and L. David Sibley)

Wild-living African apes are endemically infected with parasites that are closely related to human *Plasmodium vivax*, a leading cause of malaria outside 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 nearly 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 non-synonymous 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*, and *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 to 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.

*Plasmodium vivax* genomics malaria great apes zoonotic transmission

The protozoal parasite *Plasmodium vivax* causes over 8 million cases of human malaria per year, with the great majority occurring in Southeast Asia and South America (1). *P. vivax* is rare in humans in Africa due to the high prevalence of the Duffy-negative mutation (2), 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 (3), although this protection is not absolute (4). Until recently, *P. vivax* was thought to have emerged in Asia following the cross-species transmission of a macaque parasite (5, 6). However, the finding of closely related parasites in wild-living chimpanzees and gorillas suggested an African origin of *P. vivax* (7). Indeed, parasite sequences closely resembling *P. vivax* have been detected in western (*Pan troglodytes venus*), central (*Pan troglodytes troglodytes*), and eastern (*Pan troglodytes schweinfurthii*) chimpanzees, eastern (*Gorilla beringei graueri*) and western lowland (*Gorilla gorilla gorilla*) gorillas, and most recently in bonobos (*Pan paniscus*) (7–11). Phylogenetic analyses of available sequences revealed that ape and human parasites were nearly identical, with human *P. vivax* sequences forming a monophyletic lineage that usually fell within the radiation of the ape parasites (7). 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.

**Significance**

Chimpanzees, bonobos, and gorillas harbor close relatives of human *Plasmodium vivax*, but current knowledge of these parasites is limited to a small number of gene fragments derived almost exclusively from mitochondrial DNA. We compared nearly full-length genomes of ape parasites with a global sample of human *P. vivax* and tested the function of human and ape *P. vivax* proteins believed to be important for erythrocyte binding. The results showed that ape parasites are 10-fold more diverse than human *P. vivax* and exhibit no evidence of species specificity, whereas human *P. vivax* represents a bottlenecked lineage that emerged from within this parasite group. Thus, African apes represent a large *P. vivax* reservoir whose impact on human malaria eradication requires careful monitoring.


The authors declare no conflict of interest.

**Open access statement:** This open access article is distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

**Data deposition:** The data reported in this paper have been deposited in the GenBank database (accession nos. PRJNA447492 and MH443154-MH443228).

1 D.E.L. and L.J.P. contributed equally to this work.
2 P.M.S. and B.H.H. contributed equally to this work.
3 To whom correspondence should be addressed. Email: bhahn@pennmedicine.upenn.edu.

**See supplementary information online:** This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1810053115/-/DCSupplemental.

**Published online:** August 20, 2018.
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 as infecting vertebrate hosts (12), have successfully colonized humans (13). Thus, the circumstances that surround the emergence of each of these human pathogens are of interest, especially since most, if not all, have nonhuman primate parasites as their closest relatives (7, 9, 14). 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 reflect 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 (11), and human *P. vivax* has been transmitted to wild-living monkeys in South America, generating what has been called “Plasmodium simium” (15). If cross-species infection and recombination of ape and human *P. vivax* are possible, as appears to be the case for *P. simium* and *P. vivax* in South America (15, 16), 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 (17).

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. chimpensis* to cross the species barrier to infect humans (18, 19). 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 (7) and has not been cultured. Although removal of host leukocytes from whole-blood samples (20) and parasite nucleic acid capture (21) 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 (18) 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* (21), 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 (RBP) 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, while the current human-infecting parasites represent a lineage that had escaped out of Africa.

### 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 SWGA without the risk of generating interspecies recombinants (18, 22). 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 (18). Since SWGA can result in stochastic amplification when target templates are rare (18, 23), 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 (SI Appendix, Table S1). Individual SWGA reactions were pooled and sequenced on Illumina and PacBio platforms (SI Appendix).

Draft genomes of the chimpanzee *P. vivax* strains PV56 and PV43 were generated using iterative reference-guided assembly to the human PvP01 reference genome (24) followed by gap-filling steps. In addition, PV56 reads that did not map to the PvP01 genome were de novo assembled to obtain subtelomeric contigs. The resulting assemblies yielded 21.9 Mbp and 21.2 Mbp of sequence for PV56 and PV43, respectively (Table 1). Because sample SY43 contained at least five *P. vivax* strains (7), the PV43 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

<table>
<thead>
<tr>
<th>Table 1. Genome features of ape <em>P. vivax</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Genome attributes</td>
</tr>
<tr>
<td>Host species</td>
</tr>
<tr>
<td>Country</td>
</tr>
<tr>
<td>Chromosomal assembly‡, bp</td>
</tr>
<tr>
<td>Mean depth of coverage§</td>
</tr>
<tr>
<td>Chromosomal contigs</td>
</tr>
<tr>
<td>G + C content, %</td>
</tr>
<tr>
<td>Core protein-coding genes§ (% of PvP01)</td>
</tr>
<tr>
<td>Full length§ (% of PvP01)</td>
</tr>
<tr>
<td>Partial (% of PvP01)</td>
</tr>
<tr>
<td>Genes in hypervariable regions§</td>
</tr>
</tbody>
</table>

N/A, not applicable.

* The genome assembly of PV43 represents a consensus sequence of at least two major and three minor chimpanzee *P. vivax* variants (Fig. 2 C and D and SI Appendix, Fig. S5).

† Chimpanzee *P. vivax* genomes were compared with the human *P. vivax* reference PvP01 (24).

‡ Number of unambiguous bases.

§ Calculated by dividing the number of nucleotides in reads mapped to the assemblies by the expected genome size from PvP01.

The Duffy-negative mutation, while the current human-infecting parasites represent a lineage that had escaped out of Africa.
sequencing errors to maintain an ORF. Overall, PvSY56 and PvSY43 shared a highly conserved core genome with human P. vivax. More than 98% of PvP01 core genes (as defined in ref. 20) were identified in each chimpanzee P. vivax assembly (Table 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 as the reason 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 (Table 1).

**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, in contrast with a difference of only 0.11% between the two human P. vivax reference genomes, PvSall and PvP01 (SI Appendix, Fig. S1A). Since PvSY56 and PvSY43 were 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 Côte d’Ivoire (10), and from a western lowland gorilla (Gor3157) sampled in Cameroon (SI Appendix, Table S2). We also mined the read database from a blood pool of a Plasmodium malariae-infected sanctuary chimpanzee from Gabon (14), which we had noted contained a substantial number of ape P. vivax reads. Reads from each sample were mapped to the PvSY56 assembly, and SNPs were identified. The extent of genome coverage varied considerably among the six chimpanzee samples; however, we were able to recover between 695 and 3,005 core genes (SI Appendix, Table S2), with 65% of genes analyzed being covered in four or more parasite genomes (SI Appendix, Fig. S2). The gorilla P. vivax strain, which was derived from a partially degraded bushesheet sample, yielded only 10 genes despite repeated amplification and thus was not included in the diversity analysis. For comparison, we included sequences from seven additional human P. vivax strains (21), each from a different country (India, Myanmar, Papua New Guinea, Thailand, Colombia, Mexico, and Peru), and identified SNPs using the same methods. Diversity values were then calculated across a common set of 4,263 core genes for which we obtained sequences from two or more strains for both ape and human P. vivax (Table 2). The results revealed a mean pairwise nucleotide sequence diversity (π) among the six chimpanzee P. vivax strains of 0.698%, about eight times higher than the value (0.085%) for the global sample of human strains (Fig. 1A, Table 2, and SI Appendix, Fig. S1B). Removal of the PvSY43 sequence did not decrease this difference (SI Appendix, Fig. S1C), indicating that the inclusion of this multiply infected sample did not inflate the diversity of the chimpanzee P. vivax parasites. Furthermore, analysis of transition/transversion ratios at fourfold degenerate sites yielded nearly identical results for chimpanzee (1.08) and human (1.07) P. vivax strains, excluding the possibility of an SWGA-related increase in diversity. Thus, the much higher level of diversity among chimpanzee P. vivax strains, compared with parasites currently circulating in humans, does not appear to be an artifact.

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 an NS/S ratio of 1.37 (Table 2). The NS/S ratio among polymorphisms can be compared with (i.e., divided by) the NS/S ratio among interspecies differences to yield the neutrality index (NI) (25). This 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 3,913 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 = 0.96) (SI Appendix, Table S3). 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 (26). 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 (P < 0.05), 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 (27, 28), and 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 (Dataset S1). Such results are usually interpreted as evidence of balancing selection maintaining polymorphism, but the large overall NI value for the human strains suggests that a more pervasive factor, such as past demography, is influencing these genes. Fig. 1B shows the distributions of NI values for 1,585 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 (Fig. 1B and Dataset S1).

To look for possible human-specific adaptive changes, we also performed MK tests for 4,263 core genes comparing polymorphisms

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

<table>
<thead>
<tr>
<th>Parasites</th>
<th>n*</th>
<th>πm4</th>
<th>πm8</th>
<th>πs4</th>
<th>πs8</th>
<th>NS polymorphisms</th>
<th>S polymorphisms</th>
<th>NS/S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ape P. vivax</td>
<td>6</td>
<td>0.00698</td>
<td>0.000357</td>
<td>0.01604</td>
<td>32,364</td>
<td>47,494</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>Human P. vivax</td>
<td>9</td>
<td>0.000085</td>
<td>0.000060</td>
<td>0.00143</td>
<td>10,530</td>
<td>7,673</td>
<td>1.37</td>
<td></td>
</tr>
</tbody>
</table>

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

1 Mean pairwise diversity at coding sites from 4,263 genes (6.5 million sites).
2 Mean pairwise diversity at zerofold degenerate sites from 4,263 genes (4.0 million sites).
3 Mean pairwise diversity at fourfold degenerate sites from 4,263 genes (0.7 million sites).

Numbers 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 (PvSY56 for ape P. vivax; PvP01 for human P. vivax).
within ape *P. vivax* with fixed differences between ape and human *P. vivax*. After correcting for multiple tests, we found no genes with a significant excess of nonsynonymous fixed differences. While this may seem surprising, the timespan since the human lineage of *P. vivax* became restricted to this host may have been too short for the accumulation of a sufficient number of adaptive changes to yield significant test results, and this finding does not exclude the possibility that smaller numbers of adaptive changes have indeed occurred.

To investigate the unusual pattern of nucleotide polymorphisms in human *P. vivax* 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 fourfold and zerofold 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 zerofold degenerate sites (0.357%) was 22% of that at fourfold degenerate sites (1.604%), whereas among human *P. vivax* strains, the value at zerofold degenerate sites (0.060%) was 42% of that at fourfold degenerate sites (0.143%) (Table 2 and SI Appendix, Fig. S1 D and E). Thus, while chimpanzee parasites were 11 times more diverse than human parasites at fourfold degenerate sites, they were only six times more diverse at zerofold 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 (20, 21), so it is possible to investigate the frequencies at which SNPs are segregating within the population. Synonymous polymorphisms are expected to be neutral, so their site frequency spectrum (SFS) should reflect past demography. By comparison, many nonsynonymous polymorphisms are expected to be slightly deleterious, and their SFS thus should be more skewed toward lower frequencies (29). To examine a large number of parasite sequences from a single geographic region, we focused on SNP data from Malaria Genomic Epidemiology Network (MalariaGen) samples from Southeast Asia (20). 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 zerofold and fourfold degenerate sites are almost identical (Fig. 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 (7). This was observed for mitochondrial and apicoplast sequences as well as for three nuclear genes, while for a fourth nuclear gene the ape and human *P. vivax* sequences formed sister clades (7). 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 3,958 core genes, the chimpanzee parasite genomes PVSY56 and PVSY43 differed from one another at 0.6% of sites but differed from the human *P. vivax* reference genomes PVsul1 and PVp01 at 2.2% of sites on average. This relationship is summarized in a neighbor-joining tree constructed from a matrix of pairwise genetic distances from an alignment of 241 nuclear genes available for all six chimpanzee parasites (Fig. 24). Although this tree may not reflect the true evolutionary history of any one particular gene (due to recombination), the overall relationships were confirmed in a phylogenetic network (SI Appendix, Fig. S34), which showed that the 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 (GAI2) and Côte 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 (Fig. 2B and SI Appendix, Fig. S3B), 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.

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 (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
Fig. 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 241 nuclear genes is shown for nine human (black) and six chimpanzee (red) P. vivax strains. (B) Inset shows the human P. vivax strains in greater detail. (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 (SI Appendix, Table S2); all other sequences except for P. cynomolgi and P. knowlesi were generated by SGA and include a code identifying their geographic origin (SI Appendix, Fig. S5), ape subspecies (G.g.g., Gorilla gorilla gorilla; P.t.e., Pan troglodytes elliottii; P.t.t., Pan troglodytes troglodytes), and sample number (see SI Appendix, Table S4 for GenBank accession numbers). Bootstrap values ≥70 are shown for clades with two or more nonidentical tips. Fragment lengths in Pvp01 are indicated above the trees. The scale bars indicate substitutions per site (see also SI Appendix, Fig. S3 for phylogenetic network representations).

genes yielded tree topologies in which human and ape parasites represented sister clades (SI Appendix, Fig. S4). 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 (SI Appendix, Fig. S5). Each DNA preparation was diluted so 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 found in only two wild chimpanzees, which is distinct from but most closely related to the P. vivax clade (7, 30). 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 in which the previous topology depicted ape and human strains as sister clades, with three of these four trees including gene sequences for P. carteri (Fig. 2C and SI Appendix, Fig. S6). Thus, the 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 (Fig. 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 ORFs for rhp 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 (13). 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 (DBBP2, or EBP) (31), but no additional DBP-like genes were identified.

Variation in the complement of RBPs is thought to influence the ability of Plasmodium parasites to invade erythrocytes (13, 32). Human P. vivax, which exclusively invades reticulocytes, has full-length ORFs for five rhp genes (rhp1a, rhp1b, rhp2a, rhp2b, and
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 (33). We identified orthologs of both of these partial genes in the chimpanzee P. vivax genomes (Fig. 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, synonymous-to-nonsynonymous substitution (dN/dS)-based tests for positive selection (34) in inversion genes on the branch leading to human P. vivax (rbp2p1, rbp1a, rbp1b, rbp2a, and rbp2b) 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 Plasmodium inui, while Plasmodium fragile has an intact rbp2e gene but a rbp3 pseudogene (Fig. 3A). So far, rbp2d has been identified only in P. vivax. The two chimpanzee P. vivax genomes PvSY56 and PvSY43 contained full-length intact ORFs 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 (Fig. 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 (20, 21) to the rbp2d, rbp2e, and rbp3 reference genes and analyzed those that yielded a greater than threefold 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 (Fig. 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.

Fig. 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 by 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).
To examine whether the other chimpanzee *P. vivax* strains (*SI Appendix, Table S1*) contained any of the *rhp2d*, *rhp2e*, and *rhp3* inactivating mutations, we mapped available sequencing reads to the respective P*SYS6* genes. We also used SGA to amplify the same regions from *P. vivax*-positive gorilla samples. Although in most instances the coverage of the *rhp2d*, *rhp2e*, and *rhp3* genes was incomplete, none of the recovered sequences contained the frameshift and stop codon mutations found in all human *P. vivax* strains (*SI Appendix, Fig. S7*). This was also true for SGA-derived gorilla parasite sequences from the multiply infected SAggg3157 sample, which contained a number of polymorphisms, 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 host tropism.

**Recombinant RBP2e and RBP3 Do Not Exhibit Species-Specific Red Blood Cell Binding.** The pseudogenization of *rhp2d*, *rhp2e*, and *rhp3* 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 N-terminal domain of RBPs encoded by human *P. vivax* (RBP2a160–1000 and RBP2e161–969) have been used to characterize their erythrocyte-binding properties (35, 36). These studies showed that RBP2a binds the reticulocyte-specific transferrin receptor 1 (TIR1), also termed “CD71” (35), while RBP2a binds an unknown receptor present on both normocytes and reticulocytes (36). To examine the function of chimpanzee *P. vivax* RBP2d, RBP2e, and RBP3 proteins, we expressed their N-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 α-helical and β-sheet content similar to human *P. vivax* RBP2b (*SI Appendix, Fig. S8*). Because some RBPs bind only reticulocytes, we attempted to enrich these cells from blood samples obtained from four humans, four chimpanzees, and one gorilla using a Percoll density gradient as previously described (35, 36). 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 (*SI Appendix, SI Materials and 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, RBP2a150–1000 and RBP2e161–969 (35, 36). 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 (35, 36). Reticulocytes were stained with TO before flow cytometry (Fig. 4A). Consistent with previous results, we observed robust binding of RBP2a to both human normocytes (10.3% of TO-negative cells) (Fig. 4B) and reticulocytes (32.4% of TO-positive cells) (Fig. 4C). Interestingly, a similar binding profile was observed for gorilla and chimpanzee red blood cells (Fig. 4 and *SI Appendix, Fig. S9*). As expected, RBP2b exhibited a strong preference for reticulocytes, binding 19.9% of human reticulocytes (Fig. 4C) but only a minor fraction (0.9%) of human normocytes (Fig. 4B), likely reflecting incomplete reticulocyte staining and/or nonspecific binding. RBP2b also bound chimpanzee and gorilla reticulocytes, albeit at a reduced level (Fig. 4C). Although the TIR1 proteins of chimpanzees and gorillas differ from their human counterpart by a few amino acids (*SI Appendix, Fig. S10*),

![Fig. 4. Binding of RBPs to ape and human red blood cells.](image-url)

*Fig. 4. Binding of RBPs to ape and human red blood cells. (A) Dot plots 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-labeled) secondary antibody (y axis), and reticulocytes were identified by staining with 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. Experiments were performed as three technical replicates with the background signal from the antibody-only control subtracted from each binding result.*
none of these residues was identified as representing critical RBP2b contact sites (37). Thus, the decreased binding of RBP2b to ape reticulocytes is unlikely to be the result of sequence differences between chimpanzee, gorilla, and human TIR1 proteins and may instead reflect differences in TIR1 expression levels, posttranslational modifications, or other factors.

Having validated the experimental system, we next tested the binding of chimpanzee P. vivax RBP2e156–957 and RBP3149–966 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 (Fig. 4 B and C). Like the human P. vivax RBP2a and RBP2b proteins, RBP2e and RBP3 appeared to bind reticulocytes more efficiently than normocytes (Fig. 4 and SI Appendix, Fig. S9). However, there was a clear difference in host-specificity in that although RBP3 bound gorilla reticulocytes slightly more efficiently than chimpanzee and human reticulocytes, this result must be interpreted with caution, since only a single gorilla sample containing very few reticulocytes was available for testing (Fig. 4D). Indeed, when red blood cells from a macaque were tested, RBP2e and RBP3 were found to bind to reticulocytes from this host species also (SI Appendix, Fig. S9).

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 less than 1% of red cells in this sample yielded twice as many reticulocytes (4%), this higher fraction did not improve RBP2e and RBP3 binding (SI Appendix, Fig. S11). 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.

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 those infecting humans in Africa (11). The parasites infecting humans and apes are largely allopatric, but as yet there is no clear evidence for host specificity. Although RBP3 bound more efficiently than normocytes (Fig. S11). 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.

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 those infecting humans in Africa (11). The parasites infecting humans and apes are largely allopatric, but as yet there is no clear evidence for host specificity. Although RBP3 bound more efficiently than normocytes (Fig. S11). 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.

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 those infecting humans in Africa (11). The parasites infecting humans and apes are largely allopatric, but as yet there is no clear evidence for host specificity. Although RBP3 bound more efficiently than normocytes (Fig. S11). 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.

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 those infecting humans in Africa (11). The parasites infecting humans and apes are largely allopatric, but as yet there is no clear evidence for host specificity. Although RBP3 bound more efficiently than normocytes (Fig. S11). 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.

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 those infecting humans in Africa (11). The parasites infecting humans and apes are largely allopatric, but as yet there is no clear evidence for host specificity. Although RBP3 bound more efficiently than normocytes (Fig. S11). 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.

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 those infecting humans in Africa (11). The parasites infecting humans and apes are largely allopatric, but as yet there is no clear evidence for host specificity. Although RBP3 bound more efficiently than normocytes (Fig. S11). 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.
out of Africa. Under the bottleneck scenario, the most recent common ancestor of human \(P. \text{vivax}\) 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 y ago (44). Molecular clock estimates have placed the \(P. \text{vivax}\) most recent common ancestor (MRCA) at least 50,000–70,000 y ago (6, 45), but these relied on rate assumptions that may not be accurate. We have argued that the MRCA of human \(P. \text{vivax}\) may have left Africa in a more recent wave of human migration, although its higher levels of genetic diversity (21) suggest that human \(P. \text{vivax}\) is older than \(P. \text{falciparum}\). Once out of Africa, \(P. \text{vivax}\) spread through Asia and Europe and probably from Europe into the Americas (39, 46). Strains of \(P. \text{vivax}\) now present in Madagascar and East Africa, in areas where nonhuman apes are absent, likely reflect reintroductions from Asia (47). Given this historical isolation of ape and human \(P. \text{vivax}\) strains, substantial gene flow between the two populations is unlikely. The mixture of topologies found for different genes, with some showing separation of the ape and human parasite lineages and others having 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 \(P. \text{vivax}\) exhibit broad natural tropism (7, 11, 15, 30), 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 \(P. \text{vivax}\) infection of African humans have been increasing, including instances of infection of Duffy-negative individuals (4). 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.

**Methods**

**Sample Collection, DNA Extraction, and Plasmidum Screening.** Blood samples were obtained from chimpanzees at the Sananga-Yomping Chimpanzee Rescue Center following routine veterinary examination. Blood was also collected from a wild-living habituated chimpanzee from the Tai Forest in Côte d’Ivoire during emergency surgery (48). The gorilla blood sample (Gor3157, also referred to as “SAgg3157”) 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 permissions. DNA was extracted from whole-blood samples using the QiAmp Blood DNA Mini Kit or the Puregene International Trade in Endangered Species of Wild Fauna and Flora regulations. Strains of \(P. \text{vivax}\) now present in Madagascar and East Africa, in areas where nonhuman apes are absent, likely reflect reintroductions from Asia (47). Given this historical isolation of ape and human \(P. \text{vivax}\) strains, substantial gene flow between the two populations is unlikely. The mixture of topologies found for different genes, with some showing separation of the ape and human parasite lineages and others having 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 \(P. \text{vivax}\) exhibit broad natural tropism (7, 11, 15, 30), 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 \(P. \text{vivax}\) infection of African humans have been increasing, including instances of infection of Duffy-negative individuals (4). 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.

**Whole blood from five chimpanzees (New Iberia Research Center), one gorilla (Lincoln Park Zoo), and one rhesus macaque (BiovIT) were collected in ACD-A collection tubes (BD Biosciences). All ape blood samples were leftover technical replicates with the background signal from the antibody-only position. Experiment was performed as three biological replicates with the background signal from the antibody-only position. Experiment was performed as three biological replicates with the background signal from the antibody-only position.
ACKNOWLEDGMENTS. We thank Andrew Smith, Catherine Bahari, and Alex Shugart for performing Illumina sequencing; the University of Delaware Sequencing Core for PacBio sequencing; Dana Bellsisimo for assistance with flow cytometry; the staff of the Sanaga-Yong Rescue Center and Jane Fontenot, Melany Musso, and Francois Villinger at the New Iberia Research Center for providing blood samples from captive chimpanzees; Marisa Schender from the Lincoln Park Zoo for providing leftover gorilla blood; the Tai 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; and the Ministry of Environment and Forests and Research, Office Ivorian des Parcs et Réserves, Tai 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 NIH Grants R01 AI097137, R01 AI091593, R37 AI050529, and P01 AI050089 (B.H.H.); a grant from the Agence Nationale de Recherche (Programme Blanc, Sciences de la Vie, de la Santé et des Écosystèmes, ANR 11 BSV3 021 01, Projet PRIMAL) (to W.H.T.). D.E.L. was supported by an NIH Training Grant T32 AI 07532; O.A.M. was supported by a Feitoechromocytology and Biological Sciences Research Council Grant BB/M010996/1 (EastBIO); and A.L.K.M. was supported by Wellcome Trust PhD Programme Grant 108905/2/15/Z.