The Journal of Infectious Diseases

MAJOR ARTICLE

Plasmodium simium: Population Genomics Reveals the Origin of a Reverse Zoonosis

Thaís C. de Oliveira,[1,a] Priscila T. Rodrigues,[1,a] Angela M. Early,[2,3] Ana Maria R. C. Duarte,[4,5] Julyana C. Buery,[6] Marina G. Bueno,[7,8] José L. Catão-Dias,[7] Crispim Cerutti Jr.,[6] Luísa D. P. Rona,[9,10] Daniel E. Neafsey,[2,3,b] and Marcelo U. Ferreira[1,b]

[1]Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil, [2]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA, [3]Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, USA, [4]Laboratory of Biochemistry and Molecular Biology, Superintendency for the Control of Endemics (SUCEN), State Secretary of Health, São Paulo, Brazil, [5]Laboratory of Protozoology, Institute of Tropical Medicine of São Paulo, University of São Paulo, São Paulo, Brazil, [6]Department of Social Medicine, Center for Health Sciences, Federal University of Espírito Santo, Vitória, Brazil, [7]Laboratory of Wildlife Comparative Pathology, Department of Pathology, School of Veterinary Medicine and Animal Sciences, University of São Paulo, São Paulo, Brazil, [8]Laboratory of Comparative and Environmental Virology, Oswaldo Cruz Institute, Fiocruz, Rio de Janeiro, Brazil, [9]Department of Cell Biology, Embryology, and Genetics, Federal University of Santa Catarina, Florianópolis, Brazil, [10]National Council for Scientific and Technological Development, National Institute of Science and Technology in Molecular Entomology, Rio de Janeiro, Brazil

Background. The population history of *Plasmodium simium*, which causes malaria in sylvatic Neotropical monkeys and humans along the Atlantic Coast of Brazil, remains disputed. Genetically diverse *P vivax* populations from various sources, including the lineages that founded the species *P simium*, are thought to have arrived in the Americas in separate migratory waves.

Methods. We use population genomic approaches to investigate the origin and evolution of *P simium*.

Results. We find a minimal genome-level differentiation between *P simium* and present-day New World *P vivax* isolates, consistent with their common geographic origin and subsequent divergence on this continent. The meagre genetic diversity in *P simium* samples from humans and monkeys implies a recent transfer from humans to non-human primates – a unique example of malaria as a reverse zoonosis of public health significance. Likely genomic signatures of *P simium* adaptation to new hosts include the deletion of >40% of a key erythrocyte invasion ligand, PvRBP2a, which may have favored more efficient simian host cell infection.

Conclusions. New World *P vivax* lineages that switched from humans to platyrrhine monkeys founded the *P simium* population that infects nonhuman primates and feeds sustained human malaria transmission in the outskirts of major cities.

Keywords. *Plasmodium simium*; Neotropical monkeys; reverse zoonosis.

Despite the recent progress towards malaria elimination, more than 700 000 infections are reported each year in the Americas, 72% of them due to *Plasmodium vivax* [1]. Malaria is primarily transmitted by mosquitoes of the *Nyssorhynchus* subgenus, such as *Anopheles darlingi* in the Amazon Basin and *Anopheles albimanus* along the Pacific Coast of South America, Central America, and Mexico [2] (Figure 1). However, focal malaria transmission is sustained along the coast of Southeast and South Brazil by anopheline mosquitoes of the *Kerteszia* subgenus, mainly *Anopheles cruzii* and *Anopheles bellator*, that breed in water trapped by bromeliads in the Atlantic Forest biome [3].

Plasmodium vivax had already emerged as a human parasite in the Old World at the time the first modern humans settled in the Americas, approximately 15 000 years ago [4]. This parasite is thought to have originated from crossover infections from non-human primates or evolved from an ancestral parasite that infected great apes and hominids in Africa [5–7]. It was surprising to find that a closely related parasite named *Plasmodium simium* infects Neotropical platyrrhine monkeys [8], which diverged from African great apes and Old World catarrhine monkeys 35 million years ago [9]. Originally described in howler monkeys [10], *P simium* also infects wild woolly, spider, capuchin, and titi monkeys on the Atlantic Coast [8, 11, 12] but not in the Amazon [8, 13]. Parasites are genetically very similar to *P vivax* [14–17]. Cross-species switches are thought to be favored by the predilection of *Kerteszia* vectors to take bloodmeals both at the canopy of the trees, where monkeys live, and at the ground level, where humans may be found [18]. *Plasmodium simium* remains infectious to humans [19–21] and its zoonotic transmission challenges malaria elimination efforts in South America [22, 23].

How *P simium* arrived and subsequently evolved in the Americas remains disputed. Li et al [24] postulated 2 separate introductions of *P vivax* into the continent: one from Europe and Africa during the colonial era, founding the New World *P vivax* populations that circulate in the Amazon and along the Pacific Coast of northern South America, Central America, and Mexico, and another from Southeast Asia, founding the so-called Old

Received 19 February 2021; editorial decision 9 April 2021; accepted 14 April 2021; published online April 19, 2021.

[a]T. C. d. O. and P. T. R. are co-first authors and contributed equally to this article.

[b]D. E. N. and M. U. F. are co-senior authors and contributed equally to this article.

Presented in part: 67th Annual Meeting of the American Society of Tropical Medicine and Hygiene, October 30, 2018, New Orleans, LA; 7th International Conference on *Plasmodium vivax* Research, June 24–28, 2019, Paris, France.

Correspondence: Marcelo U. Ferreira, Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil (muferrei@usp.br).

The Journal of Infectious Diseases® 2021;224:1950–61

DOI: 10.1093/infdis/jiab214

World-type lineages, including *P simium*, that nowadays circulate on the Atlantic Coast of Brazil. Cormier [25] went further to speculate that *P simium* arrived in Rio de Janeiro with East Asian migrant workers in the early 1800s. In this study, we use population genomic approaches to reveal the close genetic affinity of *P simium* with present-day New World rather than Old World-type *P vivax*, contrary to the expectations of the Asian origin hypothesis. After their arrival in coastal Brazil, some *P vivax* lineages have adapted to locally abundant *Kerteszia* vectors and jumped from humans to sylvatic nonhuman primates, especially howler monkeys, offering an uncommon example of reverse zoonosis of public health significance.

METHODS

Study Sites and Samples

We retain the species name *P simium* to refer to monkey- and human-derived malaria parasites from southeastern Brazil that morphologically resemble *P vivax* [20, 21]. Simian parasite deoxyribonucleic acid (DNA) for genome sequencing was obtained from blood samples drawn between 2003 and 2010 during routine veterinary care of 3 brown howler monkeys (*Alouatta guariba clamitans*) [26, 27] and 1 black-fronted titi monkey (*Callicebus nigrifrons*) [11] from Atlantic Forest fragments around São Paulo, the largest city in South America (Figure 1, Supplementary Figure 1, and Supplementary Table 1). Human *P simium* DNA samples were obtained between April 2001 and March 2004 from 7 febrile patients from Atlantic Forest remnants in the state of Espírito Santo [28], 740 kilometers north of São Paulo (Figure 1, Supplementary Figure 1, and Supplementary Table 1). Consistent with locally acquired infections, patients reported no recent travel to the Amazon, the main malaria-endemic area in Brazil [29].

Genome Sequencing

We enriched DNA isolated from unprocessed blood samples for target parasite DNA by selective whole-genome amplification with the primer set *pvset1* [30] to prepare Nextera XT or TrueSeq Nano DNA libraries (Illumina, San Diego, CA). Paired-end short sequence reads from Illumina HiSeq 2500 or HiSeq X platforms that presented expected base call accuracy ≥99.9% were processed as described in Supplementary Methods. In brief, we mapped reads with ≥5× coverage onto the reference *P vivax* genome PvP01 [31] after excluding repetitive domains to which short reads are difficult to map [32]. Single-nucleotide polymorphisms (SNPs) were called, and those presenting within-sample variation in an unusually high percentage of samples were filtered out (Supplementary Methods). Heteroallelic SNP calls (0.01% to 0.05% of all SNPs per sample) were masked before downstream analyses. For comparison, we processed and analyzed raw genome sequence reads from worldwide *P vivax* isolates (Supplementary Table 2) in the same way.

Analysis of the *Plasmodium simium pvs47, pvdpb1,* and *pvrbp2a* Gene Orthologs

We next focused on the *pvs47* (PVP01_1208000) gene encoding a gamete surface protein that may have facilitated *P simium* adaptation to *Kerteszia* mosquitoes. (Given the close similarity between *P vivax* and *P simium* orthologs, throughout this paper we use the same gene names to refer to them.) This protein controls parasite-vector compatibility in *P falciparum* [33] and putatively also in *P vivax* [34]. A maximum likelihood tree was generated with full-length translated Pvs47 sequences (434 amino acids) from *P simium* and worldwide *P vivax* isolates (Supplementary Table 3) using IQ-TREE 1.5.5 [35].

Because the efficiency of simian cell invasion may constitute a further barrier to vertebrate host switching, we also analyzed the genes that encode the Duffy Binding Protein 1 (*pvdbp1*; PVP01_0623800) and the Reticulocyte Binding Protein 2a (*pvrbp2a;* PVP01_1402400), a member of the reticulocyte binding protein (*rbp*) family. These are key ligands of host erythrocyte receptors [36] that display large deletions in *P simium* [17]. To detect deletions, we polymerase chain reaction (PCR)-amplified gene fragments and determined amplicon sizes by agarose gel electrophoresis; selected PCR products were Sanger-sequenced. The sample set analyzed and laboratory protocols are described in Supplementary Methods and Supplementary Table 3.

Data Analysis

We assessed newly obtained *P simium* genomes for genetic diversity and linkage disequilibrium (LD) as described in Supplementary Methods. We measured π (average number of pairwise nucleotide differences per site) and examined how it varies within 1-kb sliding windows across the genome. Tajima's *D* values were calculated within 1-kb windows and their distribution was plotted. The expected value of Tajima's *D* is zero under a strictly neutral model. We assessed multilocus LD by calculating the V_D/V_E ratio—the variance of the distribution of empirical pairwise genetic distances between samples (V_D) divided by that expected under random association of alleles in a panmictic population (V_E) [37]. Under panmixia, the V_D/V_E ratio is expected to be 1.

We used the hmmIBD software [38] to infer relatedness between *P simium* samples from monkeys and humans and across *P simium* and *P vivax* populations using identity-by-descent (IBD) analysis. Relatedness networks were drawn to represent sample pairs that share at least 5%, 20%, or 50% of their genomes. Principal component analysis (PCA) was carried out to explore the genetic affinities between *P simium* and regional populations of *P vivax*. We used the R package *admixr* [39] to compute the *f*4 statistic, *f*4 (W, X; Y, Z) [40], where (except if otherwise stated) X is *P simium* and Y and Z are 2 *P vivax* populations, to compare the proportion of shared derived alleles of *P simium* with every pair of *P vivax* populations. The

Figure 1. Map of Latin America and the Caribbean showing the distribution of *Anopheles* species and sites of collection of malaria parasite samples analyzed. The collection sites are shown for newly sequenced *Plasmodium simium* isolates (n = 11) and the New World *P vivax* samples from Brazil ([BRA] n = 30), Peru ([PER] n = 12), Colombia ([COL] n = 14), and Mexico ([MEX] n = 15) that were used in population genomic analysis; circle diameters are proportional to sample size. Color shading indicates the spatial distribution of *Anopheles darlingi*, *Anopheles albimanus*, *Anopheles cruzii*, and *Anopheles bellator*, the primary malaria vectors in the areas of sample collection.

outgroup (W) is *P cynomolgi*, the *vivax*-like parasite of Asian macaques [41].

Data Availability

Sequence data were deposited in the Sequence Read Archive ([SRA] SAMN17926899–SAMN17926909) and GenBank Nucleotide Database (MW560996–MW561043, MW561060–MW561072, and MW561044–MW561059) of the National Center for Biotechnology Information.

RESULTS

Genome-Wide Diversity of *Plasmodium simium*

We generated 11 new genome sequences of *P simium* (4 from monkeys and 7 from human hosts), with a mean read depth of 242× and 23.3 × 10^6 reads per isolate on average (Table 1 and Supplementary Table 1; Supplementary Results). Samples displayed a low complexity of infection, with F_{WS} (within-host diversity statistic) values ranging between 0.904 and 0.988. After

Table 1. Characteristics of Newly Generated *Plasmodium simium* Genome Sequences From Southeastern Brazil[a]

Isolate	Host Species	Average Read Depth (×)	PvP01 Genome Coverage (%)	Number of Reads	Number of SNPs
97Ps	*Alouatta clamitans*	316	20.5	30 564 162	1042
D121Ps	*A clamitans*	329	11.7	31 789 162	863
P160	*Callicebus nigrifrons*	284	4.8	27 472 914	610
95	*A clamitans*	117	12.4	11 349 094	356
ALNL53	*Homo sapiens*	262	79.8	25 326 697	1877
143	*H sapiens*	83	11.2	8 033 616	1848
111	*H sapiens*	106	26.0	10 294 134	1337
MASM	*H sapiens*	365	19.0	35 279 795	775
761	*H sapiens*	278	59.3	26 838 145	721
1565MT	*H sapiens*	259	49.0	25 070 079	2080
1272MT	*H sapiens*	257	82.4	24 858 660	2098

Abbreviations: SNPs, single-nucleotide polymorphism.

[a]A complete version of this table is provided as Supplementary Table 1 available online.

excluding positions with >10% missing calls in the complete set of 11 *P simium* sequences and 121 *P vivax* sequences retrieved from SRA, we ended up with a catalog of 10 738 high-confidence variable positions that were used in further analyses of *P simium* samples. The nonreference allele was called at 356 to 2098 genotyped sites per sample (mean, 1237), with 44.3% of the SNPs in intergenic regions, 6.8% in introns, and 48.9% in coding regions; 65.0% of coding SNPs were nonsynonymous.

The average genome-wide nucleotide diversity of *P simium* ($\pi = 1.63 \times 10^{-4}$) is approximately 3 times lower than that of *P vivax* populations from the Americas (Supplementary Figure 2A). Despite the small sample size, we note that *P simium* samples of simian origin are significantly less diverse than their counterparts of human origin ($\pi = 1.18 \times 10^{-4}$ vs 1.75×10^{-4}, $P = .025$, Mann-Whitney U test), consistent with a bottleneck at the time parasites jumped into monkeys, with no barrier to infecting humans again following this original host switch. Domains with the top 1% π values (Supplementary Figure 3) include genes coding for the following: (1) a *Plasmodium* helical interspersed subtelomeric (PHIST) protein family member (PVP01_1470300) [42]; (2) 28S (PVP01_0504500) and 18S (PVP01_0202900) ribosomal RNA; and 2 proteins involved in red blood cell invasion, (3) RBP2a (PVP01_1402400) [43] and (4) cysteine-rich protective antigen (CyRPA) (PVP01_0532400) [44] (Supplementary Table 4). *Plasmodium vivax* genes from parasites circulating in the Amazon Basin of Brazil (BRA) are, on average, twice as diverse as their *P simium* orthologs; in contrast, *pvrbp2a* is 2.6-fold more diverse in *P simium* than *P vivax* (Supplementary Results and Supplementary Figure 4). No significant enrichment of any gene ontology (GO) term (http://geneontology.org/) was found among annotated genes within high-diversity domains (Supplementary Table 5).

Negative Tajima's *D* values predominate within 1-kb windows (median *D*, −0.543) (Supplementary Figure 2B), with an excess of low-frequency SNPs consistent with a *P simium* population expansion after a recent bottleneck. Coding sequences within domains with extremely high or low Tajima's *D* scores (Supplementary Figure 3), suggestive of either balancing or directional selection, are listed in Supplementary Tables 6 and 7. There is no significant enrichment of any GO term among annotated genes within domains with the top-1% or bottom-1% Tajima's *D* scores (Supplementary Tables 8 and 9).

Population Structure and Shared Ancestry

We examined LD levels to test whether simian and human *P simium* are part of a single panmictic population. We found significant multilocus LD at the genome level, consistent with *P simium* being structured into discrete, highly inbred lineages, with V_D largely exceeding the variance expected under panmixia (Supplementary Figure 5). Identity-by-descent analysis indicated substantial relatedness, including between simian and human isolates from sites >700-km apart (Figures 2B and C). Eight of the 11 P. simium isolates share >50% of their genomes (equal to the average ancestry between meiotic siblings) and 10 share >20% of their genomes with at least one sample. The exception is isolate P160, derived from a titi monkey in São Paulo, which shares <10% ancestry with all other *P simium* isolates (Figure 2A). However, because P160 genome reads covers only 4.8% of the PvP01 reference (Table 1), its genetic affinities must be interpreted with caution (Supplementary Results). We identified a cluster of closely related parasites that includes 5 Espírito Santo human samples from 2002 ro 2003 and 3 São Paulo simian samples from 2003 and 2009. These findings suggest that the P. simium population may experience a small degree of stratification through both space and time.

Geographic Origin of *Plasmodium simium*

The PCA with pairwise nucleotide distances from a global data set of genomes reproduces the clear divide between New and Old World populations of *P vivax* found in previous studies [34]. *Plasmodium simium* clusters with the New World populations of *P vivax*—especially from Mexico

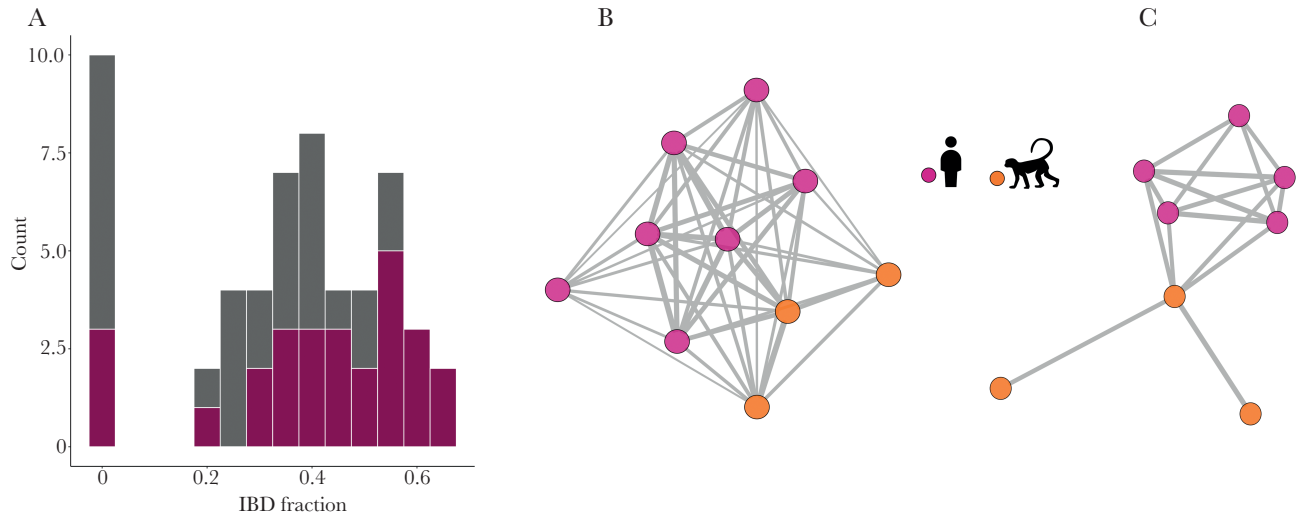

Figure 2. Shared genome ancestry in *Plasmodium simium* samples from monkeys and humans inferred by identity-by-descent (IBD) analysis. (A) Frequency distribution of the pairwise proportion of shared genome-wide ancestry between *P simium* samples. Pink bar segments represent pairwise comparisons of samples from the same vertebrate host (either humans or monkeys), whereas gray bar segments represent comparisons between different host types. (B and C) Parasite relatedness networks. Edges connect parasite pairs with shared ancestry ≥20% (B) or ≥50% (C) and unconnected samples are omitted. Edge thickness is proportional to the IBD fraction.

(MEX) (Figure 3A), a relatively inbred population that has experienced a steady decline in recent years [34]. Identity-by-descent analysis reveals qualitatively similar clustering patterns (Figure 3B), and only parasite pairs from the same geographic region share ≥5% ancestry, which indicates limited gene flow between continents. There are 2 noteworthy exceptions: (1) the low-coverage *P simium* isolate P160 seems to be connected to Ethiopian (ETH) *P vivax* isolates, and (2) the *P vivax* samples from India (IND) and ETH are fairly connected to each other. The latter finding supports the hypothesis of recolonization of Africa by South Asian *P vivax* stocks after the local fixation of the Duffy-negative phenotype, which confers resistance to blood-stage *P vivax* infection [45].

Using *f*4 statistic, we show that *P simium* shares more derived alleles with *P vivax* samples from the Americas, followed by IND, ETH, and the recently characterized European isolate (EUR) [46], compared with any population from East and Southeast Asia and Papua New Guinea (PNG) (Supplementary Table 10). These results rule out an Asian origin of *P simium*. The greatest genetic affinity of *P simium* is with *P vivax* from MEX, followed by Colombia (COL), Peru (PER), and BRA, with the lowest affinity with PNG. Figure 3C shows the west-to-east longitudinal cline of shared ancestry between *P simium* and all *P vivax* populations represented in our global sample, in relation to PNG.

We next explored the genetic affinities of individual isolates of *P simium*. Each of them displays a greater genetic similarity to MEX than with the BRA, PER, COL, ETH, IND, or EUR populations of *P vivax*, with the exception of P160, which appears to be closer to ETH than to American populations (Figure 4A). We also compared the extent of shared ancestry of parasites from Brazil—*P simium*, from the Atlantic Coast, and BRA, from the Amazon—with the non-BRA *P vivax* populations relative to PNG. We show that MEX is significantly closer to *P simium* than to BRA, whereas PER and COL are nearly equidistant of *P simium* and BRA. Conversely, all Old World *P vivax* populations are significantly closer to BRA than to the *P simium* population (Figure 4B). This suggests that parasites from diverse geographic origins founded the highly diverse present-day BRA population [16, 47]. Conversely, with the notable exception of P160, *P simium* appears to have descended primarily from less admixed New World stocks of *P vivax*, similar to the MEX population.

There is minimal genetic differentiation between *P simium* and *P vivax* populations from the Americas (Supplementary Figure 6; Supplementary Results). The average divergence between BRA and *P simium* (between-species $\pi = 4.92 \times 10^{-4}$) is identical to the pairwise polymorphism within the BRA population (within-species $\pi = 4.93 \times 10^{-4}$; $P = .8504$, Mann-Whitney *U* test). Moreover, the differentiation between BRA and MEX (within-species $\pi = 6.10 \times 10^{-4}$) is significantly greater than the average divergence between MEX and *P simium* (between-species $\pi = 4.74 \times 10^{-4}$; $P = 4.1 \times 10^{-15}$, Mann-Whitney *U* test) and that between BRA and *P simium* ($P = 2.0 \times 10^{-16}$, Mann-Whitney *U* test).

Pathway to a Reverse Zoonosis

Plasmodium simium experimentally infects anophelines from North America, Africa, South Asia, Southeast Asia, and the

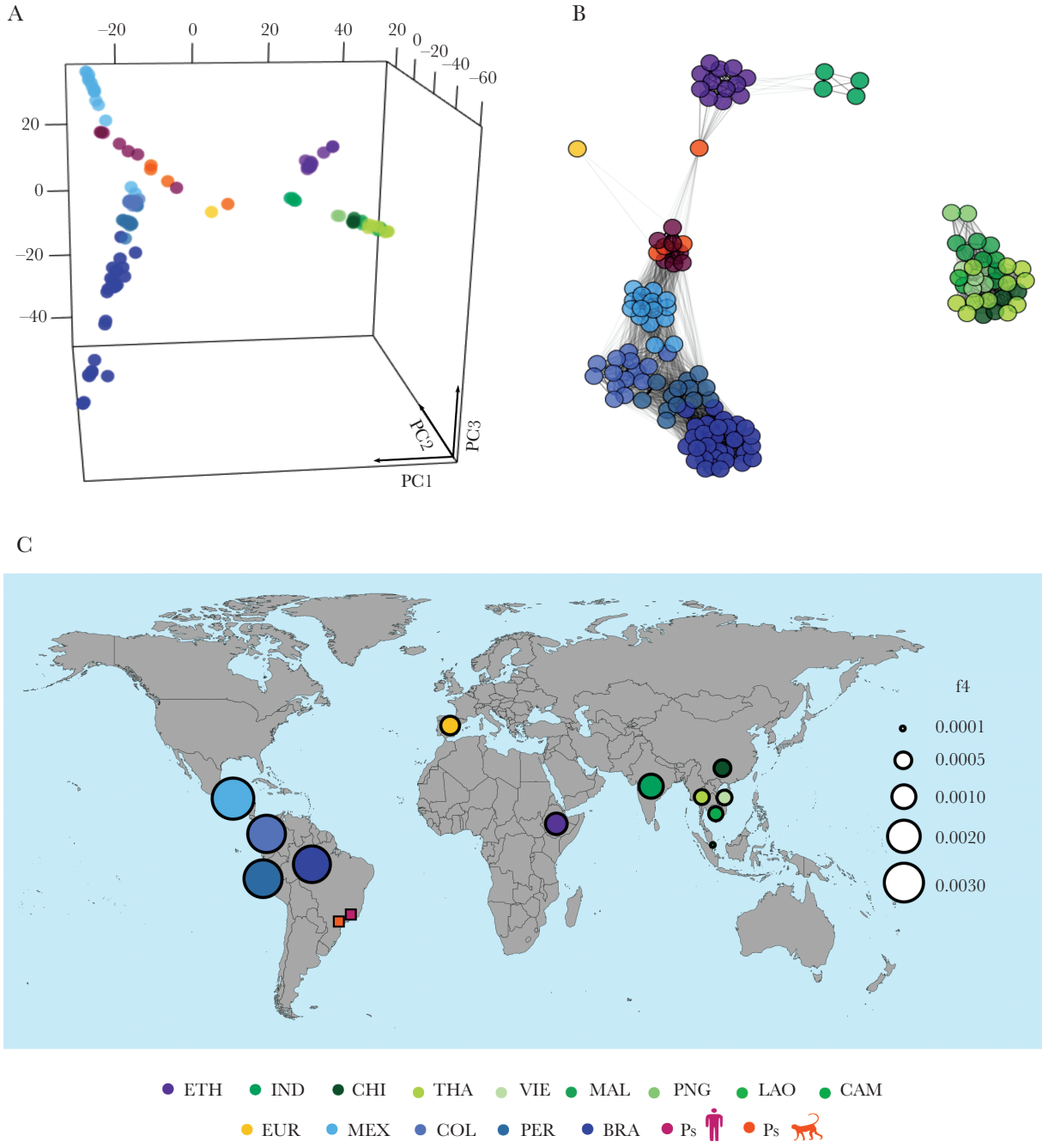

Figure 3. Genetic relatedness between *Plasmodium simium* and a global collection of *P vivax* isolates. (A) Three-eigenvector principal component analysis plot using genome-wide single-nucleotide polymorphism variation data from *P simium* and *Plasmodium vivax* isolates listed in Supplementary Table 2. Each circle represents an isolate, which is colored according to its country of origin. Populations are abbreviated as follows: BRA, Amazonian Brazil; PS, Atlantic Forest of Brazil; CAM, Cambodia; CHI, China, COL, Colombia; ETH, Ethiopia; EUR, Ebro Delta of Spain, Europe; IND, India; LAO, Laos; MAL, Malaysia, MEX, Mexico; PER, Peru; PNG, Papua New Guinea; THA, Thailand; and VIE, Vietnam. The percentage contributions of each eigenvector are as follows: PC1, 11.8%; PC2, 5.6%; and PC3, 3.9%. (B) Global parasite relatedness network inferred by identity-by-descent analysis. Edges connecting parasites are shown for pairs with ≥5% shared ancestry and unconnected isolates are omitted. Edge thickness is directly proportional to the proportion of shared ancestry, with thick edges indicating greater relatedness. (C) Values of *f*4 statistic, under the relationship *f*4 (W, X; Y, Z), where W is the macaque parasite *Plasmodium cynomolgi* (outgroup), X is *P simium*, Y is the PNG population of *P vivax*, and Z is each of 12 non-PNG populations of *P vivax*, represented as circles in the world map colored according to the country of origin. Circle sizes are directly proportional to the magnitude of the computed *f*4 statistics; greater circles indicate a closer relationship between Y and the *P simium* population from humans and monkeys (Ps, represented by squares), relative to PNG. Detailed results are shown in Supplementary Table 10.

South Pacific [48], but no study has compared its infectiousness to *Nyssorhinchus* and *Kerteszia* anophelines or investigated adaptative changes in key proteins, such as Pvs47, that may affect vector compatibility. Not surprisingly, we show that Pvs47 sequences of *P simium* cluster together with those of *P vivax* from the Pacific Coast of Colombia and Amazonian Brazil and are

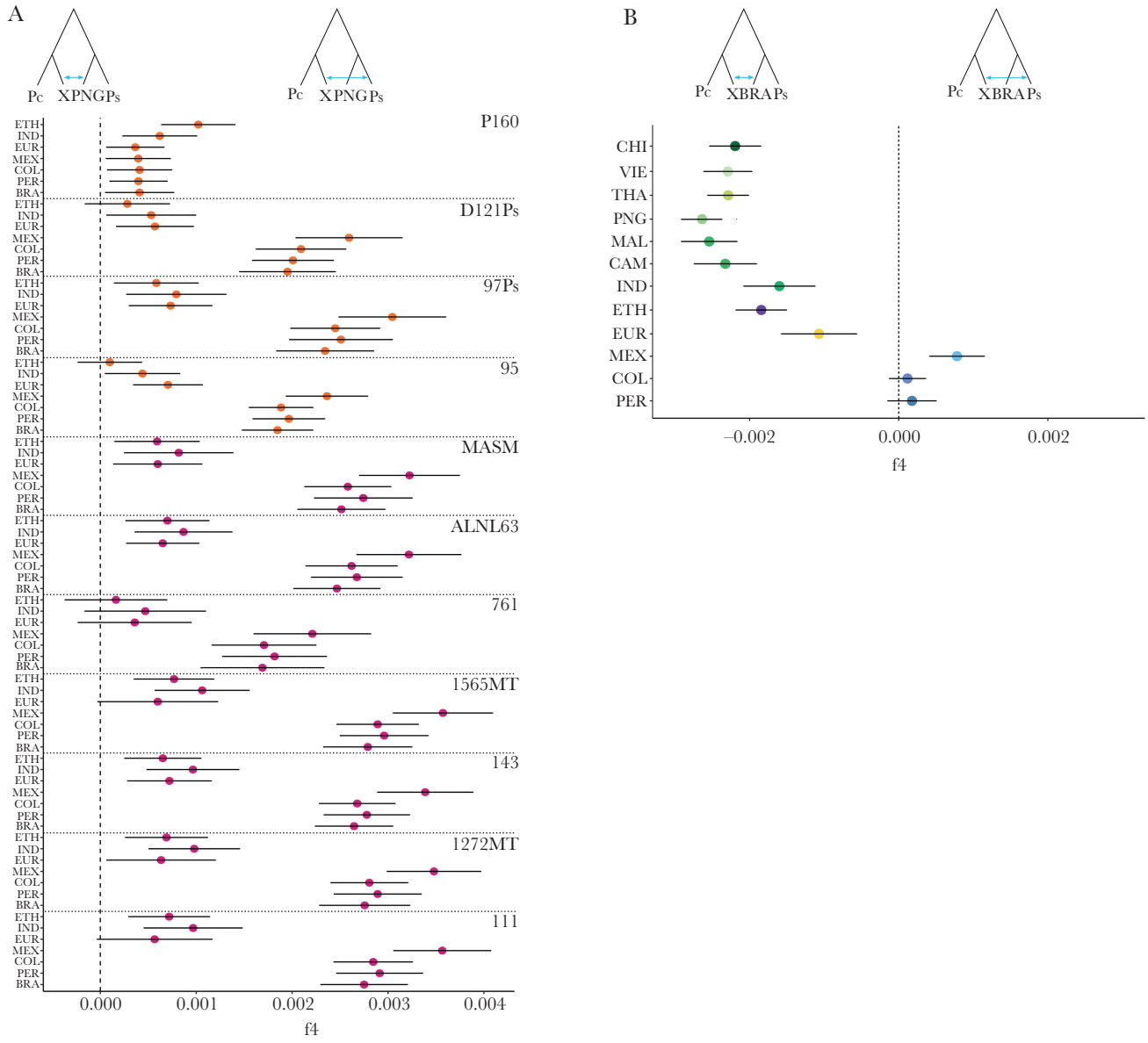

Figure 4. Genetic affinities of *Plasmodium simium* with New World isolates of *Plasmodium vivax*. (A) To explore the genetic affinities of individual isolates of *P simium*, we computed the *f*4 statistic for W corresponding to *Plasmodium cynomolgi* (Pc, outgroup), X to each of the 11 individual isolates of *P simium* (Ps), Y to the Papua New Guinean (PNG) population of *P vivax*, and Z to one of the American (Amazonian Brazil [BRA], Mexico [MEX], Peru [PER], and Colombia [COL]), Indian (IND), Ethiopian (ETH), or European (EUR) populations of *P vivax*, that are indicated on the y-axis legend. Sample codes are shown to the right. Note that P160 is the only *P simium* isolate that is genetically closer to non-American (ETH and to some extent IND) than American populations of *P vivax* (also see Supplementary Results). (B) To compare the genetic affinities of the *P simium* and BRA populations (from the Atlantic Coast and Amazon Basin of Brazil, respectively) with each of the non-BRA *P vivax* populations, we used the *f*4 statistic with W corresponding to *Plasmodium cynomolgi* (Pc), X to the *P simium* (Ps) population, Y to BRA, and Z to each of 12 non-BRA populations of *P vivax*, which are indicated on the y-axis legend.

clearly divergent from Old World sequences (Figure 5A), most likely as a result of selection imposed by adaptation to local vectors [34]. In addition, we found 2 amino acid changes present in *P simium* Pvs47 but absent in global *P vivax* populations (Figure 5B and Supplementary Table 11) that merit further investigation.

The red blood cell ligands PvDBP1 and PvRBP2a display marked divergence between *P vivax* and *P simium*. We identified the previously described 38-amino acid deletion in PvDBP1 in 7 of 16 *P simium* samples from humans and 5 of 9 samples from monkeys (Figure 6A; Supplementary Figures 7 and 8). However, the deletion is unlikely to modify the erythrocyte binding affinity of *P simium* PvDBP1, because its cysteine-rich 170-amino acid binding motif remains intact. Worldwide *P vivax* isolates present a much smaller deletion (9 amino acids) in the same PvDBP1 domain, relative to the reference PvP01 sequence. All *P simium* samples analyzed

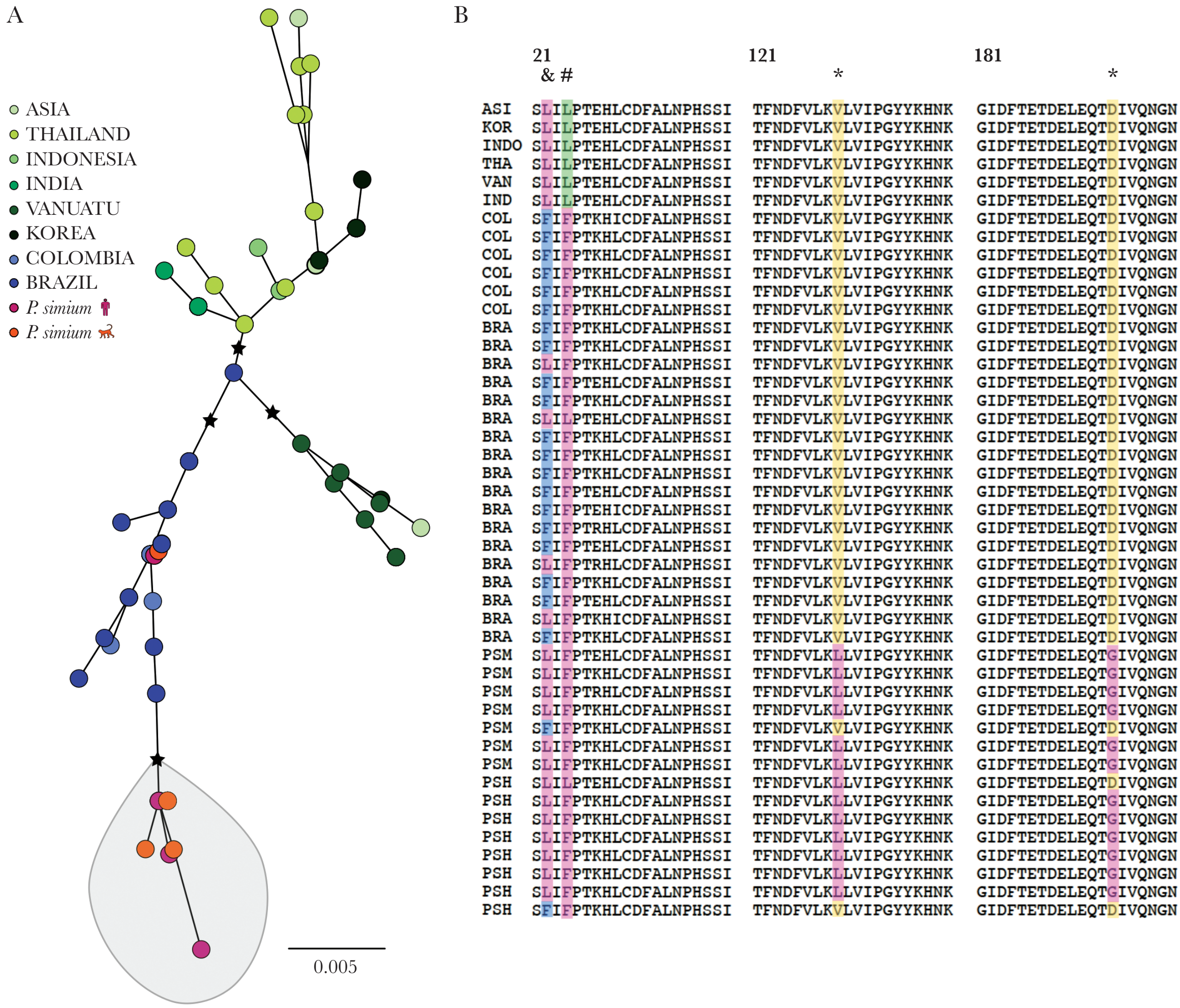

Figure 5. Global diversity in the *Plasmodium vivax* gamete surface protein Pvs47, a key determinant of parasite-vector compatibility. (A) Maximum likelihood tree of Pvs47 protein sequences from *P vivax* and *Plasmodium simium* generated with the best-fit protein evolution model (FLU + I). Black stars indicate basal nodes with bootstrap support >85% (1000 pseudoreplicates). Note that *P simium* clusters with New World *P vivax* isolates. (B) Amino acid alignment of the variable domain of Pvs47 (residues 21–200); the full-length protein has 434 amino acids. * indicates amino acid changes private of *P simium*; **#** indicates an amino acid change shared between *P simium* and New World *P vivax* lineages; and **&** indicates an amino acid change private of New World *P vivax* lineages. A single Pvs47 sequence represents all Asian lineages of *P vivax* because they are >99% identical [34]. Sample origins are abbreviated as follows: ASI, Asia (country not specified); BRA, Amazonian Brazil; COL, Colombia; IND, India; INDO, Indonesia; KOR, South Korea; PSH, *P simium* of human origin; PSM, *P simium* of simian origin; THA, Thailand; and VAN, Vanuatu. Sample origins are given in Supplementary Table 3 and Pvs47 polymorphisms are listed in Supplementary Table 11.

present the same private E→K amino acid substitution (Supplementary Figure 8).

More strikingly, all *P simium* samples from humans (n = 16) and monkeys (n = 7) display the same 1004-amino acid deletion in exon 2 of PvRBP2a [17] (Figure 6B and Supplementary Figures 6 and 7). This deletion preserves the reading frame but removes 44% of the coding sequence—therefore, it is very likely to destabilize the protein structure [43] and affect erythrocyte binding affinity. No deletion occurs in *P vivax* PvRBP2a (Supplementary Figures 8 and 9). Eight private amino acid changes and 1 private synonymous nucleotide replacement are present throughout *P simium* PvRBP2a (Figure 6B and Supplementary Table 11), which ranks among the most diverse coding sequences of this parasite (Supplementary Table 4; Supplementary Figure 4).

DISCUSSION

We have characterized, at the genome level, New World malaria parasites that infect humans and platyrrhine monkeys and sustain zoonotic malaria transmission on the Atlantic Coast of Brazil [20, 21]. The paucity of genomic diversity in *P simium* is consistent with its recent jump from humans to howler

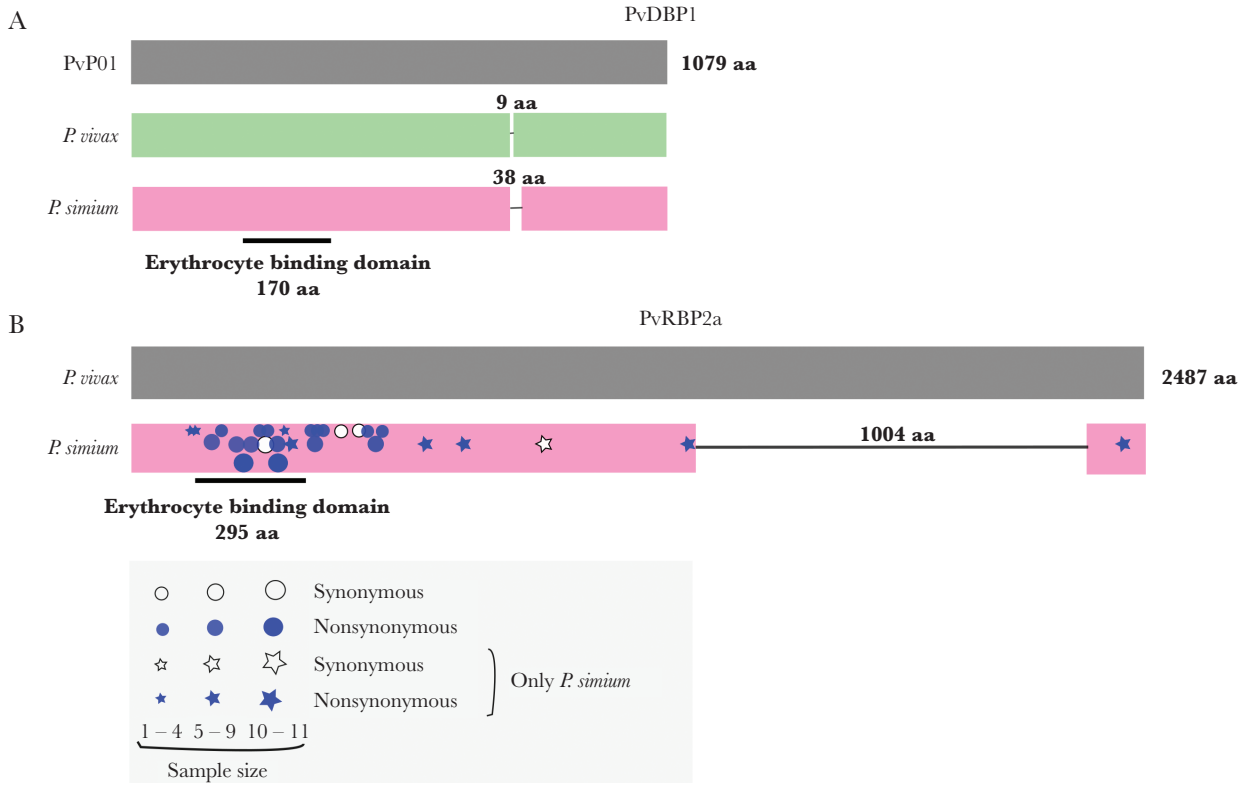

Figure 6. Deletions in the *Plasmodium simium* PvDBP1 and PvRBP2a ligands of red blood cells. (A) Schematic representation of PvDBP1 showing the 9-amino acid (aa) deletion found in the vast majority of worldwide *Plasmodium vivax* isolates, relative to the PvP01 reference sequence from Papua New Guinea, and the 38-aa deletion that may occur in *P simium* isolates, but not in *P vivax*. The location of the 170-aa erythrocyte binding motif of PvDBP is indicated. (B) Schematic representation of PvRBP2a showing the 1004-aa deletion found in all *P simium* isolates but not in worldwide *P vivax* populations. The location of the 295-aa erythrocyte binding motif of PvRBP2a is indicated [43]. The approximate location of synonymous and nonsynonymous nucleotide replacements at the *pvrbp2a* locus observed in newly sequenced *P simium* genome sequences is indicated by white and blue symbols, respectively; symbol sizes are directly proportional to the number of sequences showing the mutation. Stars and circles indicate nucleotide replacements found only in *P simium* and those also found in *P vivax* populations, respectively. See Supplementary Figure 8 and Supplementary Table 11 for further details.

monkeys and a few other Neotropical nonhuman primates [14, 15]. Although natural and experimental human infections with nonhuman primate malaria parasites are well known [49], in this study we address the much less common case of a host switch from humans to sylvatic monkeys, which characterizes a reverse zoonosis. The new hosts constitute a vast infectious reservoir that remains unaddressed by malaria elimination efforts [22, 23, 26, 29].

The *P simium* isolates from platyrrhine monkeys and humans have close genome-wide affinity with *P vivax* from the New World. We hypothesize that *P vivax* stocks from southern Europe and Africa that arrived in the Atlantic Coast during the colonial era [18] evolved to infect *Kerteszia* anophelines and local nonhuman primate species, in addition to humans. Similar species of platyrrhine monkeys are rarely found to harbor *P simium* across the Amazon Basin, although these same hosts comprise a significant sylvatic reservoir of the zoonotic malariae-like parasite *Plasmodium brasilianum* in the region [8, 13, 15]. More importantly, the meagre genome-wide divergence between *P simium* and New World *P vivax*, compared with similar or even higher levels of intraspecific polymorphism in *P vivax* populations across the continent [34, 47] (Supplementary Figure 6), argues against the current placement of *P simium* and *P vivax* in separate species.

Signatures of adaptation to new vectors are expected to occur in the *P simium* genome. As in *P falciparum* [33], the substantial genetic differentiation at the *pvs47* locus between Old World and New World *P vivax* is thought to have allowed parasites to evade the Jun N-terminal kinase (JNK) immune response of Neotropical *Nyssorhynchus* vectors [34], which diverged from the lineage leading to Old World anophelines approximately 100 million years ago. Our analysis of Pvs47 sequences in *P vivax* and *P simium* suggests that local *Kerteszia* vectors may have similarly imposed a selective pressure on parasites that, once introduced in South America, remained on the Atlantic Coast, instead of spreading to the *Nyssorhynchus*-dominated interior of the continent.

The deletion of >40% of the coding sequence of PvRBP2a, combined with numerous private amino acid changes, is the most likely signature of adaptation to nonhuman primate hosts identified in the *P simium* genome. We hypothesize that PvRBP2a-mediated erythrocyte binding has been drastically reduced or abolished in *P simium*, favoring alternative ligands

for more efficient monkey infection. This is reminiscent of the loss of EBA165, the red blood cell ligand of *Laverania* parasites that recognizes ape-specific sialic acids on erythrocytes, allowing *P falciparum* to infect humans [50]. Whether *P simium* binds to and invades human and simian erythrocytes with similar efficiency is unknown. Typically, *P simium* parasitemias are low in natural human and simian infections, with mild or absent clinical manifestations [19, 21, 22, 28]. We note, however, that we obtained human *P simium* samples from febrile hosts with patent parasitemias, which may have biased our analysis towards parasites with increased virulence. How *P simium* PvRBP2a interacts with its (undetermined) receptor on human and simian red blood cells remains unexplored [36], and experimental approaches are severely limited by the lack of a continuous culture in vitro for this parasite.

CONCLUSIONS

We conclude that *P simium* parasites from humans and monkeys comprise New World-type *P vivax* lineages that have very recently adapted to Neotropical nonhuman primates in the Atlantic Forest ecosystem. Likely genomic signatures of adaptation to these new vertebrate hosts include a massive deletion and extensive amino acid polymorphism in a key erythrocyte ligand, which may have favored more efficient simian host cell infection. Addressing the zoonotic reservoir is critical for preventing human malaria in the outskirts of major urban hubs, such as São Paulo and Rio de Janeiro, along the Atlantic Coast of Brazil [21, 22, 27].

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Acknowledgments. We are grateful to Maria José Menezes and Rebecca Kuzma for expert laboratory support, Rodrigo M. Corder for help in data analysis, Igor C. Johansen for preparing maps, and Seth Redmond for the useful insights.

Financial support. This work was funded by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brazil (2016/18740-9 [to M. U. F.], 2014/10919-4 [to A. M. R. C. D.], and 2009/51466-4 [to J. L. C.-D.]), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil (301011/2019-2; to M. U. F.), and the Wellcome Trust, United Kingdom (207486/Z/17/Z; to L. D. P. R.). This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Grant Number U19AI110818 to the Broad Institute (to D. E. N.). T. C. d. O., L. C.-D., and M. U. F. acknowledge scholarships from CNPq, Brazil. P. T. R. receives a postdoctoral scholarship from FAPESP. T. C. d. O. received a travel scholarship from the Fulbright Foreign Student Program for Brazil, USA.

Potential conflicts of interest. All authors: No reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest.

References

1. Ferreira MU, Castro MC. Malaria situation in Latin America and the Caribbean: residual and resurgent transmission and challenges for control and elimination. Methods Mol Biol **2019**; 2013:57–70.
2. Sinka ME, Rubio-Palis Y, Manguin S, et al. The dominant *Anopheles* vectors of human malaria in the Americas: occurrence data, distribution maps and bionomic précis. Parasit Vectors **2010**; 3:72.
3. Marrelli MT, Malafronte RS, Sallum MA, Natal D. *Kerteszia* subgenus of *Anopheles* associated with the Brazilian Atlantic rainforest: current knowledge and future challenges. Malar J **2007**; 6:127.
4. Harcourt AH. Human phylogeography and diversity. Proc Natl Acad Sci U S A **2016**; 113:8072–8.
5. Prugnolle F, Rougeron V, Becquart P, et al. Diversity, host switching and evolution of *Plasmodium vivax* infecting African great apes. Proc Natl Acad Sci U S A **2013**; 110:8123–8.
6. Liu W, Li Y, Shaw KS, et al. African origin of the malaria parasite *Plasmodium vivax*. Nat Commun **2014**; 5:3346.
7. Loy DE, Plenderleith LJ, Sundararaman SA, et al. Evolutionary history of human *Plasmodium vivax* revealed by genome-wide analyses of related ape parasites. Proc Natl Acad Sci U S A **2018**; 115:E8450–9.
8. Deane LM. Simian malaria in Brazil. Mem Inst Oswaldo Cruz **1992**; 87:1–20.
9. Schrago CG, Russo CA. Timing the origin of New World monkeys. Mol Biol Evol **2003**; 20:1620–5.
10. da Fonseca F. *Plasmodium* of a primate of Brazil. Mem Inst Oswaldo Cruz **1951**; 49:543–53.
11. Bueno MG. [doctoral dissertation]. São Paulo: Faculty of Veterinary Medicine of the University of São Paulo; **2012**.
12. de Alvarenga DA, de Pina-Costa A, de Sousa TN, et al. Simian malaria in the Brazilian Atlantic forest: first description of natural infection of capuchin monkeys (Cebinae subfamily) by *Plasmodium simium*. Malar J **2015**; 14:81.
13. Araújo MS, Messias MR, Figueiró MR, et al. Natural *Plasmodium* infection in monkeys in the state of Rondônia (Brazilian Western Amazon). Malar J **2013**; 12:180.
14. Lim CS, Tazi L, Ayala FJ. *Plasmodium vivax*: recent world expansion and genetic identity to *Plasmodium simium*. Proc Natl Acad Sci U S A **2005**; 102:15523–8.

15. Tazi L, Ayala FJ. Unresolved direction of host transfer of *Plasmodium vivax* v. *P. simium* and *P. malariae v. P. brasilianum*. Infect Genet Evol **2011**; 11:209–21.
16. Rodrigues PT, Valdivia HO, de Oliveira TC, et al. Human migration and the spread of malaria parasites to the New World. Sci Rep **2018**; 8:1993.
17. Mourier T, de Alvarenga DAM, Kaushik A, et al. The genome of the zoonotic malaria parasite *Plasmodium simium* reveals adaptions to host-switching [preprint]. BioRxiv 841171.
18. Deane LM, Ferreira Neto JA, Lima MM. The vertical dispersion of *Anopheles* (*Kerteszia*) *cruzi* in a forest in southern Brazil suggests that human cases of malaria of simian origin might be expected. Mem Inst Oswaldo Cruz **1984**; 79:461–3.
19. Deane LM, Deane MP, Ferreira Neto J. Studies on transmission of simian malaria and on a natural infection of man with *Plasmodium simium* in Brazil. Bull World Health Organ **1966**; 35:805–8.
20. Buery JC, Rodrigues PT, Natal L, et al. 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 **2017**; 16:437.
21. Brasil P, Zalis MG, de Pina-Costa A, et al. Outbreak of human malaria caused by *Plasmodium simium* in the Atlantic Forest in Rio de Janeiro: a molecular epidemiological investigation. Lancet Glob Health **2017**; 5:e1038–46.
22. Abreu FVS, Santos ED, Mello ARL, et al. Howler monkeys are the reservoir of malarial parasites causing zoonotic infections in the Atlantic forest of Rio de Janeiro. PLoS Negl Trop Dis **2019**; 13:e0007906.
23. Multini LC, Marrelli MT, Beier JC, Wilke ABB. Increasing complexity threatens the elimination of extra-Amazonian malaria in Brazil. Trends Parasitol **2019**; 35:383–7.
24. Li J, Collins WE, Wirtz RA, Rathore D, Lal A, McCutchan TF. Geographic subdivision of the range of the malaria parasite *Plasmodium vivax*. Emerg Infect Dis **2001**; 7:35–42.
25. Cormier LA. The historical ecology of human and wild primate malarias in the New World. Diversity **2010**; 2:256–80.
26. Duarte AM, Malafronte Rdos S, Cerutti C Jr, et al. Natural *Plasmodium* infections in Brazilian wild monkeys: reservoirs for human infections? Acta Trop **2008**; 107:179–85.
27. Yamasaki T, Duarte AM, Curado I, et al. Detection of etiological agents of malaria in howler monkeys from Atlantic Forests, rescued in regions of São Paulo city, Brazil. J Med Primatol **2011**; 40:392–400.
28. Cerutti C Jr, Boulos M, Coutinho AF, et al. Epidemiologic aspects of the malaria transmission cycle in an area of very low incidence in Brazil. Malar J **2007**; 6:33.
29. Ferreira MU, Castro MC. Challenges for malaria elimination in Brazil. Malar J **2016**; 15:284.
30. Cowell AN, Loy DE, Sundararaman SA, et al. Selective whole-genome amplification is a robust method that enables scalable whole-genome sequencing of *Plasmodium vivax* from unprocessed clinical samples. mBio **2017**;8:e02257-16.
31. Auburn S, Böhme U, Steinbiss S, et al. A new *Plasmodium vivax* reference sequence with improved assembly of the subtelomeres reveals an abundance of *pir* genes. Wellcome Open Res **2016**; 1:4.
32. Pearson RD, Amato R, Auburn S, et al. Genomic analysis of local variation and recent evolution in *Plasmodium vivax*. Nat Genet **2016**; 48:959–64.
33. Molina-Cruz A, Canepa GE, Alves E Silva TL, et al. *Plasmodium falciparum* evades immunity of anopheline mosquitoes by interacting with a Pfs47 midgut receptor. Proc Natl Acad Sci U S A **2020**; 117:2597–605.
34. Hupalo DN, Luo Z, Melnikov A, et al. Population genomics studies identify signatures of global dispersal and drug resistance in *Plasmodium vivax*. Nat Genet **2016**; 48:953–8.
35. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol **2015**; 32:268–74.
36. Kanjee U, Rangel GW, Clark MA, Duraisingh MT. Molecular and cellular interactions defining the tropism of *Plasmodium vivax* for reticulocytes. Curr Opin Microbiol **2018**; 46:109–15.
37. Haubold B, Travisano M, Rainey PB, Hudson RR. Detecting linkage disequilibrium in bacterial populations. Genetics **1998**; 150:1341–8.
38. Schaffner SF, Taylor AR, Wong W, Wirth DF, Neafsey DE. hmmIBD: software to infer pairwise identity by descent between haploid genotypes. Malar J **2018**; 17:196.
39. Petr M, Vernot B, Kelso J. admixr-R package for reproducible analyses using ADMIXTOOLS. Bioinformatics **2019**; 35:3194–5.
40. Patterson N, Moorjani P, Luo Y, et al. Ancient admixture in human history. Genetics **2012**; 192:1065–93.
41. Tachibana S, Sullivan SA, Kawai S, et al. *Plasmodium cynomolgi* genome sequences provide insight into *Plasmodium vivax* and the monkey malaria clade. Nat Genet **2012**; 44:1051–5.
42. Warncke JD, Vakonakis I, Beck HP. *Plasmodium* helical interspersed subtelomeric (PHIST) proteins, at the center of host cell remodeling. Microbiol Mol Biol Rev **2016**; 80:905–27.
43. Gruszczyk J, Lim NT, Arnott A, et al. Structurally conserved erythrocyte-binding domain in *Plasmodium* provides a versatile scaffold for alternate receptor engagement. Proc Natl Acad Sci U S A **2016**; 113:E191–200.
44. Knuepfer E, Wright KE, Kumar Prajapati S, et al. Divergent roles for the RH5 complex components, CyRPA and RIPR

in human-infective malaria parasites. PLoS Pathog **2019**; 15:e1007809.
45. Culleton R, Carter R. African *Plasmodium vivax*: distribution and origins. Int J Parasitol **2012**; 42:1091–7.
46. van Dorp L, Gelabert P, Rieux A, et al. *Plasmodium vivax* malaria viewed through the lens of an eradicated European strain. Mol Biol Evol **2020**; 37:773–85.
47. de Oliveira TC, Rodrigues PT, Menezes MJ, et al. Genome-wide diversity and differentiation in New World populations of the human malaria parasite *Plasmodium vivax*. PLoS Negl Trop Dis **2017**; 11:e0005824.
48. Collins WE, Sullivan JS, Galland GG, et al. *Plasmodium simium* and *Saimiri boliviensis* as a model system for testing candidate vaccines against *Plasmodium vivax*. Am J Trop Med Hyg **2005**; 73:644–8.
49. Martinelli A, Culleton R. Non-human primate malaria parasites: out of the forest and into the laboratory. Parasitology **2018**; 145:41–54.
50. Proto WR, Siegel SV, Dankwa S, et al. Adaptation of *Plasmodium falciparum* to humans involved the loss of an ape-specific erythrocyte invasion ligand. Nat Commun **2019**; 10:4512.