Genomic insights into *Plasmodium vivax* and *Plasmodium simium* host shifts in Latin America

Margaux J. M. Lefebvre[1,2*], Fanny Degrugillier[2], Céline Arnathau[2], Camila González[3], Silvia Rondón[3], Andrés Link[4], A. Chaves[5,6], Julio A. Benavides[2,7], Aline Alves Scarpellini Campos[8], Edmilson dos Santos[8], Rosana Huff[8], Cláudia Maria Dornelles Silva[8], Ezequiel Vanderhoeven[9,10], Benoit De Thoisy[11], Michael C. Fontaine[2,12§*], Franck Prugnolle[13,14§*] & Virginie Rougeron[13,14§*]

Authors affiliations:
1: Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
2: MiVEGEC, Univ. Montpellier, CNRS, IRD, Montpellier, France.
3: Centro de Investigaciones en Microbiología y Parasitología Tropical (CIMPAT), Universidad de los Andes, Bogotá D. C., Colombia.
4: Laboratorio de Ecología de Bosques Tropicales y Primatología, Universidad de Los Andes, Bogota D.C., Colombia.
5: Centro Nacional de Innovaciones Biotecnológicas (CENIBiot), CeNAT-CONARE, San José, Costa Rica.
6: Escuela de Biología, Universidad de Costa Rica, San Jose, Costa Rica.
7: Doctorado en Medicina de la Conservación y Centro de Investigación para la Sustentabilidad, Facultad de Ciencias de la Vida, Universidad Andrés Bello, Santiago Chile.
8: Centro Estadual de Vigilância em Saúde, Secretaria de Saúde do Rio Grande do Sul, Porto Alegre, Brazil.
9: Asociación Civil Centro de Investigaciones del Bosque Atlántico, Puerto Iguazú, Misiones 3370, Argentina.
10: Instituto de Biología Subtropical (IBS), Universidad Nacional de Misiones, CONICET, Puerto Iguazú, Misiones, Argentina.
11: Institut Pasteur de la Guyane, Laboratoire des Interactions Virus Hôtes, 16 avenue Pasteur, 97300 Cayenne, Guyane, France
12: Groningen Institute for Evolutionary Life Sciences (GELIFES), University of Groningen, Groningen, The Netherlands.
13: REHABS, International Research Laboratory, CNRS-NMU-UCBL, George Campus, Nelson Mandela University, George, South Africa.
14: Sustainability Research Unit, George Campus, Nelson Mandela University, George campus, Madiba drive, 6529 George, South Africa

§ These authors contributed to and co-supervised equally this work and should be considered as co-last authors.

* Corresponding author(s):
margaux.jm.lefebvre@gmail.com (or margaux_lefebvre@eva.mpg.de)
michael.fontaine@cnrs.fr
franck.prugnolle@cnrs.fr
rougeron.virginie@gmail.com and virginie.rougeron@cnrs.fr

Abstract

Plasmodium vivax causes >70% of malaria cases in Latin America. The emergence of *Plasmodium simium*, transmitted from humans to monkeys, increased the malaria burden, notably after an outbreak in 2015 in Brazil. To investigate *P. simium* evolutionary history, we screened 646 monkey samples and analyzed whole genome variations for 19 *P. simium* and 408 *P. vivax* isolates. *P. vivax* strains were identified in Colombian (n=3) and Brazilian monkeys (n=1), suggesting host reinvasion in Colombia and genetic exchange between *P. vivax* and *P. simium* in Brazil. Genetic analysis indicates that *P. simium* originated from a host jump approximately a century ago, possibly linked to *P. vivax* migration from Mexico to Brazil. Genome-wide scans identified selection in genes interacting with primate hosts and vectors. These findings highlight *P. simium* evolutionary history and zoonotic malaria risks, and underscore the need to include monkeys in malaria prevention measures while ensuring human-wildlife coexistence.

Keywords: *Plasmodium simium*, *Plasmodium vivax,* population genomics, host jump, zoonotic malaria.

Introduction

Plasmodium species are responsible for malaria in a wide range of vertebrate hosts, including primates, rodents, reptiles and birds[1]. Within this genus, *Plasmodium falciparum* and *Plasmodium vivax* are the most clinically significant species in terms of malaria-related morbidity and mortality in humans[2]. Other species, such as *Plasmodium malariae*, *Plasmodium ovale curtisi*, *Plasmodium ovale wallikeri*, *Plasmodium knowlesi*, *Plasmodium cynomolgi* and *Plasmodium simium*, also can infect humans, contributing to the complexity of malaria epidemiology[3]. *P. knowlesi*, *P. cynomolgi*, and *P. simium* primarily infect non-human primates (NHPs), but also humans, contributing to emerging zoonotic malaria cases[4–6]. For example, *P. knowlesi* and *P. cynomolgi*, which naturally infect Asian NHPs, have emerged as a significant public health issue in South-East Asia[7–9]. *P. simium* has more recently been implicated in human malaria cases in Latin America (Brazil)[5]. Today, zoonotic malaria, caused by malaria parasites transmitted from animals to humans, is becoming problematic due to human encroachment on forest habitats[10,11]. This increases the risk of transmission from NHP reservoirs[10,11], potentially challenging malaria elimination programs and making zoonotic malaria a potential public health issue.

In South America, *P. simium* is a malaria parasite that primarily infect platyrrhine monkeys[12]. It was initially described in 1951 from blood smears of the southern brown howler monkey (*Alouatta clamitans*) near São Paulo, Brazil[13], and since then it has been detected in other monkey species, including the howler monkeys (*Alouatta fusca* and *Alouatta caraya*), the southern muriqui (*Brachyteles arachnoides*), the three-striped night monkey (*Aotus trivirgatus*), and capuchin monkeys[3,14]. Its zoonotic potential was first described in 1966 when human infections were identified in Brazil through blood smear analysis, raising concerns about cross-species transmission[15]. Historically, *P. simium* is believed to be endemic in Brazil[5,12,16]. The detection of infections in NHPs from Colombia in 2019[17] and Costa Rica in 2022[18] suggested a broader distribution, although it remains unclear whether these infections were caused by *P. simium* or *P. vivax*.

Genetically, *P. simium* is closely related to *P. vivax*, although they can be distinguished by two mitochondrial single nucleotide polymorphisms (SNPs) and morphological differences in erythrocytic stages[5,19]. In addition, no hybrid has been identified to date[20], supporting the classification of *P. simium* as a distinct species from *P. vivax*[13]. However, due to their genetic similarities, shared host range (humans), and potential for recombination, the distinction between *P. simium* and *P. vivax* may not be as clear-cut as currently understood and further investigation are required to clarify the taxonomic status of the two species.

Plasmodium simium evolutionary origin(s) and history was largely overlooked[3] until 2017, when Brasil *et al.*[5] characterized its mitochondrial genome following a human malaria epidemic in Brazil. They showed that this epidemic was caused by *P. simium* isolates. Then whole genome sequencing (WGS) studies confirmed that *P. simium* is genetically close to the American human *P. vivax* and characterized by a lower genetic diversity[20,21], suggesting a host jump to Neotropical NHPs[14,20,21] (*i.e.,* reverse zoonosis[21]). Intriguingly, some studies showed that Brazilian *P. simium* strains are more closely related to Mexican than to Brazilian *P. vivax*, without providing any explanation for such results[20,21]. Other studies, based on mitochondrial genomes and some nuclear markers, suggested the occurence of multiple host jumps[22,23], that involved not only human *P. vivax* populations from Latin America, but also potentially from Asia[24,25]. Additionally, although the genetic proximity of these species

suggests a recent host jump and thus a recent divergence[12,14,20,21], to our knowledge, no precise date for this event has been provided. In conclusion, *P. simium* exact evolutionary history and origin(s) remain unresolved.

The mechanisms by which *P. simium* has adapted to different hosts and vectors are also poorly understood. No study has explored the population genomics of the host jump from humans to NHPs. This is likely to have exerted strong selective pressures on the parasite, driving its adaptation to new hosts and resulting in distinct phenotypes and genotypes. Only some studies targeting a few candidate genes identified some genes that may have facilitated *P. simium* successful invasion of the Neotropical primate hosts[20,21]. For instance, deletions were found in the coding regions of *pvrbpa2* (reticulocyte binding protein 2a) and *pvdbp1* (Duffy binding protein 1), two genes involved in the invasion of reticulocytes in primates[20,21]. Other than adapting to primate hosts, the parasite can also adapt to the vector responsible of its transmission. In Latin America, *P. vivax* is mainly transmitted by *Anopheles darlingi*[26,27] and *Anopheles albimanus*[27,28]. Conversely, it is thought that *P. simium* is transmitted by anopheline mosquitoes of the *Kerteszia* subgenus, primarily *Anopheles cruzii* and *Anopheles bellator*[29,30], suggesting adaptations to these distinct mosquito vectors. For example, *Pvs47*, a gene involved in the evasion of the mosquito immune system, harbors two amino acid changes in *P. simium*, but absent in global *P. vivax* populations[21]. However, a comprehensive genomic understanding of these adaptations is still lacking.

Given the zoonotic potential of *P. simium*, understanding its geographic range, evolutionary origin(s), and adaptation mechanisms is fundamental for developing effective malaria control and elimination strategies, but also for the conservation of endangered NHPs affected by this pathogen. The aim of this study was to clarify *P. simium* geographic distribution and host range in Latin America and to investigate its evolutionary history and adaptations. To address these questions, we first screened 646 NHP samples from five Latin American countries (Colombia, Brazil, Argentina, Costa Rica and French Guiana) for the presence of *P. simium*. This extensive screening identified three Colombian NHPs infected with *P. vivax* and not *P. simium.* We then used WGS data of these samples and previously published *P. simium* (n=19) and *P. vivax* (n=405) datasets to assess genetic variation, population structure, and evidence of selection (Supplementary Table 1 and Fig. 1). We found that the simian *P. vivax* samples from Colombia were more closely related to the Asian and Oceanian human *P. vivax*. The detection of a simian *P. vivax* sample in Brazil, which displayed a hybrid genetic ancestry between *P. simium* and *P. vivax*, indicated ongoing genetic exchange between these species. These results suggest the existence of an ongoing reverse zoonosis (*i.e.*, *P. vivax* transfer from humans to Neotropical NHPs), highlighting the risks of novel zoonotic malaria outbreaks and potentially affecting the health of NHPs. We also demonstrated that *P. simium* likely originated from a transfer of human *P. vivax* to Brazilian NHPs ~100-200 years ago, coinciding with the migration of human *P. vivax* from Mexico to this region. Lastly, we detected signals of positive selection in genes involved in interactions with primate hosts and mosquito vectors. Our results underline the importance of considering NHP as potential reservoirs of human malaria. This is crucial to anticipate potential emerging outbreaks and more globally to address human malaria public health issues but also the conservation of NHPs, whose populations may be threatened by the disease transmission.

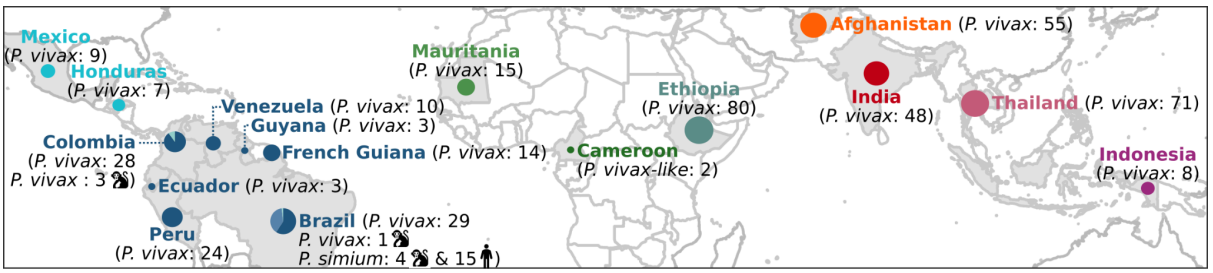

Figure 1. Geographic origins of the 408 *P. vivax* isolates, 19 *P. simium* isolates, and two African great apes *P. vivax-like* isolates. The number of samples for each species is indicated between brackets and by the circle size. The monkey pictogram indicates *P. vivax* and *P. simium* isolates from non-human primates in Latin America. For *P. simium* samples, the human pictogram indicates human samples.

Results

Evidence of *P. vivax* infections in South American non-human primates

Among 646 samples collected from NHPs across five Latin American countries (10 Colombian, 23 Brazilian, 30 Argentinian, 210 French Guianese, and 373 Costa Rican), 15 (10 from Colombia and 5 from Brazil) were positive for *Plasmodium* using the cytochrome-b based PCR assay[31]. These15 samples derived from various biological materials (feces, liver, mixed liver and kidney and blood) collected from four NHP species (*Alouatta seniculus*, *Ateles hybridus*, *Alouatta clamitans*, and *Cebus versicolor*) (see Supplementary Table 1). The sequence analysis revealed that they were all infected with *P. vivax* or *P. simium* and were all sent for WGS. Four samples, three from feces and one from blood, provided sufficient mitochondrial DNA (mtDNA) sequencing data to distinguish between *P. simium* and *P. vivax* (see Material and Methods), and were integrated with WGS data on 405 *P. vivax* and 29 *P. simium* genomes from previous studies to create a comprehensive dataset (Supplementary Table 1, Supplementary Fig. 1). Among the 33 *P. simium* or undetermined (*P. vivax* or *P. simium)* samples, 3 from Colombia (newly sequenced Col_164, Col_68 and Col_86) and one from Brazil (P160 from Ibrahim *et al.*[19]) displayed mtDNA SNPs specific to *P. vivax*[5,19], thus indicating that *P. vivax* circulates in South American NHPs.

Genetic relationships between the *P. vivax* from Neotropical non-human primates, *P. simium* and the worldwide human *P. vivax* populations

After filtering out multi-strain infections (i.e., multiple *P. vivax* or *P. simium* genotypes) and excessively related samples (for more details see Materials and Methods), the final dataset included: (i) 404 human *P. vivax* samples from 15 countries; (ii) four *P. vivax* from NHPs (two from Colombian *A. hybridus*, one from Colombian *A. seniculus* and one from Brazilian *Callicebus nigrifrons*), and (iii) 19 *P. simium* isolates (n=15 from *Homo sapiens* and n=4 from *Alouatta guariba*, all from Brazil). Two additional *P. vivax*-like genomes from

Cameroon chimpanzees (*Pan troglodytes ellioti*) served as outgroup (Fig. 1). From this dataset, we created two datasets with different filtering strategies to accommodate the requirements of different population genetic analyses (Supplementary Fig. 1). The first dataset (*dataset_GL*) was composed of 1,286,079 SNPs and all 429 genomes, with lenient filters and accounting for genotype uncertainty through genotype likelihood analysis using the ANGSD software ecosystem[32]. This was necessary to handle sequencing depth heterogeneity, especially for the *P. vivax* isolates from Colombian NHPs, which had low genome coverage (≥1X for 4.2% for the genome of Col_164, 11.5% for Col_68, and 8.9% for Col_86). The second dataset (*dataset_HF)* relied on more classic SNP calling using stringent hard filtering. It consisted of 819,652 SNPs for 426 samples, excluding the three Colombian *P. vivax* isolates from NHPs due to their low sequencing depth (median depth <1X) (Supplementary Fig. 1, and Material and Methods).

Using these datasets, we first explored the genetic relationships of the simian *P. vivax* and *P. simium* isolates within the global diversity of human *P. vivax* populations using multiple approaches. First, we performed a principal component analysis (PCA) and a model-based individual ancestry analysis using *dataset_GL* and *PCAngsd*[33] to assess the population structure. We also visualized genetic relationships among isolates using a maximum likelihood (ML) phylogenetic tree based on *dataset_HF* with IQ-TREE[34]. The results were congruent (Fig. 2) and revealed four distinct genetic clusters of human *P. vivax*, consistent with previous studies[35–37]: (1) Oceania, East and Southeast Asia, (2) Africa, (3) Middle East and South Asia, and (4) Latin America.

The four *P. vivax* isolates from NHPs did not cluster with *P. simium* isolates in the PCA (Fig. 2A). The three Colombian simian *P. vivax* isolates were closely related to South-East Asian and Oceanian *P. vivax* isolates, while the Brazilian isolate was more closely aligned with African *P. vivax*, particularly Ethiopian samples (Fig. 2A). The Colombian samples displayed admixed genetic ancestry (Fig. 2B), with contributions from South-East Asian and West African *P. vivax* populations. Conversely, the Brazilian sample showed shared genetic ancestry with *P. simium* and predominantly with Ethiopian *P. vivax* (Fig. 2B). This isolate clustered within the Ethiopian diversity in the ML tree, although the node support was not high (SH-aLRT = 95% and UFboot = 91%, using thresholds established by Guindon *et al.*[38] and Minh *et al.*[39]) (Fig. 2C).

In the PCA plot (Fig. 2A), *P. simium* strains, isolated from NHPs and humans, formed a distinct cluster, separated from human *P. vivax* along the first principal component (PC 1). *Plasmodium simium* isolates were genetically closer to American *P. vivax* populations than to other populations from the rest of the world. The genetic ancestry plots further supported this result. Indeed, *P. simium* clustered with the American *P. vivax* when testing a number of clusters (K) ranging from 2 to 4 (Supplementary Fig. 2). Only with K=5, *P. simium* started to differentiate (Supplementary Fig. 2), although it still shared some genetic ancestry with the Mexican *P. vivax* (dark blue in Fig. 2B). The ML tree supported this result: *P. simium* branched within the genetic diversity of Mexican *P. vivax*, despite limited node support (SH-aLRT = 100% and UFboot = 80%). However, the strong node support at the base of the *P. simium* cluster suggests that these isolates share a recent common ancestor with Mexican *P. vivax* (Fig. 2C).

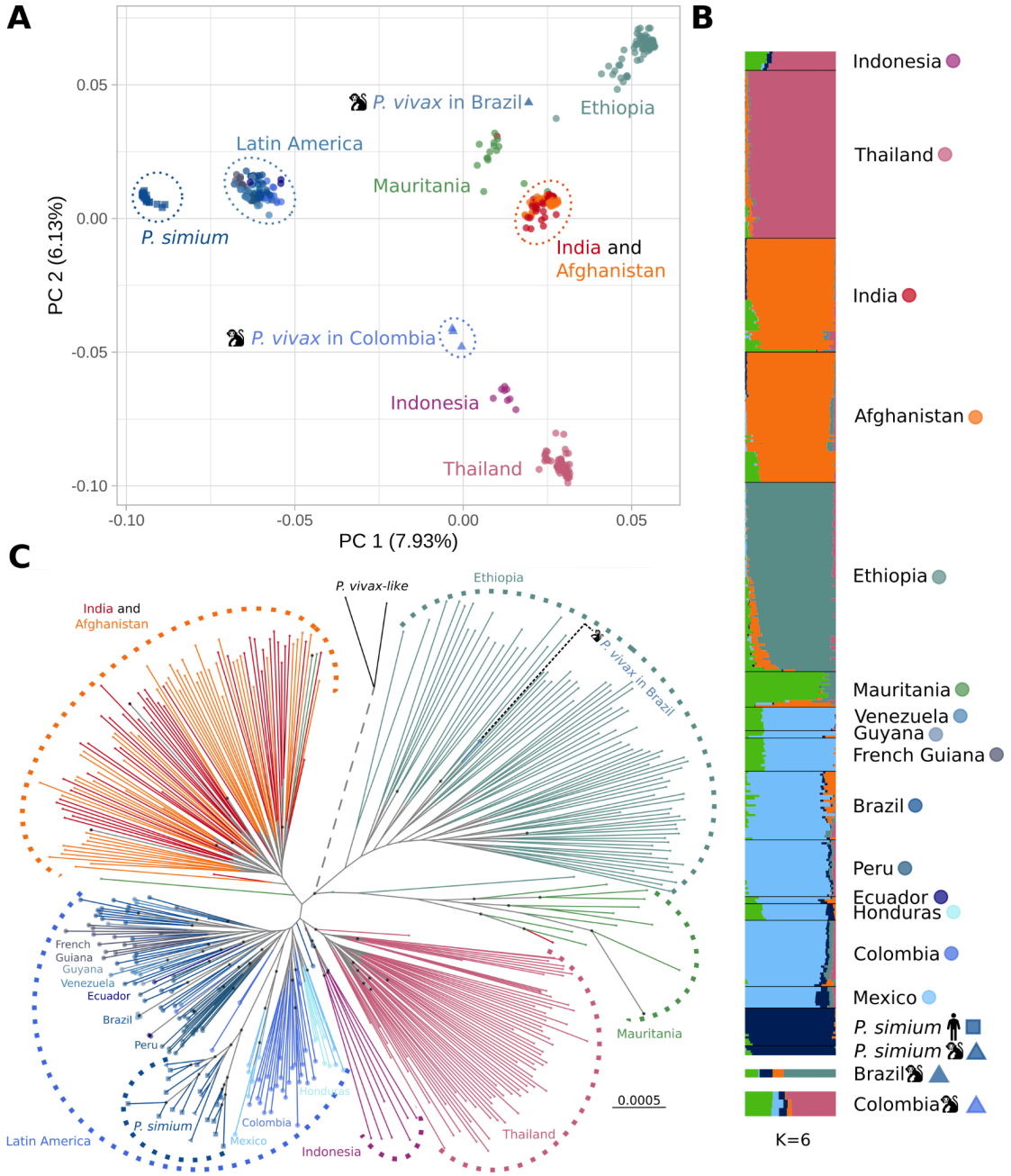

Figure 2. *P. vivax* and *P. simium* genetic structure. (A) Plot of the principal component (PC) analysis of 408 *P. vivax* and 19 *P. simium* strains showing the first and second PCs based on the genotype likelihood of 247,890 unlinked SNPs. **(B)** Individual genetic ancestry assuming K=6 genetic clusters estimated using *PCAngsd* (see Supplementary Fig. 2 for the other K values). The monkey pictogram indicates the *P. vivax* samples from South American non-human primates (NHPs). **(C)** Maximum likelihood phylogenetic tree of the 404 *P. vivax*, 19 *P. simium* and 2 *P. vivax-like* individuals. *P. simium* and *P. vivax* from NHPs are indicated by triangles. The *P. vivax* isolate from a Brazilian NHP is among the Ethiopian *P. vivax* isolates. The three *P. vivax* samples from Colombian NHPs could not be included in this analysis because of the high number of missing data. The tree includes two *P. vivax-like* strains used as outgroups. Note that the length of the outgroup branch (dotted lines) has been truncated. Black dots at nodes indicate highly supported nodes (SH-aLRT ≥ 80% and UFboot ≥ 95%, following the thresholds by Guindon *et al.*[38] and Minh *et al.*[39]).

Although *P. simium* formed a well-defined group distinct from the human *P. vivax* (Fig. 2), the strains originated from various regions in Brazil and from different host species (humans and *A. guariba*) (Fig. 3A). Therefore, we investigated whether there was a genetic sub-structuring within *P. simium*, independently of *P. vivax,* according to the host species and geography. The genetic ancestry analyses with *ADMIXTURE*[40] identified two genetic clusters within *P. simium*, without evidence of admixture between them (Fig. 3B, and Supplementary Fig. 3). This separation was also evident in the first principal component of the PCA (Fig. 3C). The Rio de Janeiro (RJ) cluster (n=13) comprised *P. simium* samples from humans in the states of Rio de Janeiro and Espírito Santo, as well as in the municipality of Peruíbe, in the State of São Paulo. Additionally, *P. simium* samples from *A. guariba* were included from Rio de Janeiro and São Paulo (Fig. 3B). Within this cluster, the three samples from São Paulo, all isolated from NHPs, were distinct from the other samples in the RJ cluster along the second principal component of the PCA (Fig. 3C). In contrast, the São Paulo (SP) cluster (n=6), located further south, included only human samples from the municipalities of Maresias, Riacho Grande, Juquitiba, Registro, and Iporanga, all within the state of São Paulo.

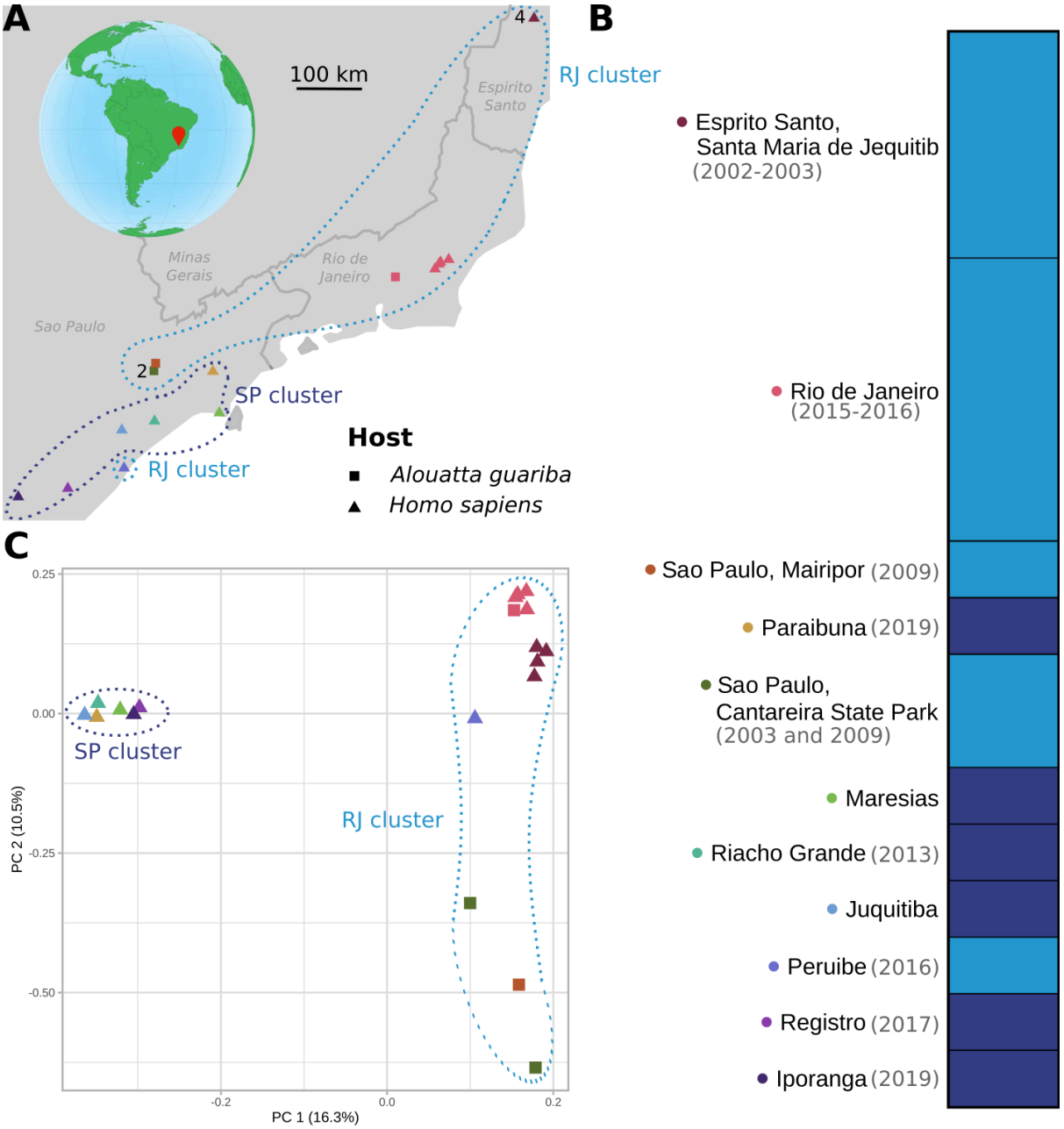

Figure 3. Geographic origins and genetic structure of the Brazilian *P. simium* isolates. (A) Geographic origins of the 19 *P. simium* isolates. The squares and triangles represent the host species (non-human primates and humans, respectively), and the colors correspond to the region of origin. The SP cluster is outlined in dark blue, the RJ cluster in light blue.The borders and names of the Brazilian states are indicated in dark gray. On the globe, the red pointer indicates the region of Brazil from which the samples originate. **(B)** Genetic ancestry estimated using *ADMIXTURE* for *K* = 2. Collection dates are indicated when known. **(C)** Genetic structure recovered from the first two PCs of the PCA. The shape (square and triangle) represents the host species (non-human primates and humans), and the color the region of origin. The SP cluster is outlined in dark blue, the RJ cluster in light blue.

Evidence of admixture between *P. simium* and *P. vivax*

We next investigated the role of admixture in shaping the genetic makeup of different populations and samples. We estimated and visualized the pairwise identity-by-descent (IBD) as a network to depict recent common ancestry among samples[41,42] (see the Materials and Methods for more details). We also constructed population graphs using *AdmixtureBayes*[43] and *TreeMix*[43,44], to obtain information on population branching order and potential admixture events (*i.e.* reticulation). Additionally, we used f_4*-statistics* to formally test for admixture[45,46]. All these analyses were based on the *dataset_HF* (Supplementary Fig. 1).

As observed in the population structure analyses (Fig. 2), the pairwise IBD network showed the same four main genetic clusters identified in *P. vivax*: Asia-Oceania, Middle East-South Asia, East Africa, and Latin America. In the IBD network, all *P. simium* isolates formed a single coherent and genetically distinct cluster, closely related to Latin American *P. vivax* (Fig. 4A). Conversely, the Brazilian simian *P. vivax* isolate was genetically distinct, showing strong connections with Ethiopian isolates, as observed in previous analyses (Fig. 2).

The *AdmixtureBayes* analysis identified a population graph with two admixture events as the best-fitting solution for our data (Supplementary Fig. 4). *P. simium* isolates formed a well-supported monophyletic group (posterior support >80%) that branched off from Mexican *P. vivax* (Fig. 4B). The first reticulation in the *AdmixtureBayes* population graph identified Thailand *P. vivax* as an admixed population between Indonesian populations and the common ancestor of the Indian and Afghan populations. This reticulation was not visible in the pairwise IBD network, which only showed recent ancestry, and in *TreeMix*, which identified only one optimal migration edge (Supplementary Fig. 5). The only reticulation detected by *TreeMix*, also present in the *AdmixtureBayes* population graph, suggested that the Brazilian simian *P. vivax* isolate resulted from an admixture event between East African *P. vivax* and *P. simium*, with ~40% of its ancestry from *P. simium* (Fig. 4B and Supplementary Fig. 5). These results were supported by the f_4-statistic tests (Fig. 4C), confirming that this simian *P. vivax* sample was genetically closer to Ethiopian *P. vivax* than to the *P. simium* cluster.

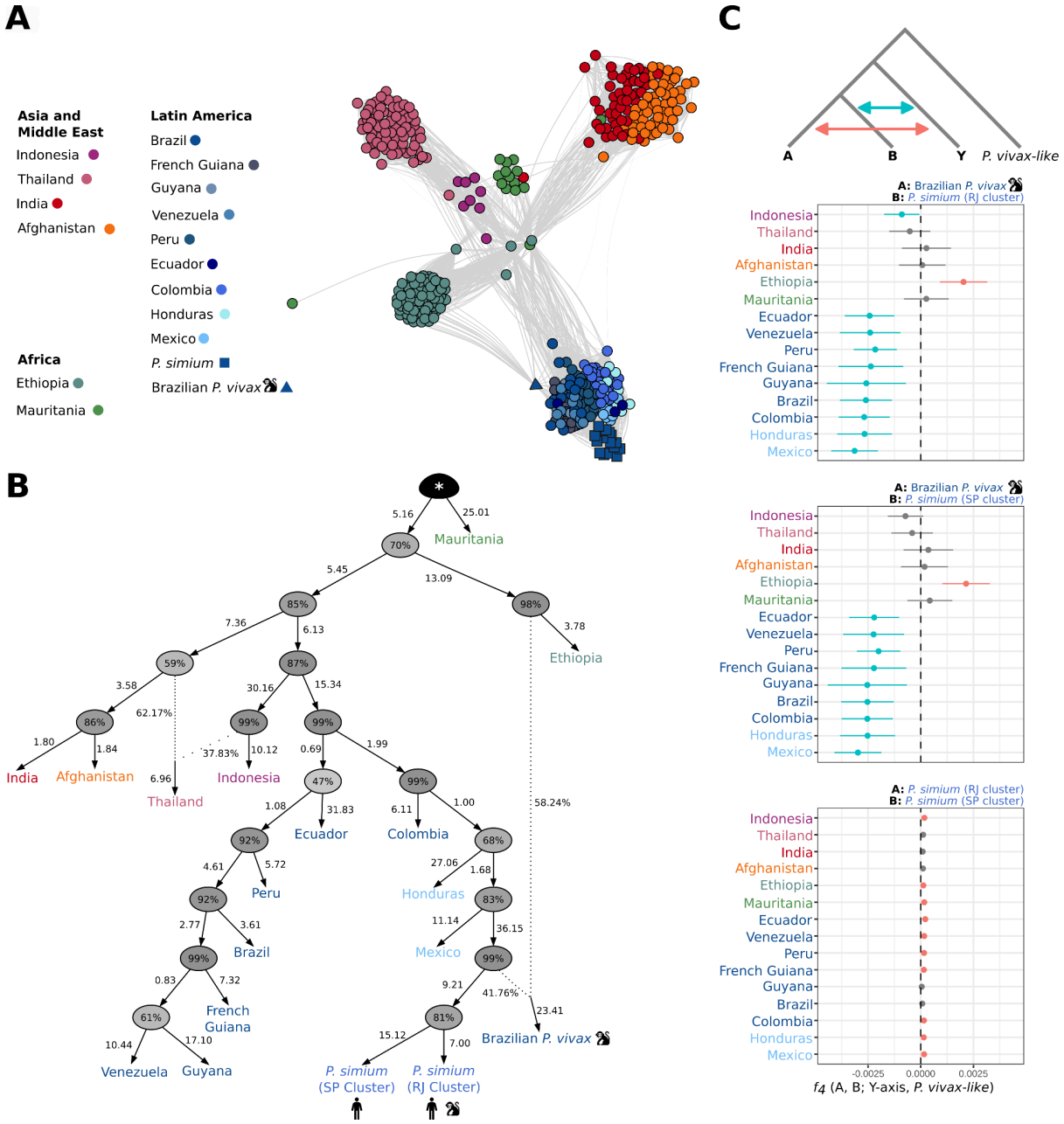

Figure 4. Population genetic relationships and admixture between *P. vivax* and *P. simium*. (A) Network visualization of the identity-by-descent (IBD) relationships among *P. vivax* and *P. simium* sample pairs. Edges represent the IBD values between each sample pair. Only edges with IBD values > 5% are displayed. **(B)** Admixture population graph topology of *P. vivax* and *P. simium* populations. The highest posterior probability was estimated with *AdmixtureBayes*[43], and rooted with the two *P. vivax-like* genomes (*). The branch length (measured in drift units) indicates the genetic divergence between populations, multiplied by 100. Percentages at nodes indicate the posterior probability that the true graph has a node with the same descendants. For each admixture event (indicated by the dotted lines), the percentages show the admixture proportions from each donor. **(C)** Admixture tests based on the f_4-*statistics* for *P. simium* clusters and the Brazilian *P. vivax* sample from a non-human primate with *P. vivax* populations. This statistic assesses the excess of shared alleles between populations. In the form f_4(A, B; Y-axis, P. vivax-like), values not significantly different from 0 (gray) indicate no gene flow between populations A

or B and the Y-axis population. Positive values (red) suggest gene flow between population A and the Y-axis population, and negative values (blue) indicate gene flow between population B and the Y-axis population. The monkey pictogram indicates the *P. vivax* isolate from Brazilian non-human primate.

Plasmodium simium demographic history

We futher investigated *P. simium* demographic history, potential founder effect and isolation dynamics. We first computed the genome-wide distribution of the nucleotide diversity (π) in windows along the core genome of populations with at least 10 samples. *P. simium* exhibited significantly lower π values than *P. vivax* (*p-value* <0.001, Wilcoxon signed-rank test with Bonferroni correction, Supplementary Fig. 6). Then, we explored the changes in effective population size (N_e) over time, estimated using coalescence rates with the *Relate* software[47]. In agreement with the π values (S6 Fig.), both *P. vivax* and *P. simium* populations worldwide exhibited a steady decline in N_e starting ~100,000 generations ago (~20,000 years ago, assuming a generation time of 5.5 generations/year and a mutation rate of 6.43 x 10^{-9} mutations/site/generation[35]), followed by a recent moderate expansion (Fig. 5A). Notably, the two *P. simium* clusters (RJ and SP) displayed lower N_e values than the *P. vivax* populations, although we did not detect any strong bottleneck signal.

To explore the isolation and migration history of *P. simium* from *P. vivax* in Latin America, we used Multiple Sequentially Markovian Coalescent-Isolation Migration program (MSMC-IM)[48] to estimate the isolation dynamics and the migration rate changes over time. Our results, based on the cross-coalescent rates (CCR) between *P. simium* and American *P. vivax* populations, suggested a relatively recent isolation of *P. simium* from *P. vivax*. Based on the point at which the CCR dropped below 50% as suggested by Wang *et al.*[48], we could estimate that in all pairwise comparisons between *P. simium* and American *P. vivax*, split times occurred between 10 and 30 years ago (Fig. 5). The only exception was between the *P. simium* clusters, where CCR values remained ~75%, indicating ongoing genetic exchange (Fig. 5B, bottom).

Migration rate estimations (Fig. 5B) suggested that the oldest migration events occurred 100-200 years ago between the Mexican and Brazilian *P. vivax* populations, and between Mexican *P. vivax* and *P. simium* from the RJ cluster. The estimated divergence times were ~30 years ago for both the Mexican-Brazilian *P. vivax* and Mexican *P. vivax*-*P. simium* RJ cluster. Similarly, the divergence time between the *P. simium* SP cluster and Mexican *P. vivax* was ~30 years ago, with minimal gene exchange over the entire period (Fig. 5B). Conversely, the *P. simium* SP cluster and Brazilian *P. vivax* populations have exchanged genetic material more recently (30-60 years ago), with divergence estimated at ~20 years ago. On the other hand, the RJ cluster showed limited recent exchanges with Brazilian *P. vivax*, and diverged ~10 years ago. Importantly, although the cumulative migration probability (*Mt*), which is analogous to the CCR[48,49], was close to zero for some comparisons between *P. vivax* and *P. simium*, none of them reached zero. This suggests that genetic exchanges may still be possible, even at a low rate (Fig. 5B).

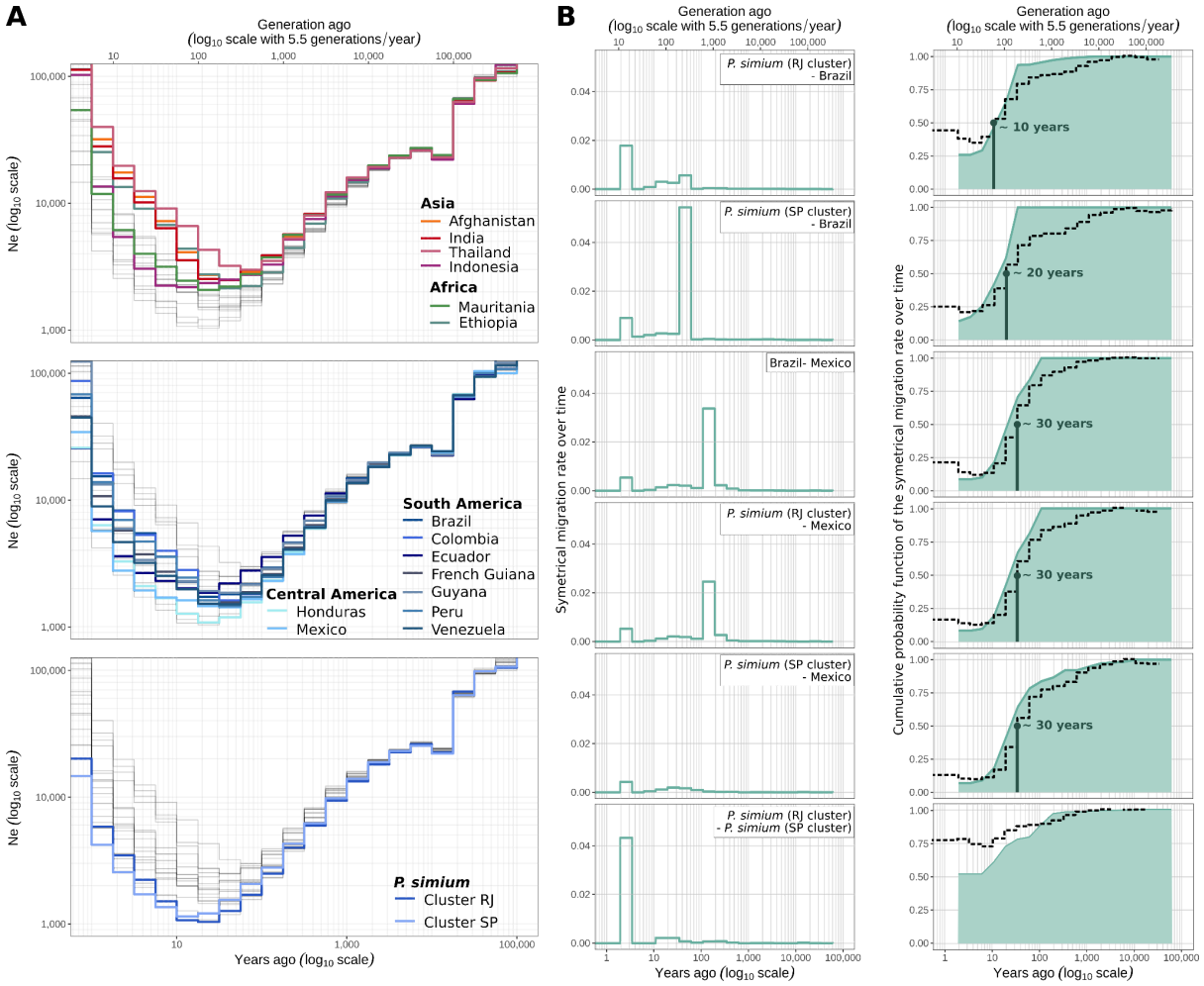

Figure 5. Coalescent-based inference of the demographic history and migration rates of *P. vivax* and *P. simium* populations. (A) The variation in effective population (*Ne*) size was estimated using *Relate* (axes are $\log_{10}$ transformed). **(B)** Timing and dynamics of the separation between the indicated groups estimated using *Relate* coalescence rates and *MSMC-IM*. On the left, the time-dependent-symmetric migration rate (x axes are $\log_{10}$ transformed). On the right, the cumulative distribution probabilities of the migration rates in function of time. This is an estimate of the proportion of ancestry already merged at time *t*, and represents the proportions of gene flow over time (x axes are $\log_{10}$ transformed). Values close to 0 indicate complete isolation between groups, while 1 shows a complete mixture as one population. Dashed lines indicate the relative cross-coalescence rate. The divergence time (when the cross-coalescent rate drops below 50%) is specified in each panel.

Genetic evidences of *P. simium* adaptation to new hosts

The host jump from humans to NHPs likely exerted strong selective pressure on *P. simium*, potentially leaving detectable signals of positive selection in its genome. Therefore, we first used haplotype-based tests (*iHS, XP-EHH,* and *Rsb* statistics[50–52]), that identify signals of positive selection specific to the target population by analyzing the haplotype lengths variations (*i.e.,* linkage disequilibrium, LD), focusing primarily on recent or ongoing selective events[50–52]. Given the limited sample size (n=19), we treated the two *P. simium* clusters as a single group and compared it to the Brazilian *P. vivax* population as the

reference population for *XP-EHH* and *Rsb* statistics. The datasets derived from *dataset_HF* included 10,304 SNPs for the 19 *P. simium* samples for the *iHS* test, and 378,441 SNPs for 48 samples (n=19 *P. simium* and n=29 *P.vivax* - Brazil) for the *XP-EHH* and *Rsb* analyses (see Material and Methods for more details). The *iHS* test identified a single significant SNP within the coding sequence (CDS) of the *PVP01_1139100* gene, an orthologue of a *Plasmodium* protein of unknown function (Supplementary Fig. 7). The *XP-EHH* and *Rsb* analyses identified two significant SNPs in the CDS of the *PTEX150* gene (Fig. 6A, Supplementary Fig. 7 and Supplementary Table 2), which encodes a protein involved in protein transport across the parasitophorous vacuolar membrane during the *Plasmodium* erythrocytic cycle[53,54].

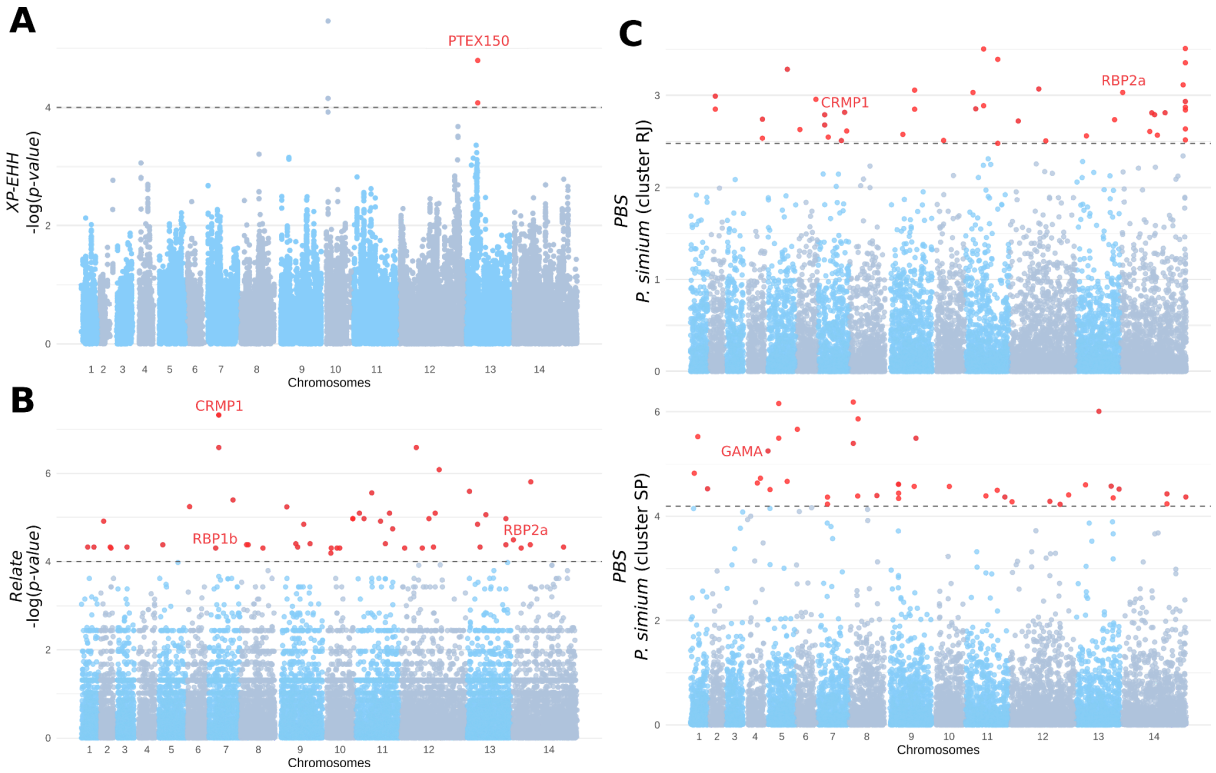

Figure 6. Evidence of selective sweeps in *P. simium*. (A) Manhattan plot showing the haplotype-based *XP-EHH* statistics between pairs of populations. The dotted line represents the significance threshold value at -log(*p-value*) = 4. Red points are the SNPs identified as affected by a selective sweep in *P. simium* (negative *XP-EHH* values). **(B)** Manhattan plot showing the *Relate* results. The dotted line represents the threshold significance value -log(*p-value*) = 4. The red points are the SNPs identified as potentially affected by positive selection in *P. simium*, on the basis of their higher-than-expected coalescence rate compared with the rest of the genome. **(C)** Manhattan plots showing the *PBS* values, an F_{ST}-type statistic that captures a difference in allele frequencies compared with two reference populations, for each *P. simium* cluster, with *P. vivax* Brazilian and Colombian samples as outgroup populations. For the *PBS* scores, all values in red represent the top 0.1% of the PBS values, suggesting higher allelic frequency differentiation compared with the reference populations, and thus indicating significant evidence of positive selection that affects these SNPs. CRMP1, cysteine repeat modular protein 1; RBP1b, reticulocyte binding protein 1b; RBP2a, reticulocyte binding protein 2a; GAMA, GPI-anchored micronemal antigen.

The haplotype-based tests may miss other genomic regions because recombination can break down LD patterns. Therefore, we used *Relate*[47] to detect positive selection by identifying increased coalescence rates in selected regions compared with the rest of the genome. By analyzing 26,565 SNPs in the 19 *P. simium* samples, we identified 49 SNPs with significant signals of selection in the CDS of 14 genes (Supplementary Table 2). Notably, three of these genes are involved in the interactions of the parasite with primate hosts (*PvRBP2a* and *PvRBP1b*[55–57]) and mosquitoes (*PvCMRP1*[58–60]).

We also investigated cluster-specific adaptations in *P. simium* by calculating the population branch statistic (*PBS*), an F_{ST}-like statistic, by comparing each *P. simium* cluster with two American *P. vivax* populations (Colombia and Brazil). Outlier values (top 0.1%) indicated significant differentiation of the *P. simium* clusters compared with American *P. vivax*. We identified 43 outlier windows in the RJ cluster and 41 in the SP cluster, corresponding to positive selection in the CDS of 32 genes (RJ cluster) and 31 genes (SP cluster) (Supplementary Table 2). Four genes showed signs of positive selection in both clusters (*PVP01_0525300*, *PVP01_0931400*, *SMC1* and *PVP01_1469500*). *SMC1* (Structural Maintenance of Chromosomes protein 1) is involved in the structural maintenance of chromosomes in eukaryotes[61,62]. Conversely, the functions of other three genes are not known. In the RJ cluster, two genes with evidence of selection have already been detected with *Relate*: *PvRBP2a* and *PvCMRP1* that play a direct role in the interactions with the primate hosts and with mosquitoes, respectively. The SP cluster showed selection signals in *PvGAMA*, a gene involved in binding to human erythrocytes[63].

Discussion

Historically, natural *P. vivax* or *P. simium* infections in NHPs (*A. guariba* and *A. caraya*) in Brazil were detected with antigenic tests[64] and/or PCR assay[16], although the two species were often indistinguishable from one another. *Plasmodium simium* was long thought to circulate exclusively in Brazil[5,64,65]. Recently, this assumption was challenged by studies that reported natural infections in Colombian and Costa Rican NHPs, attributed to *P. vivax* or *P. simium*[17,18]. This questioned the natural range of *P. simium* in Neotropical NHPs and its geographic distribution. Here, by screening 646 NHPs samples from five different Latin American countries, we confirmed that *P. simium* seemed indeed restricted to Brazil. For the first time, we identified natural infections by *P. vivax* in two American NHP species (*A. hybridus* and *A. seniculus*). The analysis of the published *P. simium* genomes allowed the identification of another *P. vivax* isolate in a NHP from Brazil (sample “P160”). These findings suggest ongoing host switches from humans to American NHPs, highlighting the reverse zoonotic potential of these parasites as already suggested[21].

The three Colombian *P. vivax* isolates from NHPs were genetically closer to South-East Asian and Oceanian *P. vivax* than to the American strains (Fig. 2). This pattern appears to be specific to simian *P. vivax* and is not observed in American isolates sampled from humans[35–37,66]. This genetic relationship may reflect historical human migrations from Asia or Oceania to Latin America[67], but it could be also linked to international animal trade between these regions[68], including the introduction of Asian NHPs transported to Brazil for experimental purposes[69]. These findings suggest a potential re-invasion of Asian human *P. vivax* into American NHPs. More simian *P. vivax* sampling in Colombia with greater sequencing depth is needed to clarify this evolutionary history.

The analysis of the *P. vivax* sample P160 from a Brazilian NHP revealed traces of admixture between *P. simium* and Ethiopian *P. vivax* (Fig. 2B and Fig. 3), consistent with prior observations[21]. The recent admixture detected in this sample demonstrates ongoing recombination between these species. The genetic proximity or even lack of distinct polymorphisms on some markers between *P. simium* and *P. vivax* had already raised questions about their classification as separate species[21,70]. This is not surprising when considering that the host jump occurred only ~80-100 years ago, and that we detected ongoing gene flow between these species (Fig. 5B). Interestingly, the P160 sample showed a mixture with Ethiopian rather than American *P. vivax*, which is unexpected given its collection area. The lack of literature explaining this Ethiopian ancestry underscores the need of more sampling to better understand the history of this sample.

The discovery of natural *P. vivax* infections in Colombian NHPs suggests that NHPs could act as anthropo-zoonotic reservoirs, thus complicating malaria elimination efforts across the continent, but also affect the survival of NHPs infected. *Plasmodium vivax* remarkable adaptability in changing hosts contrasts with *Plasmodium* species of the subgenus *Laverania* (including *P. falciparum*, the deadliest malaria agent) where host transfer is rarely observed[6]. Ongoing genetic exchanges between *P. vivax* and *P. simium* indicate porous species boundaries and suggest that recombination still occurs. These results are of major importance for malaria epidemiology and underscore the role of NHPs as potential malaria reservoirs. Therefore, it is crucial to consider NHP hosts in strategies to control and potentially eliminate malaria in Latin America. These strategies must also consider the health and conservation of already threatened NHPs[71], which have previously been killed by humans to prevent yellow fever transmission in Brazil[72]. Furthermore, the killing of NHPs may have negative impacts on public human health in both the short and long term[72].

The evolutionary processes underlying *P. simium* host shift and adaptation have not been fully explored so far (but see de Oliveira *et al.*[21]). Our PCA, ancestry plots, ML phylogenetic tree (Fig. 2), IBD networks, and population graphs (Fig. 4A-B) consistently showed a genetic structure among *P. vivax* populations that is in line with previous findings[35–37]. *P. simium* formed a unique cluster closely related to American *P. vivax*, particularly Mexican strains, as supported by its branching within the Mexican diversity in the ML tree and population graph (Fig. 2C, Fig. 4B, Supplementary Fig. 5), as previously reported[20,21]. Our results, supported by the ancestry plot (Fig. 2B), the population graph, and the f_4-statistics (Fig. 4), showed no genetic evidence of any Asian contribution to *P. simium* genetic ancestry, unlike some early studies based on mitochondrial DNA[24,25]. This supports an American origin for *P. simium*, closely linked to Mexican *P. vivax*.

A closer look at the genetic structure of *P. simium* isolates revealed two distinct genetic clusters (Fig. 3): RJ and SP. The SP cluster contains only human *P. simium* samples, whereas the RJ cluster included *A. guariba* and human samples. This is the first time this structure has been observed in *P. simium*, possibly thanks to the integration of different datasets in this study. However, the structure appears more complex. Indeed, the PCA results indicate that samples from São Paulo city are genetically distinct within the RJ cluster, while the Peruíbe sample is geographically nearer to the SP cluster, but is included in the RJ cluster. More genomic data across regions and host species are needed to clarify these patterns. For example, since the states of São Paulo and Rio de Janeiro are highly connected by public transports, the movement of infected humans can not be excluded to explain the observed pattern.

The close genetic similarity between human *P. vivax* and NHP *P. simium* suggests recent host transfer(s), as previously proposed[14,20,21], although no precise split times was estimated. Our coalescent-based analyses estimate that *P. simium* diverged from American *P. vivax* in the last 30 years (Fig. 5B). Our analyses also suggest that significant migration occurred ~100-200 years ago between Mexican and Brazilian *P. vivax* and between the *P. simium* RJ cluster and Mexican *P. vivax* (Fig. 5B). Therefore, a first host jump might have taken place ~100-200 years ago when Mexican *P. vivax* migrated to Brazil. This timing coincides with the migration of several thousand confederates from the Southern United States to Latin America, including Mexico and Brazil, at the end of the American Civil War[73,74]. It is plausible that *P. vivax* strains from the Southern United States, potentially genetically close to Mexican strains due to their geographical proximity, contributed to the origin of the RJ cluster in *P. simium*. To confirm this hypothesis Southern United States samples should be sequenced to assess their genetic connection to Mexican strains and their potential role in the emergence of *P. simium* in Brazil.

The *P. simium* SP cluster did not show any evidence of genetic exchanges with Mexican *P. vivax*, but exhibited gene flow with Brazilian *P. vivax* that occurred ~30–60 years ago. The *MSMC-IM* analysis indicated ongoing gene flow between the *P. simium* clusters, with CCR values well above 50% (threshold to suggest split time following Wang *et al.*[48]) and close to 75%, along with the high migration rate ~2 to 4 years ago. This suggests that the two *P. simium* populations have not fully diverged and may still exchange genes. Thus, the SP cluster origin remains unresolved. Indeed, population graphs showed that both *P. simium* clusters connected to the ancestor of Mexican *P. vivax* (Fig. 5), while the *MSMC-IM* results indicated that the SP cluster may have originated between 30 and 60 years ago from human *P. vivax* from Brazil (Fig. 6). The close genetic proximity between the *P. simium* clusters may reflect recent secondary contacts, as suggested by the recent high rate of migration (~2-4 years ago, see Fig. 5B). Nevertheless the SP cluster small sample size (n=6) limits conclusions about its origin and evolutionary dynamics. Previous studies proposed the possibility of multiple host jumps based on the presence of polymorphisms in few mitochondrial or nuclear markers[22,23]. An alternative hypothesis could be a single host jump that involved several *P. vivax* strains infecting NHPs. Therefore, we cannot rule out the possibility that the SP cluster derived from the RJ cluster and that the genetic distinction observed between clusters might be attributed to host adaptation dynamics. Specifically, the RJ cluster infects both simian and human hosts, whereas the SP cluster seems to infect exclusively humans. Larger samples are needed to determine the role of host specificity in the observed genetic divergence, but also the influence of human movement between these two clusters. Scenario-testing methods, such as *DIYABC-RF*[75], *dadi*[76] and *fastsimcoal*[77], would also be useful to clarify *P. simium* evolutionary history.

Importantly, the inferred divergence times may be biased by the life-history traits of *P. simium* and *P. vivax*. *Relate* and *MSMC-IM* rely on the Wright-Fisher (WF) coalescent model (Kingman's coalescent model[78]) that assumes a panmictic population with non-overlapping generations and neutral evolution[79,80]. However, this assumption does not hold for *P. vivax* and likely for *P. simium* because they exhibit dormancy and emerge when the host is infected with new strains[81,82]. Consequently, diversity is maintained, the effect of drift is attenuated, and the effective population size may be overestimated[83,84]. This could explain the absence of a detectable host-transfer bottleneck in the *P. simium* population size inferences (Fig. 5). Moreover, the Kigman's coalescent model assumes only the coalescence of two lineages per generation[78]. However *Plasmodium* life cycle alternates between asexual reproduction in primates and sexual reproduction in mosquitoes, potentially

leading to simultaneous coalescent events[85]. When such events are frequent the Kingman's model underestimates the effective population sizes, although it remains accurate if these events are rare[86]. To address these biases, we excluded clonal and related individuals (see Materials and Methods). Despite these precautions, the WF model limitations and the unique biology of *P. simium* and *P. vivax* suggest that divergence times and population size estimates should be interpreted with caution because they may not fully capture the parasite complex evolutionary dynamics.

Through its evolution, *P. simium* has experienced both human-to-wild animal and wild animal-to-human transmission, making it an interesting candidate for studying the genetics basis of host adaptation. The recent host shift is likely to have imposed significant selection pressure on *P. simium* genomes, and this study provides evidence of such selection. Specifically, we found significant signals in SNPs located in two genes (*PvRBP2a* and *PvRBP1b)* that play a crucial role in the interaction with the NHP host red blood cells (Fig. 6). These proteins, apically expressed in merozoites, facilitate *P. vivax* binding to reticulocytes[55–57]. Previous studies identified a deletion in *P. simium PvRBP2a* CDR and divergence between *P. vivax* and *P. simium*[20,21]. Our findings suggest that these selection signals may reflect linked selection near the deletion (~1,912 nucleotides away). This is the first report of positive selection in *PvRBP1b* in *P. simium*. Given the essential role of these two genes in reticulocyte invasion[55–57], our results suggest that *P. simium* has adapted to its new hosts to effectively invade their reticulocytes. It would therefore be relevant to study the impact of this parasite on the health and survival of the NHPs in this region, which are already threatened[71]. Noteworthy, within the SP cluster, which comprises *P. simium* samples that infect only humans, we discovered significant divergence in the *PvGAMA* gene compared with American human *P. vivax* (Fig. 6C). *PvGAMA* binds to human erythrocytes irrespective of their Duffy antigen status[63], suggesting that some *P. simium* genetic groups may have specifically adapted to human hosts. Functional studies on reticulocyte invasion in simian and human hosts based on the *PvGAMA* genotypes will be crucial to determine the gene role in host adaptation.

P. simium adapted not only to primate hosts but also to new mosquito vectors. Unlike *P. vivax*, *P. simium* is confined to the Atlantic coast in Brazil and transmitted primarily by anopheline mosquitoes of the *Kerteszia* subgenus, mainly *An. cruzii* and *An. bellator*[29,30]. We identified selection signals in the *PvCMRP1* gene for all *P. simium* samples and in the *PvPAT* gene only for the SP cluster (Fig. 6). Previous studies showed that *PvCMRP1* underwent positive selection in Latin American *P. vivax* populations[37,87,88], suggesting that the signal observed in *P. simium* may reflect a similar evolutionary pressure in American *P. vivax*. Although the exact functions of *PvCMRP1* and *PvPAT* are not fully known, studies in which their orthologs were knocked out showed that they are crucial for completing the mosquito life cycle and transmission to vertebrate hosts[58–60].

de Oliveira *et al.*[21] detected selection signals in the *pvs47* gene (potentially essential for avoiding the mosquito immune system), the sequence of which differs between *P. vivax* and *P. simium* by only two amino acid substitutions[89]. Although both SNPs are present in our dataset, our analyses did not find them to be under selection. The small *P. simium* sample size (n=19) may have limited the power of our analyses, particularly with methods like *Relate* and haplotype-based tests (*iHS*, *Rsb*, and *XP-EHH*). For the genome scan using *PBS*, which is more robust with small samples[90], the selection signals may not have been detected due to the similarity of *P. simium* sequences to those of Brazilian and Colombian *P. vivax*[21], the

reference populations used. Therefore, additional *P. simium* samples are necessary to identify such signals.

We found no evidence of selection in *P. simium* genes associated with resistance to antimalarial treatments, primarily chloroquine and primaquine. For *P. vivax*, no chloroquine resistance was reported in Latin America before 2000, and its spread across the continent remains unclear[91,92]. Additionally, no *in vivo* resistance to primaquine has been reported[92]. Consequently, it is possible that *P. simium* is not resistant to these treatments, which would explain the absence of selection evidence for resistance-related genes. Similarly, Brasil *et al.*[5] reported no treatment resistance in humans infected with *P. simium*. However, the limited sample size (n=19) may have been insufficient to detect such signals.

Noteworthy, we detected significant positive selection signals in 73 genes, including *PvIMC1h*, *PvG377*, and *PvGRASP* (Supplementary Table 2). These genes may have played a role in *P. simium* adaptation to the mosquito and/or primate host, but their function is currently unknown. Functional analyses could provide additional insights into *P. simium* adaptive processes.

In conclusion, zoonotic malaria is a significant issue in Latin America, as highlighted by a recent *P. simium* outbreak in the Atlantic Forest of Rio de Janeiro, Brazil. This underscores the potential role of NHPs malaria reservoirs and emphasizes the need of extensive research on their genetic diversity, evolutionary history, and zoonotic potential. Furthermore, the health and conservation impacts of these infections on NHPs, particularly endangered species in the Atlantic Forest, remain largely unexplored. Our results, based on 646 NHP samples from five Latin American countries, provide the first evidence of natural *P. vivax* infections in Colombian NHPs, highlighting the possibility of ongoing host switches and their relevance for malaria elimination strategies. The genetic analyses revealed recent admixture events and gene flow between *P. simium* and *P. vivax* populations, emphasizing the porous boundaries between these species and their adaptive processes. Using different population genomic approaches, we consistently demonstrated that *P. simium* forms a separate cluster more closely related to American *P. vivax* populations. We then identified two genetic clusters within *P. simium*: one predominantly in Rio de Janeiro state and another in neighbouring São Paulo state. The close genetic relationship between *P. simium* and human *P. vivax* suggests recent host transfers, with divergence occurring in the last 30 years and migration events dating back to 100-200 years, likely involving a host jump from Mexican *P. vivax* to Brazilian *P. simium*. Despite the limited sample size, which may have hindered the detection of some selection signals, we identified several genes potentially involved in *P. simium* adaptation to NHPs, human red blood cells, and mosquito vectors. Our findings question the distinction between *P. simium* and *P. vivax* as separate species because both parasites can infect humans and Neotropical NHPs and produce hybrids, and suggest that *P. simium* may represent distinct *P. vivax* populations. The results presented in this study have profound implications for malaria eradication efforts in Latin America. By highlighting the potential NHPs role as zoonotic malaria reservoirs and the porous genetic boundaries between *P. vivax* and *P. simium*, our study underscores the need to incorporate zoonotic dynamics into malaria control strategies, while promoting coexistence between humans and NHPs.

Materials and methods

P. simium sample collection and ethical statements

In French Guiana, non-human vertebrate samples were collected and used in accordance with an international CITES permit (Convention on International Trade in Endangered Species of Wild Fauna and Flora; permit FR973A) according to the French legislation. Following the sharing policy in French Guiana, non-human mammal samples are registered in the collection JAGUARS (https://kwata.net/gestion-collection-biologique/, CITES reference: FR973A) supported by the Kwata NGO (accredited by the French Ministry of the Environment and the Prefecture of French Guiana, Agreement R03-2019-06-19-13), Institut Pasteur de la Guyane, DGTM Guyane, Collectivité Territoriale de la Guyane, and validated by the French Guianan prefectoral decree n°2012/110.

In Costa Rica, the animal study was reviewed and approved by the Institutional Committee for the Care and Use of Animals (Comité Institucional para el Cuidado y Uso de los Animales) of the Universidad de Costa Rica, and adhered to the Costa Rica legal requirements (collection permit number: MINAET-SINAC, Costa Rica: 042-2012-SINAC).

In Brazil, the collected animals were dead and part of the epidemiology study of the Rio Grande do Sul State Health Secretaria, with a license issued by the Ministry of Health.

In Colombia, ethical approvals for the collection of fecal and blood samples as well as the collection of vectors were obtained by Universidad de los Andes, the National Environmental Licensing Authority of Colombia (ANLA) and the Centers for Disease Control and Prevention (permits numbers: 2017025578-1-000, 2017043863-1-000, 2017065795-1-000, 2017013727-1-000, 2017052943-1-000, 2017081458-1-000, 2017108650-1-000).

In Argentina, permits were obtained from National Parks to conduct this research (IF 2022 85832064 APN DRNEA # APNAC PROYECT NUMBER 541 and IF 2022 85832064 APN DRNEA # APNAC PROYECT NUMBER 544).

DNA extraction and *Plasmodium* screening

For each of the 646 samples collected from NHPs in Latin America, genomic DNA was extracted using the Qiagen DNeasy Blood and Tissue Kit according to the manufacturer's recommendations. *P. simium* and *P. vivax* samples were identified by amplification of *Plasmodium cytochrome b* using nested PCR, as described in Prugnolle *et al.*[31]. The reaction products were visualized on 1.5% agarose gels stained with EZ-vision.

For *P. simium* or *P. vivax*-positive samples, selective whole-genome amplification (sWGA) was performed to enrich submicroscopic levels of *Plasmodium* DNA, following the protocol described by Cowell *et al.*[93]. This technique preferentially amplifies *P. vivax* and *P. simium* genomes from a set of target DNAs and avoids host DNA contamination. For each sample, DNA amplification was carried out using the strand-displacing phi29 DNA polymerase and *P. vivax*-specific primers that target short (6 to 12 nucleotides) motifs commonly found in the *P. vivax* genome (PvSet1[93,94]). Although *P. vivax*-specific, these primers have already been used successfully to amplify other *Plasmodium* species phylogenetically close to *P. vivax* such as *P. vivax-like* identified in African great apes[35]. About 30 ng of input DNA was added to a 50 µL reaction mixture containing 3.5 µM of each

sWGA primer, 30 U of phi29 DNA polymerase enzyme (New England Biolabs), 1× phi29 buffer (New England Biolabs), 4 mM deoxynucleotide triphosphates (Invitrogen), 1% bovine serum albumin, and sterile water. DNA was amplified in a thermal cycler with the following program: a ramp down from 35° to 30°C (10 min per degree), 16 hours at 30°C, 10 min at 65°C, and hold at 4°C. For each sample, the products of the two amplifications (one per primer set) were purified with AMPure XP beads (Beckman Coulter) at a 1:1 ratio according to the manufacturer's recommendations and pooled at equimolar concentrations. Samples with the highest concentration of parasite genome were selected after qPCR analysis using a Roche LightCycler 96 with the following program: 95°C for 10 minutes; 40 cycles of 95°C for 15 seconds, 60°C for 20 seconds, and 72°C for 20 seconds; 95°C for 10 seconds; and 55°C for 1 minute. Each sWGA library was prepared using the two pooled amplification products and the Nextera XT DNA kit (Illumina), following the manufacturer's protocol. Then, samples were pooled, clustered, and sequenced on one lane of a Illumina Novaseq-6000 S4 with PE150 with 2 × 150-bp paired-end reads.

Short-read mapping, SNP calling, and data compilation

Three newly sequenced *P. vivax* isolates collected in Colombian monkeys were added to a compilation of several previously published genomic datasets: *i*) 621 *P. vivax* samples from Daron *et al.*[35], Benavente *et al.*[87], the MalariaGEN *P. vivax* Genome Variation project[95], Mourier *et al.*[20], and Lefebvre *et al.* [36]; *ii*) 31 *P. simium* samples from Mourier *et al.*[20], de Oliveira *et al.* [21], and Ibrahim *et al.*[19]; and *iii*) two *P. vivax-like* genomes collected on Nigeria-Cameroon chimpanzees (*Pan troglodytes ellioti*), from Loy *et al.*[96] (Supplementary Fig. 1). Short read archives (SRA) from the literature were retrieved from NCBI (accession numbers in Supplementary Table 1).

Short reads were trimmed to remove potential lingering adapters and preprocessed to eliminate low-quality reads (*−quality-cutoff = 30*) using the *cutadapt* program[97]. Reads shorter than 50 bp and containing "N" were discarded (*−minimumlength = 50 –max-n = 0*). Cleaned paired-end reads were mapped to the *P. vivax* reference genome PVP01[98] using *bwa-mem*[99] with default options. Duplicate reads were marked using the *MarkDuplicates* tool from the *Picard tools* v2.5.0 (broadinstitute.github.io/picard/) with default options. Local realignment around indels was performed using the *IndelRealigner* tool from *Genome Analysis Toolkit* (*GATK*[100], v3.8.0). Variants were called using the *HaplotypeCaller* module in *GATK* with default options. Lastly, the different isolated variant call format (VCF) files were merged using the *GATK* module *CombineGVCFs*.

Data filtering

For all samples (*P. vivax* and *P. simium*) from Mourier *et al.*[20], de Oliveira *et al.*[21], and Ibrahim *et al.*[19], whether *P. vivax* or *P. simium*, the same filters used in Lefebvre *et al.*[36] were applied. Thus samples with >50% missing data were removed. As several strains can infect the same host, the within-host infection complexity was assessed with the F_{WS} metric[101], calculated using *vcfdo* (github.com/IDEELResearch/vcfdo; last accessed July 2022). Samples with mono-clonal infections (*i.e.* F_{WS} >0.95) were kept. Highly related samples and clones could generate spurious signals of population structure, biased estimators of population genetic variation, and violated the assumptions of the model-based population genetic approaches used in this study[102]. Therefore, the relatedness between haploid

genotype pairs was measured by estimating the pairwise fraction of the genome IBD using the *hmmIBD* program[42], with the default parameters for recombination and genotyping error rates, and using the allele frequencies estimated by the program. Within each country and species, isolate pairs that shared >50% of IBD were considered highly related. Only the strain with the lowest amount of missing data was retained in each family of related samples.

The final dataset included 429 samples: 404 *P. vivax* isolates, 19 *P. simium* isolates, 4 *P. vivax* isolates from NHPs, and 2 *P. vivax-like* isolates used as outgroup species in phylogenetic analyses.

For the VCF with 426 samples (without the three newly sequenced samples), only bi-allelic SNPs with less than 50% missing data and a genotype quality (GQ) score above 30 were kept. SNPs with a minimum average depth coverage of 5x and a maximum average depth of 152x on all samples were kept. Singletons were removed because they can be sequencing errors. This dataset based on hard filters (*dataset_HF*) included 819,652 SNPs with a mean density of 38.29 SNPs per kilobase (Supplementary Fig. 1).

For the *ANGSD* software suite that uses the genotype likelihood (GL or GQ) rather than the actual genotype calling, all sites with a genotype quality >20, a maximum depth of 155x per sample, a minimum total depth of 5x on all samples, and present in at least 5 individuals were kept. Only bi-allelic SNPs with a *p-value* of 1e-6 were kept. A total of 20,844,131 sites from the core genome were considered, including 1,286,079 variants (65.22 SNPs/kb). This approach allowed considering low to very low coverage samples[32,103] and led to a dataset that included all 429 samples (*dataset _GL*). All filtration steps for both datasets are detailed in Supplementary Figure 1.

Population structure and relationships between populations

The principal component analysis (PCA) and the individual genetic ancestry analyses were carried out with *ANGSD* v.040 and *PCAngsd* v0.98, after selecting only biallelic SNPs present in the *P. vivax* core genome and excluding SNPs with a minor allele frequency (MAF) <5% from *dataset_GL*. Variants were LD-pruned to obtain a set of unlinked variants using *ngsld* v1.1.1[104], with a threshold of r^2=0.5 and windows of 0.5 kb. In total, these analyses included 247,890 SNPs for 425 individuals (the *P. vivax-like* outgroup was not considered). For the ancestry analyses, using *PCAngsd*, a number of clusters (*K*) ranging from 2 to 8 were tested. The best K value was inferred based on the broken-stick eigenvalues plus one (Supplementary Fig. 2), following the recommendations of Meisner and Albrechtsen[33]. The results were post-processed using *pong* v1.5[105] to compute the ancestry proportions.

The maximum likelihood (ML) phylogenetic tree was obtained with *IQ-TREE*[34] using the best-fitted model determined by *ModelFinder*[106]. Considering the *dataset_HF,* only SNPs with ≤10% missing data and samples with <50% missing data were retained, resulting in 48,070 SNPs in the core genome for 404 samples. As the dataset was composed only of SNPs and no invariant sites, the ascertainment bias correctionwas added to the tested models, following the *IQ-TREE* user guide recommendations. According to the Akaike information criterion[107], the best model of nucleotide evolution was the general time reversible (GTR) model that integrates unequal rates and unequal base frequency[108]. The node support with the Ultrafast Bootstrap Approximation (UFboot)[39] and SH-aLRT methods[38]. Nodes were considered as highly supported only if SH-aLRT ≥ 80% and UFboot ≥ 95%, following the thresholds by Guindon *et al.*[38] and Minh *et al.*[39].

PCA and genetic ancestry analysis were performed focusing only on the *P. simium* samples to investigate the genetic substructure within this group. As these analyses require a data set with unlinked variants, SNPs from the *dataset_HF* were LD-pruned with *PLINK* v.2[109]. All SNPs with a correlation coefficient r^2>0.5 (parameters: *–indep-pairwise 50 10 0.5*) were removed using a window size of 50 SNPs, and a sliding step of 10 bp. For these analyses, 16,008 SNPs remained for the 19 samples. The PCA was carried out with *PLINK v.2* and the following parameters: *–maf 0.01 –mind 0.99*. Then, individual genetic ancestry was estimated using the *ADMIXTURE* v 1.3.0 software[40]. Different cluster (*K*) numbers were tested from 1 to 5, with 10 replicates for each *K* to check consistency among replicates. The optimal *K* value was estimated using the cross-validation error-rate index, and convergence was checked with *pong*[105].

For the pairwise-IBD network, the complete *dataset_HF*, composed of 819,652 SNPs and 426 samples, was used. Pairwise-IBD values among pairs of isolates were estimated using the *hmmIBD* program[110], with the default parameters for recombination and genotyping error rates, and using the allele frequencies estimated by the program. The pairwise-IBD relationships among isolates were visualized as a network that displayed only pairwise IBD relationships with values ≥ 5%. The network was plotted using the *igraph* package in R[41] and default options.

Population graphs were built using *TreeMix*[44] and *AdmixtureBayes*[43]. First, the *dataset_HF* was pruned using PLINK[109] with a threshold of r^2=0.5, a sliding window of 50 SNPs and a step of 10 SNPs (396,835 SNPs remaining). All SNPs with missing data were removed. The final dataset used in these analyses included 11,907 SNPs for 18 ingroup populations and one outgroup population (*P. vivax-like*).

The number of migration events (*m*) in *TreeMix* that best fitted the data was calculated by running the analysis 15 times for each *m* value, with *m* ranging from 1 to 10. The optimal *m* value was estimated using the *OptM* R package[111] and the Evanno method (Supplementary Fig. 5). Then, a consensus tree with bootstrap node support was obtained by running 100 replicates of *TreeMix* and postprocessing them using the *BITE* R-package[112] for the *m* optimal values, following Daron *et al.*[35] and Lefebvre *et al.*[113].

For the *AdmixtureBayes* analysis, three independent runs were performed, each one including 40 coupled Monte Carlo Markov Chains (MCMC) and 250,000 steps. The results of these chains were analyzed after a burn-in of 50% and a thinning step of 10, using the default parameters. Convergence across the trees was assessed using the Gelman-Rubin (GR) statistics for the models parameters (i.e., posterior probability, branch lengths and number of admixtures), as recommended by Nielsen *et al.*[43]. Runs were deemed convergent if the GR statistics for the model parameters converged toward the value of 1 along the MCMC steps (Supplementary Fig. 4). The retained network was common among the three best networks in each chain that converged, based on their posterior probabilities.

The admixture f_4-statistics were computed with *ADMIXTOOLS 2* v2.0.0[46] and the same dataset used for *TreeMix* and *AdmixtureBayes*,.

Demographic History of American *P. simium*

The nucleotide diversity (π) values were estimated using *pixy* v1.2.7[114], in a non-overlapping sliding window of 500 pb along the genome based on the *dataset_GL*. The sample size was standardized (*i.e.*, 10 randomly chosen isolates for each population) to obtain values that could be compared.

Relate[47] was used to infer historical changes in N_e and to estimate the divergence time between populations. The ancestral/derived states of 1,267,897 sites from the *dataset_GL* were defined using the consensus sequence of the outgroup (*P. vivax-like*). Only the core genome regions were considered, the rest was masked. As the *P. simium* samples were not enough (n=19) to infer a *P. simium*-specific genetic map, the *P. vivax* genetic map from Lefebvre *et al.*[36] was used with a mutation rate of 6.43×10^{-9} mutations/site/generation and a generation time of 5.5 generations/years[35]. Coalescence rates were calculated from 10 to 1,000,000 generations ago (options *--bins 0,7,0.25*).

Then, MSMC-IM[48], based on the coalescent rates estimated within and between two populations from *Relate*[47], was used to fit an isolation-with-migration (IM) model. The scripts to convert *Relate* results to *MSMC-IM* were kindly provided by E. Patin and D. Liu from the Human Evolutionary Genetics team at the Institut Pasteur, Paris, France. The migration rates were inferred between *P. simium* clusters and two *P. vivax* populations (Brazil and Mexico). These *P. vivax* populations were selected based on their known evolutionary history and our population structure results. Brazilian and Mexican *P. vivax* were included due to their geographic proximity to *P. simium*, and genetic similarity to *P. simium*, respectively. The mutation rate was the one used for *Relate*, and the population sizes were derived from the *Relate* results.

Positive selection detection

Given the low sample size in the two *P. simium* genetic clusters (n = 13 and n = 6), they were combined for the selection scan analyses with *rehh*[115] and *Relate*[47], following Hupalo *et al.*[37] and Lefebvre *et al.*[113].

The haplotype-based tests (*iHS, XP-EHH,* and *Rsb*) were used to detect evidence of recent selective sweeps, using *rehh*[115] and *dataset_HF*. The significance threshold was set at -log(*p-value*) = 4, as recommended by Gautier *et al.*[115]. For *XP-EHH* and *Rsb*, the *P. simium* samples were compared with Brazilian *P. vivax* samples and only SNPs with negative standardized values (*i.e.*, indicating positive selection for the *P. simium* cluster rather than for *P. vivax*) were considered. The ancestral/derived states of 378,441 SNPs from *dataset_HF* were defined using the consensus sequence of the two outgroup *P. vivax-like* isolates. The Brazilian population was chosen due to its larger sample size (n=29). The Mexican population did not have a large enough sample size (n=9) to be used as reference population to detect selection signals[116].

The *Relate* program used to estimate the coalescent rates in the demographic section was also used to detect positive selection by identifying genomic regions with significantly lower coalescent rate values than the rest of the genomes. The significance threshold was set at -log(*p-value*) = 4, to take into account the multiple tests.

The population branch-site (*PBS*) values were calculated for each *P. simium* cluster, with *ANGSD* v0.40 in sliding windows of the genome (windows of 1 kb with a sliding step of 500 pb), based on *dataset_GL*. The outgroup populations were *P. vivax* from humans in Brazil and Colombia. All windows with <20 SNPs were removed to avoid extreme values caused by a low SNP number. The 0.1% most extreme values were considered evidence that the genomic regions displayed signs of selection in our *P. simium* clusters.

Once selection signals were detected, the identified genes were annotated using the general feature format (GFF) file (*PlasmoDB-50_PvivaxP01.gff*, available on *PlasmoDB*) and the intersect function of *BEDtools* v 2.26.0[117]. Additional information (*e.g.*, gene name,

function, and biological process) was retrieved from *PlasmoDB* (plasmodb.org, accessed in April 2024).

Data availability

The sequences of the newly sequenced and analyzed simian *P. vivax* samples have been deposited on NCBI (Bioproject PRJNA116416). The scripts used in this study are available in this github repository: https://github.com/MargauxLefebvre/EvoHistory_Simium

References

1. Garnham, P. C. Malaria in its various vertebrate hosts. (1980).
2. Geneva: World Health Organization. *World Malaria Report 2023.* https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2023 (2023).
3. Rougeron, V. *et al.* A population genetic perspective on the origin, spread and adaptation of the human malaria agents *Plasmodium falciparum* and *Plasmodium vivax*. *FEMS Microbiol. Rev.* (2021) doi:10.1093/femsre/fuab047.
4. Baird, J. K. Malaria zoonoses. *Travel Med. Infect. Dis.* **7**, 269–277 (2009).
5. Brasil, P. *et al.* Outbreak of human malaria caused by *Plasmodium simium* in the Atlantic Forest in Rio de Janeiro: a molecular epidemiological investigation. *Lancet Glob. Health* **5**, e1038–e1046 (2017).
6. Su, X. & Wu, J. Zoonotic transmissions and host switches of malaria parasites. *Zoonoses Burlingt. Mass* **1**, (2021).
7. Yusof, R. *et al.* High proportion of *knowlesi* malaria in recent malaria cases in Malaysia. *Malar. J.* **13**, 168 (2014).
8. Herdiana, H. *et al.* Malaria risk factor assessment using active and passive surveillance data from Aceh Besar, Indonesia, a low endemic, malaria elimination setting with *Plasmodium knowlesi*, *Plasmodium vivax*, and *Plasmodium falciparum*. *Malar. J.* **15**, 468 (2016).
9. Ta, T. H. *et al.* First case of a naturally acquired human infection with *Plasmodium cynomolgi*. *Malar. J.* **13**, 68 (2014).
10. Bueno, M. G. *et al.* Survey of *Plasmodium* spp. in Free-Ranging Neotropical Primates from the Brazilian Amazon Region impacted by Anthropogenic Actions. *EcoHealth* **10**, 48–53 (2013).
11. van de Straat, B. *et al.* Zoonotic malaria transmission and land use change in Southeast Asia: what is known about the vectors. *Malar. J.* **21**, 109 (2022).
12. Rougeron, V., Daron, J., Fontaine, M. C. & Prugnolle, F. Evolutionary history of *Plasmodium vivax* and *Plasmodium simium* in the Americas. *Malar. J.* **21**, 141 (2022).
13. Da Fonseca, F. [*Plasmodium* of a primate of Brazil] (in portugese). *Mem. Inst. Oswaldo Cruz* **49**, 543–553 (1951).
14. de Oliveira, T. C., Rodrigues, P. T., Duarte, A. M. R. C., Rona, L. D. P. & Ferreira, M. U. Ongoing host-shift speciation in *Plasmodium simium*. *Trends Parasitol.* **37**, 940–942 (2021).
15. Deane, L. M., Deane, M. P. & Neto, J. F. Studies on transmission of simian malaria and on a natural infection of man with *Plasmodium simium* in Brazil. *Bull. World Health Organ.* **35**, 805 (1966).
16. de Castro Duarte, A. M. R. *et al.* Natural *Plasmodium* infections in Brazilian wild monkeys: Reservoirs for human infections? *Acta Trop.* **107**, 179–185 (2008).
17. Rondón, S., León, C., Link, A. & González, C. Prevalence of *Plasmodium* parasites in non-human primates and mosquitoes in areas with different degrees of fragmentation in

Colombia. *Malar. J.* **18**, 276 (2019).
18. Chaves, A. *et al.* Presence and potential distribution of malaria-infected New World primates of Costa Rica. *Malar. J.* **21**, 17 (2022).
19. Ibrahim, A. *et al.* Population-based genomic study of *Plasmodium vivax* malaria in seven Brazilian states and across South America. *Lancet Reg. Health – Am.* **18**, (2023).
20. Mourier, T. *et al.* The genome of the zoonotic malaria parasite *Plasmodium simium* reveals adaptations to host switching. *BMC Biol.* **19**, 219 (2021).
21. de Oliveira, T. C. *et al. Plasmodium simium*: population genomics reveals the origin of a reverse zoonosis. *J. Infect. Dis.* **224**, 1950–1961 (2021).
22. Lim, C. S., Lim, C. S., Tazi, L. & Ayala, F. J. *Plasmodium vivax*: Recent world expansion and genetic identity to *Plasmodium simium*. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15523–15528 (2005).
23. Tazi, L. & Ayala, F. J. Unresolved direction of host transfer of *Plasmodium vivax* v. *P. simium* and *P. malariae* v. *P. brasilianum*. *Infect. Genet. Evol.* **11**, 209–221 (2011).
24. Li, J. *et al.* Geographic subdivision of the range of the malaria parasite *Plasmodium vivax*. *Emerg. Infect. Dis.* **7**, 35–42 (2001).
25. Carter, R. Speculations on the origins of *Plasmodium vivax* malaria. *Trends Parasitol.* **19**, 214–219 (2003).
26. Laporta, G. Z. *et al.* Malaria vectors in South America: current and future scenarios. *Parasit. Vectors* **8**, 426 (2015).
27. Zimmerman, R. H. Ecology of malaria vectors in the Americas and future direction. *Mem. Inst. Oswaldo Cruz* **87**, 371–383 (1992).
28. Frederick, J. *et al.* Malaria vector research and control in Haiti: a systematic review. *Malar. J.* **15**, 376 (2016).
29. Marrelli, M. T., Malafronte, R. S., Sallum, M. A. & Natal, D. Kerteszia subgenus of Anopheles associated with the Brazilian Atlantic rainforest:current knowledge and future challenges. *Malar. J.* **6**, 127 (2007).
30. Pina-Costa, A. de *et al.* Malaria in Brazil: what happens outside the Amazonian endemic region. *Mem. Inst. Oswaldo Cruz* **109**, 618–633 (2014).
31. Prugnolle, F. *et al.* African great apes are natural hosts of multiple related malaria species, including *Plasmodium falciparum*. *Proc. Natl. Acad. Sci.* **107**, 1458–1463 (2010).
32. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).
33. Meisner, J. & Albrechtsen, A. Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics* **210**, 719–731 (2018).
34. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
35. Daron, J. *et al.* Population genomic evidence of *Plasmodium vivax* Southeast Asian origin. *Sci. Adv.* **7**, eabc3713 (2021).
36. Lefebvre, M. J. M. *et al.* Genomic exploration of the complex journey of *Plasmodium vivax* in Latin America. *bioRxiv* 2024.05.08.592893 (2024) doi:10.1101/2024.05.08.592893.
37. Hupalo, D. N. *et al.* Population genomics studies identify signatures of global dispersal and drug resistance in *Plasmodium vivax*. *Nat. Genet.* **48**, 953–958 (2016).
38. Guindon, S. *et al.* New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
39. Minh, B. Q., Nguyen, M. A. T. & von Haeseler, A. Ultrafast Approximation for Phylogenetic Bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013).
40. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
41. Csardi, G. & Nepusz, T. The igraph software. *Complex Syst* **1695**, 1–9 (2006).
42. Schaffner, S. F., Taylor, A. R., Wong, W., Wirth, D. F. & Neafsey, D. E. hmmIBD:

software to infer pairwise identity by descent between haploid genotypes. *Malar. J.* **17**, 196 (2018).
43. Nielsen, S. V. *et al.* Bayesian inference of admixture graphs on Native American and Arctic populations. *PLOS Genet.* **19**, e1010410 (2023).
44. Pickrell, J. & Pritchard, J. Inference of population splits and mixtures from genome-wide allele frequency data. *Nat. Preced.* (2012) doi:10.1038/npre.2012.6956.1.
45. Patterson, N. *et al.* Ancient Admixture in Human History. *Genetics* **192**, 1065–1093 (2012).
46. Maier, R. & Patterson, N. admixtools: Inferring demographic history from genetic data. (2024).
47. Speidel, L., Forest, M., Shi, S. & Myers, S. R. A method for genome-wide genealogy estimation for thousands of samples. *Nat. Genet.* **51**, 1321–1329 (2019).
48. Wang, K., Mathieson, I., O'Connell, J. & Schiffels, S. Tracking human population structure through time from whole genome sequences. *PLOS Genet.* **16**, e1008552 (2020).
49. Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46**, 919–925 (2014).
50. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A Map of Recent Positive Selection in the Human Genome. *PLOS Biol.* **4**, e72 (2006).
51. Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
52. Tang, K., Thornton, K. R. & Stoneking, M. A New Approach for Using Genome Scans to Detect Recent Positive Selection in the Human Genome. *PLOS Biol.* **5**, e171 (2007).
53. Beck, J. R. & Ho, C.-M. Transport mechanisms at the malaria parasite-host cell interface. *PLOS Pathog.* **17**, e1009394 (2021).
54. Elsworth, B. *et al.* Proteomic analysis reveals novel proteins associated with the *Plasmodium* protein exporter PTEX and a loss of complex stability upon truncation of the core PTEX component, PTEX150. *Cell. Microbiol.* **18**, 1551–1569 (2016).
55. Galinski, M. R., Xu, M. & Barnwell, J. W. *Plasmodium vivax* reticulocyte binding protein-2 (PvRBP-2) shares structural features with PvRBP-1 and the *Plasmodium yoelii* 235 kDa rhoptry protein family. *Mol. Biochem. Parasitol.* **108**, 257–262 (2000).
56. Galinski, M. R., Medina, C. C., Ingravallo, P. & Barnwell, J. W. A reticulocyte-binding protein complex of *Plasmodium vivax* merozoites. *Cell* **69**, 1213–1226 (1992).
57. Galinski, M. R. & Barnwell, J. W. *Plasmodium vivax*: Merozoites, invasion of reticulocytes and considerations for malaria vaccine development. *Parasitol. Today* **12**, 20–29 (1996).
58. Hart, R. J., Lawres, L., Fritzen, E., Mamoun, C. B. & Aly, A. S. I. *Plasmodium yoelii* Vitamin B5 Pantothenate Transporter Candidate is Essential for Parasite Transmission to the Mosquito. *Sci. Rep.* **4**, 5665 (2014).
59. Kenthirapalan, S., Waters, A. P., Matuschewski, K. & Kooij, T. W. A. Functional profiles of orphan membrane transporters in the life cycle of the malaria parasite. *Nat. Commun.* **7**, 10519 (2016).
60. Thompson, J. *et al. Plasmodium* cysteine repeat modular proteins 1–4: complex proteins with roles throughout the malaria parasite life cycle. *Cell. Microbiol.* **9**, 1466–1480 (2007).
61. Losada, A., Hirano, M. & Hirano, T. Identification of Xenopus SMC protein complexes required for sister chromatid cohesion. *Genes Dev.* **12**, 1986–1997 (1998).
62. Yoshinaga, M. & Inagaki, Y. Ubiquity and Origins of Structural Maintenance of Chromosomes (SMC) Proteins in Eukaryotes. *Genome Biol. Evol.* **13**, evab256 (2021).
63. Cheng, Y. *et al. Plasmodium vivax* GPI-anchored micronemal antigen (PvGAMA) binds human erythrocytes independent of Duffy antigen status. *Sci. Rep.* **6**, 35581 (2016).
64. Costa, D. C. *et al. Plasmodium simium/Plasmodium vivax* infections in southern brown howler monkeys from the Atlantic Forest. *Mem. Inst. Oswaldo Cruz* **109**, 641–643 (2014).
65. Alvarenga, D. A. M. *et al.* New potential *Plasmodium brasilianum* hosts: tamarin and

marmoset monkeys (family Callitrichidae). *Malar. J.* **16**, 71 (2017).
66. Kattenberg, J. H. *et al.* Population genomic evidence of structured and connected *Plasmodium vivax* populations under host selection in Latin America. *Ecol. Evol.* **14**, e11103 (2024).
67. Bustamante-Matoma, H. & Agudelo-Ruiz, J. Local perspectives and analysis of the impact of Chinese migration in the Colombian territory. *Univ.-XXI Rev. Cienc. Soc. Humanas* 15–41 (2024).
68. Nijman, V. *et al.* Global online trade in primates for pets. *Environ. Dev.* **48**, 100925 (2023).
69. Andrade, M. C. R. *et al.* Biologic data of *Macaca mulatta, Macaca fascicularis*, and *Saimiri sciureus* used for research at the fiocruz primate center. *Mem. Inst. Oswaldo Cruz* **99**, 584–589 (2004).
70. Leclerc, M. C. *et al.* Meager genetic variability of the human malaria agent *Plasmodium vivax*. *Proc. Natl. Acad. Sci.* **101**, 14455–14460 (2004).
71. IUCN. *The IUCN Red List of Threatened Species*. Version 2024-2. (2024). [Last accessed the 19/12/2024]
72. Bicca-Marques, J. C. & de Freitas, D. S. The Role of Monkeys, Mosquitoes, and Humans in the Occurrence of a Yellow Fever Outbreak in a Fragmented Landscape in South Brazil: Protecting Howler Monkeys is a Matter of Public Health. *Trop. Conserv. Sci.* **3**, 78–89 (2010).
73. Hill, L. F. The Confederate Exodus to Latin America, I. *Southwest. Hist. Q.* **39**, 100–134 (1935).
74. Sutherland, D. E. Exiles, emigrants, and sojourners: The post-Civil War Confederate exodus in perspective. *Civ. War Hist.* **31**, 237–256 (1985).
75. Collin, F.-D. *et al.* Extending approximate Bayesian computation with supervised machine learning to infer demographic history from genetic polymorphisms using DIYABC Random Forest. *Mol. Ecol. Resour.* **21**, 2598–2613 (2021).
76. Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data. *PLOS Genet.* **5**, e1000695 (2009).
77. Excoffier, L. *et al.* fastsimcoal2: demographic inference under complex evolutionary scenarios. *Bioinformatics* **37**, 4882–4885 (2021).
78. Kingman, J. F. C. On the genealogy of large populations. *J. Appl. Probab.* **19**, 27–43 (1982).
79. Fisher, R. A. XXI.—On the dominance ratio. *Proc. R. Soc. Edinb.* **42**, 321–341 (1923).
80. Wright, S. Evolution in Mendelian Populations. *Genetics* **16**, 97–159 (1931).
81. Imwong, M. *et al.* Relapses of *Plasmodium vivax* Infection Usually Result from Activation of Heterologous Hypnozoites. *J. Infect. Dis.* **195**, 927–933 (2007).
82. White, N. J. Determinants of relapse periodicity in *Plasmodium vivax* malaria. *Malar. J.* **10**, 297 (2011).
83. Ferreira, M. U., Gunawardena, S., Kapilananda, G. M. G., Karunaweera, N. D. & Wirth, D. F. The Sri Lankan paradox: high genetic diversity in *Plasmodium vivax* populations despite decreasing levels of malaria transmission. *Parasitology* **141**, 880–890 (2014).
84. Mwima, R., Hui, T.-Y. J., Kayondo, J. K. & Burt, A. The population genetics of partial diapause, with applications to the aestivating malaria mosquito Anopheles coluzzii. *Mol. Ecol. Resour.* **24**, e13949 (2024).
85. Tellier, A. & Lemaire, C. Coalescence 2.0: a multiple branching of recent theoretical developments and their applications. *Mol. Ecol.* **23**, 2637–2652 (2014).
86. Korfmann, K., Sellinger, T. P. P., Freund, F., Fumagalli, M. & Tellier, A. Simultaneous Inference of Past Demography and Selection from the Ancestral Recombination Graph under the Beta Coalescent. *Peer Community J.* **4**, (2024).
87. Benavente, E. D. *et al.* Distinctive genetic structure and selection patterns in *Plasmodium vivax* from South Asia and East Africa. *Nat. Commun.* **12**, 3160 (2021).
88. Benavente, E. D. *et al.* Genomic variation in *Plasmodium vivax* malaria reveals regions under selective pressure. *PLOS ONE* **12**, e0177134 (2017).

89. Tachibana, M. *et al. Plasmodium vivax* gametocyte proteins, Pvs48/45 and Pvs47, induce transmission-reducing antibodies by DNA immunization. *Vaccine* **33**, 1901–1908 (2015).
90. Willing, E.-M., Dreyer, C. & van Oosterhout, C. Estimates of Genetic Differentiation Measured by FST Do Not Necessarily Require Large Sample Sizes When Using Many SNP Markers. *PLOS ONE* **7**, e42649 (2012).
91. Price, R. N. *et al.* Global extent of chloroquine-resistant *Plasmodium vivax*: a systematic review and meta-analysis. *Lancet Infect. Dis.* **14**, 982–991 (2014).
92. Ferreira, M. U. *et al.* Monitoring *Plasmodium vivax* resistance to antimalarials: Persisting challenges and future directions. *Int. J. Parasitol. Drugs Drug Resist.* **15**, 9–24 (2021).
93. Cowell, A. N. *et al.* Selective Whole-Genome Amplification Is a Robust Method That Enables Scalable Whole-Genome Sequencing of *Plasmodium vivax* from Unprocessed Clinical Samples. *mBio* **8**, e02257-16 (2017).
94. Sundararaman, S. A. *et al. Plasmodium falciparum*-like parasites infecting wild apes in southern Cameroon do not represent a recurrent source of human malaria. *Proc Natl Acad Sci U A* **110**, 7020–7025 (2013).
95. Adam, I. *et al.* An open dataset of *Plasmodium vivax* genome variation in 1,895 worldwide samples. *Wellcome Open Res.* **7**, (2022).
96. Loy, D. E. *et al.* Evolutionary history of human *Plasmodium vivax* revealed by genome-wide analyses of related ape parasites. *Proc. Natl. Acad. Sci.* **115**, E8450–E8459 (2018).
97. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
98. Auburn, S. *et al.* A new *Plasmodium vivax* reference sequence with improved assembly of the subtelomeres reveals an abundance of pir genes. *Wellcome Open Res.* **1**, (2016).
99. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics* **25**, 1754–1760 (2009).
100. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
101. Amegashie, E. A. *et al.* Population genetic analysis of the *Plasmodium falciparum* circumsporozoite protein in two distinct ecological regions in Ghana. *Malar. J.* **19**, 437 (2020).
102. Wang, J. Effects of sampling close relatives on some elementary population genetics analyses. *Mol. Ecol. Resour.* **18**, 41–54 (2018).
103. Lou, R. N., Jacobs, A., Wilder, A. P. & Therkildsen, N. O. A beginner's guide to low-coverage whole genome sequencing for population genomics. *Mol. Ecol.* **30**, 5966–5993 (2021).
104. Fox, E. A., Wright, A. E., Fumagalli, M. & Vieira, F. G. ngsLD: evaluating linkage disequilibrium using genotype likelihoods. *Bioinformatics* **35**, 3855–3856 (2019).
105. Behr, A. A., Liu, K. Z., Liu-Fang, G., Nakka, P. & Ramachandran, S. pong: fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics* **32**, 2817–2823 (2016).
106. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
107. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**, 716–723 (1974).
108. Tavaré, S. Some probabilistic and statistical problems on the analysis of DNA sequence. *Lect. Math. Life Sci.* **17**, 57 (1986).
109. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, s13742-015-0047–8 (2015).
110. Schaffner, S. hmmIBD. (2019).
111. Fitak, R. R. OptM: estimating the optimal number of migration edges on population

trees using Treemix. *Biol. Methods Protoc.* **6**, bpab017 (2021).
112. Milanesi, M. *et al.* BITE: an R package for biodiversity analyses. *bioRxiv* 181610 (2017) doi:10.1101/181610.
113. Lefebvre, M. J. M. *et al.* Population genomic evidence of adaptive response during the invasion history of *Plasmodium falciparum* in the Americas. *Mol. Biol. Evol.* msad082 (2023) doi:10.1093/molbev/msad082.
114. Korunes, K. L. & Samuk, K. pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol. Ecol. Resour.* **21**, 1359–1368 (2021).
115. Gautier, M., Klassmann, A. & Vitalis, R. rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Mol. Ecol. Resour.* **17**, 78–90 (2017).
116. Pickrell, J. K. *et al.* Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* **19**, 826–837 (2009).
117. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

Acknowledgments

The authors acknowledge the ISO 9001-certified IRD i-Trop HPC (member of the South Green Platform) at IRD Montpellier for providing HPC resources that contributed to the research results reported in this paper. URL: https://bioinfo.ird.fr/- http://www.southgreen.fr. The bioinformatics analyses were also performed on the Core Cluster of the Institut Français de Bioinformatique (IFB) (ANR-11-INBS-0013). This work was funded by the French Agence Nationale de la Recherche (ANR) in the framework of the LabEx CEBA (ANR-10-LABX-25-01), JCJC GENAD (ANR-20-CE35-0003) and MICETRAL (ANR-19-CE350010). CGR, AL and SR thank the United States (U.S.) Agency for International Development and the U.S. Centers for Disease Control and Prevention (CDC) for funding the project entitled "Evaluation of ZIKV potential to establish a sylvatic transmission cycle in Colombia". We thank Olivier Duron for facilitating the connection with Benoit De Thoisy, who provided the samples from French Guiana. We would like to thank Josquin Daron for providing the scripts for sample mapping and calling. We also thank Leo Speidel for his guidance on *Relate,* and Dang Liu and Etienne Patin for providing the scripts to convert results from *Relate* to *MSMC-IM*. Finally, we thank Elisabetta Andermarcher for proofreading the manuscript and supplementary material.

Author Contributions

Conception: V.R., F.P., M.C.F. and M.J.M.L.; funding acquisition: V.R., M.C.F., F.P.; biological data acquisition and management: V.R., B.DT., C.G., S.R., A.L., A.C., J.A.B., A.S.C., E.dS.,R.H., C.M.DdS., E.V.; sequence data acquisition: F.D. and C.A.; method development and data analysis: M.J.M.L., V.R., F.P., and M.C.F.; interpretation of the results: V.R., F.P., M.C.F., and M.J.M.L.; drafting of the manuscript: M.J.M.L., V.R., M.C.F., and F.P.; and reviewing and editing of the manuscript: V.R., F.P., M.C.F., and M.J.M.L. with the contribution from all the co-authors.