# Population genomics of *Plasmodium ovale* species in sub-Saharan Africa

Kelly Carey-Ewend[1], Zachary R. Popkin-Hall[2], Alfred Simkin[3], Meredith Muller[2], Chris Hennelly[2], Wenqiao He[2], Kara A. Moser[2], Claudia Gaither[2], Karamoko Niaré[3], Farhang Aghakanian[2], Sindew Feleke[4], Bokretsion G. Brhane[4], Fernandine Phanzu[5], Kashamuka Mwandagalirwa[6], Ozkan Aydemir[7], Colin Sutherland[8], Deus S. Ishengoma[9,10], Innocent M. Ali[11], Billy Ngasala[12], Albert Kalonji[5], Antoinette Tshefu[6], Jonathan Parr[2,13,14], Jeffrey A. Bailey[3], Jonathan J. Juliano[1,2,13,14,#], Jessica T. Lin[2,13,15,#]

[1]: Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA
[2]: Institute for Global Health and Infectious Diseases, University of North Carolina, Chapel Hill, NC, USA
[3]: Department of Pathology and Laboratory Medicine, Brown University, Providence, RI, USA
[4]: Ethiopian Public Health Institute, Addis Ababa, Ethiopia
[5]: SANRU Asbl, Kinshasa, DRC
[6]: Kinshasa School of Public Health, Kinshasa, DRC
[7]: Program in Molecular Medicine, Chan Medical School, University of Massachusetts, Worcester, MA, USA
[8]: London School of Hygiene and Tropical Medicine, London, UK
[9]: National Institute for Medical Research (NIMR), Dar es Salaam, Tanzania
[10]: Department of Biochemistry, Kampala International University in Tanzania, Dar es Salaam, Tanzania
[11]: Department of Biochemistry, Faculty of Science, University of Dschang, Cameroon
[12]: Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania
[13]: Division of Infectious Diseases, University of North Carolina School of Medicine, University of North Carolina, Chapel Hill, NC, USA
[14]: Curriculum in Genetics and Molecular Biology, University of North Carolina School of Medicine, University of North Carolina, Chapel Hill, NC, USA
[15]: Department of Microbiology and Immunology, University of North Carolina School of Medicine, University of North Carolina, Chapel Hill, NC, USA

#: Co-senior authors

## Abstract

*Plasmodium ovale curtisi* (*Poc)* and *Plasmodium ovale wallikeri* (*Pow*) are relapsing malaria parasites endemic to Africa and Asia that were previously thought to represent a single species. Amid increasing detection of ovale malaria in sub-Saharan Africa, we performed a population genomic study of both species across the continent. We conducted whole-genome sequencing of 25 isolates from central and east Africa and analyzed them alongside four previously published west and central African genomes. Isolates were predominantly monoclonal (27/29), with their genetic similarity aligning with geography. *Pow* showed lower average nucleotide diversity ($1.9 \times 10^{-4}$) across the genome compared to *Poc* ($2.8 \times 10^{-4}$) ($p < 0.0001$). Signatures of selective sweeps involving the dihydrofolate reductase gene were found in both species, as were signs of balancing selection at the merozoite surface protein 1 gene. Differences in the nucleotide diversity of *Poc* and *Pow* may reflect unique demographic history, even as similar selective forces facilitate their resilience to malaria control interventions.

## Introduction

Parasites in the genus *Plasmodium* were responsible for an estimated 249 million cases of malaria and 608,000 deaths in 2022.[1] Ninety-four percent of these cases occurred in the World Health Organization Africa Region, where control efforts have primarily focused on the predominant species, *P. falciparum* (*Pf*).[2] Yet these case counts likely underrepresent the burden of non-falciparum species, which may be rising in prevalence even where control efforts have successfully reduced *P. falciparum* transmission.[3–5] Over the last few decades, genomic studies of *P. falciparum* have enabled monitoring of drug resistance markers,[6] facilitated the identification of promising vaccine candidates,[7] uncovered the structure of parasite populations,[8] and identified evolutionary forces shaping their demography.[9,10] Much less is known about non-falciparum species, especially their comparative evolutionary history and susceptibility to malaria control interventions focused on *P. falciparum*.

*Plasmodium ovale* was first identified as a separate malaria species in 1922 based on the appearance of oval-shaped erythrocytes that contained non-ring parasite forms.[11] Hallmarks of this parasite species are its restriction to younger red cells and therefore propensity to cause low-density infections, as well as relapses from liver hypnozoites, similar to *P. vivax* and *P. cynomologi*. The species often causes co-infection alongside *P. falciparum* which, along with its low parasite densities, makes it challenging to differentiate morphologically on peripheral blood smears.[12] The advent of polymerase chain reaction (PCR)-based diagnostics has improved detection of ovale infections, but initial PCR surveys across Africa and Asia based on the small subunit rRNA gene revealed two apparent groups of *P. ovale* parasites, termed classic and variant.[13] The discovery of perfect sequence segregation of 6 genomic markers, and more recently 12 mitochondrial loci, between classic and variant *P. ovale* isolates collected across Africa and Asia has led to the conclusion that this dimorphism actually represents a true species divide in the *P. ovale* clade.[14,15] The nomenclature of these species is currently evolving but will

be referred to as *P. ovale curtisi* (*Poc*, formerly classic) and *P. ovale wallikeri* (*Pow*, formerly variant) herein.

*P. ovale curtisi and P. ovale wallikeri* have since been confirmed to circulate within the same human populations throughout Africa and Asia,[16,17] with both detected by PCR at higher rates than previously appreciated.[5,12,18,19] Limited investigation of the genetic diversity and population genetics of the two *P. ovale* species have hinted at low diversity and/or small effective population size, as few unique haplotypes have been identified at antigenic gene targets like apical membrane antigen 1 (*ama1*) and merozoite surface protein 1 (*msp1*).[17,20] There is some indication that drugs used to treat *P. falciparum* are also shaping *P. ovale* parasite populations; signs of a selective sweep involving a mutant *dhfr* allele (implicated in pyrimethamine resistance) were detected in *Poc* cases imported to China from Africa, though not in *Pow.*[21] Until now, the low density of most *P. ovale* isolates combined with lack of an *in vitro* culture system has hindered whole-genome sequencing of these parasites.[11] However, with development of strategies for parasite DNA enrichment, as well as construction of the first reference genomes in 2017, genome-wide analyses are now possible.[15,22–24]

We employed hybrid capture or leukodepletion to enrich *P. ovale spp.* DNA and perform whole-genome sequencing (WGS) of 25 clinical isolates collected from studies conducted across Ethiopia, the Democratic Republic of the Congo, Tanzania, and Cameroon. Combined with four additional public whole genomes from Senegal, Cameroon, and the Ivory Coast, we sought to better understand the comparative biology of *P. ovale curtisi*, *P. ovale wallikeri,* and co-endemic *P. falciparum* by examining their complexity of infection, population structure, nucleotide diversity, and genomic signatures of selection.

# Results

## High-quality genomic coverage of African *P. ovale* isolates

Parasite samples from 25 *P. ovale*-infected individuals collected at ten sites spanning Ethiopia, the Democratic Republic of the Congo (DRC), Tanzania, and Cameroon were selected for whole-genome sequencing (**Table 1**). These included 13 *P. ovale curtisi* and 12 *P. ovale wallikeri* isolates that were selected from six studies based on robust amplification of the *po18S* rRNA gene (Ct <36) and predominance of one *ovale* species within each isolate. The majority of the samples (n=21) underwent a custom-designed hybrid capture with RNA baits to preferentially isolate ovale DNA extracted from dried blood spots for sequencing, while four additional whole blood samples were leukodepleted (LDB) at the time of collection by CF11 filtration and directly sequenced without enrichment.[25] Finally, genomic data from four *P. ovale* cases imported to France from west and central Africa were retrieved from the European Nucleotide Archive.[26] These isolates had undergone selective whole-genome amplification (sWGA) for parasite DNA enrichment prior to sequencing. Further data on all parasite isolates are found in **Supplemental Table 1**.[27–32]

| Study | Country of Origin | Year of Collection | Study Population | # selected for sequencing (# of *Poc* and *Pow*) |
|---|---|---|---|---|
| *hrp*2/3 Deletion Survey[27] | Ethiopia | 2017-2018 | Febrile patients presenting to health facilities in the Amhara, Tigray, and Gambella regions | 7 (2 *Poc*, 5 *Pow*) |
| SANRU Rural Health Program[28] | Democratic Republic of the Congo | 2017 | Febrile patients presenting to health facilities in Sud-Kivu, Bas-Uele, and Kinshasa Provinces. | 6 (2 *Poc*, 4 *Pow*) |
| Kinshasa Malaria Longitudinal Study[29] | Democratic Republic of the Congo | 2015-2017 | Members of households participating in longitudinal study of malaria. | 5 (5 *Poc*) |
| TranSMIT[30] | Tanzania (East) | 2018-2022 | Asymptomatic children and adults attending school or health clinics in rural Bagamoyo district, eastern Tanzania | 4 (3 *Poc,* 1 *Pow*) |
| MSMT21[31] | Tanzania (West) | 2020-2022 | Tanzanian citizens at health facilities | 2 (2 *Pow*) |
| Dschang Febrile Cohort[32] | Cameroon | 2020-2021 | Febrile patients presenting to health facilities in western Cameroon | 1 (1 *Poc*) |
| Joste et al., JID 2023, France[26]* | Cameroon, Senegal, Ivory Coast | 2013-2021 | *P. ovale* infections identified in France after travel to an endemic country. | 4 (4 *Pow*) |

**Table 1. Studies of origin for 29 *P. ovale* isolates.**
*Raw sequencing data for these 4 isolates were directly incorporated into the analysis pipeline after retrieval from the European Nucleotide Archive.

Whole-genome sequencing achieved high genome coverage, with an overall average of 87% and 90% ten-fold coverage across the core genome for the 13 *P. ovale curtisi* and 16 *P. ovale wallikeri* isolates, respectively (**Figure 1**). Coverage and mapping proportion were highest when aligned to the *P. ovale* reference genome determined by the *Poc*/*Pow* species-specific qPCR assay, corroborating initial species assignment. Compared to sWGA and LDB samples, the hybrid capture method used to enrich parasite DNA in the majority of samples yielded more complete coverage across all chromosomes except for chromosome 10 (**Figure 1**). The hybrid capture was originally designed for *P. ovale wallikeri*, with additional *P. ovale curtisi* baits then selected to cover areas that differ between the two ovale genomes. Due to the smaller sized *Pow* chromosome 10 (470kb), this approach did not provide coverage for the full *Poc* chromosome 10 (1,300kb). This led to substantially lower coverage for chromosome 10 across all *Poc* hybrid capture isolates (40-60% 10x coverage vs. >85% for all other *Poc* chromosomes); thus, chromosome 10 was excluded from all genome-wide analysis in *Poc* isolates to limit error.

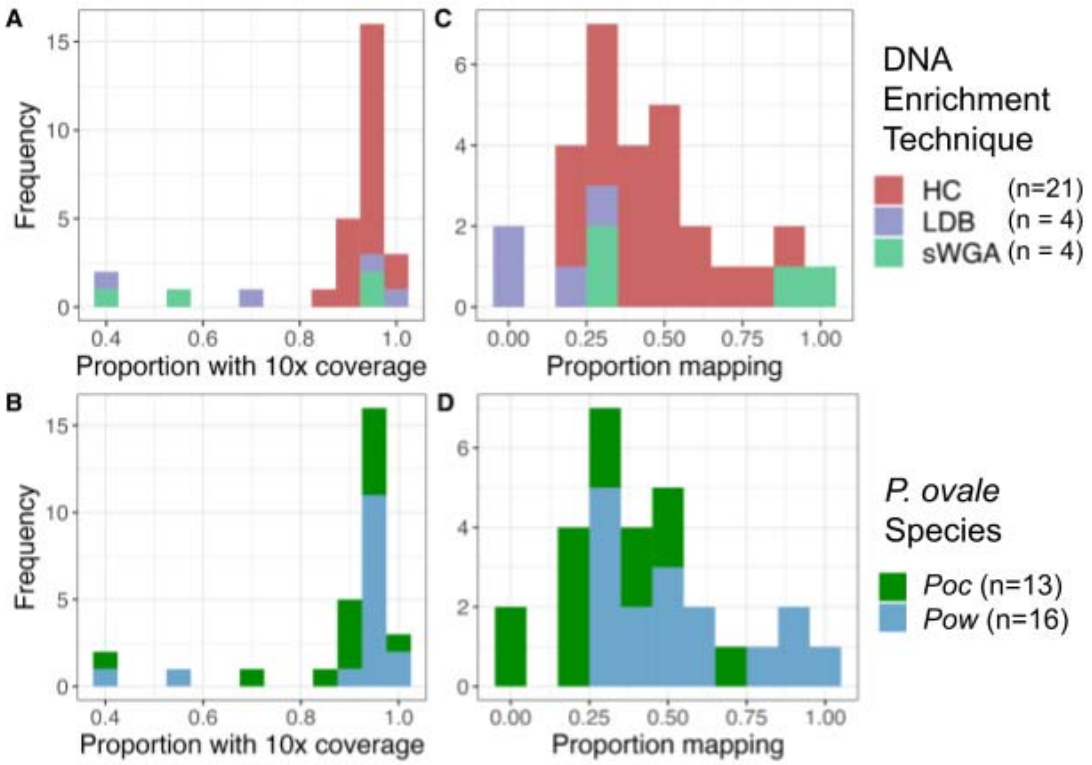


**Figure 1. Coverage and mapping among 29 isolates.** Proportion of the corresponding *P. ovale* reference genome covered by ≥10 reads by DNA enrichment technique (A) and species (B). Proportion of reads mapped to that reference genome by DNA enrichment technique (C) and species (D). For *Poc*, chromosome 10 was excluded due to incomplete coverage by hybrid capture baits. HC = hybrid capture; LDB = leukodepleted blood; sWGA = selective whole-genome amplification

As, expected, hybrid capture led to preferential sequencing of *P. ovale* DNA among samples that were co-infected with *P. falciparum* (*Pf*); *Pf*-positive isolates that underwent hybrid capture yielded only 2-11% 10x coverage of the *Pf* genome compared to >90% 10x coverage of the *Pf* genome among leukodepleted blood samples. For genomic analysis, insertions/deletions, multiallelic sites, low-quality variants, and variants within tandem repeats and expanded gene families were excluded (see **Methods**), yielding final biallelic single nucleotide polymorphism (SNP) call sets of 65,073 SNPs for *P. ovale curtisi* and 39,116 for *P. ovale wallikeri*.

## Low complexity of infection

Complexity of infection (COI), or the number of unique parasite clones present in a given isolate, was estimated 1000 times using THEREALMcCOIL for all 13 *P. ovale curtisi* and 15 *ovale wallikeri* isolates, as well as 842 geographically matched *P. falciparum* isolates downloaded from the publicly available MalariaGEN Pf6 dataset (**Figure 2**).[33] Twelve out of 13 *Poc* isolates and 15 out of 16 *Pow* isolates were estimated to be monoclonal; the remaining isolate in each *P. ovale* species was found to comprise two parasite clones. By comparison, only 539 out of 842 *P. falciparum* samples (64%) were monoclonal. The distribution of COI among the three *Plasmodium* species was significantly different ($p = 0.005$). In pairwise comparisons, both *Poc and Pow* had lower COI compared to *P. falciparum*, though the latter comparison was not quite significant ($p = 0.03$ and $p=0.07$, respectively).

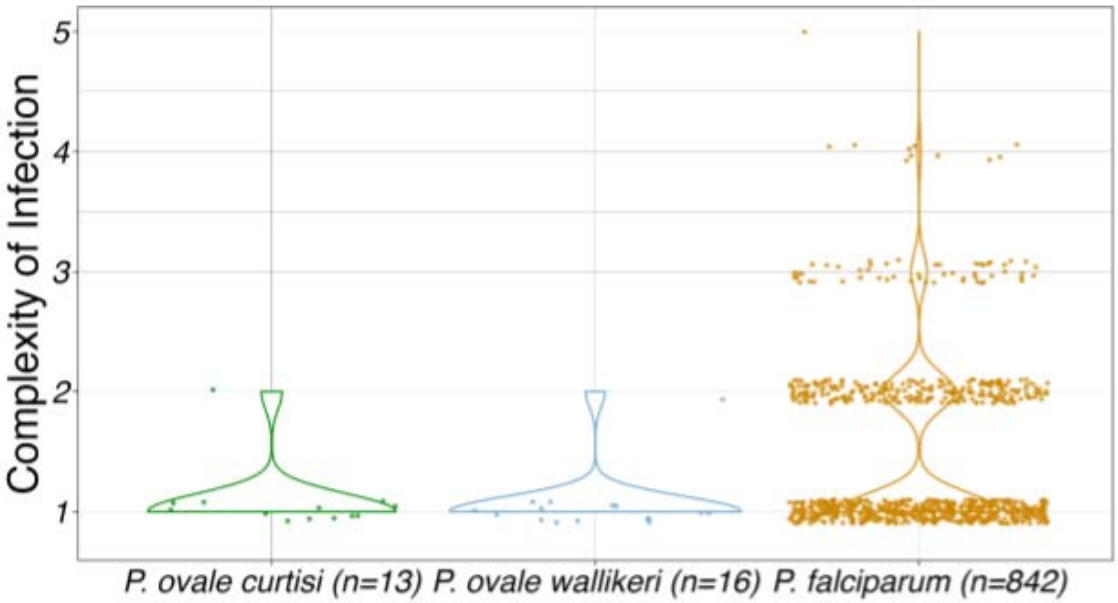


**Figure 2. Median estimated complexity of infection** (COI) among 13 *Poc* isolates, 16 *Pow* isolates, and 842 *P. falciparum* isolates geographically matched to the *P. ovale* samples by country of origin. The distributions of COI were different among the three species ($p < 0.01$) with *Pow* showing significantly lower COI than *P. falciparum* ($p = 0.03$).

The two multiclonal *P. ovale* infections each demonstrated a COI of 2 and originated from the DRC (from Kinshasa and Bas-Uele Provinces), where malaria transmission remains high.[34] In order to investigate if the two clones in these samples were distinct lineages or potentially related meiotic siblings, we investigated the distribution of heterozygous SNPs across the genome. We assumed that meiotic siblings would show restriction of heterozygous SNPs to specific regions reflecting recombination within the mosquito midgut.[35] In both samples, after filtering to high confidence SNPs based on population-wide allele frequency, we saw an even distribution of heterozygous SNPs across the genome, suggestive of two distinct parasite lineages in the same host (**Supplemental Figure 1**).

## Different nucleotide diversity and patterns of mutation in *P. ovale curtisi* and *P. ovale wallikeri*

Among a collection of 3,339 sets of one-to-one-to-one orthologous genes between the *Poc*, *Pow*, and *P. falciparum* genomes, we identified 2,003 sets that achieved high-quality sequencing coverage and had no overlap with masked genomic regions in any of the three species. The average species-specific nucleotide diversity (□) among these orthologues in the 12 monoclonal *Poc*, 15 monoclonal *Pow*, and 15 geographically-matched monoclonal *Pf* samples were $3.8 \times 10^{-4}$, $1.9 \times 10^{-4}$, and $2.6 \times 10^{-4}$, respectively. These were significantly different between species ($p < 0.0001$, $F = 173$, $df = 2$), with orthologues in *P. ovale curtisi* the most diverse among the three species and *P. ovale wallikeri* orthologues the least diverse (p-values < $10^{-4}$) (**Figure 3A**). To mitigate bias by geographic coverage and orthology with the *P. falciparum* genome, we repeated this analysis using 2,867 *Poc-Pow* orthologues among a group of

geographically-matched monoclonal *Poc* and *Pow* samples (n=8 each, **Supplemental Table 3**), revealing average nucleotide diversities of $2.8 \times 10^{-4}$ and $1.9 \times 10^{-4}$, respectively (**Figure 3B**). Nucleotide diversity was still significantly lower in *Pow* orthologues compared to *Poc* ($p < 0.0001$).

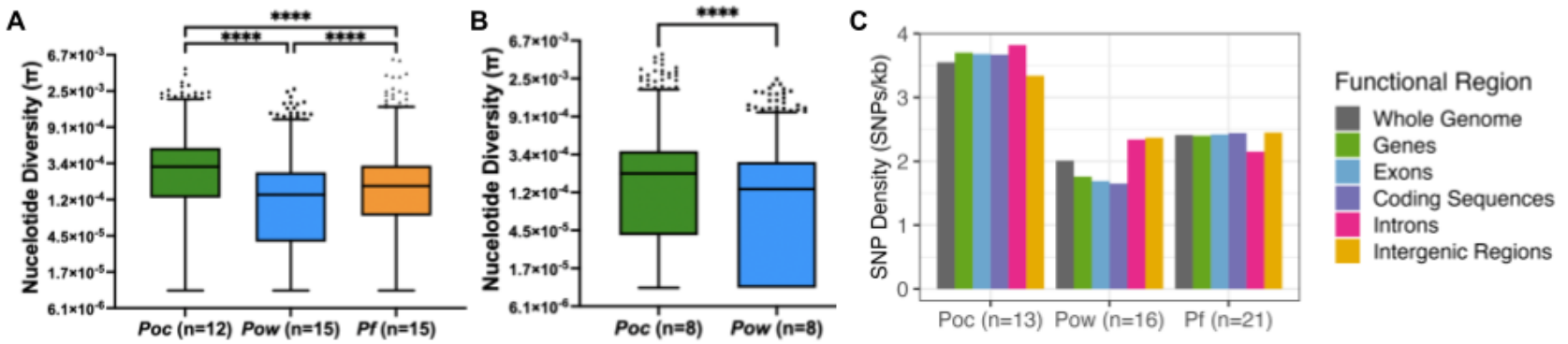


**Figure 3. Nucleotide diversity (π) of orthologous genes and SNP density by functional genomic region among *Poc*, *Pow*, and *P. falciparum* (*Pf*) isolates.** (A) Nucleotide diversity (π) per gene among 2,003 sets of orthologous genes in monoclonal *Poc*, *Pow*, and *Pf* samples. Boxes denote 25th, median, and 75th percentiles; whiskers drawn at 1st and 99th percentiles. π of 0 was coded as $1x10^{-5}$ to plot on logarithmic scale. Nucleotide diversity was significantly different between orthologues of all three species by Tukey's multiple comparisons tests, with *Poc* orthologues showing higher diversity than orthologues of *Pow* and *P. falciparum*, and *Pow* orthologues also showing lower diversity than those in *P. falciparum* (p-values < 0.0001). (B) Nucleotide diversity (π) per gene among 2,867 sets of orthologous genes in geographically-matched monoclonal *Poc* and *Pow* samples. Nucleotide diversity was significantly lower among *Pow* orthologues compared to *Poc* (p < 0.0001). (C) SNP density in different functional regions of the genome among all *Pow*, *Poc*, and *P. falciparum* isolates. SNP = single nucleotide polymorphism; kb = kilobase.

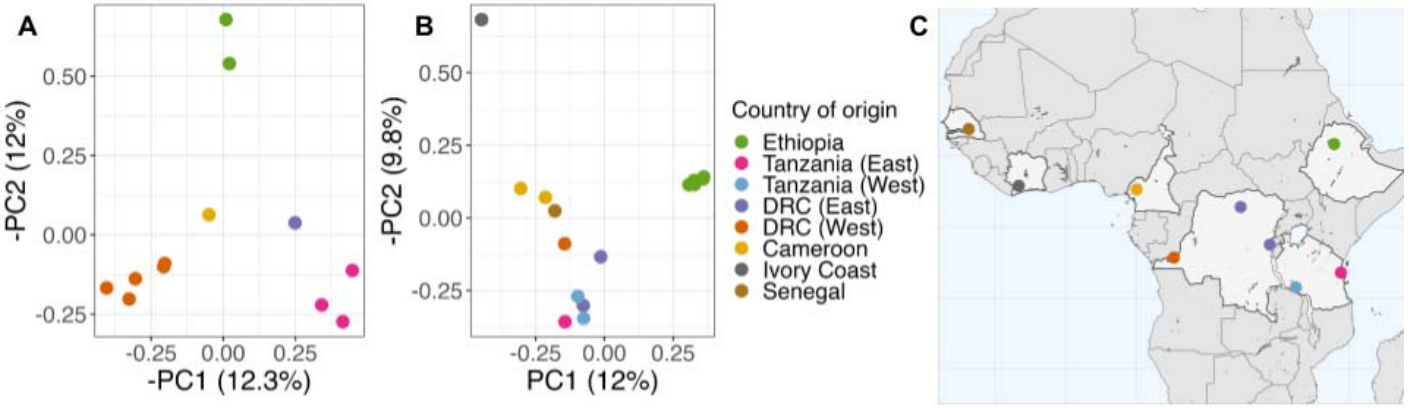


**Figure 4. Principal component analysis** showing the first two principal components among 12 monoclonal *Poc* isolates (A) and 15 monoclonal *Pow* isolates (B) using 2,984 and 3,609 biallelic SNPs, respectively. Samples colored by region of country of origin (C). Axis orientation in this analysis is arbitrary, and some components were inverted to reflect alignment with geography (represented by "-PC").

This high nucleotide diversity in *P. ovale curtisi* was corroborated by investigation of the total number and density of genome-wide SNPs. Variant calling and filtering resulted in roughly twice the number of SNPs for *P. ovale curtisi* (65,073) as for *P. ovale wallikeri* (39,116), despite a slightly smaller number of *Poc* isolates. This corresponded to a higher density of SNPs across the *Poc* genome (3.6 SNPs per kilobase in *Poc* vs. 2.0 SNPs/kilobase in *Pow*). However, a smaller proportion of SNPs in the *Poc* genome were nonsynonymous mutations, as the ratio of nonsynonymous-to-synonymous (dN/dS) mutations was 1.5 and 2.8 in *Poc* and *Pow*, respectively. Among the aforementioned geographically-matched *P. ovale* isolates, dN/dS within 2,867 orthologous genes were 1.4 for *Poc* and 2.7 for *Pow*, consistent with the broader estimates. SNP density in the *Poc* genome was relatively lower in protein-coding sequences (3.7 SNPs/kilobase[kb]), with the lowest SNP density unexpectedly found in intergenic regions of the genome (3.3 SNPs/kb) (**Figure 3C**). In contrast, *Pow* exhibited the highest SNP density in introns and intergenic regions (2.3 and 2.4 SNPs/kb, respectively), while also exhibiting the lowest SNP density in protein-coding sequences (1.7 SNPs/kb). Geographically-matched *P. falciparum* samples showed a SNP distribution more similar to that of *Pow* than *Poc*.

## Parasite genomic similarity recapitulates geographic relationships

Genome-wide principal component (PC) analysis of the monoclonal samples of each ovale species revealed spatial arrangement of related parasites along PC1 and PC2 that mirrors location of origin (**Figure 4**). These components accounted for 24.3% and 21.8% of the genetic differentiation in *Poc* and *Pow* respectively. Formal analysis by ADMIXTURE did not reveal significant clusters of samples in either species, likely limited by the sample size. However, in both species, Ethiopian parasites appeared to form a separate group from other samples along one principal component (PC2 for *Poc* and PC1 for *Pow*). The remaining parasite isolates roughly align along an East-West axis in the other principal component. Tanzanian samples were most similar to samples collected from eastern DRC (Sud-Kivu), while western DRC samples (Kinshasa) were more genetically similar to samples from Cameroon and West African nations (Ivory Coast, Senegal). In *P. ovale curtisi*, PC3 and PC4 separated samples from Cameroon, the eastern DRC, and Tanzania from others. In *P. ovale wallikeri*, PC3 and PC4 separated samples from the western DRC, Ivory Coast, and Cameroon from the others (**Supplemental Figure 2**). Examination of the top 0.5% of variants by contribution to each of the first 4 principal components did not reveal orthologs of known drug resistance markers in either species.

## Signatures of selection implicate putative drug resistance loci, proteins involved in sexual stage differentiation, and antigenic targets

We calculated $nS_L$ and Tajima's D across the genomes to identify loci under directional and balancing selection, respectively. The $nS_L$ statistic is considered robust to the currently-unknown recombination rates across the genomes of *P. ovale* species. Genetic markers of interest within 10kb of the top 0.5% absolute normalized $nS_L$ values that may be involved in selective sweeps are listed in **Table 2** and **Table 3**. Evidence of a selective sweep involving the putative bifunctional dihydrofolate reductase - thymidylate synthase (*ts-dhfr*) gene, a key gene in folate metabolism that is implicated in pyrimethamine resistance,[36,37] was found in both *P. ovale* species (**Figure 5A, 6A**). Examination of extended haplotype homozygosity (EHH) at the

selected variants show a large selective sweep in *Pow* spanning roughly 40kb as well as close proximity of the *ts-dhfr* gene to the focal variant (**Figure 6B**). In *Poc*, the positioning of the *ts-dhfr* gene lies at the edge of a smaller sweep (**Figure 5B**). However, another marker of drug resistance, multi-drug resistance associated protein 2 (*mrp2*), was found in close proximity to the highest absolute normalized $nS_L$ value in *Poc* and lies near the center of a 40kb sweep region on *Poc* chromosome 14 (**Figure 5C**).

| Statistic | Value | Chr. | Location | Closest plausible genetic driver | Distance | Gene ID |
|---|---|---|---|---|---|---|
| **$nS_L$** | -2.84 | 3 | 541949 | cysteine repeat modular protein 2, putative (*crmp2*) | 6695 | PocGH01_03022100 |
| | -2.99 | 5 | 774030 | bifunctional dihydrofolate reductase-thymidylate synthase, putative (*ts-dhfr*) | -9660 | PocGH01_05028400 |
| | -2.85 | 5 | 936176 | AP2 domain transcription factor, putative (*apiap2*) | 0 | PocGH01_05031600 |
| | -3.81 | 7 | 1217498 | merozoite surface protein, putative (*msp*) | 0 | PocGH01_07039000 |
| | -3.12 | 8 | 987227 | male development protein, putative (*md3*) | 0 | PocGH01_08032900 |
| | -2.98 | 12 | 442462 | AP2 domain transcription factor, putative (*ap2-hs*) | 0 | PocGH01_12020200 |
| | -2.81 | 12 | 703579 | merozoite surface protein 7-like protein, putative (*msp7*) | 1983 | PocGH01_12027700 |
| | -3.75 | 12 | 2760110 | male development protein, putative (*md1*) | 0 | PocGH01_12076000 |
| | -2.85 | 14 | 1579830 | AP2 domain transcription factor, putative (*ap2-g*) | -3531 | PocGH01_14048100 |
| | -4.71 | 14 | 1899110 | ABC transporter C family member 2, putative (*mrp2*) | 7715 | PocGH01_14054800 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | -3.88 | 14 | 2435520 | female development protein, putative (*fd1*) | 0 | PocGH01_14066000 |
| **Tajima's D** | 2.59 | 7 | 1152840 | merozoite surface protein 1, putative (*msp1*) | 0 | PocGH01_07037900 |

**Table 2. Selection statistics and nearby genetic markers for loci with top normalized $nS_L$ and positive Tajima's D values among monoclonal isolates of *Poc*.** Loci are described by chromosome (chr.), location on the chromosome in base pairs, and a statistical value in the top 0.5% across the genome. Nearby genetic markers were identified within 10,000 base pairs of these loci and are given alongside their gene ID and the distance of this marker to the reported locus (negative distance indicates upstream location).

| Statistic | Value | Chr. | Location | Closest plausible genetic driver | Distance | Gene ID |
|---|---|---|---|---|---|---|
| $nS_L$ | -4.43 | 2 | 537054 | dynein heavy chain, putative (*dhc*) | 0 | POWCR01_020017200 |
| | -3.91 | 2 | 554712 | transcription factor with AP2 domain(s) (*apiap2*) | 7318 | POWCR01_020017600 |
| | -3.88 | 4 | 603847 | merozoite surface protein 5, putative (*msp5*) | -5083 | POWCR01_040018700 |
| | -3.89 | 5 | 846693 | bifunctional dihydrofolate reductase-thymidylate synthase, putative (*ts-dhfr*) | -953 | POWCR01_050023500 |
| | 4.01 | 5 | 1030326 | transcription factor with AP2 domain(s), putative (*apiap2*) | 0 | POWCR01_050026700 |
| | -4.06 | 9 | 789811 | dynein heavy chain, putative (*dhc*) | 0 | POWCR01_090024400 |
| | -4.02 | 13 | 174672 | *Plasmodium* interspersed repeat protein (*pir*) | -2722 | POWCR01_130008000 |
| | -4.26 | 13 | 180100 | early transcribed membrane protein (*etramp*) | 2975 | POWCR01_130008100 |
| | 3.92 | 13 | 1393497 | surface protein *p113*, putative | 6651 | POWCR01_130035600 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Tajima's D** | 2.45 | 1 | 31900 | conserved Plasmodium protein, unknown function | 0 | POWCR01_010005700 |
| | 3.26 | 7 | 1184250 | merozoite surface protein 1, putative (*msp1*) | 0 | POWCR01_070032200 |
| | 2.13 | 8 | 1378470 | conserved Plasmodium protein, unknown function | 0 | POWCR01_080035700 |
| | 2.23 | 12 | 235570 | stromal-processing peptidase, putative | 0 | POWCR01_120010700 |
| | 2.81 | 13 | 189760 | XPA binding protein 1, putative | 0 | POWCR01_130008300 |
| | 2.28 | 13 | 198250 | cytochrome c oxidase assembly protein COX19, putative | 0 | POWCR01_130008600 |

**Table 3. Selection statistics and nearby genetic markers for loci with top normalized $nS_L$ and positive Tajima's D values among monoclonal isolates of *Pow.*** Loci are described by chromosome (chr.), location on the chromosome in base pairs, and a statistical value in the top 0.5% of absolute values across the genome. Nearby genetic markers were identified within 10,000 base pairs of these loci and are given alongside their gene ID and the distance of this marker to the reported locus (negative distance indicates upstream location).

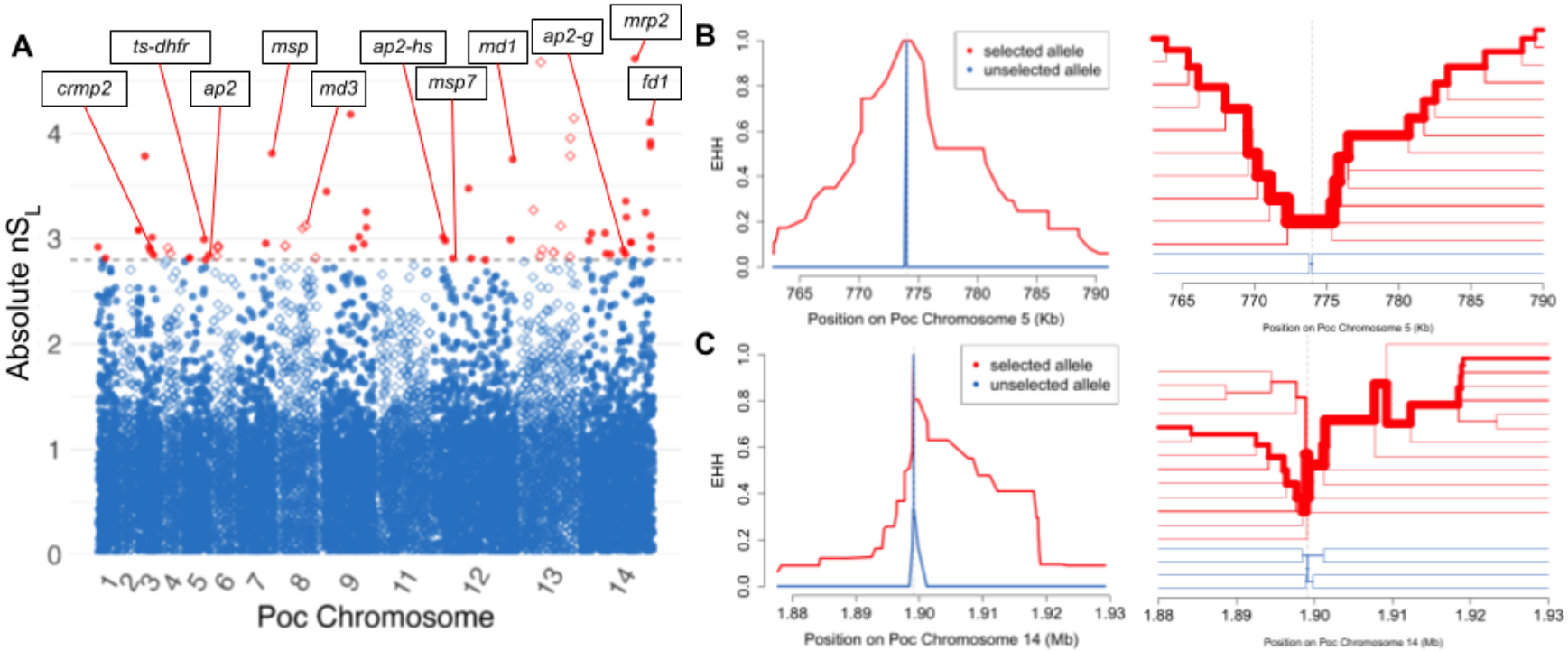


**Figure 5. Absolute $nS_L$ of 14,019 variants among monoclonal isolates across the *Poc* genome (A) and extended haplotype homozygosity (EHH) near select loci (B, C).** Individual loci are depicted using alternating shapes between chromosomes for legibility. Dotted line and red color denote the top 0.5% of loci. Putative genetic markers of note within 10,000 bases of these loci are labelled, including cysteine repeat modular protein 2 (*crmp2*), thymidylate synthase - dihydrofolate reductase (*ts-dhfr*), AP2 domain transcription factors (*ap2*, including *ap2-hs* and *ap2-g*), merozoite surface protein (*msp*), male development protein 3 (*md3*), merozoite surface protein 7-like protein (*msp7*), male development protein 1 (*md1*), multi-drug resistance associated protein 2 (*mrp2*), and female development protein 1 (*fd1*). EHH and haplotype bifurcation at a variant near the *ts-dhfr* gene (B) indicates a selective sweep spanning ~20kb, with lineage breakdown occurring first among the unselected allele haplotypes (blue) and then in the selected allele haplotypes (red) as distance from the focal variant increases. EHH and haplotype bifurcation at a variant near the *mrp2* locus (C) similarly show a sweep spanning ~40kb.

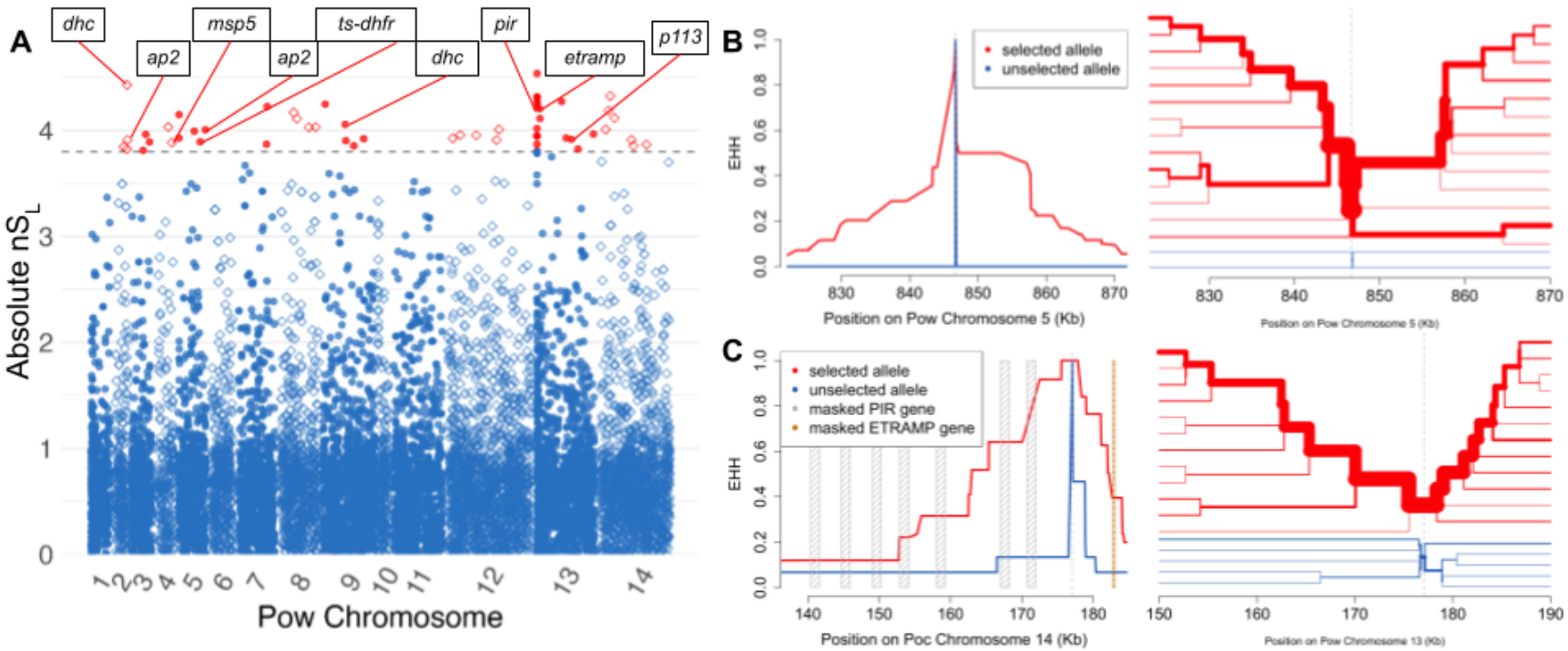


**Figure 6. Absolute $nS_L$ of 11,967 variants among monoclonal isolates across the *Pow* genome (A) and extended haplotype homozygosity (EHH) near select loci (B, C).** Individual loci are depicted using alternating shapes between chromosomes for legibility. Dotted line and red color denote the top 0.5% of loci. Putative genetic markers of note within 10,000 bases of these loci are labelled, including dynein heavy chain (*dhc*), AP2 domain transcription factors (*ap2*), merozoite surface protein 5 (*msp5*), thymidylate synthase - dihydrofolate reductase (*ts-dhfr*), *Plasmodium* interspersed repeat protein (*pir*), early transcribed membrane protein (*etramp*), and surface protein *p113*. EHH and haplotype bifurcation at a variant near the *ts-dhfr* gene (B) indicates a selective sweep spanning ~40kb, with lineage breakdown occurring first among the unselected allele haplotypes (blue) and then in the selected allele haplotypes (red) as distance from the focal variant increases. EHH and haplotype bifurcation at a variant near the both *pir* and *etramp* genes (C) similarly show a sweep spanning ~40kb. Of note, variants within the *pir* and *etramp* genes themselves (grey and orange cross-checked boxes, respectively) were excluded from analysis and create falsely extended homozygosity in both selected and unselected allele haplotypes within those particular intervals. EHH in the intervening regions still indicates directional selection.

Top absolute $nS_L$ hits were also found near *ap2* transcription factor genes that regulate apicomplexan life cycle transitions, including sexual differentiation into gametocytes (*ap2-g*), and genes involved in sex-specific development of gametes, such as those coding male development proteins 1 and 3 (*md1/3*) and female development protein 1 (*fd1*).[38–40] In *Pow*, three top $nS_L$ hits were found around genes encoding the dynein heavy chain, a cytoskeleton component highly expressed in male gametes for motility and fusion with female gametes in the mosquito blood meal.[41] And, in *Poc*, a hit near *crmp2* implicates a protein likely involved in targeting sporozoites to the salivary glands in the mosquito prior to transmission.[42]

Finally, genes encoding putative antigenic targets at the host-parasite interface, including merozoite surface protein 7 (*msp7*), merozoite surface protein 5 (*msp5*), early transcribed membrane protein (*etramp*), and surface protein *p113*, a merozoite surface protein that binds RH5, a major vaccine candidate,[43] may be under directional selection in both *P. ovale* species. A cluster of 14 top absolute $nS_L$ hits indicate a selective sweep on chromosome 13 of the *Pow* genome between *etramp* and *pir* genes, two complex multigene families involved in parasite virulence that were excluded during variant calling (**Figure 6C**).

Overall, Tajima's D in both species exhibited a negative skew across the genomic windows with an average value of -1.06 for *Poc* and -0.76 for *Pow*; this may suggest population expansion following a bottleneck or directional selection (**Figure 7**). Among the loci with positive values in the top 0.5% of absolute Tajima's D hits, the antigenic marker merozoite surface protein 1 (MSP1) was identified as a probable target of balancing or diversifying selection in both *P. ovale* species (**Tables 2, 3**).

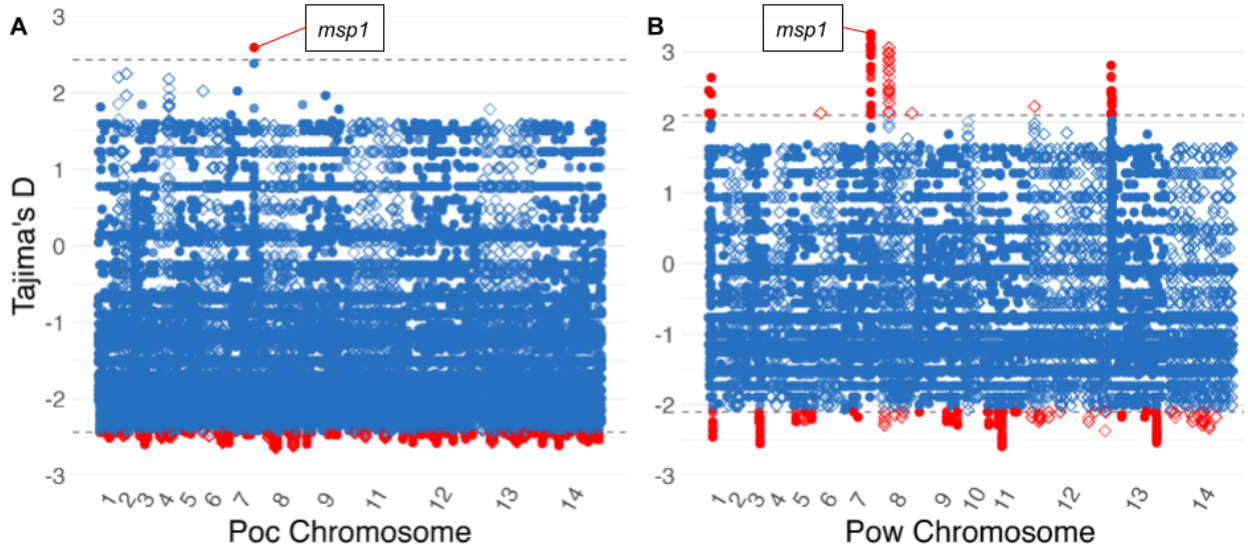


**Figure 7. Tajima's D in 350,155 and 258,065 300bp windows in genes across the *Poc* (A) and *Pow* (B) genomes among monoclonal isolates.** Individual windows are depicted using alternating shapes between chromosomes for legibility. Dotted line and red color denote the top 0.5% of loci. For both species, the top locus with a positive Tajima's D was located inside the merozoite surface protein 1 (*msp1*) gene.

## Discussion

We present the first comprehensive population genomics study of both *P. ovale* species from samples collected in sub-Saharan Africa. Our study comprises 13 *Poc* and 16 *Pow* isolates selected from seven studies, including both febrile and asymptomatic cases. Genome-wide analysis reveals differences in nucleotide diversity between *P. ovale* species, but similarity in their low complexity of infection, geographic relatedness, and signatures of selection. Our analysis was performed using the initial reference genomes from Ghana and Cameroon assembled for *P. ovale curtisi* and *P. ovale wallikeri*, respectively, in 2017.[24] Since the time of our analysis, new reference genomes have been assembled from South Sudan and Nigeria that provide improved genome contiguity, annotation, and completeness,[15] but are unlikely to change our main findings.

Compared to the "classic" *P. ovale curtisi* species, we observed significantly lower nucleotide diversity across orthologous genes among geographically-matched *P. ovale wallikeri* isolates ($2.8 \times 10^{-4}$ for *Poc* and $1.9 \times 10^{-4}$ for *Pow*, respectively). Our estimate for *Poc* is concordant with the genome-wide diversity calculated among six central African *Poc* isolates[15] as well as that derived from RNA expression data among four parasite samples from Mali.[44] However, our *Pow* estimate was substantially lower than the genome-wide estimate reported by Higgins et al. ($3.4 \times 10^{-4}$), who examined a similar number of isolates but with more representation of west Africa compared to east Africa in our collection. Our lower estimate may reflect the exclusion of higher-diversity intergenic regions, though we also found lower genome-wide and intergenic SNP density in *Pow* compared to *Poc.* Relatively low nucleotide diversity in *Pow* may indicate reduced effective population size, increased inbreeding, or a population bottleneck in the time since *Poc* and *Pow* diverged 1.3 million years ago.[14] More recent population growth is also likely in both species as indicated by the predominantly negative distribution of Tajima's D values across their protein-coding genes, a finding that can indicate population expansion following a bottleneck.[45] The high ratio of nonsynonymous-to-synonymous substitutions among these protein coding genes (2.7 for *Pow*, 1.3 for *Poc*) is similar to that seen in *P. falciparum* and *P. vivax.*[46,47] This finding may represent diversifying selection on proteins across either *P. ovale* genome, enabling maintenance of nonsynonymous substitutions. The impact of the malaria life cycle on allele frequencies has also been observed to inflate dN/dS ratios among *Plasmodium* parasites.[48] Further analysis of subpopulations of each parasite species could help to elucidate the factors driving the observed difference in genomic diversity, such as by examining *Pow* isolates from Asia to determine whether this low nucleotide diversity persists across global parasite populations.

The observed predominance of monoclonal isolates among both *P. ovale* species is consistent with low within-sample haplotypic diversity seen in previous investigations of African *P. ovale* isolates by genome-wide RNA sequencing and amplicon sequencing.[44,49] Low complexity of *P. ovale* infections may result from efficient clonal transmission[30] and/or lower transmission overall, limiting vector uptake of multiple parasite clones from either the same or different infected individuals. This low complexity is expected to limit opportunities for genetic

recombination within mosquito vectors, though multiple-clone infections were identified in Kinshasa, a region with overall higher malaria endemicity and transmission intensity.[34]

Genomic signatures of selection within both *P. ovale* species highlighted the importance of antimalarials, host-vector life cycle transitions, and human immunity as evolutionary pressures impacting parasite survival. *P. ovale* infections are frequently subclinical and go untreated,[50] but likely still face substantial drug exposure from widely used antimalarials prescribed for *P. falciparum.*[19] Additionally, malaria prophylaxis using sulfadoxine-pyrimethamine, such as intermittent preventive therapy in pregnancy (IPTp) and seasonal malaria chemoprevention for infants and schoolchildren (SMC), may be applying drug pressure on *P. ovale* parasite populations.[1] In 147 African isolates imported to China, amplicon sequencing revealed selective sweeps in the gene encoding binfunctional thymidylate synthase - dihydrofolate reductase (TS-DHFR) in *Poc*, a gene implicated in resistance to pyrimethamine, though not in *Pow*.[21] In our dataset, sweeps near the genes for TS-DHFR were among the strongest signals of directional selection in both *P. ovale* species, especially in *Pow*, possibly representing drug pressure influencing parasite survival. Functional evaluation of different alleles to determine their capacity to confer drug resistance, as well as monitoring of these alleles across the parasite populations over time, will be necessary to ensure that *P. ovale* do not develop resistance to prophylactic malaria treatment, especially given the contemporary rise in *P. ovale* prevalence observed in some African populations in recent decades.[3–5]

Finally, strong signals of balancing or diversifying selection were observed in both species within their genes encoding merozoite surface protein 1 (MSP1), the orthologues of the predominant antigen on blood-stage *P. falciparum* parasites that has been shown to induce protective immunity in some studies.[51] Diversifying selection on MSP1 has been documented in *P. falciparum*[52] as well as in focused analysis among African *P. ovale* infections imported to China.[53] Our dataset provides even stronger evidence for diversifying selection at this site, as the *msp1* gene showed the single highest Tajima's D value across all genes in both *P. ovale* species. Such immune responses may also play a role in modulating relapse potential.[54]

This study has several limitations. While it represents the largest genome-wide examination of the genomic composition of both *P. ovale curtisi* and *P. ovale wallikeri* to date, the sample size for both species is nonetheless small and limits the power to detect clustering of isolates, infer population demography, and detect selection. Additionally, the geographic coverage of the isolates employed differs between the six countries represented, with high-quality isolates of *Poc* missing from West Africa. Fortunately, this region is now represented in recently published genomes from travelers returning from Africa.[15] Whole-genome enrichment methods also differed among isolates; 21 isolates employed hybrid capture, four used selective whole-genome amplification, and four relied on leukodepletion. The two former methods may have induced amplification bias, whereas leukodepletion does not amplify *P. ovale* DNA and therefore may reduce the power to identify rare variants in those isolates. The source studies also differ by whether samples were derived from asymptomatic carriers (n=11) or febrile patients (n=18). Sample sizes were too small to analyze these populations separately. Finally, Poc chromosome 10 was excluded from analyses due to poor coverage of hybrid capture baits.

We do not expect this exclusion to systematically bias estimation of nucleotide diversity nor complexity of infection, though it does prevent us from evaluating chr. 10 loci for genomic signatures of selection.

This analysis sheds new light on the evolutionary history of the two *Plasmodium ovale* species sympatrically circulating in sub-Saharan Africa while indicating selective pressures on genes related to sexual differentiation, drug response, and immune evasion. Further population genomic studies of *Poc* and *Pow* should employ a larger selection of isolates from a greater geographic range, especially including Asia, and take advantage of new reference genome assemblies to build on these insights. Functional investigation into the genes showing signatures of selection, including via gene editing of closely related *Plasmodium* species,[55] is also a necessary step toward substantiating the biological relevance of these loci, with implications for transmission prevention, treatment strategies, and vaccine development for *P. ovale*. Finally, cataloging genome-wide diversity facilitates the design of targeted genotyping methods that can efficiently characterize the epidemiology of these understudied parasite species. Using such assays to better evaluate *P. ovale* parasite relatedness, transmission, and relapse patterns is critical for improving current malaria control strategies targeting *P. falciparum* that do not consider non-falciparum malaria species.[26,56,57]

# Methods

## Sample selection

Clinical isolates in the form of dried blood spots or leukodepleted blood were drawn from six studies shown in **Table 1**, including studies involving both asymptomatic persons and febrile patients across four countries. Across these studies, participants were screened for the presence of *P. ovale spp.* infection by a real-time polymerase chain reaction (qPCR) assay targeting the *po18S* rRNA gene.[19] Among 282 isolates with a *po18S* Ct value under 40, a species-specific (*Poc* and *Pow*) 18S rRNA qPCR assay was employed to determine the ovale species present.[58] Candidates were selected from isolates with only one species detected or mixed infections in which one species predominated by ≥3 Ct (corresponding to approximately 8 times as much DNA). Samples were also screened for presence of *P. falciparum* co-infection using a qPCR assay for the *pfldh* or *pf18S* rRNA gene.[29,34] Ultimately, samples from 25 individuals were selected for whole-genome sequencing based on higher-density *P. ovale* infection, lack of or lower-density *P. falciparum* coinfection, and balance of ovale species and geographic diversity across the sample set. Characteristics of these 25 samples, and an additional four samples from a study of *P. ovale wallikeri* in French travelers that were whole genome sequenced by Joste et al.,[26] are shown in **Supplemental Table 1**.

## Library preparation and sequencing

DNA extracted from dried blood spots using a chelex protocol[59] was sheared to 300bp using a LE220R-plus Covaris Sonicator (Covaris, Woburn, MA). Fragment size was checked with an Agilent TapeStation 4150 (Agilent, Santa Clara, CA) and DNA concentrations were tested using a Qubit Flex fluorometer (Thermo Fisher Scientific, Waltham, MA). Isolates were then prepared for sequencing using the KAPA Hyperprep kit (Kapa Biosystems, Woburn, MA). Four Tanzanian

DNA isolates extracted from blood that had been leukocyte-depleted by CF11 filtration at the time of collection[25] were directly incorporated into sequencing libraries. The remaining 21 isolates derived from dried blood spots were additionally processed using a custom-designed hybrid capture protocol to enrich for ovale DNA via thousands of RNA probes specifically designed to amplify *P. ovale* DNA without binding to human DNA (Twist Bioscience, San Francisco, CA, USA). Hybrid captures probes were designed first for the *P. ovale wallikeri* genome, with unique probes added for *P. ovale curtisi* at any sites that differed by more than 10% of bases. Since the reference genome for chromosome 10 is significantly smaller for *Poc* compared to *Pow* (roughly 1,300kb vs 470kb, respectively), this bait design approach led to a lack of baits covering 63% of *Poc* chromosome 10. Chromosome 10 was therefore excluded from analysis of *Poc* isolates. Captures were performed with four samples per capture. After preliminary sequencing on a Miseq Nano flow cell (Illumina, San Diego, CA, USA), libraries were sequenced on the Novaseq 6000 S Prime (Illumina, San Diego, CA, USA) sequencing system with 150bp paired-end chemistry.

## Sequencing data alignment and variant calling

Data processing and analysis were performed in the bash environment using a python-based *snakemake* v7.24.2 wrapper for pipeline construction, automation, and reproducibility.[60] Following trimming and quality control, a “dual” reference genome was produced for each *P. ovale* species by concatenating the reference genome of that ovale species (*P. ovale curtisi: PocGH01; P. ovale wallikeri: PowCR01*) to the *P. falciparum* strain *Pf3D7* reference genome.[24,61] Alignment and deduplication of sequencing data was performed following the *GATK* v4.4.0.0 best practices pipeline.[62] *Samtools* v1.17 was then used to select for reads that aligned to the *P. ovale* portion of the dual reference genome rather than that of *P. falciparum*, thus discarding reads from contaminating *P. falciparum* DNA present in some isolates.[63] Resulting alignments of reads that preferentially mapped to *P. ovale* were soft-clipped to reference genome edges and cleaned of unmapped reads using *GATK*, after which mapping proportion and coverage of the ovale reference genome were calculated by *samtools* and *bedtools* v2.30.[64] Additional details about sequencing data processing are available in **Supplemental Methods** and **Data Availability.**

Variant calling from aligned reads across each ovale species genome was also performed using the *GATK* best practices pipeline.[62] In the resulting callset, variants were masked if they fell outside of the 14 chromosomes of the reference genome or were part of specific expanded gene families (see **Supplemental Methods**). *GATK* hard filtering was then used to remove variants with poor quality metrics using the following filter thresholds: quality by depth <3, Fisher strand bias >50, strand odds ratio >3, mapping quality >50, mapping quality rank sum <-2.5, read position rank sum <-3. Callsets were limited to biallelic single nucleotide polymorphisms (SNPs) that were present in at least 80% of individuals. SNP density across the entire genome and within specific functional regions of each genome were calculated using custom scripts (see **Data Availability**). *SNPeff* v4.3 was used to annotate individual variants and determine the ratio of nonsynonymous-to-synonymous mutations.[24,65]

## Selection of *P. falciparum* comparison dataset

Co-endemic *P. falciparum* samples were drawn from the Pf6 dataset.[33] Of 20,705 total *P. falciparum* isolates from around the globe, 842 came from the same or nearby geographic locations as the source studies of *P. ovale* isolates described above. Twenty-one *P. falciparum* samples with over 85% base callability were randomly selected in order to have at least one co-endemic *P. falciparum* isolate for each *P. ovale curtisi* or *ovale wallikeri* isolate (**Supplemental Table 2**). Variant callsets for these *P. falciparum* samples were limited to the falciparum core genome,[61] quality filtered by Variant Quality Score Recalibration,[62] and restricted to sites present in at least 80% of individuals.

## Population genetic analyses

Complexity of infection was estimated for all samples using *THE REAL McCOIL* via the *McCOILR* package.[66] Principal components were calculated and overall genomic similarity compared among monoclonal isolates using *PLINK* v1.90b6.21.[67] Nucleotide diversity was estimated among sets of one-to-one orthologous genes between monoclonal *Poc, Pow*, and *P. falciparum* isolates using *vcftools* v0.1.15.[68] For signatures of selection, $nS_L$ was calculated for all non-missing variants in monoclonal isolates of each *P. ovale* species using *selscan* v1.2.0,[69] while Tajima's D was calculated in 300bp sliding windows in protein-coding genes using *vcf-kit* v0.2.9.[70] Additional details on statistical methodology can be found in the **Supplemental Methods**.

**Acknowledgements**: We thank all the participants of the parent studies for providing the biological material for the study. The following reagents were obtained through BEI Resources, NIAID, NIH: 1) *Plasmodium falciparum*, Strain 3D7A, MRA-151, contributed by David Walliker and 2) Diagnostic Plasmid Containing the Small Subunit Ribosomal RNA Gene (18S) from Plasmodium ovale, MRA-180, contributed by Peter A. Zimmerman. We also thank Dr. Parul Johri for her feedback on the final manuscript and Dr. Kevin Wamae for his support in bioinformatic pipeline construction.

**Ethics**: This analysis was considered non-human subjects research by the University of North Carolina. The appropriate collection of samples and IRBs involved are summarized in the original studies outlined in **Supplemental Table 1**.

**Data Availability**: All new sequence data are available at NCBI SRA (BioProject ID: PRJNA1092086). Public data used include European Nucleotide Archive Study Accession Number: PRJEB51041 (Run Accession Numbers: ERR10738334, ERR10738339, ERR10738341, ERR10738346) and additional Run Accession Numbers: ERR404145, ERR404154, ERR377533, ERR404191, ERR404207, ERR1045266, ERR1045267, ERR676479, ERR1106575, ERR1106579, ERR1106586, ERR1106587, ERR1106590, ERR449901, ERR449903, ERR405238, ERR1172603, ERR636018, ERR1172590, ERR912913, ERR1172615. All code for processing and analysis of samples is available at https://github.com/bailey-lab/Po_popgen_snakemake/tree/main/full. All other data or queries will be made available by the corresponding author on reasonable request.

**Author Contributions:** KCE, JTL, JJJ and JBP conceived and designed the study. SF, BG, FP, KM, OA, DI, IA, BB, BN, AK, and AT contributed to clinical sample collection. KCE, ZPH, MM, and CG conducted laboratory work. KCE, ZPH, AS, KN, and AS conducted data analysis. KCE drafted the first version of the manuscript. KCE, JTL, JJJ, ZPH, JB, and JP edited and finalized the manuscript.

**Competing Interests:** J.B.P. reports research support from Gilead Sciences, non-financial support from Abbott Laboratories, and consulting for Zymeron Corporation, all outside the scope of this study. The remaining authors declare no competing interests.

**Funding**: This study was funded by the National Institute for Allergy and Infectious Diseases, National Institutes of Health (R01AI137395 to JTL, R21AI152260 to JTL, R21AI148579 to JBP and JTL, R01AI65537 to JJJ, K24AI134990 to JJJ, R0AI129812 to AT, R01AI132547 to JJJ, R01TW010870 to JJJ, T32AI070114 supporting KCE), the Fogarty Center, the Global Fund to Fight AIDS, Tuberculosis, and Malaria, and the Bill and Melinda Gates Foundation (Inv. No. 002202 to DI, JJJ and JAB). Under the grant conditions of the Bill and Melinda Gates Foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the Author Accepted Manuscript version that might arise from this submission.

**Disclaimer:** The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the Bill and Melinda Gates Foundation or other funders.

## References

1. Geneva: World Health Organization. *World Malaria Report 2023*. vol. Licence: CC BY-NC-SA 3.0 IGO. (2023).
2. Lover, A. A., Baird, J. K., Gosling, R. & Price, R. N. Malaria Elimination: Time to Target All Species. *Am. J. Trop. Med. Hyg.* **99**, 17–23 (2018).
3. Betson, M., Clifford, S. & Stanton, M. Emergence of Nonfalciparum Plasmodium Infection Despite Regular Artemisinin Combination Therapy in an 18-Month Longitudinal Study of Ugandan Children and …. *The Journal of* (2018).
4. Yman, V. *et al.* Persistent transmission of Plasmodium malariae and Plasmodium ovale species in an area of declining Plasmodium falciparum transmission in eastern Tanzania. *PLoS Negl. Trop. Dis.* **13**, e0007414 (2019).

5. Akala, H. M. *et al.* Plasmodium interspecies interactions during a period of increasing prevalence of Plasmodium ovale in symptomatic individuals seeking treatment: an observational study. *The Lancet Microbe* vol. 2 e141–e150 Preprint at https://doi.org/10.1016/s2666-5247(21)00009-4 (2021).
6. Amato, R. *et al.* Genetic markers associated with dihydroartemisinin–piperaquine failure in Plasmodium falciparum malaria in Cambodia: a genotype–phenotype association study. *Lancet Infect. Dis.* **17**, 164–173 (2017).
7. Bailey, J. A. *et al.* Microarray analyses reveal strain-specific antibody responses to Plasmodium falciparum apical membrane antigen 1 variants following natural infection and vaccination. *Sci. Rep.* **10**, 3952 (2020).
8. Amambua-Ngwa, A. *et al.* Major subpopulations of Plasmodium falciparum in sub-Saharan Africa. *Science* **365**, 813–816 (2019).
9. Volkman, S. K. *et al.* A genome-wide map of diversity in Plasmodium falciparum. *Nat. Genet.* **39**, 113–119 (2007).
10. Ghansah, A. *et al.* Monitoring parasite diversity for malaria elimination in sub-Saharan Africa. *Science* **345**, 1297–1298 (2014).
11. Collins, W. E. & Jeffery, G. M. Plasmodium ovale: parasite and disease. *Clin. Microbiol. Rev.* **18**, 570–581 (2005).
12. Mueller, I., Zimmerman, P. A. & Reeder, J. C. Plasmodium malariae and Plasmodium ovale--the ‘bashful’malaria parasites. *Trends Parasitol.* **23**, 278–283 (2007).
13. Calderaro, A. *et al.* Genetic polymorphisms influence Plasmodium ovale PCR detection accuracy. *J. Clin. Microbiol.* **45**, 1624–1627 (2007).
14. Sutherland, C. J. *et al.* Two nonrecombining sympatric forms of the human malaria parasite Plasmodium ovale occur globally. *J. Infect. Dis.* **201**, 1544–1550 (2010).
15. Higgins, M. *et al.* New reference genomes to distinguish the sympatric malaria parasites, Plasmodium ovale curtisi and Plasmodium ovale wallikeri. *Sci. Rep.* **14**, 3843 (2024).

16. Oguike, M. C. *et al.* Plasmodium ovale curtisi and Plasmodium ovale wallikeri circulate simultaneously in African communities. *International Journal for Parasitology* vol. 41 677–683 Preprint at https://doi.org/10.1016/j.ijpara.2011.01.004 (2011).
17. Fuehrer, H.-P. *et al.* Plasmodium ovale in Bangladesh: genetic diversity and the first known evidence of the sympatric distribution of Plasmodium ovale curtisi and Plasmodium ovale wallikeri in southern Asia. *Int. J. Parasitol.* **42**, 693–699 (2012).
18. Fuehrer, H.-P., Campino, S. & Sutherland, C. J. The primate malaria parasites Plasmodium malariae, Plasmodium brasilianum and Plasmodium ovale spp.: genomic insights into distribution, dispersal and host transitions. *Malaria Journal* vol. 21 Preprint at https://doi.org/10.1186/s12936-022-04151-4 (2022).
19. Sendor, R. *et al.* Similar Prevalence of Plasmodium falciparum and Non–P. falciparum Malaria Infections among Schoolchildren, Tanzania. *Emerging Infectious Disease journal* **29**, 1143 (2023).
20. Lei, Y. *et al.* Low genetic diversity and strong immunogenicity within the apical membrane antigen-1 of plasmodium ovale spp. imported from africa to china. *Acta Trop.* **210**, 105591 (2020).
21. Chen, J. *et al.* Disparate selection of mutations in the dihydrofolate reductase gene (dhfr) of Plasmodium ovale curtisi and P. o. wallikeri in Africa. *PLoS Negl. Trop. Dis.* **16**, e0010977 (2022).
22. Bright, A. T. *et al.* Whole genome sequencing analysis of Plasmodium vivax using whole genome capture. *BMC Genomics* **13**, 262 (2012).
23. Joste, V., Guillochon, E., Clain, J., Coppée, R. & Houzé, S. Development and Optimization of a Selective Whole-Genome Amplification To Study Plasmodium ovale Spp. *Microbiol Spectr* e0072622 (2022).
24. Rutledge, G. G. *et al.* Plasmodium malariae and P. ovale genomes provide insights into malaria parasite evolution. *Nature* vol. 542 101–104 Preprint at

https://doi.org/10.1038/nature21038 (2017).

25. Venkatesan, M. *et al.* Using CF11 cellulose columns to inexpensively and effectively remove human DNA from Plasmodium falciparum-infected whole blood samples. *Malar. J.* **11**, 41 (2012).
26. Joste, V. *et al.* Genetic Profiling of Plasmodium ovale wallikeri Relapses With Microsatellite Markers and Whole-Genome Sequencing. *J. Infect. Dis.* **228**, 1089–1098 (2023).
27. Feleke, S. M. *et al.* Plasmodium falciparum is evolving to escape malaria rapid diagnostic tests in Ethiopia. *Nat Microbiol* **6**, 1289–1299 (2021).
28. Parr, J. B. *et al.* Analysis of false-negative rapid diagnostic tests for symptomatic malaria in the Democratic Republic of the Congo. *Sci. Rep.* **11**, 6495 (2021).
29. Mwandagalirwa, M. K. *et al.* Individual and household characteristics of persons with Plasmodium falciparum malaria in sites with varying endemicities in Kinshasa Province, Democratic Republic of the Congo. *Malar. J.* **16**, 456 (2017).
30. Tarimo, B. B. *et al.* Seasonality and transmissibility of Plasmodium ovale in Bagamoyo District, Tanzania. *Parasit. Vectors* **15**, 56 (2022).
31. Rogier, E. *et al.* Plasmodium falciparum pfhrp2 and pfhrp3 gene deletions among patients enrolled at 100 health facilities throughout Tanzania: February to July 2021. *bioRxiv* (2023) doi:10.1101/2023.07.29.23293322.
32. Ali, I. M. *et al.* Arboviruses as an unappreciated cause of non-malarial acute febrile illness in the Dschang Health District of western Cameroon. *PLoS Negl. Trop. Dis.* **16**, e0010790 (2022).
33. MalariaGEN *et al.* An open dataset of Plasmodium falciparum genome variation in 7,000 worldwide samples. *Wellcome Open Res.* **6**, 42 (2021).
34. Sendor, R. *et al.* Epidemiology of Plasmodium malariae and Plasmodium ovale spp. in Kinshasa Province, Democratic Republic of Congo. *Nat. Commun.* **14**, 6618 (2023).
35. Bright, A. T. *et al.* A high resolution case study of a patient with recurrent Plasmodium vivax

infections shows that relapses were caused by meiotic siblings. *PLoS Negl. Trop. Dis.* **8**, e2882 (2014).

36. Sirawaraporn, W., Sathitkul, T., Sirawaraporn, R., Yuthavong, Y. & Santi, D. V. Antifolate-resistant mutants of Plasmodium falciparum dihydrofolate reductase. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 1124–1129 (1997).
37. Hastings, M. D. *et al.* Dihydrofolate reductase mutations in Plasmodium vivax from Indonesia and therapeutic response to sulfadoxine plus pyrimethamine. *J. Infect. Dis.* **189**, 744–750 (2004).
38. Gomes, A. R. *et al.* A transcriptional switch controls sex determination in Plasmodium falciparum. *Nature* **612**, 528–533 (2022).
39. Russell, A. J. C. *et al.* Regulators of male and female sexual development are critical for the transmission of a malaria parasite. *Cell Host Microbe* **31**, 305–319.e10 (2023).
40. Tadesse, F. G. *et al.* Gametocyte Sex Ratio: The Key to Understanding Plasmodium falciparum Transmission? *Trends Parasitol.* **35**, 226–238 (2019).
41. Khan, S. M. *et al.* Proteome analysis of separated male and female gametocytes reveals novel sex-specific Plasmodium biology. *Cell* **121**, 675–687 (2005).
42. Douradinha, B. *et al.* Plasmodium Cysteine Repeat Modular Proteins 3 and 4 are essential for malaria parasite transmission from the mosquito to the host. *Malar. J.* **10**, 71 (2011).
43. Galaway, F. *et al.* P113 is a merozoite surface protein that binds the N terminus of Plasmodium falciparum RH5. *Nat. Commun.* **8**, 14333 (2017).
44. Tebben, K. *et al.* Malian children infected with Plasmodium ovale and Plasmodium falciparum display very similar gene expression profiles. *PLoS Negl. Trop. Dis.* **17**, e0010802 (2023).
45. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
46. Singh, G. P. & Sharma, A. South-East Asian strains of Plasmodium falciparum display

higher ratio of non-synonymous to synonymous polymorphisms compared to African strains. *F1000Res.* **5**, 1964 (2016).

47. Feng, X. *et al.* Single-nucleotide polymorphisms and genome diversity in Plasmodium vivax. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 8502–8507 (2003).
48. Chang, H.-H. *et al.* Malaria life cycle intensifies both natural selection and random genetic drift. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 20129–20134 (2013).
49. Saralamba, N. *et al.* Genetic dissociation of three antigenic genes in Plasmodium ovale curtisi and Plasmodium ovale wallikeri. *PLoS One* **14**, e0217795 (2019).
50. Faye, F. B. K. *et al.* Diagnostic criteria and risk factors for Plasmodium ovale malaria. *J. Infect. Dis.* **186**, 690–695 (2002).
51. Thomson-Luque, R., Stabler, T. C., Fürle, K., Silva, J. C. & Daubenberger, C. Plasmodium falciparum merozoite surface protein 1 as asexual blood stage malaria vaccine candidate. *Expert Rev. Vaccines* **23**, 160–173 (2024).
52. Escalante, A. A., Lal, A. A. & Ayala, F. J. Genetic polymorphism and natural selection in the malaria parasite Plasmodium falciparum. *Genetics* **149**, 189–202 (1998).
53. Chu, R. *et al.* Limited genetic diversity of N-terminal of merozoite surface protein-1 (MSP-1) in Plasmodium ovale curtisi and P. ovale wallikeri imported from Africa to China. *Parasit. Vectors* **11**, 596 (2018).
54. Joyner, C. J. *et al.* Humoral immunity prevents clinical malaria during Plasmodium relapses without eliminating gametocytes. *PLoS Pathog.* **15**, e1007974 (2019).
55. Thiam, L. G., Mangou, K., Ba, A., Mbengue, A. & Bei, A. K. Leveraging genome editing to functionally evaluate Plasmodium diversity. *Trends Parasitol.* **38**, 558–571 (2022).
56. Liu, P. *et al.* Increasing proportions of relapsing parasite species among imported malaria in China's Guangxi Province from Western and Central Africa. *Travel Med. Infect. Dis.* **43**, 102130 (2021).
57. Wångdahl, A. *et al.* Relapse of Plasmodium vivax and Plasmodium ovale Malaria With and

Without Primaquine Treatment in a Nonendemic Area. *Clin. Infect. Dis.* **74**, 1199–1207 (2022).

58. Potlapalli, V. R. *et al.* Real-time PCR detection of mixed Plasmodium ovale curtisi and wallikeri infections in human and mosquito hosts. *PLoS Negl. Trop. Dis.* **17**, e0011274 (2023).
59. Mitchell, C. L. *et al.* Under the Radar: Epidemiology of Plasmodium ovale in the Democratic Republic of the Congo. *J. Infect. Dis.* **223**, 1005–1014 (2021).
60. Mölder, F. *et al.* Sustainable data analysis with Snakemake. *F1000Res.* **10**, 33 (2021).
61. Miles, A. *et al.* Indels, structural variation, and recombination drive genomic diversity in Plasmodium falciparum. *Genome Res.* **26**, 1288–1299 (2016).
62. Van der Auwera, G. A. & O'Connor, B. D. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. ('O'Reilly Media, Inc.', 2020).
63. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
64. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
65. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
66. Chang, H.-H. *et al.* THE REAL McCOIL: A method for the concurrent estimation of the complexity of infection and SNP allele frequency for malaria parasites. *PLoS Comput. Biol.* **13**, e1005348 (2017).
67. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
68. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

69. Szpiech, Z. A. & Hernandez, R. D. selscan: An Efficient Multithreaded Program to Perform EHH-Based Scans for Positive Selection. *Molecular Biology and Evolution* vol. 31 2824–2827 Preprint at https://doi.org/10.1093/molbev/msu211 (2014).

70. Cook, D. E. & Andersen, E. C. VCF-kit: assorted utilities for the variant call format. *Bioinformatics* **33**, 1581–1582 (2017).