Article

High-throughput genotyping of *Plasmodium vivax* in the Peruvian Amazon via molecular inversion probes

Zachary R. Popkin-Hall [1] ✉, Karamoko Niaré[2], Rebecca Crudale[2], Alfred Simkin[2], Abebe A. Fola [2], Juan F. Sanchez [3], Danielle L. Pannebaker[3], David J. Giesbrecht[2], Isaac E. Kim Jr. [4,5], Özkan Aydemir [6], Jeffrey A. Bailey [2,5,10], Hugo O. Valdivia [3,10] & Jonathan J. Juliano [1,7,8,9,10]

Plasmodium vivax transmission occurs throughout the tropics and is an emerging threat in areas of *Plasmodium falciparum* decline, causing relapse infections that complicate treatment and control. Targeted sequencing for *P. falciparum* has been widely deployed to detect population structure and the geographic spread of antimalarial and diagnostic resistance. However, there are fewer such tools for *P. vivax*. Leveraging global variation data, we designed four molecular inversion probe (MIP) genotyping panels targeting geographically differentiating SNPs, neutral SNPs, putative antimalarial resistance genes, and vaccine candidate genes. We deployed these MIP panels on 866 infections from the Peruvian Amazon and identified transmission networks with clonality (IBD[identity by descent]>0.99), copy number variation in *Pvdbp* and multiple *Pvrbps*, mutations in antimalarial resistance orthologs, and balancing selection in 13 vaccine candidate genes. Our MIP panels are the broadest genotyping panel currently available and are poised for successful deployment in other regions of *P. vivax* transmission.

Plasmodium vivax is the most geographically widespread malaria species, with active transmission throughout South and Southeast Asia, Latin America, the Pacific, and parts of Africa[1]. *P. vivax* is also an emerging threat in areas where *P. falciparum* has been eliminated or is on track for elimination[2]. Transmission of *P. vivax* is more complex due to the presence of hypnozoites in the liver, which cause disease relapse without a new infectious mosquito bite[3], and earlier gametocytogenesis than *P. falciparum*[4], meaning that it can be transmitted earlier in an infection, even among asymptomatic patients[5]. Genomic epidemiology is a powerful tool for understanding parasite biology and transmission, and while some studies have been performed on *P. vivax*[6–18], most have been done in *P. falciparum*[19–24]. Given the challenging biology impeding control[25], there is an urgent need for the development of high-throughput genotyping methods to better understand this neglected parasite.

Genomic epidemiology studies of *P. falciparum* using whole genome sequencing (WGS), molecular inversion probes (MIPs), and amplicon sequencing have generated useful data to support malaria

[1]Institute for Global Health and Infectious Diseases, University of North Carolina, Chapel Hill, NC, USA. [2]Department of Pathology and Laboratory Medicine, Warren Alpert Medical School, Brown University, Providence, RI, USA. [3]Department of Parasitology, U.S. Naval Medical Research Unit SOUTH (NAMRU SOUTH), Lima, Peru. [4]The Warren Alpert Medical School of Brown University, Providence, RI, USA. [5]Center for Computational Molecular Biology, Brown University, Providence, RI, USA. [6]Program in Molecular Medicine, UMass Chan Medical School, Worcester, MA, USA. [7]Division of Infectious Diseases, University of North Carolina School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [8]Curriculum of Genetics and Molecular Biology, University of North Carolina School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [9]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [10]These authors jointly supervised this work: Jeffrey A. Bailey, Hugo O. Valdivia, Jonathan J. Juliano. ✉e-mail: zach_popkin-hall@med.unc.edu

elimination by identifying parasite population structure[19,20], importation and transmission networks and connectivity[20], and the geographical spread of antimalarial[19,20,22–24,26–28] and diagnostic resistance[29]. Such studies of *P. vivax* also exist, but are rarer[6–18]. However, they are similarly necessary, especially in South America, where *P. vivax* is now the predominant species and sympatric with *P. falciparum*[15]. Peru had the fourth largest number of malaria cases in the Americas in 2022, with the majority being *P. vivax*[30]. Most transmission occurs in the Amazonian region and department of Loreto[31,32], where morbidity and economic losses stemming from malaria are greater than in the rest of the country[33]. Cases tend to cluster at the household and community levels[34], with most malaria cases being asymptomatic and submicroscopic[34–36]. While the malaria burden in Peru has significantly declined from 1990 levels, progress has stalled and even reversed in the past decade[33], with increases from 2010 onwards reversing the successes of the early 2000s[32]. Furthermore, *P. vivax* prevalence has been slower to decline than *P. falciparum* due to its unique biology that makes it more resistant to most current control measures, which are designed to target *P. falciparum*[37].

In contrast to *P. falciparum*, where there are multiple large-scale genotyping panels in use[19,20,22–24,26–29], there are relatively few existing panels for *P. vivax*. Genotyping panels based on targeted deep sequencing have been used to study *P. falciparum* populations and their capacity to escape control interventions in Africa and Latin America, including population structure and gene flow, geographic transmission patterns, origins of imported malaria cases, spread of antimalarial drug resistance, and genetic diversity of vaccine targets[19,20,22–24,26–29]. *P. vivax* genotyping, however, has historically relied on microsatellite genotyping with SNP barcodes and multiplex amplicon sequencing panels only more recently being used in focal studies[18,38–42]. A summary of existing panels is shown in Table 1.

MIPs are a highly multiplexed targeted sequencing approach that allow thousands of loci to be captured in a single tube. They are high-throughput and cost-effective for large genomic studies[43] (see Supplementary Methods for an estimate) and have several advantages including the ability to tile across genes and integrated unique molecular identifiers (UMIs) that tag specific strands of DNA and allow bioinformatic error correction. This approach has been highly successfully used for studying *P. falciparum* epidemiology in Africa[19,20,22,23]. Thus, we sought to replicate it by creating four panels of MIPs to address multiple pertinent questions regarding *P. vivax* biology, including population structure, antimalarial resistance orthologs, and selection in potential blood-stage vaccine targets. Using publicly available *P. vivax* WGS datasets (details in Supplementary Methods, Supplementary Data 1), we designed our panels based on all known *P. vivax* genetic diversity to make them deployable in any epidemiological context. Here, we report the use of these MIP panels on 866 *P. vivax* samples collected in the Peruvian Amazon by NAMRU SOUTH between 2011 and 2018 (Supplementary Table 1) which enabled us to detect the frequency of putative antimalarial resistance alleles, identify balancing selection on red blood cell (RBC) invasion ligands, calculate complexity of infection, detect copy number variation in multiple genes, and reveal highly related (IBD ≥ 0.99) clonal transmission networks that span years.

Results

High-throughput sequencing and genotyping of Peruvian *P. vivax* isolates

Following panel optimization and MIP rebalancing on high parasitemia isolates, we performed MIP captures on 866 samples collected by NAMRU South, primarily in Loreto (Fig. 1). Two captures were performed for each sample, with one for the tiled gene panel (PvG) and one multiplexed capture for three SNP panels (PvFST, PvMAF5, PvMAF40), with a repool (only samples with average MIP < 100X UMI coverage included in repool) after initial sequencing to improve sequencing coverage and depth. We successfully extracted usable sequences from 685 samples. Additional details on sequencing output (Supplementary Fig. 1, Supplementary Fig. 2), samples successfully sequenced (Supplementary Table 1) and impact of parasitemia (Supplementary Fig. 3) are described in the Supplementary Results and Discussion. The number of samples retained for each step of the analysis is shown in Supplementary Fig. 4.

Clonal transmission detected in Belén 2012–2014

To assess relatedness between isolates, a principal component analysis (PCA) was performed for 685 samples that passed quality filters and had sufficient associated metadata for further analysis. The PCA revealed no broad geographical or temporal structure, but PC1 and PC2 included multiple clusters that were associated by both city and year (plotted by city, collection year, and both in Supplementary Fig. 5). To investigate these clusters further, pairwise identity by descent (IBD) was calculated for these samples using the MIPanalyzer R package (https://github.com/mrc-ide/MIPanalyzer). The median, mean, and maximum pairwise IBD were 0, 0.0878, 0.99, respectively. The overall distribution was bimodal, with 599 isolates being distantly related to at least one other pair (IBD < 0.25) and 448 being very highly related to at least one other pair (IBD ≥ 0.9) (Supplementary Fig. 6, Supplementary Fig. 7, Supplementary Fig. 8, Supplementary Fig. 9). To detect evidence of clonal transmission, IBD networks were generated between isolates with IBD ≥ 0.99 for five communities with a sufficient number of isolates: Belén, Indiana, Iquitos, Punchana, and San Juan Bautista (Fig. 2). In Belén, multiple clonal clusters were detected (Fig. 2), with many of these clusters spanning multiple years of collection. Smaller clonal networks were also detected in isolates collected from 2016 and 2017 in Indiana, 2012 and 2013 in Punchana, 2016 and 2017 in Punchana, and 2015–2017 in San Juan Bautista (Fig. 2).

Most infections are monoclonal

Of 697 isolates (all samples that passed quality filters, regardless of metadata) analyzed with THEREALMcCOIL, the majority (637, 91.4%) were monoclonal (Fig. 3). Among the 60 isolates estimated to be polyclonal, 44 (73.3%) were estimated to contain two clones, 14 (23.3%) were estimated to contain three, and two (3.3%) were estimated to contain four clones. The median COI estimate across all samples was 1.

Copy number variation detected in duffy and reticulocyte binding proteins

Prior to calculating copy number based on UMI depth, the larger dataset of 685 samples was filtered for missingness and high coverage (see "Methods" for details), as well as monoclonality, leaving 595 samples for copy number calculation. Copy number variants (CNV) were detected in twelve samples (2.02%) (Fig. 4). All twelve samples contained multiple copies of reticulocyte binding protein 2b (*Pvrbp2b*), eleven of which had two copies while one had three copies. Ten of the twelve samples showed CNV only in *Pvrbp2b*, while two had additional duplicated genes (Fig. 4). The sample with three copies of *Pvrbp2b* also had two copies of Duffy Binding Protein (*Pvdbp*), *Pvrbp1b*, *Pvrbp2a*, and *Pvrbp2c* (Fig. 4). The final sample had a duplication in *Pvrbp2c*. No CNV was detected in either *Pvrbp1a* or erythrocyte binding protein (*Pvebp*).

Mutations in antimalarial resistance orthologs

Allele frequencies were calculated for mutations that are found in antimalarial resistance orthologs in *P. vivax*. The full set of allele frequencies for all antimalarial resistance ortholog genes is reported in Supplementary Data 2. Mutations were detected in *Pvcrt*, *Pvdhfr-ts*, *Pvpppk-dhps*, and *Pvmdr1* (Table 2). Allele frequencies in detected mutations are mostly rare (below 15%) (Table 2), though the *Pvpppk-dhps* G383A mutation was detected at a relatively high frequency (49.6%) (Table 2).

Table 1 | Summary of Existing *P. vivax* Genotyping Panels

| Method	Key Use Case	Targets Included	Neutral Sites	Antimalarial Resistance Orthologs	Vaccine Candidates	Advantages	Disadvantages
High-resolution melting (HRM) SNP barcode[39]	Detecting importation	42 SNPs	Yes	No	No	Economical, sequencing not required, effectively differentiates samples by geography to characterize population genomics and identify importation, requires minimal sample volumes and concentrations	Does not incorporate information on biologically significant genes (antimalarial resistance orthologs or erythrocyte invasion ligands), limited ability to detect selection or population dynamics
Papua New Guinea SNP barcode[41]	Detecting subnational population structure and transmission networks	146 SNPs	Yes	No	No	Reveals subnational population structure by geography, allows detection of within-country transmission networks, can detect polyclonal infections	Does not incorporate information on biologically significant genes (antimalarial resistance orthologs or erythrocyte invasion ligands)
vivaxGEN-geo[40]	Detecting importation	33, 50, or 55 SNPs (three options available)	No	Limited	No	Accurately identifies importation on a global level, reflects global population diversity, easily modifiable, works with polyclonal infections	Non-neutral SNPs may need to be replaced over time, limited South Asian data incorporated, does not systematically address biologically significant genes (antimalarial resistance orthologs or erythrocyte invasion ligands)
Pv AmpliSeq[38]	Detecting subnational population structure and transmission networks, importation, resistance genotyping	229 amplicons (includes 33 vivaxGEN-geo SNPs and 42-SNP Vietnam barcode, 11 putative drug resistance genes)	No	Yes	Limited to *Pvama1*	Provides high resolution within a high-transmission country with high gene flow (Vietnam), incorporates vivaxGEN-geo to identify importation, includes 11 putative antimalarial resistance genes, can be used to calculate identity-by-descent (IBD), works with low parasitemia (5 parasites/μL) samples, incorporates *Pvama1* to facilitate study of polyclonal infections and antigen diversity, easily modifiable	Needs to be updated for new geographic contexts, does not include the full suite of antimalarial resistance orthologs, unable to detect selection in erythrocyte invasion ligands, identification of importation origin not always successful
Pv AmpliSeqv2[18]	Detecting subnational population structure and transmission networks, importation, resistance genotyping	33 vivaxGEN-geo SNPs, 41-SNP Peru barcode, 11 putative drug resistance genes	No	Yes	Limited to *Pvama1*	Builds on Pv AmpliSeq v1 to characterize Peruvian population structure, incorporates the core SNPs and associated advantages from PvAmpiSeq v1, likely to work elsewhere in Latin America	Needs to be updated for new geographic contexts, does not include the full suite of antimalarial resistance orthologs, unable to detect selection in erythrocyte invasion ligands, identification of importation origin not always successful
Microhaplotype Panels[42]	Detecting subnational population structure and transmission networks, importation, detecting relapse infections	100 microhaplotypes (3-10 SNPs per each)	Yes	No	No	Works reliably on paired samples to distinguish relapse and recrudescence from reinfection, incorporates global diversity for a standardized barcode that can be used anywhere and enable the detection of importation, enables high-resolution population structure analysis based on IBD, can detect polyclonal infections	Markers were selected with relatively small number of samples from South Asia and Latin America, does not incorporate putative vaccine candidates or antimalarial resistance orthologs
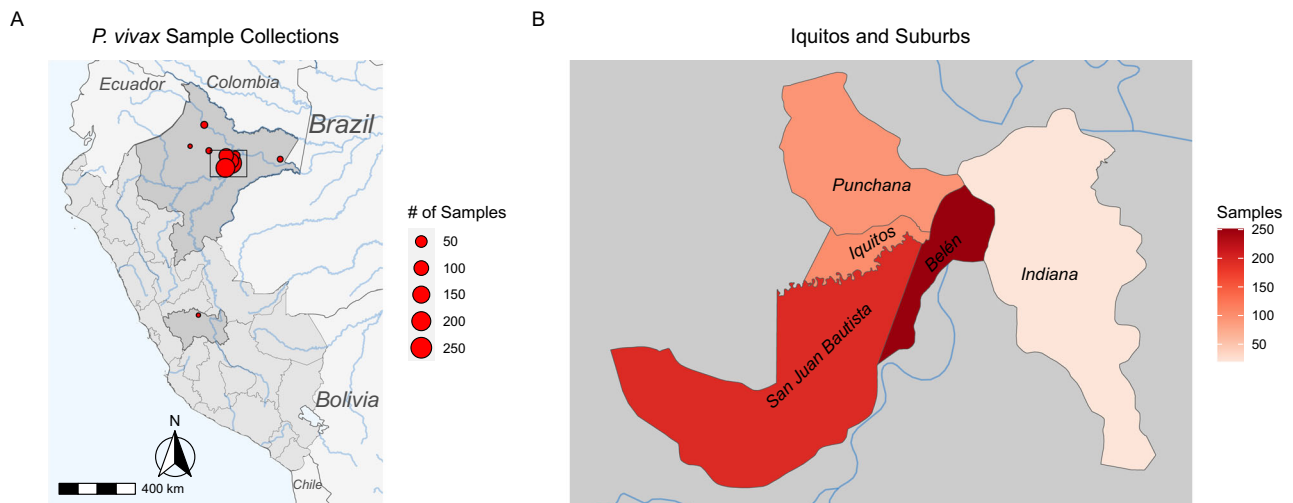
Fig. 1 | Study sites and sample size. A Locations of sample collection by community within Peru, scaled by size of circle. Darker gray boundaries identify the regions and departments of Loreto (where all but one of the circles are located) and Junín. Blue rectangle defines Iquitos, expanded in (**B**) Inset map depicting region around Iquitos, scaled by color of shading. Map created in R using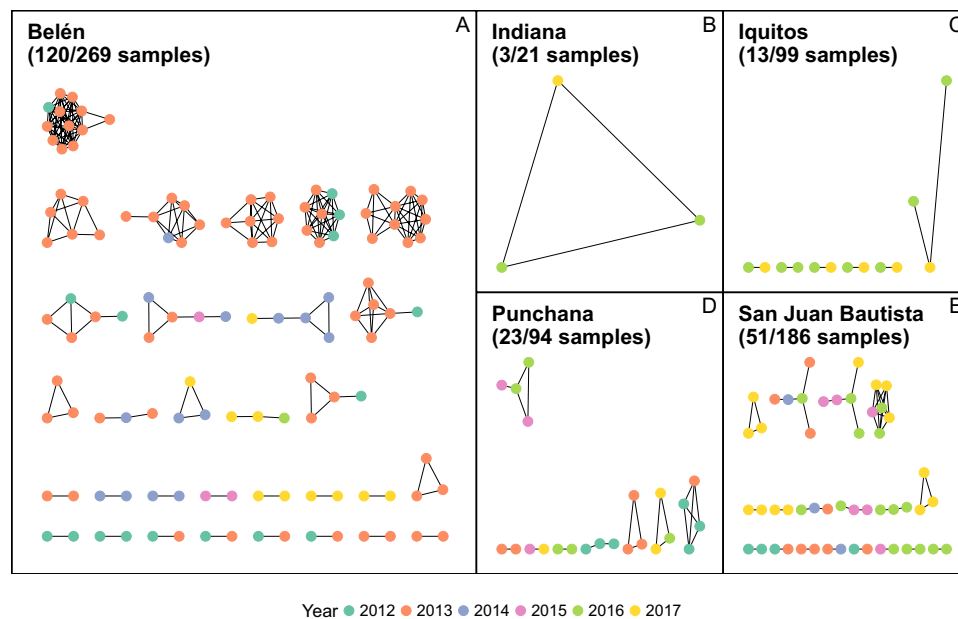 open data from GADM database, v 3.6. www.gadm.org and the Sf package (https://cran.r-project.org/package=sf) with GPL-2 license (https://cran.r-project.org/web/licenses/GPL-2).

Fig. 2 | Identity by descent (IBD) networks for five communities within the vicinity of Iquitos. Only monoclonal samples are included. Each network node is a sample, with the edges representing IBD estimates. Nodes are color coded by collection year. An IBD threshold of 0.99 was used to generate networks. The Belén network shows evidence of clonal transmission in 2013 that is closely related to isolates collected in 2012 and 2014. In all five cities, there are networks connecting closely related isolates across years. **A** Belén, **B** Indiana, **C** Iquitos, **D** Punchana, and **E** San Juan Bautista.

Potential balancing selection in multiple RBC invasion ligands

In addition to determining the allele frequencies of antimalarial resistance ortholog markers, we also calculated Tajima's D in 100 bp bins with a 50 bp step size across loci encoding merozoite surface and secreted proteins that are involved in the red blood cell invasion that we tiled with the PvG MIP panel. These ligands are potential *P. vivax* vaccine targets (see Supplementary Methods). The strongest signatures of balancing selection (Tajima's $D > 2$), were detected in 13 of 20 genes analyzed, including apical membrane antigen 1 (*Pvama1*), *Pvdbp, Pvebp,* merozoite surface protein 1 (*Pvmsp1*), *Pvrbp1a, Pvrbp1b, Pvrbp2a, Pvrbp2b, Pvrbp2c,* tryptophan-rich antigen 111 (*Pvtrag11*),

Pvtrag19, Pvtrag22, Pvtrag34, and *Pvtrag36* (Fig. 5). The median value across all RBC invasion ligands was −0.577, while the minimum was −1.39 (*Pvrbp1a*) and the maximum was 3.14 (*Pvtrag11*). No Tajima's D values <−2, consistent with strong directional selection, were detected in any RBC invasion ligand. All polymorphisms detected in RBC invasion ligands are reported in Supplementary Data 3 and all Tajima's D values are shown in Supplementary Fig. 10.

Discussion

In this study, we demonstrate that the high-throughput MIP genotyping approach used to characterize the genomic epidemiology of *P.*
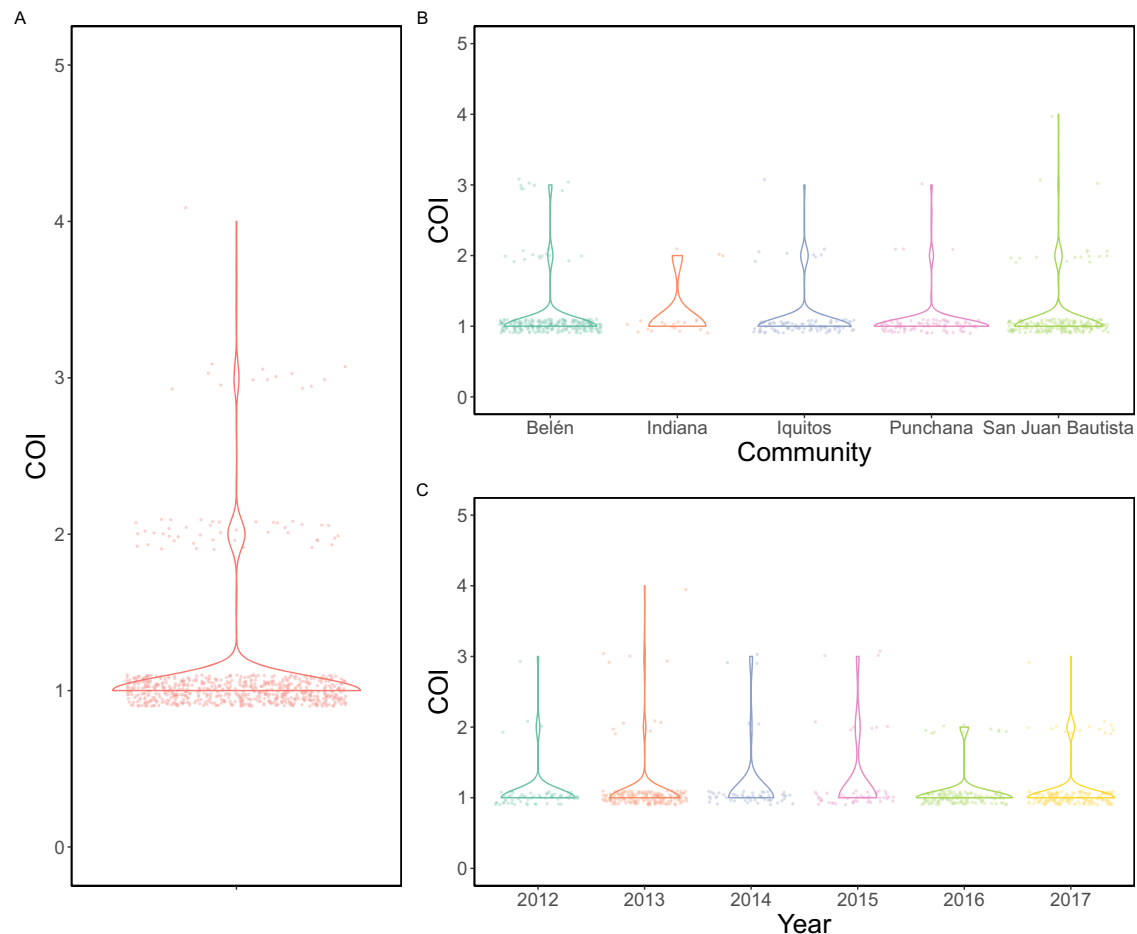
Fig. 3 | Violin plot depicting the complexity of infection (COI) of *P. vivax* isolates included in the analysis. A all *P. vivax* isolates analyzed. Each dot represents the mean COI estimate for a given isolate. The majority of isolates (91.4%, *n* = 637/697) are monoclonal, with 44 isolates (6.31%) containing two clones, 14 isolates (2.01%) containing three clones, and two isolates (0.287%) containing four clones. **B** *P. vivax* isolates from five communities with largest sample numbers. COI does not differ significantly by community. **C** *P. vivax* isolates from the same five communities ordered by collection year. COI does not differ significantly by year.

falciparum in Africa is also effective and informative for South American *P. vivax* parasite populations. Combined, our novel genotyping panels are the broadest genotyping panel currently available, working reliably on *P. vivax* isolates with parasitemias ≥100 parasites/µL to characterize population structure, relatedness, antimalarial resistance ortholog mutation frequency, and identify signatures of selection in candidate vaccine targets. By deploying them on samples from Peru, we were able to simultaneously detect highly related (IBD ≥ 0.99) samples within the same community across years, detect CNV in *Pvdbp* and multiple *Pvrbps*, estimate COI, identify mutations in multiple antimalarial resistance orthologs, and detect balancing selection in 13 vaccine candidate genes.

These new MIP panels build on both the success of *P. falciparum* MIPs and other existing *P. vivax* genotyping panels to enable a thorough analysis of multiple key elements of *P. vivax* biology. MIPs are more easily multiplexed than amplicon panels and incorporate unique molecular identifiers that enable correction of amplification bias, but are typically less sensitive than amplicons. Crucially, like the microhaplotype panels[42] and unlike the existing amplicon panels[18,38], our MIP panels should be deployable in any geographic context without modifications. In addition, unlike any existing panels, our MIP panels contain both geographically discriminating and neutral SNPs in addition to 21 vaccine-candidate genes and five genes that are antimalarial resistance orthologs. The 1,200 SNPs genotyped across the three panels are more than sufficient to generate robust IBD estimates[20,44].

This single set of panels therefore enables the simultaneous detection of population structure and transmission networks along with surveillance of potential variants of concern for antimalarial resistance and mutations that may decrease the effectiveness of vaccines.

While our MIP panels are efficient, reliable, and high-throughput, they nevertheless will require further refinement to better meet every need for a widely deployed *P. vivax* genotyping panel. As with the amplicon, SNP, and microhaplotype panels, MIPs require sophisticated molecular laboratories with sequencing capabilities, which are rare in malaria-endemic regions. In addition, the computational analyses necessary to analyze the genotyping results require substantial bioinformatic resources and expertise, which are also in short supply in malaria-endemic regions. Another limitation of both the MIP and amplicon panels is that the incorporation of antimalarial resistance orthologs is of questionable significance to malaria control programs since none of these mutations are validated. In addition, unlike the microhaplotype panels, the MIP panels have not been validated for use in treatment efficacy studies (TES), which are critical to distinguish reinfection from recrudescence and relapse. Finally, a major limitation of MIPs is inconsistent performance in low-parasitemia samples, which hampers highly-multiplex amplicon panels as well. Since *P. vivax* typically causes low-density infections, many of which are undiagnosed in the Amazon[36], we are currently limited to analyzing the high-density minority of infections, which could mean missing vital population data.
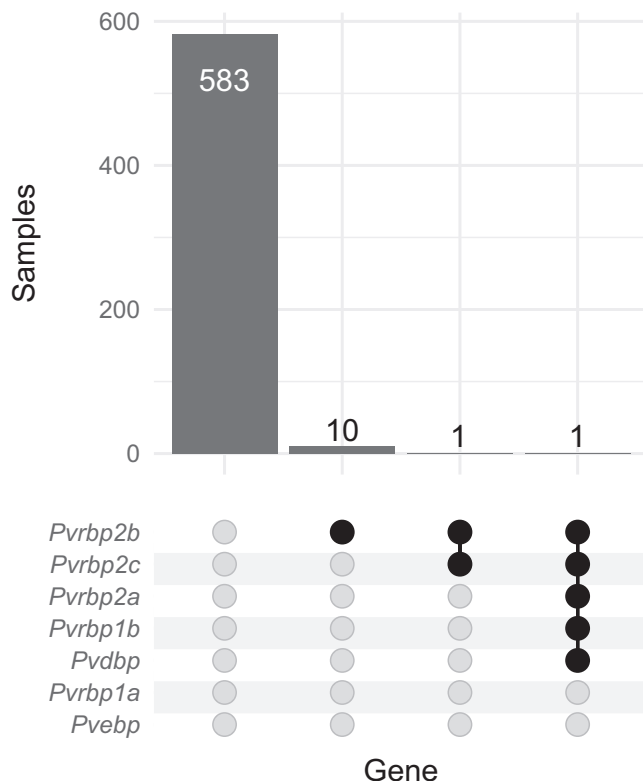
Fig. 4 | Upset plot showing the frequency of copy number variation (CNV) among monoclonal isolates. The majority of isolates (98.0%) do not exhibit CNV. CNV in *Pvrbp2b* was observed in twelve (2.01%) samples, with one isolate having three copies, as well as duplications in *Pvdbp*, *Pvrbp1b*, *Pvrbp2a*, and *Pvrbp2c*. One other isolate showed duplications in both *Pvrbp2b* and *Pvrbp2c*. No CNV was detected in either *Pvebp* or *Pvrbp1a*.

Table 2 | Frequencies of mutations in antimalarial resistance orthologs detected in the sample set

Gene	Mutation	Drug(s)	Study Frequency	Global Frequency[78]
Pvcrt	K9 duplication	CQ	11.0%	23.9%
Pvdhfr	F57L	SP, AQ	1.6%	5.5%
Pvdhfr	N50I	SP	0.0%	0.7%
Pvdhfr	S58R	SP, AQ	0.0%	6.2%
Pvdhfr	N117S	SP, AQ	3.3%	8.4%
Pvdhfr	I173L	SP	5.7%	1.1%
Pvdhps	G383A	SP	49.6%	29.3%
Pvdhps	553 G	SP	0.0%	17.8%
Pvmdr1	S513R	CQ	0.0%	18.2%
Pvmdr1	G698S	CQ	0.0%	22.9%
Pvmdr1	L908M	CQ	7.9%	2.2%
Pvmdr1	M958T	CQ	1.9%	0.5%
Pvmdr1	Y976F	CQ, SP, AQ	5.0%	48.9%
Pvmdr1	F1076L	CQ, SP, AQ	8.7%	22.0%

In addition to demonstrating the viability of MIPs to characterize and surveil *P. vivax* in a highly endemic region, we also identified multiple aspects of Peruvian *P. vivax* biology with potential implications for malaria control and elimination plans in the country. Within the immediate vicinity of Iquitos, we identified multiple clonal (IBD ≥ 0.99) networks with temporal-geographic connections. The Amazon region is known for its high clonality[18], which contributes to a high proportion of submicroscopic infections[34]. Most notably in

our study, the community of Belén appears to have experienced clonal *P. vivax* transmission from 2012-2014, which has also been observed elsewhere in the Peruvian Amazon[18]. Iquitos is known to be a source of *P. vivax* transmissions in the region, due to the highly mobile population[37,45]. The presence of clonal transmission in the suburb of Belén suggests that this strain likely spread elsewhere as well, meaning that monitoring transmission in the Iquitos region should provide a representative picture of most *P. vivax* within the Peruvian Amazon.

Several genomic epidemiology studies of *P. vivax* have occurred in Loreto with variable findings. Transmission patterns are complex, with a heterogenous spatio-temporal malaria distribution that is correlated with human mobility patterns, with travel history, and occupation influencing prevalence[17,34]. Urban areas have a similar proportion of monoclonal and polyclonal infections, but monoclonal infections predominate in rural areas and complexity of infection (COI) is higher in urban areas[45], although most isolates regardless of urbanization are monoclonal[17,18,37,46,47]. While heterozygosity (He) is similar in all localities, multiple studies have identified substantial differentiation by geography[17,18,46], which is also reflected in principal component analysis that separates rural and urban isolates[45]. These studies have also identified substantial substructure and clonal propagation in the parasite population[18], with Iquitos functioning as the main source of parasites for the rest of the region[45]. Clonal populations reflected by isolates from the same spatio-temporal context clustering with high identity by descent (IBD) have been identified[18]. However, a study of isolates collected between 2015 and 2019 identified three subpopulations but no geographic substructure[37].

CNV was rare in these samples and most frequently found in *Pvrbp2b*, although one isolate exhibited CNV in five genes. Although CNV has been previously observed in *Pvebp*[48], we did not detect any. CNV has previously been identified in *Pvrbp2a* and *Pvrbp2b*, with duplications detected in two Thai isolates[49]. However, duplications in *Pvrbp1b* and *Pvrbp2c* have not been previously described. The *Pvrbp* gene family plays an important role in invading reticulocytes[50–52], is immunogenic, and is a potential vaccine target. While there are multiple functional *Pvrbp* genes, the family also contains pseudogenes[50]. The *Pvrbp*s vary in their degree of polymorphism, e.g., *Pvrbp2c* is more polymorphic than *Pvrbp1*[49], and their different domains may be under differential selective pressure[49]. Previous analyses have not identified deviations from neutrality in the *Pvrbp*s[53,54].

We also detected a duplication in *Pvdbp* in one isolate. *Pvdbp* plays a critical role in red blood cell (RBC) invasion in Duffy antigen/receptor for chemokines (DARC)-positive individuals, who make up the vast majority of people living in areas with high *P. vivax* transmission[55]. As such, PvDBP is a promising blood-stage vaccine candidate[55]. However, studies have demonstrated low *Pvdbp* immunogenicity in the Amazon region, and *Pvdbp* is highly polymorphic[55]. *Pvdbp* copy number variation (CNV) has been detected globally in isolates from both DARC-positive and DARC-negative individuals[56,57]. The implications of the CNV we detected in 14 isolates are somewhat unclear, but may suggest that vaccines targeting DBP or the RBPs that do not account for this variability may have reduced efficacy[58].

While none of the mutations in antimalarial resistance orthologs in *P. vivax* are validated, they have been associated with resistance to chloroquine (CQ)[59,60], sulfadoxine-pyrimethamine (SP)[61–63], and other drugs[64,65]. Since CQ is still the first-line treatment for *P. vivax* in Peru and many other contexts[30], monitoring putatively resistance-conferring mutations to this drug is critical to maintaining efficacious treatment programs. Previous studies have identified mutations in chloroquine resistance transporter (*Pvcrt*), which have been associated with vivax CQ resistance[37], that were nearly at fixation[17]. However, since these mutations are in intronic regions, the functional significance is unclear[17,59]. Polymorphisms in multidrug resistance protein 1 (*Pvmdr1*) have been detected at heterogeneous levels

throughout the Peruvian Amazon[60], including in the CQ resistance-associated Y976F and F1076L mutations[18,60]. In our analysis, we find the *Pvcrt* K9 duplication at a relatively low level (11%), while both *Pvmdr1* Y976F and F1076L are lower frequency than the 15.2% found with amplicon sequencing elsewhere in Loreto[18].

SP is used as a partner drug in artemisinin combination therapy for *P. falciparum*. In addition, while SP is not a preferred treatment for *P. vivax*, it is sometimes given in areas where *P. falciparum* and *P. vivax* co-occur, although it was removed as a first-line treatment in Peru in 2001[66]. As with *Pvcrt*, previous studies have identified mutations in dihydrofolate reductase (*Pvdhfr*) and dihydropteroate synthase (*Pvdhps*) ranging in frequency from very low (1.3%) to fixation in Peru[17,18,37], which may have occurred while SP was still used for *P. falciparum* treatment in Peru[37]. The frequency of the *Pvdhps* A383G mutation is similar in our analysis (49.6%) to that found with amplicon sequencing (61.7%)[18]. We also found similar frequencies of the *Pvdhfr* S117N mutation (3.3% vs. 1.3%), but identified no isolates with the S58R mutation, in contrast to the relatively high frequency (54.3%) found in the other study[18]. Conversely, we identified the F57L mutation at a low frequency (1.6%), while the other study did not find it at all[18]. Our findings also contrast with data from PNG that found the F57L and S58R mutations to be consistently linked[61]. Although the clinical and operational significance of these mutations associated with reduced CQ and SP susceptibility is unclear, monitoring them may enable Peru's malaria elimination goals, e.g., by increasing surveillance intensity in regions where their frequency is rising.

We also demonstrate evidence of balancing selection in 13 RBC invasion ligands which are potential vaccine targets, including *Pvama1*, *Pvdbp*, *Pvebp*, *Pvmsp1*, four members of the *Pvrbp* family, and four members of the *Pvtrag* family. The biological aspects of balancing selection in these targets are discussed in the Supplementary Results and Discussion. While the PvG panel identified potential balancing selection in 13 genes, it is limited by our ability to successfully design MIPs to target the whole genes. For some genes, such as *Pvama1*, we were unable to design MIPs targeting the beginning of the gene, while for others, such as *Pvrbp2a*, there are multiple smaller regions within the gene that we were unable to target. Other genes, such as *Pvcsp*, contain repeats that are difficult to target with the short length of the capture design. A future redesign of the PvG panel may improve coverage within these genes of interest. Nonetheless, even with this limitation, we were able to find signatures of selection.

Our four MIP panels, along with existing amplicon and microhaplotype panels, are poised to begin replicating the successful genomic epidemiology studies of *P. falciparum* in *P. vivax*. While their performance is less reliable with lower parasitemia (a common feature of *P. vivax* infections), this shortcoming may be overcome by incorporating enrichment methods such as selective whole genome amplification prior to MIP capture. In this analysis of Peruvian isolates, we demonstrate their utility to simultaneously detect population structure, CNV, COI, mutation frequencies in antimalarial resistance orthologs, and signatures of selection. As malaria control programs move toward the elimination of *P. falciparum*, such a tool for *P. vivax* will be critical if indeed it expands to fill the niche left by *P. falciparum*.
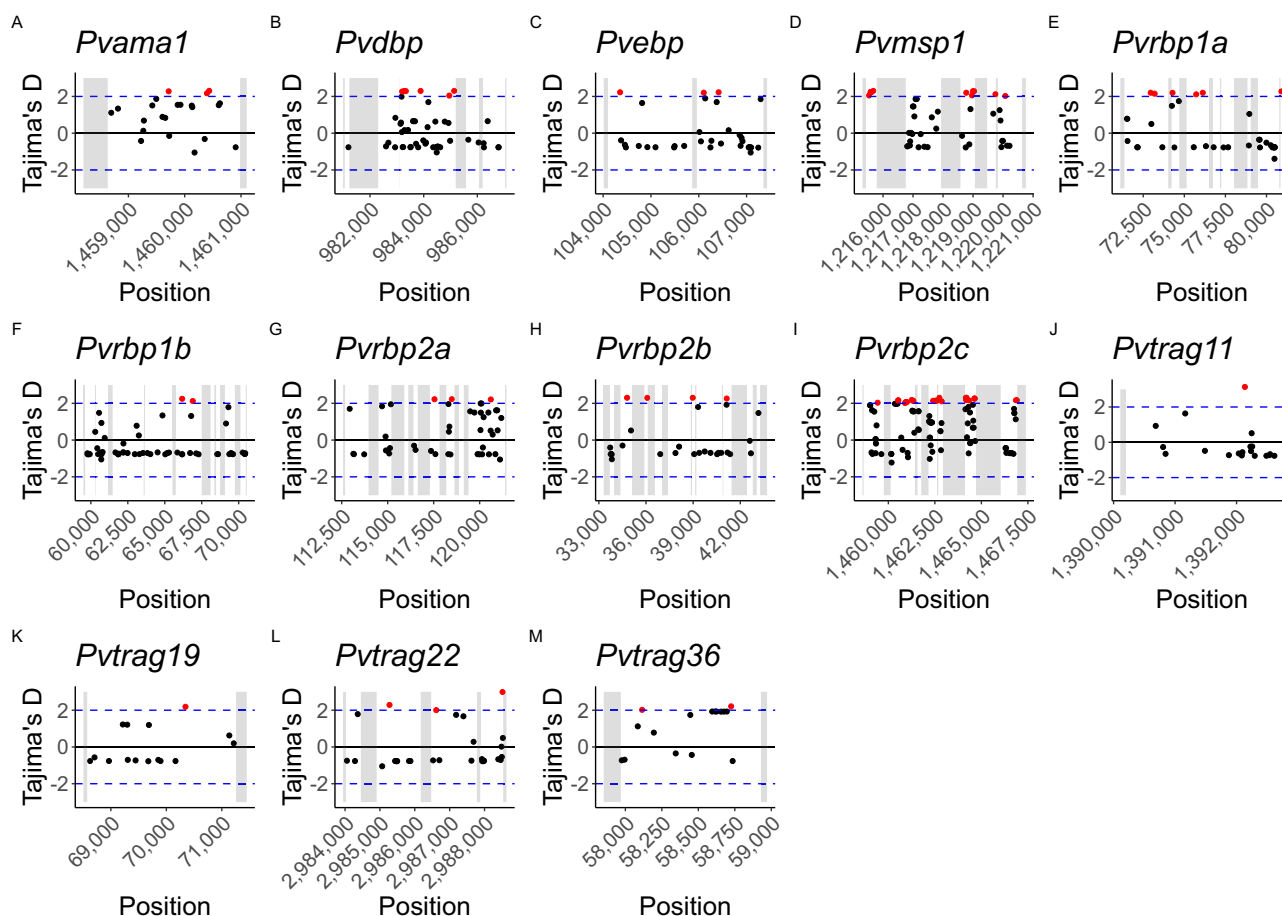

Fig. 5 | Scatter plots depicting Tajima's D values in 100 bp bins with a 50 bp step size across potential vaccine target genes. Gray boxes indicate gene regions that were not captured by MIPs. Only genes with Tajima's D values > 2, i.e., showing strong signatures of balancing selection, are depicted. Dots are placed in the midpoint of each bin, with values >|2| highlighted in red. **A** *Pvama1*, **B** *Pvdbp*, **C** *Pvebp*, **D** *Pvmsp1*, **E** *Pvrbp1a*, **F** *Pvrbp1b*, **G** *Pvrbp2a*, **H** *Pvrbp2b*, **I** *Pvrbp2c*, **J** *Pvtrag11*, **K** *Pvtrag19*, **L** *Pvtrag22*, **M** *Pvtrag36*.

Future MIP analyses of isolates from other geographical regions and/or collection years will likely enable a robust characterization of global and/or temporal trends in *P. vivax* biology, and refinement of the panel for low density infections.

Methods

Ethics approval

The protocol for this study was approved by the Institutional Review Board of the U.S Naval Medical Research Unit SOUTH (NAMRU SOUTH) in compliance with all applicable federal regulations governing the protection of human subjects (protocol NMRCD.2007.0004). Informed consent was obtained from all participants. For participants aged 8 to 17 years, written consent from parents or guardians and assent from participants were obtained.

Sample collection

Samples were collected between 2011 and 2018 as part of a malaria passive surveillance study conducted in Iquitos, the main city in the Peruvian Amazon Basin, and its surrounding areas in the Loreto region. Loreto consistently contributed 90–95% of all malaria cases reported in Peru during the study period (2011–2018) with an approximate distribution of *P. vivax* and *P. falciparum* of 80% and 20%, respectively. The mean API in Loreto region for vivax malaria during the study period was 33.0 cases/1000 population with a range from 9.2 (2011) to 48.9 (2014) and malaria incidence was mainly homogeneous in all the study sites (https://www.dge.gob.pe/portalnuevo/publicaciones/boletines-epidemiologicos/).

We enrolled individuals older than 1 year old, with fever or history of fever during the previous 72 h in the absence of another cause of fever. Enrolled participants provided a written and/or assent form. We collected clinical, demographic and epidemiological information[67] and up to 8 milliliters of venous whole blood. Two thin and thick smears were prepared and stained with 10% Giemsa for each participant. Slides were read independently by two dedicated microscopists; discordant results were solved by a third microscopist. All participants were microscopy positive to *Plasmodium vivax* with a parasitemia range of 12 to 141,373 par/ul during the study period[68,69].

Sample processing and sequencing

Dried blood spots were extracted using the Chelex method, as previously described[20,43]. Briefly, DNA was extracted from three 6-mm punches, which represent approximately 25 µL of blood. Extracted DNA was used for molecular inversion probe (MIP) captures, which were sequenced as previously described for *P. falciparum*[20,43]. The human MIP protocol was used instead of the parasite MIP protocol due to the GC richness of the *P. vivax* genome. When comparing the parasite and human MIP protocols, the differing buffer and enzyme in the respective capture and PCR steps, in addition to the increased temperatures in the PCR cycling conditions allow for more efficient amplification. Due to the difficulty in capturing the *P. vivax* genome from human samples, the amount of DNA input into the capture reaction was increased from 3 µL to 5 µL and the volume of water in the capture master mix was decreased, in exchange. Four new MIP panels were designed for *P. vivax* based on the PvP01 genome assembly using variation data from 918 published *P. vivax* genomes (Supplementary Table 2): 1) PvG, a tiled gene panel designed to densely genotype genes of interest, including antimalarial resistance ortholog markers and invasion ligands that are potential vaccine targets (Supplementary Table 3); 2) PvFST, a panel designed to capture geographically discriminating SNPs between and within continents; 3) PvMAF5, a panel designed to capture rare neutral (non-coding) SNPs (minor allele frequency <0.05); and 4) PvMAF40, a panel designed to capture common neutral SNPs (MAF > 0.4). The details of the MIP design scheme are given in Supplementary Methods and probe sequences are provided in Supplementary Data 4–7. The generated libraries were sequenced

on an Illumina Nextseq 550 platform using 150 bp paired-end sequencing with dual indexing using Nextseq 500/550 Mid-output Kit v2 in the case of the gene panels and 75 bp paired-end sequencing in the case of the SNP panels.

MIP variant calling and filtering

MIP sequencing data was processed using *MIPWrangler* within *MIP-Tools* (https://github.com/bailey-lab/MIPTools), which normalizes read counts based on unique molecular identifiers (UMIs) that are incorporated in the MIPs. Variant calling for all panels was performed using *FreeBayes*[70] within *MIPTools*. For the SNP panels, we specified a minimum UMI depth of 3, a within-sample allele frequency threshold of 0.5, and a minimum alternate read count of 2 to obtain 616 variant SNPs across 760 samples (PvFST), 437 variant SNPs across 763 samples (PvMAF5), and 377 variant SNPs across 750 samples (PvMAF40). *Bcftools*[71] v1.9 was used to filter the resulting VCF to include only targeted SNPs. The filtered VCFs were imported into R using vcfR[68] (version 1.14), where they were assessed for sample and variant missingness. Samples with >50% missingness on any panel were excluded from the VCFs, which were then sorted and concatenated with *bcftools*. Finally, *bcftools* was used to subset the resulting VCF to include only biallelic SNPs and exclude SNPs with >50% missingness. The tiled gene panel was processed in a similar fashion, yielding 13,927 variant SNPs across 740 samples.

Population structure analysis

MIPanalyzer (https://github.com/mrc-ide/MIPanalyzer) was used to assess parasite relatedness and detect population structure. We first performed PCA of all samples. Given the unwieldy size and lightly structured nature of the dataset, we subsetted samples by city for plotting, with samples coded by collection year. To more thoroughly determine population structure and relatedness, we calculated identity by descent (IBD) of monoclonal isolates using maximum likelihood estimation. IBD networks connecting samples from each city with IBD ≥ 0.99 were generated using *ggraph*[69] (version 2.1).

Complexity of Infection (COI), i.e., the number of parasite clones present in each sample, was calculated with *McCOILR* (https://ojwatson.github.io/McCOILR/index.html, version 1.3), an R implementation of *THE REAL McCOIL* (v2) categorical method[72].

Analysis of antimalarial resistance orthologs and potential vaccine candidates

All amino acid changes detected with the tiled gene panel and their respective frequencies were summarized using miplicoRn (https://github.com/bailey-lab/miplicorn, v0.2.1) and flextable[73] (v0.9). We prioritized the analysis of mutations that have previously been associated with drug resistance.

To detect potential selection on the potential vaccine candidates, we used vcf-kit (https://github.com/AndersenLab/VCF-kit) to calculate Tajima's D in 100 bp sliding windows with a 50 bp step size.

Prior to CNV calculation, samples and MIPs were filtered for missingness as follows: samples with <5 unique molecular identifier (UMI) counts for more than 50% of MIPs were excluded, as were MIPs with <5 UMI counts for more than 50% of samples sequenced. For CNV calculation at loci of interest, the goal was to mitigate technical variations in the read coverage while maintaining true biological modifications. To correct the raw read data for systematic errors, z-score normalization of UMI coverage of each MIP was performed by accounting for potential sample and locus effects. Individual UMI coverages were transformed into z-score as $z_{ij} = \frac{|UMI_{ij} - \underline{X}_{UMI}|}{\sigma_{UMI}}$, where UMI_{ij} is the UMI coverage of a given MIP (or microhaplotype) i within a sample j and $\underline{X}_{UMI}$ and σ_{UMI} represent the mean of per sample means and standard deviations of UMI coverage, respectively. For each sample, the mean and standard deviation of UMI coverages were initially calculated across all loci included in the MIP panel.

Consequently, this step generated distributions of means and standard deviations for the entire sample set in which we computed averages to obtain the parameters $\underline{X}_{UMI}$ and σ_{UMI}. The relative read abundance of a candidate gene within a sample, which corresponds to CNV, was estimated by dividing the z-score (z_{ij}) of each MIP targeting this gene by the averaged z-score of all the other genes captured that are known not to be subject to CNV. As this method deals with sample and locus-derived errors and not batch effects, it was only applied to calculate CNV in samples that are sequenced together in the same run. The approach was previously validated in laboratory strains of *P. falciparum*, including 3D7, 7G8, and Dd2, in independent runs. Unlike 3D7 and 7G8, Dd2, the positive control with multiple copies of *mdr1*[74,75] consistently showed CNV > 2 for this gene. We have also previously used MIP data to generate precise copy number estimates[76].

Reporting summary

Further information on research design is available in Nature Portfolio Reporting Summary linked to this article.

Data availability

Parasite sequence data is available through SRA (BioProject ID PRJNA1117913). Demographic and clinical data for the patients sampled in this study are available on UNC Dataverse (https://doi.org/10.15139/S3/H0YPJD)[67].

Code availability

Code used for analysis and MIP probe sequences as well as Excel-formatted files for probe orders are available from GitHub at https://github.com/IDEELResearch/PvMIPs/[77].

References

1. Battle, K. E. & Baird, J. K. The global burden of *Plasmodium vivax* malaria is obscure and insidious. *PLoS Med.* **18**, e1003799 (2021).
2. World Health Organization. *World Malaria Report 2022* (WHO, 2022).
3. Commons, R. J., Simpson, J. A., Watson, J., White, N. J. & Price, R. N. Estimating the proportion of plasmodium vivax recurrences caused by relapse: a systematic review and meta-analysis. *Am. J. Trop. Med. Hyg.* **103**, 1094–1099 (2020).
4. Bousema, T. & Drakeley, C. Epidemiology and infectivity of *Plasmodium falciparum* and *Plasmodium vivax* Gametocytes in relation to malaria control and elimination. *Clin. Microbiol Rev.* **24**, 377–410 (2011).
5. Tarimo, B. B. et al. Seasonality and transmissibility of *Plasmodium ovale* in Bagamoyo District, Tanzania. *Parasit. Vectors* **15**, 56 (2022).
6. Quah, Y. W. et al. Molecular epidemiology of residual *Plasmodium vivax* transmission in a paediatric cohort in Solomon Islands. *Malar. J.* **18**, 106 (2019).
7. Brazeau, N. F. et al. The epidemiology of *Plasmodium vivax* among adults in the Democratic Republic of the Congo. *Nat. Commun.* **12**, 1–10 (2021).
8. Zhong, D. et al. Multiplicity and molecular epidemiology of *Plasmodium vivax* and *Plasmodium falciparum* infections in East Africa. *Malar. J.* **17**, 185 (2018).
9. Pearson, R. D. et al. Genomic analysis of local variation and recent evolution in *Plasmodium vivax*. *Nat. Genet.* **48**, 959–964 (2016).
10. Popovici, J. et al. Genomic analyses reveal the common occurrence and complexity of *Plasmodium vivax* relapses in Cambodia. *mBio* **9**, e01888–17 (2018).
11. Auburn, S. et al. Genomic analysis of *Plasmodium vivax* in southern Ethiopia reveals selective pressures in multiple parasite mechanisms. *J. Infect. Dis.* **220**, 1738–1749 (2019).
12. Valdivia, H. O. et al. Genomic surveillance of *Plasmodium falciparum* and *Plasmodium vivax* cases at the University Hospital in Tegucigalpa, Honduras. *Sci. Rep.* **10**, 20975 (2020).
13. Gartner, V. et al. Genomic insights into Plasmodium vivax population structure and diversity in central Africa. *Malar. J.* **23**, 1–9 (2024).
14. Brashear, A. M. et al. Population genomics identifies a distinct *Plasmodium vivax* population on the China-Myanmar border of Southeast Asia. *PLoS Negl. Trop. Dis.* **14**, e0008506 (2020).
15. 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**, 100420 (2023).
16. Auburn, S. et al. Genomic analysis of a pre-elimination Malaysian Plasmodium vivax population reveals selective pressures and changing transmission dynamics. *Nat. Commun.* **9**, 1–12 (2018).
17. Cowell, A. N., Valdivia, H. O., Bishop, D. K. & Winzeler, E. A. Exploration of Plasmodium vivax transmission dynamics and recurrent infections in the Peruvian Amazon using whole genome sequencing. *Genome Med.* **10**, 1–12 (2018).
18. Kattenberg, J. H. et al. *Plasmodium vivax* genomic surveillance in the Peruvian Amazon with Pv AmpliSeq assay. *PLoS Negl. Trop. Dis.* **18**, e0011879 (2024).
19. Moser, K. A. et al. Describing the current status of *Plasmodium falciparum* population structure and drug resistance within mainland Tanzania using molecular inversion probes. *Mol. Ecol.* **30**, 100–113 (2021).
20. Verity, R. et al. The impact of antimalarial resistance on the genetic structure of *Plasmodium falciparum* in the DRC. *Nat. Commun.* **11**, 2107 (2020).
21. Connelly, S. V. et al. Strong isolation by distance and evidence of population microstructure reflect ongoing Plasmodium falciparum transmission in Zanzibar. *Elife* **12**, RP90173 (2023).
22. Mensah, B. A. et al. Antimalarial drug resistance profiling of Plasmodium falciparum infections in Ghana using molecular inversion probes and next-generation sequencing. *Antimicrob. Agents Chemother.* **64**, e01423–19 (2020).
23. Fola, A. A. et al. Plasmodium falciparum resistant to artemisinin and diagnostics have emerged in Ethiopia. Nat. Microbiol. **2023**, 1–9 https://doi.org/10.1038/S41564-023-01461-4 (2023).
24. MalariaGEN Plasmodium falciparum Community Project. Genomic epidemiology of artemisinin resistant malaria. *Elife* **5**, e08714 (2016).
25. Neafsey, D. E. et al. The malaria parasite *Plasmodium vivax* exhibits greater genetic diversity than *Plasmodium falciparum*. *Nat. Genet* **44**, 1046–1050 (2012).
26. Huang, F. et al. Genomic epidemiology of antimalarial drug resistance in *Plasmodium falciparum* in Southern China. *Front Cell Infect. Microbiol* **10**, 610985 (2021).
27. Coonahan, E. et al. Whole-genome surveillance identifies markers of Plasmodium falciparum drug resistance and novel genomic regions under selection in Mozambique. *mBio* **14**, e0176823 (2023).
28. Talundzic, E. et al. Molecular epidemiology of Plasmodium falciparum kelch13 mutations in Senegal determined by using targeted amplicon deep sequencing. *Antimicrob. Agents Chemother.* **61**, e02116 (2017).
29. Feleke, S. M. et al. Plasmodium falciparum is evolving to escape malaria rapid diagnostic tests in Ethiopia. *Nat. Microbiol.* **6**, 1289–1299 (2021).
30. World Health Organization. *World Malaria Report 2023*. (2023).
31. Parker, B. S. et al. Hyperendemic malaria transmission in areas of occupation-related travel in the Peruvian Amazon. *Malar. J.* **12**, 1–15 (2013).
32. World Health Organization. *World Malaria Report 2017*. (World Health Organization, 2017).
33. Sanchez-Castro, E. E. et al. Health and economic burden due to malaria in Peru over 30 years (1990–2019): Findings from the global burden of diseases study 2019. *Lancet Reg. Health - Am.* **15**, 100347 (2022).

34. Carrasco-Escobar, G. et al. Micro-epidemiology and spatial heterogeneity of P. vivax parasitaemia in riverine communities of the Peruvian Amazon: a multilevel analysis. *Sci. Rep.* **7**, 1–17 (2017).

35. Carrasco-Escobar, G. et al. High prevalence of very-low Plasmodium falciparum and Plasmodium vivax parasitaemia carriers in the Peruvian Amazon: insights into local and occupational mobility-related transmission. *Malar. J.* **16**, 1–15 (2017).

36. Ferreira, M. U. et al. Relative contribution of low-density and asymptomatic infections to Plasmodium vivax transmission in the Amazon: pooled analysis of individual participant data from population-based cross-sectional surveys. *Lancet Reg. Health - Am.* **9**, 100169 (2022).

37. Villena, F. E. et al. Drug resistance and population structure of Plasmodium falciparum and Plasmodium vivax in the Peruvian Amazon. *Sci. Rep.* **12**, 1–11 (2022).

38. Kattenberg, J. H. et al. Novel highly-multiplexed AmpliSeq targeted assay for Plasmodium vivax genetic surveillance use cases at multiple geographical scales. *Front. Cell Infect. Microbiol.* **12**, 1–24 (2022).

39. Baniecki, M. L. et al. Development of a single nucleotide polymorphism barcode to genotype Plasmodium vivax infections. *PLoS Negl. Trop. Dis.* **9**, e0003539 (2015).

40. Trimarsanto, H. et al. A molecular barcode and web-based data analysis tool to identify imported Plasmodium vivax malaria. *Commun. Biol.* **5**, 1–10 (2022).

41. Fola, A. A. et al. SNP barcodes provide higher resolution than microsatellite markers to measure *Plasmodium vivax* population genetics. *Malar. J.* **19**, 375 (2020).

42. Siegel, S. V. et al. Lineage-informative microhaplotypes for recurrence classification and spatio-temporal surveillance of Plasmodium vivax malaria parasites. *Nat. Commun. 2024 15:1* **15**, 1–16 (2024).

43. Aydemir, O. et al. Drug-resistance and population structure of *Plasmodium falciparum* across the Democratic Republic of Congo using high-throughput molecular inversion probes. *J. Infect. Dis.* **218**, 946–955 (2018).

44. Taylor, A. R., Jacob, P. E., Neafsey, D. E. & Buckee, C. O. Estimating relatedness between malaria parasites. *Genetics* **212**, 1337–1351 (2019).

45. Delgado-Ratto, C. et al. Population genetics of Plasmodium vivax in the Peruvian Amazon. *PLoS Negl. Trop. Dis.* **10**, e0004376 (2016).

46. Delgado-Ratto, C. et al. Population structure and spatio-temporal transmission dynamics of Plasmodium vivax after radical cure treatment in a rural village of the Peruvian Amazon. *Malar. J.* **13**, 1–12 (2014).

47. Manrique, P. et al. Microsatellite analysis reveals connectivity among geographically distant transmission zones of Plasmodium vivax in the Peruvian Amazon: a critical barrier to regional malaria elimination. *PLoS Negl. Trop. Dis.* **13**, e0007876 (2019).

48. Roesch, C. et al. Genetic diversity in two *Plasmodium vivax* protein ligands for reticulocyte invasion. *PLoS Negl. Trop. Dis.* **12**, e0006555 (2018).

49. Kosaisavee, V. et al. Genetic diversity in new members of the reticulocyte binding protein family in Thai Plasmodium vivax isolates. *PLoS One* **7**, 1–6 (2012).

50. Chan, L.-J. et al. Plasmodium vivax reticulocyte binding proteins for invasion into reticulocytes. *Cell Microbiol* **22**, e13110 (2020).

51. Han, J. H. et al. Identification of a reticulocyte-specific binding domain of Plasmodium vivax reticulocyte-binding protein 1 that is homologous to the PfRh4 erythrocyte-binding domain. *Sci. Rep.* **6**, 1–12 (2016).

52. Gupta, E. D. et al. Naturally acquired human antibodies against reticulocyte-binding domains of Plasmodium vivax proteins, PvRBP2c and PvRBP1a, exhibit binding-inhibitory activity. *J. Infect. Dis.* **215**, 1558–1568 (2017).

53. Nourani, L., Abouie, A., Id, M., Zakeri, S. & Djadid, N. D. Untangling population structure and genetic diversity of reticulocyte binding protein 2b (PvRBP2b) erythrocytic stage vaccine candidate in worldwide Plasmodium vivax isolates. https://doi.org/10.1371/journal.pone.0266067 (2022).

54. Núñez, A., Ntumngia, F. B., Guerra, Y., Adams, J. H. & Sáenz, F. E. Genetic diversity and natural selection of Plasmodium vivax reticulocyte invasion genes in Ecuador. *Malar. J.* **22**, 1–17 (2023).

55. de Sousa, T. N., Kano, F. S., de Brito, C. F. A. & Carvalho, L. H. The Duffy binding protein as a key target for a *Plasmodium vivax* vaccine: lessons from the Brazilian Amazon. *Mem. Inst. Oswaldo Cruz* **109**, 608–617 (2014).

56. Menard, D. et al. Whole genome sequencing of field isolates reveals a common duplication of the duffy binding protein gene in Malagasy Plasmodium vivax strains. *PLoS Negl. Trop. Dis.* **7**, e2489 (2013).

57. Hostetler, J. B. et al. Independent origin and global distribution of distinct Plasmodium vivax duffy binding protein gene duplications. *PLoS Negl. Trop. Dis.* **10**, e0005091 (2016).

58. Benavente, E. D. et al. Global genetic diversity of var2csa in Plasmodium falciparum with implications for malaria in pregnancy and vaccine development. *Sci. Rep. 2018 8:1* **8**, 1–8 (2018).

59. Sá, J. M. et al. Plasmodium vivax chloroquine resistance links to pvcrt transcription in a genetic cross. *Nat. Commun.* **10**, 1–10 (2019).

60. Villena, F. E. et al. Molecular surveillance of the Plasmodium vivax multidrug resistance 1 gene in Peru between 2006 and 2015. *Malar. J.* **19**, 1–10 (2020).

61. Marfurt, J. et al. Molecular markers of in vivo plasmodium vivax resistance to amodiaquine plus sulfadoxine-pyrimethamine: mutations in pvdhfr and pvmdr1. *J. Infect. Dis.* **198**, 409–417 (2008).

62. Auliff, A. et al. Amino acid mutations in Plasmodium vivax DHFR and DHPS from several geographical regions and susceptibility to antifolate drugs. *Am. J. Trop. Med. Hyg.* **75**, 617–621 (2006).

63. Imwong, M. et al. Association of genetic mutations in Plasmodium vivax dhfr with resistance to sulfadoxine-pyrimethamine: geographical and clinical correlates. *Antimicrob. Agents Chemother.* **45**, 3122–3127 (2001).

64. Khim, N. et al. Effects of mefloquine use on plasmodium vivax multidrug resistance. *Emerg. Infect. Dis.* **20**, 1637–1644 (2014).

65. Li, J. et al. Ex vivo susceptibilities of Plasmodium vivax isolates from the China-Myanmar border to antimalarial drugs and association with polymorphisms in Pvmdr1 and Pvcrt-o genes. *PLoS Negl. Trop. Dis.* **14**, e0008255 (2020).

66. Ruebush, T. K., Neyra, D. & Cabezas, C. Modifying national malaria treatment policies in Peru. *J. Public Health Policy* **25**, 328–345 (2004).

67. Popkin-Hall, Z. R., et al. "Study Participant Data", https://doi.org/10.15139/S3/HOYPJD, UNC Dataverse, V1, UNF:6:QRA39Ul896HR5Eb-qEG50MA = = [fileUNF] (2024).

68. Knaus, B. J. & Grünwald, N. J. vcfr: a package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resour.* **17**, 44–53 (2017).

69. Pedersen, T. L. ggraph: An Implementation of Grammar of Graphics for Graphs and Networks. R package version 2.1 (2022).

70. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing (2012).

71. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, 1–4 (2021).

72. 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).

73. Gohel, D. & Skintzos, P. flextable: functions for tabular reporting. R package (2023).

74. Ferreira, I. D., Rosário, V. Edo & Cravo, P. V. L. Real-time quantitative PCR with SYBR Green I detection for estimating copy numbers of

nine drug resistance candidate genes in Plasmodium falciparum. *Malar. J.* **5**, 1 (2006).

75. Srisutham, S., Suwannasin, K., Sugaram, R., Dondorp, A. M. & Imwong, M. Measurement of gene amplifications related to drug resistance in Plasmodium falciparum using droplet digital PCR. *Malar. J.* **20**, 1–12 (2021).

76. Tumwebaze, P. K. et al. Drug susceptibility of Plasmodium falciparum in eastern Uganda: a longitudinal phenotypic and genotypic study. *Lancet Microbe* **2**, e441–e449 (2021).

77. Popkin-Hall, Z. R. et al. High-throughput genotyping of *Plasmodium vivax* in the Peruvian Amazon via molecular inversion probes, PvMIPs, https://doi.org/10.5281/zenodo.14042814 (2024).

78. MalariaGEN et al. An open dataset of *Plasmodium vivax* genome variation in 1895 worldwide samples. *Wellcome Open Res.* **7**, 136 (2022).

Acknowledgements

Author contributions

Z.R.P.H., J.A.B., J.J.J., and H.O.V. conceptualized and designed the study. J.F.S., D.L.P., and H.O.V. supervised sample collection. Z.R.P.H. designed the MIP panels with input from K.N., A.A.F., D.J.G., J.A.B., and J.J.J. using programs designed by Ö.A. Z.R.P.H., R.C., A.A.F., and D.J.G. performed laboratory analyses, data collection, and cleaning. Z.R.P.H., K.N., A.S., and I.K. performed computational analyses. Z.R.P.H., K.N., A.A.F., J.A.B., J.J.J., and H.O.V. performed data analysis and interpretation. Z.R.P.H., K.N., J.A.B., J.J.J., and H.O.V. wrote the manuscript draft. All authors reviewed and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-54731-y.

Correspondence and requests for materials should be addressed to Zachary R. Popkin-Hall.

Peer review information *Nature Communications* thanks Nathalia Rammé Medeiros de Albuquerque and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints