Selective sweep suggests transcriptional regulation may underlie *Plasmodium vivax* resilience to malaria control measures in Cambodia

Christian M. Parobek[a], Jessica T. Lin[b], David L. Saunders[c], Eric J. Barnett[d], Chanthap Lon[c,e], Charlotte A. Lanteri[c], Sujata Balasubramanian[b], Nicholas Brazeau[f], Derrick K. DeConti[g], Deen L. Garba[b], Steven R. Meshnick[f], Michele D. Spring[c], Char Meng Chuor[h], Jeffrey A. Bailey[g,i], and Jonathan J. Juliano[a,b,f,1]

[a]Curriculum in Genetics and Molecular Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; [b]Division of Infectious Diseases, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; [c]US Army Medical Component, Armed Forces Research Institute of Medical Sciences, Bangkok, Thailand 10400; [d]Upstate Medical University, State University of New York, Syracuse, NY 13210; [e]US Army Medical Component, Armed Forces Research Institute of Medical Sciences, Phnom Penh, Cambodia; [f]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; [g]Program in Bioinformatics and Integrative Biology, University of Massachusetts School of Medicine, Worcester, MA 01655; [h]National Center for Parasitology, Entomology and Malaria Control, Phnom Penh, Cambodia; and [i]Division of Transfusion Medicine, University of Massachusetts School of Medicine, Worcester, MA 01655

Edited by Francisco J. Ayala, University of California, Irvine, CA, and approved November 3, 2016 (received for review June 2, 2016)

Cambodia, in which both *Plasmodium vivax* and *Plasmodium falciparum* are endemic, has been the focus of numerous malaria-control interventions, resulting in a marked decline in overall malaria incidence. Despite this decline, the number of *P. vivax* cases has actually increased. To understand better the factors underlying this resilience, we compared the genetic responses of the two species to recent selective pressures. We sequenced and studied the genomes of 70 *P. vivax* and 80 *P. falciparum* isolates collected between 2009 and 2013. We found that although *P. falciparum* has undergone population fracturing, the coendemic *P. vivax* population has grown undisrupted, resulting in a larger effective population size, no discernable population structure, and frequent multiclonal infections. Signatures of selection suggest recent, species-specific evolutionary differences. Particularly, in contrast to *P. falciparum*, *P. vivax* transcription factors, chromatin modifiers, and histone deacetylases have undergone strong directional selection, including a particularly strong selective sweep at an AP2 transcription factor. Together, our findings point to different population-level adaptive mechanisms used by *P. vivax* and *P. falciparum* parasites. Although population substructuring in *P. falciparum* has resulted in clonal outgrowths of resistant parasites, *P. vivax* may use a nuanced transcriptional regulatory approach to population maintenance, enabling it to preserve a larger, more diverse population better suited to facing selective threats. We conclude that transcriptional control may underlie *P. vivax*'s resilience to malaria control measures. Novel strategies to target such processes are likely required to eradicate *P. vivax* and achieve malaria elimination.

Plasmodium | malaria | vivax | transcription | genome

During the last decade, western Cambodia has been the focus of numerous and multimodal interventions to control the spread of artemisinin-resistant *Plasmodium falciparum* (1, 2). Such interventions, including increased vector control, increased surveillance, and improved access to quality artemisinin-combination therapy (ACT), would be expected to curtail coendemic *Plasmodium vivax* as well. However, even as *P. falciparum* infections in Cambodia decreased by 81% between 2009 and 2013, *P. vivax* cases have increased, making it the predominant species in the Mekong region (3–6). This scenario, repeated in Brazil and other areas of coendemicity, has led to growing awareness that *P. vivax*, although infecting the same populations and transmitted by the same mosquito vectors, will likely be the more challenging species to eradicate (6–9). In this study, we use population genomics to gain insight into the evolutionary factors underlying *P. vivax*'s resilience to malaria control measures.

Population genetic studies have previously hinted at the resilience of *P. vivax* populations in comparison with *P. falciparum*. Studies of microsatellites and highly variable antigens of sympatric *P. vivax* and *P. falciparum* populations in Southeast Asia and the Southwest Pacific have consistently shown *P. vivax* populations to be more diverse, with a higher effective population size (N_{eff}), more stable transmission, and increased gene flow between geographic islands, whereas *P. falciparum* populations tend to be clonal with episodic transmission and structure-by-geography (10–15). We hypothesized that the species have evolved disparate responses to selective pressures and that genomic studies of sympatric *Plasmodium* sp. populations would highlight key differences in their population structures, demographic histories, and genomic selective signatures, helping elucidate the basis for these observed differences.

To understand the genome-wide species-specific patterns of selection in sympatric *P. vivax* ($n = 70$) and *P. falciparum* ($n = 80$) populations in Cambodia, we conducted whole-genome sequencing of coendemic parasites sampled between 2009 and 2013 at a primary site and two satellite sites in western Cambodia (Fig. S1).

Significance

In Cambodia, where *Plasmodium vivax* and *Plasmodium falciparum* are coendemic and intense multimodal malaria-control interventions have reduced malaria incidence, *P. vivax* malaria has proven relatively resistant to such measures. We performed comparative genomic analyses of 150 *P. vivax* and *P. falciparum* isolates to determine whether different evolutionary strategies might underlie this species-specific resilience. Demographic modeling and tests of selection show that, in contrast to *P. falciparum*, *P. vivax* has experienced uninterrupted growth and positive selection at multiple loci encoding transcriptional regulators. In particular, a strong selective sweep involving an AP2 transcription factor suggests that *P. vivax* may use nuanced transcriptional approaches to population maintenance. Better understanding of *P. vivax* transcriptional regulation may lead to improved tools to achieve elimination.

Author contributions: C.M.P., J.T.L., D.L.S., C.L., C.A.L., S.B., N.B., M.D.S., C.M.C., J.A.B., and J.J.J. designed research; C.M.P., J.T.L., D.L.S., E.J.B., C.L., C.A.L., S.B., N.B., D.K.D., D.L.G., M.D.S., and J.J.J. performed research; C.M.C. contributed new reagents/analytic tools; C.M.P., J.T.L., D.L.S., E.J.B., S.R.M., J.A.B., and J.J.J. analyzed data; and C.M.P., J.T.L., D.L.S., C.L., S.R.M., J.A.B., and J.J.J. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the National Center for Biotechnology Information Sequence Read Archive (SRA) database. For a list of SRA accession numbers, see Table S9.

[1]To whom correspondence should be addressed. Email: jjuliano@med.unc.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1608828113/-/DCSupplemental.

These sites were designated as "zone 2" during the recent artemisinin-resistance containment campaign and as such were subject to intensified malaria control efforts (16). We initially assessed the relative diversity, within-host complexity, population structure, and demographic histories of the two parasite populations, finding that the *P. vivax* population remains less structured, more diverse, and more rapidly expanding than the sympatric *P. falciparum* population. We then evaluated genome-wide signatures of selection in both populations using haplotype-based tests of directional selection, allele frequency-based tests of selection, and copy-number analysis. Differences in genomic loci under directional selection in the two populations highlight different mechanistic responses to selective pressures, suggesting that more nuanced transcriptional control may underlie the resilience of the *P. vivax* population.

Results

Sequencing Sympatric *P. vivax* and *P. falciparum* Populations. Whole-genome sequencing identified 61,448 high-quality *P. vivax* SNPs and 6,734 *P. falciparum* SNPs from 70 and 80 samples, respectively. All *P. vivax* isolates had fivefold or greater coverage in ≥99% of coding regions, whereas all *P. falciparum* isolates had fivefold or greater coverage in ≥94% of coding regions. Additional information about sequence quality and coverage is provided in *SI Materials and Methods*, *SI Sequencing Sympatric* P. vivax *and* P. falciparum *Populations*.

***P. vivax* Infections Have Higher Within-Host Diversity than Sympatric *P. falciparum* Infections.** Because *Plasmodia* infections are frequently multiclonal, we investigated the extent of multiclonality among our sequenced field isolates. When applied to whole-genome sequencing data, the F_{WS} statistic accurately predicts polyclonality (17). *P. vivax* infections were more polyclonal (defined as F_{WS} <0.95) than *P. falciparum* infections ($P < 0.0001$; Fisher's exact test), a finding that remained unchanged after bootstrapping to account for the difference in the number of SNPs identified in the two species (Fig. 1*A*). Using the accepted standard of F_{WS} <0.95 as the marker of a multiclonal infection, 60% of *P. vivax* isolates and 22.5% of *P. falciparum* isolates in our cohort were polyclonal (Fig. 1*B*). To confirm the reliability of this method for identifying polyclonal infections, we conducted amplicon deep sequencing of *P. vivax* merozoite surface protein 1 (*pvmsp1*) in 47 isolates, finding a high degree of agreement with F_{WS} (*SI Materials and Methods*, *SI* P. vivax *Infections Have Higher Within-Host Diversity than Sympatric* P. falciparum *Infections*) (18). Subsequently, most analyses were performed both for the 28 *P. vivax* infections that were monoclonal by the F_{WS} metric and for all *P. vivax* samples (i.e., multiclonal and monoclonal infections) to assess the effect of multiclonality on our results.

***P. vivax* Has Less Population Substructuring than Sympatric *P. falciparum*.** Principal component analysis (PCA) revealed no population substructuring among *P. vivax* isolates. In contrast, *P. falciparum* parasites were partitioned into subpopulations (Fig. 2). These partitions did not correspond to collection site, date of collection, or multiplicity of infection (MOI). We used *k*-means clustering to confirm that all *P. vivax* isolates were part of a single cluster and that the *P. falciparum* population was subdivided into four clusters (Fig. S2) (19). One cluster represents a central "ancestral-like" *P. falciparum* population from which subpopulations of drug-resistant parasites have undergone epidemic expansion. Analysis of additional projections supports these differences in population structure (Fig. S3). Therefore most analyses were performed using all *P. falciparum* samples and, separately, using the 18 parasites of the central ancestral-like population.

***P. vivax* Has Undergone More Rapid Expansion than Sympatric *P. falciparum*.** To explore the differences in demographic histories, we examined the allele-frequency spectra (AFS) of the *P. vivax* and *P. falciparum* populations. Spectra were calculated by variant type and compared with the spectrum expected in a simulated coalescent population with no natural selection, constant population size, and complete random mating (Fig. 3). We observed an excess of low-frequency derived alleles in the Cambodian *P. vivax* AFS [both for the entire sample and for the monoclonal infections only (Fig. 3 and Fig. S4)], suggesting population expansion. In contrast, the overall *P. falciparum* population had no excess of low-frequency alleles, suggesting limited or absent population expansion. However the overall *P. falciparum* population did exhibit an excess of intermediate-frequency derived alleles, which likely reflected the presence of multiple *P. falciparum* subpopulations and which disappeared upon analysis of only the central, ancestral-like population (Fig. 3 and Fig. S4).

Next, we fit various demographic scenarios to the observed allele-frequency spectra to identify a best-fit model and specific population parameters. Using a diffusion approximation paradigm, we tested scenarios of constant population size, population decline, exponential increase, two-epoch increase, and bottleneck with subsequent exponential growth (Fig. 4) (20). We used the Akaike information criterion (AIC) to inform model selection. For *P. vivax* (all samples and monoclonals only), models of parasite expansion strongly outperformed the other models, with the

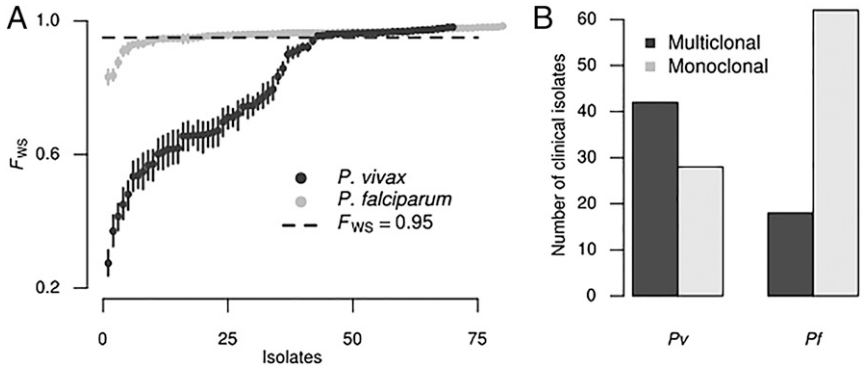

Fig. 1. Higher MOI in *P. vivax* infections compared with *P. falciparum* infections. (*A*) F_{WS} calculated for *P. vivax* and *P. falciparum*. Points represent the point estimate of F_{WS} for each sample in the respective population. Vertical bars represent the maximum and minimum value in 1,000 bootstrap replicates, which downsampled the number of SNPs to be equal for *P. vivax* and *P. falciparum*, to correct for the increased number of *P. vivax* SNPs. (*B*) Summary bar graph representing the number of *P. vivax* (Pv) and *P. falciparum* (Pf) clinical isolates considered monoclonal or polyclonal, with a cutoff of F_{WS} <0.95 being considered polyclonal.

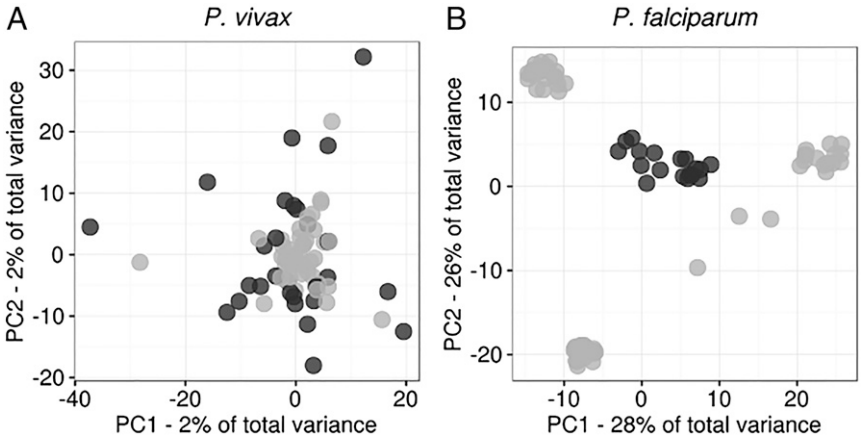

Fig. 2. PCA reveals striking differences in population structure between Cambodian *P. vivax* and *P. falciparum*. PCA was performed from single-nucleotide genetic variant data for each of 70 *P. vivax* clinical isolates (*A*) and 80 *P. falciparum* clinical isolates (*B*). Monoclonal *P. vivax* isolates are shaded dark gray. Among *P. falciparum* isolates, members of the ancestral-like population are shaded dark gray. Noise was added to the *P. falciparum* panel to mitigate overplotting.

model of positive exponential growth having the best fit (Table S1). For the ancestral-like *P. falciparum* population, exponential growth models marginally outperformed the other models, suggesting only modest expansion of this parasite population (Table S2). Comparison of best-fit models suggests that *P. vivax* has expanded more dramatically [factor of population contraction (η_G) = 20.00] and over a short time span (T = 1.03) than the ancestral-like *P. falciparum* population (η_G = 1.94, T = 4.99), resulting in a larger N_{eff} for *P. vivax* than for *P. falciparum* [ancestral mutation rate (θ) = 850 and 240, respectively] (Tables S1 and S2).

To assess the goodness-of-fit of these inferred parameters, we selected the best-fit models to parameterize coalescent simulations. Tajima's D was calculated for each simulated gene and compared with observed values of Tajima's D (Fig. S5). Simulated and observed values of Tajima's D for *P. vivax* and *P. falciparum* were concordant, with a negative mean value and a strong right skew, supporting the inferred population histories for both *P. vivax* and *P. falciparum* (Fig. S6). Cell-surface proteins in *P. falciparum* have been described previously as being under strong balancing selection (high Tajima's D) because of selection by human immunity. Our analysis confirmed enrichment for cell-surface protein exons among targets of strong balancing selection after Bonferroni correction (P = 0.00124) (Table 1). Genome-wide assessment of balancing selection with Tajima's D for *P. vivax* has not been reported previously, and we found several instances of modest Gene Ontology (GO)-term enrichment among targets of strong balancing selection, including chromatin modifiers (Table 1). Additional details about the assessment of Tajima's D are provided in *SI Materials and Methods, SI Assessment of Tajima's* D.

***P. vivax* Shows Stronger Evidence of Recent Directional Selection on Transcriptional Regulators than *P. falciparum*.** The population structure of *P. falciparum* in Cambodia has been shaped by the intensive use of artemisinins and their partner drugs, bed nets, and improved diagnostics (21, 22). However, little is known about how coendemic *P. vivax* populations have responded to these same selective forces. We used linkage disequilibrium-based tests to identify genomic regions that have undergone recent directional selection consistent with selective sweeps. Because these tests are haplotype based, we focused on the monoclonal *P. vivax* isolates (n = 28) and the

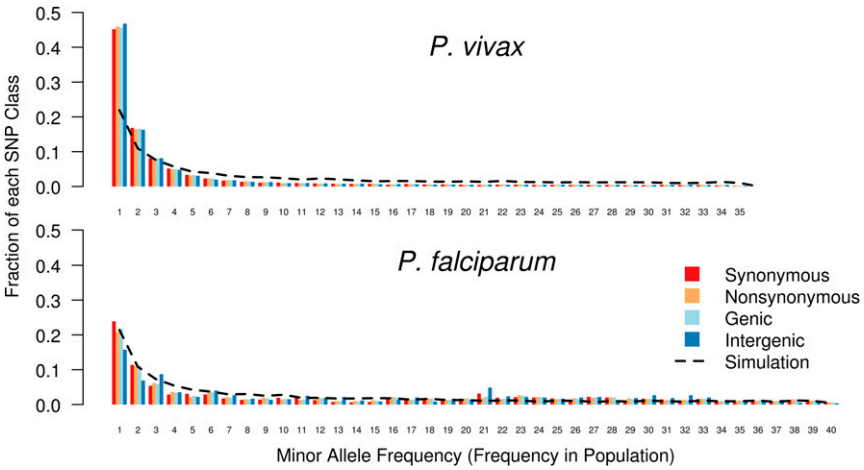

Fig. 3. Allele-frequency spectra suggest greater population expansion among *P. vivax* than among *P. falciparum*. Each column indicates the fraction of the total SNPs that fall into a particular frequency class. Columns are color coded by type of SNP. The *x* axis represents the number of samples within which each SNP occurs. (*Upper*) We observed a preponderance of low-frequency alleles in the *P. vivax* population, compared with coalescent simulations of a no-growth population (black line). Under a Wright–Fisher model of genetic evolution, these data suggest that the *P. vivax* population has undergone a recent expansion. (*Lower*) In contrast, for *P. falciparum*, we observed an excess in intermediate-frequency minor alleles, reflecting the subdivided population structure.

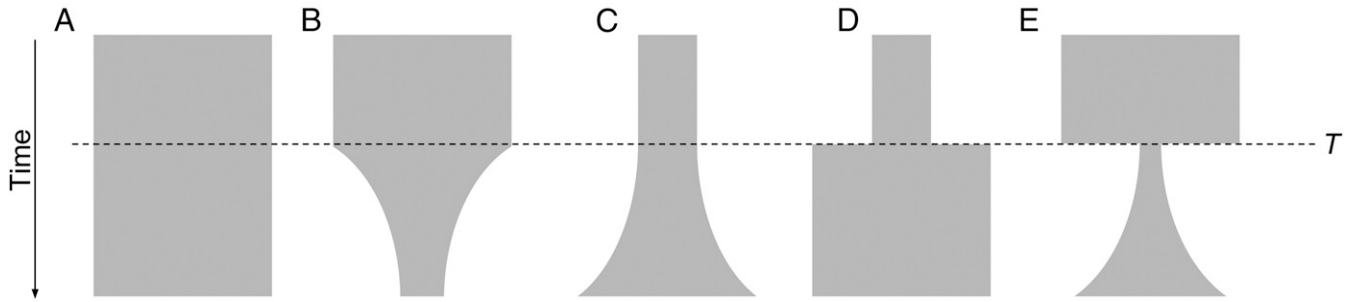

Fig. 4. Five one-population models were used for demographic inference. The following models were fitted to observed *P. vivax* and *P. falciparum* allele-frequency spectra using a diffusion approximation of population evolution: (*A*) a model of no change in N_{eff} through time; (*B*) a model of population decline beginning at time *T*; (*C*) a model of exponential population expansion, beginning at time *T*; (*D*) a two-epoch model of sudden population expansion at time *T*; (*E*) a model of rapid population decline followed by exponential growth, beginning at time *T*. In all cases, the exponential-growth model (*C*) was at least marginally the best fit scenario. For the *P. vivax* population, growth models (*C*–*E*) far outperformed the static (*A*) or decline (*B*) models. See Tables S1 and S2.

ancestral-like *P. falciparum* group (n = 18), which were predominantly monoclonal (15/18).

Using 45,701 high-quality SNPs occurring in the monoclonal *P. vivax* infections, we performed the nS_L test for directional selection (23). This haplotype-based testing offers the advantage of not requiring a genome recombination map and has proven sensitive for detecting incomplete selective sweeps. Strikingly, among the 15 strongest signals of directional selection, six were in close proximity to regulators of gene expression (four AP2 domain-containing transcription factors and two proteins containing the SET domain, which is a histone modulator) (Fig. 5*A* and Table 2). The strongest signal was on chromosome 14, in close proximity to an AP2 domain-containing transcription factor (PVX_122680). Analysis of the entire *P. vivax* population using nS_L yielded qualitatively similar, although blunted, results because of the artifactual breakdown of linkage disequilibrium caused by multiple infections (Fig. S7). Similar results were found when the integrated haplotype score (iHS), which has been used extensively in malaria studies to assess directional selection, was calculated for these same loci (Fig. S7). In contrast, when nS_L testing was performed for the 5,158 SNPs in the *P. falciparum* ancestral-like population, only a single transcription factor was identified as being under moderate directional selection (Table 3 and Fig. S7). This difference does not appear to be attributable to the smaller number of SNPs used in the *P. falciparum* analysis (likely resulting from a more clonal population as well as from a decreased ability to call SNPs in intergenic regions of the AT-rich *P. falciparum* genome). As evidence that results are not significantly impacted by detection bias between genic and intergenic SNPs, when we repeated nS_L testing in the *P. vivax* population using only the 28,746 genic SNPs, no appreciable differences in nS_L signals were identified. The high level of coverage achieved in the coding regions of both populations suggests an equal opportunity to detect signatures of selection among transcription factors in both species.

To determine the extent of haplotype homozygosity around the predominant selective sweep, we performed extended haplotype homozygosity (EHH) testing, centering our analysis on the SNP with the highest nS_L score. We identified a 100-kb region of strong linkage disequilibrium around the principle chromosome 14 locus (Fig. 5 *B* and *C*) in isolates with the selected allele. In our Cambodian *P. vivax* population sample, this ApiAP2 transcription factor contains 25 polymorphisms (Table S3), including two high-frequency nonsynonymous changes, one of which occurs within the AP2 DNA-binding domain, based on comparisons with its

Table 1. GO analysis of observed Tajima's *D* values uncovers evidence for enrichment of GO terms in genes under strong directional and balancing selection

Dataset*	Feature†	*D* percentile‡	GO term	GO ID	*P* value§
P. vivax (M/A)	Exon	First	Histone deacetylase complex	GO:0000118	0.0423
P. vivax (M)	Gene	First	ATP binding	GO:0005524	0.0493
P. vivax (A)	Gene	First	Cellular component	GO:0005575	0.00941
P. vivax (M)	Exon	99th	Metallo-sulfur cluster assembly	GO:0031163	0.0440
P. vivax (M)	Exon	99th	Iron-sulfur cluster assembly	GO:0016226	0.04402
P. vivax (M/A)	Gene/exon	99th	Plastid large ribosomal subunit	GO:0000311	0.0166
P. vivax (M/A)	Gene/exon	99th	Plastid ribosome	GO:0009547	0.0306
P. vivax (M/A)	Gene/exon	99th	Plastid part	GO:0044435	0.0306
P. vivax (M/A)	Gene/exon	99th	Plastid stroma	GO:0009532	0.0306
P. falciparum	Gene	First	Retrograde transport, endosome to Golgi	GO:0042147	0.0431
P. falciparum	Gene	First	Endosomal transport	GO:0016197	0.0431
P. falciparum	Gene/exon	99th	Cell surface	GO:0009986	0.00124

Tajima's *D* was calculated on a per-gene and per-exon basis for the entire *P. vivax* population sample, the *P. vivax* monoclonal subset, and the central, ancestral-like *P. falciparum* subpopulation. For genewise and exonwise Tajima's *D* values, the first percentile (largest negative values) and 99th percentile (largest positive values) were included in GO term enrichment analysis.

*(M) indicates the result was found among *P. vivax* monoclonals; (A) indicates result found among the entire *P. vivax* population sample.

†GO-term enrichment was found in exonwise analysis, genewise analysis, or both exon- and genewise analyses.

‡First percentile indicates largest negative Tajima's *D* values, consistent with directional selection; 99th percentile indicates largest positive Tajima's *D* values, consistent with balancing selection.

§Bonferroni-corrected *P* values; if a term was significant (to a Bonferroni-corrected $P \leq 0.05$) in both monoclonal (M) and all (A) analyses, the more conservative (higher) of the two *P* values is reported.

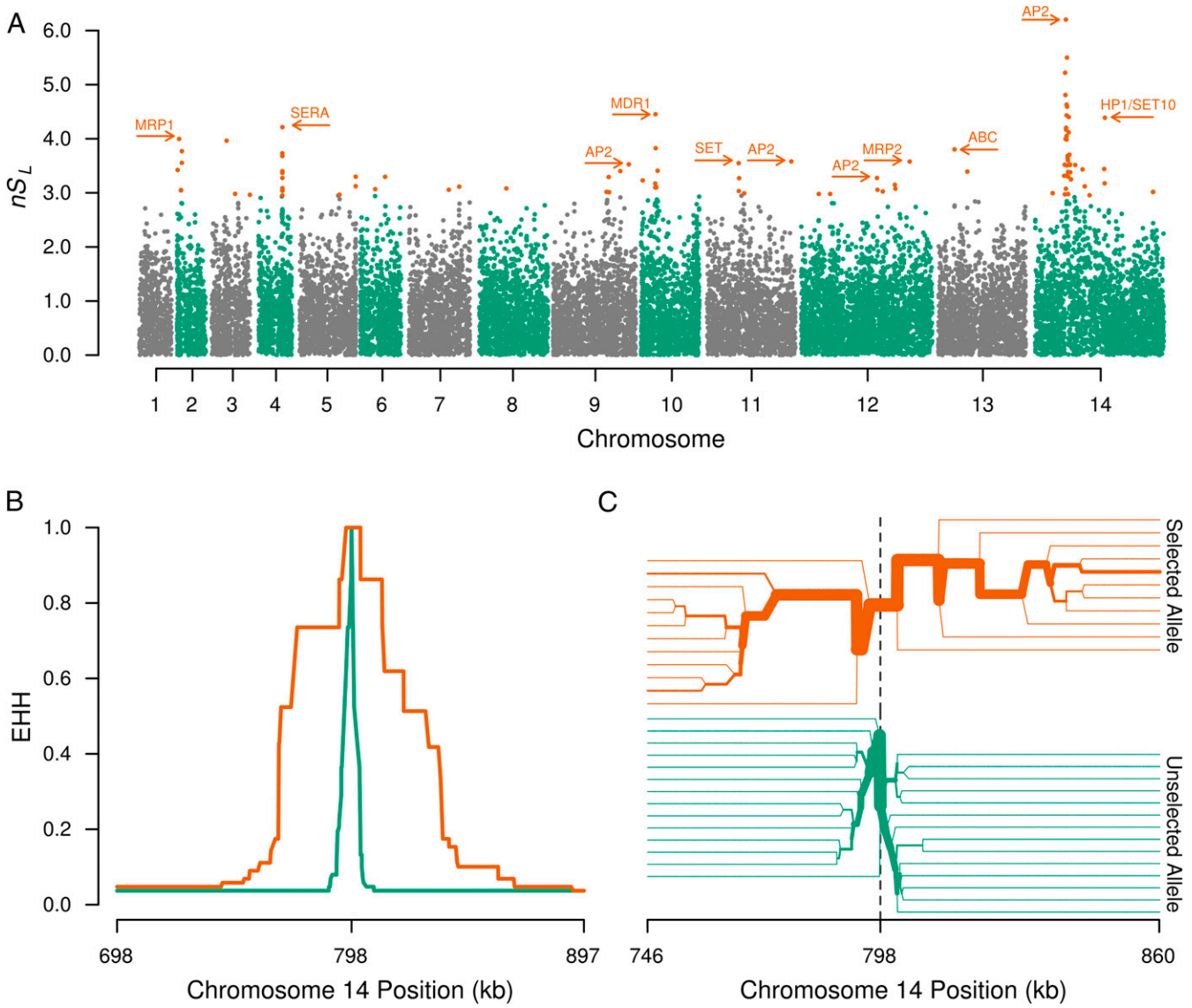

Fig. 5. Evidence for strong selective sweeps in Cambodian *P. vivax*. (*A*) A Manhattan plot of normalized nS_L values. Each point corresponds to an SNP, and the top 0.5% values (under strong directional selection) are rendered in orange. Polymorphisms without evidence of strong directional selection are rendered in either gray or green, according to chromosome. This view suggests that several genomic regions are under positive selection, including areas near transcription factors (AP2 domain-containing), chromatin regulators (SET10 and HP1), antigens under known positive selection (SERA4 and 5), and drug-resistance genes (MDR1 and MRP1). (*B* and *C*) The extent of EHH in the strongest sweep, within chromosome 14 in *A*. (*B*) EHH decay for the haplotypes around selected (orange) and unselected (green) alleles. (*C*) A haplotype bifurcation diagram is shown centered on the focal variant. Line thickness represents the number of identical haplotypes flanking the selected (orange) or unselected (green) alleles. Linkage breaks down with increasing distance from the focal variant. *B* and *C* provide evidence that the strong sweep on chromosome 14 extends ~50 kb in each direction.

P. falciparum ortholog (PF3D7_1317200) (24). In addition to the AP2-domain transcription factor, this region contains 25 genes (Table S4).

We found complementary evidence that *P. vivax* transcriptional regulators are under strong directional selection using allele frequency-based testing for selection. We calculated genewise Tajima's *D* (an allele frequency-based test, in contrast to nS_L and iHS, which are haplotype-based tests of selection) for *P. vivax* and *P. falciparum*. We selected the first and 99th percentile of observed genewise Tajima's *D* values and investigated these genes for functional enrichment. We observed enrichment of histone deacetylase complex members among the first percentile of genes, but this observation did not reach significance after Bonferroni correction. Because strong local values of Tajima's *D* can be obscured when considering an entire gene, we also performed this analysis in an exonwise manner for both *P. vivax* and *P. falciparum* (18). In this way we found statistically significant enrichment for histone deacetylase complex members among the first-percentile exons after strict Bonferroni correction ($P = 0.0423$) (Table 1).

Both *P. vivax* and *P. falciparum* Show Evidence of Recent Directional Selection on Known and Putative Drug-Resistance Genes. Four of the strongest 15 strongest nS_L signals in *P. vivax* were in close proximity to transporters (*pvmdr1*, *pvmdr2*, *pvmrp1*, and an ABC transporter), all of which are potential drug-resistance loci (Table 2). A prominent sweep did encompass the *pvmdr1* locus. We compared key drug-resistance SNP frequencies in *pvmdr1* to frequencies in Cambodian isolates collected several years earlier, during 2006 and 2007, from Kâmpôt province (25). Two key mutations (Y976F and F1076L) existed at roughly the same frequency (89% in previously

Table 2. Top hits in nS_L analysis of *P. vivax* monoclonal infections

	Focal SNP statistics		Sweep region statistics		
Chr	Location	nS_L	Closest plausible genetic driver	Gene ID	Distance*
14	797,870	6.203	Transcription factor with AP2 domain(s), putative (apiap2)	PVX_122680	−4,462
10	351,691	4.453	Multidrug resistance protein 1, putative (MDR1)	PVX_080100	10,010
14	1,650,897	4.389	Heterochromatin protein 1 (HP1); H3 lysine methyltransferase (SET10)	PVX_123682; PVX_123685	999; 9,918
04	577,543	4.215	Serine-repeat antigen 4 (SERA) Serine-repeat antigen 5 (SERA)	PVX_003825; PVX_003830	789; −1,691
02	85,446	3.997	Multidrug resistance-associated protein 1, putative (MRP1)	PVX_097025	−68,196
03	368,737	3.964	Hypothetical proteins	—	—
13	396,427	3.805	ABC transporter B family member 7, putative (ABCB7)	PVX_084521	−7,326
11	1,900,445	3.581	Transcription factor with AP2 domain(s), putative (apiap2)	PVX_113370	−31,184
12	2,421,187	3.578	Multidrug resistance protein 2 (MDR2)	PVX_118100	−4,361
11	749,030	3.549	SET domain containing protein	PVX_114585	46,877
09	1,693,066	3.527	Transcription factor with AP2 domain(s), putative (AP2-O)	PVX_092760	7,201
14	1,002,439	3.515	Hypothetical proteins	—	—
09	1,506,397	3.403	Transcription factor with AP2 domain(s), putative (apiap2)	PVX_092570	9,327
13	676,297	3.393	Hypothetical protein	—	—
05	1,303,321	3.299	Vir and fam antigens	—	—

Haplotype-based tests of selection were performed in monoclonal samples to avoid false breakdown of linkage caused by mixed genotypes from sequencing. Results are presented by ranking on absolute normalized nS_L score.
*Distance in bases from focal SNP to the putative driver gene; a negative sign indicates putative driver gene occurs upstream of the focal SNP.

collected samples vs. 77% in more recently collected samples and 87% in previously collected samples vs. 90% in more recently collected samples, respectively). This high similarity in allele frequency in samples collected before and during the artemisinin-resistance containment efforts suggests that the sweep encompassing the *pvmdr1* locus was not driven by ACT drug pressure or resistance-containment efforts and may result from more longstanding chloroquine pressure (similar to *pfmdr1* mutations in *P. falciparum*) that preceded containment efforts. Allele frequencies for all polymorphisms in known or putative drug-resistance genes are summarized in Table S5.

Similar to previous studies of directional selection in *P. falciparum*, we identified drug-resistance genes (e.g., *pfcrt* and *kelch* K13) with evidence of strong directional selection (Table 3) (26–29). Allele frequencies for all polymorphisms in known drug-resistance genes are summarized in Table S6. Of note, both nS_L and iHS statistics revealed a chromosome 11 locus with strong and extended directional selection (Table 3 and Fig. S7). This locus encompassed the *pfama1* and phosphatidylinositol-4-phosphate 5-kinase genes (PF3D7_1129600). Interestingly, PF3D7_1129600 catalyzes the phosphorylation of phosphatidylinositol 4-phosphate to form phosphatidylinositol 4,5-bisphosphate (30). The drug target PI(4)K (PF3D7_0509800) alters the intracellular distribution of phosphatidylinositol-4-phosphate in the parasite, placing these genes on the same biologic pathway (31). Although the functional significance of this signal is unclear, genome scans in other populations also have identified strong directional selection at this locus (27, 32).

Copy-Number Variants Are Not Associated with Detected Selective Sweeps. Copy-number variants (CNVs) are important in the evolution of parasite populations (33). Using two complementary

Table 3. Top hits in nS_L analysis of *P. falciparum* ancestral-like population

	Focal SNP statistics		Sweep region statistics		
Chr	Location	nS_L	Closest plausible genetic driver	Gene ID	Distance*
04	531,138	3.220	Hypothetical proteins	—	—
07	384,302	3.106	Chloroquine resistance transporter (CRT)	PF3D7_0709000	18,920
11	1,089,187–1,347,798	3.024	Phosphatidylinositol-4-phosphate 5-kinase; apical membrane antigen 1 (AMA1)	PF3D7_1129600; PF3D7_1133400	51,935; −93,151
13	2,550,178	2.924	Hypothetical proteins	—	—
06	1,093,476	2.871	Hypothetical proteins	—	—
14	610,371	2.805	Hypothetical proteins	—	—
06	825,473	2.730	Merozoite surface protein 10 (MSP10)	PF3D7_0620400	−25,905
09	203,308	2.656	Hypothetical proteins	—	—
12	124,649	2.652	Dynein heavy chain, putative	PF3D7_1202300	0
13	639,213	2.597	Transcription factor MYB1 (MYB1)	PF3D7_1315800	−20,805
02	374,197	2.589	Serine repeat antigen 1 (SERA1)	PF3D7_0208000	48,474
03	361,403	2.571	Hypothetical proteins	—	—
13	1,775,996	2.563	Kelch protein K13 (K13)	PF3D7_1343700	48,999
01	226,924	2.505	Hypothetical proteins	—	—
09	513,729	2.489	Cysteine repeat modular protein 1 (CRMP1)	PF3D7_0912000	0

Results are presented by ranking on absolute normalized nS_L score.
*Distance in bases from focal SNP to the putative driver gene; a negative sign indicates putative driver gene occurs upstream of the focal SNP; a zero indicates focal SNP occurs within the putative driver gene.

in silico methods, we identified a handful of high-confidence CNVs in the *P. vivax* and *P. falciparum* populations (Table S7). Notably, we identified Duffy-binding protein (DBP) duplication events in 17 of 70 *P. vivax* isolates and multidrug-resistance protein 1 (MDR1) duplication events in 12 of 80 *P. falciparum* isolates. As in previous work in *P. falciparum* (33), no high-confidence CNV occurred in close proximity with the selective sweeps identified above.

Discussion

In this study we compared whole-genome sequence data of 70 *P. vivax* and 80 coendemic *P. falciparum* infections to show that the *P. vivax* population in western Cambodia has experienced more rapid and uninterrupted growth than the sympatric *P. falciparum* population, resulting in less substructuring and a larger N_{eff}. Clues to the causes underlying this population-level resilience of *P. vivax* in the face of malaria-control measures can be found in the different genomic signatures of selection between the two species. Evidence for the importance of transcriptional regulation to the success of *P. vivax* was found both by haplotype-based and allele frequency-based tests of selection. Although known and putative drug-resistance genes were found at the center of selective sweeps in both *P. falciparum* and *P. vivax* in this and other studies (Tables 2 and 3), the strongest selective sweep in the *P. vivax* population occurred in close proximity to an AP2 transcription factor (PVX_122680), suggesting that *P. vivax* is responding to selective pressures by altering its transcriptional profile (29, 34, 35).

Although studies of transcriptional regulation in *P. vivax* are just beginning (36–38), some of the key biological processes that are thought to help *P. vivax* evade traditional control measures may be under tight transcriptional control. Unlike *P. falciparum*, *P. vivax* gametocytogenesis occurs early and concomitantly with the asexual cycle, so by the time a person is symptomatic or is diagnosed with *P. vivax* malaria, that person already harbors infectious gametocytes capable of infecting mosquitoes (39). It appears that a key step in *Plasmodium* gametocytogenesis is accomplished by transcriptional repression via an AP2 family transcription factor that blocks asexual replication and promotes conversion to the sexual stage (40). In fact, the *P. falciparum* ortholog of the principal AP2 gene identified in our study has been associated with transcriptional regulation of gametocytogenesis (41). Similarly, processes governing hypnozoite dormancy and activation that underlie *P. vivax* relapse may be under epigenetic regulation. Histone deacetylase inhibitors accelerated the rate of hypnozoite activation in long-term primary simian hepatocyte cultures, suggesting that histone methylation may maintain hypnozoite dormancy by suppressing transcription (42, 43). Finally, *P. vivax* chloroquine resistance has not been found to correlate with DNA changes in *pvcrt* but may correlate with the expression of this transporter protein, whose ortholog mediates chloroquine resistance in *P. falciparum* (44–46).

Thus, an ability to modulate transcriptional regulation may be more central to *P. vivax*'s adaptation to selective pressures than for *P. falciparum*. Notably, a previous comparison of a 1990s era Peruvian *P. vivax* isolate to the reference Sal1 strain collected 35 y earlier found an increased dN/dS signal at two AP2-containing transcription factors, suggesting that evolutionary changes in these genes are not confined to *P. vivax* in Cambodia (47). Because this was a within-species comparison of nonsynonymous to synonymous SNP ratio (dN/dS), it is difficult to determine the time-scale (e.g., whether ancient or recent) associated with this selection (48). The top hit in our selective sweep, a sign of recent and strong directional selection, was an AP2 factor that was specifically cited among genes with >1.0 dN/dS in this previous study.

The potential reliance on transcriptional changes is not an absolute difference between species, however. CNV of *pfmdr1*, which likely leads to increased gene transcription, is known to modulate drug efficacy. However, a closer look at previously published *P. falciparum* scans for selection in multiple locations in Africa also reveals a role for adaptation through the modification of transcriptional regulation. The *P. falciparum* ortholog of the SET-domain protein on chromosome 11 (PVX_114585) identified in our nS_L analysis lies near the center of a selective sweep that also has occurred in Senegal, the Gambia, and Ghana (16, 49, 50). In addition, it is known that transcriptional timing can affect *P. falciparum* drug-resistance responses, in particular to artemisinins (51). These findings suggest an underappreciated role for the modification of transcriptional regulation in *P. falciparum* fitness.

Further investigation of the potential role of transcriptional modification in *P. vivax* drug resistance and other biological processes will require continued development of in vitro models. Such research strategies include modifications to *Plasmodium knowlesi* homologs, the use of humanized mice infected with transgenic *P. vivax*, allowing transcriptional analysis through the liver stage, and the use of monkeys infected with transgenic *P. vivax*, allowing transcriptional analysis through the blood stage (52–56). To date, descriptions of transcriptional modification in *Plasmodium* sp. appear to be limited to associations with genomic structural variants, for example, gene amplification as in *pfmdr1* and deletions or repeat-length polymorphisms in promoter regions (57, 58). One could hypothesize that SNPs in an AP2 transcription factor DNA-binding domain may alter its motif binding and stage-specific expression of downstream genes. More likely, we have not identified causal variants; rather, our selective sweep is detecting variants that are in linkage disequilibrium with the true functional variants, which could be larger deletions or insertions not readily detected by short-read sequencing. If the selected variants found in our analyses are replicated in other cohorts, further in vitro experimentation is needed to determine whether they associate with transcriptional regulation. Such findings would provide key evidence of an advanced parasitic response to selective pressure in *P. vivax*. They also would suggest that tracking genetically fit parasites could be complicated, supporting the idea that *P. vivax* will be the more challenging species to eliminate.

Our findings of a diverse *P. vivax* population without a discernible population structure are similar to microsatellite- and gene-based studies in Papua New Guinea, Indonesia, Venezuela, and Cambodia that have observed more genetic structuring among *P. falciparum* populations than among coendemic *P. vivax* populations (Fig. 2) (10–15). We also have previously characterized the within-host diversity of coendemic *P. vivax* and *P. falciparum* in this region by amplicon deep sequencing, with results supporting the difference in polyclonality (59, 60). The high proportion of polyclonal *P. vivax* infections and the lack of parasite substructuring also were observed in a recent report from Cambodia that used deep sequencing of more than 100 SNPs across the *P. vivax* genome (18, 61).

Whole-genome sequencing allowed us to extend beyond descriptions of diversity and to make inferences about the demographic history of sympatric *Plasmodium* in the region. The best-fit demographic models indicated that *P. vivax* in western Cambodia has undergone steady exponential growth, expanding more rapidly than the ancestral-like *P. falciparum* population, for which exponential growth was only marginally the best-fit model (Tables S1 and S2). The demography-adjusted estimate of the ancestral mutation rate (θ), a proxy for N_{eff}, found that the *P. vivax* N_{eff} is substantially larger than the ancestral-like *P. falciparum* population. These demographic scenarios match epidemiological observations that *P. vivax* cases have increased even while *P. falciparum* incidence has rapidly declined over the same time period (6). Interestingly, although the total number of *P. falciparum* cases in Cambodia has decreased, our demographic models show a slowly increasing N_{eff}, at least for the ancestral-like population. Although it is possible that our models are detecting ancient rather than recent trends, another study of isolates in this area reached similar conclusions, lending credence to our findings and suggesting that recent control efforts

have not significantly decreased *P. falciparum* genetic diversity in the region (62).

We should note that we have largely assumed that selective pressures have been imposed by human control interventions, primarily antimalarial treatment. However, we cannot rule out the possibility that other selective pressures have also affected parasite population structuring, for example, widespread deforestation and climate change altering the diversity of Anopheline vectors or evolving host–parasite interactions within mosquito vectors. In *P. vivax*, a strong New World vs. Old World divide correlates with genetic variation in *pvs47*, the ortholog of *pfs47*, which has been associated with differential infectivity in different mosquito species (63, 64). However, both *Anopheles dirus s.s.* and *Anopheles minimus s.s.*, the main malaria vectors in Cambodia, are adept at transmitting both *P. falciparum* and *P. vivax* (65, 66). Therefore we do not think that the species-specific differences observed in population structure and signatures of selection are attributable solely to unmeasured mosquito factors.

Finally, the population structuring we observed among contemporary Cambodian *P. vivax* and *P. falciparum* isolates should be interpreted in light of their ancient demographic histories (67). Evidence has shown that both species of malaria are significantly less diverse than primate malarias and likely have undergone genetic bottlenecks associated with host switching and emergence out of Africa, likely less than 10,000 y ago (63, 68–72). As these species have emerged from the bottleneck, they likely have undergone both sustained and recent selection. In the case of *P. vivax*, previous work has demonstrated that long-term diversifying and directional selection has shaped its genetic diversity (63, 68). Our data are consistent with this finding, showing minimal structuring in the *P. vivax* population but evidence of strong directional selection in multiple regions of the genome. Additional investigation is required to determine if these adaptive signatures are geographically restricted (68) or occurred as or before *P. vivax* expanded out of Africa. In the case of *P. falciparum*, reductions in diversity immediately around loci associated with drug resistance (73, 74) are commonly reported, and the current substructuring seen in Cambodian *P. falciparum* parasites has been linked to drug-resistant subtypes (22). This evidence suggests that more recent selective forces have shaped population structure in this species. Previous studies also suggest that *P. falciparum* infections in Southeast Asia were more multiclonal two decades ago, again suggesting that recent forces have constrained its genetic diversification (75–77). This evidence is consistent with our findings that *P. falciparum* in Cambodia has undergone large demographic shifts much more recently than *P. vivax*.

Applying population genomic tools to *Plasmodium* parasites comes with caveats. Malaria parasites have a complex lifecycle, including human and mosquito stages, with multiple clonal generations occurring within the human bloodstream and frequent bottlenecks during transmission (78). Such realities violate the assumptions of the Wright–Fisher model and complicate inference from genetic data. These peculiarities of the malaria lifecycle may skew the allele-frequency spectrum toward increased singletons, even at neutral sites, leading to quantitatively or even qualitatively inappropriate demographic conclusions (79). However, our demographic conclusions are supported by epidemiologic observation as well as by the results of other population genetic studies (6, 13). Because extensive recognized and unrecognized paralogous families in the *P. vivax* genome present significant mapping and variant-calling challenges, we curated our data carefully, performing extensive tests to determine the best alignment, filtering, and variant-calling approaches. In addition, none of our samples underwent hybrid selection, giving us greater confidence in the quantitative accuracy of calls in mixed infections and structural variants (80, 81).

In summary, we present evidence that sympatric *P. vivax* and *P. falciparum* populations in Cambodia have responded in substantively different ways to the intense selective pressure imposed by the recent artemisinin-resistance containment campaign and national antimalarial drug policies. Although *P. falciparum* has experienced population splitting, the *P. vivax* population remains admixed, with strong growth, high genetic diversity, and frequent polyclonal infections. These findings match epidemiologic observations of relative *P. vivax* resilience to current control measures. Our comparative genomic analysis hints at the mechanisms behind these different responses. Although we found that both *P. vivax* and *P. falciparum* have experienced selective sweeps around known or putative antimalarial-resistance genes, the strongest signatures of directional selection in *P. vivax* occur near genes involved in transcriptional regulation. These findings highlight important differences between *P. vivax* and *P. falciparum* biology that are relevant to the direction of future malaria elimination efforts. *P. vivax* elimination will require a deeper understanding of the ways in which this species exerts transcriptional control. A clearer picture of *P. vivax* adaptive mechanisms will guide drug and vaccine strategies that target early gametocytogenesis, hypnozoite biology, and the next generation of drug resistance.

Materials and Methods

Sample Collection. Clinical isolates were collected between 2009 and 2013 by the Armed Forces Research Institute of Medical Sciences in three Cambodian provinces, Oddar Meanchey, Battâmbâng, and Kâmpôt (Fig. S1). Uncomplicated *P. vivax* or *P. falciparum* malaria patients presenting to study-site clinics gave written informed consent for their participation in this study. Study staff collected and leukodepleted venous blood and administered treatment in accordance with Cambodian National Malaria Control Program guidelines. Molecular studies were approved by the Institutional Review Board at the University of North Carolina, the Walter Reed Army Institute of Research Institutional Review Board, and the Cambodian National Ethical Committee for Health Research. Additional details of the participants are included in *SI Materials and Methods*, *SI Participant Characteristics* and Table S8.

***P. vivax* and *P. falciparum* Sample Sequencing.** For *P. vivax*, whole blood was leukodepleted using Plasmodipur filters (Euro-Diagnostica). The ratio of parasite DNA to host DNA was determined using a quantitative PCR (qPCR) assay, and isolates with ≥20% *P. vivax* DNA were sequenced (82). Clinical isolates with high plasmodium:human DNA content were sequenced on the HiSeq 2000 or HiSeq 2500 sequencing system (Illumina) using 100- or 125-bp paired-end chemistry. Data are available at the Sequence Read Archive, and accession numbers are listed in Table S9. For *P. falciparum*, data from previously reported isolates as well as from two previously unidentified isolates were used and reanalyzed in this study to allow comparable analysis methods between species.

Sequence Analysis. Sequence reads were aligned to the *P. falciparum* 3D7 (v3) and *P. vivax* Sal1 (v3) genomes using *bwa mem*, which allows a hybrid end-to-end and local alignment approach (83). To increase sensitivity through hypervariable regions, we raised the base-match bonus (A = 2) and the clip penalty for local alignment (L = 15). PCR and optical duplicates were removed from alignments using the Picard Tools MarkDuplicates utility, and local realignment of highly entropic regions was performed using the GATK IndelRealigner utility (broadinstitute.github.io/picard/). Isolates with fivefold coverage at ≥80% of the genome and ≥90% of genes in the case of *P. vivax* and with ≥60% of the genome and ≥90% of genes in the case of *P. falciparum* were considered for variant calling and further analyses.

Variant calling for both species is described in *SI Materials and Methods*, *SI Sequencing Sympatric* P. vivax *and* P. falciparum *Populations*. Variants were called for each species independently but jointly for all samples using the GATK UnifiedGenotyper (84, 85). Additional details of sequencing and bioinformatic analysis are provided in *SI Materials and Methods*, *SI Sequencing Sympatric* P. vivax *and* P. falciparum *Populations*.

Within-Host Diversity. To test whether infections were monoclonal or polyclonal, we used the F_{WS} statistic (86, 87). To enable direct F_{WS} comparisons between *P. falciparum* and *P. vivax* isolates, which had different numbers of loci, we bootstrapped each calculation 1,000 times, randomly selecting 5,000 variable sites for each isolate and each replicate. The maximum and minimum F_{WS} bootstrap values were identified to provide a generous upper and lower confidence interval for each F_{WS} point estimate. Additionally, *P. vivax*

isolates were screened for MOI using ultra-deep sequencing of the highly polymorphic *pvmsp1* locus (60). This screening was performed using the Ion Torrent platform, and the reads for each individual were clustered using SeekDeep, an iterative clustering algorithm (baileylab.umassmed.edu/SeekDeep/). Experimental procedures were performed as previously described (60). For the purposes of the present study, a sample was deemed multiclonal if a minor clone existed at ≥10% read frequency.

Population Structure and Demographic Inferences. We determined population substructuring using PCA, and cluster assignments were determined using a nonparametric *k*-means approach. PCA was calculated using *adegenet* (19). Five one-population demographic scenarios were fit to the observed site-frequency spectra at synonymous sites in both the *P. vivax* and *P. falciparum* populations (20). For each model, 100 independent runs were performed for each dataset. Interrun parameter values were compared to assess model convergence, and the iteration with the highest log-likelihood was selected. Details pertaining to the models and parameter space explored are available in *SI Materials and Methods*, *SI Demographic Modeling Using a Diffusion Approximation Paradigm*.

Assessment of Tajima's *D*. Gene loci with extreme high and low Tajima's *D* were identified for both *P. vivax* and *P. falciparum*, excluding genes from highly paralogous families and in chromosomal telomeres (80). Nonexcluded genes were simulated for corresponding Tajima's *D* values using the R package *coala*, a wrapper for *ms* (88, 89). Inferred parameters for the best-fit one-population demographic scenario for the *P. falciparum* ancestral-like population, the entire *P. vivax* population, and the *P. vivax* monoclonal population were used to parameterize coalescent simulations. Mutation rates per gene were determined from gene length and the calculated genome mutation rate, because the mutation rate per gene considered in the *ms* model is proportional to gene length (26). From these simulations, we established a null distribution of Tajima's *D* values for both *P. vivax* populations and for the *P. falciparum* ancestral-like population to aid in identifying genes under unexpectedly strong balancing or directional selection. Genes identified in the first and 99th percentile of the observed Tajima's *D* distribution for each of the three populations were investigated with the GO analysis tool from PlasmoDB. Genes with a Bonferroni-corrected *P* value <0.05 were considered significant by the GO analysis.

Haplotypic Scans of Positive Selection. Because there is as yet no fine-scale map of recombination for the *P. vivax* genome, we initially sought a map-independent haplotype-based approach. nS_L, a modification of iHS, obviates the need for a genetic map, reduces its dependence on recombination and demographic events, and may afford increased sensitivity to detect soft selective sweeps (23). This statistic has proved sensitive in identifying selection in other nonmodel organisms (90). As a secondary test, we constructed recombination maps using LDhat interval and performed the iHS haplotype-based test for directional selection (91, 92). iHS was calculated using iHH_0/iHH_1, in which the subscripts 0 and 1 denote alleles, agnostic to ancestral or derived status. For both nS_L and iHS, because of the lack of outgroup sequences, we discarded the sign. We plotted iHS and nS_L scores that were normalized according to allele frequency bins. For selected intervals, EHH was calculated for the selected and unselected allele. These three haplotype-based tests for selection were performed using the program *selscan* (93).

Copy Number Analysis. We identified CNVs using a tailored two-step approach for identifying segmental duplications using a custom search approach tuned to the AT-rich genome of *Plasmodium* sp. followed by a probabilistic framework for detecting variants that incorporates multiple types of evidence (*SI Materials and Methods*, *SI CNV in* P. vivax *and* P. falciparum). A subset of *P. falciparum md1* CNVs was confirmed using qPCR as described previously (1, 94).

ACKNOWLEDGMENTS. We thank the study participants and Kristina De Paris, Corbin Jones, and Praveen Sethupathy for review of the manuscript. This research was supported by NIH Grants R01AI089819 and R21AI111108 (to J.J.J.) and R01AI099473 (to J.A.B.). C.M.P. was supported by NIH Training Grants T32GM007092, T32GM008719, and F30AI109979. J.T.L. was supported by NIH Grant K08AI110651. The views expressed in this presentation are those of the authors and do not reflect official policy of the Department of the Army, Department of Defense, or the United States Government.

1. Spring MD, et al. (2015) Dihydroartemisinin-piperaquine failure associated with a triple mutant including kelch13 C580Y in Cambodia: An observational cohort study. *Lancet Infect Dis* 15(6):683–691.
2. Dondorp AM, et al. (2009) Artemisinin resistance in Plasmodium falciparum malaria. *N Engl J Med* 361(5):455–467.
3. Zhou G, et al. (2005) Spatio-temporal distribution of Plasmodium falciparum and p. Vivax malaria in Thailand. *Am J Trop Med Hyg* 72(3):256–262.
4. Cui L, et al. (2012) Malaria in the Greater Mekong subregion: Heterogeneity and complexity. *Acta Trop* 121(3):227–239.
5. Wangroongsarb P, Sudathip P, Satimai W (2012) Characteristics and malaria prevalence of migrant populations in malaria-endemic areas along the Thai-Cambodian border. *Southeast Asian J Trop Med Public Health* 43(2):261–269.
6. Maude RJ, et al. (2014) Spatial and temporal epidemiology of clinical malaria in Cambodia 2004-2013. *Malar J* 13:385.
7. World Health Organization (2015) *Control and Elimination of Plasmodium vivax Malaria – A Technical Brief* (WHO, Geneva).
8. Ould Ahmedou Salem MS, et al. (2015) Increasing prevalence of Plasmodium vivax among febrile patients in Nouakchott, Mauritania. *Am J Trop Med Hyg* 92(3):537–540.
9. Vitor-Silva S, et al. (2016) Declining malaria transmission in rural Amazon: Changing epidemiology and challenges to achieve elimination. *Malar J* 15(1):266.
10. Gray K-A, et al. (2013) Population genetics of Plasmodium falciparum and Plasmodium vivax and asymptomatic malaria in Temotu Province, Solomon Islands. *Malar J* 12:429.
11. Jennison C, et al. (2015) Plasmodium vivax populations are more genetically diverse and less structured than sympatric Plasmodium falciparum populations. *PLoS Negl Trop Dis* 9(4):e0003634.
12. Noviyanti R, et al. (2015) Contrasting transmission dynamics of co-endemic Plasmodium vivax and P. falciparum: Implications for malaria control and elimination. *PLoS Negl Trop Dis* 9(5):e0003739.
13. Orjuela-Sánchez P, et al. (2013) Higher microsatellite diversity in Plasmodium vivax than in sympatric Plasmodium falciparum populations in Pursat, Western Cambodia. *Exp Parasitol* 134(3):318–326.
14. Ord RL, Tami A, Sutherland CJ (2008) ama1 genes of sympatric Plasmodium vivax and P. falciparum from Venezuela differ significantly in genetic diversity and recombination frequency. *PLoS One* 3(10):e3366.
15. Arnott A, et al. (2014) Distinct patterns of diversity, population structure and evolution in the AMA1 genes of sympatric Plasmodium falciparum and Plasmodium vivax populations of Papua New Guinea from an area of similarly high transmission. *Malar J* 13(1):233.
16. World Health Organization (2011) *Global Plan for Artemisinin Resistance Containment* (WHO, Geneva).
17. Murray L, et al. (2016) Microsatellite genotyping and genome-wide single nucleotide polymorphism-based indices of Plasmodium falciparum diversity within clinical infections. *Malar J* 15(1):275.
18. Parobek CM, et al. (2014) Differing patterns of selection and geospatial genetic diversity within two leading Plasmodium vivax candidate vaccine antigens. *PLoS Negl Trop Dis* 8(4):e2796.
19. Jombart T, Ahmed I (2011) adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* 27(21):3070–3071.
20. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* 5(10):e1000695.
21. Miotto O, et al. (2013) Multiple populations of artemisinin-resistant Plasmodium falciparum in Cambodia. *Nat Genet* 45(6):648–655.
22. Miotto O, et al. (2015) Genetic architecture of artemisinin-resistant Plasmodium falciparum. *Nat Genet* 47(3):226–234.
23. Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R (2014) On detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol Biol Evol* 31(5):1275–1291.
24. Painter HJ, Campbell TL, Llinás M (2011) The Apicomplexan AP2 family: Integral factors regulating Plasmodium development. *Mol Biochem Parasitol* 176(1):1–7.
25. Lin JT, et al. (2013) Plasmodium vivax isolates from Cambodia and Thailand show high genetic complexity and distinct patterns of P. vivax multidrug resistance gene 1 (pvmdr1) polymorphisms. *Am J Trop Med Hyg* 88(6):1116–1123.
26. Chang H-H, et al. (2012) Genomic sequencing of Plasmodium falciparum malaria parasites from Senegal reveals the demographic history of the population. *Mol Biol Evol* 29(11):3427–3439.
27. Mu J, et al. (2010) Plasmodium falciparum genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. *Nat Genet* 42(3):268–271.
28. Nwakanma DC, et al. (2014) Changes in malaria parasite drug resistance in an endemic population over a 25-year period with resulting genomic evidence of selection. *J Infect Dis* 209(7):1126–1135.
29. Mobegi VA, et al. (2014) Genome-wide analysis of selection on the malaria parasite Plasmodium falciparum in West African populations of differing infection endemicity. *Mol Biol Evol* 31(6):1490–1499.
30. Leber W, et al. (2009) A unique phosphatidylinositol 4-phosphate 5-kinase is activated by ADP-ribosylation factor in Plasmodium falciparum. *Int J Parasitol* 39(6):645–653.
31. McNamara CW, et al. (2013) Targeting Plasmodium PI(4)K to eliminate malaria. *Nature* 504(7479):248–253.
32. Amambua-Ngwa A, et al. (2012) SNP genotyping identifies new signatures of selection in a deep sample of West African Plasmodium falciparum malaria parasites. *Mol Biol Evol* 29(11):3249–3253.

33. Cheeseman IH, et al. (2016) Population structure shapes copy number variation in malaria parasites. *Mol Biol Evol* 33(3):603–620.
34. Ocholla H, et al. (2014) Whole-genome scans provide evidence of adaptive evolution in Malawian Plasmodium falciparum isolates. *J Infect Dis* 210(12):1991–2000.
35. Park DJ, et al. (2012) Sequence-based association and selection scans identify drug resistance loci in the Plasmodium falciparum malaria parasite. *Proc Natl Acad Sci USA* 109(32):13052–13057.
36. Bozdech Z, et al. (2008) The transcriptome of Plasmodium vivax reveals divergence and diversity of transcriptional regulation in malaria parasites. *Proc Natl Acad Sci USA* 105(42):16290–16295.
37. Zhu L, et al. (2016) New insights into the Plasmodium vivax transcriptome using RNA-Seq. *Sci Rep* 6:20498.
38. Hoo R, et al. (2016) Integrated analysis of the Plasmodium species transcriptome. *EBioMedicine* 7:255–266.
39. Bousema T, Drakeley C (2011) Epidemiology and infectivity of Plasmodium falciparum and Plasmodium vivax gametocytes in relation to malaria control and elimination. *Clin Microbiol Rev* 24(2):377–410.
40. Yuda M, Iwanaga S, Kaneko I, Kato T, Tomomi K (2015) Global transcriptional repression: An initial and essential step for Plasmodium sexual development. *Proc Natl Acad Sci USA* 112(41):12824–12829.
41. Ikadai H, et al. (2013) Transposon mutagenesis identifies genes essential for Plasmodium falciparum gametocytogenesis. *Proc Natl Acad Sci USA* 110(18):E1676–E1684.
42. Dembélé L, et al. (2014) Persistence and activation of malaria hypnozoites in long-term primary hepatocyte cultures. *Nat Med* 20(3):307–312.
43. Barnwell JW, Galinski MR (2014) Malarial liver parasites awaken in culture. *Nat Med* 20(3):237–239.
44. Fernández-Becerra C, et al. (2009) Increased expression levels of the pvcrt-o and pvmdr1 genes in a patient with severe Plasmodium vivax malaria. *Malar J* 8(1):55.
45. Melo GC, et al. (2014) Expression levels of pvcrt-o and pvmdr-1 are associated with chloroquine resistance and severe Plasmodium vivax malaria in patients of the Brazilian Amazon. *PLoS One* 9(8):e105922.
46. Pava Z, et al. (2015) Expression of Plasmodium vivax crt-o is related to parasite stage but not ex vivo chloroquine susceptibility. *Antimicrob Agents Chemother* 60(1): 361–367.
47. Dharia NV, et al. (2010) Whole-genome sequencing and microarray analysis of ex vivo Plasmodium vivax reveal selective pressure on putative drug resistance genes. *Proc Natl Acad Sci USA* 107(46):20045–20050.
48. Kryazhimskiy S, Plotkin JB, Plotkin JB (2008) The population genetics of dN/dS. *PLoS Genet* 4(12):e1000304.
49. Duffy CW, et al. (2015) Comparison of genomic signatures of selection on Plasmodium falciparum between different regions of a country with high malaria endemicity. *BMC Genomics* 16:527.
50. Daniels RF, et al. (2015) Modeling malaria genomics reveals transmission decline and rebound in Senegal. *Proc Natl Acad Sci USA* 112(22):7067–7072.
51. Mok S, et al. (2015) Drug resistance. Population transcriptomics of human malaria parasites reveals the mechanism of artemisinin resistance. *Science* 347(6220): 431–435.
52. Zeeman A-M, der Wel AV, Kocken CHM (2013) Ex vivo culture of Plasmodium vivax and Plasmodium cynomolgi and in vitro culture of Plasmodium knowlesi blood stages. *Methods Mol Biol* 923:35–49.
53. Vaughan AM, Kappe SHI, Ploss A, Mikolajczak SA (2012) Development of humanized mouse models to study human malaria parasite infection. *Future Microbiol* 7(5): 657–665.
54. Mikolajczak SA, et al. Plasmodium vivax liver stage development and hypnozoite persistence in human liver-chimeric mice. *Cell Host Microbe* 17(4):526–35.
55. Joyner C, Barnwell JW, Galinski MR (2015) No more monkeying around: Primate malaria model systems are key to understanding Plasmodium vivax liver-stage biology, hypnozoites, and relapses. *Front Microbiol* 6:145.
56. Moraes Barros RR, et al. (2015) Editing the Plasmodium vivax genome, using zinc-finger nucleases. *J Infect Dis* 211(1):125–129.
57. Mok S, et al. (2014) Structural polymorphism in the promoter of pfmrp2 confers Plasmodium falciparum tolerance to quinoline drugs. *Mol Microbiol* 91(5):918–934.
58. Gonzales JM, et al. (2008) Regulatory hotspots in the malaria parasite genome dictate transcriptional variation. *PLoS Biol* 6(9):e238.
59. Mideo N, et al. (2016) A deep sequencing tool for partitioning clearance rates following antimalarial treatment in polyclonal infections. *Evol Med Public Health* 2016(1):21–36.
60. Lin JT, et al. (2015) Using amplicon deep sequencing to detect genetic signatures of Plasmodium vivax relapse. *J Infect Dis* 212(6):999–1008.
61. Friedrich LR, et al. (2016) Complexity of infection and genetic diversity in Cambodian Plasmodium vivax. *PLoS Negl Trop Dis* 10(3):e0004526.
62. Nkhoma SC, et al. (2013) Population genetic correlates of declining transmission in a human pathogen. *Mol Ecol* 22(2):273–285.
63. Hupalo DN, et al. (2016) Population genomics studies identify signatures of global dispersal and drug resistance in Plasmodium vivax. *Nat Genet* 48(8):953–958.
64. Molina-Cruz A, et al. (2013) The human malaria parasite Pfs47 gene mediates evasion of the mosquito immune system. *Science* 340(6135):984–987.
65. Durnez L, et al. (2013) Outdoor malaria transmission in forested villages of Cambodia. *Malar J* 12:329.
66. Sinka ME, et al. (2011) The dominant Anopheles vectors of human malaria in the Asia-Pacific region: Occurrence data, distribution maps and bionomic précis. *Parasit Vectors* 4(1):89.
67. Anderson TJ, et al. (2000) Microsatellite markers reveal a spectrum of population structures in the malaria parasite Plasmodium falciparum. *Mol Biol Evol* 17(10): 1467–1482.
68. Leclerc MC, et al. (2004) Meager genetic variability of the human malaria agent Plasmodium vivax. *Proc Natl Acad Sci USA* 101(40):14455–14460.
69. Lim CS, Tazi L, Ayala FJ (2005) Plasmodium vivax: Recent world expansion and genetic identity to Plasmodium simium. *Proc Natl Acad Sci USA* 102(43):15523–15528.
70. Prugnolle F, et al. (2013) Diversity, host switching and evolution of Plasmodium vivax infecting African great apes. *Proc Natl Acad Sci USA* 110(20):8123–8128.
71. Liu W, et al. (2014) African origin of the malaria parasite Plasmodium vivax. *Nat Commun* 5:3346.
72. Sundararaman SA, et al. (2016) Genomes of cryptic chimpanzee Plasmodium species reveal key evolutionary events leading to human malaria. *Nat Commun* 7:11078.
73. Nair S, et al. (2003) A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. *Mol Biol Evol* 20(9):1526–1536.
74. Volkman SK, et al. (2007) A genome-wide map of diversity in Plasmodium falciparum. *Nat Genet* 39(1):113–119.
75. Snounou G, et al. (1999) Biased distribution of msp1 and msp2 allelic variants in Plasmodium falciparum populations in Thailand. *Trans R Soc Trop Med Hyg* 93(4): 369–374.
76. Paul RE, et al. (1998) Transmission intensity and Plasmodium falciparum diversity on the northwestern border of Thailand. *Am J Trop Med Hyg* 58(2):195–203.
77. Jongwutiwes S, Putaporntip C, Hughes AL (2010) Bottleneck effects on vaccine-candidate antigen diversity of malaria parasites in Thailand. *Vaccine* 28(18): 3112–3117.
78. Smith RC, Vega-Rodríguez J, Jacobs-Lorena M (2014) The Plasmodium bottleneck: Malaria parasite losses in the mosquito vector. *Mem Inst Oswaldo Cruz* 109(5):644–661.
79. Chang H-H, Hartl DL (2015) Recurrent bottlenecks in the malaria life cycle obscure signals of positive selection. *Parasitology* 142(Suppl 1):S98–S107.
80. Neafsey DE, et al. (2012) The malaria parasite Plasmodium vivax exhibits greater genetic diversity than Plasmodium falciparum. *Nat Genet* 44(9):1046–1050.
81. Hester J, et al. (2013) De novo assembly of a field isolate genome reveals novel Plasmodium vivax erythrocyte invasion genes. *PLoS Negl Trop Dis* 7(12):e2569.
82. Beshir KB, et al. (2010) Measuring the efficacy of anti-malarial drugs in vivo: Quantitative PCR measurement of parasite clearance. *Malar J* 9:312.
83. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv.org*. Available at https://arxiv.org/abs/1303.3997. Accessed September 30, 2016.
84. DePristo MA, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5):491–498.
85. McKenna A, et al. (2010) The Genome Analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297–1303.
86. Auburn S, et al. (2012) Characterization of within-host Plasmodium falciparum diversity using next-generation sequence data. *PLoS One* 7(2):e32891.
87. Manske M, et al. (2012) Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. *Nature* 487(7407):375–379.
88. Staab PR, Metzler D (2016) Coala: An R framework for coalescent simulation. *Bioinformatics* 32(12):1903–1904.
89. Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18(2):337–338.
90. Schlamp F, et al. (2016) Evaluating the performance of selection scans to detect selective sweeps in domestic dogs. *Mol Ecol* 25(1):342–356.
91. Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. *PLoS Biol* 4(3):e72.
92. McVean G, Awadalla P, Fearnhead P (2002) A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* 160(3):1231–1241.
93. Szpiech ZA, Hernandez RD (2014) selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. *Mol Biol Evol* 31(10):2824–2827.
94. Layer RM, Chiang C, Quinlan AR, Hall IM (2014) LUMPY: A probabilistic framework for structural variant discovery. *Genome Biol* 15(6):R84.
95. Zimmerman PA, Mehlotra RK, Kasehagen LJ, Kazura JW (2004) Why do we need to know more about mixed Plasmodium species infections in humans? *Trends Parasitol* 20(9):440–447.
96. Benson G (1999) Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res* 27(2):573–580.
97. Köster J, Rahmann S (2012) Snakemake–a scalable bioinformatics workflow engine. *Bioinformatics* 28(19):2520–2522.
98. Robinson JD, Coffman AJ, Hickerson MJ, Gutenkunst RN (2014) Sampling strategies for frequency spectrum-based population genomic inference. *BMC Evol Biol* 14:254.
99. Chang H-H, et al. (2013) Malaria life cycle intensifies both natural selection and random genetic drift. *Proc Natl Acad Sci USA* 110(50):20129–20134.
100. Menard D, et al. (2013) 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(11):e2489.
101. Trenholme KR, et al. (2000) clag9: A cytoadherence gene in Plasmodium falciparum essential for binding of parasitized erythrocytes to CD36. *Proc Natl Acad Sci USA* 97(8):4029–4033.