



# Single-Nucleotide Polymorphism, Linkage Disequilibrium and Geographic Structure in the Malaria Parasite *Plasmodium vivax*: Prospects for Genome-Wide Association Studies

## Citation

Orjuela-Sánchez, Pamela, Nadira D. Karunaweera, Mônica da Silva-Nunes, Natal S. da Silva, Kézia K. G. Scopel, Raquel M. Gonçalves, Chanaki Amaratunga, et al. 2010. Single-nucleotide polymorphism, linkage disequilibrium and geographic structure in the malaria parasite *Plasmodium vivax*: prospects for genome-wide association studies. *BMC Genetics* 11:65.

## Published version

<https://doi.org/10.1186/1471-2156-11-65>

## Link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:4551790>

## Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

## Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

## Share Your Story

The Harvard community has made this article openly available. Please share how this access benefits you. [Submit a story](#)

RESEARCH ARTICLE

Open Access

# Single-nucleotide polymorphism, linkage disequilibrium and geographic structure in the malaria parasite *Plasmodium vivax*: prospects for genome-wide association studies

Pamela Orjuela-Sánchez<sup>1</sup>, Nadira D Karunaweera<sup>2</sup>, Mônica da Silva-Nunes<sup>1,3</sup>, Natal S da Silva<sup>1</sup>, Kézia KG Scopel<sup>1,4</sup>, Raquel M Gonçalves<sup>1</sup>, Chanaki Amaratunga<sup>5</sup>, Juliana M Sá<sup>5</sup>, Duong Socheat<sup>6</sup>, Rick M Fairhurst<sup>5</sup>, Sharmini Gunawardena<sup>2</sup>, Thurasamy Thavakodirasah<sup>7</sup>, Gawrie LN Galapaththy<sup>7</sup>, Rabindra Abeysinghe<sup>7</sup>, Fumihiko Kawamoto<sup>8</sup>, Dyann F Wirth<sup>9</sup> and Marcelo U Ferreira\*<sup>1</sup>

## Abstract

**Background:** The ideal malaria parasite populations for initial mapping of genomic regions contributing to phenotypes such as drug resistance and virulence, through genome-wide association studies, are those with high genetic diversity, allowing for numerous informative markers, and rare meiotic recombination, allowing for strong linkage disequilibrium (LD) between markers and phenotype-determining loci. However, levels of genetic diversity and LD in field populations of the major human malaria parasite *P. vivax* remain little characterized.

**Results:** We examined single-nucleotide polymorphisms (SNPs) and LD patterns across a 100-kb chromosome segment of *P. vivax* in 238 field isolates from areas of low to moderate malaria endemicity in South America and Asia, where LD tends to be more extensive than in holoendemic populations, and in two monkey-adapted strains (Salvador-I, from El Salvador, and Belem, from Brazil). We found varying levels of SNP diversity and LD across populations, with the highest diversity and strongest LD in the area of lowest malaria transmission. We found several clusters of contiguous markers with rare meiotic recombination and characterized a relatively conserved haplotype structure among populations, suggesting the existence of recombination hotspots in the genome region analyzed. Both silent and nonsynonymous SNPs revealed substantial between-population differentiation, which accounted for ~40% of the overall genetic diversity observed. Although parasites clustered according to their continental origin, we found evidence for substructure within the Brazilian population of *P. vivax*. We also explored between-population differentiation patterns revealed by loci putatively affected by natural selection and found marked geographic variation in frequencies of nucleotide substitutions at the *pvmdr-1* locus, putatively associated with drug resistance.

**Conclusion:** These findings support the feasibility of genome-wide association studies in carefully selected populations of *P. vivax*, using relatively low densities of markers, but underscore the risk of false positives caused by population structure at both local and regional levels.

**See commentary:** <http://www.biomedcentral.com/1741-7007/8/90>

## Background

*Plasmodium vivax*, the most widespread of the four human malaria parasites, causes 132 to 391 million epi-

sodes of disease each year [1], with 2.6 billion people at risk of infection worldwide [2]. Outside of Africa, *P. vivax* is the main cause of malaria morbidity, with enormous public health burden. The recent emergence of drug-resistant strains and severe (sometimes fatal) disease challenges the traditional view of vivax malaria as a benign infection and calls for new strategies to examine

\* Correspondence: [muferre@usp.br](mailto:muferre@usp.br)

<sup>1</sup> Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, 05508-900 São Paulo, São Paulo, Brazil

Full list of author information is available at the end of the article

molecular mechanisms underlying drug resistance and increased virulence [3].

Linkage analysis of laboratory crosses and population-based genome-wide association studies are powerful approaches to map genetic loci contributing to drug resistance and virulence in malaria parasites [4]. Experimental crosses of *P. vivax* are currently limited by the lack of practical methods for its long-term propagation and cloning in vitro [5], which are required to characterize phenotypes in the progeny [4]. The ideal populations for allelic association studies are those with high prevalences of the phenotype of interest and relatively high levels of genetic diversity, allowing for numerous informative markers across the genome. Association studies are usually most cost-effective when low meiotic recombination rates are present in these populations, allowing for strong linkage between markers and phenotype-determining loci. The phenotype-associated allele is presumed to have emerged on a single or a few haplotype backgrounds. Crossover events during meiosis tend to randomize the initial haplotype(s) and to reduce population-level associations, known as linkage disequilibrium (LD), between flanking markers and candidate phenotype-determining loci. Therefore, to determine the feasibility of association studies we need a detailed picture of the overall diversity and LD landscape across the genome of *P. vivax*. However, patterns of genetic diversity and LD remain little characterized in field populations of this parasite and most nucleotide polymorphism data currently available are for antigen-coding genes [6]. These data may not be informative of genome-wide patterns because of potential biases introduced by natural selection on particular phenotypes [7]. Due to the lack of continuous culture in vitro, only field samples of *P. vivax*, which are heavily contaminated with host's DNA, are available for use in next-generation sequencing projects for large-scale characterization of new genetic markers.

Microsatellite markers have revealed clear geographic differences in levels of genetic diversity and LD in *P. falciparum* isolates sampled from four continents. Diversity was highest and LD lowest in populations from holoendemic Africa, while diversity was lowest and LD highest in populations from hypoendemic South America, with intermediate patterns seen in Southeast Asia. Parasite populations clustered according to their continental origins, with most variation found within locations in highly endemic areas. Nevertheless, substantial divergence was seen between subpopulations in South America [8]. Putatively neutral microsatellite markers, sampled from across the genome [9], have recently been used to characterize field populations of *P. vivax* [10-12], suggesting that a spectrum of population structures also exists for this species. Again, South American parasite populations sampled from nearby sites are highly divergent [10,11].

However, because the microsatellite markers analyzed map to different chromosomes, prior studies could not examine chromosome-level LD and infer recombination rates in *P. vivax*. In addition, microsatellites mutate at very high rates and other neutral genetic markers with different rate and mode of evolution, such as intergenic or synonymous single-nucleotide polymorphisms (SNPs), could offer a different picture [13]. Significant differences detected with rapidly evolving markers, which may remain undetected with more conserved markers, do not necessarily translate into biologically meaningful differences among populations [14].

To explore the potential for future genome-wide association studies, we examined SNP diversity and LD across a 100-kb chromosome segment of *P. vivax*. We sampled parasites from areas of low to moderate endemicity in South America and Asia, where LD tends to be more extensive than in holoendemic populations. We show varying levels of SNP diversity and LD across populations (highest diversity and LD in the area of lowest malaria transmission, Sri Lanka), with substantial genetic differentiation among populations. Frequencies of nucleotide substitutions at *pvm-dr-1* gene, putatively associated with drug resistance, varied markedly across locations. Although these findings support the use of genome-wide association approaches to map genes underlying drug resistance and other traits in *P. vivax*, they underscore the risk of false positives if population structure, at both local and regional levels, is left uncorrected in association studies.

## Results and Discussion

### Chromosome-level SNP diversity

We examined chromosome-level SNP diversity and LD in 238 field isolates of *P. vivax* from areas of low to intermediate levels of malaria endemicity and two monkey-adapted strains, Belém and Salvador-I. The field isolates originated from three sites in the Amazon Basin of Brazil (Granada, Plácido de Castro and Porto Velho) and three sites (Pursat, Cambodia; Bao Loc, Vietnam; Tricomalee, Sri Lanka) across South and Southeast Asia (Additional file 1 Table S1). We assayed 85 SNPs across 100-kb of contiguous DNA sequence on chromosome 8. Of them, 57 (67.1%) segregated in at least one of the study locations. The number of SNPs that segregated in parasites from Brazil, Cambodia, Sri Lanka and Vietnam were 43, 35, 44 and 24, respectively. Only 13 (15.3%) segregated in all six locations; 8 segregated in Brazil alone and 14 segregated in Asia alone.

We measured SNP  $\pi$ , the average proportion of pairwise differences at assayed SNP loci [15], to compare diversity across study sites; monkey-adapted strains were not considered in this analysis. We obtained the following SNP  $\pi$  values (standard errors in parentheses): Granada =

0.1264 (0.0006); Plácido de Castro = 0.1298 (0.0026); Porto Velho = 0.1589 (0.0110); Brazil (three sites combined) = 0.1364 (0.0005); Cambodia = 0.1163 (0.0012); Vietnam = 0.0842 (0.0034); Sri Lanka = 0.15476 (0.0063); Asia (three sites combined) = 0.1401 (0.0010). The overall SNP  $\pi$  value for Brazilian populations of *P. vivax* is identical to the estimate obtained by Neafsey and colleagues for 11 *P. falciparum* isolates from this country that were assayed for 1638 SNPs across the whole genome [15].

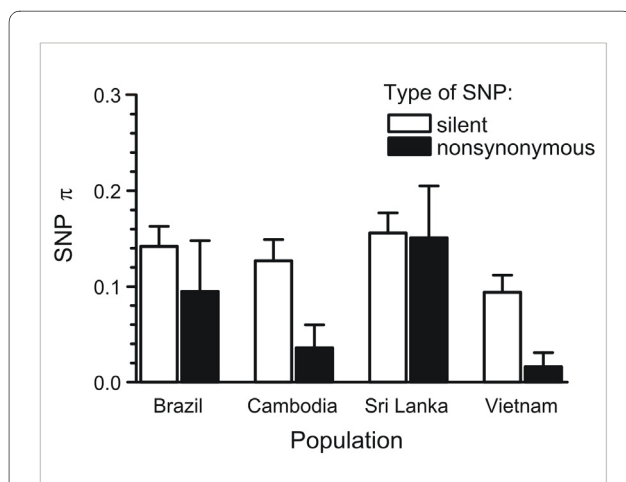
The observed ranking of country-specific numbers of segregating sites and SNP  $\pi$  values (Sri Lanka > Brazil > Cambodia > Vietnam) does not match the ranking of malaria endemicity at the time of sample collection (Vietnam > Cambodia > Brazil > Sri Lanka). The trend towards a negative correlation between levels of genetic diversity and malaria transmission contrasts with the pattern observed for *P. falciparum*, for which highest diversity is seen in high-transmission settings [8,15]. This contrast is surprising and may reflect differences in the demographic history and biology of these major human parasites. In all countries but Sri Lanka, SNP  $\pi$  values were significantly higher for 75 silent (synonymous or noncoding) SNPs than for 10 nonsynonymous SNPs (Figure 1), consistent with nonsynonymous SNPs being often subject to purifying selection in *P. vivax* populations, as previously suggested for *P. falciparum* [15].

Parasites were systematically sampled over 30 months (April 2004 through October 2006) in one of the study sites in Brazil, Granada [11,16]. We hypothesized that local parasite populations gradually diversify over time as a result of migration, genetic drift, mutation and recombination. Accordingly, the proportion of pairwise SNP

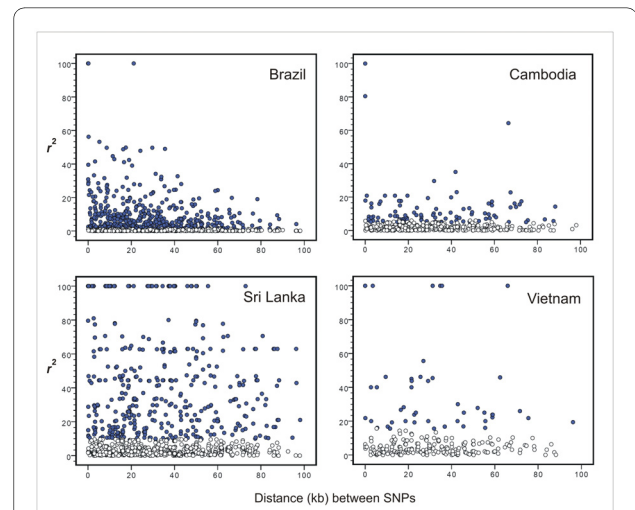
differences correlated positively with the temporal distance between dates of collection of pairs of isolates in Granada ( $r = 0.138$ ,  $P < 0.001$ , Mantel correlation test). The correlation remained significant when only silent SNPs were considered ( $r = 0.146$ ,  $P < 0.001$ ), but not when only nonsynonymous sites were considered ( $r = 0.007$ ,  $P = 0.343$ ). The little variation at nonsynonymous SNP loci over time underscores the potential biases introduced by natural selection in studies of malaria parasite diversity. Whether or not similar temporal patterns of genetic divergence occur in other endemic areas remains to be investigated.

#### Linkage disequilibrium and haplotype blocks

The extent of LD between pairs of markers across a chromosome is expected to decline at a rate that is proportional to the population recombination rate. In fact, LD (measured with the  $r^2$  statistic) decayed with increasing physical distance between pairs of segregating sites in the Brazilian population (Figure 2), with similar results when two subpopulations from this country, Granada and Plácido de Castro, were considered separately to remove the putative effect of population substructuring on LD levels (Additional file 2 Figure S1). However, no significant correlation between  $r^2$  and intermarker distance was found in Cambodia, Sri Lanka, or Vietnam (Figure 2), with significant LD often extending over the entire chromosome segment analyzed. Accordingly, the proportion of pairs of



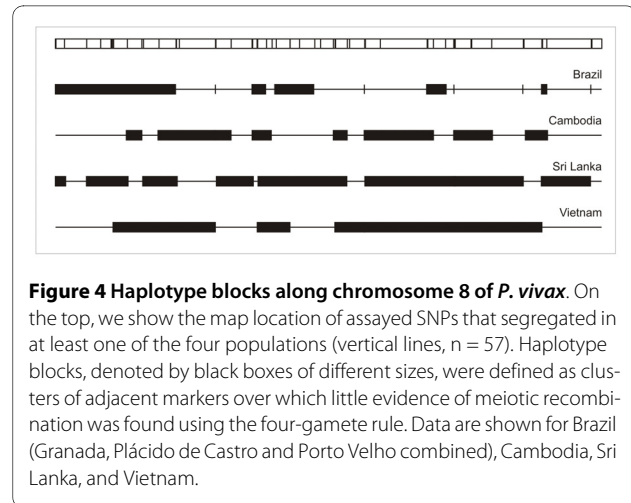
**Figure 1 Diversity at silent and nonsynonymous SNPs (SNP  $\pi$ ) across chromosome 8 by *P. vivax* population.** Error bars indicate standard errors of the means, derived from bootstrapping. Significantly lower nonsynonymous SNP diversity was found in all populations but Sri Lanka ( $P < 0.001$  by Wilcoxon test).



**Figure 2 Variation in LD in relation to physical distance between SNP loci along chromosome 8 of *P. vivax*.** Data are shown for Brazil (Granada, Plácido de Castro and Porto Velho combined), Cambodia, Sri Lanka, and Vietnam. LD was measured with the  $r^2$  statistic (values on y-axis are multiplied by 100); closed (blue) circles denote significant LD ( $\chi^2$  statistic,  $P < 0.05$ ), while open circles denote nonsignificant LD. A significant negative correlation between  $r^2$  values and physical distance between SNPs (in kb) was found only for Brazil (Pearson correlation coefficient  $r = -0.195$ ,  $P < 0.0001$ ).

segregating sites with significant LD declined with increasing map distance in the Brazilian population, but not in populations from areas with the highest endemicity, Cambodia and Vietnam (Figure 3). These results suggest that initial genome-wide association mapping, using relatively low densities of marker loci, is feasible in natural *P. vivax* populations from areas of low to moderate malaria endemicity with the features observed in Brazil, where LD can persist over several kb but clearly declines with increasing intermarker distance. Among parasites from the other sites analyzed, recombination rates may be too low to allow for cost-effective association studies, since the persistence of significant LD over long chromosome segments may lead to frequent false-positive associations.

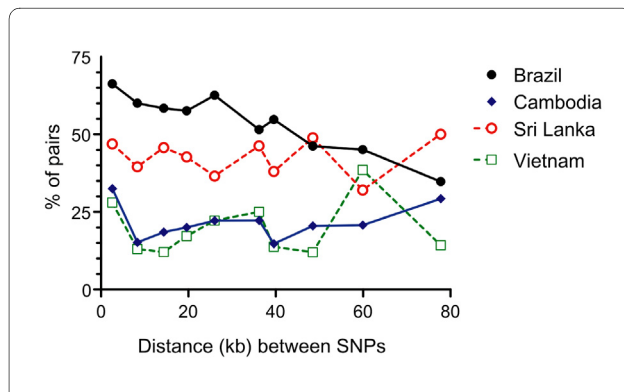
The 100-kb chromosome segment analyzed comprised several clusters of adjacent markers over which little evidence of meiotic recombination was found using the four-gamete rule. These haplotype blocks varied in number and length across the four populations examined (Figure 4). The greatest number of blocks and the smallest average block size (3.6 kb [range, < 100 bp to 21 kb] and 3.4 kb [range, < 100 bp to 9 kb]) were found in Brazil and Cambodia, while the average block sizes for Sri Lanka and Vietnam were 7.4 (range, 1 to 14 kb) and 14.0 kb (range, 5 to 21 kb). Whether the relatively large blocks in these two populations are artefacts resulting from the small sample size remains to be investigated. Haplotype blocks can be further used to define the minimal set of segregating SNPs required to capture most variation in each population, by selecting a single tagging marker within each block (and assaying all SNP loci outside blocks). These minimal sets would comprise 56% (24 of



**Figure 4 Haplotype blocks along chromosome 8 of *P. vivax*.** On the top, we show the map location of assayed SNPs that segregated in at least one of the four populations (vertical lines,  $n = 57$ ). Haplotype blocks, denoted by black boxes of different sizes, were defined as clusters of adjacent markers over which little evidence of meiotic recombination was found using the four-gamete rule. Data are shown for Brazil (Granada, Plácido de Castro and Porto Velho combined), Cambodia, Sri Lanka, and Vietnam.

43) segregating markers in Brazil, 63% (22 of 35) in Cambodia, 43% (19 of 44) in Sri Lanka and 50% (12 of 24) in Vietnam.

We next examined whether the overall haplotype structure was conserved in different populations. To determine to which extent haplotype block boundaries were shared across populations, we calculated the proportions of pairs of markers assigned to the same block (i. e., non-recombining sites) and to different blocks or no block (i. e., recombining sites) in pairwise population comparisons. This analysis was limited to SNPs that were segregating in both populations analyzed (99 to 227 SNP assignments compared). Overall, the vast majority of SNP pairs had a concordant assignment, especially in comparisons involving Brazil and Cambodia (93%), Vietnam and Sri Lanka (91%), and Brazil and Sri Lanka (83%). These data are consistent with a conserved haplotype structure across *P. vivax* populations with varying levels of LD, as previously shown for *P. falciparum* [17]. Lower proportions of concordant assignments were found, however, in comparisons between parasite populations from Vietnam and Cambodia (73%) and Brazil and Vietnam (60%). Most discrepancies were due to marker pairs that were in LD in Sri Lanka and Vietnam but not in Brazil and Cambodia. The vast majority (81-100%) of SNPs with concordant assignment between pairs of populations were silent. We conclude that haplotype block boundaries are shared by parasite populations with different geographic origins, suggesting the existence of conserved recombination hotspots in the genomic region analyzed, with clear implications for future association studies.



**Figure 3 Proportions of informative SNP pairs that display significant LD plotted at various intermarker distances.** Data are shown for Brazil (Granada, Plácido de Castro and Porto Velho combined), Cambodia, Sri Lanka, and Vietnam. Pairwise physical distances between markers (in kb) were divided into deciles; the average distance within each decile is used to plot values. Note that the proportion of pairs with significant LD decreases with intermarker distance only in Brazil.

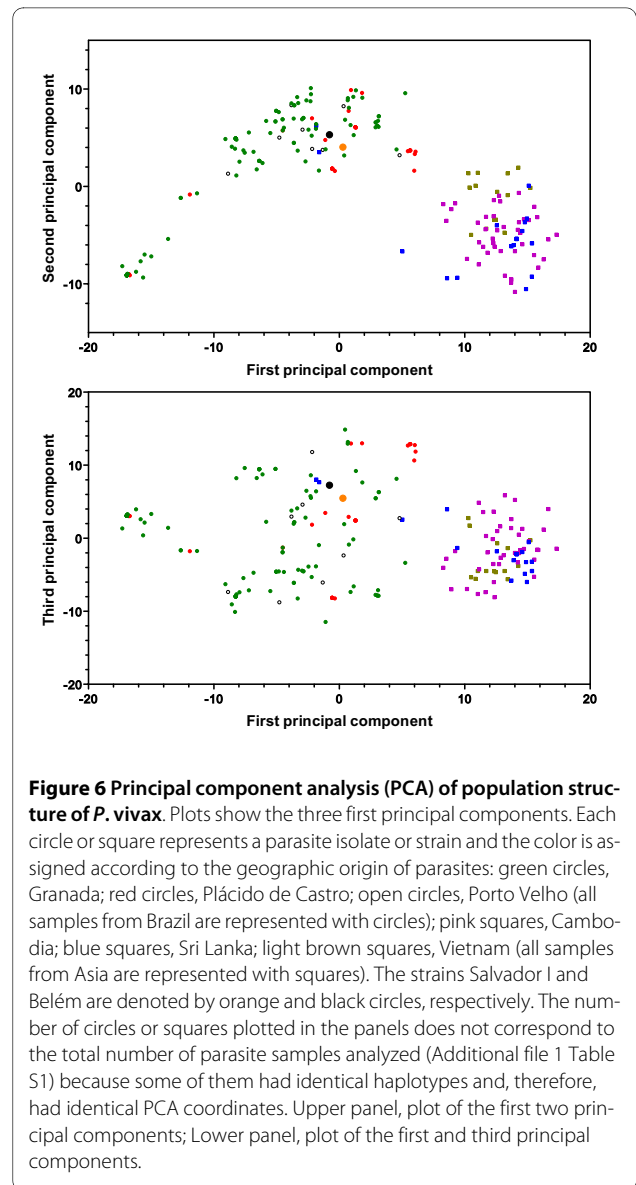
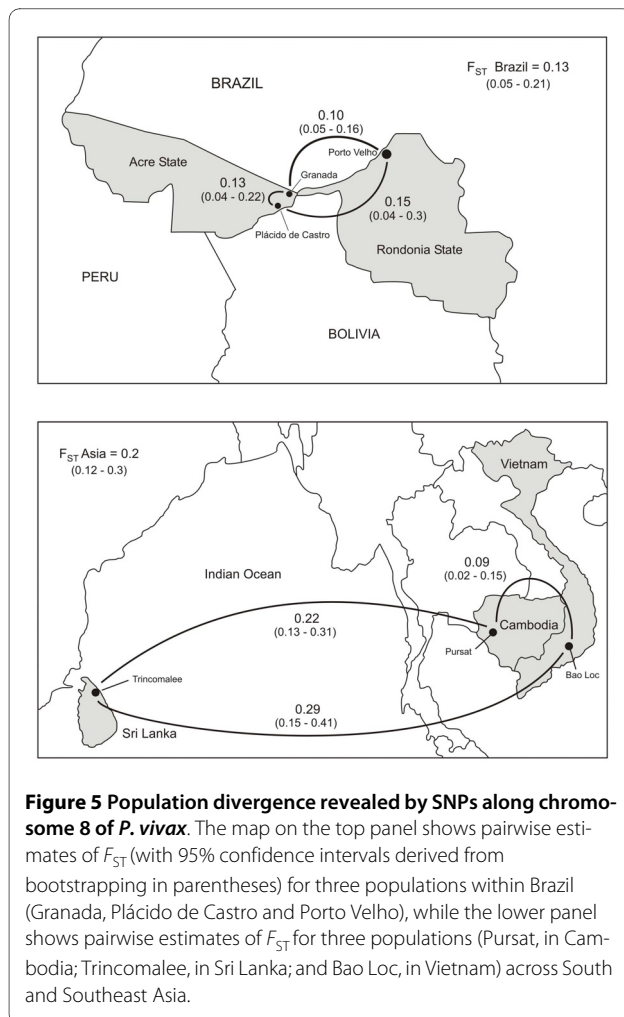
#### Population structure at local and regional levels

Analysis of population differentiation using the  $\theta$  estimator of  $F_{ST}$  statistics revealed substantial chromosome-level differentiation between Brazilian and Asian samples ( $F_{ST} = 0.228$ , 95% confidence interval [CI] 0.132-0.247).



Significant divergence was found in all pairwise comparisons within Brazil and Asia (Figure 5), although relatively low divergence was found between Cambodian and Vietnamese samples, which were collected more than 10 years apart. The overall  $F_{ST}$  value of 0.393 (all locations considered) indicates that a considerable proportion (~40%) of the diversity at assayed SNP loci results from differentiation among geographic populations of *P. vivax*. For all between-population comparisons, except Cambodia versus Sri Lanka, we observed a greater  $F_{ST}$  for silent SNPs relative to nonsynonymous SNPs (Additional file 3 Table S2), suggesting that purifying selection may constrain estimates of population differentiation.

Not surprisingly, principal component analysis (PCA) [18] defined two major clusters that reflect the continental origin of samples (Figure 6). The first major cluster, characterized with the first and second principal components (which, together, explain 38% of the variance), comprised all Brazilian samples, the monkey-adapted strains Belém (from Brazil) and Salvador-I (from El Salvador), and two samples from Sri Lanka, whereas the second



clusters comprised the remaining Asian samples. By combining the first and third principal components (35% of the variance explained), we observed a greater dispersal of Brazilian samples, a few of which clustered together with Asian samples (Figure 6). We repeated the analysis with a set of 29 SNPs selected to minimize intermarker LD, with quite similar results (Additional file 4 Figure S2), indicating that the observed clustering pattern was not an artifact arising from the interdependence of segregating sites. We also repeated the analysis after excluding the populations with the smallest sample size, Porto Velho and Vietnam, but observed the same clustering pattern, with two isolates from Sri Lanka grouped with those from Granada and Plácido de Castro (data not shown).

As previously shown for *P. falciparum* populations [15,19], PCA provided evidence for substantial substructure

ture within the Brazilian population of *P. vivax*. We thus analyzed separately this population and found that parasites did not cluster according to their collection site (Granada, Plácido de Castro or Porto Velho). In fact, most minor clusters in Brazil comprised samples from at least two of these locations (Additional file 5 Figure S3). We also carried a separate analysis of Asian samples to determine whether further substructuring was apparent. We were unable to differentiate between Cambodian and Vietnamese parasites, but this analysis revealed a large dispersal, with some substructuring, in the Sri Lankan population (Additional file 6 Figure S4). The heterogeneity of parasites from nearby locations in Brazil and from a single outbreak in Sri Lanka indicates that correcting for population structure may be required in future association studies in these endemic sites. Otherwise, there may be a spurious association between a phenotype of interest with varying prevalence among subpopulations (for example, from different locations within the same country) and any candidate genetic marker that display allele frequency differences across these subpopulations.

The finding of population structure within Brazil was further supported by the Bayesian clustering procedure implemented by STRUCTURE software [20]. This detected, with strong statistical support (posterior probability > 0.999), three major populations in the whole dataset. Nearly all field isolates from Brazil (in addition to Belém and Salvador-I strains) had predominant ancestry in one of two populations, represented in blue and red in Figure 7. In contrast, nearly all parasites from Asia had their predominant ancestry in a third population, represented in green in Figure 7. Only four isolates failed to cluster according to their continent of origin; exceptions were two isolates from Brazil with a predominant membership in the Asian (green) population and two isolates

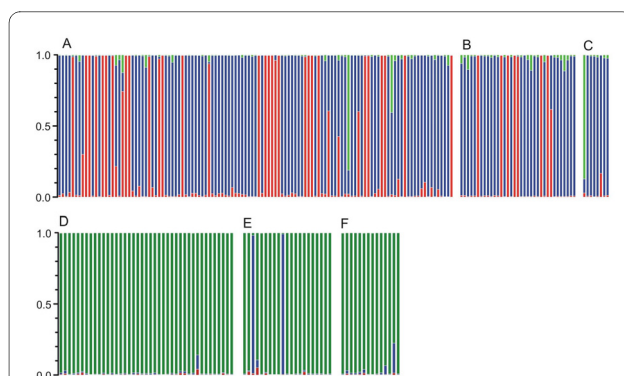
from Sri Lanka with clear membership in the blue subpopulation from Brazil. Granada and Plácido de Castro, the largest populations in Brazil, comprised parasites with predominant membership in the blue or red populations, while Porto Velho included only parasites with predominant membership in the blue population. A separate STRUCTURE analysis of the Brazilian population confirmed the subdivision into two major populations, without clear further substructuring (data not shown).

#### Natural selection and population differentiation

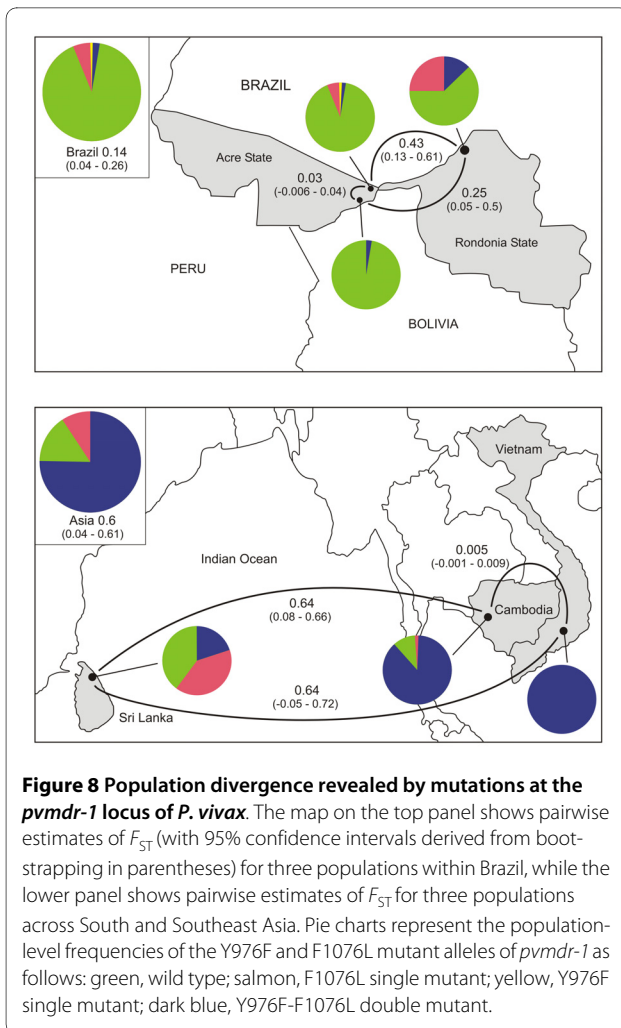
We next explored between-population differentiation patterns revealed by loci putatively affected by natural selection. We scored SNPs at two loci, *pvcr-t-o* and *pvmdr-1*, encoding digestive-vacuole membrane proteins that can be involved in chloroquine (CQ) resistance in *P. vivax* (Additional file 7 Figure S5). None of the five non-synonymous SNPs assayed in the *pvcr-t-o* gene was found to segregate in any of the parasite populations examined. In contrast, the six SNP sites analyzed within the *pvmdr-1* gene segregated in most populations. Allele frequencies at *pvmdr-1* varied markedly between locations, with an overall  $F_{ST}$  of 0.705. We found very little differentiation between neighboring locations within Brazil (Granada versus Plácido de Castro) and Asia (Cambodia versus Vietnam) (Figure 8), with large differentiation in inter-continental comparisons (Brazil versus Cambodia,  $F_{ST}$  = 0.746; Brazil versus Vietnam,  $F_{ST}$  = 0.755; Brazil versus Sri Lanka,  $F_{ST}$  = 0.528). The low  $F_{ST}$  value estimated for the comparison between Cambodian and Vietnamese parasites is particularly noteworthy, given the long time interval between dates of sample collection in these sites.

We focused on two nonsynonymous substitutions in *pvmdr-1* thought to be associated with CQ resistance, Y976F and F1076L [21-23]. Double-mutant alleles accounted for all or nearly all samples from Cambodia and Vietnam, while wild-type alleles predominated in Brazil. An intermediate pattern was found in Sri Lanka (Figure 8). All but one single-mutant allele carried the F1076L change; the only allele carrying the Y976F change alone (i. e., not co-occurring with the F1076L change) came from Granada, Brazil. These findings lend further support to the hypothesis that a two-step mutational trajectory (F1076L followed by Y976F) at the *pvmdr-1* locus leads to CQ resistance [21]. If this hypothesis is correct, molecular detection of F1076L single mutants may provide an early warning about the risk of emerging CQ resistance before the drug-resistant phenotype itself can be detected in populations.

The finding of parasites with wild-type *pvmdr-1* alleles that are CQ-resistant [22,24] suggests that other polymorphisms may contribute to this phenotype. Similar to *P. falciparum* [25], in vitro susceptibility to CQ in *P. vivax* appears to be modulated by *pvmdr-1* copy number. Gene



**Figure 7** STRUCTURE analysis of population structure of *P. vivax*. Each bar represents a parasite isolate and bar partitioning represents the proportion of ancestry of each isolate assigned to three populations, represented with different colors (blue, red or green). **A**, Granada; **B**, Plácido de Castro; **C**, Porto Velho (**A**, **B** and **C** are in Brazil); **D**, Cambodia; **E**, Sri Lanka; **F**, Vietnam.



**Figure 8 Population divergence revealed by mutations at the *pvmdr-1* locus of *P. vivax*.** The map on the top panel shows pairwise estimates of  $F_{ST}$  (with 95% confidence intervals derived from bootstrapping in parentheses) for three populations within Brazil, while the lower panel shows pairwise estimates of  $F_{ST}$  for three populations across South and Southeast Asia. Pie charts represent the population-level frequencies of the Y976F and F1076L mutant alleles of *pvmdr-1* as follows: green, wild type; salmon, F1076L single mutant; yellow, Y976F single mutant; dark blue, Y976F-F1076L double mutant.

amplification correlates with increased susceptibility to CQ and decreased susceptibility to amodiaquine, artesunate and mefloquine [26]. Different country-specific drug policies may therefore favor parasites with increased *pvmdr-1* copy number or select for Y976F alleles, further complicating geographic comparisons of allele frequencies and associated phenotypes.

#### Prospects for genome-wide population association studies

Here we describe long-ranging chromosome-level LD and relatively conserved haplotype blocks in *P. vivax* populations from areas with low to moderate levels of malaria transmission. The most favorable conditions for association studies with relatively low marker density were observed in Brazil, where parasites were reasonably diverse (SNP  $\pi$  values comparable to those estimated for local *P. falciparum* populations) and strong LD was observed at relatively short intermarker distances (~40 kb) but gradually declined with increasing physical distance between pairs of markers. In contrast, the similar levels of LD, along the whole 100-kb region analyzed, in

the Asian populations may increase the probability of detecting false-positive associations between phenotype-associated loci and genetic markers located at considerable map distances. These findings, although limited to a single chromosome segment that comprises only 0.4% of the whole genome of the parasite, indicate that genome-wide association studies can represent a feasible strategy to map genetic regions associated with drug resistance, virulence and other phenotypes of interest in carefully selected *P. vivax* populations. Genome-wide studies with higher marker density are required to confirm these findings.

We have also observed substantial genetic differentiation among populations, at both local and regional levels. Consistent with previous studies of *P. falciparum* [27], we found significant geographic structure revealed by synonymous SNPs, which are putatively free of strong directional selection. In addition, we found large differences in the frequencies of nucleotide substitutions at the *pvmdr-1* locus among populations. We note that, because of the underlying geographic structure, allele frequency differences observed among populations may be unrelated to the genetics of the particular phenotype under study, resulting in false-positive results or reduced power of association studies. Even when studies are restricted to a single continental origin, false positives may still result from major differences in the ancestry of local parasites, such as those found in Brazil. Accounting for the geographic structure is a major practical issue in future population-based studies of genetic determinants of *P. vivax* traits. Statistical methods have now been developed [28] to address this issue when ideally homogeneous (unstructured) parasite populations are not available for sampling.

#### Conclusion

We found varying levels of SNP diversity and LD across *P. vivax* populations, with the most favorable conditions for genome-wide association studies, relatively high diversity and strong LD that declines with increasing intermarker distance, observed in Brazil. However, we have also observed substantial genetic differentiation among populations at both local and regional levels, especially in the Brazilian population. Although these results suggest that association studies are feasible in selected *P. vivax* populations, they highlight the need for correcting for population structure to avoid false-positive associations.

#### Methods

##### Geographical parasite sampling

We collected venous or finger-prick blood samples from 432 patients with slide-confirmed *P. vivax* infection from three locations in Brazil and one location each in three Asian countries (Cambodia, Vietnam, and Sri Lanka (Figure 5). All sites in Brazil are characterized by year-round



but hypoendemic malaria transmission, with *P. vivax* prevalence rates typically below 1% [29,30]. The 249 samples from Brazil were collected in the rural settlement of Granada and the town of Plácido de Castro (50 km south of Granada), both in Acre State, and in the city of Porto Velho (500 km east of Granada), in Rondônia State. The samples from Granada (n = 193) were collected during prospective cohort studies between 2004 and 2006 [11,16,30], those from Plácido de Castro (n = 38) were collected in 2008 from patients attending the town's malaria clinic [31], and those from Porto Velho (n = 17) were collected from patients attending the Oswaldo Cruz Outpatient Clinic in June-July 1995 [32]. The 70 samples from Cambodia were collected between June and December 2008 in Pursat town from individuals who became infected in the nearby forests (ClinicalTrials.gov identifier, NCT00663546). In April 2008, a cross-sectional survey of 1056 individuals living in the forest fringe of Pursat province, near the border with Thailand, estimated a *P. vivax* prevalence rate of 2.2% (CA and RMF, unpublished results). The 23 samples from Vietnam were collected in January-December 1995 from patients attending the outpatient clinic of the Lam Dong Provincial Hospital, in the town of Bao Loc, 150 km northwest of Ho Chi Minh City, Lam Dong Province [33]. Prevalence rates of *P. vivax* infection in the rural communities surrounding Bao Loc, on the southern highlands of Vietnam, were estimated to be around 2.5-7.5% at the time of sample collection [33]. The 29 samples from Sri Lanka were collected during a malaria outbreak in Trincomalee, Eastern Province [34]. Between January and August 2007, a total of 87 cases of *vivax* malaria were reported from Trincomalee (Anti-Malaria Campaign, Ministry of Health, Sri Lanka and TT, unpublished information). Over the past few years malaria transmission has declined steadily in Sri Lanka, from 210,039 cases of malaria in 2001 to only 196 cases country-wide in 2007 [35]. In contrast, no drastic reduction in malaria transmission has been documented over the past decade in the other Asian sites included in this study.

DNA from field samples was extracted using standard protocols referenced in the original publications. Additional DNA samples, from monkey-adapted strains, Belém (isolated in Brazil, 1980) and Salvador-I (El Salvador, 1969), were provided by the Malaria Research and Reference Reagent Resource Center (MR4), ATCC (Manassas, United States). To obtain adequate DNA concentrations for SNP assays, parasite DNA was submitted to whole-genome amplification (WGA) prior to typing. WGA was performed on 10 ng of genomic DNA, with high-fidelity multiple displacement technology [36], using a REPLI-g Minikit (Qiagen, Valencia, United States) according to the manufacturer's instructions.

### Single-nucleotide polymorphisms (SNPs)

We identified SNPs across chromosome 8 by aligning 100 kb of contiguous DNA sequence from five *P. vivax* isolates [37,38] (GenBank accession numbers [AY003872](#) and [AY216936-AY216939](#)). DNA sequences were derived from two isolates from Brazil (Belém strain [39] and a field isolate from Rondônia [40]), one from El Salvador (Salvador-I strain [41]), one from India (India VII strain [42]), and one from Thailand (Thai NYU strain [43]). SNPs were selected to fit two criteria: that they are located in nonrepetitive regions and that they are surrounded by 200 bp of upstream sequence and 200 bp of downstream sequence with no significant similarity to human sequences, as determined by BLAST search against the human genome, to prevent cross-amplification of human DNA present in the field-collected test samples. We examined a single chromosome region (spanning ~0.4% of the whole genome) because other genomic regions of *P. vivax* have not been systematically screened for SNPs using a worldwide parasite sample [37]. We designed assays to 108 candidate SNPs meeting the criteria above, which were screened in 50 field-collected samples. Of them, 22 were discarded because alleles were called in less than 10% of these 50 test samples and one was discarded because of cross-amplification from human DNA. The final set comprised 85 markers across chromosome 8, with 75 silent SNPs (39 intergenic, 1 intronic, 9 located in 5' or 3' untranslated regions (UTR) and 26 synonymous nucleotide replacements in open reading frames of genes encoding annotated or hypothetical proteins) and 10 nonsynonymous SNPs (Additional file 8 Table S3).

We also examined nucleotide replacements at two loci encoding digestive-vacuole membrane proteins that are potentially involved in chloroquine (CQ) resistance in *P. vivax* (Additional file 7 Figure S5). Since parasites are exposed to different drug treatment regimes in each country, local adaptation may theoretically result in more geographic structure revealed by these polymorphisms than neutral markers [27]. Although the molecular mechanisms of CQ resistance in *P. vivax* remain unknown, we focused on the two most likely candidate genes. We typed five nonsynonymous SNPs (L47 S, K76T, S250P, F276V, and L384F) in the *pvcr-t-o* gene [44] that were recently found in field isolates of *P. vivax* [44,45]. K76T mutant alleles of the *P. falciparum* orthologue of *pvcr-t-o*, which encodes the protein chloroquine resistance transporter (PfCRT), confer CQ resistance to this species [46], but the limited data available to date fail to support associations between mutations in *pvcr-t-o* alleles and CQ resistance in *P. vivax* [44,45]. We also typed one synonymous (at codon 4065) and five nonsynonymous (N89 S, N500 D, M908L, Y976E, and F1076L) SNPs previously

described at the *multidrug resistance 1* gene of *P. vivax* (*pvmdr-1*) [21,22,24,45], which encodes a P-glycoprotein of the family of ATP binding cassette (ABC) transporters. The Y976F mutation (TACTTTC) has been associated with CQ resistance in Southeast Asia [22] and Papua New Guinea [23]. Interestingly, the Y976F change is rarely, if ever, observed in alleles that do not carry the F1076L (TTTCTT) change, suggesting a two-step mutation pathway leading to CQ resistance [21]. All SNPs were genotyped, under contract, by K-Biosciences (Cambridge, UK), with an amplifluor assay [47,48]. Primer sequences for amplifying the SNPs are provided in the additional file 9 Table S4; the annealing temperature for all primers was 60°C. Accuracy of genotyping was empirically assessed as 99.8% in blind replicate analyses of 5493 SNPs.

#### Data analysis

Estimation of allele frequencies in malaria parasite populations is complicated by the co-occurrence of multiple clones within infections. Counting all alleles identified within an infection results in overestimation of frequencies of rare alleles and underestimation of common alleles. To minimize bias, we excluded 101 infections in which > 1 allele was observed in any of the SNP loci. For the analysis of chromosome-level SNP diversity and LD patterns, we also excluded 93 infections with allele calls for < 60 markers across chromosome 8. Analysis of the *pvcr1-o* and *pvmdr-1* loci was based on complete genotypes; infections with one or more SNPs without allele calls were excluded. The number of isolates considered for further analysis is shown in the additional file 1 Table S1.

Population-level diversity was measured with the SNP  $\pi$  statistic, defined as the average number of pairwise differences at assayed SNPs between all members of a population [15]. SNP  $\pi$  values were also calculated separately for silent (synonymous or noncoding) and nonsynonymous SNPs, with standard errors estimated by bootstrapping, and compared with nonparametric Wilcoxon tests. To test whether average SNP  $\pi$  values increased with increased distance between dates of collection of sympatric isolates, we used Poptools (version 2.7.1) software [49] to run a Mantel matrix correlation test [50], with 1000 permutations, on the Granada subpopulation dataset (isolates systematically sampled between 2004 and 2006).

A key factor in the success of association studies is the level of LD observed within and across populations. We examined the evidence for LD within each country and in two subpopulations from Brazil (Granada and Plácido de Castro) with enough samples. The LD statistic  $r^2$  [51] was calculated for all pairs of SNPs across chromosome 8, within populations, using LDA software [52]. Statistical

significance of LD was tested, at the 5% level, using  $\chi^2$  tests. Correlation between the physical distance between markers and LD was assessed using the Pearson's coefficient of correlation ( $r$ ). We defined chromosome 8 haplotype blocks as clusters of adjacent markers over which evidence of meiotic recombination was minimal, using the four-gamete rule [53]. For each pair of markers, the frequencies of all four combinations of alleles were computed; blocks were built with consecutive markers where only three or less combinations were observed with a frequency  $\geq 0.01$ . Haplotype blocks were generated with Haploview (version 4.1) software [54]. To compare haplotype block boundaries across different populations, we examined the proportions of pairs of markers assigned to the same block (i. e., non-recombining sites) and to different blocks of no block (i. e., recombining sites) in each population. A SNP pair was considered concordant if the assignment was the same in both populations analyzed and discordant if the assignments disagreed [55]. These comparisons were made for pairs of SNPs spaced up to 31 kb, since this is the length of the largest haplotype block found in our study populations.

We assessed population differentiation using the  $\theta$  estimator [56] of  $F_{ST}$  statistic using FSTAT software [57]; 95% confidence intervals were derived by bootstrapping to determine whether values differed significantly from zero. We used principal component analysis (PCA) to determine whether isolates could be regarded as randomly chosen from a single, genetically homogeneous population or whether they were clustered, defining subpopulations [18]. We carried out separate analysis for the whole dataset and for the populations from Brazil and Asia, using the MeV software [58]. Because strong inter-marker LD may distort PCA results [18], we compared clustering patterns obtained with all informative SNPs (e.g., those that segregated in the population under analysis) with the patterns obtained with a filtered SNP set, in which a single marker was selected from every haplotype block detected by Haploview (defined as above). We also employed STRUCTURE 2.2 software [20] to examine parasite population structure. This software uses a Bayesian clustering approach to assign isolates to  $K$  populations characterized by a set of allele frequencies at each locus. We run the program 10 times each for  $K$  values between 1 and 6. Each analysis involved 100,000 iterations, with 50,000 burn-in cycles. We used the linkage model to account for the LD among markers across the same chromosome. We computed the posterior probability for each  $K$  and show here clustering patterns associated with the strongest statistical support. Again, separate analyses were made for the whole dataset and for Brazilian and Asian samples.

## Ethics statement

The parasitized blood samples described in this paper were collected under protocols approved by the relevant ethical review committees in the respective countries. Patients or their parents or legal guardians provided written informed consent before donating samples. Ethical clearance was also obtained from the Institutional Review Board (IRB) of the University of São Paulo, Brazil; the IRB of the NIAID, USA; the National Research Ethics Committee of the Brazilian Ministry of Health; and the Cambodian National Ethics Committee for Health Research.

## Additional material

**Additional file 1 Table S1.** Origin of field isolates of *P. vivax* analyzed in this study.

**Additional file 2 Figure S1.** Variation in LD in relation to distance between SNP loci along chromosome 8 of *P. vivax* in two Brazilian subpopulations. Data are shown for Granada and Plácido de Castro. LD was measured with the  $r^2$  statistic (values on y-axis are multiplied by 100); closed (blue) circles denote significant LD ( $\chi^2$  statistic,  $P < 0.05$ ), while open circles denote nonsignificant LD. A significant negative correlation between  $r^2$  values and physical distance between SNPs was found for both Granada (Pearson correlation coefficient  $r = -0.180$ ,  $P < 0.0001$ ) and Plácido de Castro ( $r = -0.096$ ,  $P = 0.028$ ).

**Additional file 3 Table S2.** Matrix of  $F_{ST}$  values showing pairwise comparisons for 10 nonsynonymous SNPs (above diagonal) and 75 silent SNPs (below diagonal).

**Additional file 4 Figure S2.** Principal component analysis (PCA) of population structure of *P. vivax*. This analysis considered a subset of 29 SNPs selected to minimize intermarker LD. Plots show the three first principal components. Each circle or square represents a parasite isolate or strain and the color is assigned according to the geographic origin of parasites: green circles, Granada; red circles, Plácido de Castro; open circles, Porto Velho (all samples from Brazil are represented with circles); pink squares, Cambodia; blue squares, Sri Lanka; light brown squares, Vietnam (all samples from Asia are represented with squares). The strains Salvador I and Belém are denoted by orange and black circles, respectively. Upper panel, plot of the first two principal components; lower panel, plot of the first and third principal components. Note that the clustering patterns revealed by the subset of 29 SNPs are quite similar to that detected with 57 informative SNPs (Figure 6).

**Additional file 5 Figure S3.** Principal component analysis (PCA) of population structure of *P. vivax* in Brazil. Plots show the three first principal components. Each circle represents a parasite isolate or strain and the color is assigned according to the geographic origin of parasites: green, Granada; red, Plácido de Castro; open, Porto Velho. The strains Salvador I and Belém are denoted by orange and black colors, respectively. Upper panel, plot of the first two principal components; lower panel, plot of the first and third principal components.

**Additional file 6 Figure S4.** Principal component analysis (PCA) of population structure of *P. vivax* in Asia. Plots show the three first principal components. Each square represents a parasite isolate and the color is assigned according to the geographic origin of parasites: pink, Cambodia; blue, Sri Lanka; light brown, Vietnam. Upper panel, plot of the first two principal components; lower panel, plot of the first and third principal components.

**Additional file 7 Figure S5.** Predicted structure and location of amino acid replacements in PvMDR-1 and PvCRT-O. PvMDR-1 (upper panel) has two hydrophobic domains, each with six transmembrane  $\alpha$ -helices, and a cytosolic domain harboring nucleotide-binding domains 1 (NBD<sub>1</sub>) and NBD<sub>2</sub>, each containing an ATP-binding site. PvCRT-O (lower panel) has 10 predicted transmembrane helices, with C- and N-terminal domains located in the parasite cytoplasm. Black dots represent nonsynonymous SNPs assayed in this study; the only open dot in PvMDR-1 represents a synonymous SNP that was also assayed.

**Additional file 8 Table S3.** Characterization of the 85 biallelic single-nucleotide polymorphisms (SNPs) along *P. vivax* chromosome 8 assayed in this study.

**Additional file 9 Table S4.** Sequences of primers used to amplify 85 single-nucleotide polymorphisms (SNPs) along chromosome 8 of *P. vivax*, 5 SNPs in the *pvcr-t-o* gene and 6 SNPs in the *pvmdr-1* gene.

## Authors' contributions

Conceived and designed the experiments: POS, NDK, DFW, MUF. Performed the experiments: POS. Contributed parasite samples and epidemiologic data: MdSN, NDK, NSdS, KKGS, RMG, CA, JMS, DS, RMF, SG, TT, GG, RA, FK, MUF. Analyzed the data: POS, MUF. Drafted the paper: POS, MUF. All authors read and approved the final version of the manuscript.

## Acknowledgements

This research was supported by funds from the National Institutes of Health (NIH) grants RO1 AI 075416-01 to MUF and 5R03TW007966-02 to NK and DFW, the Intramural Research Program of the National Institute of Allergy and Infectious Diseases (NIAID), NIH, to RMF, the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) grant 470570/2006-7 to MUF, and the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) grants 05/51988-0 and 07/51199-0 to MUF. POS, NSS and MUF receive or received scholarships from CNPq and MdSN, KKGS and RMG receive or received scholarships from FAPESP. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank the patients from all field sites for their enthusiastic cooperation, the field work collaborators Rosely S. Malafrente and Adamilson Luis de Souza for their support in Brazil, Suon Seila and Sreng Sokunthea for their support in Cambodia, and H. Van Thien for his support in Vietnam, Maria José Menezes, Melissa da Silva Bastos and Michelle Cristina Brandi (University of São Paulo, Brazil) for excellent laboratory support, Apuã Paquola for help in computational analyses and Casiano Nunes Pereira for artwork.

## Author Details

<sup>1</sup>Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, 05508-900 São Paulo, São Paulo, Brazil, <sup>2</sup>Department of Parasitology, Faculty of Medicine, University of Colombo, Kinsey Road, Colombo 8, Sri Lanka, <sup>3</sup>Center of Health Sciences, Federal University of Acre, 69915-900 Rio Branco, Acre, Brazil, <sup>4</sup>Institute of Biological Sciences, Federal University of Juiz de Fora, 36036-330 Juiz de Fora, Minas Gerais, Brazil, <sup>5</sup>Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, MD 20852, USA, <sup>6</sup>National Malaria Center, Phnom Penh 1, Cambodia, <sup>7</sup>Anti-Malaria Campaign, Ministry of Health, Colombo 10, Sri Lanka, <sup>8</sup>Department of Social and Environmental Medicine, Institute of Scientific Research, Oita University, Oita 879-5593 Japan and <sup>9</sup>Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA 02115, USA

Received: 20 January 2010 Accepted: 13 July 2010

Published: 13 July 2010

## References

1. Hay SI, Guerra CA, Tatem AJ, Noor AM, Snow RW: **The global distribution and population at risk of malaria: past, present, and future.** *Lancet Infect Dis* 2004, **4**:327-336.
2. Guerra CA, Snow RW, Hay SI: **Mapping the global extent of malaria in 2005.** *Trends Parasitol* 2006, **22**:353-358.
3. Price RN, Douglas NM, Anstey NM: **New developments in *Plasmodium vivax* malaria: severe disease and the rise of chloroquine resistance.** *Curr Opin Infect Dis* 2009, **22**:430-435.
4. Su X, Hayton K, Welles TE: **Genetic linkage and association analyses for trait mapping in *Plasmodium falciparum*.** *Nat Rev Genet* 2007, **8**:497-506.
5. Udomsangpetch R, Kaneko O, Chotivanich K, Sattabongkot J: **Cultivation of *Plasmodium vivax*.** *Trends Parasitol* 2008, **24**:85-88.
6. Cui L, Escalante AA, Imwong M, Snounou G: **The genetic diversity of *Plasmodium vivax* populations.** *Trends Parasitol* 2003, **19**:220-226.
7. Cornejo OE, Rojas A, Udhayakumar V, Lal AA: **Assessing the effect of natural selection in malaria parasites.** *Trends Parasit* 2004, **20**:388-395.

8. Anderson TJC, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, Mollinedo R, Bockarie M, Mokili J, Mharakurwa S, French N, Whitworth J, Velez ID, Brockman AH, Nosten F, Ferreira MU, Day KP: **Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum***. *Mol Biol Evol* 2000, **17**:1467-1482.
9. Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, Crabtree J, Angiuoli SV, Merino EF, Amedeo P, Cheng Q, Coulson RM, Crabb BS, del Portillo HA, Essien K, Feldblyum TV, Fernandez-Becerra C, Gilson PR, Gueye AH, Guo X, Kang'a S, Kooij TW, Korsinczky M, Meyer EV, Nene V, Paulsen I, White O, Ralph SA, Ren Q, Sargeant TJ, Salzberg SL, Stoeckert C, Sullivan SA, Yamamoto MM, Hoffman SL, Wortman JR, Gardner MJ, Galinski MR, Barnwell JW, Fraser-Liggett CM: **Comparative genomics of the neglected human malaria parasite *Plasmodium vivax***. *Nature* 2008, **455**:757-763.
10. Imwong M, Nair S, Pukrittayakamee S, Sudimack D, Williams JT, Mayxay M, Newton PN, Kim JR, Nandy A, Osorio L, Carlton JM, White NJ, Day NPJ, Anderson TJ: **Contrasting genetic structure in *Plasmodium vivax* populations from Asia and South America**. *Int J Parasitol* 2007, **37**:1013-1022.
11. Ferreira MU, Karunaweera ND, da Silva-Nunes M, Silva NS, Wirth DF, Hartl DL: **Population structure and transmission dynamics of *Plasmodium vivax* in rural Amazonia**. *J Infect Dis* 2007, **195**:1218-1226.
12. Karunaweera ND, Ferreira MU, Munasinghe A, Barnwell JW, Collins WE, King CL, Kawamoto F, Hartl DL, Wirth DF: **Extensive microsatellite diversity in the human malaria parasite *Plasmodium vivax***. *Gene* 2008, **410**:105-112.
13. Schlötterer C: **The evolution of molecular markers - just a mater of fashion?** *Nat Rev Genet* 2004, **5**:63-69.
14. Hedrick PW: **Perspective: Highly variable loci and their interpretation in evolution and conservation**. *Evolution* 1999, **53**:313-318.
15. Neafsey DE, Schaffner SF, Volkman SK, Park D, Montgomery P, Milner DA Jr, Lukens A, Rosen D, Daniels R, Houde N, Cortese JF, Tyndall E, Gates C, Stange-Thomann N, Sarr O, Ndiaye D, Ndir O, Mboup S, Ferreira MU, Moraes S do L, Dash AP, Chitnis CE, Wiegand RC, Hartl DL, Birren BW, Lander ES, Sabeti PC, Wirth DF: **Genome-wide SNP genotyping highlights the role of natural selection in *Plasmodium falciparum* population divergence**. *Genome Biol* 2008, **9**:R171.
16. Orjuela-Sánchez P, da Silva NS, da Silva-Nunes M, Ferreira MU: **Parasitemia recurrences and population dynamics of *Plasmodium vivax* polymorphisms in rural Amazonia**. *Am J Trop Med Hyg* 2009, **81**:961-968.
17. Mu J, Awadalla P, Duan J, McGee KM, Joy DA, McVean GA, Su XZ: **Recombination hotspots and population structure in *Plasmodium falciparum***. *PLoS Biol* 2005, **3**:e335.
18. Patterson N, Price AL, Reich D: **Population structure and eigenanalysis**. *PLoS Genet* 2006, **2**:e190.
19. Machado RL, Povoá MM, Calvosa VS, Ferreira MU, Rossit AR, dos Santos EJ, Conway DJ: **Genetic structure of *Plasmodium falciparum* populations in the Brazilian Amazon region**. *J Infect Dis* 2004, **190**:1547-1555.
20. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data**. *Genetics* 2000, **155**:945-959.
21. Brega S, Meslin B, de Monbrison F, Severini C, Gradoni L, Udonsangpetch R, Sutanto I, Peyron F, Picot S: **Identification of the *Plasmodium vivax mdr-like* gene (*pvmdr1*) and analysis of single-nucleotide polymorphisms among isolates from different areas of endemicity**. *J Infect Dis* 2005, **191**:272-277.
22. Suwanarusk R, Russell B, Chavchich M, Chalfein F, Kenangalem E, Kosaisavee V, Prasetyorini B, Piera KA, Barends M, Brockman A, Lek-Uthai U, Anstey NM, Tjitra E, Nosten F, Cheng Q, Price RN: **Chloroquine resistant *Plasmodium vivax*: in vitro characterisation and association with molecular polymorphisms**. *PLoS One* 2007, **2**:e1089.
23. Marfurt J, de Monbrison F, Brega S, Barbollat L, Müller I, Sie A, Goroti M, Reeder JC, Beck HP, Picot S, Genton B: **Molecular markers of in vivo *Plasmodium vivax* resistance to amodiaquine plus sulfadoxine-pyrimethamine: mutations in *pvdhfr* and *pvmdr1***. *J Infect Dis* 2008, **198**:409-417.
24. Sá JM, Nomura T, Neves J, Baird JK, Wellems TE, del Portillo HA: ***Plasmodium vivax*: allele variants of the *mdr1* gene do not associate with chloroquine resistance among isolates from Brazil, Papua, and monkey-adapted strains**. *Exp Parasitol* 2005, **109**:256-259.
25. Duraisingh MT, Refour P: **Multiple drug resistance genes in malaria - from epistasis to epidemiology**. *Mol Microbiol* 2005, **57**:874-877.
26. Suwanarusk R, Chavchich M, Russell B, Jaidee A, Chalfein F, Barends M, Prasetyorini B, Kenangalem E, Piera KA, Lek-Uthai U, Anstey NM, Tjitra E, Nosten F, Cheng Q, Price RN: **Amplification of *pvmdr1* associated with multidrug-resistant *Plasmodium vivax***. *J Infect Dis* 2008, **198**:1558-1564.
27. Anderson TJ, Nair S, Sudimack D, Williams JT, Mayxay M, Newton PN, Guthmann JP, Smithuis FM, Tran TH, van den Broek IV, White NJ, Nosten F: **Geographical distribution of selected and putatively neutral SNPs in Southeast Asian malaria parasites**. *Mol Biol Evol* 2005, **22**:2362-2374.
28. Tian C, Gregersen PK, Seldin MF: **Accounting for ancestry: population substructure and genome-wide association studies**. *Human Mol Genet* 2008, **17**:R143-R150.
29. Camargo LMA, Ferreira MU, Krieger H, Camargo EP, Pereira da Silva L: **Unstable hypoendemic malaria in Rondônia (Western Brazilian Amazon): epidemic outbreaks and work-associated incidence in an agro-industrial rural settlement**. *Am J Trop Med Hyg* 2004, **51**:16-25.
30. da Silva-Nunes M, Codeço CT, Malafrente RS, da Silva NS, Juncansen C, Muniz PT, Ferreira MU: **Malaria on the Amazonian frontier: transmission dynamics, risk factors, spatial distribution, and prospects for control**. *Am J Trop Med Hyg* 2008, **79**:624-635.
31. Orjuela-Sánchez P, da Silva-Nunes M, da Silva NS, Scopel KKG, Gonçalves RM, Malafrente RS, Ferreira MU: **Population dynamics of genetically diverse *Plasmodium falciparum* lineages: community-based prospective study in rural Amazonia**. *Parasitology* 2009, **136**:1097-1105.
32. da Silveira LA, Dorta ML, Kimura EAS, Katzin AM, Kawamoto F, Tanabe K, Ferreira MU: **Allelic diversity and antibody recognition of *Plasmodium falciparum* merozoite surface protein 1 during hypoendemic malaria transmission in the Brazilian Amazon region**. *Infect Immun* 1999, **67**:5906-5916.
33. Ferreira MU, Liu Q, Zhou M, Kimura M, Kaneko O, Van Thien H, Isomura S, Tanabe K, Kawamoto F: **Stable patterns of allelic diversity at the merozoite surface protein-1 locus of *Plasmodium falciparum* in clinical isolates from southern Vietnam**. *J Eukaryot Microbiol* 1998, **45**:131-136.
34. Gunawardena S, Karunaweera ND, Ferreira MU, Phone-Kyaw M, Pollack RJ, Alifrangis M, Rajakaruna RS, Konradsen F, Amerasinghe PH, Schousboe ML, Galappaththy GNL, Abeyasinghe RR, Hartl DL, Wirth DF: **Geographic structure of *Plasmodium vivax*: microsatellite analysis of parasite populations from Sri Lanka, Myanmar and Ethiopia**. *Am J Trop Med Hyg* 2010, **82**:235-242.
35. Fernando SD, Abeyasinghe RR, Galappaththy GNL, Rajapaksa LC: **Absence of asymptomatic malaria infections in previously high endemic areas of Sri Lanka**. *Am J Trop Med Hyg* 2009, **81**:763-767.
36. Dean FB, Hosono S, Fang L, Wu X, Farugi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J, Driscoll M, Song W, Kingsmore SF, Egholm M, Lasken RS: **Comprehensive human genome amplification using multiple displacement amplification**. *Proc Natl Acad Sci USA* 2002, **99**:5261-5266.
37. Tchavtchitch M, Fischer K, Huestis R, Saul A: **The sequence of a 200 kb portion of a *Plasmodium vivax* chromosome reveals a high degree of conservation with *Plasmodium falciparum* chromosome 3**. *Mol Biochem Parasitol* 2001, **118**:211-222.
38. Feng X, Carlton JM, Joy DA, Mu J, Furuya T, Suh BB, Wang Y, Barnwell JW, Su XZ: **Single-nucleotide polymorphisms and genome diversity in *Plasmodium vivax***. *Proc Natl Acad Sci USA* 2003, **100**:8502-8507.
39. del Portillo HA, Longacre S, Khouri E, David PH: **Primary structure of the merozoite surface antigen 1 of *Plasmodium vivax* reveals sequences conserved between different *Plasmodium* species**. *Proc Natl Acad Sci USA* 1991, **88**:4030-4034.
40. Camargo AA, Fischer K, Lanzer M, del Portillo HA: **Construction and characterization of a *Plasmodium vivax* genomic library in yeast artificial chromosomes**. *Genomics* 1997, **42**:467-473.
41. Collins WE, Contacos PG, Krotoski WA, Howard WA: **Transmission of four Central American strains of *Plasmodium vivax* from monkey to man**. *J Parasitol* 1972, **58**:332-335.
42. Sullivan JS, Strobert E, Yang C, Morris CL, Galland GG, Richardson BB, Bounngaseng A, Kendall J, McClure H, Collins WE: **Adaptation of a strain of *Plasmodium vivax* from India to New World monkeys, chimpanzees, and anopheline mosquitoes**. *J Parasitol* 2001, **87**:1398-1403.
43. Arnot DE, Stewart MJ, Barnwell JW: **Antigenic diversity in Thai *Plasmodium vivax* circumsporozoite proteins**. *Mol Biochem Parasitol* 1990, **43**:147-149.
44. Nomura T, Carlton JM, Baird JK, del Portillo HA, Fryauff DJ, Rathore D, Fidock DA, Su X, Collins WE, McCutchan TF, Wootton JC, Wellems TE: **Evidence for different mechanisms of chloroquine resistance in 2 *Plasmodium* species that cause human malaria**. *J Infect Dis* 2001, **183**:1653-1661.

45. Orjuela-Sánchez P, de Santana Filho FS, Machado-Lima A, Chehuan YF, Costa MR, Alecrim MG, del Portillo HA: **Analysis of single-nucleotide polymorphisms in the *crt-o* and *mdr1* genes of *Plasmodium vivax* among chloroquine-resistant isolates from the Brazilian Amazon region.** *Antimicrob Agents Chemother* 2009, **53**:3561-3564.
46. Fidock DA, Nomura T, Talley AK, Cooper RA, Dzekunov SM, Ferdig MT, Ursos LM, Sidhu AB, Naudé B, Deitsch KW, Su XZ, Wootton JC, Roepe PD, Wellem TE: **Mutations in the *P. falciparum* digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance.** *Mol Cell* 2000, **6**:861-871.
47. Nazarenko IA, Bhatnagar SK, Hohman RJ: **A closed tube format for amplification and detection of DNA based on energy transfer.** *Nucleic Acids Res* 1997, **25**:2516-2521.
48. Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, Smith JC, Markham AF: **Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS).** *Nucleic Acids Res* 1989, **17**:2503-2516.
49. **Poptools (version 2.7.1)** [<http://www.cse.csiro.au/poptools>]
50. Mantel N: **The detection of disease clustering and a generalized regression approach.** *Cancer Res* 1967, **27**:209-220.
51. Hill WG, Robertson A: **Linkage disequilibrium in finite populations.** *Theor Appl Genet* 1968, **38**:226-231.
52. Ding K, Zhou K, He F, Shen Y: **LDA - a java based linkage disequilibrium analyzer.** *Bioinformatics* 2003, **19**:2147-2148.
53. Wang N, Akey JM, Zhang K, Chakraborty R, Jin L: **Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation.** *Am J Hum Genet* 2002, **71**:1227-1234.
54. Barrett JC, Fry B, Maller J, Daly MJ: **Haploview: analysis and visualization of LD and haplotype maps.** *Bioinformatics* 2005, **21**:263-265.
55. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D: **The structure of haplotype blocks in the human genome.** *Science* 2002, **296**:2225-2229.
56. Weir BS, Cockerham CC: **Estimating *F*-statistics for the analysis of population structure.** *Evolution* 1984, **38**:1358-1370.
57. **FSTAT version 2.9.3.2** [<http://www2.unil.ch/popgen/softwares/fstat.htm>]
58. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J: **TM4: a free, open-source system for microarray data management and analysis.** *Biotechniques* 2003, **34**:374-378.

doi: 10.1186/1471-2156-11-65

**Cite this article as:** Orjuela-Sánchez *et al.*, Single-nucleotide polymorphism, linkage disequilibrium and geographic structure in the malaria parasite *Plasmodium vivax*: prospects for genome-wide association studies *BMC Genetics* 2010, **11**:65

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

