

Worldwide Genetic Variability of the Duffy Binding Protein: Insights into *Plasmodium vivax* Vaccine Development

Taís Nóbrega de Sousa, Luzia Helena Carvalho, Cristiana Ferreira Alves de Brito*

Laboratory of Malaria, Centro de Pesquisa Rene Rachou/Fiocruz, Belo Horizonte, Brazil

Abstract

The dependence of *Plasmodium vivax* on invasion mediated by Duffy binding protein (DBP) makes this protein a prime candidate for development of a vaccine. However, the development of a DBP-based vaccine might be hampered by the high variability of the protein ligand (DBP_{II}), known to bias the immune response toward a specific DBP variant. Here, the hypothesis being investigated is that the analysis of the worldwide DBP_{II} sequences will allow us to determine the minimum number of haplotypes (MNH) to be included in a DBP-based vaccine of broad coverage. For that, all DBP_{II} sequences available were compiled and MNH was based on the most frequent nonsynonymous single nucleotide polymorphisms, the majority mapped on B and T cell epitopes. A preliminary analysis of DBP_{II} genetic diversity from eight malaria-endemic countries estimated that a number between two to six DBP haplotypes (17 in total) would target at least 50% of parasite population circulating in each endemic region. Aiming to avoid region-specific haplotypes, we next analyzed the MNH that broadly cover worldwide parasite population. The results demonstrated that seven haplotypes would be required to cover around 60% of DBP_{II} sequences available. Trying to validate these selected haplotypes per country, we found that five out of the eight countries will be covered by the MNH (67% of parasite populations, range 48–84%). In addition, to identify related subgroups of DBP_{II} sequences we used a Bayesian clustering algorithm. The algorithm grouped all DBP_{II} sequences in six populations that were independent of geographic origin, with ancestral populations present in different proportions in each country. In conclusion, in this first attempt to undertake a global analysis about DBP_{II} variability, the results suggest that the development of DBP-based vaccine should consider multi-haplotype strategies; otherwise a putative *P. vivax* vaccine may not target some parasite populations.

Citation: Nóbrega de Sousa T, Carvalho LH, Alves de Brito CF (2011) Worldwide Genetic Variability of the Duffy Binding Protein: Insights into *Plasmodium vivax* Vaccine Development. PLoS ONE 6(8): e22944. doi:10.1371/journal.pone.0022944

Editor: Mauricio Martins Rodrigues, Federal University of São Paulo, Brazil

Received: May 27, 2011; **Accepted:** July 8, 2011; **Published:** August 2, 2011

Copyright: © 2011 Nóbrega de Sousa et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the Research Foundation of Minas Gerais State (FAPEMIG), The Brazilian National Research Council (CNPq), and Oswaldo Cruz Foundation (FIOCRUZ, PAPES V) and Malaria Network/Support Program for Centers of Excellence - Pronex (MS/DECIT, CNPq, FAPEMIG, FAPEMAT and FAPERJ). Scholarships from CNPq (TNS, LHC and CFAB) are also acknowledged. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: cristiana@cpqrr.fiocruz.br

Introduction

After more than a century of development of malaria control measures, *Plasmodium vivax* remains more widely distributed than *Plasmodium falciparum* and it is a potential cause of morbidity and mortality. Around 2.85 billion people live at risk of infection by *P. vivax*, with the greatest burden occurring in Middle East, Asia, the Western Pacific, Central and South America [1,2]. Although neglected, *P. vivax* causes important socioeconomic loss, with the overall global cost of *vivax* infection estimated as being between US\$1.4–4.0 billion per year [3]. The recent emergence of drug-resistant strains and severe (sometimes fatal) disease challenges the traditional view of *P. vivax* malaria as a benign infection [4,5]. Consequently, the malaria eradication research agenda (MaERA) placed the *P. vivax* in the top list of priorities [www.ploscollections.org/malERA2011].

The complex life cycle of the *Plasmodium* includes an erythrocytic phase that is responsible for clinical symptoms of human malaria. While *P. falciparum* invades mature as well immature red blood cells (RBC) through multiple invasion pathways, *P. vivax* invades preferentially reticulocytes [6] and requires mainly the interaction of parasite ligand to Duffy antigen/

receptor for chemokines (DARC) on RBC membrane [7]. The *P. vivax* ligand is a 140 kDa micronemal type I membrane protein, called the Duffy binding protein (DBP), and gene-deletion experiment showed that DBP plays an important role in the irreversible junction of the merozoite with host erythrocytes, a key step of human infection [8]. Cysteine-rich region II of the DBP (DBP_{II}) comprises erythrocyte binding motif known as Duffy-binding-like domain (DBL) [9], which is also found in other erythrocyte binding proteins (erythrocyte binding antigen 175 - EBA-175, EBA-140 and EBA-181) and in cytoadherent proteins (*Plasmodium falciparum* erythrocyte membrane protein 1) [10]. The crystal structure of the orthologous DBP ligand domain of the simian malaria *Plasmodium knowlesi* provided insight into the molecular basis for receptor recognition of the PvDBP. The proposed DARC-recognition site of DBP lies in a solvent-accessible groove on a fairly flat surface and exposed site for DARC recognition in subdomain 2 of DBP_{II} [11]. Recently, Bolton and Garry (2011) demonstrated an additional region on subdomain 1 of DBP_{II} that might be also necessary for DARC binding [12].

DBP_{II} is an important anti-*P. vivax* vaccine candidate since antibodies against this molecule: (i) inhibit *in vitro* their binding to

DARC; (ii) reduce merozoite invasion of human erythrocytes; and (iii) confer protection against blood-stage infection [13,14,15,16]. The important role of DBP-DARC interaction is reinforced by previous studies that showed individuals without DARC on their erythrocytes surface are highly resistant to *P. vivax* invasion [17,18]. In addition, studies developed in Brazil and Papua New Guinea showed reduced susceptibility to *P. vivax* infection in heterozygous carriers of one DARC-negative allele compared to two DARC-positive allele carriers [19,20]. However, the paradigm of the absolute dependence on the presence of Duffy on the red cell for *P. vivax* infection has been recently questioned for some findings indicating that *P. vivax* can infect and cause disease in Duffy-negative people [21,22,23]. Nevertheless, this situation seems to occur in specific areas and/or a small proportion of the populations, thus the epidemiological importance of this alternative pathway seem to be restricted to specific endemic areas, such as Madagascar [23].

Analysis of genetic variability from field parasites showed that the DBP binding domain (region II) is highly polymorphic [24,25,26,27,28,29,30,31], therefore it might hamper the vaccine development as some variable residues alter immune recognition of protein [32,33,34]. The excess of non-synonymous substitutions observed in DBP_{II} is consistent with the hypothesis of positive selective pressure acting on this protein domain, and suggests allelic variation as a mechanism of immune evasion [35]. As antigenic variation presents a major limitation in successful vaccine design, we analyzed all DBP_{II} sequences recorded in GenBank in order to identify the main haplotypes shared among *P. vivax* isolates from different malaria-endemic areas and undertake a detailed analysis of the nucleotide diversity of the *dbpII* gene. Here we show the need to include a minimum number of DBP_{II} haplotypes in a DBP-based vaccine; otherwise no broad coverage against worldwide *P. vivax* isolates might be reached.

Materials and Methods

P. vivax isolates

The following 511 Duffy binding protein gene sequences of *P. vivax* deposited in GenBank (last update on 13th April 2011) were downloaded for genetic analyses: the sequence from a reference strain Sal-I (access number: NC_009911); 113 sequences from Papua New Guinea isolates (PNG) (DQ156519; Wosera area in East Sepik Province: AF289480–AF289483, AF289635–AF289653 and AF291096; Madang town and rural villages within Madang Province: AY970837–AY970925, AF469515–AF469602); 17 sequences from Colombia isolates (COL) obtained from the villages of Villavicencio and Tumaco (U50575–U50590, DQ156513); 15 sequences from South Korea isolates (SK) (DQ156515, DQ156522–DQ156523, AF215737–AF215738, AF220657, AF220659–AF220667); 102 sequences from different parts of India (IND) (FJ491142–FJ491241); 30 sequences from Thailand (THAI) (EF219451, EF368159–EF368180, EF379127–EF379132, EF379134); 100 sequences from Sri Lanka (SLK) (GU143914–GU144013); 123 sequences from different parts of Brazil (BRA) (DQ156520, EU812839–EU812960) and eleven sequences from Iran (IRA) (EU860428–EU860438). Only unique haplotypes for each individual were analyzed.

Data Analysis

Genetic diversity analysis. DBP_{II} sequences were aligned and compared using the Clustal W multiple alignment algorithm in BioEdit Sequence Alignment editor [36] to identify the single nucleotide polymorphisms (SNPs). Gaps were removed from alignments because indels (insertions/deletions) and repeats evolve

by different mechanisms than SNPs and might result in false estimates of biologically significant diversity. The number of segregating sites (S), haplotypes (H), nucleotide diversity (π), haplotype diversity (H_d), and the corresponding standard deviations (SD) were estimated using DnaSP 5.10 software [37]. Between-population differentiation using the pairwise fixation index F_{ST} was measured with Arlequin 3.5 software [38]. **Haplotype construction.** Haplotypes (combinations of nucleotides with no particular weight placed upon any position) were constructed by using DnaSP 5.10. We removed synonymous SNPs to focus the analysis only on the protein diversity. **Cluster (population) analysis.** To determine whether our sample could be grouped into genetic clusters and to infer the number of clusters (K) that best fit the data, we used the Bayesian clustering method implemented in the Structure 2.3 software [39,40]. Structure was run 10 times for $K=1-10$ for 30,000 Monte Carlo Markov Chain (MCMC) iterations after burn-in period of 10,000 using the admixture model and correlated allele frequencies. We did not use prior information about population origin for each individual (USEPOPINFO = 0). The mean log probability of the data ($\ln P[D]$) and its standard deviation was plotted to predict the optimal value for K . Graphs of Structure results were produced by using the DISTRUCT program [41].

Results

Polymorphism and genetic differentiation

We compiled 511 sequences from GenBank for the gene fragment encoding Duffy binding protein region II (DBP_{II}). The population dataset included sequences from the natural parasite populations of eight countries (Table 1): Brazil [35], Colombia [25], India and Iran [30], Papua New Guinea [27,42,43], South Korea [28], Sri Lanka [31] and Thailand [44]. The average of sample was 64 sequences (ranged from 11 to 123) per country. By analyzing a region of the DBP_{II} of 676 bp that is available for the overall dataset, 127 polymorphic sites were identified (ranged from 16 to 73 per country) with a nucleotide diversity (π) varying between 0.006 and 0.0109 (South Korea and Thailand, respectively). Most of these SNPs (55%) are singletons, i.e. observed only in one sequence. Despite the wide range of number of haplotypes per country (ranged from 9 to 73, mean of 24), the levels of haplotype diversity (H_d) among them were equally high and quite similar (ranged from 0.922 to 0.993 in Sri Lanka and Colombia, respectively). In order to remove most potential sequencing errors that could interfere with the analysis and interpretation of the results, additional analyses were performed excluding singleton polymorphisms (Table S1). By comparing both analyses a significant bias could be detected in the number of segregating sites and haplotypes in South Korea and PNG samples. However, nucleotide (π) and haplotype diversity (H_d) were not significantly affected by this further analysis because these diversity parameters exclude polymorphisms that are present at low frequencies.

To determine how the observed diversity was distributed among geographic regions, population structure was inferred by measuring genetic differentiation among countries (F_{ST}). The highest differentiation was identified between Colombia and South Korea ($F_{ST} = 0.384$) and the lowest differentiation was detected between Iran and Brazil ($F_{ST} = -0.011$) (Table 1). We repeated the analyses for F_{ST} using the dataset in which singletons were excluded and no significant differences were observed for F_{ST} values among countries (Table S1).

Haplotype diversity

To focus the analysis on the putative antigenic diversity (i.e. polymorphisms that change protein sequence), the nonsynon-

Table 1. Estimates of diversity and genetic differentiation for PvDBP_{II} encoding gene among *P. vivax* isolates.

| Population (N) | S | π (SD) | H | Hd (SD) | F_{ST} | | | | | | | |
|----------------|-----|-----------------|-----|---------------|----------|--------|--------|--------|--------|-------|--------|---|
| | | | | | BRA | COL | PNG | SK | THAI | IRA | SLK | |
| BRA (123) | 21 | 0.0082 (0.0003) | 35 | 0.935 (0.012) | - | | | | | | | |
| COL (17) | 16 | 0.0085 (0.0007) | 16 | 0.993 (0.023) | 0.163* | - | | | | | | |
| PNG (113) | 73 | 0.0106 (0.0004) | 73 | 0.981 (0.005) | 0.117* | 0.196* | - | | | | | |
| SK (15) | 18 | 0.0060 (0.0013) | 10 | 0.924 (0.053) | 0.197* | 0.384* | 0.234* | - | | | | |
| THAI (30) | 29 | 0.0109 (0.0005) | 24 | 0.982 (0.014) | 0.083* | 0.234* | 0.127* | 0.177* | - | | | |
| IRA (11) | 17 | 0.0094 (0.0016) | 9 | 0.964 (0.051) | -0.011 | 0.136* | 0.063* | 0.169* | 0.040 | - | | |
| SLK (100) | 27 | 0.0097 (0.0005) | 39 | 0.922 (0.014) | 0.029* | 0.186* | 0.129* | 0.200* | 0.101* | 0.018 | - | |
| IND (102) | 38 | 0.0088 (0.0005) | 36 | 0.923 (0.016) | 0.011* | 0.183* | 0.120* | 0.200* | 0.077* | 0.004 | 0.018* | - |
| All (511) | 127 | 0.0101 (0.0002) | 193 | 0.970 (0.004) | | | | | | | | |

N: number of isolates; S: number of segregating sites; π : average number of nucleotide substitutions per 1000 sites between pairs of sequences (standard deviation, SD); Hd: haplotype diversity (SD); H: number of haplotypes; F_{ST} : Fixation index, a measure of genetic differentiation between populations;

*: F_{ST} values with $P < 0.05$. BRA – Brazil, COL – Colombia, IRA – Iran, SK – South Korea, SLK – Sri Lanka, THAI – Thailand, IND – India, PNG – Papua new Guinea.
doi:10.1371/journal.pone.0022944.t001

ymous single nucleotide polymorphism (nsSNP) haplotypes were derived for DBP_{II} sequences. In total 46 nsSNPs were identified among all sequences with most of them being rare (19 nsSNPs with frequency $\leq 1\%$ and 17 with frequency between 1–10%). Only 10 nsSNPs showed allele frequency above 10% in the whole world (R308S, K371E, G384D, E385K, K386N, H390R, N417K, L424I, W437R, and I503K). All but one nsSNP were found at the eight countries with DBP genetic diversity data available, except for the SNP R308S that was absent in Colombia and South Korea (Table 2). Moreover we investigated if these polymorphisms were localized in regions previously identified or predicted as T- or B-cell epitopes. All but one nsSNP (K371E) are in regions predicted or experimentally identified as epitope in the DBP_{II} (Table 2).

In further analysis, those 10 most frequent polymorphisms were used to build predominant DBP_{II} haplotypes circulating in malaria-endemic areas. Seventy-three haplotypes were defined

for the whole dataset, ranging from 7 to 29 haplotypes per country (Table S2). In order to determine a minimum number of DBP_{II} haplotypes (MNH) required to be included in anti-*P. vivax* vaccine, we next sought to identify the number of haplotypes per country able to cover at least 50% of local parasite population. In the whole dataset, 17 out of 73 haplotypes fitted this criterion: two haplotypes in South Korea and Sri Lanka; three in Papua New Guinea, Iran and India; four haplotypes in Brazil and Colombia and six haplotypes in Thailand (Figure 1 and Table S2). Together, the 17 haplotypes covers about 70% of worldwide parasite population. Among them, seven were shared between two or more areas, being one haplotype (Hap 23, colored in red) found in high frequency in four countries from two different continents (America and Asia). This result agrees with those from F_{ST} analysis, which estimated low genetic differentiation between DBP_{II} sequences from Brazil and Iran, India or Sri Lanka (Table 1, Figure 1 and

Table 2. Description of the most frequent nonsynonymous (nsSNPs) polymorphisms used to construct DBP_{II} haplotypes, their frequencies in each geographic population and presence in described epitopes.

| nsSNP ^a | Frequency | | | | | | | | | |
|--------------------|-----------|-------|-------|-------|-------|-------|-------|-------|-------|----------------------|
| | BRA | COL | IRA | SK | SLK | THAI | IND | PNG | World | Epitope ^b |
| R308S | 0.073 | 0.000 | 0.182 | 0.000 | 0.130 | 0.267 | 0.098 | 0.690 | 0.235 | H1, 5 |
| K371E | 0.260 | 0.176 | 0.182 | 0.467 | 0.340 | 0.200 | 0.324 | 0.115 | 0.254 | - |
| G384D | 0.187 | 0.412 | 0.455 | 0.467 | 0.060 | 0.233 | 0.137 | 0.345 | 0.211 | H3, 45 |
| E385K | 0.203 | 0.176 | 0.182 | 0.067 | 0.200 | 0.467 | 0.304 | 0.097 | 0.209 | H3, 45, 48 |
| K386N | 0.228 | 0.176 | 0.182 | 0.067 | 0.200 | 0.400 | 0.304 | 0.097 | 0.211 | H3, 45, 48 |
| H390R | 0.504 | 0.941 | 0.636 | 0.533 | 0.340 | 0.433 | 0.353 | 0.504 | 0.456 | H3, 45, 48 |
| N417K | 0.398 | 0.412 | 0.364 | 0.933 | 0.360 | 0.400 | 0.373 | 0.336 | 0.387 | M2, la * |
| L424I | 0.480 | 0.412 | 0.545 | 1.000 | 0.490 | 0.867 | 0.451 | 0.681 | 0.558 | M2, 66, la |
| W437R | 0.488 | 0.118 | 0.455 | 1.000 | 0.370 | 0.633 | 0.373 | 0.327 | 0.417 | M3 * |
| I503K | 0.431 | 0.059 | 0.636 | 1.000 | 0.550 | 0.567 | 0.569 | 0.425 | 0.497 | Id, 1638* |

^a: Amino acid numbers according to SAL-1 sequence [39] with the first letter representing the amino acid present in Sal-1 sequence and the latter representing the polymorphism;

^b: H1, H3, M2 and M3 are B-cell epitopes [51]; 5 – T and B-cell epitope; 45 and 48 are B-cell epitopes; 66 – T-cell epitope [52,53]; la and Id – MHC class I *in silico* predicted promiscuous epitopes [35]; 1638 – MHC class II universal epitope [54];

*- polymorphisms that alter the antigenic character of DBP_{II} [32].

doi:10.1371/journal.pone.0022944.t002

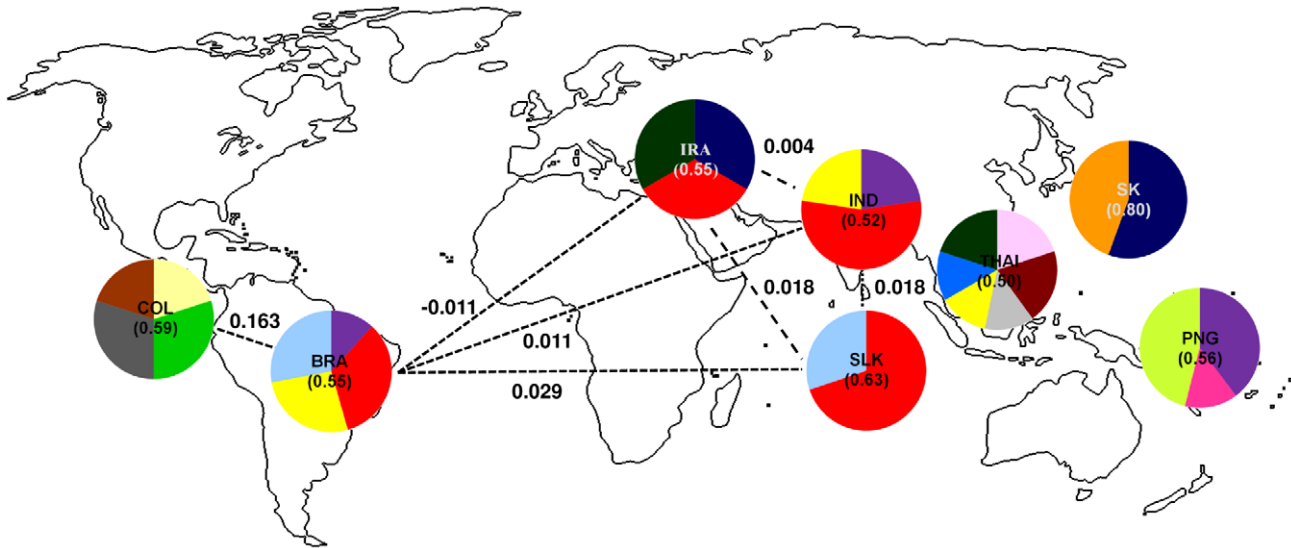


Figure 1. Frequency by country of the DBP_{II} haplotypes that cover at least fifty percent of local parasite population. Among the 73 haplotypes defined using the 10 most prevalent nonsynonymous single nucleotide polymorphisms (nsSNPs), 17 haplotypes (each haplotype was represented by a color) were able to cover at least 50% of each country parasite population: 2 in South Korea (Hap 4 – dark blue and Hap 40 – orange) and in Sri Lanka (Hap 23 – red and Hap 44 – light blue); 3 in Iran (Hap 4, Hap 23 and Hap 59 – dark green), in India (Hap 12 – purple, Hap 23 and Hap 24 – yellow) and in PNG (Hap 12, Hap 55 – pink and Hap 64 – light green); 4 in Colombia (Hap 9 – light yellow, Hap 27 – dark gray, Hap 30 – green and Hap 37 – brown) and 4 in Brazil (Hap 12, Hap 23, Hap 24 and Hap 44); 6 in Thailand (Hap 8 – dark brown, Hap 15 – light pink, Hap 21 – blue, Hap 24, Hap 36 – light gray and Hap 59). The haplotype from Sal-1 reference strain was represented by purple (Hap 12), for haplotype sequences please see Table S2. The number in parentheses indicates the frequency of the selected haplotypes in the respective country (range from 0.52 to 0.63). Some F_{ST} values among countries are showed and represented by dashed lines. doi:10.1371/journal.pone.0022944.g001

Table S2). The reference strain Sal-1 sequence, which has been used to develop a DBP_{II}-based vaccine, covers only 10% of worldwide samples (Hap 12, colored in purple in Figure 1). So far, Sal-1 DBP_{II} was detected only in three endemic areas (Brazil, India and PNG), being in very low frequency in other countries (Figure 1 and Table S2).

Concerning a more global approach for DBP-based vaccine, we further sought to determine the MNH that will be able to cover the majority (~50%) of worldwide parasite population indepen-

dent of the region of origin. Seven haplotypes fitted this criterion and were found in 60% of 511 DBP_{II} sequences of *P. vivax* deposited in GenBank (Haplotypes 4, 12, 23, 24, 44, 59, 64 of Table S2). Considering the distribution of these 7 selected haplotypes by locality, it was possible to categorize those localities in two groups (Figure 2). The first group includes Sri Lanka, Iran, Brazil, India and PNG, where around 50% of the parasites population will be covered by these haplotypes. The second group includes South Korea, Thailand and Colombia, where about only 24% (range 12–33%) of parasite isolates will be covered by these seven selected haplotypes.

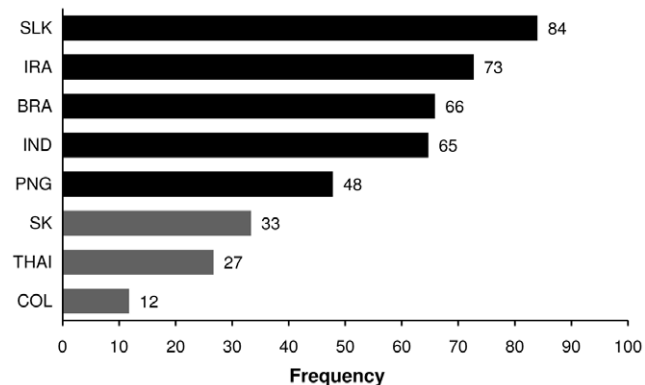


Figure 2. Frequency of the seven nsSNP haplotypes of DBP_{II} that cover at least fifty percent of DBP_{II} sequences deposited in GenBank. Coverage of the selected haplotypes in each country: countries that at least 50% of parasites are covered by the selected haplotypes are represented by black bars and countries that less than 50% of the parasites are covered by the selected haplotypes are represented by grey bars. The selected haplotypes were: 4, 12, 23, 24, 44, 59, 64 (for haplotype sequences see Table S2). doi:10.1371/journal.pone.0022944.g002

Clustering

Aiming to identify related subgroups of DBP_{II} sequences among parasites circulating in the study countries, individual samples were clustered to population on the basis of their genotypes, independent of their geographic location. For that, we used the clustering method that uses departures from Hardy-Weinberg equilibrium to detect population structure. The algorithm groups related individuals into a predefined number of clusters (K), herein $K=1-10$. A Bayesian approach is taken to infer the K value that provides the best fit to the data as measured by the log-likelihood score. Each individual is then assigned a membership coefficient (Q) to each of the clusters with majority of the haplotypes being assigned to only one cluster at “true” K . The estimated log probability of our data [$\ln P(D)$] plateaued between $K=4-6$ (Figure 3A). Simulation studies have shown that once the real K has been reached, $\ln P(D)$ will typically plateau or continue to increase slightly, indicating that $K=6$ provides the best fit to our data. We show clustering results for K values of 2–6, being each individual represented by a vertical line, and each ancestral population in a different color (Figure 3B). To determine whether the above-defined subgroups were geographically restricted we

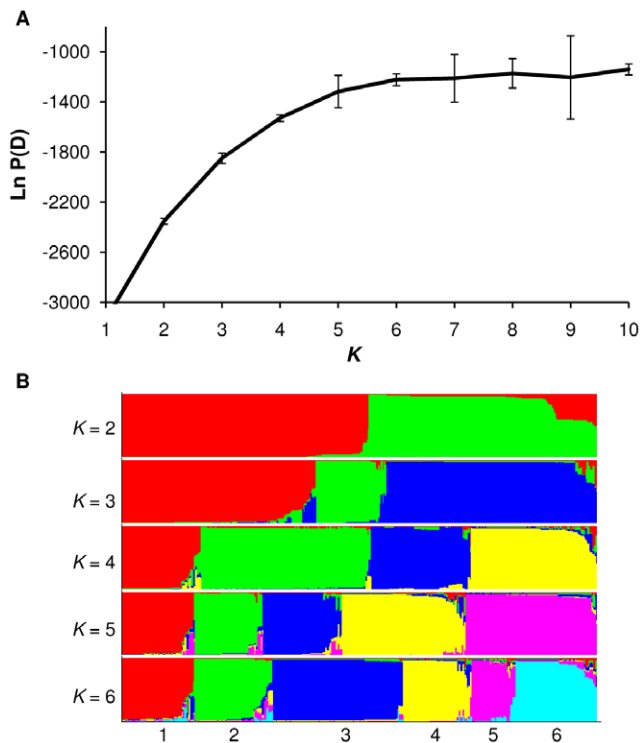


Figure 3. Population structure of the gene encoding DBP_{II} based on Bayesian cluster analysis using Structure 2.3 software. (A) Plot of the log probability of the data [$\ln P(D)$] and its standard deviation (vertical bars) given values for K of 1–10. (B) Population clustering for $K=2$ –6. Each individual is represented by a thin vertical color bar. Each color represents an ancestral population (pop), and the color of individual represents their proportional membership in the different populations. Ancestral populations: red, pop1; green, pop2; blue, pop3; yellow, pop4; pink, pop5 and light blue, pop6. The figure shown for a given K is based on the highest probability run at that K .

doi:10.1371/journal.pone.0022944.g003

plotted the average Q for each country. For all countries but South Korea the analysis supported low levels of differentiation among geographical regions, with three to six ancestral populations present in each country (Figure 4).

Discussion

The major pathway used by *P. vivax* to invade human reticulocytes depends on the interaction of the DBP with its cognate receptor, which makes DBP a high priority anti-*P. vivax* vaccine candidate. A significant challenge for the development of DBP-based vaccines is its highly polymorphic nature, which can bias antibodies response toward a specific DBP variant

[32,33,34]. Engineering vaccines that combine all potential haplotypes that might be circulating in malaria-endemic areas could circumvent polymorphism's limitation; however, it is not feasible for a highly polymorphic antigen like DBP. Therefore, the purpose of this study was to identify predominant DBP_{II} haplotypes that could be included in a putative vaccine, with potential to induce an immune response against *P. vivax* circulating around the world.

By comparing the DBP_{II} diversity found in different countries worldwide, we demonstrated high levels of haplotype diversity among *P. vivax* isolates. The profile of DBP_{II} genetic diversity was not related with the levels of malaria endemicity, since similar pattern was observed from areas with low and unstable malaria transmission, such as Brazil, as well as from highly endemic areas such as Papua New Guinea. These findings suggested that recombination plays an important role in determining the haplotype structure of DBP_{II}, as we recently demonstrated [35]. Genetic recombination of parasites takes place in the vector, as part of the *Plasmodium* life cycle, and is likely facilitated by multiplicity of infections, i.e. the simultaneous infection of a host by more than one parasite variant. For *P. falciparum*, the levels of multiplicity of infection are partially correlated with the levels of transmission intensities [45]. Nevertheless for *P. vivax*, strikingly high values of multiplicity of infection are reported even in regions of low endemicity such as Brazil and Thailand. Thus, the proportion of multiple-clone *P. vivax* infections, estimated by using microsatellites analyses, range from 49–57% in Brazil [46,47], 10–47% in India [48,49], 9–60% in Sri Lanka [50] and 52–63% in Thailand [48,49]. This difference in pattern of multiplicity of infections between *P. vivax* and *P. falciparum* could be related with specific biological features of the *P. vivax* parasite, such as earlier gametocytogenesis [51] and relapse [52]. Early gametocytogenesis might allow for a more efficient transmission to the mosquito vector before symptoms appear and, thus, before drug treatment is initiated, while relapses would also enhance transmission and increase the probability of detecting mixed infections as further inoculations occur.

Besides the role of recombination we have recently demonstrated that natural selection has an important role in the generation and maintenance of the genetic diversity of DBP_{II} [35]. The maintenance of this variability by diversifying selection presumably helps parasite to evade the host immune recognition and favors a low-medium frequency of distinct haplotypes. As we can expect due differences in the endemicity spectrum and immune response profile, both recombination and selection seem to be acting differentially among distinct geographical areas. If no recombination is assumed, we would expect that the number of described polymorphic sites observed would be arranged into a maximum of $n+1$ haplotypes. Here, this profile was found in the majority of studied countries, some of them with the number of haplotypes lower than the number of segregating sites (South Korea and Iran). However, it is important to highlight that the small number of DBP_{II} sequences available for those two countries

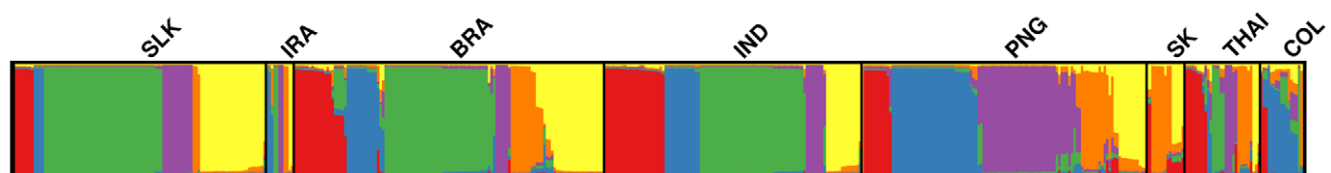


Figure 4. Population structure for *P. vivax* sequences of DBP_{II} at $K=6$. Graphic of Structure results were produced by using the DISTRUCT program. Individual *P. vivax* isolate are represented by a thin vertical line from each geographic population showed above the graphic and separated by a thin black line.

doi:10.1371/journal.pone.0022944.g004

might biased the number of haplotypes observed for those regions; specially, because there was a significant correlation between the number of haplotypes and sample size (Spearman correlation coefficient: $r_s = 0.6628$, $P = 0.0139$). Interestingly, in two countries, Brazil and Sri Lanka, the number of haplotypes was higher than the number of segregating sites, suggesting a major role of recombination in these areas. At this time, it is not possible to conclude if the differences on haplotype number could be consequence of different recombination rates or due to the different levels of selective pressure.

For the purpose of rationalizing vaccine design will be necessary to define polymorphisms that represent antigenically distinct haplotypes. Here, we selected the most frequent nonsynonymous single nucleotide polymorphisms (nsSNPs) to derive DBP_{II} haplotypes. Nine out of 10 nsSNPs lay in regions of DBP_{II} that are immunologically relevant, mapping on previously defined T- and B-cell epitopes [32,35,53,54,55,56]. Based on these nsSNPs, we identified a number between two and six DBP_{II} haplotypes, which will be required to cover 50% of parasite population in each studied country. Unfortunately, most of those haplotypes were geographically restricted and we could not find a single high-frequency haplotype covering all endemic areas. Of note, the DBP_{II} Sal-1 variant that is currently being used to develop a DBP_{II}-based vaccine was found so far in low frequency (10%) and it seems to be restricted to specific geographic areas. Aiming to avoid geographically restrict DBP_{II} haplotypes, in the next analysis we determined the MNH that covers at least fifty percent of DBP_{II} sequences available in GenBank independent of the region of origin. By using this approach, seven frequent haplotypes were identified, that broadly cover parasite populations from 5 out of 8 endemic-countries, i.e., Sri Lanka, Iran, Brazil, India and PNG. These same selected haplotypes were not able to cover parasite populations from South Korea, Colombia and Thailand; however, the results for South Korea and Colombia must be interpreted with caution because of the limited number of sequences available. Together, these results show that the development of DBP-based vaccine should considered multi-haplotype strategies. Similar strategy has been proposed to another polymorphic malaria vaccine candidate, the *P. falciparum* Apical Membrane Antigen 1 (PfAMA-1) [57]. By using a similar approach to that used here, the authors demonstrated that a worldwide collection of PfAMA1 haplotypes could be clustered into six populations that were independent of geographic location. Because continuous culture of *P. falciparum* has been successfully automated for over 30 years [58], the same authors were able to demonstrate that immunization with one member of PfAMA-1 elicited antibodies that block *in vitro* parasite invasion against the same subgroup [57]. Altogether, these two studies provide an important proof of concept that vaccine against malaria polymorphic antigens should consider multi-haplotype strategies. Now it remains to be determined whether those DBP_{II} haplotypes described here could elicit broadly immune response. Although essential, those studies will be a difficult task due to limitation on *P. vivax* culture. Recent progress in short-term culture of *P. vivax* field isolates [59,60] and *in vitro* DBP_{II}-DARC binding assays [15,61] present new opportunities to improve understanding of the ability of DBP_{II} haplotypes to elicit cross-reactive responses against those that are genetically similar. In addition, these assays might help to define which polymorphisms determine antigenically different DBP_{II} haplotypes.

An additional consideration in the malaria vaccines design is the geographic structuring of the parasite populations because significant variation among regions would suggest a need for

vaccines to be tailored accordingly. Overall, we detected low to medium differentiation between countries, except to sequences from Colombia and South Korea, which showed high differentiation compared to the other regions. Remarkably, the four haplotypes that covers more than half Colombian parasites were not shared with any other parasite population, even with the Brazilian, another Latin America parasite population. The contrast among-region structuring was confirmed by F_{ST} analysis with whole data set and clustering analysis using the STRUCTURE software with the ten selected nsSNPs. The Bayesian clustering tool identified six subgroups of related DBP_{II} sequences in the worldwide parasite population. Nevertheless, clustering was not related to the geographical origin of *P. vivax* isolates and all subgroups were present in the eight endemic countries studied, although in different proportions. Moreover, the large majority of isolates (>75%) was formed predominantly by only one ancestral population. Similar pattern was described for *P. falciparum* merozoite antigens in a meta-population analysis [62]. In that study, the merozoite antigens generally had lower levels of differentiation among countries. In addition, cluster analysis suggested that haplotypes formed subgroups independent of geographic origin with high levels of diversity within population. Thus, to improve probability of broad efficacy, vaccines based on merozoite antigens as DBP_{II} may incorporate predominant haplotypes from different regions. Further analysis will be required to define the stability of the haplotypes in a region over time, as natural fluctuations can result from frequency dependent selection or new haplotypes can appear as result of recombination. For DBP, a single study so far has investigated the dynamic of asymptomatic *P. vivax* infections in a cohort of PNG children [63]. By following those children over six-month period, it was possible to demonstrate that the DBP_{II} dominant haplotypes remained relatively stable throughout the study. However, it is difficult to determine the period of time necessary to observe such changes.

The present study comprises the first attempt to undertake a global analysis about DBP_{II} variability and provides new insights on future design of a broad-spectrum *P. vivax* vaccine. Overcome the limitation of DBP diversity may require the inclusion of representative haplotypes otherwise a large proportion of the *P. vivax* population will be not target.

Supporting Information

Table S1 Estimates of genetic diversity and differentiation for PvDBP_{II} encoding gene among *P. vivax* isolates in the absence of singleton sequences.

(DOC)

Table S2 Characterization of DBP_{II} haplotypes present in the eight endemic countries studied.

(DOC)

Acknowledgments

We are grateful to all of the researchers whose publically available sequences made this study possible. We also thank the Program for Technological Development in Tools for Health - PDTIS platform (FIOCRUZ) for DNA sequencing facilities.

Author Contributions

Conceived and designed the experiments: TNdS CFAdB. Performed the experiments: TNdS. Analyzed the data: TNdS LHC CFAdB. Contributed reagents/materials/analysis tools: LHC CFAdB. Wrote the paper: TNdS LHC CFAdB.

References

- Mendis K, Sina BJ, Marchesini P, Carter R (2001) The neglected burden of *Plasmodium vivax* malaria. *Am J Trop Med Hyg* 64: 97–106.
- Guerra CA, Howes RE, Patil AP, Gething PW, Van Boeckel TP, et al. (2010) The international limits and population at risk of *Plasmodium vivax* transmission in 2009. *PLoS Negl Trop Dis* 4: e774.
- Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, et al. (2007) Vivax malaria: Neglected and not benign. *Am J Trop Med Hyg* 77: 79–87.
- Price RN, Douglas NM, Anstey NM (2009) New developments in *Plasmodium vivax* malaria: severe disease and the rise of chloroquine resistance. *Curr Opin Infect Dis* 22: 430–435.
- Alexandre MA, Ferreira CO, Siqueira AM, Magalhães BL, Mourão MP, et al. (2010) Severe *Plasmodium vivax* malaria, Brazilian Amazon. *Emerg Infect Dis* 16: 1611–1614.
- Mons B, Collins WE, Skinner JC, Vanderstar W, Croon J, et al. (1988) *Plasmodium vivax* - *in vitro* growth and reinvasion in red blood-cells of aotusnancymai. *Exp Parasitol* 66: 183–188.
- Horuk R, Chitnis C, Darbonne W, Colby T, Rybicki A, et al. (1993) A receptor for the malarial parasite *Plasmodium vivax*: the erythrocyte chemokine receptor. *Science* 261: 1182–1184.
- Singh AP, Ozwara H, Kocken CHM, Puri SK, Thomas AW, et al. (2005) Targeted deletion of *Plasmodium knowlesi* Duffy binding protein confirms its role in junction formation during invasion. *Mol Microbiol* 55: 1925–1934.
- Adams JH, Sim BKL, Dolan SA, Fang XD, Kaslow DC, et al. (1992) A family of erythrocyte binding-proteins of malaria parasites. *Proc Natl Acad Sci U S A* 89: 7085–7089.
- Michon P, Stevens JR, Kaneko O, Adams JH (2002) Evolutionary relationships of conserved cysteine-rich motifs in adhesive molecules of malaria parasites. *Mol Biol Evol* 19: 1128–1142.
- Singh S, Hora R, Belrhali H, Chitnis C, Sharma A (2006) Structural basis for Duffy recognition by the malaria parasite Duffy-binding-like domain. *Nature* 439: 741–744.
- Bolton MJ, Garry RF (2011) Sequence similarity between the erythrocyte binding domain 1 of the *Plasmodium vivax* Duffy binding protein and the V3 loop of HIV-1 strain MN reveals binding residues for the Duffy Antigen Receptor for Chemokines. *Virology* 418: 45.
- Ceravolo IP, Souza-Silva FA, Fontes CJF, Braga EM, Madureira AP, et al. (2008) Inhibitory properties of the antibody response to *Plasmodium vivax* Duffy binding protein in an area with unstable malaria transmission. *Scand J Immunol* 67: 270–278.
- Grimberg BT, Udumsangpetch R, Xainli J, McHenry A, Panichakul T, et al. (2007) *Plasmodium vivax* invasion of human erythrocytes inhibited by antibodies directed against the Duffy binding protein. *PLoS Med* 4: 1940–1948.
- Michon P, Fraser T, Adams JH (2000) Naturally acquired and vaccine-elicited antibodies block erythrocyte cytoadherence of the *Plasmodium vivax* Duffy binding protein. *Infect Immun* 68: 3164–3171.
- King C, Michon P, Shakri A, Marcotty A, Stanic D, et al. (2008) Naturally acquired Duffy-binding protein-specific binding inhibitory antibodies confer protection from blood-stage *Plasmodium vivax* infection. *Proc Natl Acad Sci U S A* 105: 8363–8368.
- Miller L, Mason S, Clyde D, McGinniss M (1976) The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, FyFy. *N Engl J Med* 295: 302–304.
- Welch SG, McGregor IA, Williams K (1977) The Duffy blood group and malaria prevalence in Gambian West Africans. *Trans R Soc Trop Med Hyg* 71: 295–296.
- Sousa TN, Sanchez BAM, Ceravolo IP, Carvalho LH, Brito CFA (2007) Real-time multiplex allele-specific polymerase chain reaction for genotyping of the Duffy antigen, the *Plasmodium vivax* invasion receptor. *Vox Sang* 92: 373–380.
- Kaschagen L, Mueller I, Kiniboro B, Bockarie M, Reeder J, et al. (2007) Reduced *Plasmodium vivax* erythrocyte infection in PNG Duffy-negative heterozygotes. *PLoS One* 2: e336.
- Ryan JR, Stoute JA, Amon J, Dunton RF, Mtalib R, et al. (2006) Evidence for transmission of *Plasmodium vivax* among a Duffy antigen negative population in western Kenya. *Am J Trop Med Hyg* 75: 575–581.
- Cavasin CE, de Mattos LC, Couto AAD, Bonini-Omingos CR, Valencia SH, et al. (2007) *Plasmodium vivax* infection among Duffy antigen-negative individuals from the Brazilian Amazon region: an exception? *Trans R Soc Trop Med Hyg* 101: 1042–1044.
- Menard D, Barnadas C, Bouchier C, Henry-Hallidin C, Gray LR, et al. (2010) *Plasmodium vivax* clinical malaria is commonly observed in Duffy-negative Malagasy people. *Proc Natl Acad Sci U S A* 107: 5967–5971.
- Sousa TN, Ceravolo IP, Fontes CJF, Couto A, Carvalho LH, et al. (2006) The pattern of major polymorphisms in the Duffy binding protein ligand domain among *Plasmodium vivax* isolates from the Brazilian Amazon area. *Mol Biochem Parasitol* 146: 251–254.
- Ampudia E, Patarroyo MA, Patarroyo ME, Murillo LA (1996) Genetic polymorphism of the Duffy receptor binding domain of *Plasmodium vivax* in Colombian wild isolates. *Mol Biochem Parasitol* 78: 269–272.
- Tsuboi T, Kappe SH, Alyaman F, Prickett MD, Alpers M, et al. (1994) Natural variation within the principal adhesion domain of the *Plasmodium vivax* Duffy binding-protein. *Infect Immun* 62: 5581–5586.
- Xainli J, Adams J, King C (2000) The erythrocyte binding motif of *Plasmodium vivax* duffy binding protein is highly polymorphic and functionally conserved in isolates from Papua New Guinea. *Mol Biochem Parasitol* 111: 253–260.
- Kho W, Chung J, Sim E, Kim D, Chung W (2001) Analysis of polymorphic regions of *Plasmodium vivax* Duffy binding protein of Korean isolates. *Korean J Parasitol* 39: 143–150.
- Gosi P, Fukuda M, Schaecher K, Lanar D, Khushmith S, et al. (2007) Polymorphisms of *Plasmodium vivax* duffy binding protein in isolates from Thai patients. *Am J Trop Med Hyg* 77: 247–247.
- Babaeckho L, Zakeri S, Djadid ND (2009) Genetic mapping of the duffy binding protein (DBP) ligand domain of *Plasmodium vivax* from unstable malaria region in the Middle East. *Am J Trop Med Hyg* 80: 112–118.
- Premaratne PH, Aravinda BR, Escalante AA, Udagama-Randeniya PV (2011) Genetic diversity of *Plasmodium vivax* Duffy Binding Protein II (PvDBPII) under unstable transmission and low intensity malaria in Sri Lanka. *Infect Genet Evol*: Apr 28. [Epub ahead of print].
- VanBuskirk KM, Cole-Tobian JL, Baisor M, Sevova ES, Bockarie M, et al. (2004) Antigenic drift in the ligand domain of *Plasmodium vivax* Duffy binding protein confers resistance to inhibitory antibodies. *J Infect Dis* 190: 1556–1562.
- Ceravolo IP, Sanchez BA, Sousa TN, Guerra BM, Soares IS, et al. (2009) Naturally acquired inhibitory antibodies to *Plasmodium vivax* Duffy binding protein are short-lived and allele-specific following a single malaria infection. *Clin Exp Immunol* 156: 502–510.
- Cole-Tobian JL, Michon P, Baisor M, Richards JS, Beeson JG, et al. (2009) Strain-specific duffy binding protein antibodies correlate with protection against infection with homologous compared to heterologous *Plasmodium vivax* strains in Papua New Guinean children. *Infect Immun* 77: 4009–4017.
- Sousa TN, Tarazona-Santos EM, Wilson DJ, Madureira AP, Falcao PR, et al. (2010) Genetic variability and natural selection at the ligand domain of the Duffy binding protein in Brazilian *Plasmodium vivax* populations. *Malar J* 9: 334.
- Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41: 95–98.
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
- Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol Bioinformatics Online* 1: 47–50.
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567–1587.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Rosenberg N (2004) DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes* 4: 137–138.
- Gosi P, Khushmith S, Khalambaheti T, Lanar D, Schaecher K, et al. (2008) Polymorphism patterns in Duffy-binding protein among Thai *Plasmodium vivax* isolates. *Malar J* 7: 112.
- Cole-Tobian JL, Cortes A, Baisor M, Kastens W, Xainli J, Bockarie M, Adams JH, King CL (2002) Age-acquired immunity to a *Plasmodium vivax* invasion ligand, the duffy binding protein. *J Infect Dis* 186: 531–539.
- Cole-Tobian JL, Baisor M, King CL (2005) High complexity of *Plasmodium vivax* infections in Papua New Guinean children. *Am J Trop Med Hyg* 73: 626–633.
- Anderson T, Haubold B, Williams J, Estrada-Franco J, Richardson L, et al. (2000) Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol Biol Evol* 17: 1467–1482.
- Rezende A, Tarazona-Santos E, Fontes C, Souza J, Couto A, et al. (2010) Microsatellite loci: determining the genetic variability of *Plasmodium vivax*. *Trop Med Int Health* 15: 718–726.
- Ferreira MU, Karunaweera ND, da Silva-Nunes M, da Silva NS, Wirth DF, et al. (2007) Population structure and transmission dynamics of plasmodium vivax in rural amazonia. *J Infect Dis* 195: 1218–1226.
- Imwong M, Nair S, Pukrittayakamee S, Sudimack D, Williams JT, et al. (2007) Contrasting genetic structure in *Plasmodium vivax* populations from Asia and south America. *Int J Parasitol* 37: 1013–1022.
- Imwong M, Snounou G, Pukrittayakamee S, Tanomsing N, Kim JR, et al. (2007) Relapses of *Plasmodium vivax* infection usually result from activation of heterologous hypnozoites. *J Infect Dis* 195: 927–933.
- Karunaweera ND, Ferreira MU, Hard DL, Wirth DF (2007) Fourteen polymorphic microsatellite DNA markers for the human malaria parasite *Plasmodium vivax*. *Mol Ecol Notes* 7: 172–175.
- McKenzie FE, Jeffery GM, Collins WE (2007) Gametocytemia and fever in human malaria infections. *J Parasitol* 93: 627–633.
- Krotoski WA (1989) The hypnozoite and malarial relapse. *Prog Clin Parasitol* 1: 1–19.
- Chootong P, Ntumngia F, Vanbuskirk K, Xainli J, Cole-Tobian J, et al. (2010) Mapping Epitopes of the *Plasmodium vivax* Duffy Binding Protein with Naturally Acquired Inhibitory Antibodies. *Infect Immun* 78: 1089–1095.
- Cole-Tobian J, King CL (2003) Diversity and natural selection in *Plasmodium vivax* Duffy binding protein gene. *Mol Biochem Parasitol* 127: 121–132.
- Xainli J, Baisor M, Kastens W, Bockarie M, Adams J, et al. (2002) Age-dependent cellular immune responses to *Plasmodium vivax* Duffy binding protein in humans. *J Immunol* 169: 3200–3207.

56. Saravia C, Martinez P, Granados D, Lopez C, Reyes C, et al. (2008) Identification and evaluation of universal epitopes in *Plasmodium vivax* Duffy binding protein. *Biochem Biophys Res Commun* 377: 1279–1283.
57. Duan J, Mu J, Thera MA, Joy D, Kosakovsky Pond SL, et al. (2008) Population structure of the genes encoding the polymorphic *Plasmodium falciparum* apical membrane antigen 1: implications for vaccine design. *Proc Natl Acad Sci U S A* 105: 7857–7862.
58. Trager W, Jensen JB (1976) Human malaria parasites in continuous culture. *Science* 193: 673–675.
59. Russell BM, Udomsangpetch R, Rieckmann KH, Kotecka BM, Coleman RE, et al. (2003) Simple in vitro assay for determining the sensitivity of *Plasmodium vivax* isolates from fresh human blood to antimalarials in areas where *P. vivax* is endemic. *Antimicrob Agents Chemother* 47: 170–173.
60. Udomsangpetch R, Somsri S, Panichakul T, Chotivanich K, Sirichaisinthop J, et al. (2007) Short-term in vitro culture of field isolates of *Plasmodium vivax* using umbilical cord blood. *Parasitol Int* 56: 65–69.
61. Hans D, Pattnaik P, Bhattacharyya A, Shakri AR, Yazdani SS, et al. (2005) Mapping binding residues in the *Plasmodium vivax* domain that binds Duffy antigen during red cell invasion. *Mol Microbiol* 55: 1423–1434.
62. Barry AE, Schultz L, Buckee CO, Recder JC (2009) Contrasting population structures of the genes encoding ten leading vaccine-candidate antigens of the human malaria parasite, *Plasmodium falciparum*. *PLoS One* 4: e8497.
63. Cole-Tobian JL, Michon P, Dabod E, Mueller I, King CL (2007) Dynamics of asymptomatic *Plasmodium vivax* infections and Duffy binding protein Polymorphisms in relation to Parasitemia levels in Papua new guinean children. *Am J Trop Med Hyg* 77: 955–962.