



Identifying Potential *Plasmodium vivax* Sporozoite Stage Vaccine Candidates: An Analysis of Genetic Diversity and Natural Selection

Diego Garzón-Ospina^{1,2}, Sindy P. Buitrago¹, Andrea E. Ramos¹ and Manuel A. Patarroyo^{1,3*}

¹ Molecular Biology and Immunology Laboratory, Fundación Instituto de Inmunología de Colombia, Bogotá, Colombia, ² PhD Programme in Biomedical and Biological Sciences, School of Medicine and Health Sciences, Universidad del Rosario, Bogotá, Colombia, ³ Basic Sciences Department, School of Medicine and Health Sciences, Universidad del Rosario, Bogotá, Colombia

OPEN ACCESS

Edited by:

José M. Álvarez-Castro,
Universidade de Santiago de
Compostela, Spain

Reviewed by:

Marcelo R. S. Briones,
Federal University of São Paulo, Brazil
Gonzalo Gajardo,
University of Los Lagos, Chile

*Correspondence:

Manuel A. Patarroyo
mapatarr.fidic@gmail.com

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 24 October 2017

Accepted: 09 January 2018

Published: 25 January 2018

Citation:

Garzón-Ospina D, Buitrago SP,
Ramos AE and Patarroyo MA (2018)
Identifying Potential *Plasmodium vivax*
Sporozoite Stage Vaccine Candidates:
An Analysis of Genetic Diversity and
Natural Selection. *Front. Genet.* 9:10.
doi: 10.3389/fgene.2018.00010

Parasite antigen genetic diversity represents a great obstacle when designing a vaccine against malaria caused by *Plasmodium vivax*. Selecting vaccine candidate antigens has been focused on those fulfilling a role in invasion and which are conserved, thus avoiding specific-allele immune responses. Most antigens described to date belong to the blood stage, thereby blocking parasite development within red blood cells, whilst studying antigens from other stages has been quite restricted. Antigens from different parasite stages are required for developing a completely effective vaccine; thus, pre-erythrocyte stage antigens able to block the first line of infection becoming established should also be taken into account. However, few antigens from this stage have been studied to date. Several *P. falciparum* sporozoite antigens are involved in invasion. Since 77% of genes are orthologous amongst *Plasmodium* parasites, *P. vivax* sporozoite antigen orthologs to those of *P. falciparum* might be present in its genome. Although these genes might have high genetic diversity, conserved functionally-relevant regions (ideal for vaccine development) could be predicted by comparing genetic diversity patterns and evolutionary rates. This study was thus aimed at searching for putative *P. vivax* sporozoite genes so as to analyse their genetic diversity for determining their potential as vaccine candidates. Several DNA sequence polymorphism estimators were computed at each locus. The evolutionary force (drift, selection and recombination) drawing the genetic diversity pattern observed was also determined by using tests based on polymorphism frequency spectrum as well as the type of intra- and inter-species substitutions. Likewise, recombination was assessed both indirectly and directly. The results showed that sporozoite genes were more conserved than merozoite genes evaluated to date. Putative domains implied in cell traversal, gliding motility and hepatocyte interaction had a negative selection signal, being conserved amongst different species in the genus. PvP52, PvP36, PvSPATR, PvPLP1, PvMCP1, PvTLP, PvCelTOS, and PvMB2 antigens or functionally restricted regions within them would thus seem promising vaccine candidates and could be used when designing a pre-erythrocyte and/or multi-stage vaccine against *P. vivax* to avoid allele-specific immune responses that could reduce vaccine efficacy.

Keywords: *Plasmodium vivax*, genetic diversity, sporozoite, vaccine, natural selection, hepatocyte invasion

INTRODUCTION

Plasmodium vivax (Pv) is one of the five *Plasmodium* species causing malaria in human beings [Coatney and National Institute of Allergy and Infectious Diseases (U.S.), 1971; Rich and Ayala, 2003]. Outdoor biting of less-anthropophilic mosquitos (than the main *Plasmodium falciparum* vectors) transmitting it, and endemic regions' social-economic conditions make *P. vivax* an emergent public health problem (Mueller et al., 2015). This parasite exclusively invades reticulocytes and is characterized by relapses from dormant liver stages; it produces early and continuous gametocytes (Price et al., 2009; Patarroyo et al., 2012; Adams and Mueller, 2017) and has great genetic diversity throughout its genome (Neafsey et al., 2012; Winter et al., 2015). All these features make *P. vivax* control and elimination a great challenge.

Vaccine development has been considered as one of the most cost-effective interventions for controlling malaria. Designing a vaccine against this disease has focused on selecting antigens able to induce an effective immune response that block invasion of target cells. The *Plasmodium* life-cycle should be considered when designing an anti-malarial vaccine. Malarial infection begins with an infected female mosquito's bite. *Plasmodium* sporozoites (Spz or pre-erythrocyte stage) in the vertebrate bloodstream must migrate to the host's liver, traversing the endothelial and Kupffer cells that form the sinusoidal barrier. They then migrate through some hepatocytes before infecting one of them (Menard, 2001; Frevert, 2004). Inside hepatocytes, Spz differentiate into thousands of merozoites (Mrz) which after their release proceed to invade red blood cells (RBC), initiating the erythrocyte or blood stage. Within RBC the Mrz could differentiate in new Mrz which will infect new RBC or into gametocytes which can be taken by the mosquito vector to start the sexual stage.

As mentioned above, proteins involved in parasite-host cell interactions are the main targets for vaccine development. However, the genetic diversity found in the parasite has become

a challenge for designing a fully-effective vaccine (Patarroyo et al., 2012; Barry and Arnott, 2014). Such polymorphisms are typically found within functionally irrelevant gene/protein regions enabling the evasion of host immune responses, whilst functionally important regions remain conserved due to functional/structural constraints (Garzón-Ospina et al., 2012; Baquero et al., 2017); these regions could therefore be taken into account for vaccine design in order to avoid allele-specific immune responses. Several studies have measured potential vaccine candidates' genetic diversity (Putaporn et al., 1997, 2009, 2010; Gomez et al., 2006; Garzón-Ospina et al., 2010, 2011, 2012, 2014; Dias et al., 2011; Premaratne et al., 2011; Chenet et al., 2013; Barry and Arnott, 2014; Forero-Rodriguez et al., 2014a,b; Buitrago et al., 2016; Chaurio et al., 2016; Mehrizi et al., 2017). Likewise, the evolutionary forces (mutation, natural selection, genetic drift, recombination, and migration) modulating polymorphism (Casillas and Barbadilla, 2017) have also been determined. This has been used for monitoring anti-malarial vaccine targets (Barry and Arnott, 2014) but might also be used for predicting functional regions which are usually conserved amongst species (Kimura, 1983; Graur et al., 2013). Accordingly, promising vaccine candidates must likely be parasite proteins playing an important role during target cell invasion but displaying limited genetic diversity or, at least, a domain having such pattern. These genes or domains must thus have a negative selection signal ($\omega < 1$ evolutionary rate). Moreover, vaccine candidates should be able to induce an immune response in natural or experimental infection (Patarroyo et al., 2012; Barry and Arnott, 2014; Weiss et al., 2015).

Despite there being three intervention points [pre-erythrocyte, blood and gametocyte stages (Barry and Arnott, 2014)], potential *P. vivax* candidates characterized to date have mainly been described for the blood stage to avoid Mrz entry to RBC, preventing the disease's typical symptomatology (Patarroyo et al., 2012). On the contrary, few pre-erythrocyte phase antigens have been studied, in spite of the fact that blocking Spz interaction with hepatocytes would greatly reduce the possibility of developing the disease. This strategy would also avoid *P. vivax* dormant form formation in the liver and hence infected patients' constant relapses (Price et al., 2009; Hulden, 2011). Several Spz antigens involved in invasion have been characterized in *P. falciparum* and other parasite species (Sultan et al., 1997; Wengelnik et al., 1999; Menard, 2001; Matuschewski et al., 2002; Romero et al., 2004; Kariu et al., 2006; Labaied et al., 2007a,b; Moreira et al., 2008; Rosado et al., 2008; Engelmann et al., 2009; Alba et al., 2011; Curtidor et al., 2011; Aldrich et al., 2012; Annoura et al., 2014; Ferguson et al., 2014; Jimah et al., 2016; Kublin et al., 2017; Manzoni et al., 2017; Yang et al., 2017a). Since around 77% of *Plasmodium* genes are orthologous (Carlton et al., 2008), many being essential for parasite survival (Bushell et al., 2017), orthologous Spz antigens might also be present in the *P. vivax* genome. Given the *P. vivax* genome's high genetic diversity (Neafsey et al., 2012), *P. vivax* Spz genes might also be highly polymorphic; however, as functionally important regions tend to have low polymorphism (Kimura, 1983; Graur et al., 2013), *P. vivax* and related *Plasmodium* parasites' highly conserved regions could

Abbreviations: BepiPred, B-cell epitope prediction; CBI, codon bias index; ENC, effective number of codons; FEL, fixed effects likelihood; FUBAR, fast, unconstrained Bayesian approximation for inferring selection; GARD, genetic algorithm recombination detection; GPI, glycosylphosphatidylinositol membrane anchor; GTP, guanosine triphosphate; IEDB, immune epitope database and analysis resource; IFEL, internal fixed effects likelihood; MEME, mixed effects model of evolution; Mrz, merozoite; *pvspect1*, gene encoding *Plasmodium vivax* sporozoite protein essential for cell traversal; *pvsiap1*, gene encoding *Plasmodium vivax* sporozoite invasion-associated protein-1; *pvp52*, gene encoding *Plasmodium vivax* 6-cysteine protein P52; *pvp36*, gene encoding *Plasmodium vivax* 6-cysteine protein P36; *pvspatr*, gene encoding *Plasmodium vivax* secreted protein with altered thrombospondin repeat domain; *pvtrsp*, gene encoding *Plasmodium vivax* thrombospondin-related sporozoite protein; *pvttrap*, gene encoding *Plasmodium vivax* thrombospondin-related anonymous protein; *pvsiap2*, gene encoding *Plasmodium vivax* sporozoite invasion-associated protein-2; *pymaabl*, gene encoding *Plasmodium vivax* merozoite adhesive erythrocytic binding protein; *pvp1p1*, gene encoding *Plasmodium vivax* perforin-like protein 1; *pvmcp1*, gene encoding *Plasmodium vivax* merozoite capping protein 1; *pvtlp*, gene encoding *Plasmodium vivax* thioredoxin-like protein; *pvceltos*, gene encoding *Plasmodium vivax* cell traversal protein for ookinetes and sporozoites; *pymb2*, gene encoding *Plasmodium vivax* MB2 protein; REL, random effects likelihood; SLAC, single likelihood ancestor counting; SNP, single nucleotide polymorphisms; Spz, sporozoite.

be functionally important and should therefore be taken into account for vaccine development (Patarroyo et al., 2012). As few *P. vivax* pre-erythrocyte stage antigens have been described to date (Menard, 2001; Castellanos et al., 2007; Barry and Arnott, 2014), and because characterizing antigens from this stage is laborious, this study involved a search for orthologous genes in *P. vivax* to those previously studied in *P. falciparum* for *in silico* characterization so as to assess their genetic diversity. The evolutionary forces modulating the variation pattern observed were analyzed for identifying conserved and functional regions; then proteins' antigenic potential was predicted for identifying those potentially able to induce an immune response and thus determine which of these antigens should be prioritized and taken into account when designing a pre-erythrocyte and/or multi-stage vaccine against malaria caused by *P. vivax*.

METHODOLOGY

Sequences Data Set and *in Silico* Characterization of *P. vivax* Loci

Few genes encoding putative *Plasmodium vivax* Spz stage vaccine antigens have been assessed regarding their genetic diversity. Fifteen *P. falciparum* proteins have been suggested as promising vaccine candidates (Curtidor et al., 2011). Orthologous sequences to these *P. falciparum* vaccine candidates were sought in the *P. vivax* Salvador I (Sal-I) strain from the PlasmoDB database (Release 32). One hundred and seventy-one *P. vivax* natural strain DNA sequences from regions worldwide, analyzed by whole genome sequencing (Chan et al., 2012; Neafsey et al., 2012; Hester et al., 2013; Hupalo et al., 2016) available in the PlasmoDB database, were also obtained for these putative antigens. Sal-I sequences from these genes were then used as query for searching orthologous sequences in closely-related species [*Plasmodium cynomolgi* (GCA_000321355.1), *Plasmodium inui* (GCA_000524495.1), *Plasmodium fragile* (GCA_000956335.1), *Plasmodium knowlesi* (GCA_000006355.1) and *Plasmodium coatneyi* (GCA_000725905.1)], using the whole genome data available in GenBank. Orthologous sequences from less related *Plasmodium* species (*Plasmodium berghei*, *Plasmodium yoelii*, *Plasmodium chabaudi*, *Plasmodium vinckei*, *Plasmodium falciparum*, *Plasmodium reichenowi*, *Plasmodium gaboni*, and *Plasmodium gallinaceum*) were obtained from the PlasmoDB database (Release 32).

Potential vaccine candidates described in Mrz are characterized by having a signal peptide and some have membrane anchoring structures [i.e., transmembrane helices and/or glycosylphosphatidylinositol (GPI) anchor], whilst some others have binding and/or protein-protein interaction domains (Patarroyo et al., 2012), therefore, the Sal-I Spz protein sequences were used for *in silico* characterization using several bioinformatics tools. SignalP (Nielsen, 2017) and Phobius (Kall et al., 2007) predictors were used for ascertaining signal peptide presence; the BaCelLo algorithm (Pierleoni et al., 2006) was used for predicting antigen location. Transmembrane and/or GPI domains were evaluated with Phobius, TMHMM (Sonnhammer

et al., 1998) and GPI-SOM (Fankhauser and Maser, 2005) and the Pfam database was searched for putative domains.

Alignment and Sequence Analysis

The 171 sequences from different locations worldwide obtained for each gene were screened to rule out gene sequences having missing data or ambiguous nucleotides; introns were removed from those genes having them. Multiple DNA alignment for each gene was performed based on amino acid (aa) information using the TranslatorX (Abascal et al., 2010) server with the Muscle algorithm (Edgar, 2004).

DnaSP v5 software (Librado and Rozas, 2009) was then used for calculating several genetic diversity estimators for 14 Spz loci. On the other hand, the effective number of codons (ENC) and codon bias index (CBI) were obtained as a measure of selective pressure at translational level (Novembre, 2002). Tajima (1989), Fu and Li (1993), and Fay and Wu (2000) tests were used for evaluating a neutral model of molecular evolution. Repeat regions or those having insertions/deletions were not taken into account for analysis. Tests based on the type of nucleotide substitution were used to infer natural selection signals within genes. The Nei-Gojobori modified method (Zhang et al., 1998) was used for calculating the difference between non-synonymous and synonymous substitution rates at intra-species level (d_N-d_S). The difference between non-synonymous and synonymous divergences rates (K_N-K_S) was calculated by modified Nei-Gojobori method with Jukes-Cantor correction (Jukes and Cantor, 1969), using *P. vivax* sequences together with orthologous sequences from phylogenetically-closely related species for determining natural selection signals at inter-species level. MEGA v6 software (Tamura et al., 2013) was used for all such analysis. A sliding window for omega rates ($\omega = d_N/d_S$ and/or K_N/K_S) was used for evaluating the effect of natural selection throughout the gene. Likewise, individual sites (codons) under selection were identified by calculating synonymous and non-synonymous substitution rates per codon using SLAC, FEL, REL (Kosakovsky Pond and Frost, 2005), MEME (Murrell et al., 2012) and FUBAR methods (Murrell et al., 2013) in the Datamonkey online server (Delport et al., 2010). The McDonald-Kreitman (MK) test (McDonald and Kreitman, 1991) with Jukes-Cantor correction was also used for evaluating neutrality deviations by using the <http://mkt.uab.es/mkt/MKT.asp> web server (Egea et al., 2008).

Lineage-Specific Positive Selection

Lineages under episodic diversifying selection were assessed for each gene using the random effects likelihood (REL)-branch-site method (Kosakovsky Pond et al., 2011). Orthologous sequences from 13 *Plasmodium* species were aligned using the MUSCLE algorithm; this was then used for inferring the best evolutionary model using JModelTest (Posada, 2008). Phylogeny was then inferred using the Bayesian method (Ronquist et al., 2012) with a corresponding evolutionary model; these were used as reference when analyzing lineage-specific positive selection using the HyPhy package (Pond et al., 2005). CIPRES Science Gateway (Miller et al., 2010) web application was used for choosing the evolutionary model and Bayesian analysis.

Linkage Disequilibrium and Recombination

Linkage disequilibrium (LD) was assessed by using the Z_{ns} estimator (Kelly, 1997), followed by linear regression between LD and nucleotide distance to ascertain whether intra-gene recombination could have taken place regarding any gene. Recombination was also evaluated using the ZZ estimator (Rozas et al., 2001), the minimum number of recombination events (Rm) (Hudson and Kaplan, 1985) and the GARD algorithm (Kosakovsky Pond et al., 2006).

Predicting Proteins' Antigenic Potential

A previous study has shown a correlation between predicting potential B-cell epitopes and antigenic protein regions in a *P. vivax* antigen (Rodrigues-da-Silva et al., 2017). Potential linear B-cell epitopes were therefore predicted by using the immune epitope database (IEDB) server (Kolaskar and Tongaonkar, 1990) for each protein, using Sal-I reference sequences as query.

RESULTS

P. vivax Spz Loci *in Silico* Characterization

The currently available information for the Sal-I strain published in PlasmoDB database was used as the source for obtaining sequences from Spz genes orthologous to those identified as vaccine candidates in *P. falciparum* (Table 1). An ortholog search in other *Plasmodium* species showed that all genes (except for *siap2*, which was only present in species infecting primates) had an ortholog in the *Plasmodium* species analyzed here. The Sal-I protein sequence for each gene was then inferred from searching protein features within them (Figure 1). A positive secretion signal sequence was predicted for all antigens, except PvP36 (6-Cys protein family member), PvSIAP2 (sporozoite invasion-associated protein 2) and PvMCP1 (merozoite capping protein) proteins. In addition to signal peptide, a post-translational modification (consisting of a C-terminal GPI anchor sequence) was predicted for PvP52 (6-Cys protein family member) and PvTRSP (thrombospondin-related sporozoite protein), whilst several proteins seemed to have a transmembrane helix (Figure 1 and Table 1); transmembrane helices were found at the N-terminal end in PvP36, PvPLP1 (perforin-like protein 1) and PvSIAP2. A signal peptide and transmembrane helix were predicted at the same position in PvPLP1. The transmembrane helices in PvP36, PvPLP1 and PvSIAP2 could thus have resulted from misidentification and they could actually have been signal peptide sequences.

PvTRSP, PvTRAP (thrombospondin-related anonymous protein) and PvTLP (thioredoxin-like protein) proteins, having a transmembrane region, also had a thrombospondin type 1 putative domain, whilst PvTRAP and PvTLP had a von Willebrand factor type A putative domain. A membrane attack complex/perforin (MACPF) putative domain was found in PvPLP1 and the AhpC/TSA domain in PvMCP1. A single sexual stage antigen s48/45 domain was predicted in PvP52 and PvP36. Both elongation factor Tu GTP (guanosine triphosphate)-binding domain and translation-initiation factor 2 were predicted for PvMB2 (Figure 1 and Table 1).

Genetic Diversity at *P. vivax*-Sporozoite Loci

Sequences from different regions worldwide were analyzed for quantifying Spz antigen genetic diversity (Table 2); these parasite antigens have limited genetic diversity ($\pi < 0.003$). According to the diversity parameters estimated here, *pvtrap*, *pvsia2* and *pvceltos* (cell traversal protein for ookinetes and sporozoites) genes had the highest nucleotide and protein diversity values ($\pi > 0.0015$; $\rho > 0.0039$); *pvspect1* (sporozoite protein essential for cell traversal), *pvplp1*, *pvspatr* (secreted protein with altered thrombospondin repeat domain), *pvtlp*, *pvsia1*, and *pvmaebl* (merozoite adhesive erythrocytic binding protein) genes formed part of the most conserved genes/proteins ($\pi < 0.0009$; $\rho < 0.0020$) (Table 2). The haplotype number was low in most loci. However, *pvtrap*, *pvsia2*, *pvplp1*, *pvmcp1*, *pvtlp*, and *pymb2* were the Spz loci with the highest haplotype number.

Assessing Neutral Evolution in *P. vivax*-Sporozoite Loci

The ENC and CBI parameters were evaluated to assess whether codon bias had taken place in Spz genes. These parameters gave values higher than 47 for ENC and lower than 0.47 for CBI (Table 3). Tests based on polymorphism frequency spectrum were used with the 14 Spz *P. vivax* genes to assess any departure from neutral expectations. Six of these 14 genes had an overall negative statistically significant value for at least one test (Table 3), suggesting that natural selection could have been acting regarding these genes; *pvp36* and *pymb2* genes had specific regions within each gene having a statistical significant negative value (Supplementary Material 1).

Regarding d_N-d_S rates, negative values were found for *pvmaebl*, *pvplp1*, and *pymb2* whilst positive values was observed for the *pvsia2* gene (Table 3). When species divergence was assessed, there was evidence of inter-specific negative selection for all genes, except *trsp*, *siap2*, and *celtos*; the sliding window for ω rate (Supplementary Material 2) gave values lower than 1 for these genes. Likewise, codon-based methods identified several codons under negative selection with few codons under positive selection. Many codons under negative selection were located in regions encoding putative domains (e.g., thrombospondin type 1 domain, Supplementary Material 2). It was found that *siap1* and *trap* loci had statistically significant values higher than 1 when polymorphism and divergence were compared by MK test, whilst a value lower than 1 was observed for *maebl*.

Lineage-Specific Positive Selection

Plasmodium species orthologous sequences were used for constructing phylogenies to assess whether positive natural selection had taken place during antigen evolutionary history (Supplementary Material 3). Topologies gave three monophyletic clusters (*siap2* did not, since it is only present in primate-infecting parasites); the first involved monkey-malaria parasites, the second clustered *Plasmodium* species infected rodents and the third was formed by hominid-malaria parasites. The latter cluster and the rodent clade represent *Plasmodium* phylogenetic relationships. However, the monkey-malaria

TABLE 1 | *In silico* characterization of 14 *P. vivax* sporozoite proteins.

Gene	Signal peptide (position)	Transm helix (position)	GPI	Domain (position)	Location
<i>siap1</i>	Yes (1–22)	No	No	–	Secretory
<i>p52</i>	Yes (1–18)	No	C-ter	Sexual stage antigen s48/45 domain (159–281)	Secretory/membrane
<i>p36</i>	No	Yes (10–32 and 37–59)	No	Sexual stage antigen s48/45 domain (204–332)	Secretory/membrane
<i>spatr</i>	Yes (1–20)	No	No	–	Secretory
<i>trsp</i>	Yes (1–18)	Yes (131–154)	C-ter	Thrombospondin type 1 domain (59–105)	Secretory/membrane
<i>trap</i>	Yes (1–24)	Yes (494–515)	No	von Willebrand factor type A domain (44–225) and Thrombospondin type 1 domain (241–284)	Secretory/membrane
<i>spect</i>	Yes (1–19)	No	No	–	Secretory
<i>siap2</i>	No	Yes (7–28)	No	–	Secretory
<i>maebl</i>	Yes (1–19)	Yes (1,799–1,816)	No	–	Secretory/membrane
<i>plp1</i>	Yes (1–23)	Yes (7–26)	No	Membrane attack complex/perforin (MACPF) (232–581)	Secretory
<i>mcp1</i>	No	No	No	AhpC/TSA family (9–134)	Non cytoplasmic
<i>tlp</i>	Yes (1–23)	Yes (1,426–1,445)	No	Thrombospondin type 1 domain (264–311) and von Willebrand factor type A domain (334–508)	Secretory/membrane
<i>celtos</i>	Yes (1–35)	No	No	–	Non cytoplasmic
<i>mb2</i>	Yes (1–22)	No	No	Elongation factor Tu GTP binding domain (766–931) and Translation-initiation factor 2 (1,151–1,263)	Secretory

The key structures for a potential vaccine candidate (such as a signal peptide, a transmembrane helix or glycosylphosphatidylinositol (GPI) anchor and functional domains) were described from the *P. vivax* Sal-I strain aa sequence. A protein's type of cell location is also indicated. The numbers in brackets indicate the position where a determined structure was predicted.

parasite topology had different branch patterns regarding malaria-species relationships which could have resulted from positive selection (Sawai et al., 2010). Nine of these topologies had lineages (branches) where some sites were under positive selection. Most branches having evidence of positive selection lead to a particular species, whilst few of them were ancestral lineages (Supplementary Material 3).

Linkage Disequilibrium and Recombination

Linkage disequilibrium (LD) at intra-gene level was assessed by Zns estimator (Table 4). Non-random associations between SNPs were found for *pvmaebl* but not for the remaining genes. However, there was a decreasing linear regression tendency between LD and nucleotide distance in all genes, suggesting recombination action. Likewise, most of these genes had at least one recombination event (RM), while the ZZ estimator just gave statistically significant values for *pvp52* and *pvspatr*. The GARD algorithm gave one recombination breakpoint for *pv52*, *pvtrap*, *pvplp1*, *pvtlp*, and *pvceltos* and two recombination breakpoints for *pvmmb2* (Table 4).

P. vivax-Spz Proteins' Antigenic Potential

Previous studies have shown a B-epitope and solvent accessibility prediction correlation with antigenic regions in natural infections (Rodrigues-da-Silva et al., 2017). The Sal-I sequences for each Spz protein studied here were thus analyzed by the BepiPred server for determining their antigenic potential (Supplementary Material 4). Large-sized proteins, such as MAEBL, TLP, TRAP, P52, and MCP1, had regions toward the C-terminal which could be recognized as linear B-epitopes; these aa regions were

the regions having greater solvent accessibility (Supplementary Material 4). It was seen that the most exposed PLP1 region was located toward the N-terminal. Smaller proteins, such as SPECT1, SPART, and SIAP1/2, did not seem to have clearly-defined recognition regions all along their sequences, whilst 3 regions having antigen potential were observed for CelTOS (Supplementary Material 4). Many P36, TRSP, PLP1, MB2 antigen sequences seemed to be exposed and several regions having potential B-epitopes were found.

DISCUSSION

A prospective *Plasmodium vivax*-malarial vaccine has been delayed regarding *P. falciparum*; however, knowledge acquired concerning the latter species could be useful for designing a fully-effective anti-*P. vivax* vaccine. Vaccine development involves several challenges; for instance, high antigen diversity has made vaccines not fully-protective since polymorphism provokes allele-specific immune responses; genetic diversity is therefore an immune avoidance mechanism. Consequently, conserved antigens (or regions within them) should be used as vaccine candidates to avoid this kind of response (Richie and Saul, 2002; Patarroyo et al., 2012). Even more, just one antigen might not be enough to produce full protection regarding a particular vaccine, so several antigens would be necessary. Since malaria parasites have multiple stages, a fully-effective vaccine must have several conserved antigens (or regions containing them) from different parasite stages.

Antigen identification is not an easy task due to *P. vivax* having a complex biology. Most antigens described to date

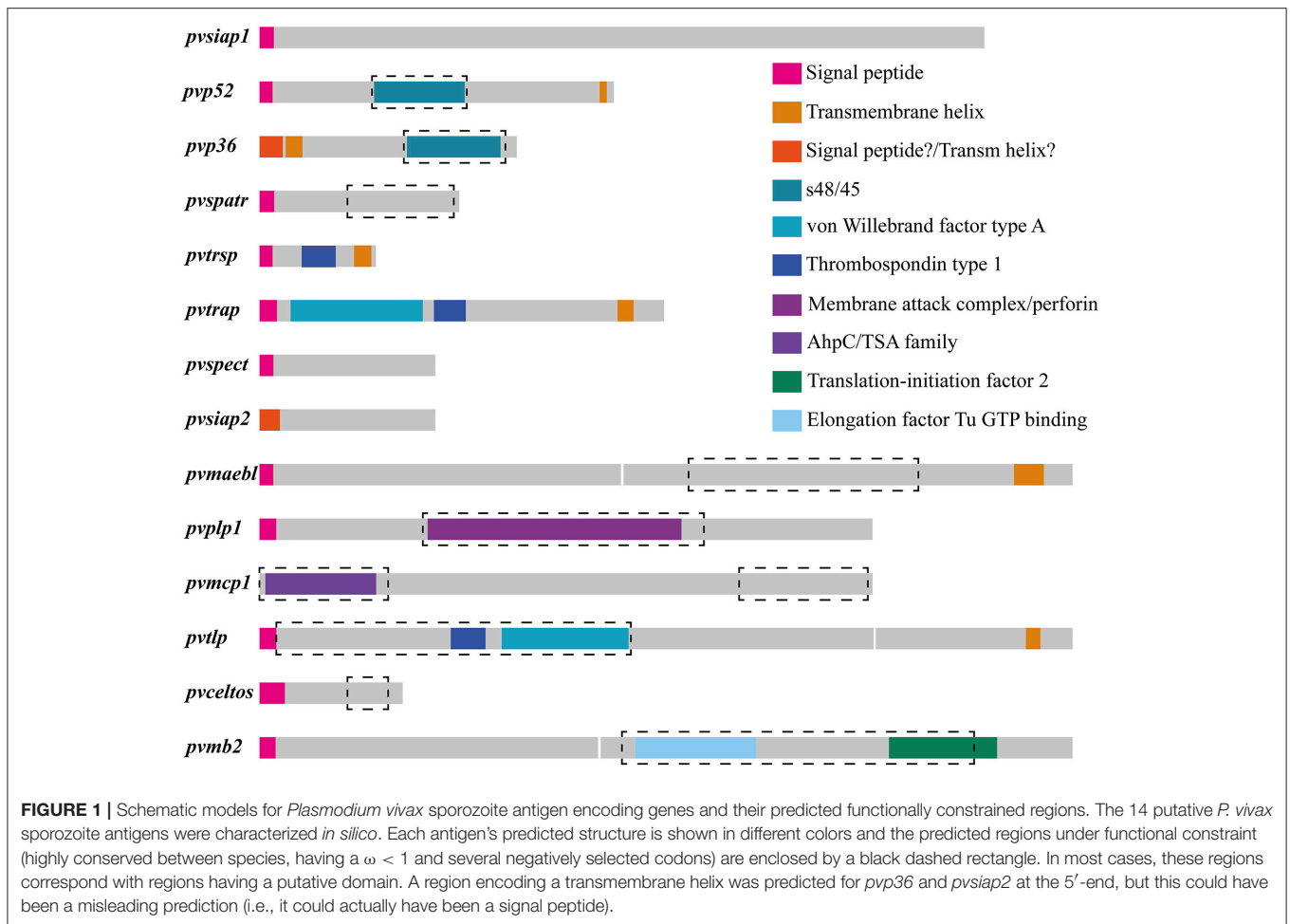


TABLE 2 | Estimating 14 *P. vivax* sporozoite genes' genetic diversity.

<i>n</i>	Gene	Sites	Ss	S	Ps	H	\bar{K}^{DNA}	θ_W (SD)	π^{DNA} (SD)	\bar{d}^{AA}	ρ^{AA}
15	<i>siap1</i>	2,991	12	8	4	8	2.55	0.0012 (0.0005)	0.0008 (0.0002)	2.00	0.0020
86	<i>p52</i>	1,461	9	2	7	13	1.98	0.0012 (0.0012)	0.0014 (0.0001)	1.49	0.0031
102	<i>p36</i>	1,059	6	3	3	9	1.13	0.0011 (0.0005)	0.0011 (0.0001)	0.67	0.0019
91	<i>spatr</i>	822	4	2	2	5	0.21	0.0009 (0.0005)	0.0003 (0.0001)	0.09	0.0003
63	<i>trsp</i>	480	2	0	2	3	0.59	0.0009 (0.0006)	0.0012 (0.0002)	0.59	0.0037
81	<i>trap</i>	1,668	28	9	19	20	5.11	0.0036 (0.0011)	0.0031 (0.0002)	3.91	0.0070
86	<i>spect1</i>	723	1	0	1	2	0.05	0.0003 (0.0003)	0.0001 (0.0000)	0.05	0.0002
79	<i>siap2</i>	1,239	25	8	17	25	2.09	0.0041 (0.0013)	0.0017 (0.0002)	1.99	0.0048
12	<i>maebi</i>	5,598	25	22	3	6	5.15	0.0015 (0.0006)	0.0009 (0.0005)	2.49	0.0013
73	<i>plp1</i>	2,529	13	5	8	14	0.77	0.0011 (0.0004)	0.0003 (0.0001)	0.40	0.0004
92	<i>mcp1</i>	1,455	13	4	9	19	1.44	0.0020 (0.0007)	0.0009 (0.0001)	0.84	0.0017
57	<i>tlp</i>	4,308	25	11	14	16	2.45	0.0013 (0.0004)	0.0006 (0.0001)	1.68	0.0012
101	<i>celtos</i>	588	6	3	3	10	0.89	0.0020 (0.0009)	0.0015 (0.0002)	0.76	0.0039
45	<i>mb2</i>	4,044	35	16	19	30	5.08	0.0020 (0.0006)	0.0012 (0.0001)	2.28	0.0017

The genetic diversity estimators were calculated from a variable amount of sequences for each antigen. *n*: amount of sequences analyzed; Sites, total of sites analyzed, excluding gaps; Ss, amount of segregating sites; S, amount of singleton sites; Ps, amount of informative-parsimonious sites; H, amount of haplotypes; \bar{K}^{DNA} , The average amount of nucleotide differences; θ_W , Watterson estimator; π^{DNA} , The average amount of nucleotide differences per site between two sequences; \bar{d}^{AA} , The average amount of aa differences; ρ^{AA} , The average amount of aa differences per site between two sequences; SD, standard deviation.

TABLE 3 | Neutrality and codon usage bias tests.

Gene	Tajima		Fu & Li		Fay & Wu	$d_N - d_S$ (ES)	$K_N - K_S$ (ES)	MKT	ENC	CBI
	D	D	F	H	NI					
<i>siap1</i>	-1.20	-1.71	-1.84	-0.44	0.000 (0.000)	-0.153 (0.010) ^{††}	13.6 ^{††}	52.0	0.27	
<i>p52</i>	0.27	1.29	0.94	-2.93*	0.000 (0.000)	-0.023 (0.003)**	2.87	53.6	0.22	
<i>p36</i>	-0.36	-1.80	-1.79	0.44	-0.001 (0.002)	-0.023 (0.003) ^{††}	3.82	54.2	0.23	
<i>spatr</i>	-1.42	-1.47	-1.71	-0.22	-0.001 (0.001)	-0.034 (0.005) ^{††}	1.56	49.6	0.39	
<i>trsp</i>	0.64	0.72	0.81	0.46	0.002 (0.001)	-0.007 (0.005)	Null	50.7	0.45	
<i>trap</i>	-0.48	-0.24	-0.36	-11.61**	0.002 (0.001)	-0.014 (0.004) ^{††}	6.46 ^{††}	55.8	0.20	
<i>spect1</i>	-0.91	0.50	0.09	0.04	0.000 (0.000)	-0.014 (0.003) ^{††}	Null	47.5	0.29	
<i>siap2</i>	-1.80*	-0.75	-1.35	-2.07	0.002 (0.001)**	-0.003 (0.003)	Null [†]	47.5	0.36	
<i>maebl</i>	-1.68	-1.20	-1.64	-10.21**	-0.001 (0.000)**	-0.066 (0.007) ^{††}	0.39 [†]	49.1	0.30	
<i>plp1</i>	-2.11*	-0.78	-1.46	-3.16*	-0.002 (0.001)*	-0.033 (0.003)*	2.09	52.3	0.28	
<i>mcp1</i>	-1.19	-1.44	-1.64	0.57	-0.001 (0.001)	-0.045 (0.005) ^{††}	1.64	47.8	0.28	
<i>tlp</i>	-1.76	-1.39	-1.95	-5.22**	-0.0001 (0.000)	-0.051 (0.004) ^{††}	1.81	52.7	0.20	
<i>celtos</i>	-0.80	-0.82	-0.87	-1.49	0.001 (0.001)	-0.036 (0.0027)	6.36	53.8	0.35	
<i>mb2</i>	-1.31	-1.67	-1.84	-3.41	-0.001 (0.001)*	-0.074 (0.004) ^{††}	1.26	53.4	0.25	

Selective pressure on each gene was inferred from neutrality tests and estimating effective codon usage. Non-synonymous substitution rate (d_N) and synonymous substitution rate (d_S) in *P. vivax*. Non-synonymous (K_N) and synonymous (K_S) divergence between *P. vivax* and the phylogenetically closest specie. Neutrality index (NI) estimated by McDonald-Kreitman test, using Jukes-Cantor correction. The preferential use of the synonymous codons was evaluated by estimating the effective amount of codons (ENC) and the codon bias index (CBI). Genes under selection in the tests had * $p < 0.05$, ** $p < 0.01$, [†] $p < 0.03$, *** $p < 0.006$, and ^{††} $p < 0.0001$.

regarding this parasite have been from the blood stage; few genes/proteins from others stages have been characterized. Although *P. vivax* biology cannot be assessed directly, some new technologies enable making inferences about it. Whole genome sequences could be used to infer the genes in a particular parasite and thus gene ontology could provide clues about its biology. At least 15 *P. falciparum* Spz stage proteins are involved in parasite invasion and could thus become vaccine candidates. Since many genes are shared amongst *Plasmodium* species (Carlton et al., 2008), and several of them seem to be essential for parasite survival (Bushell et al., 2017), orthologs to the 15 *P. falciparum* Spz antigens might be present in the *P. vivax* genome and could thus be taken into account for vaccine development. Fourteen of the 15 *P. falciparum* Spz genes found in *P. vivax* had similar gene/protein structures, suggesting that they could be involved in a conserved *Plasmodium* Spz invasion pathway. However, gene/protein identification is the first step in vaccine design. As *P. vivax* genetic diversity represents an even greater challenge than in *P. falciparum* concerning vaccine design, the next step in this work was to assess the genetic diversity of these Spz loci using available whole genome sequences to find out which of them could be promising vaccine candidates against *P. vivax*.

P. vivax Spz Loci Genetic Diversity

It has been suggested that the *P. vivax* genome's large genetic diversity would hinder *P. vivax* control and elimination. Several *P. vivax* blood-stage antigens have been considered as vaccine candidates; however, they have high genetic diversity (Putaporntip et al., 1997, 2010; Gomez et al., 2006; Dias et al., 2011; Premaratne et al., 2011; Garzón-Ospina et al., 2012, 2014, 2015), representing one of the challenges to be overcome

when designing a completely-effective anti-malarial vaccine. Contrasting with the aforementioned antigens, the Spz loci analyzed here had low genetic diversity ($\pi < 0.003$) which is a desirable feature when designing a fully-effective antimalarial vaccine. Few segregating sites were found at each locus; some were singleton sites. This could have resulted from genetic differentiation amongst *P. vivax* populations worldwide, as has been shown previously (Taylor et al., 2013; Hupalo et al., 2016). Nevertheless, fully-conserved regions were also identifiable in all loci. The observed polymorphism in Spz genes was comparable to that found in the most of the conserved Mrz genes described to date (Putaporntip et al., 2009; Garzón-Ospina et al., 2010, 2011, 2015; Pacheco et al., 2012; Chenet et al., 2013; Forero-Rodriguez et al., 2014a,b; Buitrago et al., 2016) being *pvspect*, *pvspar*, *pvsia1*, *pvplp1*, and *pvtlp* the loci with the lowest diversity. The most polymorphic gene was *pvtrap*; its diversity pattern has previously been reported for the *P. falciparum* ortholog (Ohashi et al., 2014).

Evolutionary Forces Modulating *P. vivax* Sporozoite Diversity

Contrasting with Mrz proteins (Putaporntip et al., 1997, 2010; Gomez et al., 2006; Dias et al., 2011; Garzón-Ospina et al., 2012, 2014, 2015), Spz antigens evaluated here displayed low diversity. This could have resulted from various evolutionary forces. Taking into account that a low amount of synonymous and non-synonymous polymorphism was found, low diversity could have been a consequence of codon usage bias. ENC and CBI parameters were thus estimated for evaluating such hypothesis. ENC values close to 61 indicated that all synonymous codons for each aa were used equitably (values close to 0 suggest bias or preferential codon use). CBI values range from 0, meaning

TABLE 4 | LD and intra-gene recombination estimators.

Gene	Zns	ZZ	RM	GARD
<i>siap1</i>	0.12	-0.00	1	-
<i>p52</i>	0.08	-0.06*	2	369 [†]
<i>p36</i>	0.01	-0.00	1	-
<i>spatr</i>	0.54	0.18*	0	-
<i>trsp</i>	0.04	0.00	0	-
<i>trap</i>	0.10	0.06	22	1,200 [†]
<i>spect1</i>	-	-	-	-
<i>siap2</i>	0.03	-0.02	2	-
<i>maebi</i>	0.73*	0.16	0	-
<i>plp1</i>	0.03	-0.01	2	1,275**
<i>mcp1</i>	0.02	-0.02	2	-
<i>tlp</i>	0.08	0.12	2	2,551 [†]
<i>celtos</i>	0.02	0.01	0	355 [†]
<i>mb2</i>	0.04	0.04	7	1,368 [†] and 2,605 [†]

GARD, recombination breakpoint position. * $p < 0.04$, ** $p < 0.005$, [†] $p < 0.001$.

uniform synonymous codon usage, to 1 (i.e., maximum codon bias) (Morton, 1993). Spz genes' ENC and CBI values suggested that all of them had random synonymous codon usage. Such values were similar to those previously described for genes participating in Mrz invasion and the value reported for the complete genome (Cornejo et al., 2014). This meant continuous transcription related to these proteins' level of expression during invasion (Gajbhiye et al., 2017; Uddin, 2017) which was not affected by any type of selective pressure or preferential codon usage. The low genetic diversity found in Spz genes was therefore not a consequence of selection at translation level and thus other evolutionary forces must be causing low genetic diversity in these antigens.

Test of neutral molecular evolution (e.i. Tajima, Fu & Li, Fay & Wu) gave negative values, just a few being statistically significant; the neutrality could not thus be ruled out in antigens lacking significant values. Since these genes were highly conserved, a functional/structural constraint was likely (Kimura, 1983; Graur et al., 2013). Although some genes had no statistically significant values, some of them had specific regions where neutrality could be ruled out, suggesting that negative selection was acting in such regions. Likewise, genes having an overall statistically significant negative value in these tests suggested that negative selection was responsible for low protein diversity and consequently functional/structural constraints would also be expected under this kind of selection. This was confirmed when non-synonymous and synonymous rates, as well as evolutionary rate (ω) sliding windows, were computed. Synonymous mutations seemed to fix at a higher rate than non-synonymous ones after speciation involving monkey-malaria parasites; this thus agreed with a hypothesis of functional/structural constraint. Furthermore, several negative selected sites (codons) were found; many were located in putative functional domains, suggesting that negative selection is an important force for maintaining protein domain integrity. These

regions had statistically significant negative values in the tests based on polymorphism frequency spectrum (e.i. Tajima, Fu & Li, Fay & Wu). These patterns could thus be used to predict functionally important regions. A recent report regarding *P. vivax* showed that regions having low ω rates having several negatively selected sites are regions used by the parasite to recognize host cells (Baquero et al., 2017). Consequently, the regions from the 14 antigens considered here having low ω rates and several codons under negative selection could be those used to recognize hepatocytes. However, these regions might not necessarily be involved in host-parasite interaction and they could have other functions. However, they could be taken into account for *P. vivax* vaccine development since they had low diversity and were under functional/structural constraint. The aforementioned results could then also be used for elucidating aspects of *P. vivax* sporozoite invasion of target cells by assessing the domains having functional constraint in invasion assays.

Few positively selected sites were found; they could have been fixed to adapt to different selective pressures. Molecular adaptations could play an important role during parasite evolutionary history. Monkey-malaria clades diversify three to four times more rapidly than those infecting other mammals (Muehlenbein et al., 2015). Taking host species' rapid diversification into account (Ziegler et al., 2007), adaptive radiation in monkey-malaria parasites could explain this accelerated cladogenesis and therefore several molecular adaptations could have arisen during such radiation. Phylogenies inferred for Spz antigens showed that monkey-malaria parasite topology did not agree with malaria-species relationships which could have resulted from positive selection (Sawai et al., 2010). There was evidence of selective sweep in five *P. vivax* genes; this pattern has already been observed for some of these genes in previous studies (Shen et al., 2017). Some mutations fixed in this parasite would thus allow it to adapt to a new host after host-switch decreasing genetic diversity.

Additionally, phylogenetically-based analysis could provide greater insight into the role of selection (at these loci) during a parasite's evolutionary history if this was the result of adaptations to new hosts and/or environments (Muehlenbein et al., 2015). There was evidence of episodic positive selection for nine of these 14 antigens. Throughout phylogenies, several lineages had some sites under positive selection; however, few of them were ancestral lineages. Since most branches under selection lead to particular species, episodic selection could have resulted from adaptation to different host, for instance, to avoid immune response to a particular host during sympatric speciation or to recognize a new host receptor. However, this behavior seems to be common in the *Plasmodium* genus and not just for monkey-malaria parasites.

Recombination is an evolutionary force which can increase genetic diversity. It could be acting on Mrz blood-stage antigens leading to new haplotypes to arise, being maintained in the parasite population to evade the host's immune responses (Garzón-Ospina et al., 2012). Even though low genetic diversity was observed, some Spz genes had a large amount of haplotypes. Nevertheless, larger-sized genes accumulated a greater amount

of single nucleotide polymorphisms in their sequences; these characteristics were not related to the amount of haplotypes. Some haplotypes could thus have arisen by recombination. Evidence of intra-gene recombination was found for *pvp52*, *pvp1p1*, *pvt1p*, *pvceltos*, and *pvm2* genes. Recombination is thus an evolutionary force in these loci increasing genetic diversity.

***P. vivax* Spz Antigens' Putative Roles**

The mechanism by which these Spz proteins act in *P. vivax* is not certain; however, their possible role could be elucidated from them having putative domains. SPECT, PLP1, and TLP seem to be essential in different *Plasmodium* species for traversing host cells (Ishino et al., 2004; Yang et al., 2017a) since deleting them has reduced parasite capability to traverse the sinusoidal barrier and thus gain access to hepatocytes (Ishino et al., 2004; Kaiser et al., 2004). No described domains were identified for PvSPECT and PvSPART and their role thus requires further investigation. Nevertheless, orthologs to these proteins have been implicated in parasite interaction with epithelial or hepatic host cells in *P. falciparum* (Curtidor et al., 2011). Even though putative protein-protein interaction domains were not identified, the regions involved in such interaction can be predicted regarding their degree of conservation between species, as can their evolutionary rates (ω) (Graur et al., 2013; Baquero et al., 2017). According to the sliding window for the gene encoding to SPART in *P. vivax*, the C-terminal region seemed to be functionally restricted as low genetic diversity (intra- e inter-species) was observed, as well as lower than 1 ω rate and various sites under negative selection (at inter-specific level); this region could be implicated in interaction with hepatocytes. Additionally, PfSPART is antigenic in natural infections and antibodies against it have blocked interaction with hepatocytes *in vitro* (Palaeya et al., 2013). Similar to PfSPART, PvSPART has antigenic potential given the 4 regions along its sequence having high solvent accessibility values and potential linear B-epitopes, suggesting that PvSPART is a promising antigen when designing an anti-*P. vivax* vaccine.

Similar to the aforementioned antigens, particular domains were not found in the whole PvSIAP1 sequence. Orthologs for this protein in *Plasmodium* spp, are involved in Spz exit from oocysts, as well as their colonization in a mosquito's salivary glands (Engelmann et al., 2009). SIAP1 in a vertebrate host seems to be mediated by pathogen-host interaction (Curtidor et al., 2011). This protein's ortholog in *P. vivax* has been shown to have low diversity ($\omega < 1$) throughout its sequence, as well as several negatively-selected sites. However, putative functional regions have not been clearly defined due to a large amount of sites under positive selection all along the sequence of the gene encoding this protein. Nevertheless, the C-terminal region was the region where most codons were found to be under negative selection and could thus be the region used for *P. vivax* interaction with a particular host.

PLP1 is not required for entry to hepatocytes, but does play an important role in exit from transitory vacuoles during cell traversal (Risco-Castillo et al., 2015). A putative MAC/perforin domain was identified in *P. vivax* which had a high sequence conservation, several codons under negative selection and a $\omega < 1$; this functionally restricted region could thus be mediating

membrane destabilization and pore formation in *P. vivax*, as has been suggested in other species (Rosado et al., 2008; Patarroyo et al., 2016; Yang et al., 2017a).

It is well known that the TLP (Moreira et al., 2008) and TRAP (Sultan et al., 1997) are essential for Spz gliding motility (Sultan et al., 1997). *P. falciparum* TLP is the most conserved member of the TRAP/MIC2 family and the first to be seen to play a role traversing cells (Moreira et al., 2008). *P. vivax* TLP was one of the most conserved proteins of those evaluated here; the N-terminal region was highly conserved within and amongst species. Thrombospondin type 1 (TSP1) and von Willebrand factor type A (vWa) domains were observed in this region. The *P. falciparum*, TSP1 domain mediates glycosaminoglycan binding whilst the vWa domain is involved in cell-cell, cell-matrix, matrix-matrix interactions and includes a metal-ion dependent adhesion site. Proteins containing such domains might be associated with parasite invasion ability (Wengelnik et al., 1999; Matuschewski et al., 2002; Mongui et al., 2010); the PvTLP N-terminal region could thus be mediating *P. vivax* Spz invasion of hepatocytes.

Unlike a PvTLP, PvTRAP (having the same domains toward the N-terminal region) was the protein having the greatest diversity of those evaluated here. This gene is highly polymorphic in *P. falciparum* and is under positive selection (Ohashi et al., 2014); the region encoding the vWA domain is where *pftrap* diversity is concentrated (Moreira et al., 2008). Such pattern seems to be similar in *P. vivax* and its related species; the sliding window for this gene showed that this region evolved rapidly, having several sites under positive selection amongst species. This could have resulted from lineage-specific adaptations for the recognition of particular receptors for each host or is the target region for immune responses, similar to that which occurs in PvDBP (VanBuskirk et al., 2004; Chootong et al., 2014). A prediction of B-epitopes and solvent accessibility showed that TRAP should have a potentially antigenic region toward the N-terminal where the vWa domain would be located. However, the C-terminal region seemed to have greater antigenic potential. According to previous studies PvTRAP can induce an IgG1 and IgG3 response following natural infection; producing this type of antibody is usually correlated with protection against disease in a hyper-endemic region (Nazeri et al., 2017). This protein has thus been proposed as being an important immune response target regarding pre-erythrocyte stages of malaria caused by both *P. falciparum* and *P. vivax* (Ohashi et al., 2014). Nevertheless, the diversity observed in this protein could produce allele-specific immune responses and would thus not be a good vaccine candidate.

Unlike TRAP and TLP, PvTRSP only has the TSP1 domain, this protein is involved in *P. berghei* invasion where the deletion of *pbtrsp* reduces mutant parasite capability to enter hepatocytes (Labaied et al., 2007a). This gene was highly conserved in *P. vivax*; however, when the sequences from *P. vivax* and related species were compared, the TSP1 domain was seen to have a $\omega > 1$ and few sites under negative selection. On the other hand, the C-terminal region seemed to be this protein's antigenic region.

MAEBL has been described as being a type I membrane protein in *P. falciparum*, having erythrocyte-binding activity

and seeming to fulfill a function similar to that of AMA1, given the high sequence similarity (Yang et al., 2017b). This protein is necessary for Spz invasion of a mosquito's salivary glands and for them to traverse vertebrate host cells (Yang et al., 2017b). MAEBL has 4 tandem repeats in *P. vivax*, located toward the gene's 3' region, between positions 3,427 and 4,756. Each repeat has 90% similarity; 9 imperfect copies of GCTAGAAGGGCTGAGGAGT residues, 3 copies of AGAAAGGCGGAAGAGGCA, 17 copies of GCAAGGAAGGCAGAGGATGCTAGAAAGGCAGAGGCGGCTA and 6 copies of GCTAAAAGGCTGAA GCAGCAAGGAAGGCAGAGGCA residues were found. In spite of an accumulation of repeats being responsible for size polymorphism, they did not show such polymorphism when sequences from different isolates were analyzed. Calculating the omega rate between *P. vivax* isolates and related species revealed that such repeats were highly conserved in *Plasmodium*, as has been suggested previously (Leite et al., 2015). The fact that this repeat region was found to be under negative selection suggested an important role, similar to that already reported for CSP (Aldrich et al., 2012; Ferguson et al., 2014). A prediction of linear epitopes suggested that the C-terminal region where these repeats were located could be antigenic. Even though the repeats have been suggested as targets distracting an immune response, this is still not clear because they are highly conserved, even between species. Conversely, the *maebl* 5'-end was highly divergent amongst species, having several sites under positive selection, possibly resulting from species-specific adaptations.

MB2 (just like MAEBL) is one of the largest proteins expressed in *Plasmodium* spp Spz and is involved in hepatocyte invasion (Nguyen et al., 2009). It contains a characteristic GTP-binding domain (Nguyen et al., 2001; Romero et al., 2004), that is also present in its *P. vivax* counterpart; in addition, a translation-initiation factor domain was predicted for PvMB2. Both were functionally restricted, being conserved in *P. vivax* as well as between phylogenetically-related species. Even though it is not clear whether these domains are mediated by pathogen-host interaction, they do seem to be important for this protein's function. This protein is antigenic in *P. falciparum* and has been recognized by patients showing protection against *Plasmodium* infection following experimental immunization (Nguyen et al., 2009). The region so recognized is the N-terminal region, which had antigenic potential in *P. vivax*, given its solvent accessibility and the prediction of B-linear epitopes. Given low PvMB2 diversity and its antigenic potential, it could also be taken into account when designing a vaccine against *P. vivax*.

Members of the 6-cys family are expressed during different *Plasmodium* spp. stages. Two members of this family (P36 and P52) are expressed during the pre-erythrocyte stage and seem to play an important role in invasion (Labaied et al., 2007b; Annoura et al., 2014; Kublin et al., 2017; Manzoni et al., 2017). Like 6-Cys members expressed in *P. vivax* Mrz (Forero-Rodriguez et al., 2014a,b), *pvp36* and *pvp52* displayed low genetic diversity, accompanied by low evolutionary rates. The s48/45 domains, characteristic of this family, seemed to be functionally restricted (limited diversity, $\omega < 1$ and various sites under negative selection) and thus might have been responsible for the interaction between *P. vivax* Spz and hepatocytes. The proteins

encoded by these genes also have antigenic potential; the PvP52 C-terminal region seemed to be exposed, having potential linear B-epitopes, whilst the PvP38 central region and C-terminal could be antigenic regions. Thus, the same as other 6-Cys family members (Forero-Rodriguez et al., 2014a,b), PvP36 and PvP52 would seem to be promising antigens when designing a vaccine.

In addition to TRAP, SIAP2, and CELTOS were the antigens having the highest diversity values amongst those evaluated here. These proteins are predominantly expressed during the Spz stage in other species where they cover parasite surface (Siau et al., 2008). Unlike the other proteins evaluated here, SIAP2 only had orthologs in *Plasmodium* species infecting primates. This is a potential vaccine candidate as it seems to interact specifically with heparin sulfate and chondroitin sulfate-type membrane receptors (Siau et al., 2008; Alba et al., 2011). Potential functional regions could not be defined for SIAP2 (like TRAP) since it had many sites under positive selection amongst species throughout its sequence. Predicted results regarding PvSIAP2 antigenicity suggested that this could be exposed to the immune system, having various potential B-epitopes along its sequence. Its use as vaccine candidate is limited due to its diversity and the large amount of haplotypes found.

CeLTOS has been considered a potential vaccine candidate given its association with clinical protection regarding naturally-acquired immunity (Kanoi et al., 2017). This protein has a sole specificity for phosphatidic acid (a lipid found predominantly on plasma membrane inner face) and breaks liposomes consisting of phosphatidic acid through pore formation. It has been shown that parasites lacking CeLTOS can enter target cells but remain trapped inside. The foregoing supposes that CeLTOS targets cell membrane inner leaflet which can burst due to pore formation inside infected cells and favors parasite exit (Kariu et al., 2006; Jimah et al., 2016). Even though its function seems to be intracellular, predictive analysis suggested that PvCeLTOS could be exposed to the immune system, having potential B-epitopes toward the C-terminal region.

A recent study has shown a correlation between the prediction and this protein's antigenic regions. Naturally-infected patients generate immune responses against PvCeLTOS, predominantly toward the protein's C-terminal region (Rodrigues-da-Silva et al., 2017), coinciding with the most exposed regions and having potential B-epitopes. This would suggest that PvCeLTOS is a potential antigen for inclusion in a vaccine against *P. vivax*. The antigenic region coincided with the gene's region where various sites under negative selection were found, suggesting that the CeLTOS functional region could be located in the C-terminal region; however, some sites under positive inter-species selection were found toward this region. If this region is functionally restricted, but is also the region toward which the immune response is directed in different species, then some sites could have been positively selected in a species-specific manner fixing specific mutations in each species, resulting in a positive selection signal. In addition to its antigenic potential, PvCeLTOS has been shown to have limited diversity within *P. vivax*. The gene encoding this protein in Iran had, on average, less than one mutation in paired comparisons (Mehrizi et al., 2017); such results were similar to those reported here. PvCeLTOS

might thus be considered a vaccine candidate, given its antigenic characteristics and limited diversity.

CONCLUSIONS

Designing a completely effective vaccine against *P. vivax* must include antigens or highly conserved regions containing them to abolish allele-specific immune responses; antigens from different stages must also be included. As most research has been focused on the blood stage, new strategies must be implemented for identifying potential vaccine candidates from other parasite stages. As genetic diversity is a characteristic to be born in mind when designing a vaccine, the present work could be taken as a basis for selecting Spz antigens which might become potential vaccine candidates.

P. vivax genetic diversity would be expected to be high as it has been shown in whole genome analysis (Neafsey et al., 2012) and in several Mrz antigens (Gomez et al., 2006; Putaporntip et al., 2010; Dias et al., 2011; Premaratne et al., 2011; Garzón-Ospina et al., 2012). However, this study's results showed that some Spz antigens had limited genetic diversity and, consequently, they could be good vaccine candidates, since low genetic diversity is ideal for avoiding an allele-specific immune response. The analysis described above enabled determining which regions in these proteins could be functionally important. Recent studies have shown that regions predicted to have functional restriction coincide with regions used by the parasite to bind to a host cell (Baquero et al., 2017). Therefore, regions in Spz antigens which were conserved between species having low ω values and codons under negative selection in the parts of a protein might be thus implicated in these antigens' function. This could help to elucidate aspects of *P. vivax* Spz invasion. Further assays regarding interaction with hepatocytes are thus required for determining whether such regions are involved in pathogen-host interaction.

Promising antigens to be taken into account for designing a fully effective vaccine are those having a limited genetic

diversity or at least one domain with such pattern. These genes or domains must have a negative selection signal, as well as a $\omega < 1$ (Garzón-Ospina et al., 2015). Thus, the results reported here suggest that the PvP52, PvP36, PvSPATR, PvPLP1, PvMCP1, PvTLP, PvCelTOS, and PvMB2 antigens (or functionally restricted regions of them) are promising vaccine candidates and thus should be prioritized in future studies aimed at developing a completely effective vaccine against *P. vivax*.

AUTHOR CONTRIBUTIONS

DG-O, SB, and AR acquired the genomic data and carried out the genetic diversity analyses. DG-O performed the general interpretation of the data and together with SB and AR participated in writing the paper. MP coordinated the study and helped to write the manuscript. All the authors have read and approved the final version of the manuscript.

FUNDING

This work was financed by the Departamento Administrativo de Ciencia, Tecnología e Innovación (COLCIENCIAS), through grant RC # 0309-2013.

ACKNOWLEDGMENTS

We would like to thank Jason Garry for translating and reviewing the manuscript and also thank the team in charge of updating the PlasmoDB database for providing the genetic information from which this work was carried out.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00010/full#supplementary-material>

REFERENCES

- Abascal, F., Zardoya, R., and Telford, M. J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* 38, W7–W13. doi: 10.1093/nar/gkq291
- Adams, J. H., and Mueller, I. (2017). The biology of *plasmodium vivax*. *Cold Spring Harb. Perspect. Med.* 7, 1–13. doi: 10.1101/cshperspect.a025585
- Alba, M. P., Almonacid, H., Calderón, D., Chacón, E. A., Poloche, L. A., Patarroyo, M. A., et al. (2011). 3D structure and immunogenicity of *Plasmodium falciparum* sporozoite induced associated protein peptides as components of fully-protective anti-malarial vaccine. *Biochem. Biophys. Res. Commun.* 416, 349–355. doi: 10.1016/j.bbrc.2011.11.039
- Aldrich, C., Magini, A., Emiliani, C., Dottorini, T., Bistoni, F., Crisanti, A., et al. (2012). Roles of the amino terminal region and repeat region of the *Plasmodium berghei* circumsporozoite protein in parasite infectivity. *PLoS ONE* 7:e32524. doi: 10.1371/journal.pone.0032524
- Annoura, T., van Schaijk, B. C., Ploemen, I. H., Sajid, M., Lin, J. W., Vos, M. W., et al. (2014). Two Plasmodium 6-Cys family-related proteins have distinct and critical roles in liver-stage development. *FASEB J.* 28, 2158–2170. doi: 10.1096/fj.13-241570
- Baquero, L. A., Moreno-Pérez, D. A., Garzón-Ospina, D., Forero-Rodríguez, J., Ortiz-Suárez, H. D., and Patarroyo, M. A. (2017). PvGAMA reticulocyte binding activity: predicting conserved functional regions by natural selection analysis. *Parasit. Vectors* 10:251. doi: 10.1186/s13071-017-2183-8
- Barry, A. E., and Arnott, A. (2014). Strategies for designing and monitoring malaria vaccines targeting diverse antigens. *Front. Immunol.* 5:359. doi: 10.3389/fimmu.2014.00359
- Buitrago, S. P., Garzón-Ospina, D., and Patarroyo, M. A. (2016). Size polymorphism and low sequence diversity in the locus encoding the *Plasmodium vivax* rhoptry neck protein 4 (PvRON4) in Colombian isolates. *Malar. J.* 15:501. doi: 10.1186/s12936-016-1563-4
- Bushell, E., Gomes, A. R., Sanderson, T., Anar, B., Girling, G., Herd, C., et al. (2017). Functional profiling of a plasmodium genome reveals an abundance of essential genes. *Cell* 170, 260–272 e268. doi: 10.1016/j.cell.2017.06.030
- Carlton, J. M., Adams, J. H., Silva, J. C., Bidwell, S. L., Lorenzi, H., Caler, E., et al. (2008). Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature* 455, 757–763. doi: 10.1038/nature07327
- Casillas, S., and Barbadailla, A. (2017). Molecular population genetics. *Genetics* 205, 1003–1035. doi: 10.1534/genetics.116.196493

- Castellanos, A., Arévalo-Herrera, M., Restrepo, N., Gulluso, L., Corradin, G., and Herrera, S. (2007). *Plasmodium vivax* thrombospondin related adhesion protein: immunogenicity and protective efficacy in rodents and *Aotus monkeys*. *Mem. Inst. Oswaldo Cruz* 102, 411–416. doi: 10.1590/S0074-02762007005000047
- Chan, E. R., Menard, D., David, P. H., Ratsimbasa, A., Kim, S., Chim, P., et al. (2012). Whole genome sequencing of field isolates provides robust characterization of genetic diversity in *Plasmodium vivax*. *PLoS Negl. Trop. Dis.* 6:e1811. doi: 10.1371/journal.pntd.0001811
- Chaurio, R. A., Pacheco, M. A., Cornejo, O. E., Durrego, E., Stanley, C. E. Jr., Castillo, A. I., et al. (2016). Evolution of the transmission-blocking vaccine candidates Pvs28 and Pvs25 in *Plasmodium vivax*: geographic differentiation and evidence of positive selection. *PLoS Negl. Trop. Dis.* 10:e0004786. doi: 10.1371/journal.pntd.0004786
- Chenet, S. M., Pacheco, M. A., Bacon, D. J., Collins, W. E., Barnwell, J. W., and Escalante, A. A. (2013). The evolution and diversity of a low complexity vaccine candidate, merozoite surface protein 9 (MSP-9), in *Plasmodium vivax* and closely related species. *Infect. Genet. Evol.* 20, 239–248. doi: 10.1016/j.meegid.2013.09.011
- Chootong, P., McHenry, A. M., Ntumngia, F. B., Sattabongkot, J., and Adams, J. H. (2014). The association of Duffy binding protein region II polymorphisms and its antigenicity in *Plasmodium vivax* isolates from Thailand. *Parasitol. Int.* 63, 858–864. doi: 10.1016/j.parint.2014.07.014
- Coatney, G.R., and National Institute of Allergy and Infectious Diseases (U.S.) (1971). *The Primate Malariae*. Bethesda, MD: U.S. National Institute of Allergy and Infectious Diseases; for sale by the Supt. of Docs., U.S. Government Printing off Washington.
- Cornejo, O. E., Fisher, D., and Escalante, A. A. (2014). Genome-wide patterns of genetic polymorphism and signatures of selection in *Plasmodium vivax*. *Genome Biol. Evol.* 7, 106–119. doi: 10.1093/gbe/evu267
- Curtidor, H., Vanegas, M., Alba, M. P., and Patarroyo, M. E. (2011). Functional, immunological and three-dimensional analysis of chemically synthesised sporozoite peptides as components of a fully-effective antimalarial vaccine. *Curr. Med. Chem.* 18, 4470–4502. doi: 10.2174/092986711797287575
- Delpont, W., Poon, A. F., Frost, S. D., and Kosakovsky Pond, S. L. (2010). Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26, 2455–2457. doi: 10.1093/bioinformatics/btq429
- Dias, S., Longacre, S., Escalante, A. A., and Udagama-Randeniya, P. V. (2011). Genetic diversity and recombination at the C-terminal fragment of the merozoite surface protein-1 of *Plasmodium vivax* (PvMSP-1) in Sri Lanka. *Infect. Genet. Evol.* 11, 145–156. doi: 10.1016/j.meegid.2010.09.007
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Egea, R., Casillas, S., and Barbadilla, A. (2008). Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Res.* 36, W157–W162. doi: 10.1093/nar/gkn337
- Engelmann, S., Silvie, O., and Matuschewski, K. (2009). Disruption of *Plasmodium* sporozoite transmission by depletion of sporozoite invasion-associated protein 1. *Eukaryotic Cell* 8, 640–648. doi: 10.1128/EC.00347-08
- Fankhauser, N., and Mäser, P. (2005). Identification of GPI anchor attachment signals by a Kohonen self-organizing map. *Bioinformatics* 21, 1846–1852. doi: 10.1093/bioinformatics/bti299
- Fay, J. C., and Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. *Genetics* 155, 1405–1413.
- Ferguson, D. J., Balaban, A. E., Patzewitz, E. M., Wall, R. J., Hopp, C. S., Poulin, B., et al. (2014). The repeat region of the circumsporozoite protein is critical for sporozoite formation and maturation in *Plasmodium*. *PLoS ONE* 9:e113923. doi: 10.1371/journal.pone.0113923
- Forero-Rodríguez, J., Garzón-Ospina, D., and Patarroyo, M. A. (2014a). Low genetic diversity and functional constraint in loci encoding *Plasmodium vivax* P12 and P38 proteins in the Colombian population. *Malar. J.* 13:58. doi: 10.1186/1475-2875-13-58
- Forero-Rodríguez, J., Garzón-Ospina, D., and Patarroyo, M. A. (2014b). Low genetic diversity in the locus encoding the *Plasmodium vivax* P41 protein in Colombia's parasite population. *Malar. J.* 13:388. doi: 10.1186/1475-2875-13-388
- Frevort, U. (2004). Sneaking in through the back entrance: the biology of malaria liver stages. *Trends Parasitol.* 20, 417–424. doi: 10.1016/j.pt.2004.07.007
- Fu, Y. X., and Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics* 133, 693–709.
- Gajbhiye, S., Patra, P. K., and Yadav, M. K. (2017). New insights into the factors affecting synonymous codon usage in human infecting *Plasmodium* species. *Acta Trop.* 176, 29–33. doi: 10.1016/j.actatropica.2017.07.025
- Garzón-Ospina, D., Forero-Rodríguez, J., and Patarroyo, M. A. (2014). Heterogeneous genetic diversity pattern in *Plasmodium vivax* genes encoding merozoite surface proteins (MSP)-7E, -7F and -7L. *Malar. J.* 13:495. doi: 10.1186/1475-2875-13-495
- Garzón-Ospina, D., Forero-Rodríguez, J., and Patarroyo, M. A. (2015). Inferring natural selection signals in *Plasmodium vivax*-encoded proteins having a potential role in merozoite invasion. *Infect. Genet. Evol.* 33, 182–188. doi: 10.1016/j.meegid.2015.05.001
- Garzón-Ospina, D., López, C., Forero-Rodríguez, J., and Patarroyo, M. A. (2012). Genetic diversity and selection in three *Plasmodium vivax* merozoite surface protein 7 (Pvmsp-7) genes in a Colombian population. *PLoS ONE* 7:e45962. doi: 10.1371/journal.pone.0045962
- Garzón-Ospina, D., Romero-Murillo, L., and Patarroyo, M. A. (2010). Limited genetic polymorphism of the *Plasmodium vivax* low molecular weight rhoptry protein complex in the Colombian population. *Infect. Genet. Evol.* 10, 261–267. doi: 10.1016/j.meegid.2009.12.004
- Garzón-Ospina, D., Romero-Murillo, L., Tóbon, L. F., and Patarroyo, M. A. (2011). Low genetic polymorphism of merozoite surface proteins 7 and 10 in Colombian *Plasmodium vivax* isolates. *Infect. Genet. Evol.* 11, 528–531. doi: 10.1016/j.meegid.2010.12.002
- Gomez, A., Suarez, C. F., Martinez, P., Saravia, C., and Patarroyo, M. A. (2006). High polymorphism in *Plasmodium vivax* merozoite surface protein-5 (MSP5). *Parasitology* 133(Pt 6), 661–672. doi: 10.1017/S0031182006001168
- Graur, D., Zheng, Y., Price, N., Azevedo, R. B., Zufall, R. A., and Elhaik, E. (2013). On the immortality of television sets: “function” in the human genome according to the evolution-free gospel of ENCODE. *Genome Biol. Evol.* 5, 578–590. doi: 10.1093/gbe/evt028
- Hester, J., Chan, E. R., Menard, D., Mercereau-Puijalon, O., Barnwell, J., Zimmerman, P. A., et al. (2013). De novo assembly of a field isolate genome reveals novel *Plasmodium vivax* erythrocyte invasion genes. *PLoS Negl. Trop. Dis.* 7:e2569. doi: 10.1371/journal.pntd.0002569
- Hudson, R. R., and Kaplan, N. L. (1985). Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111, 147–164.
- Hulden, L. (2011). Activation of the hypnozoite: a part of *Plasmodium vivax* life cycle and survival. *Malar. J.* 10:90. doi: 10.1186/1475-2875-10-90
- Hupalo, D. N., Luo, Z., Melnikov, A., Sutton, P. L., Rogov, P., Escalante, A., et al. (2016). Population genomics studies identify signatures of global dispersal and drug resistance in *Plasmodium vivax*. *Nat. Genet.* 48, 953–958. doi: 10.1038/ng.3588
- Ishino, T., Yano, K., Chinzai, Y., and Yuda, M. (2004). Cell-passage activity is required for the malarial parasite to cross the liver sinusoidal cell layer. *PLoS Biol.* 2:e4. doi: 10.1371/journal.pbio.0020004
- Jimah, J. R., Salinas, N. D., Sala-Rabanal, M., Jones, N. G., Sibley, L. D., Nichols, C. G., et al. (2016). Malaria parasite CelTOS targets the inner leaflet of cell membranes for pore-dependent disruption. *Elife* 5:e20621. doi: 10.7554/eLife.20621
- Jukes, T. H., and Cantor, C. R. (1969). “Evolution of protein molecules,” in *Mammalian Protein Metabolism*, ed H. N. Munro (New York, NY: Academic Press), 21–132.
- Kall, L., Krogh, A., and Sonnhammer, E. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* 35, W429–W432. doi: 10.1093/nar/gkm256
- Kaiser, K., Camargo, N., Coppens, I., Morrissey, J. M., Vaidya, A. B., and Kappe, S. H. I. (2004). A member of a conserved *Plasmodium* protein family with membrane-attack complex/perforin (MACPF)-like domains localizes to the micronemes of sporozoites. *Mol. Biochem. Parasitol.* 133, 15–26. doi: 10.1016/j.molbiopara.2003.08.009
- Kanoi, B. N., Takashima, E., Morita, M., White, M. T., Palacpac, N. M., Ntege, E. H., et al. (2017). Antibody profiles to wheat germ cell-free system synthesized

- Plasmodium falciparum* proteins correlate with protection from symptomatic malaria in Uganda. *Vaccine* 35, 873–881. doi: 10.1016/j.vaccine.2017.01.001
- Kariu, T., Ishino, T., Yano, K., Chinzei, Y., and Yuda, M. (2006). CelTOS, a novel malarial protein that mediates transmission to mosquito and vertebrate hosts. *Mol. Microbiol.* 59, 1369–1379. doi: 10.1111/j.1365-2958.2005.05024.x
- Kelly, J. K. (1997). A test of neutrality based on interlocus associations. *Genetics* 146, 1197–1206.
- Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge; New York: Cambridge University Press. doi: 10.1017/CBO9780511623486
- Kolaskar, A. S., and Tongaonkar, P. C. (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett.* 276, 172–174. doi: 10.1016/0014-5793(90)80535-Q
- Kosakovsky Pond, S. L., and Frost, S. D. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* 22, 1208–1222. doi: 10.1093/molbev/msi105
- Kosakovsky Pond, S. L., Murrell, B., Fourment, M., Frost, S. D., Delpont, W., and Scheffler, K. (2011). A random effects branch-site model for detecting episodic diversifying selection. *Mol. Biol. Evol.* 28, 3033–3043. doi: 10.1093/molbev/msr125
- Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., and Frost, S. D. (2006). Automated phylogenetic detection of recombination using a genetic algorithm. *Mol. Biol. Evol.* 23, 1891–1901. doi: 10.1093/molbev/msl051
- Kublin, J. G., Mikolajczak, S. A., Sack, B. K., Fishbaugher, M. E., Seilie, A., Shelton, L., et al. (2017). Complete attenuation of genetically engineered *Plasmodium falciparum* sporozoites in human subjects. *Sci. Transl. Med.* 9:eaa9099. doi: 10.1126/scitranslmed.aad9099
- Labaied, M., Camargo, N., and Kappe, S. H. (2007a). Depletion of the *Plasmodium berghei* thrombospondin-related sporozoite protein reveals a role in host cell entry by sporozoites. *Mol. Biochem. Parasitol.* 153, 158–166. doi: 10.1016/j.molbiopara.2007.03.001
- Labaied, M., Harupa, A., Dumpit, R. F., Coppens, I., Mikolajczak, S. A., and Kappe, S. H. (2007b). *Plasmodium yoelii* sporozoites with simultaneous deletion of P52 and P36 are completely attenuated and confer sterile immunity against infection. *Infect. Immun.* 75, 3758–3768. doi: 10.1128/IAI.00225-07
- Leite, J. A., Bargieri, D. Y., Carvalho, B. O., Albrecht, L., Lopes, S. C., Kayano, A. C., et al. (2015). Immunization with the MAEBL M2 domain protects against lethal *Plasmodium yoelii* infection. *Infect. Immun.* 83, 3781–3792. doi: 10.1128/IAI.00262-15
- Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187
- Manzoni, G., Marinach, C., Topçu, S., Briquet, S., Grand, M., Tolle, M., et al. (2017). *Plasmodium P36* determines host cell receptor usage during sporozoite invasion. *Elife* 6. doi: 10.7554/eLife.25903
- Matuschewski, K., Nunes, A. C., Nussenzweig, V., and Ménard, R. (2002). *Plasmodium* sporozoite invasion into insect and mammalian cells is directed by the same dual binding system. *EMBO J.* 21, 1597–1606. doi: 10.1093/emboj/21.7.1597
- McDonald, J. H., and Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351, 652–654. doi: 10.1038/351652a0
- Mehrizi, A. A., Torabi, F., Zakeri, S., and Djadid, N. D. (2017). Limited genetic diversity in the global *Plasmodium vivax* Cell traversal protein of Ookinetes and Sporozoites (CelTOS) sequences; implications for PvCelTOS-based vaccine development. *Infect. Genet. Evol.* 53, 239–247. doi: 10.1016/j.meegid.2017.06.005
- Ménard, R. (2001). Gliding motility and cell invasion by Apicomplexa: insights from the *Plasmodium* sporozoite. *Cell. Microbiol.* 3, 63–73. doi: 10.1046/j.1462-5822.2001.00097.x
- Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). “Creating the CIPRES science gateway for inference of large phylogenetic trees,” in *Gateway Computing Environments Workshop (GCE)* (New Orleans, LA: IEEE), 1–8.
- Mongui, A., Angel, D. I., Moreno-Perez, D. A., Villarreal-Gonzalez, S., Almonacid, H., Vanegas, M., et al. (2010). Identification and characterization of the *Plasmodium vivax* thrombospondin-related apical merozoite protein. *Malar. J.* 9:283. doi: 10.1186/1475-2875-9-283
- Moreira, C. K., Templeton, T. J., Lavazec, C., Hayward, R. E., Hobbs, C. V., Kroeze, H., et al. (2008). The plasmodium TRAP/MIC2 family member, TRAP-Like Protein (TLP), is involved in tissue traversal by sporozoites. *Cell. Microbiol.* 10, 1505–1516. doi: 10.1111/j.1462-5822.2008.01143.x
- Morton, B. R. (1993). Chloroplast DNA codon use: evidence for selection at the psb A locus based on tRNA availability. *J. Mol. Evol.* 37, 273–280. doi: 10.1007/BF00175504
- Muehlenbein, M. P., Pacheco, M. A., Taylor, J. E., Prall, S. P., Ambu, L., Nathan, S., et al. (2015). Accelerated diversification of nonhuman primate malarial parasites in Southeast Asia: adaptive radiation or geographic speciation? *Mol. Biol. Evol.* 32, 422–439. doi: 10.1093/molbev/msu310
- Mueller, I., Shakri, A. R., and Chitnis, C. E. (2015). Development of vaccines for *Plasmodium vivax* malaria. *Vaccine* 33, 7489–7495. doi: 10.1016/j.vaccine.2015.09.060
- Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L., et al. (2013). FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol. Biol. Evol.* 30, 1196–1205. doi: 10.1093/molbev/mst030
- Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., and Kosakovsky Pond, S. L. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* 8:e1002764. doi: 10.1371/journal.pgen.1002764
- Nazeri, S., Zakeri, S., Mehrizi, A. A., and Djadid, N. D. (2017). Naturally acquired immune responses to thrombospondin-related adhesion protein (TRAP) of *Plasmodium vivax* in patients from areas of unstable malaria transmission. *Acta Trop.* 173, 45–54. doi: 10.1016/j.actatropica.2017.05.026
- Neafsey, D. E., Galinsky, K., Jiang, R. H., Young, L., Sykes, S. M., Saif, S., et al. (2012). The malaria parasite *Plasmodium vivax* exhibits greater genetic diversity than *Plasmodium falciparum*. *Nat. Genet.* 44, 1046–1050. doi: 10.1038/ng.2373
- Nguyen, T. V., Fujioka, H., Kang, A. S., Rogers, W. O., Fidock, D. A., and James, A. A. (2001). Stage-dependent localization of a novel gene product of the malaria parasite, *Plasmodium falciparum*. *J. Biol. Chem.* 276, 26724–26731. doi: 10.1074/jbc.M103375200
- Nguyen, T. V., Sacci, J. B. Jr., de la Vega, P., John, C. C., James, A. A., and Kang, A. S. (2009). Characterization of immunoglobulin G antibodies to *Plasmodium falciparum* sporozoite surface antigen MB2 in malaria exposed individuals. *Malar. J.* 8:235. doi: 10.1186/1475-2875-8-235
- Nielsen, H. (2017). Predicting secretory proteins with signalP. *Methods Mol. Biol.* 1611, 59–73. doi: 10.1007/978-1-4939-7015-5_6
- Novembre, J. A. (2002). Accounting for background nucleotide composition when measuring codon usage bias. *Mol. Biol. Evol.* 19, 1390–1394. doi: 10.1093/oxfordjournals.molbev.a004201
- Ohashi, J., Suzuki, Y., Naka, I., Hananantachai, H., and Patarapotikul, J. (2014). Diversifying selection on the thrombospondin-related adhesive protein (TRAP) gene of *Plasmodium falciparum* in Thailand. *PLoS ONE* 9:e90522. doi: 10.1371/journal.pone.0090522
- Pacheco, M. A., Elango, A. P., Rahman, A. A., Fisher, D., Collins, W. E., Barnwell, J. W., et al. (2012). Evidence of purifying selection on merozoite surface protein 8 (MSP8) and 10 (MSP10) in *Plasmodium* spp. *Infect. Genet. Evol.* 12, 978–986. doi: 10.1016/j.meegid.2012.02.009
- Palaya, V., Lau, Y. L., Mahmud, R., Chen, Y., and Fong, M. Y. (2013). Cloning, expression, and immunocharacterization of surface protein containing an altered thrombospondin repeat domain (SPATR) from *Plasmodium knowlesi*. *Malar. J.* 12:182. doi: 10.1186/1475-2875-12-182
- Patarroyo, M. A., Calderón, D., and Moreno-Pérez, D. A. (2012). Vaccines against *Plasmodium vivax*: a research challenge. *Expert Rev. Vaccines* 11, 1249–1260. doi: 10.1586/erv.12.91
- Patarroyo, M. E., Arévalo-Pinzón, G., Reyes, C., Moreno-Vranich, A., and Patarroyo, M. A. (2016). Malaria parasite survival depends on conserved binding peptides’ critical biological functions. *Curr. Issues Mol. Biol.* 18, 57–78. doi: 10.21775/cimb.018.057
- Pierleoni, A., Martelli, P. L., Fariselli, P., and Casadio, R. (2006). BaCelLo: a balanced subcellular localization predictor. *Bioinformatics* 22, e408–e416. doi: 10.1093/bioinformatics/btl222
- Pond, S. L., Frost, S. D., and Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21, 676–679. doi: 10.1093/bioinformatics/bti079
- Posada, D. (2008). jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25, 1253–1256. doi: 10.1093/molbev/msn083

- Premaratne, P. H., Aravinda, B. R., Escalante, A. A., and Udagama, P. V. (2011). Genetic diversity of *Plasmodium vivax* Duffy Binding Protein II (PvDBP-II) under unstable transmission and low intensity malaria in Sri Lanka. *Infect. Genet. Evol.* 11, 1327–1339. doi: 10.1016/j.meegid.2011.04.023
- Price, R. N., Douglas, N. M., and Anstey, N. M. (2009). New developments in *Plasmodium vivax* malaria: severe disease and the rise of chloroquine resistance. *Curr. Opin. Infect. Dis.* 22, 430–435. doi: 10.1097/QCO.0b013e32832f14c1
- Putaporntip, C., Jongwutiwes, S., Ferreira, M. U., Kanbara, H., Udomsangpetch, R., and Cui, L. (2009). Limited global diversity of the *Plasmodium vivax* merozoite surface protein 4 gene. *Infect. Genet. Evol.* 9, 821–826. doi: 10.1016/j.meegid.2009.04.017
- Putaporntip, C., Jongwutiwes, S., Tanabe, K., and Thaithong, S. (1997). Interallelic recombination in the merozoite surface protein 1 (MSP-1) gene of *Plasmodium vivax* from Thai isolates. *Mol. Biochem. Parasitol.* 84, 49–56. doi: 10.1016/S0166-6851(96)02786-7
- Putaporntip, C., Udomsangpetch, R., Pattanawong, U., Cui, L., and Jongwutiwes, S. (2010). Genetic diversity of the *Plasmodium vivax* merozoite surface protein-5 locus from diverse geographic origins. *Gene* 456, 24–35. doi: 10.1016/j.gene.2010.02.007
- Rich, S. M., and Ayala, F. J. (2003). Progress in malaria research: the case for phylogenetics. *Adv. Parasitol.* 54, 255–280. doi: 10.1016/S0065-308X(03)54005-2
- Richie, T. L., and Saul, A. (2002). Progress and challenges for malaria vaccines. *Nature* 415, 694–701. doi: 10.1038/415694a
- Risco-Castillo, V., Topçu, S., Marinach, C., Manzoni, G., Bigorgne, A. E., Briquet, S., et al. (2015). Malaria Sporozoites Traverse Host Cells within Transient Vacuoles. *Cell Host Microbe* 18, 593–603. doi: 10.1016/j.chom.2015.10.006
- Rodrigues-da-Silva, R. N., Soares, I. F., Lopez-Camacho, C., Martins da Silva, J. H., Perce-da-Silva, D. S., Têva, A., et al. (2017). *Plasmodium vivax* cell-traversal protein for ookinetes and sporozoites: naturally acquired humoral immune response and B-cell epitope mapping in Brazilian Amazon inhabitants. *Front. Immunol.* 8:77. doi: 10.3389/fimmu.2017.00077
- Romero, L. C., Nguyen, T. V., Deville, B., Ogunjumo, O., and James, A. A. (2004). The MB2 gene family of *Plasmodium* species has a unique combination of S1 and GTP-binding domains. *BMC Bioinformatics* 5:83. doi: 10.1186/1471-2105-5-83
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029
- Rosado, C. J., Kondos, S., Bull, T. E., Kuiper, M. J., Law, R. H., Buckle, A. M., et al. (2008). The MACPF/CDC family of pore-forming toxins. *Cell. Microbiol.* 10, 1765–1774. doi: 10.1111/j.1462-5822.2008.01191.x
- Rozas, J., Gullaud, M., Blandin, G., and Aguadé, M. (2001). DNA variation at the rp49 gene region of *Drosophila simulans*: evolutionary inferences from an unusual haplotype structure. *Genetics* 158, 1147–1155.
- Sawai, H., Otani, H., Arisue, N., Palacpac, N., de Oliveira Martins, L., Pathirana, S., et al. (2010). Lineage-specific positive selection at the merozoite surface protein 1 (msp1) locus of *Plasmodium vivax* and related simian malaria parasites. *BMC Evol. Biol.* 10:52. doi: 10.1186/1471-2148-10-52
- Shen, H. M., Chen, S. B., Wang, Y., Xu, B., Abe, E. M., and Chen, J. H. (2017). Genome-wide scans for the identification of *Plasmodium vivax* genes under positive selection. *Malar. J.* 16:238. doi: 10.1186/s12936-017-1882-0
- Siau, A., Silvie, O., Franetich, J. F., Yalaoui, S., Marinach, C., Hannoun, L., et al. (2008). Temperature shift and host cell contact up-regulate sporozoite expression of *Plasmodium falciparum* genes involved in hepatocyte infection. *PLoS Pathog.* 4:e1000121. doi: 10.1371/journal.ppat.1000121
- Sonnhammer, E. L., von Heijne, G., and Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 6, 175–182.
- Sultan, A. A., Thathy, V., Frevert, U., Robson, K. J., Crisanti, A., Nussenzweig, V., et al. (1997). TRAP is necessary for gliding motility and infectivity of plasmodium sporozoites. *Cell* 90, 511–522. doi: 10.1016/S0092-8674(00)80511-5
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595.
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Taylor, J. E., Pacheco, M. A., Bacon, D. J., Beg, M. A., Machado, R. L., Fairhurst, R. M., et al. (2013). The evolutionary history of *Plasmodium vivax* as inferred from mitochondrial genomes: parasite genetic diversity in the Americas. *Mol. Biol. Evol.* 30, 2050–2064. doi: 10.1093/molbev/mst104
- Uddin, A. (2017). Codon usage bias: a tool for understanding molecular evolution. *J. Proteomics Bioinform.* 10:e32. doi: 10.4172/jpb.1000e32
- VanBuskirk, K. M., Sevova, E., and Adams, J. H. (2004). Conserved residues in the *Plasmodium vivax* Duffy-binding protein ligand domain are critical for erythrocyte receptor recognition. *Proc. Natl. Acad. Sci. U.S.A.* 101, 15754–15759. doi: 10.1073/pnas.0405421101
- Weiss, G. E., Gilson, P. R., Taechalertrapisarn, T., Tham, W. H., de Jong, N. W., Harvey, K. L., et al. (2015). Revealing the sequence and resulting cellular morphology of receptor-ligand interactions during *Plasmodium falciparum* invasion of erythrocytes. *PLoS Pathog.* 11:e1004670. doi: 10.1371/journal.ppat.1004670
- Wengelnik, K., Spaccapelo, R., Naitza, S., Robson, K. J., Janse, C. J., Bistoni, F., et al. (1999). The A-domain and the thrombospondin-related motif of *Plasmodium falciparum* TRAP are implicated in the invasion process of mosquito salivary glands. *EMBO J.* 18, 5195–5204. doi: 10.1093/emboj/18.19.5195
- Winter, D. J., Pacheco, M. A., Vallejo, A. F., Schwartz, R. S., Arevalo-Herrera, M., Herrera, S., et al. (2015). Whole genome sequencing of field isolates reveals extensive genetic diversity in *Plasmodium vivax* from Colombia. *PLoS Negl. Trop. Dis.* 9:e0004252. doi: 10.1371/journal.pntd.0004252
- Yang, A. S. P., O'Neill, M. T., Jennison, C., Lopaticki, S., Allison, C. C., Armistead, J. S., et al. (2017a). Cell traversal activity is important for *Plasmodium falciparum* liver infection in humanized mice. *Cell Rep.* 18, 3105–3116. doi: 10.1016/j.celrep.2017.03.017
- Yang, A. S. P., Lopaticki, S., O'Neill, M. T., Erickson, S. M., Douglas, D. N., Kneteman, N. M., et al. (2017b). AMA1 and MAEBL are important for *Plasmodium falciparum* sporozoite infection of the liver. *Cell. Microbiol.* 19:e12745. doi: 10.1111/cmi.12745
- Zhang, J., Rosenberg, H. F., and Nei, M. (1998). Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. U.S.A.* 95, 3708–3713. doi: 10.1073/pnas.95.7.3708
- Ziegler, T., Abegg, C., Meijaard, E., Perwitasari-Farajallah, D., Walter, L., Hodges, J. K., et al. (2007). Molecular phylogeny and evolutionary history of Southeast Asian macaques forming the *M. silenus* group. *Mol. Phylogenet. Evol.* 42, 807–816. doi: 10.1016/j.ympev.2006.11.015

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Garzón-Ospina, Buitrago, Ramos and Patarroyo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.