ON THE EVOLUTION AND FUNCTION OF PLASMODIUM VIVAX RETICULOCYTE BINDING SURFACE ANTIGEN (PVRBSA)

PAOLA ANDREA CAMARGO-AYALA1,2, DIEGO GARZÓN-OSPINA1,3, DARWIN ANDRÉS MORENO-PÉREZ1,4, LAURA ALEJANDRA RICAURTE-CONTRERAS1, OSCAR NOYA5 AND MANUEL A. PATARROYO1,6*

1 Department of Molecular Biology and Immunology, Fundación Instituto de Inmunología de Colombia (FIDIC), Bogotá, Colombia; 2 Microbiology Postgraduate Programme, Universidad Nacional de Colombia, Bogotá, Colombia; 3 PhD Programme in Biomedical and Biological Sciences, Universidad del Rosario, Bogotá, Colombia; 4 Livestock Sciences Faculty, Universidad de Ciencias Aplicadas y Ambientales, Bogotá, Colombia; 5 Instituto de Medicina Tropical, Facultad de Medicina, Universidad Central de Venezuela, Caracas, Venezuela; 6 School of Medicine and Health Sciences, Universidad del Rosario, Bogotá, Colombia

The RBSA protein is encoded by a gene described in Plasmodium species having tropism for reticulocytes. Since this protein is antigenic in natural infections and can bind to target cells, it has been proposed as a potential candidate for an anti-Plasmodium vivax vaccine. However, genetic diversity (a challenge which must be overcome for ensuring fully effective vaccine design) has not been described at this locus. Likewise, the minimum regions mediating specific parasite-host interaction have not been determined. This is why the rbsa gene’s evolutionary history is being here described, as well as the P. vivax rbsa (pvrbsa) genetic diversity and the specific regions mediating parasite adhesion to reticulocytes. Unlike what has previously been reported, rbsa was also present in several parasite species belonging to the monkey-malaria clade; paralogs were also found in Plasmodium parasites invading reticulocytes. The pvrbsa locus had less diversity than other merozoite surface proteins where natural selection and recombination were the main evolutionary forces involved in causing the observed polymorphism. The N-terminal end (PvRBSA-A) was conserved and under functional constraint; consequently, it was expressed as recombinant protein for binding assays. This protein fragment bound to reticulocytes whilst the C-terminus, included in recombinant PvRBSA-B (which was not under functional constraint), did not. Interestingly, two PvRBSA-A-derived peptides were able to inhibit protein binding to reticulocytes. Specific conserved and functionally important peptides within PvRBSA-A could thus be considered when designing a fully-effective vaccine against P. vivax.

Keywords: Plasmodium vivax, rbsa, genetic diversity, evolutionary forces, protein binding, parasite-host interaction, antimalarial vaccine

Abbreviations: ω, omega rate (dN/dS and/or Kd/Ks); BY, Bayesian method; Chr, chromosome; LD, linkage disequilibrium; LRR, leucine-rich region; ML, maximum likelihood; Mrz, merozoite; RBSA, reticulocyte binding surface antigen; rRBSA, recombinant RBSA protein; WB, Western blot.
INTRODUCTION

*Plasmodium vivax* is a parasite which emerged in Asia (Escalante et al., 2005; Carlton et al., 2013) (although an African origin is also likely, Liu et al., 2014), involving a host-switch from monkeys to humans (Mu et al., 2005) around 1.3–2.9 million years ago (Pacheco et al., 2011), *Plasmodium cynomolgi* being the most phylogenetically related species (Mu et al., 2005; Pacheco et al., 2011; Tachibana et al., 2012). *P. vivax* then reached countries on the 5 continents through human migration (Rodrigues et al., 2018), currently predominating in Asia and America (Guerra et al., 2010; Getting et al., 2012). This species is considered the second most important human-malaria parasite worldwide, due to the morbidity it causes (WHO, 2017).

*Plasmodium vivax* incidence has decreased since 2010. An estimated 6–11 million cases were attributed to this parasite in 2016 (around 7 million less than in 2010) (WHO, 2017). Despite that, the social and economic burden of malaria in endemic countries (Suh et al., 2004) could still be huge. Even though control measures against this parasite have been shown to be useful, *P. vivax* elimination (and/or malarial elimination) is not an easy task. The most relevant challenges for eliminating malaria concern the social and economic conditions of the places most affected by this illness, the social and political conflicts in several endemic areas and the anomalous weather patterns (WHO, 2017). These, together with the spread of insecticide-resistant mosquitoes and drug-resistant parasites, could bring about a recurrence of this disease (Maxmen, 2012; Price et al., 2014; Huijben and Paaijmans, 2018). Consequently, new interventions must be designed which can reduce the parasite reservoir, limiting the time that a human (or mosquito) host is infectious (Barry and Arnott, 2014). This goal could be achieved by developing a fully effective antimalarial vaccine which, together with existing control measures, could contribute towards a malaria-free world (Barry and Arnott, 2014; White et al., 2014).

Malaria-related vaccine research and development efforts were exclusively focused on *Plasmodium falciparum* for a long time (Galinski and Barnwell, 2008). During the last few decades, *P. vivax* research has been increased and more and more groups are studying approaches to an anti-*P. vivax* vaccine (Galinski and Barnwell, 2008; Chen et al., 2010; Valencia et al., 2011; The malERA Consultative Group on Vaccines, 2011; Arnott et al., 2012; Patarroyo et al., 2012; Nanda Kumar et al., 2016). Initial research was based on the knowledge acquired regarding *P. falciparum* (Patarroyo et al., 2012); however, given *P. vivax*’s unique attributes (such as hypnozoite forms in the liver, invasion preference for reticulocytes or the rapid gametocyte formation) (Galinski et al., 2013) and since *P. falciparum* and *P. vivax* have had different evolutionary paths, research regarding *P. vivax*-exclusive antigens could be relevant. For instance, the Duffy antigen, which is not present in *P. falciparum*, is essential for *P. vivax* invasion (Chitnis and Miller, 1994). PfDBP-conserved epitopes can trigger strong neutralizing antibodies; it has therefore been suggested as one of the main *P. vivax* vaccine candidates (Chen et al., 2016; Ntumngia et al., 2017).

One antigen shared just by *P. vivax* and *P. cynomolgi* (both having reticulocyte predilection) has been characterized recently (Moreno-Perez et al., 2017). This antigen (named reticulocyte binding surface antigen – RBSA) is encoded by a two-exon gene (Moreno-Perez et al., 2017) located on Chr 3. The protein product has a signal peptide and two transmembrane helices and is located on mature Mrz membrane. This protein seems to be involved in Mrz invasion, since recombinant *P. vivax* RBSA (rPvRBSA) binds to a subpopulation of immature reticulocytes having a Duffy positive phenotype; it is also recognized by the human immune system in natural infections. It has thus been suggested as a putative vaccine candidate for anti-*P. vivax* malaria vaccine development (Moreno-Perez et al., 2017).

However, since parasite’s high genetic diversity is one of the challenges to be overcome during *P. vivax* vaccine design (Neafsey et al., 2012); the antigens’ genetic diversity must therefore be assessed (Arnott et al., 2012; Barry and Arnott, 2014) for selecting those having limited polymorphism or the conserved functional regions within them (Richie and Saul, 2002). *pvrbsa* genetic diversity has not been assessed to date and, although PVRBSA binds to reticulocytes (Moreno-Perez et al., 2017), the regions involved in Mrz-host specific interactions have not been established yet.

PvRBSA is an Mrz membrane protein and is recognized by the human immune system in natural infection; it could therefore be a highly polymorphic antigen. Nevertheless, functionally important parts of the protein (i.e., regions involved in host-parasite interaction) should be functionally constrained, being maintained by negative selection as highly conserved within and between species (Graur et al., 2013; Garzon-Ospina et al., 2015). Consequently, inferring this kind of selection on *pvrbsa* could be used to predict regions under functional constraint which then might be used as putative vaccine candidates. This study evaluated *pvrbsa* genetic diversity, assessed the evolutionary forces involved in causing the observed polymorphism and established minimum regions involved in protein–cell interaction.

MATERIALS AND METHODS

Ethics Approval and Consent to Participate

All *P. vivax*-infected patients who provided us with the blood samples were informed about the purpose of the study and all gave their written consent. Regarding newborn umbilical cord blood samples, the progenitors signed an informed consent form after having received detailed information regarding the study’s goals. All procedures carried out in this study were approved by the ethics committees of the Fundación Instituto de Inmunología de Colombia (IRB number: ACTA N° 037-CEEPAA), the Universidad del Rosario (IRB number: CEI-ABN026-0001061) and the Bioethics’ committee from the Instituto de Medicina Tropical “Dr. Félix Piñano” at the Universidad Central de Venezuela (IRB number: CEC-IMT11/2018).

Parasite DNA

One hundred and sixty-seven peripheral blood samples and/or blood drops collected on FTA cards from patients proving
positive for *P. vivax* malaria by microscope examination were collected in different Colombian (Chocó, *n* = 37, Córdoba, *n* = 39, and Amazonas, *n* = 41) and Venezuelan (Bolivar, *n* = 29 and Venezuela’s coastal area, *n* = 19) endemic areas between 2010 and 2016 (Supplementary Figure S1). A Wizard Genomic DNA Purification kit (Promega) was used for obtaining DNA from Chocó and Córdoba samples, following the manufacturer's instructions. A Pure Link Genomic DNA mini kit (Invitrogen) was used for extracting DNA from Amazon samples from each drop of blood collected on the FTA cards, according to the manufacturer's specifications whilst blood were extracted from Venezuelan samples by salting-out, following Welsh and Bunce modifications (Welsh and Bunce, 1999). The recovered DNAs were stored at −20°C until use. All samples had a single *P. vivax* msp3 allele, suggesting that these samples came from single *P. vivax* infection.

**Identifying the rbsa Gene in Plasmodium Monkey Lineage**

A Blast search (using the 1,506 bp from Sal-I-*P. vivax* rbsa gene sequence as query) was carried out using available whole genome sequences from *C. cymomolgi* (GenBank accession number: GCA_000321355.1), *P. inui* (GenBank accession number: GCA_000524495.1), *P. knowlesi* (GenBank accession number: GCA_000063555.1), *P. coatneyi* (GenBank accession number: GCA_000725905.1) and *P. fragile* (GenBank accession number: GCA_000956335.1) to identify orthologous in these Plasmodium species. The MUSCLE method (Edgar, 2004) was used for aligning the recovering contigs having high identity sequences (>60). The best model for DNA substitutions was then selected using the JModelTest v.2.1.3 (Posada, 2008) with the Akaike's information criterion and MEGA v.6 software (Tamura et al., 2013). Phylogenetic trees were inferred through ML and Bayesian (BY) methods, using the GTR and/or the TVM models (selected as the best models). MEGA v.6 software was used for ML analysis and topology reliability was evaluated by bootstrap, using 1,000 iterations. Bayesian phylogenetic analysis was conducted with MrBayes v.3.2 software (Ronquist et al., 2012) and the analysis was run until reaching a lower than 0.01 standard deviation of split frequencies value; sump and sumt commands were then used for tabulating posterior probabilities and building a consensus tree.

**PCR Amplification and Sequencing**

One hundred and sixty-seven parasite DNAs from endemic Colombian (*n* = 117) and Venezuelan (*n* = 50) regions were used for amplifying the *pvrbsa* locus by nested-PCR, as follows. The first reaction involved 3.3 μL of ultrapure water, 5.4 μL of KAPA HiFi HotStart ReadyMix, 0.3 μM of each primer (rbsafwd: 5′-TTTATTTCTATTGAGCTTGAACCTTG-3′ and rbsarev: 5′-TGAAGGTTGTGTGGATCTCC-3′) and 1 μL of DNA. The PCR was carried out as follows: one 5 min step at 95°C, a second 35 cycle step for 20 s at 98°C, 15 s at 57°C and 45 s at 72°C, followed by a final step of 10 min at 72°C. Two μL of the first PCR product were added to the second PCR reaction containing 7.5 μL of ultrapure water, 12.5 μL of KAPA HiFi HotStart Readymix and 0.3 μM of each primer (rbsafwd: 5′-GAAATACAGATGAAAGGGTTATG-3′ and rbsarev: 5′-GATTCCCACTCGGTTTATTCC-3′). The thermal conditions were the same as those used in the first PCR reaction. *PvRBSA* fragments towards the amino and carboxyl (*PvRBSA*-A and *PvRBSA*-B, respectively) encoding regions were amplified using KAPA-HiFi HotStart ReadyMix (KAPA Biosystems) and the genomic DNA previously extracted from the Vivax Colombia Guaviare-I (VCG-I) strain (Moreno-Perez et al., 2014). The 25 μL PCR reaction contained 7.5 μL ultrapure water, 12.5 μL enzyme, 0.3 μM primer (designed taking into account the natural selection signatures observed for *PvRBSA*). *PvRBSA*-A: rbsa-a-fwd: 5′-GGGGTACACAGCAATGAGTGTCTCTG-3′ and rbsa-a-rev: 5′-CCCTCGAGTCATACATTCTCCACACAATTAA-3′; *PvRBSA*-B: rbsb-b-fwd: 5′-GGGGTGACCAATATGATAGTCTGGAG-3′ and rbsb-b-rev: 5′-CCCTCGAGCAAATCGGA-3′. Singleton SNPs were confirmed by an independent PCR amplification. The sequences obtained here were deposited in the GenBank database (accession numbers MH391806-MH391972).

**Obtaining Recombinant Plasmids With RBSA Fragments**

*pvrbsa*-A and *pvrbsa*-B PCR products were digested with KpnI (New England Biolabs) and Aval (NEB) enzymes and then ligated into the pET32b+ vector using T4 ligase (NEB). Briefly, 0.5 μg of each purified product, 1x of CutSmart buffer and 1 U/μL of each enzyme were used in a 25 μL reaction which was incubated for 1 h at 37°C and then inactivated at 80°C for 20 min. Ligation to pET32b+ vector was performed in a 20 μL volume containing 1x T4 buffer, 30 U/μL T4 ligase and vector/product in a 1:3 ratio. The reaction was incubated at 16°C for 16 h and then inactivated at 65°C for 20 min. Each recombinant plasmid was transformed into *Escherichia coli* JM109 cells (Invitrogen) according to the manufacturer's recommendations and recombinant colonies were then confirmed by PCR using the primers from each product. Three positive colonies were used for extracting plasmids with an UltraClean 6 Minute Mini Plasmid Prep kit (MOBIO), following the manufacturer's recommendations and then sequenced bidirectionally using pet32b-fwd: 5′-GGGTGAAGTGCGCGGCAA-3′ and pet32-Rev: 5′-CCGGGGTTATTAGCTAGT-3′ primers.

**pvrbsa Gene DNA Diversity and Evolutionary Analysis**

Three chromatograms (forward, reverse, and internal primer) were obtained from Colombian and Venezuelan samples from
sequencing; they were assembled using CLC Main workbench v.3 software (CLC bio, Cambridge, MA, United States). Colombian and Venezuelan sequences were compared and aligned [using the MUSCLE method (Edgar, 2004)] with reference strain sequences (Sal-I GenBank accession number: AAKM01000020.1, Brazil-I GenBank accession number: AFMK01000195.1/AFMK01000194.1, India-VII GenBank accession number: AFBK01001271.1 and North Korean GenBank accession number: AFNJ01003313.1) (Carlton et al., 2008; Neafsey et al., 2012) and with sequences obtained from several sequencing projects (Chan et al., 2012; Hester et al., 2013; Hupalo et al., 2016) available in PlasmoDB database (sequences were screened to rule out those having missing data or ambiguous nucleotides); 232 sequences from different regions around the world were thus used. Afterwards the intron was ruled out from all sequences and codon alignments were inferred using the TranslatorX web server (Abascal et al., 2010). This alignment was manually edited (Supplementary Data Sheet 1A) to ensure correct repeat alignment for further analysis.

DnaSP v.5 software (Librado and Rozas, 2009) was used for calculating the amount of singleton sites (s), the amount of parsimony-informative sites (Ps), the amount of haplotypes (H), the haplotype diversity (Hd), the nucleotide polymorphism (or Watterson estimator, θw) and the nucleotide diversity per site (π) for all available sequences (worldwide samples), as well as for the Colombian and Venezuelan populations and for the subpopulations within both populations. Departure from the neutral model of molecular evolution was assessed by Tajima (1989); Fu and Li (1993), and Fay and Wu (2000) estimators. These are frequency spectrum-based tests for comparing two estimators of the population mutation parameter θ which characterizes mutation–drift equilibrium (neutral model). Under neutrality, several unbiased estimators of θ should be equal. Rejection of the neutral expectations suggests that selection or a demographic process could be taking place, the Fay and Wu test being suitable for detecting selective sweep (Fay and Wu, 2000). On the other hand, since each new polymorphic site has a high probability of delineating a new haplotype (Depaulis and Veuille, 1998), tests based on haplotype distribution have been developed [K- and Hd-test (Depaulis and Veuille, 1998) as well as Fu’s Fs (Fu, 1997)]. Similar to frequency spectrum-based tests, departures from neutral expectation could be the consequence of selection or demographic history, Fu’s Fs estimator being a more sensitive indicator of population expansion than Tajima’s test (Fu, 1997). DnaSP v.5 and/or ALLELIX software were used for these tests, coalescent simulations being used for obtaining confidence intervals (Librado and Rozas, 2009). Sites containing gaps or repeats in the alignment were not taken into account.

The aforementioned methods do not consider the classes of mutations (non-synonymous and synonymous); natural selection signatures were therefore also assessed using different methods which classified mutations as non-synonymous or synonymous. The aBSREL method (Smith et al., 2015) was used to test whether positive selection had occurred on a percentage of branches regarding rbsA phylogeny. The modified Nei-Gojobori method (Zhang et al., 1998) with Jukes-Cantor correction (Jukes and Cantor, 1969) was then used for calculating the non-synonymous and synonymous substitution difference rate (dN-dS) within P. vivax, using a Z-test available in MEGA software v.6 (Tamura et al., 2013) to identify significant statistical values. Two tests were performed for assessing natural selection signals comparing different species; the McDonald–Kreitman test (McDonald and Kreitman, 1991) was calculated using a web server (Egea et al., 2008) using P. cynomolgi orthologous sequences. Likewise, non-synonymous divergence and synonymous divergence substitution difference rates (Kn-Ks) were also calculated, using the Z-test for identifying statistically significant values. Both tests were carried out taking Jukes-Cantor divergence correction into account.

A sliding window for omega (ω) rates (dN/dS and/or Kn/Ks) was then used for assessing how natural selection acts throughout the gene. The Datamonkey web server (Delport et al., 2010) was used for assessing codon sites under positive or negative selection by using codon-based ML and BYs [IFEL (Pond et al., 2006), FEL, SLAC, REL (Kosakovsky Pond and Frost, 2005), MEME (Murrell et al., 2012) and FUBAR (Murrell et al., 2013)]. A < 0.1 p-value was considered significant for IFEL, FEL, SLAC and MEME methods and a > 0.9 posterior probability for FUBAR. Since ignored recombination can bias dN/dS estimation (Anisimova et al., 2003; Arenas and Posada, 2010, 2014), the GARD method (Kosakovsky Pond et al., 2006) was considered regarding recombination before running these tests.

Linkage disequilibrium was evaluated by calculating Zns (Kelly, 1997) to assess whether recombination was/is taking place in pvrbsA; this was followed by linear regression between LD and nucleotide distances. Evidence of recombination was also assessed by the GARD method (Kosakovsky Pond et al., 2006), as well as by ZZ (Rozas et al., 2001) and RM (Hudson and Kaplan, 1985) tests. The degree of genetic differentiation amongst P. vivax malaria-endemic regions (subpopulations) regarding the pvrbsa locus was evaluated by analysis of molecular variance (AMOVA) and by computing Wright’s fixating index (FSt), using Arlequin population genetics data analysis v3.1 software (Excoffier et al., 2007). The mutational pathways giving rise to pvrbsa haplotypes, their distribution and frequencies were inferred by median Joining method, using Network v.5 software (Bandelt et al., 1999).

### Assessing Functional Regions in the PvRBSA Protein

Whole PvRBSA, as well as its A and B regions, were recombinantly expressed and purified, as previously described (Moreno-Perez et al., 2017) with minor modifications. For example, the protein was expressed for 4 h at 37°C (for complete PvRBSA) or 30°C (for PvRBSA-A and PvRBSA-B regions) adding 0.2 mM IPTG. The cell pellet was homogenized in BI buffer (20 mM Tris-Cl, 500 mM NaCl and 1 mM EDTA, pH 8.5) during native protein extraction and then lysed by sonication on ice for 3 min 30 s, with 0.2 s pulses ON, 0.2 s OFF, at 40% amplitude. Protein purification and cell binding assays involved using the same protocols described for PvGaMA and PvRBSA proteins, including PvDBP-RII as positive control and PvDBP-RII/IV as negative control (Baquero et al., 2017; Moreno-Perez et al., 2017). The competition assay was performed by pre-incubating...
Aligning *P. cynomolgi* *rbsa* (*pcrbsa*, Ch3 GenBank accession number: BAEJ01000151.1) and *pcrbsa* paralogue (*pcrbspa*, Ch10 GenBank accession number: BAEJ01000874.1) DNA fragments revealed a 10 Mb DNA region between chr3 and ch10 (Figure 1, Supplementary Data Sheet 1D and Supplementary Figure S2), having 98.59% identity.

### pvrbsa Gene PCR Amplification and Sequencing

One hundred and sixty-seven parasite DNAs from endemic Colombian and Venezuelan regions were used for amplifying the *pvrbsa* locus. The amplicons had 1,454–1,580 bp sizes. These PCR fragments were then purified and sequenced. At least three chromatograms were assembled for each sample, thereby obtaining a consensus sequence.

### pvrbsa Gene DNA Diversity and Evolutionary Analysis

Together with the 167 sequences obtained here, 65 additional sequences recovered from PlasmoDB (Colombia *n* = 18, Peru *n* = 16, Brazil *n* = 3, Mexico *n* = 5, Madagascar *n* = 2, China *n* = 2, Cambodia *n* = 2, Thailand *n* = 8, Papua New Guinea *n* = 6, India *n* = 1, North Korea *n* = 1 and the Sal-I sequences, Supplementary Data Sheet 1A) were used to assess genetic diversity in *pvrbsa*. The *pvrbsa* encoding region had 44 segregating sites whilst just 3 polymorphic sites were found at intron worldwide. *pvrbsa* had an intermediate diversity according to the nucleotide diversity estimator (*π*) (Table 1, *π* < 0.01). The Venezuelan parasite population had a higher *π* than the Colombian population; both subpopulations within Venezuela (Bolivar and Venezuela's Coastal area) had similar *π* values. Regarding Colombian subpopulations, parasites from Meta (sequences available in PlasmoDB) had the lowest diversity, whilst Córdoba had the highest diversity value for Colombian subpopulations (Table 1). The 232 sequences could be clustered into 80 different haplotypes (70 regarding CDS, Supplementary Data Sheet 1E). The Colombian population had 37 haplotypes whilst 25 haplotypes were observed in Venezuela (Table 1).

Since different evolutionary forces (drift, selection, recombination and migration) could determine the genetic diversity pattern (Casillas and Barbadilla, 2017) observed in natural populations, several evolutionary tests were performed to infer which of them were modulating the *pvrbsa* gene diversity observed here. Tests based on a neutral molecular evolutionary model were performed for population and subpopulations to assess drift/selection force. The Colombia population had negative statistically significant values just for the Fay and Wu test (Table 2) indicating a possible selective sweep. When subpopulations were analyzed separately, Meta had positive values for the Fu and Li test, being statistically significant (*p* < 0.05) which could have resulted from balancing selection or a decrease in population. The Chocó subpopulation had a value lower than 0 (*p* < 0.03) for Fay and Wu's *H* test (Table 2). A sliding window for frequency spectrum-based tests showed regions inside *pvrbsa* having significant values (Supplementary Figure S3). On the other hand, the Fu and Li's *F* estimator gave

### RESULTS

**Identifying the rbsa Gene in *Plasmodium* Monkey Lineage**

The Sal-I-*P. vivax* *rbsa* gene (*pvrbsa*) sequence was used for a Blast search of *P. vivax* phylogenetically related species to identify orthologs in the monkey-parasite lineage. A contig was found in *P. inui* and *P. fragile* having higher than 60% identity whilst two different contigs were found in *P. cynomolgi* having 80% identity with Sal-I *pvrbsa* (Supplementary Data Sheet 1B). Interestingly, both *P. cynomolgi* contigs had the same identity but they did not belong to the same chr; one of the two fragments occurred at the chr3 and the other one at chr10. There was 99.57% identity between these *P. cynomolgi* DNA fragments.

A similar Blast search using Sal-I *pvrbsa* sequence was then performed against other *P. vivax* reference stains’ whole genome sequences (Neafsey et al., 2012), revealing extra DNA fragments having 63% identity at the same chr, lacking the first exon and the intron sequences (Supplementary Data Sheet 1C), probably derived from duplication. The duplicated fragment was around 4,000 bp (Figure 1, Supplementary Data Sheet 1C and Supplementary Figure S2), taking upstream and downstream DNA regions into account (Sal-I Ch3 GenBank accession number: AAKM01000020.1) from *pcrbsa* and the *pvrbsa* paralog (*pcrbspa*). Counterparts to this 4,000 bp-DNA fragment were also identified in *P. cynomolgi* (at both chr), *P. inui* and *P. fragile*, which were aligned and used to infer a phylegetic tree (Supplementary Figure S2). Three well-supported groups were identified in the tree; the first one placed *P. vivax* sequences together, the second group clustered *P. cynomolgi* sequences and in the third, *P. inui* and *P. fragile* sequences were found (Supplementary Figure S2).
FIGURE 1 | Phylogenetic tree mapping the presence/absence of rbsa locus and duplication events. The phylogenetic tree represents Plasmodium evolutionary history. The branches are colored regarding their vertebrate host. The rbsa locus was found in parasites from monkey-malaria clade, but it was not clear whether the rbsa gene became lost from the other clades or arose exclusively in the monkey-malaria clade. rbsa became lost in P. knowlesi and P. coatneyi whilst duplication events occurred in P. vivax and P. cynomolgi. Models of putative ancestral chromosome 3 are displayed, indicating the arrangement after the duplication event (chr derived). PlasmoDB genome fragment annotation is also shown.

TABLE 1 | Genetic diversity estimators.

|          | n  | Aligned length | S  | Ps | H       | θw (sd)         | π  (sd)      | Hd  (sd)  |
|----------|----|----------------|----|----|---------|----------------|-------------|----------|
| Worldwide isolates | 232 | Gene | 1,354 | 15 | 32 | 80 | 0.0067 (0.0010) | 0.0085 (0.0002) | 0.970 (0.005) |
|          |    | CDS            | 1,113 | 14 | 30 | 70 | 0.0066 (0.0010) | 0.0080 (0.0002) | 0.962 (0.006) |
| Colombia | 135 | Gene | 1,354 | 13 | 26 | 37 | 0.0061 (0.0010) | 0.0079 (0.0002) | 0.951 (0.009) |
|          |    | CDS            | 1,113 | 12 | 24 | 32 | 0.0059 (0.0010) | 0.0074 (0.0002) | 0.937 (0.010) |
| Amazonas | 41  | Gene | 1,391 | 6  | 20 | 13 | 0.0050 (0.0010) | 0.0072 (0.0004) | 0.890 (0.028) |
|          |    | CDS            | 1,170 | 6  | 18 | 13 | 0.0049 (0.0010) | 0.0067 (0.0004) | 0.890 (0.028) |
| Chocó    | 37  | Gene | 1,354 | 14 | 20 | 18 | 0.0072 (0.0012) | 0.0079 (0.0005) | 0.964 (0.024) |
|          |    | CDS            | 1,113 | 13 | 18 | 16 | 0.0069 (0.0012) | 0.0074 (0.0005) | 0.936 (0.024) |
| Córdoba  | 39  | Gene | 1,405 | 3  | 24 | 19 | 0.0083 (0.0010) | 0.0080 (0.0006) | 0.966 (0.023) |
|          |    | CDS            | 1,170 | 3  | 22 | 17 | 0.0051 (0.0010) | 0.0076 (0.0006) | 0.944 (0.024) |
| Meta     | 18  | Gene | 1,506 | 0  | 16 | 8  | 0.0036 (0.0009) | 0.0055 (0.0003) | 0.961 (0.033) |
|          |    | CDS            | 1,290 | 0  | 14 | 8  | 0.0034 (0.0008) | 0.0050 (0.0003) | 0.961 (0.033) |
| Venezuela| 50  | Gene | 1,363 | 2  | 28 | 25 | 0.0056 (0.0010) | 0.0069 (0.0003) | 0.958 (0.023) |
|          |    | CDS            | 1,149 | 2  | 26 | 25 | 0.0055 (0.0010) | 0.0067 (0.0003) | 0.958 (0.023) |
| Bolivar  | 29  | Gene | 1,363 | 0  | 27 | 17 | 0.0059 (0.0011) | 0.0068 (0.0005) | 0.975 (0.027) |
|          |    | CDS            | 1,149 | 0  | 25 | 17 | 0.0057 (0.0011) | 0.0065 (0.0005) | 0.975 (0.027) |
| Coastal area | 19 | Gene | 1,363 | 6  | 22 | 13 | 0.0070 (0.0013) | 0.0094 (0.0005) | 1.000 (0.036) |
|          |    | CDS            | 1,149 | 5  | 21 | 13 | 0.0069 (0.0013) | 0.0093 (0.0005) | 1.000 (0.036) |

Genetic diversity estimators were calculated using all the pvrbsa sequences available (worldwide isolates) as well as for populations (Colombia and Venezuela) and subpopulations (Amazonas, Chocó, Córdoba, Meta, Bolivar, and Venezuela’s coastal area). n, amount of isolates analyzed; aligned length, total sites analyzed, excluding gaps; S, amount of segregating sites; Ps, amount of informative parsimonious sites; H, amount of haplotypes; θw, nucleotide polymorphism; π, nucleotide diversity per site; Hd, haplotype diversity. sd, standard deviation. Diversity estimates were corrected for sample size by multiplying by n/(n-1). Although nucleotide diversity was not affected by sample size, it was also corrected. 1Populations. 2Subpopulations.

statistically significant values for the Venezuelan population; this pattern was also observed for the Bolivar subpopulation (Table 2). The sliding window analysis revealed regions inside pvrbsa having significant values for Venezuelan subpopulations. Tests based on haplotype distribution did not have statistically significant values (Table 2).

A test based on non-synonymous and synonymous mutations was also computed to search natural selection signals in pvrbsa. The aBSREL method found evidence of episodic diversifying selection on rbsa phylogeny (Figure 2A). The dN−dS difference had positive values within P. vivax; however, they were not statistically significant, except for the Meta subpopulation (Table 3). A sliding window was then inferred to assess how dN and dS have been accumulated throughout the pvrbsa gene (Figure 2B). The 3′-end showed omega (ω) values higher than 1 (Figure 2B) which is an indicator of positive selection. The region from nucleotide 200–1,100 had ω = 0 which is expected under negative selection. Although, the McDonald–Kreitman test did not have significant values, the K_N−K_S differences (comparing P. vivax sequences to P. cynomolgi ones) had
sequences from phylogenetically related species were analyzed. A similar result was obtained when *P. vivax* and all orthologous sequences were analyzed. The tests were carried out with just the rbsa encoding region (CDS). Statistically significant values.

**TABLE 2** | Neutrality tests for Colombia, Venezuela, and subpopulations.

| n     | Tajima D (p-value) | Fu and Li D (p-value) | Fu's Fs H (p-value) | K-test (p-value) | H-test (p-value) |
|-------|--------------------|-----------------------|---------------------|-----------------|-----------------|
| CDS   | 135                | 0.637 (p > 0.1)       | −0.779 (p > 0.1)    | −12.819* (p < 0.01) | −3.447 (p > 0.1) | 32 (p > 0.05) | 0.930 (p > 0.05) |
| 41    | Amazonas          | 1.192 (p > 0.1)       | 0.262 (p > 0.1)     | −4.821 (p > 0.1)  | 1.192 (p > 0.1)  | 13 (p > 0.05) | 0.868 (p > 0.05) |
| 37    | Chocó             | 0.174 (p > 0.1)       | −0.460 (p > 0.1)    | −11.339* (p < 0.03) | −0.755 (p > 0.1) | 16 (p > 0.05) | 0.911 (p > 0.05) |
| 39    | Córdoba           | 1.424 (p > 0.1)       | 1.046 (p > 0.1)     | −5.036 (p > 0.1)  | −0.758 (p > 0.1) | 17 (p > 0.05) | 0.920 (p > 0.05) |
| 18    | Meta              | 1.856 (p > 0.1)       | 1.543* (p < 0.05)   | 2.118* (p < 0.02)  | −1.673 (p > 0.1) | 8 (p > 0.05)  | 0.908 (p > 0.05) |
| 50    | Venezuela         | 1.865 (p > 0.1)       | −5.036 (p < 0.05)   | 2.168* (p < 0.02)  | −4.140 (p > 0.1) | −3.851 (p > 0.1) | 25 (p > 0.05) | 0.939 (p > 0.05) |
| 29    | Bolivar           | 1.735 (p > 0.05)      | 1.806* (p < 0.02)   | 2.236* (p < 0.02)  | −5.197 (p > 0.1) | −1.969 (p > 0.1) | 17 (p > 0.05) | 0.941 (p > 0.05) |
| 19    | Coastal area      | 1.426 (p > 0.1)       | 0.910 (p > 0.1)     | 1.392 (p > 0.1)   | −4.520 (p > 0.1) | −1.196 (p > 0.1) | 13 (p > 0.05) | 0.947 (p > 0.05) |

The tests were carried out with just the rbsa encoding region (CDS). * Statistically significant values.

**FIGURE 2** | Identifying natural selection signatures in the rbsa gene. (A) A phylogeny tree inferred from pvrbsa CDS and its orthologous CDS was analyzed by the aBSREL method. The shade of each color on branches indicates selection strength [red indicates positive selection (ω > 1) and green negative selection (ω < 1)]. The size of each colored segment represents the percentage of selected sites in the corresponding ω class. The Figure shows the ω values, the percentage of positive selected sites (Pr [ω > ω+]) and p-values. Branches have been classified as undergoing episodic diversifying selection, by the ω rate. The ω rates within *P. vivax* (Kv/Kv) and between the species (Kv/Ks) are shown. A gene model of the rbsa gene is given below the sliding window, indicating the sites under purifying selection (green) and positive selection (red). Gene model and numbering were based on the Supplementary Data Sheet 1E alignment.

Positive selection signals (positive significant values, **Table 4**). A similar result was obtained when *P. vivax* and all orthologous sequences from phylogenetically related species were analyzed (Table 4). The sliding ω (Kv/Ks) window for this dataset showed positive selection signatures (ω > 1) towards the 5′-end, as well as in a region between nucleotide 600 – 1,572 (hereinafter called PvrBSA-B). The region between nucleotide 200 – 600 (hereinafter called PvrBSA-A) had a negative
selection signal (\(\omega < 1\), Figure 2B). Consequently, the gene was split into two regions (PvRBSA-A and PvRBSA-B) and \(d_N - d_S\) (as well as \(K_N - K_S\)) were computed again for each one. No statistically significant values were observed for \(d_N - d_S\) in either of the two regions (except for the Meta subpopulation, Table 3). Conversely, positive values (indicator of positive selection) having a \(p\)-value \(\leq 0.045\) were found in PvRBSA-B when \(K_N - K_S\) was computed; by contrast, PvRBSA-A had a negative selection signature (negative values having \(p > 0.05\), Table 4).

\(\omega\) rate was then computed for each codon using codon-based methods. Twenty-one positive selected codons were found between species whilst another 30 were under negative selection. Several of the negative selected sites were found in PvRBSA-A (Figure 2B). A protein blast (using NCBI database with Sal-I PvRBSA haplotype as query) gave two putative domains in the

### Table 3 | \(d_N - d_S\) difference within \(P. vivax\).

| Population         | PvRBSA-A (196 - 600 bp) | PvRBSA-B (601 - 1,353 bp) | Full length gene |
|---------------------|-------------------------|--------------------------|------------------|
|                     | \(d_N - d_S\) (se)      | \(d_N - d_S\) (se)      | \(d_N - d_S\) (se) |
| Worldwide isolates  | 0.0014 (0.0022)         | 0.0060 (0.0048)         | 0.0044 (0.0028)  |
|                     | \(p = 0.255\)           | \(p = 0.102\)           | \(p = 0.062\)    |
| Colombian\(^a\)    | 0.0011 (0.0002)         | 0.0050 (0.004)          | 0.0036 (0.003)   |
|                     | \(p = 0.318\)           | \(p = 0.145\)           | \(p = 0.099\)    |
| Amazonas\(^b\)    | 0.0016 (0.001)          | 0.0033 (0.005)          | 0.0027 (0.003)   |
|                     | \(p = 0.083\)           | \(p = 0.251\)           | \(p = 0.177\)    |
| Chocó\(^b\)       | 0.0023 (0.002)          | 0.0050 (0.005)          | 0.0039 (0.003)   |
|                     | \(p = 0.161\)           | \(p = 0.133\)           | \(p = 0.078\)    |
| Córdoba\(^b\)     | 0.0006 (0.004)          | 0.0036 (0.005)          | 0.0027 (0.003)   |
|                     | \(p = 0.441\)           | \(p = 0.811\)           | \(p = 0.173\)    |
| Meta\(^b\)        | −0.0024 (0.004)         | 0.0082 (0.003\(^*\))   | 0.0045 (0.002\(^*\)) |
|                     | \(p = 0.248\)           | \(p = 0.009\)           | \(p = 0.023\)    |
| Venezuela\(^a\)   | 0.0015 (0.0002)         | 0.0064 (0.005)          | 0.0046 (0.003)   |
|                     | \(p = 0.269\)           | \(p = 0.120\)           | 0.070            |
| Bolívar\(^b\)     | 0.0023 (0.0002)         | 0.0058 (0.005)          | 0.0045 (0.003)   |
|                     | \(p = 0.166\)           | \(p = 0.124\)           | \(p = 0.076\)    |
| Coastal area\(^b\)| 0.0002 (0.003)          | 0.0070 (0.006)          | 0.0047 (0.003)   |
|                     | \(p = 0.472\)           | \(p = 0.099\)           | \(p = 0.081\)    |

Non-synonymous substitution rate (\(d_N\)) and synonymous substitution rate (\(d_S\)) within \(P. vivax\). se, standard error. \(^*\)Statistically significant values. \(^a\)Populations. \(^b\)Subpopulations.

### Table 4 | \(K_N - K_S\) difference between \(Plasmodium\) species.

| Population           | RBSA-A (196 - 600 bp) | RBSA-B (601 - 1,353 bp) | Full length gene |
|----------------------|-----------------------|-------------------------|------------------|
| \(P. vivax/P. cynomolgi\) | \(K_N - K_S\) (se) | \(K_N - K_S\) (se) | \(K_N - K_S\) (se) |
| Worldwide isolates   | −0.0010 (0.003)       | 0.0080 (0.005\(^*\))   | 0.0053 (0.003\(^*\)) |
|                      | \(p = 0.372\)         | \(p = 0.045\)           | \(p = 0.033\)    |
| Colombian            | −0.0030 (0.004)       | 0.0080 (0.005\(^*\))   | 0.0053 (0.003\(^*\)) |
|                      | \(p = 0.249\)         | \(p = 0.033\)           | \(p = 0.032\)    |
| Venezuela            | −0.0092 (0.010)       | 0.0150 (0.006\(^*\))   | 0.0090 (0.004\(^*\)) |
|                      | \(p = 0.188\)         | \(p = 0.005\)           | \(p = 0.012\)    |
| \(P. vivax/Plasmodium sp.\) |           |                         |                  |
| Worldwide isolates   | −0.0031 (0.004)       | 0.0101 (0.005\(^*\))   | 0.0068 (0.003\(^*\)) |
|                      | \(p = 0.234\)         | \(p = 0.017\)           | \(p = 0.012\)    |
| Colombian            | −0.0064 (0.007)       | 0.0118 (0.005\(^*\))   | 0.0077 (0.003\(^*\)) |
|                      | \(p = 0.184\)         | \(p = 0.008\)           | \(p = 0.007\)    |
| Venezuela            | −0.0174 (0.002)       | 0.0238 (0.006\(^*\))   | 0.0152 (0.005\(^*\)) |
|                      | \(p = 0.167\)         | \(p = 0.0001\)          | \(p = 0.001\)    |

Non-synonymous (\(K_N\)) and synonymous (\(K_S\)) divergence between \(P. vivax/P. cynomolgi\) and \(P. vivax/\) phylogenetically (\(P. cynomolgi/P. inui/P. fragile\)) related species. se, standard error. \(^*\)Statistically significant values.
regions having $\omega < 1$ (PvRBSA-A). The first belonged to the DMP1 domain and the other to a domain containing a LRR.

Recombination is an evolutionary force which can provide fresh opportunities for overcoming selective pressures to adapt to new environments and/or hosts by linking (within the same DNA region) independently arising variants (Perez-Losada et al., 2015). In fact, it has been observed that diversity levels increase with recombination rate (Hellmann et al., 2003; Kulathinal et al., 2008; Rao et al., 2011). Several tests were thus performed to assess whether this evolutionary force takes/has taken place in pvrbsa. A linear regression between LD and nucleotide distance showed that LD decreased regarding increased nucleotide distance, a pattern expected under recombination. Statistically significant ZZ values and several Rm were observed for all populations, thereby confirming recombination action (Supplementary Data Sheet 2). Two breakpoints were identified by the GARD method at nucleotides 618 and 1058 ($p = 0.0004$, nucleotide number base in the Sal-I sequence) (Figure 2). Recombination was thus taking place at this locus.

A haplotype network was inferred (Figure 3) and AMOVA used for assessing whether migration was also involved in shaping pvrbsa locus diversity (Table 5). Several haplotypes were shared between populations (Colombia and Venezuela, i.e. H_4) but also amongst subpopulations (Chocó, Córdoba, Amazonas, Meta, Bolivar, and the Venezuelan coastal area) lacking clear population structure (Figure 3). However, some haplotypes were restricted to particular subpopulations (i.e.
TABLE 5 | Analysis of molecular variance (AMOVA) analysis and inter-population $F_{ST}$ values.

| Source of variation | % of variation | $p$-value |
|---------------------|----------------|-----------|
| Between populations ($F_{CT}$) | 2.81 | 0.064 |
| Amongst subpopulations within populations ($F_{SC}$) | 4.39 | 0.000* |
| Amongst subpopulations ($F_{ST}$) | 92.79 | 0.001* |

$F_{ST}$ was calculated for parasite subpopulations within Colombian and Venezuela. Values close to 0 indicate low genetic differentiation whilst values close to 1 indicate high genetic differentiation. Values below the diagonal correspond to the $F_{ST}$ value and those above the diagonal represent the respective $p$-values. *Statistically significant values.

Assessing PvRBSA Functional Regions

It has been shown that immature reticulocytes are *P. vivax* target cells, expressing the CD71 receptor abundantly on their surface (Malleret et al., 2015). Therefore, a reticulocyte-binding assay was performed by flow cytometry to evaluate whether the PvRBSA region predicted under negative selection (PvRBSA-A having a $\omega < 1$) was able to bind to target cells (labeled with anti-CD71 antibody). First, the full PvRBSA protein and the PvRBSA-A and PvRBSA-B regions were recombinantly expressed and obtained in soluble form. Each protein was purified and recognized by WB using a monoclonal anti-polyhistidine antibody (Figure 4A). When proteins were incubated with umbilical cord blood (containing 6–7% reticulocytes), only the rPvRBSA ($m \pm SD = 5.5 \pm 1.8$) and rPvRBSA-A regions ($m \pm SD = 13 \pm 1.7$) bound to a CD71+CD45− cell population (reticulocytes), unlike the rPvRBSA-B region (Figure 4B). As can be observed, rPvRBSA and rPvRBSA-A binding activity had a statistically significant difference compared to negative control ($m \pm SD = 2.4 \pm 1.3$) (Kruskal–Wallis: $p = 0.014$) (Figure 4B). Once the PvRBSA functional region’s reticulocyte interaction activity had been determined, a binding inhibition experiment was performed with PvRBSA-A-derived peptides to search for specific regions within this region involved in interaction with target cells (Figure 4C). Peptide 40898 was able to inhibit protein-cell interaction by 18% whilst peptide 40894 produced a 35% reduction. Only one peptide (40897) inhibited rPvRBSA-reticulocyte interaction by more than 50%. The *Mycobacterium* peptide was unable to inhibit rPvRBSA binding activity to reticulocytes. Interestingly, synthesized peptides 40894 and 40897 were conserved in several *P. vivax* isolates and were located in the PRK15370 putative domain (Figure 2B).
DISCUSSION

Malaria remains a public health problem in several tropical and subtropical regions worldwide (WHO, 2017). Although an antimalarial vaccine appears to be a good cost-effective intervention which would help in controlling malaria, a fully effective vaccine has not been developed yet. Since P. vivax has a complex biology (Galinski et al., 2013) (for instance, it has tropism for reticulocytes, making it difficult to obtain a continuous in vitro culture), antigen identification and characterization for an antimalarial vaccine is a slow task (Patarroyo et al., 2012). Several antigens suggested as potential P. vivax vaccine candidates have been characterized, taking into account the candidates proposed for P. falciparum or other Plasmodium species (Arevalo-Pinzon et al., 2011, 2013; Patarroyo et al., 2012; Moreno-Perez et al., 2013). However, P. vivax and P. falciparum have different features and this could be the consequence of their different evolutionary paths. Consequently, several species-specific antigens could be found in both species. These species-specific antigens could thus be taken into account when designing an antimalarial vaccine.

The RBSA has recently been identified in species invading reticulocytes (P. vivax and P. cynomolgi) suggesting that this antigen could be specific for invading this kind of host cell and could therefore be considered a vaccine candidate when designing an anti-P. vivax malaria vaccine (Moreno-Perez et al., 2017). However, this antigen appears not to be exclusive for P. vivax and P. cynomolgi. A Blast search in whole monkey-malaria lineage genomes showed that pvrbsa orthologs were also present in species invading normocytes, such as P. inui and P. fragile (Supplementary Data Sheet 1B). Since the rbsa gene was present in P. fragile (the most basal species in monkey-malaria lineage, Figure 1) (Carlton et al., 2013; Muehlenbein et al., 2015), this gene could be present in the ancestor of all monkey-malaria parasites and, because orthologs were not found in P. knowlesi and P. coatneyi, then it has been lost in some species within this clade.

The rbsa gene was found as a single copy gene in P. inui (pirbsa) and P. fragile (pfirbsa), but this gene appears to be duplicated in P. vivax (pvrbsa) and P. cynomolgi (pcrbsap). The phylogenetic tree (Supplementary Figure S2) showed that pvrbsa was closer to pvrbsap than pcrbsa. This could be the consequence of concerted evolution which can homogenize duplicate gene fragments (Nei and Rooney, 2005). However, duplication was incomplete or was found at different chromosomes; the duplication event must therefore have taken place after P. vivax and P. cynomolgi divergence. A 4,000 bp fragment was duplicated at chr3 in P. vivax (it is present in all P. vivax strain genomes available in the PlasmoDB database; duplication and pseudogenisation should therefore have been taking place during early P. vivax evolutionary history); this duplicate (pcrbsap) lacked the first exon and the intron (Figure 1) and could, consequently, be a pseudogene. By contrast, both pcrbsa copies were complete in P. cynomolgi; in fact, they had 99.91% identity. Furthermore, pcrbsap was not found at chr3; instead, it was located at chr10. At least a 10 Mbp were duplicated for this species; however, the mechanism involved in this large fragment’s duplication at a different chr is not clear. Likewise, it is not yet clear whether the high identity found in pcrbsa and pcrbsap has been due to negative selection acting on both copies or has been due to a recent duplication event. According to the aforementioned results, two independent duplication events must have happened.

On the other hand, the pvrbsa gene was amplified in 167 samples from Colombian and Venezuelan populations. The derived sequences were analysed together with 65 sequences from different regions around the world. This gene had length polymorphism due to repeats located at the 5’-end. Worldwide, the pvrbsa encoding region had 44 segregating sites (47, taking encoding and non-encoding regions into account) giving 70 haplotypes (Supplementary Data Sheet 1E). The pvrbsa π value (0.0080 ± 0.0002) was lower than that observed for other Mrz surface proteins [pvmsp1, π > 0.05200 (Valderrama-Aguirre et al., 2011; Garzon-Ospina et al., 2015); pvmsp3a, π > 0.0349 (Mascorro et al., 2005; Garzon-Ospina et al., 2015); pvmsp7C, π = 0.0548; pvmsp7H, π = 0.0357; pvmsp7I, π = 0.0430 (Garzon-Ospina et al., 2012); pvmsp7E, π = 0.0573 (Garzon-Ospina et al., 2014)], but similar to that found in pvama1 [π = 0.0067 (Gunasekera et al., 2007; Garzon-Ospina et al., 2015)] and pvdbp [π = 0.0101 (Nobrega de Sousa et al., 2011)] and higher than pvmsp7? (−A π = 0.0002, −K π = 0.0025, −F π = 0.0008, and −L π = 0.0006) (Garzon-Ospina et al., 2011; Garzon-Ospina et al., 2014), pvmsp8 (π = 0.0022) (Pacheco et al., 2012), pvmsp10 (π = 0.0002) (Garzon-Ospina et al., 2011; Pacheco et al., 2012), pv12 (π = 0.0004), pv38 (π = 0.0026) and pv41 (π = 0.0037) (Forero-Rodriguez et al., 2014a,b; Wang et al., 2014) or rhoptry proteins [rap1, π = 0.00088, rap2, π = 0.00141, and ron4, π = 0.0004 (Garzon-Ospina et al., 2010; Pacheco et al., 2010; Buitrago et al., 2016)].

The Venezuelan subpopulations had more genetic diversity than subpopulations within Colombia regarding this gene (Table 1); however, the forces causing such diversity seemed to be similar. The sliding windows for frequency spectrum-based tests (Supplementary Figure S3) had statistically significant values within pvrbsa, suggesting that natural selection seems to be determining the pattern of diversity. The negative values (p < 0.02) found between nucleotides 100 and 300 (Supplementary Figure S3, numbers based on Supplementary Data Sheet 1E) in populations and subpopulations, suggested directional (negative or positive) selection; this kind of selection decreased diversity and could have been a consequence of functional constraint. On the other hand, the 750–940 nucleotide region was under balancing selection, as suggested by positive values for the frequency spectrum-based tests for these positions (Supplementary Figure S3). Directional selection was also identified between nucleotides 1,000–1,090 in the Colombian subpopulations whilst balancing selection was also found between nucleotides 1,090–1,200 in the Venezuelan subpopulations (Supplementary Figure S3).
Tests based on non-synonymous and synonymous mutations confirmed natural selection. A percentage of sites under positive selection were identified in \textit{P. vivax} and other species regarding the aBSREL method (Figure 2A). This pattern has also been observed in other genes, suggesting species-specific adaptation during parasite evolution (Muehlenbein et al., 2015; Buitrago et al., 2016; Garzon-Ospina et al., 2016). Evidence of positive selection was found in PvRBSA-B (Figure 2 and Tables 3, 4), agreeing with the frequency spectrum-based test results. All this data (Figure 2, Tables 3, 4, and Supplementary Figure S3) suggested that non-synonymous mutations were maintained in populations by balancing (or diversifying) selection, providing the parasite with an advantage to avoid host immune responses, as has been proposed for other antigens (Garzon-Ospina et al., 2012).

Whilst PvRBSA-B could be involved in host immune evasion, PvRBSA-A might be involved in parasite-host interaction. Negative values were found in dN-dS and K_{N}-K_{S} tests; however, they were not statistically significant (Tables 3, 4). Nevertheless, statistically significant negative frequency spectrum-based test values were observed. The \( \omega \) rate was lower than 1 in this region, several codons being under negative selection (Figure 2) according codon-based methods; this suggested that negative selection is/has been operating in PvRBSA-A and this region therefore seems to be under functional constraint. This premise was also supported because two putative domains were inferred within this region (Figure 2B): one was the DMP1 superfamily domain which is found in dentin matrix protein 1 acting as transcriptional component in mammals (Narayanan et al., 2003). However, how this domain could act in the parasite is not clear yet.

More interesting was the finding of the PRK15370 domain. This domain belongs to the NEL superfamily where family members have an LRR. The LRR has been involved in host-pathogen interaction (Kedzierski et al., 2004). Just PvRBSA-A bound when PvRBSA-A and -B recombinant protein fragments were assessed regarding their ability to bind to human reticulocytes (Figure 4). The binding percentage for PvRBSA-A was higher than that for the complete protein, which has been observed for other invasion-related \textit{Plasmodium} parasite antigens (Chitnis and Miller, 1994; Fraser et al., 2001; Kato et al., 2005; Arevalo-Pinzon et al., 2017). An inhibitory assay was then performed to identify minimum specific regions within PvRBSA-A able to inhibit rPvRBSA-reticulocyte interaction. Two peptides (FPEIRENLTASEESLTSCEE and ESLAASRESLDFCGSEESV) were able to decrease rPvRBSA-reticulocyte interaction. Both peptides were located in the PRK15370 putative domain, the first located towards the N-terminal domain whilst the other one was at the end of this. Remarkably, this region was found in the PvRBSA repeat region; repeats are usually used by the parasite as an immune evasion mechanism. Nevertheless, repeats could also be functionally important, as has been observed in the CS protein (Aldrich et al., 2012; Ferguson et al., 2014). The PvRBSA region involved in parasite-host interaction was thus located between amino acids 76 to 176 (numbers based on Sal-I protein sequences), FPEIRENLTASEEULSCEE and ESLAASRESLDFCGSEESV peptides (which were fully conserved in all \textit{P. vivax} sequences analyzed here) being critical for binding, meaning that these peptides could be considered in vaccine development.

The aforementioned results thus suggest that natural selection is an evolutionary force modulating \textit{pvrbsa} genetic diversity whilst negative selection acts at the gene’s 5' end, the 3' end is under diversifying or balancing selection. Nonetheless, other forces could also be involved. Recombination might increase diversity by interchanging DNA fragments during sexual reproduction. LD decreased in \textit{pvrbsa} as nucleotide distance increased (Supplementary Data Sheet 2), suggesting that recombination was/has been taking place in this gene. Recombination was confirmed by using ZZ, Rm and GARD method (Supplementary Data Sheet 2), suggesting that recombination can increase genetic diversity by intra-gene recombination.

A previous study has suggested that American populations are structured (Taylor et al., 2013); this pattern has also been observed in other \textit{P. vivax} antigens in Colombia (Forero-Rodriguez et al., 2014a,b; Buitrago et al., 2016), hence migration was the last force evaluated here. The haplotype network inferred by using all available sequences did not have clear structuring; \textit{pvrbsa} haplotypes were shared by different countries, even around the world. Several haplotypes found on the American continent were also present in Asia; consequently, most haplotypes should have arisen before \textit{P. vivax} spread worldwide. Therefore, due to different \textit{P. vivax} introductions to America by human migration (Rodrigues et al., 2018), \textit{pvrbsa} haplotypes have reached different American counties. A structured population was not found when a haplotype network was inferred. In fact, several haplotypes were shared between Colombia and Venezuela, as well as amongst all subpopulations (Figure 4); however, AMOVA suggested that subpopulations were genetically different (Table 5). Statistically significant values were found for most comparisons (Table 5) when the fixation index (\( F_{ST} \) based on haplotype diversity) was computed. This could have been the consequence of limited gene flow amongst Colombian subpopulations or due to local adaptation. Synonymous and non-synonymous mutations were analysed independently to try to determine which event(s) led to this structure (Supplementary Data Sheet 3). Since synonymous mutations are typically silent, they are considered to follow neutral expectations and, therefore, they could represent the demographic history (migration). On the other hand, non-synonymous mutations are subject to natural selection and could consequently represent local adaptations. The synonymous data set showed that Bolivar and Venezuela’s coastal area in Venezuela as well as the Chocó, Córdoba, Amazonas and Meta departments in Colombia seem to be genetically similar populations since non statistically significant \( F_{ST} \) were found. However, the non-synonymous data set showed that all (except Bolivar and Venezuela’s coastal area as well as Chocó and Meta) subpopulations were genetically different. This suggested that local adaptation is responsible for the
observed structure. However, limited gene flow (partly) might also have provoked such structuring in Colombia. Nevertheless, migration in Venezuela seems to be/have been an important force modulating pvrbsa diversity. Further analysis using neutral markers within Colombia and Venezuela populations could confirm this issue.

CONCLUSION

Although the RBSA protein has previously been identified as an antigen exclusive to Plasmodium species invading reticulocytes (Moreno-Perez et al., 2017), it is actually present in several monkey-malaria lineage species. The encoding gene should have been present in the last monkey-malaria parasite common ancestor and it then became lost in some species (i.e. P. knowlesi and P. coatneyi). An independent duplication event took place in P. vivax and P. cynomolgi. The pvrbsa paralog appears to be a pseudogene whilst the pcrbsa paralog is a functional gene having just one mutation between pcrbsa and pcrbsap.

The pvrbsa locus has lower genetic diversity (π = 0.008) than other Mrz surface proteins (Mascorro et al., 2005; Valderrama-Aguirre et al., 2011; Garzon-Ospina et al., 2012, 2014); this diversity is modulated by natural selection, recombination and migration (the latter for Venezuela but not for Colombia). According to Tajima, Fu and Li, K_N-K_S and codon-based tests the RBSAs C-terminal end (PVRBSA-B) is under balancing (or diversifying) selection, likely due to this region being involved in immune response evasion whilst PVRBSA-A is under directional selection due to a functional/structural constraint (ω < 1). The latter region has the PRK15370 putative domain (characterized by an LRR) and is involved in host-parasite interaction according to binding assays. Inhibition assays showed that two PRK15370 domain-derived peptides which were conserved in P. vivax isolates have been particularly involved in the specific interaction between PVRBSA and reticulocytes. Thus, these minimum regions could be considered when designing a fully effective anti-P. vivax vaccine.

AUTHOR CONTRIBUTIONS

PC-A performed the molecular evolutionary, recombinant expression and binding assays, and wrote the manuscript. DG-O devised and designed the study, performed the molecular evolutionary analysis, and wrote the manuscript. DM-P devised and designed the study, performed recombinant expression and binding assays, and wrote the manuscript. LR-C performed recombinant expression and binding assays and helped in writing the manuscript. ON and 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 the Instituto de Ciencia, Biotecnología e Innovación en Salud (IDCBIS) in Bogotá for supplying the umbilical cord blood and Jason Garry for translating and reviewing the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2018.00372/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(Web Server issue), W7–W13. doi: 10.1093/nar/gkq291

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

Anisimova, M., Nielsen, R., and Yang, Z. (2003). Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 164, 1229–1236.

 Arenas, M., and Posada, D. (2010). Coalescent simulation of intracodon recombination. Genetics 184, 429–437. doi: 10.1534/genetics.109.109736

 Arenas, M., and Posada, D. (2014). “The influence of recombination on the estimation of selection from coding sequence alignments,” in Natural Selection: Methods and Applications, ed. M. A. Fares (Boca Raton, FL: CRC Press/Taylor & Francis), 112–125.

 Arevalo-Pinzon, G., Bermudez, M., Curtidor, H., and Patarroyo, M. A. (2015). The Plasmodium vivax rhoptry neck protein 5 is expressed in the apical pole of Plasmodium vivax VCG-1 strain schizonts and binds to human reticulocytes. Malar. J. 14:106. doi: 10.1186/s12936-015-0619-1

 Ateu, C., Ateu, C., and Patarroyo, M. A. (2017). Plasmodium vivax ligand-receptor interaction: PvAMA-1 domain I contains the minimal regions for specific interaction with CD71+ reticulocytes. Sci. Rep. 7:9616. doi: 10.1038/s41598-017-10025-6

 Arevalo-Pinzon, G., Curtidor, H., Abril, J., and Patarroyo, M. A. (2013). Annotation and characterization of the Plasmodium vivax rhoptry neck protein 4 (PvRON4). Malar. J. 12:356. doi: 10.1186/1475-2875-12-356

 Arevalo-Pinzon, G., Curtidor, H., Patino, L. C., and Patarroyo, M. A. (2011). PVRON2, a new Plasmodium vivax rhoptry neck antigen. Malar. J. 10:60. doi: 10.1186/1475-2875-10-60

 Arnott, A., Barry, A. E., and Reeder, J. C. (2012). Understanding the population genetics of Plasmodium vivax is essential for malaria control and elimination. Malar. J. 11, 14. doi: 10.1186/1475-2875-11-14

 Bandelt, H. J., Forster, P., and Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16, 37–48. doi: 10.1093/oxfordjournals.molbev.a026036

 Baquero, L. A., Moreno-Perez, D. A., Garzon-Ospina, D., Forero-Rodriguez, J., Ortiz-Suarez, 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.2014.01067
Buitrago, S. P., Garzon-Ospina, D., and Patarroyo, M. A. (2016). Size
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.2014.01067
Buitrago, S. P., Garzon-Ospina, D., and Patarroyo, M. A. (2016). Size
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.2014.01067
Buitrago, S. P., Garzon-Ospina, D., and Patarroyo, M. A. (2016). Size
Murrell, B., Moola, S., Mabona, A., Weighill, T., Scheffler, K., and Kosakovsky Pond, S. L. (2012). Detecting individual sites subject to episodic diversifying selection. PLoS Genet 8(9), e1002764. doi: 10.1371/journal.pgen.1002764

Nanda Kumar, Y., Jayakodi, G., Gunasekaran, K., and Jambulingam, P. (2016). Computational screening and characterization of putative vaccine candidates of Plasmodium vivax. J. Biomol. Struct. Dyn. 34, 1736–1750. doi: 10.1080/07391102.2015.1090344

Narayan, K., Ramachandran, A., Hao, J., He, G., Park, K. W., Cho, M., et al. (2003). Dual functional roles of dentin matrix protein 1, implications in bone mineralization and gene transcription by activation of intracellular Ca2+ store. J. Biol. Chem. 278, 17500–17508. doi: 10.1074/jbc.M212700200

Neafsey, D. E., Galinsky, K., Jiang, R. H., Young, L., Sykes, S. M., Safi, 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

Nei, M., and Rooney, A. P. (2005). Combined and birth-and-death evolution of multigene families. Annu. Rev. Genet. 39, 121–152. doi: 10.1146/annurev.genet.39.070303.112240

Nobre da Sousa, T., Carvalho, L. H., and Alves de Brito, C. F. (2011). Worldwide genetic variability of the Duffy binding protein: insights into Plasmodium vivax vaccine development. PLoS One 6(6), e22944. doi: 10.1371/journal.pone.0022944

Ntumngia, F. B., Pires, C. V., Barnes, S. J., George, T. M., Thomsen-Luque, R., Kano, F. S., et al. (2017). An engineered vaccine of the Plasmodium vivax Duffy binding protein enhances induction of broadly neutralizing antibodies. Sci. Rep. 7, 13779. doi: 10.1038/s41598-017-13891-2

Pacheco, M. A., Battistuzzi, F. E., Junge, R. E., Conrero, O. E., Williams, C. V., Landau, I., et al. (2011). Timing the origin of human malaria: the lemur puzzle. BMC Biol. 11:299. doi: 10.1186/1471-2148-11-299

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

Pacheco, M. A., Ryan, E. M., Poe, A. C., Basco, L., Udahayakumar, V., Collins, W. E., et al. (2010). Evidence for negative selection on the gene encoding rhoptry-associated protein 1 (RAP-1) in Plasmodium spp. Infect. Genet. Evol. 10, 655–661. doi: 10.1016/j.meegid.2010.03.013

Patarroyo, M. A., Calderon, D., and Moreno-Perez, D. A. (2012). Vaccines against Plasmodium vivax: a research challenge. Exp. Rev. Vaccines 11, 1249–1260. doi: 10.1586/14767965.11.11.1249

Perez-Losada, M., Arenas, M., Galan, J. C., Palero, F., and Gonzalez-Candelas, F. (2015). Recombination in viruses: mechanisms, methods of study, and evolutionary consequences. Infect. Genet. Evol. 30, 296–307. doi: 10.1016/j.meegid.2014.12.022

Pond, S. L., Frost, S. D., Grossman, Z., Gravenor, M. B., Richman, D. D., and Brown, A. J. (2006). Adaptation to different human populations by HIV-1 revealed by codon-based analyses. PLoS Comput. Biol. 2(6), e62. doi: 10.1371/journal.pcbi.002062

Pozo, C. A., (2008). [ModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256. doi: 10.1093/molbev/msn083

Price, R. N., von Seidlein, L., Valecha, N., Nosten, F., Baird, I. K., and White, N. J. (2014). Global extent of chloroquine-resistant Plasmodium vivax: a systematic review and meta-analysis. Lancet Infect. Dis. 14, 982–991. doi: 10.1016/S1473-3099(14)70855-2

Rao, Y., Sun, L., Nie, Q., and Zhang, X. (2011). The influence of recombination on SNP diversity in chickens. Hereditas 148, 63–69. doi: 10.1111/j.1601-5223.2010.02210.x

Richie, T. L., and Saul, A. (2002). Progress and challenges for malaria vaccines. Nature 415, 694–701. doi: 10.1038/415694a

Rodrigues, P. T., Valdivia, H. O., de Oliveira, T. C., Alves, I. M. P., Duarte, A., Cerutti-Junior, C., et al. (2018). Human migration and the spread of malaria parasites to the New World. Sci. Rep. 8:1993. doi: 10.1038/s41598-018-19554-0

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Hohna, 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/syr029

Rozas, J., Gullaud, M., Blandin, G., and Agudo, M. (2001). DNA variation at the rp49 gene region of Drosophila simulans: evolutionary inferences from an unusual haplotype structure. Genetics 158, 1147–1155.
Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., and Kosakovsky Pond, S. L. (2015). Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* 32, 1342–1353. doi: 10.1093/molbev/msv622

Suh, K. N., Kain, K. C., and Keystone, J. S. (2004). *Malaria.* CMAJ 170, 1693–1702. doi: 10.1503/cmaj.1030418

Tachibana, S., Sullivan, S. A., Kawai, S., Nakamura, S., Kim, H. R., Goto, N., et al. (2012). *Plasmodium cynomolgi* genome sequences provide insight into *Plasmodium vivax* and the monkey malaria clade. *Nat. Genet.* 44, 1051–1055. doi: 10.1038/ng.2375

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

The malERA Consultative Group on Vaccines (2011). A research agenda for malaria eradication: vaccines. *PLoS Med* 8:e1000398. doi: 10.1371/journal.pmed.1000398

Valderrama-Aguirre, A., Zuniga-Soto, E., Marino-Ramirez, L., Moreno, L. A., Escalante, A. A., Arevalo-Herrera, M., et al. (2011). Polymorphism of the Pv200L fragment of merozoite surface protein-1 of *Plasmodium vivax* in clinical isolates from the Pacific coast of Colombia. *Am. J. Trop. Med. Hyg.* 84, 64–70. doi: 10.4269/ajtmh.2011.09-0517

Valencia, S. H., Rodriguez, D. C., Acero, D. L., Ocampo, V., and Arevalo-Herrera, M. (2011). Platform for *Plasmodium vivax* vaccine discovery and development. *Mem. Inst. Oswaldo Cruz* 106(Suppl 1), 179–192. doi: 10.1590/S0074-02762011000900023

Wang, Y., Ma, A., Chen, S. B., Yang, Y. C., Chen, J. H., and Yin, M. B. (2014). Genetic diversity and natural selection of three blood-stage 6-Cys proteins in *Plasmodium vivax* populations from the China-Myanmar endemic border. *Infect. Genet. Evol.* 28, 167–174. doi: 10.1016/j.meegid.2014.09.026

Welsh, K., and Bunce, M. (1999). Molecular typing for the MHC with PCR-SSP. *Rev. Immunogenet.* 1, 157–176.

White, N. J., Puokrittayakamee, S., Hien, T. T., Faiz, M. A., Mokuolu, O. A., and Dondorp, A. M. (2014). *Malaria.* *Lancet* 383, 723–735. doi: 10.1016/S0140-6736(13)60024-0

WHO (2017). *World Malaria Report 2017.* Geneva: World Health Organization.

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

**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 Camargo-Ayala, Garzón-Ospina, Moreno-Pérez, Ricaurte-Contreras, Noya 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) and the copyright owner(s) 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.