Infection, Genetics and Evolution xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Short communication

Inferring natural selection signals in *Plasmodium vivax*-encoded proteins having a potential role in merozoite invasion

Diego Garzón-Ospina [a,b], Johanna Forero-Rodríguez [a], Manuel A. Patarroyo [a,b,*]

[a] Molecular Biology and Immunology Department, Fundación Instituto de Inmunología de Colombia (FIDIC), Carrera 50 No. 26-20, Bogotá DC, Colombia
[b] Basic Sciences Department, School of Medicine and Health Sciences, Universidad del Rosario, Carrera 24 No. 63C-69, Bogotá DC, Colombia

ARTICLE INFO

Article history:
Received 5 February 2015
Received in revised form 30 April 2015
Accepted 2 May 2015
Available online xxxx

Keywords:
Plasmodium vivax
Anti-malarial vaccine
Natural selection signal
Allele-specific response

ABSTRACT

Detecting natural selection signals in *Plasmodium* parasites antigens might be used for identifying potential new vaccine candidates. Fifty-nine *Plasmodium vivax*-Sal-I genes encoding proteins having a potential role in invasion were used as query for identifying them in recent *P. vivax* strain genome sequences and two closely-related *Plasmodium* species. Several measures of DNA sequence variation were then calculated and selection signatures were detected by using different approaches. Our results may be used for determining which genes expressed during *P. vivax* merozoite stage could be prioritised for further population genetics or functional studies for designing a *P. vivax* vaccine which would avoid allele-specific immune responses.

1. Introduction

Malaria is a disease caused by *Plasmodium* parasites (Cox, 2010). *Plasmodium falciparum* is the best characterised species, whereas research into *Plasmodium vivax* has been more limited (Arnott et al., 2012; Patarroyo et al., 2012). Likewise, anti-*P. vivax* vaccine development is behindhand and few vaccine candidates have been proposed to date. Characterising potential new candidates involved the search for sequences having a high level of identity with *P. falciparum* antigens (Patarroyo et al., 2012). Recently, Restrepo-Montoya et al. (2011) has led to categorising several *P. vivax* proteins having a potential role in invasion by bioinformatics approaches.

The proteins characterised in the aforementioned studies could be used for *P. vivax* asexual-blood vaccine development since they seem to be implicated in invasion; however, these molecules' genetic diversity and evolutionary forces must be ascertained by population genetics analysis to design a completely effective vaccine (Arnott et al., 2012; Barry and Arnott, 2014). The most commonly used tests in population genetics are based on the allele frequency spectrum and require the sequencing of many isolates; therefore, performing such studies for all these genes would involve much time and resources. However, Cornejo et al. (2014), using a limited sample size (Genomes from 5 isolates) have identified genes having signatures consistent with selection. This kind of analysis could be a starting point for detecting potential new vaccine candidates (Weedall and Conway, 2010), similar to the approach adopted for *P. falciparum* (Ochola et al., 2010; Tetteh et al., 2009).

The present study has used three different approaches for detecting selection signals within 59 previously-characterised merozoite antigens using the sequences from five *P. vivax* isolates and two closely-related species. The results may be used for determining which antigens might be prioritised and evaluated in further studies aimed at designing a completely effective vaccine.

2. Material and methods

2.1. Target sequences and alignments

Sequences were obtained for 59 protein-encoding genes from the Salvador I isolate (Sal-I); these genes had been previously characterised by adopting a molecular approach (Arevalo-Pinzon et al., 2011, 2013; Moreno-Perez et al., 2013b; Patarroyo et al., 2012) or suggested as promising vaccine candidates by having a potential role in invasion (Restrepo-Montoya et al., 2011) (Supplementary data 1). Forty-eight genes had not been subjected to previous

* Corresponding author at: Molecular Biology and Immunology Department, Fundación Instituto de Inmunología de Colombia (FIDIC), Carrera 50 No. 26-20, Bogotá DC, Colombia.
E-mail addresses: degarzon@gmail.com (D. Garzón-Ospina), lady2007_10@hotmail.com (J. Forero-Rodríguez), mapatarr.fidic@gmail.com (M.A. Patarroyo).

http://dx.doi.org/10.1016/j.meegid.2015.05.001
1567-1348/

population genetic analysis and 11 have been previously evaluated. These sequences were used as query for searching for them in the available genomic *P. vivax* isolate sequences (Neafsey et al., 2012) and two closely-related species (*Plasmodium cynomolgi* and *Plasmodium knowlesi*) (Pain et al., 2008; Tachibana et al., 2012) using the tBlastn tool from the protozoa genomic NCBI (Richie and Saul, 2002) database. A tBlastn search in GenBank database was made regarding sequences reported for other stains (VCG-I, Belen or South Korea).

Some genes belong to multigene families; therefore orthologous identification should be performed. A combination of criteria was used for identifying putative orthologues, including a phylogenetic signal (tree topology), sequence similarity (genetic distance) and synteny (similar genomic position), as previously described (Arisue et al., 2011; Garzon-Ospina et al., 2010, 2014; Rice et al., 2014). *P. vivax*, *P. cynomolgi* and *P. knowlesi sera*, *msp-3*, *msp-7*, *clag*, *pfam-a* and *pfam-d* genes were aligned with all members of their families, respectively using MUSCLE (Edgar, 2004), followed by manual edition. The best evolutionary model was selected for each alignment by Bayesian Information Criterion, using MEGA software (Tamura et al., 2011). Maximum likelihood phylogenetic trees were then inferred using the respective model; all gaps and ambiguously-aligned regions were removed. Topology reliability was evaluated by bootstrapping (1000 iterations). Multiple alignments were then made (by MUSCLE) for single-copy genes using sequences from isolates, together with *P. cynomolgi* and *P. knowlesi* orthologous sequences.

2.2. Genetic diversity and natural selection analysis

DnaSP software (Librado and Rozas, 2009) was used for estimating several measures regarding DNA sequence variation. Cornejo et al. (2014) had previously identified patterns consistent with natural selection acting across the *P. vivax* genome by using the two-dimensional Hudson, Kreitman and Aguade (HKA) test, the genome-wide version of the McDonald–Kreitman (MK) test and Tajima D estimator; however, natural selection signals were not found for several genes involved in merozoite invasion. We assessed natural selection by conventional MK test (McDonald and Kreitman, 1991) and π/K ratio. The MK test was performed taking the Jukes–Cantor divergence correction into account (Jukes, 1969) by using a web server (Egea et al., 2008). The π/K ratio was evaluated for identifying genes having a high value correlated with balancing selection (Ochola et al., 2010; Tetteh et al., 2009). MEGA software was used to assess selection signals within *P. vivax* by calculating the non-synonymous substitution per site rate (d_N) and synonymous substitution per site rate (d_S) by the modified Nei–Gojobori method (Zhang et al., 1998). Likewise, to infer natural selection signatures which could have prevailed during *Plasmodium* evolutionary history (using *P. vivax*, *P. cynomolgi* and *P. knowlesi* sequences as data set) the difference between the average number of non-synonymous divergence substitutions per non-synonymous site rate and of synonymous divergence substitutions per synonymous site rate (K_N–K_S) was inferred using the modified Nei–Gojobori method with Jukes–Cantor correction. Significant differences were evaluated by *Z*-test (when non-synonymous and synonymous substitutions > 10) or Fisher's exact test (when non-synonymous and synonymous substitutions < 10). Furthermore, a sliding window for d_N/d_S and K_N/K_S ratios (ω) was performed; gaps and ambiguously aligned regions were removed for analysis. Genetic diversity and selection were assessed by Sal-I annotation as reference.

The most suitable antigens for vaccine development regarding our approach should have limited diversity or at least a domain having this pattern. Such genes/domains should have a natural negative selection signal (and $\omega < 1$). However, genes under positive selection might be taken into account if provided with domains having both limited diversity and low ω values.

3. Results

3.1. Diversity analysis

Sequences from 59 previously identified *P. vivax* protein-encoding genes having a potential role in invasion were analysed here. Sequence analysis revealed premature stop codons in PVX_096990 and PVX_097710 genes in Mauritania-I and Brazil-I isolates, respectively. Few genes were absent or incomplete in some isolates; i.e. the PVX_003825 gene was not found in a North Korean isolate whereas PVX_092425 was incomplete at the 5′-end, PVX_086850 and PVX_086930 were missing in the Brazil-I isolate, PVX_097700 was absent in the Mauritania-I isolate and PVX_097710 appeared not to be present in the India-VII isolate in which the PVX_096990 gene was incomplete at the 5′-end.

Genetic diversity measurement revealed 16 highly polymorphic genes ($\pi > 0.01$), 35 with intermediate polymorphism ($0.009 < \pi > 0.001$) and 8 having low genetic diversity ($\pi < 0.001$) (Table 1 and Supplementary data 1). Fig. 1 shows the nucleotide polymorphism distribution within the aforementioned 59 genes.

Phylogenetic trees were then inferred to determinate putative orthologous relationships for the multigene families (Supplementary data 2). Putative orthologues had to be clustered in a clade in a one-to-one relationship and had to have a similar genomic position. The clades formed in some families agreed with previous reports (Arisue et al., 2011; Garzon-Ospina et al., 2010; Rice et al., 2014). Hence, 34 of these 59 genes were found in both *P. cynomolgi* and *P. knowlesi* species, whereas another 15 genes were only present in *P. cynomolgi*; 10 genes appeared to be exclusive to *P. vivax*.

3.2. Natural selection signatures in P. vivax genes

Three different approaches were used for screening natural selection signals. The neutral index (NI) from the MK test showed that 16 genes had excess polymorphism regarding divergence (NI > 1) and only one had NI < 1, while the π/K ratio showed 11 genes which might be under balancing selection (Table 2 and Supplementary data 1).

A statistically significant $d_N > d_S$ was found in 12 genes while 9 had significant $d_N < d_S$ values (Table 2 and Supplementary data 1). The K_N–K_S difference gave negative selection between species for 35 genes whereas another 10 displayed positive selection (Supplementary data 1). Some genes had a different natural selection signal from that previously reported (Tachibana et al., 2012), probably since we used sequences from 5 isolates, unlike Tachibana et al., who only used the Sal-I isolate. Some genes evaluated here were not assessed in the aforementioned report.

Since the Nei–Gojobori method is a conservative test, we performed a sliding window for the ω rate (d_N/d_S and/or K_N/K_S) for identifying specific domains within genes having a determined selective signal (Supplementary data 3). Several genes lacking significant d_N or d_S rates displayed a particular domain having $d_N/d_S \pm 1$ but also $K_N/K_S < 1$ throughout sequences. Members of *pvmsp-3* and *pvsera* multigene families had $K_N/K_S > 1$ values throughout all genes, suggesting high divergence between *P. vivax* and related species.

4. Discussion

A vaccine focusing on *P. vivax* is urgently needed for malaria control; however, its design has been delayed, mainly due to slow

Please cite this article in press as: Garzón-Ospina, D., et al. Inferring natural selection signals in *Plasmodium vivax*-encoded proteins having a potential role in merozoite invasion. Infect. Genet. Evol. (2015), http://dx.doi.org/10.1016/j.meegid.2015.05.001