Infection, Genetics and Evolution 10 (2010) 298–303

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Evolutionary dynamics of the immunodominant repeats of the *Plasmodium vivax* malaria-vaccine candidate circumsporozoite protein (CSP)

Aarti Patil [a,b], Pamela Orjuela-Sánchez [a], Mônica da Silva-Nunes [a,c], Marcelo U. Ferreira [a,*]

[a] Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, Av. Prof. Lineu Pestes 1374, 05508-900 São Paulo (SP), Brazil
[b] Mount Sinai School of Medicine, New York, NY, USA
[c] Center for Health Sciences and Sports, Federal University of Acre, Rio Branco (AC), Brazil

ARTICLE INFO

Article history:
Received 19 October 2009
Received in revised form 12 January 2010
Accepted 15 January 2010
Available online 25 January 2010

Keywords:
Plasmodium vivax
Malaria
Circumsporozoite protein
Genetic diversity
Repetitive domains
Recombination
Mismatch distribution analysis
Vaccine

ABSTRACT

The circumsporozoite protein (CSP) of *Plasmodium vivax*, a major target for malaria vaccine development, has immunodominant B-cell epitopes mapped to central nonapeptide repeat arrays. To determine whether rearrangements of repeat motifs during mitotic DNA replication of parasites create significant CSP diversity under conditions of low effective meiotic recombination rates, we examined *csp* alleles from sympatric *P. vivax* isolates systematically sampled from an area of low malaria endemicity in Brazil over a period of 14 months. Nine unique *csp* types, comprising six different nonapeptide repeats, were observed in 45 isolates analyzed. Identical or nearly identical repeats predominated in most arrays, consistent with their recent expansion. We found strong linkage disequilibrium at sites across the chromosome 8 segment flanking the *csp* locus, consistent with rare meiotic recombination in this region. We conclude that CSP repeat diversity may not be severely constrained by rare meiotic recombination in areas of low malaria endemicity. New repeat variants may be readily created by nonhomologous recombination even when meiotic recombination is rare, with potential implications for CSP-based vaccine development.

1. Introduction

Plasmodium vivax is the most widespread of the four human malaria parasites, causing 132–391 million episodes of disease each year (Hay et al., 2004), with 2.6 billion people at risk of infection worldwide (Guerra et al., 2006). Outside of Africa, *P. vivax* is the main cause of malaria morbidity, with enormous public health burden. However, since *P. vivax* usually causes less severe symptoms than *Plasmodium falciparum*, it has received relatively little attention and limited funds for research and control (Price et al., 2007).

A notable feature of several malaria surface antigens (including major vaccine-candidate molecules) is the presence of tandem arrays of relatively short amino acid motifs. The circumsporozoite protein (CSP), an abundant antigen on the surface of sporozoites, has been extensively used as a vaccine development target (Nardin and Zavala, 1998). CSP has immunodominant B-cell epitopes mapped to central repeats (CR) bracketed between nonrepetitive sequences. *P. vivax* CSP displays two major types of nonapeptide repeats (most commonly, GDRA[D/A]GQPA and ANGAGNQPG), which define the variants known as VK210 and VK247, respectively (Rosenberg et al., 1989). Although VK210- and VK247-type sequences often occur in sympatric parasite populations, no example is known of hybrid CR array with both repeat types (Lim et al., 2005).

Insertions and deletions in the CR domain, resulting from either sexual recombination during meiosis or intrahelical strand-slippage events during mitotic DNA replication (McConkey et al., 1990), generate novel CSP variants that may be positively selected if mutant parasites evade host's immunity. Extensive variation occurs in the CR of *P. falciparum* CSP, which consists of variable numbers of copies of 4-mer NANP and NVDP motifs, without breaking down the tight linkage between polymorphic sites in flanking sequences as interhelical exchanges during meiosis would do (Rich et al., 1997). This pattern suggests that CR, as other short repeats such as microsatellite- or minisatellite-type sequences (Levinson and Gutman, 1987), undergo frequent intrahelical recombination. Whether similar mechanisms create significant variation in CR arrays of *P. vivax* CSP, which consist of longer (9-mer) repetitive motifs, remains unknown. Frequent mitotic recombination coupled with positive selection of new variants might accelerate CR evolution even when meiotic

* Corresponding author. Tel.: +55 11 30917746; fax: +55 11 30917417.
E-mail address: muferrei@usp.br (M.U. Ferreira).

1567-1348/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2010.01.006

recombination and outcrossing are relatively uncommon in malaria parasite populations (Rich et al., 2000). Whether or not length variation in CR arrays affects the recognition of B-cell epitopes is uncertain, but the conformational nature of CR epitopes in *P. falciparum* CSP (Monette et al., 2001) supports this hypothesis.

Here, we examine patterns of CR sequence diversity in *csp* alleles from sympatric *P. vivax* isolates from an area of low malaria endemicity. We use single-nucleotide polymorphism (SNP) typing to examine, in these same isolates, the haplotype structure of chromosome 8, where the *csp* gene is located. We sought to determine whether significant CR diversity at the *P. vivax csp* locus occurs under conditions of low malaria endemicity, which reduces effective meiotic recombination rates in local parasites.

2. Materials and methods

2.1. Study area and parasite population

Between March 2004 and May 2005, we obtained 54 isolates of *P. vivax* from subjects living in the eastern corner of Acre State, Western Amazon Basin of Brazil. Study subjects participated in a population-based cohort study in the rural community of Granada (9°41′S to 94°9′S, 67°05′W to 67°07′W), which is part of the frontier agricultural settlement known as Pedro Peixoto Settlement, situated 50 km northwest of Acrelândia, the nearest town, and 150 km east of Rio Branco, the capital of Acre. Both *P. falciparum* and *P. vivax* are transmitted year-round, with average incidence rates of 30 slide-confirmed *P. vivax* infections/100 person-years at risk and 16.3 slide-confirmed *P. falciparum* infections/100 person-years at risk during the study period (Da Silva-Nunes et al., 2008). Eight pairs of *P. vivax* isolates were collected from the same individuals during consecutive infections (40–210 days apart); all others were collected from unrelated subjects. The collected isolates represent 36.7% of all slide-confirmed symptomatic *P. vivax* infections diagnosed in Granada between March 2004 and May 2005. Previous microsatellite analysis of these isolates indicated that 49% of them comprised more than one genetically distinct clone (i.e., they contained multiple-clone infections) (Ferreira et al., 2007). Patients or guardians provided written informed consent, and the research protocol was approved by the ethics review board of the Institute of Biomedical Sciences, University of São Paulo (318/2002 and 538/2004). *P. vivax* infections were treated with chloroquine and primaquine (Ministry of Health of Brazil, 2001). Genomic DNA was isolated as described (Ferreira et al., 2007).

2.2. Amplification and sequencing of circumsporozoite protein (CSP) repeats of P. vivax

The CR domain of *P. vivax csp* was amplified by nested polymerase chain reaction (PCR) essentially as described by Imwong et al. (2005). Briefly, the external oligonucleotide primers VCS-OF (5′-ATGTAGATCTGTCCAAGGCCATAAA-3′) and VCS-OR (5′-TAATTGAATAATGCTAGGACTAACAATATG-3′), which target conserved domains at the nonrepetitive 5′ and 3′ domains flanking the repeats, were used to amplify a fragment of ~1100 base pairs (bp), from 1 μl of purified parasite DNA, with 25 cycles and annealing temperature set at 58 °C. The internal primers VCS-NF (5′-GCAGAACCAAAAAATCCACGTGAAAATAAG-3′) and VCS-NR (5′-CCAACGGTAGCTCTAACTTTATCTAGGTAT-3′) amplified a ~680-bp fragment, starting with 1 μl of the primary PCR product, with 30 cycles and annealing temperature set at 62 °C. Sequencing was performed on both DNA strands directly on secondary PCR products, which were purified with HiYield Gel/PCR DNA Extraction kits (Real Biotech, Taipei, Taiwan), by using the internal primers VCS-NF and VCS-NR. BigDye v3.1 terminator chemistry (Applied Biosystems, Foster City, CA) was used in sequencing reactions and products were analyzed on a 3100 automated DNA sequencer (Applied Biosystems). Sequences of poor quality or those displaying superimposed electropherogram peaks, indicating mixed-genotype infections, were discarded, leaving a total of 45 *csp* sequences analyzed. Nucleotide variants that were seen in a single isolate were confirmed by independent reamplification and resequencing from purified genomic DNA of the relevant isolate. New *csp* sequences obtained in this study were deposited in GenBank database under the accession numbers FJ845383–FJ845391.

2.3. Single-nucleotide polymorphism (SNP) typing along chromosome 8 of P. vivax

To characterize the haplotype structure of the chromosome segment surrounding the *csp* locus, we typed 31 biallelic SNP markers across a 100-kilobase (kb) nontelomeric region of chromosome 8 that comprises the 1.2-kb *csp* locus (Feng et al., 2003; Carlton et al., 2008). We identified SNPs and characterized flanking sequences by aligning orthologous sequences from five *P. vivax* isolates (GenBank accession numbers AY003872, AY216936, AY216937, AY216938, AY216939). Markers were selected to meet the following criteria: (a) no SNP could be located in repetitive domains; (b) 200 bp of sequence upstream and downstream of each SNP should have no significant similarity to human sequences (as determined by BLAST search against the human genome), to prevent cross-amplification of human DNA present in the sample; and (c) both alleles of each SNP should be present in our parasite population (i.e., no monomorphic locus was selected). Thirty-one SNPs meeting these criteria (including eight synonymous and two nonsynonymous SNPs in coding sequences and 21 SNPs in intergenic regions), which were successfully typed in a large proportion of field isolates, were selected for this analysis (Supplementary Table 1 online). Supplementary Fig. 1 (published online) shows the map location of these SNPs in relation to the *csp* locus. The physical distance between pairs of SNP markers ranged between 174 and 96315 kb. SNP typing was performed, under contract, by K-Biosciences (Cambridge, UK), with an amplifluor assay (Nazarenko et al., 1997; Newton et al., 1989). To obtain adequate DNA concentrations, parasite DNA was submitted to whole-genome amplification (WGA) prior to SNP typing (Wang et al., 2009). WGA was performed on 10 ng of genomic DNA, with high-fidelity multiple displacement technology (Dean et al., 2002), using a REPLI-g Minikit (Qiagen, Valencia, CA) according to the manufacturer's instruction. All sequences of oligonucleotide primers (reverse-strand sequences as in Feng et al., 2003) used for SNP typing are given in Supplementary Table 2 online; the annealing temperature for all primers was 60 °C. Haplotypes characterized with SNPs (unique combinations of alleles at 31 sites on chromosome 8) are referred to, throughout this article, as Chr-8 haplotypes. Thirty-one isolates had all SNPs typed, with a single allele at each site (i.e., no apparent mixed-clone infection), and had their *csp* gene successfully sequenced.

2.4. Data analysis

Nucleotide sequences were aligned manually. Parasites sharing the same nucleotide sequence in the CR array were assigned to the same *csp* type. The different nucleotide sequences encoding the same nonapeptide motif, termed repeat allotypes (RATs; Rich et al., 1997), were recorded. To examine the mode of evolution of CR, we used the mismatch distribution analysis suggested by Hughes (2004). We aligned the nucleotide sequences encoding each nonapeptide repeat unit in each *csp* type and calculated the following parameters: *p*, the average proportion of nucleotide