DRAMATIC DIFFERENCE IN DIVERSITY BETWEEN PLASMODIUM FALCIPARUM AND PLASMODIUM VIVAX RETICULOCYTE BINDING-LIKE GENES

JULIAN C. RAYNER,* TUAN M. TRAN,* VLADIMIR CORREDOR, CURTIS S. HUBER, JOHN W. BARNWELL, AND MARY R. GALINSKI

Division of Parasitic Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia; Emory Vaccine Center, Yerkes National Primate Research Center, and Department of Medicine, Division of Infectious Diseases, Emory University, Atlanta, Georgia; Departamento de Ciencias Fisiológicas, Facultad de Medicina, Universidad Nacional de Colombia, Bogotá, Colombia

Abstract. Malaria parasite proteins involved in erythrocyte invasion are considered important vaccine targets. Members of the reticulocyte binding-like (RBL) family of Plasmodium merozoite proteins are found in human, simian, and rodent malaria parasites and function in the initial steps of erythrocyte selection and invasion. The RBL genes are large, ranging in size from 7.7 to 10 kb, and the extent of any sequence diversity in parasite populations is unknown. We present the first assessment of sequence diversity within RBL genes from the two major human malaria parasites: Plasmodium falciparum and P. vivax. Polymorphism within the RBL genes is generally limited, except for P. vivax reticulocyte binding protein 2 (PvRBP2), which has nucleotide diversity levels 25-fold higher than the other RBL genes. The PvRBP2 haplotypes appear to fall into two distinct classes of alleles, suggesting large-scale dimorphism in this gene. Polymorphisms were frequently clustered, suggesting that different RBL domains may be evolving under different selection and functional pressures.

INTRODUCTION

Malaria is caused by parasites of the genus Plasmodium and kills an estimated 1–3 million people each year. Of the four Plasmodium species that infect humans, P. falciparum causes most malaria mortality, while the distantly related P. vivax species causes an estimated 80 million malaria cases annually, mostly outside sub-Saharan Africa. The overwhelming burden of malaria and the appearance and spread of drug-resistant parasites has focused attention on the development of malaria vaccines for both species. Although vaccines targeting all stages of the Plasmodium life cycle have been proposed, the symptoms and pathology of malaria are caused by the erythrocytic stage, during which Plasmodium merozoites invade and then develop within erythrocytes, culminating in erythrocyte lysis and the release of daughter merozoites. Merozoite proteins that function in the recognition and invasion of erythrocytes are therefore being intensively studied as vaccine candidates. Studies of how erythrocyte invasion proteins diverge between Plasmodium isolates is an important step in assessing their utility as vaccine candidates, and can provide evidence of the selection and functional pressures acting on specific sub-domains.

Invasion of erythrocytes is a complex, multi-step process and P. falciparum and P. vivax erythrocyte invasion pathways clearly differ in specificity. Plasmodium vivax merozoites invade only reticulocytes, whereas P. falciparum merozoites can invade mature erythrocytes as well as reticulocytes. However, many of the invasion ligands used by the two species are related, supporting the hypothesis that fundamental biologic steps in the invasion process are conserved. The P. vivax Duffy binding protein (PvDBP), which adheres to the Duffy antigen receptor for chemokines on the surface of erythrocytes, is a homolog of erythrocyte binding antigen-175 (PfEBA-175), the P. falciparum ligand that binds to glyco- phorin A. Similarly, orthologs of apical membrane antigen 1 (AMA1), a vaccine candidate in both Plasmodium species, appear to be involved in re-orienting the parasite during invasion. Geographic diversity in these vaccine candidates has been studied extensively, with specific emphasis on functionally important sub-domains.

Members of the reticulocyte binding-like (RBL) superfamily of invasion proteins have been identified in rodent, simian, and human Plasmodium parasites. The gene family is named after the P. vivax reticulocyte binding proteins 1 and 2 (PvRBP1 and PvRBP2), which bind specifically to reticulocytes and have been proposed to select them for invasion. Homologs have been characterized in P. falciparum, and are referred to as both normocyte binding proteins (PINBPs) as originally proposed and RBP homologs (PRHs). The former nomenclature is used in the remainder of this paper. PINBP1, PINBP2a, and PINBP2b have been shown to be involved in invasion, with PINBP1 binding directly to an unidentified trypsin-resistant receptor on the erythrocyte surface and PINBP2a and PINBP2b playing non-overlapping roles during invasion.

This paper presents the first assessment of DNA sequence diversity in P. falciparum and P. vivax RBL genes from Plasmodium isolates originating from geographically distinct locations. Given the potential for important structural and functional domains throughout these proteins, the RBL genes, which range in size from 7.7 to 10 kb, were sequenced and analyzed in their entirety. The presence of related RBL genes in P. falciparum and P. vivax allowed us to compare diversity and selection pressures both within and between Plasmodium species, with surprising results. Diversity levels were generally comparable to other genes involved in erythrocyte invasion, but polymorphism in PvRBP2 was exceptionally high and PvRBP2 sequences could be assigned into two dimorphic classes of alleles. Clustering of polymorphisms was observed in several genes, suggesting that different selection pressures may be acting on different domains. The implications for RBL function, evolution, and consideration as vaccine candidates are discussed.

MATERIALS AND METHODS

Plasmodium falciparum strains, culture, and extraction of DNA. The P. falciparum Vietnam Oak Knoll (FVO) strain

* Both authors contributed equally to this work.
was obtained from a parasite line adapted to *Aotus* monkeys; 7G8. Dd2 and HB3 are clonal lines of the Ituxi (Brazil), Indochina III (southeast Asia), and Honduras I (Honduras) strains, respectively, and were obtained from the Malaria Research and Reference Reagent Resource Center (Manassas, VA) (MR4, www.mr4.org). Clone 3D7 (probably African in origin) was obtained from the Walter Reed Army Institute of Research (Silver Spring, MD), and the Malayan Camp K- (MC) strain (Malaysia) was derived from stocks maintained at the National Institutes of Health (Bethesda, MD). All *P. falciparum* strains were cultured in human O+ parasites as described and DNA was prepared as described or using a QIAamp DNA blood mini kit (Qiagen, Valencia, CA).

**Plasmodium falciparum NBP amplification and cloning.** *PfNBP1.* The polymerase chain reaction (PCR) products were amplified using either the Expand High Fidelity PCR System (Roche, Indianapolis, IN) or Pfx DNA Polymerase (Invitrogen, Carlsbad, CA). The complete gene was cloned in five or six overlapping fragments using either a TA cloning kit (Invitrogen) or PCR-Script AMP cloning kit (Stratagene, La Jolla, CA). The primers used in amplifications are as follows: 1F (5’-GATGTATATTGTTTGATATTCTTTG-3’) and 1R (5’-ACTGCTTGATGTGTAATATGGTTTG-3’); 2F (5’-GCTAAAGCTTTGAAGAGCACATCAAAC-3’) and 2R (5’-GTCCTATCATTTAGCTACATACGTG-3’); 3F (5’-CCAATGATAGGATGATGACC-3’) and 3R (5’-GGTCTGGATATTTTTATGCCTGAC-3’); 4F (5’-GAGAACACGTAGAGAAGAGGA-3’) and 4R (5’-ACTTACATAAAGAGATTGATC-3’); 5F (5’-GATGATCAATTTTTATGCCTGAC-3’); and 5R (5’-GGGATAAGCTGACAGATG-3’); 6F (5’-GGAAGTAATGTCTACATGCATGTG-3’); and 6R (5’-CTTATAGTTGATATTTCTTGATG-3’); and 7F (5’-CTTACATCCTCAACCGTCTC-3’); and 7R (5’-GGGATAAGCTGACAGATG-3’). For some samples, 3F was paired with 4R; 5F (5’-GGGATAAGCTGACAGATG-3’); and 5R (5’-GGGATAAGCTGACAGATG-3’); and 6F (5’-GTAATACATCTCGGATG-3’); and 6R (5’-CTTACATCCTCAACCGTCTC-3’); and 7F (5’-CTTACATCCTCAACCGTCTC-3’); and 7R (5’-GGGATAAGCTGACAGATG-3’). P*NBP2a* and P*NBP2b* Given the extensive region of sequence identity between these two genes, 18–22 P*NBP2a* and *PfNBP2b* could not be amplified in small fragments because such fragments could not be assigned unequivocally to one gene or the other. Instead complete P*NBP2a* and P*NBP2b* genes were amplified with a combination of shared and gene-specific primer pairs using the Expand Long Template PCR system (Roche). The primers used were SF9BM (5’-TTCCAGGATCGGACGCTTTGATATGATG-3’) and N2ARBGM (5’-TTCCAGGATCGGATCAATACTCGTATG-3’) for *PfNBP2a* and SF8BM (5’-TTCCAGGATCGGACGCTTTGATATGATG-3’) for *PfNBP2a* and *PfNBP2b.* Complete P*NBP2a* and P*NBP2b* genes were gel purified and digested with Eco RI (there are two internal *Eco* RI sites in the region shared by both *PfNBP2a* and *PfNBP2b*) and Bam HI (*Bam* HI sites were included in all primers). This digest split each gene into three fragments, which were cloned into *Eco* RI- (for the internal fragment) or *Eco* RI-*Bam* HI (for the two end fragments)–digested pBluescript (Stratagene) using T4 DNA Ligase (New England Biolabs, Beverly, MA).

**Plasmodium vivax RBP amplification and cloning.** Complete *PvRBP1* and *PvRBP2* genes were amplified as eight or five overlapping fragments, respectively, using the Expand High Fidelity PCR system (Roche). Fragments were cloned using the TA cloning system (Invitrogen).

**Plasmodium falciparum NBP amplification and cloning.** *PfRBP2.* Primers used in *PfRBP2* amplification were as follows: 1.0FP (5’-TTCCATCTTTGAAACGAGACAT-3’) and 1.0RP (5’-TTCCTATCAACCACACGGTG-3’); 1.1FP (5’-TCTTATATTCCCCACTTGAGG-3’) and 1.1RP (5’-TTGCTTTTGGTTTCTCTCTTG-3’); 1.2FP (5’-TATACCTGGAAGGTGATCAGTGAAGGGTGTG-3’) and 1.2RP (5’-GCTGACTAAAGGAGGTG-3’); and 1.3XPFP (5’-GCAGGTTAACATGAAATTTCG-3’) and 1.3XRPRP (5’-GCACGGTCGAAACTTCCGTT-3’); 1.4FP (5’-GAACGGTTAACATGAAATTTCG-3’) and 1.4RP (5’-TCTTCCCTCTTTGTCAAGTGACG-3’); 1.5FP (5’-GCCCATTTATTAAGCTCTCTGTC-3’) and 1.5RP (5’-CTCAATATCAAAAGGAAAGATC-3’); and 1.7FP (5’-GAGGTGGAGAGAGAAGGACGG-3’) and 1.7RP (5’-GGGAATGATGCTGATGCTG-3’).

**Plasmodium falciparum NBP amplification and cloning.** *PfRBP2.* Primers used in *PfRBP2* amplification were as follows: 2.1FP (5’-GATGATCAATTATTTTAGCTTGAC-3’) and 2.1RP (5’-CAGAATGGGCTATCTCTGACG-3’); 2.2FP (5’-CAGAATGGGCTATCTCTGACG-3’) and 2.2RP (5’-CAGAATGGGCTATCTCTGACG-3’); 2.3FP (5’-GGAACACGTAGAGAAGAGGA-3’) and 2.3RP (5’-GGAACACGTAGAGAAGAGGA-3’); 2.4FP (5’-GGAACACGTAGAGAAGAGGA-3’) and 2.4RP (5’-GGAACACGTAGAGAAGAGGA-3’); and 2.5FP (5’-GGAACACGTAGAGAAGAGGA-3’) and 2.5RP (5’-GGAACACGTAGAGAAGAGGA-3’). For some samples, 2.3F was paired with 2.4R; 2.3F was paired with 2.4R; 2.3R was paired with 2.4R; 2.5F was paired with 2.5R; and 2.5R was paired with 2.5R. For some samples, 2.3F was paired with 2.4R; 2.3R was paired with 2.4R; 2.5F was paired with 2.5R; and 2.5R was paired with 2.5R. For some samples, 2.3F was paired with 2.4R; 2.3R was paired with 2.4R; 2.5F was paired with 2.5R; and 2.5R was paired with 2.5R. For some samples, 2.3F was paired with 2.4R; 2.3R was paired with 2.4R; 2.5F was paired with 2.5R; and 2.5R was paired with 2.5R.
RESULTS

Limitation of PfNBP1 diversity and clustered polymorphisms. The full-length 9-kb PfNBP1 gene was cloned in overlapping fragments from six P. falciparum strains of diverse geographical origin: FVO (Vietnam), 3D7 (Africa), Dd2 (southeast Asia), 7G8 (Brazil), HB3 (Honduras), and MC (Malaysia), and sequenced in its entirety. Each isolate encoded a distinct PfNBP1 haplotype and 40 polymorphic sites were identified comprising 31 non-synonymous and 9 synonymous substitutions (Table 1). Most of the observed polymorphisms were dimorphic, with one trinomorphic site, and polymorphisms tended to be clustered, with the 3′ region being the most conserved (Figure 1). Synonymous mutations are relatively rare in P. falciparum genes, a feature that has been used to argue that all extant P. falciparum populations share a recent common ancestor, yet 22.5% of the PfNBP1 polymorphisms were synonymous and eight of the nine synonymous mutations were clustered in a region spanning only 1.7 kb (Figure 1).

We previously reported that the 3D7 strain PfNBP1 gene contained a premature stop codon, the result of an additional adenine residue in a poly-adenine stretch near the 5′ end of the gene. This report was based on sequencing several independent clones of this region, all of which confirmed the additional adenine and was consistent with immunoblots that appeared to show a truncated PfNBP1 translation product in this strain, but others subsequently found no evidence for such an insertion. In generating the data for this paper, we sequenced multiple amplification products of this region generated using several different DNA polymerases. We observed that even so-called high-fidelity DNA polymerases have a tendency to slip at this poly-adenine stretch, and clones differing by ± 1–3 adenines were regularly observed with all polymerases, including those used in our original analysis and those used by our colleagues. However, the majority of sequences did not contain an additional adenine; thus, our final 3D7 PfNBP1 sequence is in agreement with that of Taylor and others. We also found no evidence for a premature termination codon in the 7G8 PfNBP1 gene, despite differences from immunoblots that the PfNBP1 translation product in this strain was also truncated. We are currently analyzing the post-translational-processing of PfNBP1 to see whether this differs between strains and could explain our previously reported immunoblot data.

The DNA diversity for PfNBP1, estimated using π, the average nucleotide diversity between sequences, and ω, which corrects for bias in transition-transversion ratios, was 0.00194 and 0.00195, respectively. A pair of PfNBP1 alleles therefore differed on average at only 0.19% of their nucleotide positions. Amino acid repeats that differ in copy number among isolates are common in Plasmodium proteins, and three were noted in PfNBP1 (Figure 1): an HN repeat near the C-terminus reported previously (x = 4, 6, or 9), and two others, Q(K/T)ₙ (x = 5, 7, 8, or 9), and DIDEINₙ (x = 3, 4, or 5). A variable stretch of poly-asparagine was also found with five of the strains having 5 Ns at this site, whereas the Brazilian strain 7G8 has 10.

High conservation of PfNBP2a and PfNBP2b. The PfNBP2a and PfNBP2b genes are related in an unusual manner, with nearly 8 kb of almost identical sequence (referred to as the shared domain), followed by a series of repeats and 3′ domains that are unique to each gene. To ensure that we were characterizing contiguous PfNBP2a or PfNBP2b genes, the full-length 9.5–10-kb genes were amplified separately by long-range PCR, gel purified, and cloned in fragments (see Materials and Methods). PfNBP2a and PfNBP2b cloned in this manner from four P. falciparum isolates, FVO (Vietnam), 3D7 (Africa), Dd2 (southeast Asia), and 7G8 (Brazil), were sequenced in their entirety and each strain represented a distinct haplotype. Twenty-seven polymorphic nucleotides were identified in PfNBP2a, with six in the series of repeats that marks the boundary between the shared and unique regions, while 23 polymorphisms were identified in PfNBP2b, none of which were in the repeat domain. Diversity in PfNBP2a and PfNBP2b is therefore slightly lower than in the PfNBP1 gene, with π values of 0.00159 and 0.00132, respectively (Table 1). In both cases, the polymorphic residues clustered at the 5′ end of the gene (Figure 1). All of the PfNBP2b polymorphisms were non-synonymous and only one PfNBP2a polymorphism was synonymous (Table 1). Although the repeats that mark the boundary between the shared and unique regions of these proteins vary in number and repeat pattern between strains, only one gene, PfNBP2a from the Brazilian isolate 7G8, had a repeat sequence that differed from the two versions already described. A poly-asparagine tract (7, 8, or 9 Ns) was the only other polymorphic region in the coding sequence.

Twenty-one of the PfNBP2a polymorphisms were in the region that is present in both genes (Figure 1) and 20 of these same polymorphisms were found in the PfNBP2b sequences. Remarkably for 18 of these polymorphic sites, both PfNBP2a and PfNBP2b in a given strain always shared the same nucleotides (Figure 1). For example, at amino acid position 249, both PfNBP2a and PfNBP2b from the Brazilian strain 7G8 encode an asparagine, while both genes from the other three strains all encode a lysine. Similarly, both of the genes in the

### Table 1

Diversity within the reticulocyte-binding-like (RBL)

<table>
<thead>
<tr>
<th>Gene</th>
<th>n</th>
<th>Sites</th>
<th>S</th>
<th>Nonsyn</th>
<th>Syn</th>
<th>ω (SD)</th>
<th>d (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PfNBP1</td>
<td>6</td>
<td>8853</td>
<td>40</td>
<td>31</td>
<td>9</td>
<td>0.00194 (0.0023)</td>
<td>0.00195 (0.0034)</td>
</tr>
<tr>
<td>PfNBP2a</td>
<td>4</td>
<td>9237</td>
<td>27</td>
<td>26</td>
<td>1</td>
<td>0.00159 (0.0033)</td>
<td>0.00159 (0.0030)</td>
</tr>
<tr>
<td>PfNBP2b</td>
<td>4</td>
<td>9606</td>
<td>23</td>
<td>22</td>
<td>0</td>
<td>0.00132 (0.00028)</td>
<td>0.00132 (0.00029)</td>
</tr>
<tr>
<td>PrRBP1</td>
<td>4</td>
<td>8499</td>
<td>28</td>
<td>25</td>
<td>3</td>
<td>0.00182 (0.00045)</td>
<td>0.00183 (0.00037)</td>
</tr>
<tr>
<td>PrRBP2</td>
<td>4</td>
<td>8454</td>
<td>646</td>
<td>489</td>
<td>151</td>
<td>0.04808 (0.01329)</td>
<td>0.05063 (0.00230)</td>
</tr>
</tbody>
</table>

* n = number of sequences sampled; Sites = sites analyzed excludes non-coding sequences and alignment gaps; S = number of sites that are polymorphic; Nonsyn = number of non-synonymous substitutions; Syn = number of synonymous substitutions; ω = average pairwise nucleotide diversity calculated using Jukes-Cantor correction with standard deviation of ω in parentheses; d = nucleotide diversity calculated using Tamura’s three-parameter model with standard error of d in parentheses; PfNBP = Plasmodium falciparum normocytic binding protein; PrRBP = P. vivax reticulocyte binding protein.
FVO and 7G8 strains encode a glycine at amino acid position 269, while both genes in the 3D7 and Dd2 strains encode a serine. Of the four polymorphic nucleotides in the shared regions that were not consistent between both genes in a single strain, three are present near the repeated motifs that mark the division between the shared and unique regions, and one is in the short exon 1, which encodes a signal sequence (Figure 1).

Limited polymorphism in *PvRBP1* and dimorphism with a high degree of polymorphism in *PvRBP2*. The complete *PvRBP1* gene was sequenced from four strains (Belem, Thai-NYU, Sal I, and Thai III). Each strain contained a distinct haplotype, with the Belem and Thai-NYU haplotypes being the most similar, differing at only three sites. Within the 8,499-basepair (bp) coding sequence there were 28 polymorphic sites, resulting in diversity estimates comparable to *PfNBP1*, *PfNBP2a*, and *PfNBP2b*, with \( \pi \) and \( d \) values of 0.00182 and 0.00183, respectively (Table 1). Twenty-five of the 28 polymorphisms were non-synonymous, and 18 of these (67% of the total polymorphisms) cluster within a 1,748-bp region (21% of the coding sequence) near the 5' end of exon 2, similar to the clustering of polymorphism observed in the *PfNBP2a* and 2b genes (Figure 1).

*PvRBP-2* exhibited a significantly greater degree of polymorphism compared with the other *RBL* genes. Only three haplotypes were identified in four complete *PvRBP2* sequences, with the Belem and Thai-NYU isolates containing identical alleles, despite their widely different geographic origins (Brazil and Thailand, respectively). To confirm that the identity between Belem and Thai-NYU *PvRBP1* sequences was not the result of sample error, other gene sequences known to differ between the Belem and Thai-NYU strains...
were amplified from the same sources of genomic DNA and found to differ in sequence as expected. Furthermore, haplotype patterns also suggest that Belem and Thai-NYU isolates are closely related. However, the Belem/Thai-NYU PvRBP2 allele differed radically from the PvRBP2 alleles found in the other two isolates, Sal I and Thai-III (derived from El Salvador and Thailand, respectively), which were similar to each other but not identical. At the deduced amino acid level, PvRBP2 sequences are 96.1% identical between Sal I and Thai-III, whereas the Belem/Thai-NYU PvRBP2 sequence is 83.5% and 84.7% identical to Sal I and Thai-III, respectively. The PvRBP2 sequences could therefore be split into two distinct classes, the Belem/Thai-NYU class and the Sal I/Thai-III class. This apparent dimorphism is shown in Figure 2, both by clustal alignment of the first 800-811 amino acid residues of the encoded PvRBP2 (Figure 2A) and schematically (Figure 2B). There are three regions that are conserved between the two classes, corresponding to nucleotide positions 1294-1574, 7054-7805, and 8367-8604 of the Belem coding sequence (unshaded regions in Figure 2B), but the majority of the encoded protein diverges significantly. Southern blot and other unpublished data have shown that the PvRBP2 gene is present as a single copy in the P. vivax genome (Belem strain) and in multiple strains of P. cynomolgi (Okenu DM and others, unpublished data), a simian parasite closely related to P. vivax. These data support the likelihood that the alleles we report here represent a single PvRBP2 gene, rather than sequences from two closely related but divergent PvRBP2 paralogs.

Overall, there were 640 polymorphic sites within the complete 8457-bp PvRBP2 coding sequence, all of which are dimorphic except for a single trimorphic site. Of these polymorphic sites, 489 were non-synonymous and 151 were synonymous. As a result, diversity values were significantly higher than in all other RBL genes (P < 0.005, by t-test), with \( \pi \) and \( d \) values of 0.04808 and 0.05063, respectively (Table 1). Although the Sal I and Thai-III alleles are much more similar to each other than they are to the Belem/Thai-NYU sequence, there are still 147 nucleotide differences between these two alleles, 107 of which are non-synonymous and 40 are synonymous. The differences between the Sal I and Thai-III PvRBP2 sequences tended to occur in the regions that were otherwise more conserved between all three haplotypes (Figure 2B). Overall, the Sal I and Thai-III PvRBP2 alleles differ at 1.7% of their nucleotide positions (\( \pi \) and \( d \) values = 0.01735 [SD = 0.00868] and 0.01759 [SD = 0.00157], respectively). As in PlnBP1, PvRBP2 repeated motifs that varied in number between isolates were also observed. PvRBP2 in the Sal I and Thai-III strains have one or two copies of a PXQKK motif near the N-terminus, a 24 amino acid deletion near the C-terminus, and a DTHD repeat near the C-terminus, while the Belem and Thai-NYU strains have seven copies of the DTHD repeat.

**Selection pressures on RBL genes.** The rates of non-synonymous and synonymous mutations can be used as a measure of whether genes are under selective pressure, with a predominance of non-synonymous mutations implying positive, diversifying selection, whereas a predominance of synonymous mutations would suggest negative, purifying selection. When normalized, the rate of non-synonymous substitutions in PvrBP1, PfnBP2a, and PfnBP2b were all significantly greater than the rate of synonymous substitutions, suggesting diversifying selection (Table 2). In contrast, the rates of synonymous and non-synonymous mutations in PfnBP1 and PvrBP2 were roughly equivalent (\( K_s/K_a \) ratios 1; Table 2), suggesting that they are not under significant selective pressure. The synonymous/non-synonymous substitution rates thus differ more significantly between RBL gene family members within each *Plasmodium* species than they do between the two species, implying that selective pressures are not uniform across RBL genes. However, because the occurrence of synonymous mutations is particularly low in *P. falciparum* and both non-synonymous and synonymous mutations were found to cluster in different RBL genes, implying that different selective pressures may be acting on different domains, these results should be interpreted cautiously.

Non-synonymous and synonymous substitution rates can also be used to infer selection pressures using the McDonald-Kreitman test, which is based on the hypothesis that under neutrality, synonymous and non-synonymous substitution rates should be the same both between and within species. Given the high divergence between the *P. falciparum* and *P. vivax* RBL genes direct comparisons between the RBL genes in these two species would not be informative. However, the homolog of PfnBP2b has recently been cloned from the chimpanzee parasite *P. reichenowi*, which is closely related to *P. falciparum*. There are 97 synonymous and 321 non-synonymous substitutions between PfnBP2b and PfnBP2b, compared with 0 synonymous and 23 non-synonymous substitutions within the four PfnBP2b haplotypes sequenced here. There are therefore a significant excess of non-synonymous substitutions within PfnBP2b sequences compared with between PfnBP2b and PfnBP2b (\( P = 0.0039, \) by Fisher’s exact test). This test implies that diversifying selection is acting on PfnBP2b. The *P. reichenowi* NBP1 homolog is a pseudogene, and only a fragment of the *P. reichenowi* NBP2a homolog has been cloned, so analysis was not possible for these genes. However, the homolog of PvrBP1 was characterized recently from *P. cynomolgi*, and although there is an excess of non-synonymous mutations within *P. vivax* RBP1 sequences compared with between PvrBP1 and PvrBP1, this excess is not significant (Okenu, NM and others, unpublished data).

**DISCUSSION**

The *P. falciparum* and *P. vivax* RBL gene sequences are reported here from a series of distinct geographic isolates, representing the first assessment of geographic diversity in the RBL gene family in human malaria parasites. Given that little is known about the relevant functional domains in these large proteins, this study used a comprehensive approach and analyzed full-length RBL sequences, totaling more than 185 kb of DNA sequence. By necessity, given the size of the genes in question, this study does not contain a large enough sample size to allow for the meaningful application of population genetic tests. However, it does show several important features of diversity in these genes, and paves the way for future more detailed analysis of specific sub-domains.

Polymorphism in merozoite invasion proteins is generally much lower than in merozoite surface proteins (MSPs), and is dictated by a balance between the opposite forces of positive selective pressure (such as by the immune system) and
Figure 2. Clustal X assisted alignment of the first 800 amino acids of the Plasmodium vivax reticulocyte binding protein 2 (PvRBP2) deduced peptides from A, four geographically diverse P. vivax isolates and B, a schematic illustration of the extent of the observed dimorphic polymorphism and interspecies-conserved regions of PvRBP2. Regions that diverge between the Sal I/Thai-III and the Belem/Thai-NYU allelic classes are indicated by gray shading, and the positions of nucleotide substitutions within the Thai-III/Sal I allelic class are noted by thin vertical lines. Del = deletion.
purifying functional constraints. Among *P. falciparum* genes encoding merozoite invasion proteins, *PfAMA1* has been the most studied for diversity, with π values of 0.01642 and 0.01402. Studies of *PfEBA-175* have concentrated on the region encoding the erythrocyte binding domain, Region II, with π values for Region II of 0.00366 and 0.003. The homologous region of *PfEBA-140/BAEBL*, an erythrocyte invasion antigen related to *PfEBA-175* that binds to glycophorin C, has a lower π value of 0.001. The *PfNBP* genes, with π values of 0.00132–0.00194, therefore appear less diverse than either *PfAMA1* or the region encoding the *PfEBA-175* binding domain, but are similar to the *PfEBA-140* Region II.

Fewer analogous studies have been performed using *P. vivax* genes. A large global study of a 461-bp fragment representing the more variable first half of *PvAMA1* produced a π value of 0.0174, similar to π values for *PfAMA1*. A recent study on the *PvDBP* gene analyzed diversity in distinct regions and found π values ranging from 0.0184 in the erythrocyte binding domain to 0.0086 in Region IV. The *PvRBP1* gene, with a π value of 0.00175, is 10-fold less diverse than *PvAMA1*, and 4-fold less diverse than the most conserved region of *PvDBP*. In contrast, *PvRBP2* has an extremely high π value of 0.04808, comparable with the variable block 4 of the *P. vivax* MSP1 (π value = 0.0451). However, of the four *PvRBP2* sequences analyzed, two are identical while the other two are much more closely related to each other than either is to the first pair, suggesting that two quite distinct *PvRBP2* allelic classes may exist, just as dimorphic alleles exist for both *PfAMA1* and *PfEBA-175* at the boundary between the shared and unique regions. Cross-allelic selection may reflect their importance in erythrocyte invasion and force of selective pressures acting on genes, which has been previously considered. Of the four mutations in the *PvRBP2a/b* shared region that are not identical in both genes in a given strain, it is notable that three of them are within 500 basepairs of the repeated sequences that mark the boundary between the shared and unique regions. Crossing over might be less favored at the unique regions because it is more likely to lead to genes with mixed unique domains. The fact that both *PvRBP2a* and *PvRBP2b* in *P. falciparum* persist in the *P. vivax* genome implies that the two unique domains have similar functions, so a mixed domain might have a significant functional impact. However, is not yet known whether *PvRBP2a* and *PvRBP2b* have distinct or redundant functions, since each can be knocked out independently and at least one strain of *P. falciparum* lacks one of the two genes, implying that even if they do have distinct functions, they are not both strictly essential.

Sequence analysis can provide information about the nature and force of selective pressures acting on genes, which can in turn provide information about function. Diversifying selection appears to be operating in *PvRBP2*, with higher rate of non-synonymous than synonymous substitution and a significant excess of non-synonymous substitutions compared with its ortholog in *P. reichenowi*. The same trend was seen in *PvRBP2a* and *PvRBP1*. Similar selection pressures operating on these genes may reflect their importance in erythrocyte invasion in both parasite species, and may be caused by the human immune response or co-evolution of these ligands with their erythrocyte receptors. Such conclusions could not be drawn significantly for the other *RBL* genes, given the limited number of sequences generated for each 7.7-10-kb *RBL* gene in this initial study, which was aimed primarily to assess the general level of diversity over the complete genes. This study provides a clear framework for future investigations involving more isolates, which will be necessary to generate sufficient sample size for detailed population genetic analyses.

The distribution of nucleotide polymorphisms across a gene can also provide clues about function. There is a cluster of eight synonymous mutations in a region of 1.7 kb near the middle of the *PfNBP2* gene, for example, which may indicate a domain under purifying selection pressure to minimize diversity. In *PvNBP2a*, *PvNBP2b*, and *PvRBP1*, the majority of the polymorphisms were found within the first 2 kb of exon 2, and most of those substitutions were non-synonymous. This pattern of polymorphism has interesting parallels with *PvDBP*, where Region II, which contains the erythrocyte-


24. Duraisingham MT, Triglia T, Ralph SA, Rayner JC, Barnwell JW, McFadden GI, Cowman AF, 2003. Phenotypic variation of


