Available online at www.sciencedirect.com
SCIENCE DIRECT
Molecular & Biochemical Parasitology 149 (2006) 10–16
MOLECULAR & BIOCHEMICAL PARASITOLOGY

Multi-character population study of the *vir* subtelomeric multigene superfamily of *Plasmodium vivax*, a major human malaria parasite

Emilio F. Merino [a,1], Carmen Fernandez-Becerra [a], Alan M. Durham [b], João E. Ferreira [b], Vanessa F. Tumilasci [a], Joana d'Arc-Neves [c], Monica da Silva-Nunes [a], Marcelo U. Ferreira [a], Thilan Wickramarachchi [d,e], Preethi Udagama-Randeniya [e], Shiroma M. Handunnetti [d], Hernando A. del Portillo [a,*]

[a] *Departamento de Parasitologia, Instituto de Ciências Biomédicas, Universidade de São, Paulo, Av. Lineu Prestes 1374, São Paulo, SP 05508-900, Brazil*
[b] *Departmento de Ciência da Computação, Instituto de Matemática e Estatística, Universidade de São Paulo, R. do Matão 1010, São Paulo, SP 05508-900, Brazil*
[c] *Instituto de Apoio a Pesquisa em Patologias Tropicais, Estrada BR 364 – Km 4,5, Porto Velho, Rondônia 78900-000, Brazil*
[d] *Institute of Biochemistry, Molecular Biology and Biotechnology, University of Colombo, Sri Lanka*
[e] *Department of Zoology, Faculty of Science, University of Colombo, Sri Lanka*

Received 9 October 2005; received in revised form 29 March 2006; accepted 4 April 2006
Available online 11 May 2006

Abstract

Plasmodium vivax, the most widely distributed human malaria parasite, contains the subtelomeric multigene *vir* superfamily corresponding to circa 10% of its coding genome. In this work, we used a multi-character strategy to study the *vir* gene repertoire circulating in natural parasite populations obtained directly from 32 human patients from endemic regions of Brazil and Sri Lanka. Cladistic analysis confirmed the existence of *vir* subfamilies, which varied in size and allele polymorphisms. Moreover, different motifs, protein domain, and secondary structures were predicted for each subfamily. Of importance, not all *vir* sequences possess a recognizable Pexel motif recently shown to be important, though not essential, signal for transportation to the cell membrane of infected red blood cells. Furthermore, subfamilies A and D display common structural features with the recently described *P. falciparum* SURFIN and *Pfmc-2tm* subtelomeric multigene families. These results suggest that VIR proteins can have different subcellular localizations and functions. This is the first study on a population level of the *P. vivax vir* subtelomeric multigene superfamily.

Keywords: Variant genes; *vir* subtelomeric multigene family; *Plasmodium vivax*; Malaria; Multi-character in silico analysis

1. Introduction

Malaria is a major human parasitic disease caused by four species of *Plasmodium* protozoa. Of these, *Plasmodium vivax* is the most widely distributed and responsible for 70–80 million clinical cases each year and large socio-economical burdens for countries such as Brazil where it is the most prevalent species [1]. Research on *P. vivax* remains largely neglected because it cannot be maintained under in vitro continuous culture conditions.

Malaria parasites have clustered multigene families in subtelomeric regions where high recombination rates facilitate their evolution and diversity [2]. The different subtelomeric multigene families vary in number, number of genes and organization depending on the parasite species. Thus, *P. falciparum* the most deadly human species contains, among others, the *var* [3], *rif* [4], *stevor* [5], *clag* [6], *Pf60* [7] and the more recently described *Pfmc-2tm* [8] and *surf* [9] multigene families. In contrast, *P. vivax* the most widely distributed human species, contains a major subtelomeric multigene superfamily termed *vir* (*P. vivax* variant genes), which corresponds to approximately 10% of coding sequences and is composed of different subfamilies (called A–F) predicted by sequence similarities [10]. Rodent malaria parasites display homologous *vir* multigene families termed *cir/bir/yir* [11], in addition to several other subfamilies [12].

* Corresponding author. Tel.: +55 11 3091 7209; fax: +55 11 3091 7417.
E-mail address: hernando@icb.usp.br (H.A. del Portillo).
[1] Present address: Department of Medical Parasitology, NYU, New York, NY 10010, USA.

0166-6851/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.molbiopara.2006.04.002

Some of these multigene families are species-specific, others are shared between some, though not all species, and some are common gene families. Indeed, it was recently proposed that *vir* genes should be included within a new variant gene superfamily (*Plasmodium* interspersed repeats, *pir*) together with *rif/stevor* in *P. falciparum*, *kir* in *P. knowlesi*, and the *cir/yir/bir* family in *P. chabaudi*, *P. yoelii* and *P. berghei* [12,13]. In spite of this knowledge, the function of most of these multigene families is presently unknown although in silico approaches, continuous in vitro culture systems and in vivo rodent models are facilitating molecular studies of some of them. In contrast, few studies on the fluxing repertoire of any of these multigene families in natural infections, in particular those from human malaria parasites, are presently available [14,15]. This information is essential to understand the diversity and evolution of these multigene families, which in turn will contribute to unveil chronicity in *Plasmodium*.

In this study, we carried out a multi-character analysis on *vir* gene sequences from natural parasite populations obtained directly from 32 human patients from two regions of the Brazilian Amazon (Porto Velho and Acrelandia) and one from Sri Lanka (Kataragama).

2. Materials and methods

2.1. Human patients and parasites isolates

P. vivax isolates used in this study were obtained from 28 patients from the Brazilian Amazon (States of Rondonia and Acre) and 4 from the village of Kataragama in Sri Lanka (Table 1). All patients consent to participate of this study and donated their blood. These studies received ethical clearance from the local and University of São Paulo Ethics Committee as well as from the Ethics committee of the Faculty of Medicine, University of Colombo, Sri Lanka.

2.2. vir genes

To study the *vir* gene repertoire, *vir* parasite genomic DNA was extracted using GFX Genomic Blood DNA Extraction kits

Table 1
Information of *P. vivax* samples

		***vir* genes**				
		Subf. A	**Subf. B**	**Subf. C**	**Subf. D**	**Subf. E**
Region	**ID MalDB**	Nº of clones	Nº of clones	Nº of clones	Nº of clones	Nº of clones
RONDÔNIA (BR)	MalDB36	24	13	27	32	31
	MalDB37	23	24	30	31	31
	MalDB39	30	24		32	
	MalDB41	16	20	26	21	30
	MalDB54	29	23	30	28	32
	MalDB55	25	7	30	31	
	MalDB56	28	8	28		32
	MalDB57	30	18	28		32
	MalDB58	19	5	20	28	31
	MalDB59	20	6		31	27
	MalDB60	6			27	
	MalDB61	30		30	31	26
	MalDB62	31		30	24	
	MalDB63				22	32
	MalDB64	30		25	27	20
	MalDB65	16	1	26	31	32
	MalDB66	27			32	
	MalDB80[a]	11	14	12	17	10
	MalDB81[a]	11	5	8	10	3
	MalDB100[a]	13	6	11	4	11
ACRE (BR)	MalDB176	16	14	13	11	9
	MalDB177	14		1	15	11
	MalDB178	16	16	15	6	
	MalDB179	16		14	16	
	MalDB180	16		6	3	
	MalDB181	13		3	7	
	MalDB182	16		15	16	12
	MalDB186	14		14		
SRI LANKA (ASIA)	MalDB187	16	14	8		14
	MalDB183	16		9	15	
	MalDB184	16		15	16	
	MalDB185	16		9	13	
TOTAL	**(32)**	**604 (31)**	**218 (17)**	**483 (27)**	**577 (28)**	**426 (13)**

Not sequenced. ( ) Denotes total number of *P. vivax* samples [16].

(Amersham Pharmacia, Piscataway, NJ) and stored at 4 °C. *vir* genes were amplified with degenerate *vir*-specific oligonucleotides representing subfamilies A–E. Primers for all subfamilies have been previously described [10,16]. We used annealing temperatures 5 °C below the calculated Tm value of the primer pairs. PCR products were cloned using in pGEM-T Easy Vector System (Promega) or TOPO TA Cloning Kit (Invitrogen). Sequencing reactions were performed with T3, SP6 and T7 (reverse) primers and the ABI PRISM BigDye terminator cycle sequencing kit version 3.1 (Applied Biosystems). Samples were resolved and analyzed in an ABI PRISM 3100 16-capillary DNA sequencer (Applied Biosystems).

2.3. Sequence analyses of vir genes

DNA sequences were deposited into ClinMalDB (http://clinmaldb.usp.br/) a field-research oriented relational clinical database to study human malaria. ClinMalDB is open source and integrates clinical histories of malaria patients, sequence information of coding genes, their automated analyses and versatile tools that can be used locally or through internet to relate this information.

Sequences were automatically processed to remove contaminants and low quality reads and to mask vector sequences using the EGene platform [17]. Briefly, the trace files were initially submitted to Phred [18] for basecalling and quality assignment. Then, sequences were sequentially submitted to a quality filter where accepted reads had to present at least 85 bases with a phred quality above 20 in a sliding window of 300 bp. After vector sequences were masked using default parameters, low quality and masked bases at both 5′ and 3′ ends were trimmed, reads with <300 bp were discarded, and sequences were checked for bacterial, yeast and human contamination and filtered when positive. Graphic outputs from some of these analyses can be seen at (http://clinmaldb.usp.br/pipes). GenBank accession numbers of *vir* subfamilies sequences corresponding to samples MalDB36, 37, 39, 41, 54–66, 176–187 are A: DQ160297–160865, B: DQ160866–161058, C: DQ161059–161510, D: DQ161511–162056, E: DQ162057–162458. GenBank accession numbers of *vir* sequences corresponding to samples MalDB80, 81, 100 are AY608742–608887.

After preprocessing, *vir* sequences were translated using the SIXPACK program, part of the EMBOSS package [19]. Amino acid sequences were aligned by DIALIGN [20]. Unrooted neighbor-joining trees were constructed using MEGA2 [21]. Position-specific variability in amino acid sequence and the overall pattern of conservation within whole protein sequences among members of the *vir* gene subfamilies was analyzed using the Plotcon program in the EMBOSS package. Overall (average) identity among sequences of the *vir* gene subfamilies was measured using the BLOSUM62 scoring matrix as implemented by the Infoalign program in the EMBOSS package. To find identical sequences within and between isolates, pairwise distance were calculated using MEGA2. Pairwise distance equal to 0 was considered identical. Consensus secondary structures of the protein sequences of each *vir* subfamily were generated by the JNET [22] and DSC [23] algorithms. Transmembrane regions (TM) were predicted using the SOSUI (http://sosui.proteome.bio.tuat.ac.jp/~sosui/proteome/sosuiG/sosuigsubmit.html) and TMHMM (http://www.cbs.dtu.dk/services/TMHMM-2.0) servers.

Conserved amino acid motifs were identified in the protein sequences of each *vir* subfamily using the MEME and MAST programs [24,25]. Hidden Markov Models (HMMs) were used to develop a profile to predict *vir* genes. HMMs were built from alignments generated by DIALIGN. HMMbuild, part of the HMMER 2.2 package was used to generate the HMMs [26]. HMMs and MEME motifs were used to search putative related sequences against sequences from *P. falciparum*, *P. yoelii*, *P. chabaudi*, *P. berghei* obtained from ftp://ftp.sanger.ac.uk/pub/pathogens/P_knowlesi/peptides/falc.chab.berb.Yoel.proteins.fasta (release date: 28/7/2004) and from the non-redundant (nr) database available from NCBI at ftp://ftp.ncbi.nlm.nih.gov/blast/db/ (release date: 24/7/2005). All searches were performed locally. Gblocks program [27] was used to remove divergent regions from the global alignment of each subfamily as to locate conserved blocks. HMMs produced from these blocks were compared with the HMMs obtained from the global alignment (including variant domains). To find the Pexel motif [28] (strict: [KR][GAVLIMFWPSTCYNQ][LI][GAVLIMFWPSTCYNQ][DEQ] and the Pexel-like motif: [KR].[LI].[DEQ]), an in-house Perl program was written. All HMMs, MEME motifs and scripts are freely available upon request.

3. Results

3.1. vir genes contain different non-redundant subfamilies in field isolates

We analyzed the *vir* gene repertoire circulating in natural parasite populations obtained directly from 32 human patients from endemic regions of Brazil and Sri Lanka (Table 1). To do so, genomic DNA was extracted from parasites of these patients and used as templates in PCR using degenerate *vir*-specific oligonucleotides representing subfamilies termed A–E. These subfamilies were predicted in silico by similarity analysis of the sequences from the telomeric YAC used in the original description of this multigene superfamily [10]. Amplified fragments were cloned into bacterial vectors and 1–32 clones from each amplification were fully sequenced generating ~1.3 Mbp of *vir* gene sequences. Contaminants and low quality reads were automatically eliminated. Remaining 2308 sequences corresponded to 604 clones amplified with oligonucleotides of subfamily A, 218 clones with primers of subfamily B, 483 clones with primers of subfamily C, 577 clones with primers of subfamily D and 426 clones with primers of subfamily E (Table 1).

As these oligonucleotides were highly degenerate, it was of relevance to determine if these sequences indeed represented different subfamilies. To do so, cladistic analysis was performed in a subset of 1883 translated sequences comprising 508 sequences from subfamily A, 185 from subfamily B, 400 from subfamily C, 441 from subfamily D and 349 from subfamily E. Remaining sequences revealed the presence of an in-frame stop codon likely

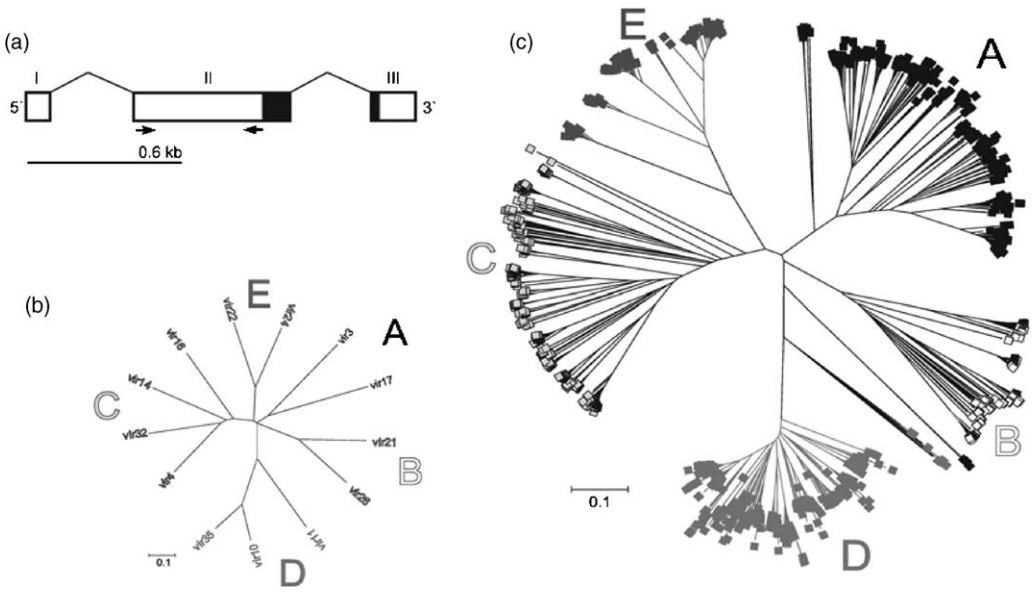

Fig. 1. (a) Schematic of the consensus three-exon *vir* gene structure. Arrows represent relative localization of primers used for amplifications. (b) Bootstrap consensus tree of *vir* amino acid sequences predicted from *P. vivax* subtelomeric YAC clone IVD10 [10]. (c) Bootstrap consensus tree of deduced *vir* amino acid sequences from *P. vivax* isolates based on pairwise distances. A–E: *vir* subfamilies. Scales at the bottom of (a and b) represent distance scale.

representing pseudogenes as previously noticed [10] and were excluded from these analysis. Results confirmed the existence of *vir* A–E subfamilies in natural parasite populations of *P. vivax* (Fig. 1). Moreover, similarity plots for each subfamily showed that subfamilies A–C were highly polymorphic with variant and conserved blocks distributed along their sequences. In contrast, subfamily D showed two large highly conserved blocks and a highly variable block, whereas subfamily E showed a large conserved N-terminus block and a polymorphic C-terminus block (Fig. 2). Overall *vir* genes showed ∼10% sequence identity at the amino acid level; yet, when considering individual subfamilies the identity percentages increased to ∼47% for subfamily A, ∼50% for B, ∼41% for subfamily C, ∼63% for subfamily D and 56% for subfamily E (Table 2).

3.2. Predicted motifs, structural domains and secondary structures of VIR subfamilies

We next used different algorithms to predict motifs, structural domains and secondary structures among *vir* subfamilies. Firstly, we searched for the Pexel motif which was recently shown to be involved in trafficking of malarial variant proteins to the infected red blood cell membrane [28], and found that 38% of all sequences contained it. This percentage was increase to 69% (1298 sequences) corresponding to 41% subfamily A, 76% subfamily B, 95% subfamily C, 48% subfamily D and 100% subfamily E, when the search was made for a Pexel-like motif. Secondly, multilevel consensus sequences corresponding to other conserved motifs of the different *vir* subfamilies were found using the MEME program (Supplementary Table SI and Fig. 2). Thirdly, we used the SOSUI and TMUMM programs to predict TM domains within the portion of exon II PCR-amplified here and which excludes the predicted TM domain common to all *vir* subfamilies [10]. Only subfamily D displayed two other transmembrame domains (Fig. 2D). In addition, this subfamily shared consensus structural characteristics with a new described multigene family of *P. falciparum*, *Pfmc-2tm* [8]. These include an N-terminal weakly hydrophobic region, a conserved pair of cysteine residues a lysine-rich C-terminal tail after the second TM segment, and proline residues in at least one of the predicted TM domains (Supplementary Fig. SI). Conserved structural characteristics between VIRs and SURFINS, another recently described subtelomeric multigene family of *P. falciparum* [9], were also observed.

SURFIN proteins were originally described as a subtelomeric multigene family of *P. falciparum* having structural features shared with exported proteins of several *Plasmodium* species, including *virs*, *SICAvar*, *pvstpl*, Pf332 and PfEMPl [9]. Our analysis extended these observations by showing that only members of subfamily A share these structural features over the same N-terminal region, which includes a cysteine-rich putative globular domain. Fourthly, secondary structure predictions showed that VIR proteins consist of a string of sparsely distributed alpha helices (numbering 5–9), averaging 2–26 amino acids in length, separated by coiled-coil regions and that beta strand predicted

Table 2
Overall and region-specific percentage amino acid identities of 1883 *vir* gene sequences from subfamilies A–E in isolates of Brazil (Rondonia—RO and Acrelandia—AC) and Sri Lanka (Kataragama—KT)

	RO	AC	KT	Overall
Subfamily A	46.75	45.63	49.16	46.72
Subfamily B	49.06	54.21	50.43	49.99
Subfamily C	39.52	44.26	40.39	40.44
Subfamily D	62.83	62.40	63.18	62.81
Subfamily E	56.22	53.37	57.69	56.04
All subfamily	9.40	9.83	10.07	9.52

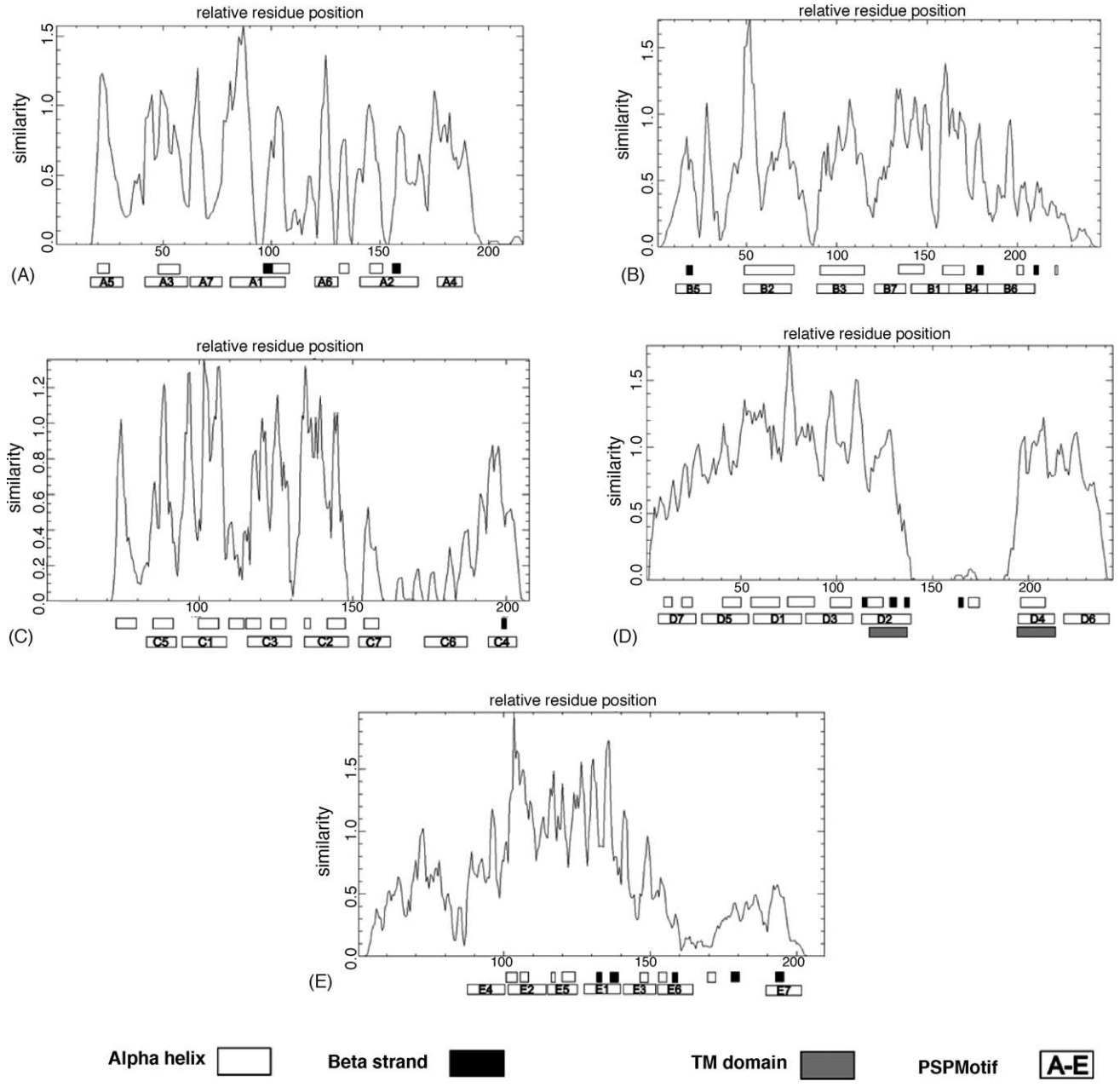

Fig. 2. Plots of sequence conservation in alignments of amino acid sequences of: (A) 508 *vir* subfamily A sequences, (B) 185 *vir* subfamily B sequences, (C) 400 *vir* subfamily C sequences, (D) 441 *vir* subfamily D sequences and (E) 349 *vir* subfamily E sequences, generated by Plotcon using default parameters. Predicted conserved secondary structures are given below each plot. Conserved sequence motifs generated by MEME (Supplementary Table SI), are shown as open boxes identifying subfamily and motif number. Predicted TM domains are shown in grey boxes.

also varied between the subfamilies being the subfamily E the most beta strand rich (Fig. 2).

3.3. Probabilistic vir models predict the malaria multigene pir superfamily

Next we took advantage of this large number of *vir* sequences to generate HMMs models and discover MEME motifs based on complete alignments for each subfamily and searched the nr and GeneDB databases for related sequences. In the nr database, HMMs and MEME motifs for each subfamily only found similar sequences with significant *E*-values from its own subfamily except for a few instances in which HMMs from subfamily E identified related sequences in subfamily C, and vice versa. In contrast, in the GeneDB database, orthologous genes included in the recently described *pir* gene superfamily were identified from all subfamilies albeit with expected low *E*-values. To refine these results, new HMMs and MEME motifs were generated using conserved blocks from the different *vir* subfamilies as defined by the default parameters of the Gblocks program [27]. Of notice, using these parameters no conserved blocks of *vir* subfamily A were found in contrast to the other subfamilies (B: 128 aa distributed in four blocks, C: 45 aa in three blocks, D: 94 aa in two blocks and E: 94 in one block). In spite of using conserved

blocks for these analyses, we observed similar results as those recorded from complete alignments. In fact, HMMs from complete aligments of subfamily C predicted *pir* gene members with higher *E*-values than any other HMMs. We also found sequences corresponding to *surf* genes from HMMs of subfamily A, in contrast to HMMs from subfamily D that failed to detect sequences corresponding to *Pfmc-2tm* genes. Similar results were observed using MAST or MEME motifs. All together, these results firmly established that there is an abundant fluxing repertoire of *vir* genes in natural infections grouped into different subfamilies some of which share common structural domains and motifs with other multigene families of *Plasmodium*. Local alignment from all *vir* subfamilies can be obtained in supplementary information (Supplementary Fig. S2).

4. Discussion

Unlike falciparum and rodent malarias from which in vitro culture systems and in vivo models allow studies to try to unveil the function of multigene families, *P. vivax* cannot be continuously cultured in vitro and studies on nonhuman primates are associated with high cost and availability. Moreover, although orthologs of *vir* genes have been described in rodent malaria parasites [11], knowledge on the fluxing repertoire of human malaria multigene families in natural infections is essential to understand their diversity and evolution. Thus, we decided to use a multi-character analysis to further our knowledge on *vir* genes. We generated a total of 2308 *vir* gene sequences of parasites obtained from 32 human patients from two different countries in different continents. Our studies clearly demonstrated a large repertoire of *vir* genes among parasites circulating in these human patients from three different geographical regions. These results are even more compelling if considering that due to the highly variant nature of these genes we were able to design degenerate oligonucleotides covering only 45% of the *vir* gene sequences originally described [10]. Unfortunately, their elevated variability precluded the use of similarity analysis to design more "universal" *vir* primers. Of notice, individual open reading frame analysis revealed that 425 sequences contained in-frame stop codons likely representing pseudogenes. The presence of such large number of pseudogenes in natural infections is very common among multigene families of parasitic protozoa [29–33]. These results thus confirmed and expand that there is a large fluxing repertoire of *vir* genes in natural infections.

Our results also confirmed the original in silico prediction that *vir* genes are organized into different subfamilies [10]. Interestingly, only genes from subfamily A had common structural domains with the recently described *surf* multigene family of *P. falciparum*, whose members can be located at the surface of the merozoites and infected red blood cells [9]. Moreover, domain structural analysis of members from subfamily D demonstrated that it displays many common features with other multigene family of *P. falciparum* termed *Pfmc-2tm*, encoding two transmembrane domains and being localized in Maurer's clefts [8]. Furthermore, in silico analysis predicted that 585 sequences lack the Pexel motif recently shown to be involved in transporting different malarial variant proteins to the surface of infected red blood cells [28]. It is important to recall however, that for these analyses only 45% of the *vir* gene repertoire found on a chromosome end and corresponding to the second variant exon were used [10]. With the imminent release of the complete genome sequence of the *P. vivax* Sal I strain (http://www.tigr.org/conf/vivax/index.shtml), it will be of relevance to relate these findings and subfamilies with complete *vir* genes from this reference strain. Regardless, these results suggest that not all VIR proteins are exported to the surface of infected red blood cells as originally reported using peptide antibodies of subfamily D [10]. If so, final functional annotation and classification of *vir* genes will have to take into account their subcellular localization in addition to predicted motifs and domain structures. Most important, under this scenario it is likely that *vir* genes have different functions related to immune evasion [16,34,35].

Probabilistic models have been recently used to predict a malaria multigene superfamily termed *Plasmodium* interspersed repeats (*pir*) including *rif/stevor* genes from *P. falciparum*, *kir* genes from *P. knowlesi*, *cir/bir/yir* genes from rodent malaria parasites, and *vir* genes from *P. vivax* [12,13]. We construct HMMs from complete alignments or conserved G-blocks for *vir* gene subfamilies based on 1883 sequences representing the highly variable exon II and found that these models can identify their own subfamilies in the nr database with significant *E*-values. Moreover, these models predicted *pir* gene subfamily members albeit with not significant *E*-values interspersed with other coding sequences in malarial databases other than *P. vivax*. Similar results were found when we used MEME motifs to search for related sequences in the nr and malarial databases. Of interest, HMMs based on complete alignments from *vir* subfamily C were the best models to predict *yir/cir/bir* and *rif* genes suggesting closer structural elements of this vivax subfamily to rodent and falciparum malaria parasites. These HMMs and MEME motifs can thus be used with the complete genome sequence of *P. vivax* to try to develop more accurate models and algorithms capable of predicting *vir* genes and orthologs.

In conclusion, this work represents the first study on a population level of the *P. vivax vir* subtelomeric multigene superfamily. Unfortunately, very few of such studies are presently available for other multigene families of *P. falciparum* [14,15] and none are available for multigene variant families of Trypanosomes and *Giardia*. Insight into these different repertoires in natural infections will be essential to understand the evolution of persistence in human diseases caused by parasitic protozoa to ultimately and hopefully lead to alternative control strategies.

Acknowledgements

We are particularly grateful to the anonymous reviewers of this manuscript whose suggestions improved significantly the quality of it. To Professor Luiz Hildebrando Pereira da Silva for facilities at IAPEPATRO. To Dr. Osmar de Souza for helpful discussion on protein structure, and to Ms. Ariane Machado Lima for helpful discussion on Perl programming. To Daniele Barbosa and Marcio M. Yamamoto for technical assistance. To G.M. Kapilananda for samples from SL. EFM was a post-doctoral fel-

low from FAPESP (02/11350-8). CFB is a ProDoc from CAPES (AUX-PRODOC-711/2002). The laboratory of HAP receives support from FAPESP (01/09401-0) and CNPq (302572/2002-3).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.molbiopara.2006.04.002.

References

[1] Mendis K, Sina BJ, Marchesini P, Carter R. The neglected burden of *Plasmodium vivax* malaria. Am J Trop Med Hyg 2001;64:97–106.

[2] Scherf A, Figueiredo LM, Freitas-Junior LH. Chromosome structure and dynamics of *Plasmodium* subtelomeres. In: Waters AP, Janse CJ, editors. Malaria parasites: genomes and molecular biology, vol. 1. Caister Academic Press; 2004. p. 187–204.

[3] Su XZ, Heatwole VM, Wertheimer SP, Guinet F, Herrfeldt JA, Peterson DS, et al. The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. Cell 1995;82:89–100.

[4] Weber JL. Interspersed repetitive DNA from *Plasmodium falciparum*. Mol Biochem Parasitol 1988;29:117–24.

[5] Cheng Q, Cloonan N, Fischer K, Thompson J, Waine G, Lanzer M, et al. stevor and rif are *Plasmodium falciparum* multicopy gene families which potentially encode variant antigens. Mol Biochem Parasitol 1998;97:161–76.

[6] Holt DC, Gardiner DL, Thomas EA, Mayo M, Bourke PF, Sutherland CJ, et al. The cytoadherence linked asexual gene family of *Plasmodium falciparum*: are there roles other than cytoadherence? Int J Parasitol 1999;29:939–44.

[7] Carey B, Bonnefoy S, Guillotte M, Le Scanf C, Grellier P, Schrevel J, et al. A large multigene family expressed during the erythrocytic schizogony of *Plasmodium falciparum*. Mol Biochem Parasitol 1994;68:221–33.

[8] Sam-Yellowe TY, Florens L, Johnson JR, Wang T, Drazba JA, Le Roch KG, et al. A *Plasmodium* gene family encoding Maurer's cleft membrane proteins: structural properties and expression profiling. Genome Res 2004;14:1052–9.

[9] Winter G, Kawai S, Haeggstrom M, Kaneko O, von Euler A, Kawazu S, et al. SURFIN is a polymorphic antigen expressed on *Plasmodium falciparum* merozoites and infected erythrocytes. J Exp Med 2005;201:1853–63.

[10] del Portillo HA, Fernandez-Becerra C, Bowman S, Oliver K, Preuss M, Sanchez CP, et al. A superfamily of variant genes encoded in the subtelomeric region of *Plasmodium vivax*. Nature 2001;410:839–42.

[11] Janssen CS, Barrett MP, Turner CM, Phillips RS. A large gene family for putative variant antigens shared by human and rodent malaria parasites. Proc R Soc Lond B Biol Sci 2002;269:431–6.

[12] Hall N, Karras M, Raine JD, Carlton JM, Kooij TW, Berriman M, et al. A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses. Science 2005;307:82–6.

[13] Janssen CS, Phillips RS, Turner CM, Barrett MP. *Plasmodium* interspersed repeats: the major multigene superfamily of malaria parasites. Nucleic Acids Res 2004;32:5712–20.

[14] Kirchgatter K, Mosbach R, del Portillo HA. *Plasmodium falciparum*: DBL-1 var sequence analysis in field isolates from central Brazil. Exp Parasitol 2000;95:154–7.

[15] Fowler EV, Peters JM, Gatton ML, Chen N, Cheng Q. Genetic diversity of the DBLalpha region in *Plasmodium falciparum* var genes among Asia-Pacific isolates. Mol Biochem Parasitol 2002;120:117–26.

[16] Fernandez-Becerra C, Pein O, de Oliveira TR, Yamamoto MM, Cassola AC, Rocha C, et al. Variant proteins of *Plasmodium vivax* are not clonally expressed in natural infections. Mol Microbiol 2005;58:648–58.

[17] Durham AM, Kashiwabara AY, Matsunaga FT, Ahagon PH, Rainone F, Varuzza L, et al. EGene: a configurable pipeline generation system for automated sequence analysis. Bioinformatics 2005;21:2812–3.

[18] Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred I. Accuracy assessment. Genome Res 1998;8:175–85.

[19] Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. Trends Genet 2000;16:276–7.

[20] Morgenstern B, Freeh K, Dress A, Werner T. DIALIGN: finding local similarities by multiple sequence alignment. Bioinformatics 1998;14:290–4.

[21] Kumar S, Tamura K, Jakobsen IB, Nei M. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 2001;17:1244–5.

[22] Cuff JA, Barton GJ. Application of multiple sequence alignment profiles to improve protein secondary structure prediction. Proteins 2000;40:502–11.

[23] King RD, Saqi M, Sayle R, Sternberg MJ. DSC: public domain protein secondary structure predication. Comput Appl Biosci 1997;13:473–4.

[24] Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol 1994;2:28–36.

[25] Bailey TL, Gribskov M. Combining evidence using *p*-values: application to sequence homology searches. Bioinformatics 1998;14:48–54.

[26] Eddy SR. Profile hidden Markov models. Bioinformatics 1998;14:755–63.

[27] Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 2000;17:540–52.

[28] Marti M, Good RT, Rug M, Knuepfer E, Cowman AF. Targeting malaria virulence and remodeling proteins to the host erythrocyte. Science 2004;306:1930–3.

[29] Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 2002;419:498–511.

[30] Carlton JM, Angiuoli SV, Suh BB, Kooij TW, Pertea M, Silva JC, et al. Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. Nature 2002;419:512–9.

[31] Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, et al. The genome of the African trypanosome *Trypanosoma brucei*. Science 2005;309:416–22.

[32] El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, et al. Comparative genomics of trypanosomatid parasitic protozoa. Science 2005;309:404–9.

[33] El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science 2005;309:409–15.

[34] del Portillo HA, Lanzer M, Rodriguez-Malaga S, Zavala F, Fernandez-Becerra C. Variant genes and the spleen in *Plasmodium vivax* malaria. Int J Parasitol 2004;34:1547–54.

[35] Suwanarusk R, Cooke BM, Dondorp AM, Silamut K, Sattabongkot J, White NJ, et al. The deformability of red blood cells parasitized by *Plasmodium falciparum* and *P. vivax*. J Infect Dis 2004;189:190–4.