

Francis B. Ntumngia · Noemi Bahamontes–Rosa  
Jürgen F. J. Kun

## Genes coding for tryptophan-rich proteins are transcribed throughout the asexual cycle of *Plasmodium falciparum*

Received: 11 March 2005 / Accepted: 7 April 2005 / Published online: 28 May 2005  
© Springer-Verlag 2005

**Abstract** Multigene families are a common feature in *Plasmodia* spp. and constitute a substantial content of the parasite genome. Here, we analyse the structural organisation and sequence diversity of two further members of the Trp-rich multigene family of *P. falciparum*. The complete DNA sequence of both genes was determined from a series of laboratory adapted and field isolates. Based on the amino acid sequences, we have termed them tryptophan-rich antigen-3 (TrpA-3) and lysine-tryptophan-rich antigen (LysTrpA). Analysis of the genes using reverse transcriptase-polymerase chain reaction (RT-PCR), showed that both genes are transcribed and that introns are spliced out at predicted positions. Gene expression profiles obtained from microarray analysis indicate that both genes are expressed in the mid-stages of the asexual cycle. In-frame stop codons were detected which interrupted the reading frame of LysTrpA. Whereas the number of the Trp-rich proteins is rather low in *P. falciparum*, *P. chabaudi*, *P. berghei* and *P. yoelii*, this family seems to have 15 or more members in *P. knowlesi* and *P. vivax*.

**Keywords** Tryptophan-rich protein · Multigene family · Pseudogene · *Plasmodium falciparum* · Malaria · *Plasmodium vivax*

### Introduction

*Plasmodium falciparum* has a complex life cycle with alternating development in mosquitoes and vertebrate hosts. This complex multi-step multiple-host cycle necessitates adequate gene sets and appropriately regulated expression. It is hoped that a better understanding of the organisation and expression constraints of the genes within its genome will nurture the rational design of novel intervention strategies, which is greatly needed in view of the rapidly deteriorating efficiency of existing control tools.

Advances in the fields of genomics, proteomics and molecular immunology offer tremendous opportunities for the development of novel interventions against this public health threat. Today, researchers can make use of a huge amount of genome sequence data accompanied by expression data that might pave the way for the hoped-for anti-malarial therapies (Bozdech et al. 2003; LeRoch et al. 2003; Hall et al. 2005).

Recent advances in the genomics of autonomously living organisms have highlighted the widespread occurrence of multigene families, which account for up to 50% of certain genomes (Brenner et al. 1995; Labedan and Riley 1995; Tatusov et al. 1997). The diversification of individual genes within gene families provides a set of related molecules with a large array of recognition properties, specific functions or expression profiles, facilitating the fine tuning of biological activity according to specific cellular developmental or environmental conditions (Henikoff et al. 1997). These multigene families constitute a substantial part of the overall genome content in bacteria (Blattner et al. 1997) as well as lower and higher eukaryotes (Rubin et al. 2000). Multigene families are usually a collection of related genes that are presumed to share a common ancestor and might have originated from each other by duplication and subsequent divergence through gene recombination. These sets of genes are known to optimise fitness by providing a substantial part of the overall genome content of the organism. The

The sequence data reported in this paper are available in the GenBank data base under the accession numbers AY898647 and AY898648 for TrpA-3 and LysTrpA respectively

Francis B. Ntumngia and Noemi Bahamontes–Rosa contributed equally to this work

F. B. Ntumngia · N. Bahamontes–Rosa · J. F. J. Kun (✉)  
Department of Parasitology, Institute of Tropical Medicine,  
Wilhelmstr. 27, 72074 Tübingen, Germany  
E-mail: juergen.kun@uni-tuebingen.de  
Tel.: +49-7071-2982191  
Fax: +49-7071-294684

number and the size of multigene families vary greatly from one organism to another (Rubin et al. 2000). With most genomes, the majority of the families contain few (up to five) members (Salanoubat et al. 2000). There are however examples of large multigene families with more than 50–100 members, including the large multigene families encoding surface antigens in plasmodia species (Su et al. 1995; Kyes et al. 1999; Portillo 2001) or apical organelle proteins (Owen et al. 1999; Preiser et al. 1999; 2002).

While functional divergence of duplicated copies can occur, some copies can result in a loss of function by the generation of stop codons or a gradual accumulation of disabling mutations (Little 1982; Wilde 1986), thereby suffering the less glamorous fate of fixing a null allele and becoming a pseudogene. Pseudogenes usually act as a reservoir of sequences that recombine with their functional paralogous genes and thus generate antigenic diversity (Restrepo 1994; Balakirev et al. 2003).

In a previous study, an antigen was identified in membrane fractions of *P. falciparum*-infected erythrocytes (Kun et al. 1991). Sequence analysis of the full gene revealed a protein rich in Trp residues and with a domain containing threonine repeats, which was termed TryThrA (Uhlemann et al. 2001). A blast search of the *P. falciparum* genome database PlasmoDB based on gene structure and amino acid sequence identified related sequences on chromosomes 1, 4, 10 and 13 (<http://plasmodb.org>). The respective gene product on chromosome 1 has recently been characterised as a merozoite-associated protein (Ntumngia et al. 2004). The *P. falciparum* Trp-rich antigens were envisaged to be homologues of the *P. yoelii* secreted antigens which have been shown to be successful in vaccine studies in mice (Burns et al. 1999; Burns et al. 2000). The level of identity on the amino acid level is rather low and limited to the Trp-rich domains. In the present study, we describe the characterisation of the sequences on chromosomes 10 and 13. On the basis of their gene structure and amino acid composition, we have termed them tryptophan-rich antigen-3 (TrpA-3) and lysine-tryptophan-rich antigen (LysTrpA) respectively.

## Materials and methods

### Parasites materials and cultures

The parasites used in this study were three *Plasmodium falciparum* laboratory strains (the Kenyan isolate Binh1, the FCR-3 strain isolated in The Gambia and the Gabonese isolate, Cys007) and a series of field isolates from malaria patients attending the Albert Schweitzer Hospital in Lambaréné, Gabon. The laboratory-adapted parasites were maintained in continuous culture using standard methods with erythrocytes ( $0^+$ ) at 5% hematocrit and AB<sup>+</sup> Rh<sup>+</sup> serum in buffered medium (Cranmer et al. 1997).

### PCR and RT-PCR

Genomic DNA and total RNA were extracted from asynchronous cultures and blood samples of field isolates using the QIAamp DNA blood kit and the RNeasy mini kit (Qiagen, Hilden, Germany) respectively as described in the user manuals. The respective genomic DNAs from all the isolates and total RNA from the Binh1 strain were used as templates for PCR and reverse transcriptase-PCR (RT-PCR) respectively. Sequence data from the database PlasmoDB (<http://plasmodb.org>), were used to design specific oligonucleotide primers for PCR and RT-PCR. The TrpA-3 gene was divided into four different regions and the various fragments were PCR amplified using the following primer sets:

F1: 5'-ATG CAA ATT AAT CCA ATG GAA-3';  
 RI: 5'-TTC TAT ATC TTT TCT TGT-3';  
 F2: 5'-CAA CTT ATG ACA TTA GAA-3';  
 R2: 5'-TCC ATT ATT GTC GAA TTT ATT-3';  
 F3: 5'-TAT ACT CAT GAT GAT TTG TAT-3';  
 R3: 5'-CAC ACG CTC CAT TGT TTT TCC-3';  
 F4: 5'-CAA TGG AGC GTG TGG TTA GAA-3'; and  
 R4: 5'-TTA TTC TTG GTT TCT AAT TAA-3'

while for LysTrpA, the gene sequence was divided into four overlapping blocks by using the corresponding primer combinations BaF/BaR, T4F/T4R, May/Ros4 and Fb3/Ros4. The primers used were:

BaF: 5'-CTG GAA CGA GAA GAA GAA GG-3';  
 BaR: 5'-CCT TTT CTT TAT GAC TGC-3';  
 T4F: 5'-CAT AAG ATG CTA CCA CAA AAA-3';  
 May: 5'-CAT AAA TGC AGT CAT AAA-3';  
 Fb3: 5'-TAT AAT CAA GAT GTG AAT-3';  
 Ros4: 5'-ATA TTC AAA CAA AAC AGC TAG-3'; and  
 T4R: 5'-TTC ATT GAT ATT TAC ATT CAC-3'.

All amplifications were initiated with a denaturing step of 94°C for 5 min, followed by 35 cycles of 94°C for 1 min, 1 min annealing time (the temperature was adjusted to the  $T_m$  of the oligonucleotides used) and 1.5 min at 72°C for elongation. The reaction was stopped after a final elongation step of 72°C for 10 min. After purification, the respective PCR products were sequenced with internally generated specific primers using BigDye terminator chemistry on an ABI PRISM 3100 Genetic analyser (Applied Biosystems, Foster City, USA). The sequences were analysed for polymorphisms after DNA alignment of the sequenced products using the Bioedit alignment program (North Carolina State University, USA). The quality of the individual SNPs were assessed by visual inspection with reference to the electropherograms. cDNA was synthesised from total RNA by RT-PCR using the Titan one Tube RT-PCR kit (Roche, Mannheim, Germany) according to the manufacturer's instructions. The respective products were also sequenced as described above.

## Gene expression profiles

The gene expression profile of the five members of the Trp-rich multigene family was done using microarray data generated by Bozdech et al. (2003). We generated a heat map using the Genesis 1.2.2 software (<http://genome.tugraz.at>).

## Results

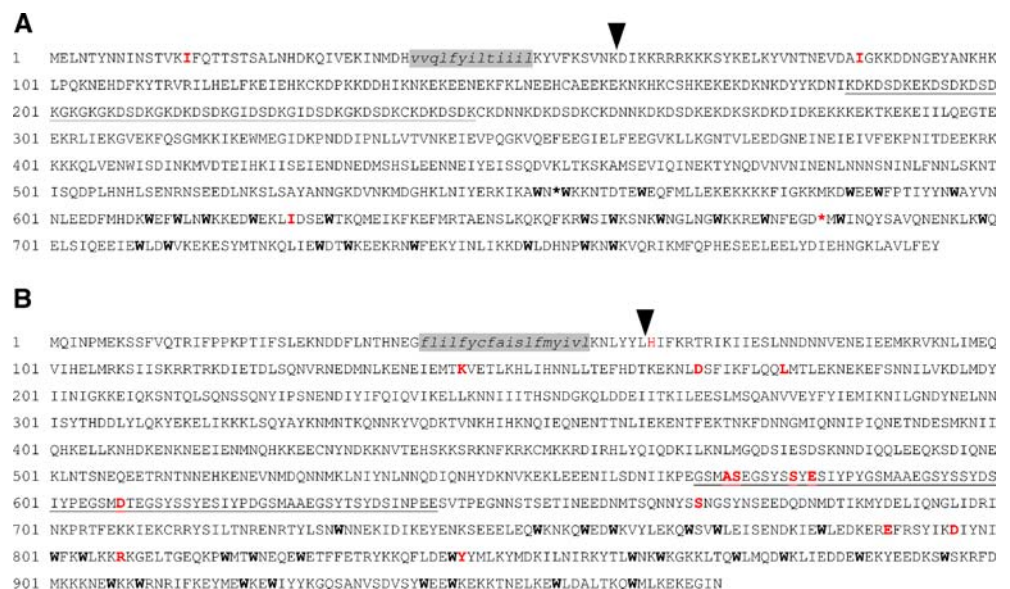
### Genomic organisation and polymorphism

We sequenced genes for TrpA-3 and LysTrpA from three laboratory-adapted isolates and seven field isolates from Lambaréné, Gabon. The complete nucleotide sequence of TrpA-3 in the Binh1 isolate consists of 3,075 nucleotides and contains an open-reading frame, which encodes for a predicted protein with 971 amino acid residues and a molecular weight of about 107 KDa, while the complete gene sequence in LysTrpA consists of a nucleotide sequence of 2,543 base pairs and an open reading frame encoding a predicted protein with 797 amino acid residues and a predicted molecular weight of about 88 KDa (Fig. 1). Total RNA was isolated from an asynchronous culture which was used to produce cDNA by RT-PCR. Transcription of both genes was shown by analysis of the PCR products generated from cDNA and genomic DNA which revealed that short introns of 159 bps and 143 bps at the 5' end were spliced out of the mature messages from both TrpA-3 and LysTrpA genes respectively (Fig. 3a, b). Towards the 3' end of LysTrpA, at the beginning of the Trp-rich domain (at position 1,812), a stop codon is situated. To investigate the possibility that the stop codon was part of a second short intron as suggested in PlasmoDB, we generated RT-PCR products from FCR-3 using specific

primers flanking the stop codon. A sequence analysis of RT-PCR products revealed that the internal stop codon was not spliced out. Therefore we concluded that both genes have a two exon structure. The first exons are very short, consisting of 189 and 180 nucleotides for the TrpA-3 and LysTrpA respectively, while the second exons account for the majority of the total gene sequences with 2,730 and 2,220 nucleotides for TrpA-3 and LysTrpA respectively. The intron regions, like for most *P. falciparum* intron sequences, are predominantly AT-rich and show length variation among the studied isolates. A closer observation of the cDNA sequences in TrpA-3 reveals an N-terminal unique region which is preceded by a sequence which codes for a hydrophobic signal sequence and a C-terminal end characterised by a cluster rich in Trp residues. In the LysTrpA gene, there are two prominent domains: the 5' highly charged domain which is rich in Lysine residues and a 3' end similar to TrpA-3 characterised by a cluster of Trp residues. In the N-terminus there is also a series of residues which code for a putative hydrophobic signal sequence. The Trp-rich domains are highly conserved both in number and position among the various isolates (not shown). When Trp is replaced by other amino acids, this is mainly by tyrosine or phenylalanine, both of which also belong to the bulky aromatic amino acids. Just upstream of the Trp region in TrpA-3 (between nucleotides 1,863 and 2,079) is a region containing four tandem repeat sequences of 18 amino acid residues. We also detected repetitive elements consisting of a number of pentapeptide repeats (KSDSK) in the 3' end lysine-rich domain of the LysTrpA gene in all the studied isolates. These differ in number between the different isolates resulting in a length polymorphism within this region.

To study the diversity of the two genes, we sequenced genomic DNA from a series of laboratory and field isolates. Using the sequence from the Binh1 isolate as

**Fig. 1** Sequence analysis of LysTrpA and TrpA-3. **a** The deduced amino acid sequence of LysTrpA and **b** TrpA-3 genes are shown. The tryptophan residues are indicated in *bold*. The repeat sequences are *underlined*, signal peptides are indicated in *lower case in gray*, while the position of introns are indicated by *arrow heads*. The stop codons in the LysTrpA sequence are highlighted with *asterisks (\*)* while the positions of mutations are highlighted in *red*



our reference, we determined the variability of these genes among the study isolates. A representative alignment demonstrated a high conservation of the genes between the laboratory and field isolates (not shown). However, we detected 14 SNPs in the coding region of TrpA-3, four of which were synonymous and the rest were biased towards non-synonymous mutations. Five of the mutations were located within the repeat region. One of the laboratory clones (Cys007) contained all the observed SNPs except for that at position 2,526 (Table 1). In the LysTrpA gene, only four SNPs were detected, all resulting in non-synonymous mutations (Table 1). The most remarkable mutation was observed at position 2,202 of the nucleotide sequence. At this position, a substitution reaction converts glutamic acid to a stop codon in all three laboratory strains and in four of the seven field isolates sequenced. Comparing TrpA-3 and LysTrpA to the other members of the gene family in *P. falciparum*, the results indicate that the different members of the family have about a 31% similarity to each other with homology restricted to the Trp-rich region as the position of the Trp residues are conserved in all the isolates (not shown). The aromatic amino acids seem to build a scaffold, which can also be found in other plasmodial species (Fig. 2). The N-terminal portion follows a regularity of W  $x_2$ W  $x_7$ W  $x_2$  F  $x_{10}$  W  $x_7$  W  $x_2$  W  $x_7$  W  $x_2$  (Y/F)  $x_7$ (Y/F) whereby  $x$  can be any residues. In the middle of the molecule this pattern is largely lost but towards the C-terminus the W  $x_2$  W  $x_7$  repeats appear again. Not only Trp residues but also some positions of charged residues like Glu, Asp and Lys are also highly conserved between species.

Between species one can find up to 85, 71 and 38% homology (chrPy1\_01062 (*P. yoelii*) versus PB\_116c08q1c (*P. berghei*); Pc\_5923 (*P. chabaudi*); Pk\_1813d08q1c (*P. knowlesi*) respectively, but among paralogues the homology is between 25 and 30% in *P. yoelii* proteins and around 40% in *P. knowlesi* (not shown) and largely restricted to the Trp scaffold. When

performing a TBLAST search on PlasmoDB on all available plasmodial sequences, all species contain proteins with a Trp-rich scaffold in various numbers: in *P. falciparum* five are found. In the murine parasites the numbers are in the same range. In the preliminary *P. vivax* and *P. knowlesi* sequences more than 15 can be found. These Trp-scaffold proteins are not related to any of the multigene families described so far.

#### Expression profiles of the family members

Gene expression analysis of the five members of the gene family was carried out using the Genesis 1.2.2 software (<http://genome.tugraz.at>). Raw data for this analysis was obtained from microarray analysis of tightly synchronised parasite cultures of the reference genome strain (3D7) of *P. falciparum*. A general picture of the gene expression profile within 48 h corresponding to the asexual intra-erythrocytic developmental cycle shows that both TrpA-3 and LysTrpA are expressed mainly in the mid-stages of the asexual cycle whereas TryThrA expression is more pronounced in earlier stages (Fig. 3). This corresponds well with the observation that TryThrA expression as shown by immunoprecipitation starts very early in the parasite development (Uhlemann et al. 2001). The expression of two other genes (PFA0135w and PFD0100c) is highest in later stages. The gene product of PFA0135w, the merozoite-associated tryptophan-rich Antigen (MaTrA), was shown by Western blot analysis to be expressed in later stages and merozoites (Ntumngia et al. 2004).

#### Discussion

The sequences of two Trp-rich proteins of a novel *Plasmodium falciparum* multi-gene family have been reported (Uhlemann et al. 2001, Ntumngia et al. 2004).

**Table 1** Polymorphisms and point mutations in the TrpA-3 and LysTrpA genes. Observed point mutations in the TrpA-3 using the Binh1 isolate as reference sequence and LysTrpA with the genome sequence of 3D7 as reference sequence. F = FCR-3, B = Binh1, S = Cys007, while D<sub>1</sub>–D<sub>7</sub> and \* represent field isolates

Position	Isolate	Affected codon	Affected amino acid
<b>TrpA-3</b>			
350	S, D <sub>6</sub>	CAC → TAC	His → Thr
589	S, D <sub>2</sub> , D <sub>3</sub> , D <sub>6</sub>	AAA → CAA	Lys → Gln
666	S, D <sub>2</sub> , D <sub>3</sub> , D <sub>6</sub>	GAT → GAA	Asp → Glu
702	S, D <sub>2</sub> , D <sub>3</sub>	TTG → TTA	Leu → Leu
1,875	S	GCG → GCT	Ala → Ala
1,876	S	TCC → GCA	Ser → Ala
1,896	S	TCC → TCA	Ser → Ser
1,902	S	GAA → GAT	Glu → Asp
1,980	S, F, D <sub>3</sub>	GAT → GCT	Asp → Ala
2,164	S, D <sub>1</sub> , D <sub>2</sub> , D <sub>6</sub> , D <sub>7</sub>	TCA → TCT	Ser → Ser
2,526	D <sub>1</sub> , D <sub>3</sub>	GAA → GAT	Glu → Asp
2,545	S, D <sub>1</sub> , D <sub>3</sub> , D <sub>4</sub>	GAT → TAT	Asp → Tyr
2,582	S, D <sub>1</sub>	AGG → ATG	Arg → Met
2,689	S, F, D <sub>1</sub> -D <sub>9</sub>	TAT → GAT	Tyr → Asp
<b>LysTrpA</b>			
43	243*	ATA → CTA	Ile → Leu
400	216*	ATA → AAA	Ile → Lys
2,034	236*	ATA → GTA	Ile → Val
2,202	F, B, S, 155*, 216*, 258*, 221*	GAA → TAA	Glu → Stop

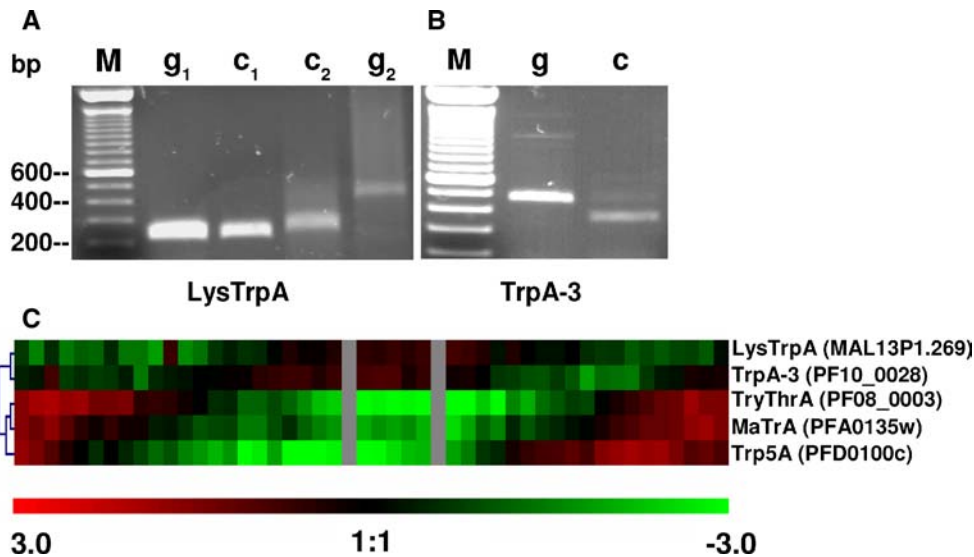


**Fig. 2** Multiple alignment of the tryptophan-rich domains of 12 genes belonging to different *Plasmodium* species. LysTrpA, TryThrA, MaTrA and Trp3 are members of the Trp-rich family in *P. falciparum*, pypAg1, pypAg3 are from *P. yoelii*, P.v396893, P.v 398040, P.v 398132, P.v 398701 are from *P. vivax*, while P. k1689e08q1c and P.b 116c08q1c correspond to *P. knowlesi* and *P. berghei* respectively. The shading indicates the consensual position of the Trp residues in the different *Plasmodium* genes. In black Trp (W), Phe (F) or Tyr (Y) residues are indicated at conserved positions. In the consensus sequence **h** represents conserved hydrophobic residues (A, C, F, I, L, M, V, W and Y), **C** indicates charged residues (D, E, K, R and H) and **+** or **-** indicate positively or negatively charged residues (K and R, E, D) in at least 70% of the residues

P. f LysTrpA	ERKIKAWN*W	KKNTDEWEQ	FMLLEKEK	KFIGKMKMD	EWFPTIYYN	WLAYVNNLEE
P. f TryThrA	DWKKNEWIKW	MDETEEWQL	LKLWLEGEKN	KWLEGKNKEY	DIWLNHMNSK	WTNYNKDIDE
P. f MaTrA	KWKASEWNMF	.NHIRNEWED	FYLVLQNDER	NILQEKHSNW	NEWIQHLENK	WGNINENINP
P. f Trp3	QWKNKQWEDW	KVYLEKQWSV	WLEISENDKI	EWLEDKERE	RSYIKYIYNI	WFKWLKRRKG
P. y pypAg1	~WKHNARRW	ILNLEDKWS	FNRNVDKNKS	EWMDQKDAQ	IEWVKELQTK	WMHFNPNIHQ
P. y pypAg3	~WKANEWETW	KEKIEAEWQT	FNDVIERERL	KWLTAKERDW	SEYTKVVEYK	WTHYSPDLEK
P. v 396893	~~~~~W	MKQLEEQWKS	FHDQMNERRK	QWLEKKDAEW	EKWKKYTEIK	YMSYCVDINN
P. v 398040	~WKRKEWSNW	LKKLEQDWKV	FNEKLQNEKK	TFLEEKEDDW	NTWIKSVEKK	WTHFPNPNMDK
P. v 398132	ELKKYAWN	FMKLQSDWKY	FNSALENEKQ	TFDEKEKEW	QEWLKS MENR	WTHYENMDA
P. v 398701	EWKAKEWKKW	INNTEKQWGN	FTQSFQHKQ	KWVNKKESEW	EEWLNTNIHYN	WIGFTNKLEF
P. k 1689e08q1c	~WKVNEWKKW	MNNTENDWEK	FMQSFERHKQ	MWIQEKENEW	EEWLNTNIHYN	WLGFTTKLEG
P. b 116c08q1c	~~~~~WEVW	KEKTETEWO	FNDIIEKERL	KWLAAKERDW	NEYMKYIEFK	WTHYSPDLEK
Consensus	W+ W W	h - -W	h - c+	Wh +--+W	-Wh hc +	W h h-
P. f LysTrpA	DFMHDKWEF	LNVKKEDEK	LIDSEWTKQM	EIKFKEFMRT	AENSLKQKQF	KRSIWSKSNK
P. f TryThrA	EYDSNVFKDS	YKWNKEQWEQ	WMKTEGKEFM	LQDFKRWLED	SEGYLKSALI	KQWIQWKNMK
P. f MaTrA	EYKTHLLHIS	LTSNEKQWEH	WAYNTLKYFL	ENDWNIFIQE	ITEEINTHID	QQWIKWLHEK
P. f Trp3	ELTGEQ.KPW	MTWNEQWET	FFETRYKKQF	LDEWYDMLKY	MDKILNIRKY	TLWNKWKGGK
P. y pypAg1	EYKIEFLSKS	ENWNDDKWKI	WMLTEAQQLL	ENDLKEWLSQ	AEADYCKWMT	DAWTQWKSQ
P. y pypAg3	LIKSDILKRN	QTLDEREWE	WMRSEGKELI	RKDVEEWIAY	NESYLNWWSI	KEWVSRNQI
P. v 396893	EFESENTLKDS	STWNRNQWED	WITTVGKEMI	KSEYEKWIYE	NENYLDEWKV	REWVEWKKQN
P. v 398040	EFEHTNMRRS	INWTESQWRE	WIQTEGRLLY	DIEWKKWFFE	NQSRLELIV	KKWIQWKKDK
P. v 398132	KFKSYILKNS	QGWDDNQWET	WIKNDGKKFM	ETNYHKWIDE	NYSNYNAWV	KRWEQWKNK
P. v 398701	DYINDKVN	TKWGEKEWKG	IIEMEWKRPM	KVKWTKLVEK	NEKIWDGKLF	NYWDNWKKDK
P. k 1689e08q1c	DYISDKTKY	KKWDENEWKG	IIEMEWTRPM	KVKWAKLVEK	NEKIWDGKFF	NYWDNWKKDK
P. b 116c08q1c	IIKSDILKRN	QTLDEREWE	WMRTEGKELI	RKDDEWASY	NESYLNWWSI	KEWLKWRNQI
Consensus	-hc c h+	W -c Wc	Wh - +chh	c c/ccWh	- h Wh h	c W W+ +
P. f LysTrpA	WNGLNWKKR	EWNFEGD*MW	INQYSAV...	.....QNEN	KLKWQELSIQ	EEIEWLDWVK
P. f TryThrA	...ILECLMN	EWRRREDEKW	SNLEDDQ.I	RVLNHNKDRKE	WLRWKERVTR	EKLEWKKHWE
P. f MaTrA	...NSMWLNS	DWIIIDENSW	EKMIKLDSSK	YAWTQDVVKY	WIKWKERINH	QNFNMNNWIK
P. f Trp3	...LTQWLMQ	DWKLIEDDEW	EKYEDKSW	KRFDMKKKNE	WKKWRNRIFK	EYIEWKEWYI
P. y pypAg1	...IKEWITL	DWKYEEDTFF	SKIDELT.V	EVLPAAEQRL	WYAWKERIYK	EGSDWKYWTS
P. y pypAg3	...ITKWISS	EWKREDEYW	ARWEE.NW.V	KHYNCEERNS	WIAWARINK	EMVEWNETK
P. v 396893	...ISSWLSS	EWKCKEDDSW	SKWEQTVW.G	KWLFTENRE	WTAWKERTLR	EAEWLKWWQ
P. v 398040	...IINWLMS	DWKRAEQEHW	EEFEKSWSS	KFFQIFEKRN	YEDFKDRVSD	EWEDWFEWVK
P. v 398132	...ILTWLLK	DWRNREFEYW	YKFKNMTLPE	PLYE.RAKNN	WNKWNKRLSK	EKEQWKKWVA
P. v 398701	W...NEWKNM	NWKKKEEEKW	NNSEKRR...	...DLNKDEN	WKEWKKRLLR	EKKEWENWVN
P. k 1689e08q1c	W...NEWKNK	NWKKKEEQW	NNFEEGK...	...DLNNDEN	WKEWNERLLR	EKQEWKNWVN
P. b 116c08q1c	...ITKWISS	EWKREEDDQW	AKWEE.NW.V	KHYNCEERND	WIAWARINK	EMVEWNDWSM
Consensus	W h Wh	-W+ c - -W	+h-c W	+hh c +	W cw+c+h +	- -W Wh
P. f LysTrpA	EKESYMTNKQ	LIEWDTWKEE	KRNWFKEYIN	LIKKDWLDHN	PWKNWVKV	
P. f TryThrA	MKNMNMIYNK	WKKWIKWKKN	KLANFNWWSK	NFIEKWIREK	QWNNWIN	
P. f MaTrA	NKNQVINHMK	KDKLEDWAKD	KYDSFNKWR	IFLQDWTANQ	KWKQLAK	
P. f Trp3	YKQSANVSD	VSYWEWKEK	KTNELKEWLD	ALTKQWMLQG	KW~~~~	
P. y pypAg1	LKESKYVNE	WELWEWKN	KQEMFSEWLQ	LYTNKWIQQK	QWRLLWIG	
P. y pypAg3	KKEEQVIDHR	SSSWGWRWTE	KQALFDMWLD	SFVDEWIKKE	HWFIWTS	
P. v 396893	LKEYVYINDE	WIPWTKWKN	NRLAFIDWIG	SFMNEWINNK	QWNEW~~	
P. v 398040	RKDNIFITNV	LDQWIKWKEE	KNLLYNNWAD	AFVTNWINKK	QWVW~~	
P. v 398132	AKYELYENSE	CKQWKKWDD	REVLFNWIE	SFINTWVAEK	KWN~~~~	
P. v 398701	EKEDFLMDSE	ETHWEKWKY	KWNYLNEWMK	QMKVDWLKSK	PWEVW~~	
P. k 1689e08q1c	EKEDSLMDAE	ETHWKKWKEY	KWNYLNEWMK	QVKVWKLKAK	PWEVW~~	
P. b 116c08q1c	KKEQQIIDHK	SSSWGWRWAE	KQTLFDMWLD	SFIDVW~~~~	~~~~~	
Consensus	++ hh c	W W W+	- +W Wh	cWh h cWh +	W W	

These are thought to be orthologues of two *Plasmodium yoelii* antigens. In this study, we have identified and characterised two further paralogues of these genes. They contain two exons each and typically for non-coding regions in *P. falciparum*, the 3' untranslated regions have a high A+T content. Both have a Trp-rich domain and a repetitive region. Despite the similarity in structure, the two antigens exert different ways of antigenic variation: LysTrpA contains a highly polymorphic repetitive region and very few single nucleotide polymorphisms (SNP), whereas TrpA-3 seems to be invariant in length but has accumulated more SNPs. Based on this, we suggest that TrpA-3 could be under greater selection pressure than LysTrpA.

TrpA-3 and LysTrpA, like their paralogues, TryThrA and MaTrA, contain highly conserved Trp-rich domains in all field and laboratory isolates studied so far. The main function of this domain in the parasite is not yet known. These sequences are very distinct from the Trp-containing domains of PfEMP1 and EBA-175. However, multiple Trp residues appear in functional domains of at least three malarial proteins that are involved in key processes including the invasion of hepatocytes (CSP and TRAP) (Goundis et al. 1988), cytoadherence (PfEMP1) (Baruch et al. 1997), and merozoite invasion of iRBCs (EBA) (Sim et al. 1994; Adams et al. 1992). Trp-rich proteins are found to be expressed throughout the asexual phase of the parasite life cycle as a clear



**Fig. 3 a** Photography of a 1.5% agarose gel showing PCR products from genomic DNA (g) and cDNA (c). g<sub>1</sub> and c<sub>1</sub> represent products from PCR and RT-PCR with primers flanking the stop codon at position 1,812 respectively while g<sub>2</sub> and c<sub>2</sub> represent products with primers flanking the predicted intron at the 5' end in the LysTrpA gene. Molecular weight markers are indicated in base pairs (bp). **b** g and c represent products from PCR and RT-PCR with primers flanking the predicted intron at the 5' end in the TrpA-3 gene. Differences in band sizes between g and c indicate the sizes of the introns. **c** Expression profiles of the various genes which constitute the Trp-rich protein gene family of *P. falciparum*. Each column indicates a different time point after merozoite invasion, covering the entire asexual 48 h cycle. *Red* and *green* represent positive and negative values of median respectively. The intensity of the red or green colour is a measure of the level of expression of the genes at the different time points (3 is a threefold induction; -3 is a threefold repression compared to a standard), as given below the time course. At the time point given in grey no samples were taken. Raw data were taken from the publication by Bozdech et al. 2003

indication that these proteins may play an essential role in the development of the parasite within its host.

Multigene families have been shown to be a common phenomenon in plasmodia and many other organisms and it is well established that the multiple paralogues generated through the process of amplification and divergence provide an organism with a set of related genes allowing the fine tuning of the biological activity based on the existing cellular developmental or environmental conditions. Through this process, some of the genes develop disabling elements resulting in loss of function and become pseudogenes. Pseudogenes in *P. falciparum* have been described in the large PfEMP1 family, the family of erythrocyte binding proteins (EBA) and for the ring-infected erythrocyte membrane antigen RESA. RESA-2 is a homologue of the ring-infected erythrocyte surface antigen RESA from *P. falciparum*, which contains a stop codon in the reading frame (Cappai et al. 1992). It was shown that RESA-2 was transcribed and that the mRNA was correctly spliced (Vazeux et al. 1993). Translation however was not shown in either report. Two further reports describe two

pseudogenes, which generate stop codons by frameshift mutations (Triglia et al. 2001; Taylor et al. 2001). Both genes belonging to different gene families were transcribed but no protein was produced. Knock-out parasites lacking PfRH3 or PsiEBA165 were viable and phenotypically not distinguishable from wild-type parasites.

LysTrpA contains two stop codons, one of which is polymorphic. The existence of these stop codons in parasites never been in culture rules out that these mutations have been introduced by artificial conditions. Translation through internal stop codons has long been known to occur in bacteria and viruses (Gesteland and Atkins 1996) and has also been shown in *P. falciparum* (Bischoff et al. 2000). The molecular mechanism underlying translation readthrough is still unknown.

Recently comparative genomic data were published from malarial parasites (Hall et al. 2005). The genes coding for Trp-rich proteins were either located in subtelomeric regions (TryThrA, MaTrA, TrpA-3) or they are classified as "inserts" (LysTrpA) within chromosome 13 of *P. falciparum*. The subtelomeric regions are pivotal locations for genes involved in antigenic variation in *P. falciparum*, *P. chabaudi* and *P. vivax* (Rubio et al. 1996; Fischer et al. 2003, Portillo et al. 2001). Given the limited number of Trp-rich proteins in *P. falciparum*, it is unlikely that they play a major role in antigenic variation in this species. But given the expression in later stages and the localisation of the members of the Trp-rich family on merozoites it is attractive to speculate about a possible role of Trp-rich proteins in antigenic variation in *P. vivax* or *P. knowlesi* merozoites.

**Acknowledgements** We are grateful to Andrea Weierich and Silvelia Grummes for running the ABI 3100 Genetic analyser. We also thank Angelica B. W. Boldt for help with the expression data. This work received financial support from the *fortune-programme* of the Medical Faculty of the University of Tübingen and a fellowship to F.B.N. from the Deutsche Akademische Austauschdienst.

## References

- Adams JH, Sim BK, Dolan SA, Fang X, Kaslow DC, Miller LH (1992) A family of erythrocyte binding proteins of malaria parasites. *Proc Natl Acad Sci* 89:7,085–7,089
- Balakiyev ES, Ayala FJ (2003) Pseudogenes: are they “junk” or functional DNA? *Annu Rev Genet* 37:123–151
- Baruch DI, Ma XC, Singh HB, Bi X, Pasloske BL, Howard RJ (1997) Identification of a region of PfEMP1 that mediates adherence of *Plasmodium falciparum* infected erythrocytes to CD36: conserved function with variant sequence. *Blood* 90:3,766–3,775
- Bischoff E, Guillolette M, Mercereau-Puijalon O, Bonnefoy S (2000) A member of the *Plasmodium falciparum* Pf60 multigene family codes for a nuclear protein expressed by readthrough of an internal stop codon. *Mol Microbiol* 35:1,005–1,016
- Blattner FR, Plunkett G3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277:1,453–1,474
- Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu J, DeRisi JL (2003) The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol* 1:E5
- Brenner B, Riley M (1995) Gene products of *Escherichia coli*: sequence comparisons and common ancestries. *Mol Biol Evol* 12:980–987
- Burns JM Jr, Adeeku EK, Dunn PD (1999) Protective immunization with a novel membrane protein of *Plasmodium yoelii*-infected erythrocytes. *Infect Immun* 67:675–680
- Burns JM, Adeeku EK, Belk CC, Dunn PD (2000) An unusual tryptophan-rich domain characterizes two secreted antigens of *Plasmodium yoelii*-infected erythrocytes. *Mol Biochem Parasitol* 110:11–21
- Cappai R, Kaslow DC, Peterson MG, Cowman AF, Anders RF, Kemp DJ (1992) Cloning and analysis of the RESA-2 gene: a DNA homologue of the ring-infected erythrocyte surface antigen gene of *Plasmodium falciparum*. *Mol Biochem Parasitol* 54:213–221
- Cranmer SL, Magowan C, Liang J, Coppel RL, Cooke BM (1997) An alternative to serum for cultivation of *Plasmodium falciparum* in vitro. *Trans R Soc Trop Med Hyg* 91:363–365
- Fischer K, Chavchich M, Huestis R, Wilson DW, Kemp DJ, Saul A (2003) Ten families of variant genes encoded in subtelomeric regions of multiple chromosomes of *Plasmodium chabaudi*, a malaria species that undergoes antigenic variation in the laboratory mouse. *Mol Microbiol* 48:1,209–1,223
- Gesteland RF, Atkins JF (1996) Recoding: dynamic reprogramming of translation. *Annu Rev Biochem* 65:741–768
- Goundis D, Reid KB (1988) Properdin, the terminal complement components, thrombospondin and the circumsporozoite protein of malaria parasites contain similar sequence motifs. *Nature* 335:82–85
- Hall N, Karras M, Raine JD, Carlton JM, Kooij TW, Berriman M, Florens L, Janssen CS, Pain A, Christophides GK, James K, Rutherford K, Harris B, Harris D, Churcher C, Quail MA, Ormond D et al (2005) A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses. *Science* 7:82–86
- Henikoff S, Greene EA, Pietrokovski S, Bork P, Attwood TK, and Hood, L (1997) Gene families: the taxonomy of protein paralogs and chimeras. *Science* 278:609–614
- Kun J, Hesselbach J, Schreiber M, Scherf A, Gysin J, Mattei D, Pereira da Silva L, Muller-Hill B (1991) Cloning and expression of genomic DNA sequences coding for putative erythrocyte membrane-associated antigens of *Plasmodium falciparum*. *Res Immunol* 142:199–210
- Kyes SA, Rowe JA, Kriek N, Newbold CI (1999) Rifins: a second family of clonally variant proteins expressed on the surface of red cells infected with *Plasmodium falciparum*. *Proc Natl Acad Sci* 96:9,333–9,338
- Labedan B, Riley M (1995) Gene products of *Escherichia coli*: sequence comparisons and common ancestries. *Mol Biol Evol* 12:980–987
- LeRoch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD, De La Vega P, Holder AA, Batalov S, Carucci DJ, Winzler EA (2003) Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* 12:1,503–1,508
- Little PF (1982) Globin pseudogenes. *Cell* 28:83–84
- Ntumngia FB, Bouyou-Akotet MK, Uhlemann AC, Mordmuller B, Kremsner PG, Kun JF (2004) Characterisation of a tryptophan-rich *Plasmodium falciparum* antigen associated with merozoites. *Mol Biochem Parasitol* 137:349–353
- Owen CA, Sinha KA, Keen JK, Ogun SA, Holder AA (1999) Chromosomal organisation of a gene family encoding rho-tryptophan proteins in *Plasmodium yoelii*. *Mol Biochem Parasitol* 99:183–192
- Portillo HA, Fernandez-Becerra C, Bowman S, Oliver K, Preuss M, Sanchez CP, Schneider NK, Villalobos JM, Rajandream MA, Harris D, Pereira da Silva, LH, Barrell B, Lanzer M (2001) A superfamily of variant genes encoded in the subtelomeric region of *Plasmodium vivax*. *Nature* 410:839–842
- Preiser PR, Jarra W, Capiod T, Snounou GA (1999) Rho-tryptophan-associated mechanism of clonal phenotypic variation in rodent malaria. *Nature* 398:618–622
- Preiser PR, Khan S, Costa FT, Jarra W, Belnoue E, Ogun S, Holder AA, Voza T, Landau, I, Snounou G, Renia L (2002) Stage-specific transcription of distinct repertoires of a multigene family during *Plasmodium* life cycle. *Science* 295:342–345
- Restrepo BI, Carter CJ, Barbour AG (1994) Activation of a *vmp* pseudogene in *Borrelia hermsii*: an alternate mechanism of antigenic variation during relapsing fever. *Mol Microbiol* 13:287–299
- Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W, Cherry JM, Henikoff, S, Skupski MP et al (2000) Comparative genomics of the eukaryotes. *Science* 287:2,204–2,215
- Rubio JP, Thompson JK, Cowman AF (1996) The var genes of *Plasmodium falciparum* are located in the subtelomeric region of most chromosomes. *EMBO J* 15:4,069–4,077
- Salanoubat M, Lemcke K, Rieger M, Ansorge W, Unseld M, Fartmann B, Valle G, Blocker H, Perez-Alonso M, Obermaier B, Delseny M, Boutry M et al (2000) Sequence and analysis of chromosome 3 of the plant *Arabidopsis thaliana*. *Nature* 408:820–822
- Sim BK, Chitnis CE, Wasniowska K, Hadley TJ, Miller LH (1994) Receptor and ligand domains for invasion of erythrocytes by *Plasmodium falciparum*. *Science* 264:1,941–1,944
- Su XZ, Heatwole VM, Wertheimer SP, Guinet F, Herrfeldt JA, Peterson DS, Ravetch JA, Wellems TE (1995) The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. *Cell* 82:89–100
- Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. *Science* 278:631–637
- Taylor HM, Triglia T, Thompson J, Sajid M, Fowler R, Wickham ME, Cowman AF, Holder AA (2001) *Plasmodium falciparum* homologue of the genes for *Plasmodium vivax* and *Plasmodium yoelii* adhesive proteins, which is transcribed but not translated. *Infect Immun* 69:3,635–3,645
- Triglia T, Thompson JK, Cowman AF (2001) An EBA175 homologue which is transcribed but not translated in erythrocytic stages of *Plasmodium falciparum*. *Mol Biochem Parasitol* 116:55–63
- Uhlemann AC, Oguariri, RM, McColl DJ, Coppel RL, Kremsner PG, Anders RF, Kun JF (2001) Properties of the *Plasmodium falciparum* homologue of a protective vaccine candidate of *Plasmodium yoelii*. *Mol Biochem Parasitol* 118:41–48
- Vazeux G, Le Scanf C, Fandeur T (1993) The RESA-2 gene of *Plasmodium falciparum* is transcribed in several independent isolates. *Infect Immun* 61:4,469–4,472
- Wilde CD (1986) Pseudogenes. *CRC Crit Rev Biochem* 19:323–352