

A global view of the nonprotein-coding transcriptome in *Plasmodium falciparum*

Carsten A. Raabe¹, Cecilia P. Sanchez², Gerrit Randau¹, Thomas Robeck¹, Boris V. Skryabin¹, Suresh V. Chinni¹, Michael Kube³, Richard Reinhardt³, Guey Hooi Ng⁴, Ravichandran Manickam⁴, Vladimir Y. Kuryshev¹, Michael Lanzer², Juergen Brosius^{1,*}, Thean Hock Tang^{4,*} and Timofey S. Rozhdestvensky^{1,*}

¹Institute of Experimental Pathology, ZMBE, University of Muenster, Von-Esmarch-Str. 56, 48149 Muenster, ²Abteilung Parasitologie Hygiene Institut, Im Neuenheimer Feld 324, 69120 Heidelberg, ³MPI Molecular Genetics, Ihnestrasse 63-73, 14195 Berlin-Dahlem, Germany and ⁴Advanced Medical and Dental Institute (AMDI), Institute for Molecular Research (INFORMM), USM 11800 Malaysia

Received August 25, 2009; Revised October 2, 2009; Accepted October 5, 2009

ABSTRACT

Nonprotein-coding RNAs (npcRNAs) represent an important class of regulatory molecules that act in many cellular pathways. Here, we describe the experimental identification and validation of the small npcRNA transcriptome of the human malaria parasite *Plasmodium falciparum*. We identified 630 novel npcRNA candidates. Based on sequence and structural motifs, 43 of them belong to the C/D and H/ACA-box subclasses of small nucleolar RNAs (snoRNAs) and small Cajal body-specific RNAs (scaRNAs). We further observed the exonization of a functional H/ACA snoRNA gene, which might contribute to the regulation of ribosomal protein L7a gene expression. Some of the small npcRNA candidates are from telomeric and subtelomeric repetitive regions, suggesting their potential involvement in maintaining telomeric integrity and subtelomeric gene silencing. We also detected 328 *cis*-encoded antisense npcRNAs (asRNAs) complementary to *P. falciparum* protein-coding genes of a wide range of biochemical pathways, including determinants of virulence and pathology. All *cis*-encoded asRNA genes tested exhibit lifecycle-specific expression profiles. For all but one of the respective sense–antisense pairs, we deduced concordant patterns of expression. Our findings have important implications for a better understanding of gene regulatory mechanisms in *P. falciparum*, revealing an extended and sophisticated npcRNA

network that may control the expression of housekeeping genes and virulence factors.

INTRODUCTION

Malaria is a devastating, life-threatening parasitic disease causing an estimated 515 million clinical cases and 1.0 million deaths annually (1). More than half of the world's human population lives in areas where malaria is endemic. The protozoan parasite *Plasmodium falciparum* is the etiological agent of the most virulent form of malaria in humans. *Plasmodium falciparum* has acquired resistance to many antimalarial drugs and the mosquito vectors have developed resistance to available insecticides.

The *P. falciparum* genome was recently sequenced (2), and consists of 23 Mb distributed over 14 chromosomes with sizes ranging from 0.643 to 3.29 Mb. In addition to the nuclear genome, *P. falciparum* contains a 5.9 kb mitochondrial and a 35 kb apicoplast genome. Approximately 5300 protein-coding genes have been identified. The (A + T) content ranges from 80.6% to ~90% in introns and intergenic regions.

Genomes typically encode two different types of RNA molecules: messenger RNAs (mRNAs) and various classes of nonprotein-coding RNAs (npcRNAs). The mRNAs provide templates for protein synthesis, whereas the npcRNAs do not code for proteins but rather perform various regulatory functions exerted by the RNA itself or in complexes with proteins (RNPs). npcRNAs participate in the regulation of diverse biochemical pathways, including chromosome modification, transcription and translation, splicing, developmental timing, cell

*To whom correspondence should be addressed. Tel: +49 251 8358607; Fax: +49 251 8352134; Email: rozhdest@uni-muenster.de
Correspondence may also be addressed to Juergen Brosius. Tel: +49 251 8358511; Fax: +49 251 8358512; Email: RNA.world@uni-muenster.de
Correspondence may also be addressed to Thean-Hock Tang. Tel: +604 5622302; Fax: +6045792960; Email: tangth@amdi.usm.edu.my

differentiation, proliferation, apoptosis and organ development [for reviews see (3–7)].

Although the small npcRNA transcriptome has been experimentally identified and successfully verified in several model organisms, very few systematic attempts to uncover the diversity and functions of small npcRNAs in pathogenic organisms have been undertaken to date. Recently, and based mainly on comparative genomics, Chakrabarti *et al.* (8) detected a number of structural RNAs, e.g. telomerase RNA, 35 small nucleolar RNAs (snoRNAs), spliceosomal small nuclear RNAs (snRNAs), MRP RNA and RNase P RNA, in *Plasmodium*. In addition, the authors described six new npcRNAs, as of yet, unknown functions. Similarly, Mourier *et al.* (9) performed computational screens in intronic and intergenic regions of *P. falciparum* to identify conserved RNA secondary structures between distantly related *Plasmodium* species, yielding a set of 604 putative npcRNAs; 33 of which they verified experimentally. Notably, 29 RNAs from this dataset overlapped with those of the aforementioned set of Chakrabarti *et al.* (8). Li *et al.* (10) reported centromeric expression of small npcRNA candidates (75 and 175 nt) in *P. falciparum*. Their data suggested bidirectional promoter activities within the centromeres of the parasite. This subclass of npcRNA candidates is localized within the nucleus and appears to associate with the centromeric chromatin (10). Taken together, these observations support the hypothesis that these npcRNAs function in the coordinated organizational assembly of chromatin. Interestingly and consistent with the absence of genes encoding argonaute and dicer proteins (11), so far there were no *bona fide* miRNAs experimentally verified in *Plasmodium*. However, there might be different and evolutionarily distinct classes of small npcRNAs that compensate for this apparent lack of RNA interference machinery.

Despite the aforementioned studies, the total number of small npcRNAs in the *P. falciparum* genome, their importance and the variety of functions they serve are still largely unknown. This is partly due to the inherent computational limitations of the algorithms used for RNA predictions. Here we report the experimental identification and analysis of the global small npcRNA transcriptome in *P. falciparum* and provide its functional annotation. Our data point to the potential regulation of *P. falciparum* gene expression and virulence by small stable npcRNAs. Conceptual similarities to the RNome of *Saccharomyces cerevisiae* are discussed. Moreover, given the broad mechanistic spectrum within which npcRNAs act and the low degree of host parasite conservation, small npcRNAs might be considered as potential drug targets.

MATERIALS AND METHODS

cDNA library construction

Total RNA extracted from all human erythrocyte stages of *P. falciparum* 3D7 strain was size fractionated (10–60 nt and 60–500 nt) on an 8% (w/v) denaturing polyacrylamide gel (7M urea, 1× TBE buffer). Passive elution was

performed in 0.3M NaOAc (pH 5.3) overnight at 4°C. Subsequently, 5 µg of size-fractionated RNA was treated with tobacco acid pyrophosphatase (Epicenter) for 1 h at 37°C and C-tailed with poly-A polymerase (Invitrogen) for 2 h at 37°C (12). A 5'-DNA *SalI* adapter (5'-CAAC GCGTCTGACTACGTGAGATTTGAGGTTTC-3') was then ligated to the 5'-end of the RNA using T4 RNA ligase (Fermentas) at 4°C overnight. First-strand cDNA synthesis was performed using the Transcriptor Reverse Transcriptase (Roche) and a *NotI* DNA oligonucleotide (5'-GACTAGTTCTAGATCGCGAGCGGCCGCCCG GGGGGGGGGGGGGG-3'), followed by single-stranded PCR reactions using the *SalI* oligonucleotide. The PCR product was then double digested with *SalI* (Roche) and *NotI* (Roche) restriction endonucleases, ligated (T4 DNA ligase, Roche) into the pSPORT1 vector (Invitrogen) and then transformed into *Escherichia coli* Top10 (Invitrogen) electrocompetent cells.

Northern blot analysis

Plasmodium falciparum total RNA was separated on 8% polyacrylamide/7M urea gels and electrotransferred to positively charged nylon membranes (Ambion BrightStar-Plus membranes). After UV cross-linking, the membranes were hybridized with npcRNA-specific ³²P-end-labeled oligonucleotide probes (Supplementary Table S12) overnight at temperatures ranging from 46°C to 58°C in 1 M sodium phosphate pH 6.2, 7% sodium dodecyl sulfate (SDS). After hybridization, the membranes were washed twice with 0.2 M sodium phosphate, 1% (w/v) SDS buffer for 10 min at 58°C and exposed to MS film (Kodak) at –80°C overnight.

Real-time PCR analyses

Plasmodium falciparum total RNA (500 ng input), isolated from different intra-erythrocyte stages of the parasite, was DNase treated (Roche) and reverse transcribed using Transcriptor Reverse Transcriptase (Roche) and a gene-specific primer (MWG) at 55°C for 1 h (Supplementary Table S12). Real-time PCR with 1:100 diluted cDNAs was performed using the LightCycler 480 (Roche) and SYBR Green I Master (Roche) in 12.5 µl reaction volume (at least in triplicates) under the following cycling conditions: 95°C, 5 min, followed by 45 cycles of 95°C, 20 s; 53°C, 40 s; and 68°C, 50 s. For relative quantification, the seryl-tRNA synthetase (*s-tRNA syn*) housekeeping gene was used as reference control. The integrities of the obtained products were verified by gel electrophoresis on either 8% polyacrylamide (in 1× TBE buffer) or 2% agarose (in 1× TAE buffer) gels. In addition, a melting curve analysis was carried out for each reaction. The calculations of average C_p values and SDs for each candidate were performed using the Roche LightCycler 480 software. Ratios were calculated as $C_p \text{ s-tRNA syn} / C_p \text{ candidate}$.

Computational analysis

Annotation of the npcRNA candidates. The 5'-adapter and C-tail sequences were removed from cDNAs using UNIX scripts. cDNAs longer than 17 bp were processed further.

The sequences were assembled into contigs using the Seqman program of the Lasergene software package. The assembly parameters were set to a 10 bp minimum overlap with sequence similarity of $\geq 90\%$. The resulting contigs were also manually inspected. The genomic coordinates of all npcRNA candidates were computed by WU-BLAST stand-alone searches (<http://blast.wustl.edu/>) against the latest *P. falciparum* genome project release (2007), accessed via the Malaria Genome Browser (http://areslab.ucsc.edu/data_downloads/). For npcRNA candidates that matched multiple locations, the longest matches with the highest identity percentages were selected (allowing for many but identical matches). Only matches with a minimum identity of 95% were analyzed further. Regions of overlapping transcription between an npcRNA candidate and any known or predicted gene were identified using AWK-based UNIX scripts. The genomic coordinates and respective orientations of all transcripts (or predicted genes) were downloaded via the Malaria Table Browser site (<http://areslab.ucsc.edu/cgi-bin/hgTables?command=start>). Each result was manually validated using web-based BLAST or BLAT genomic browsers (<http://blast.ncbi.nlm.nih.gov/Blast.cgi> and <http://areslab.ucsc.edu/cgi-bin/hgBlat?command=start>). A custom track for the Malaria genome browser (<http://areslab.ucsc.edu/cgi-bin/hgCustom?hgsid=13740>) containing all novel npcRNA candidates is provided as Supplementary file. For additional methods, see 'Materials and Methods' in Supplementary Data.

RESULTS AND DISCUSSION

Construction and analysis of *P. falciparum* cDNA libraries

Two specialized small size cDNA libraries were generated from size-fractionated total RNA from the *P. falciparum* clone 3D7 (see Results and Discussion in Supplementary Data). A total of 25 000 cDNA clones were selected at random, sequenced and subsequently analyzed, resulting in 3588 distinct contigs. A majority of the cDNA sequences analyzed could be assigned to known

npcRNAs and represented various structural RNA molecules of the *P. falciparum* transcriptome. Among them, we detected fragments of rRNAs (48%), tRNAs (5%) and the U spliceosomal snRNAs (4%). Apart from known nuclear npcRNAs, we also identified known rRNA fragments encoded by the *P. falciparum* mitochondrion (see below and also Results and Discussion in Supplementary Data) (13,14) and the plastid-like, circular 35 kb structure. On the other hand, 630 distinct cDNA contigs that matched either the nuclear or mitochondrial genomes of *P. falciparum* could not be assigned to any known npcRNA or known protein-coding gene, and are, therefore, likely to represent novel members of the small npcRNA transcriptome (Figure 1; Supplementary Tables S1–11; Results and Discussion in Supplementary Data).

Novel *P. falciparum* snoRNAs and small Cajal body-specific RNAs

In eukaryotes, snoRNAs and small Cajal body-specific RNAs (scaRNAs) are divided into two subclasses, the C/D-box sno/scaRNAs and the H/ACA-box sno/scaRNAs, with some of the scaRNAs harboring both boxes. C/D-box sno/scaRNAs target 2'-*O*-methylation of riboses, while H/ACA-box sno/scaRNAs guide the conversion of uridine into pseudouridine within ribosomal RNA (rRNA) and snRNA targets [for review see (15)]. Here we report 43 npcRNAs that belonged to the C/D and H/ACA-box subclasses of sno/scaRNAs. Among them, we identified 15 novel sno/scaRNAs that escaped previous computational screens (8) (Supplementary Tables S1 and S2; Supplementary Figures S1 and S2). For the majority of the computationally predicted sno/scaRNA candidates, we could determine precise 5'- and 3'-ends (Supplementary Table S1). Most of the novel *P. falciparum* snoRNAs have functional homologues within the human and yeast RNomes and may target corresponding regions in the 18S and 28S rRNAs (25S rRNA in yeast), respectively (16,17) (Supplementary Table S2; Supplementary Figures S1 and S2). The observed conservation of the novel *P. falciparum* snoRNAs may be an indication of the functional importance of the corresponding rRNA modifications. However, it is also

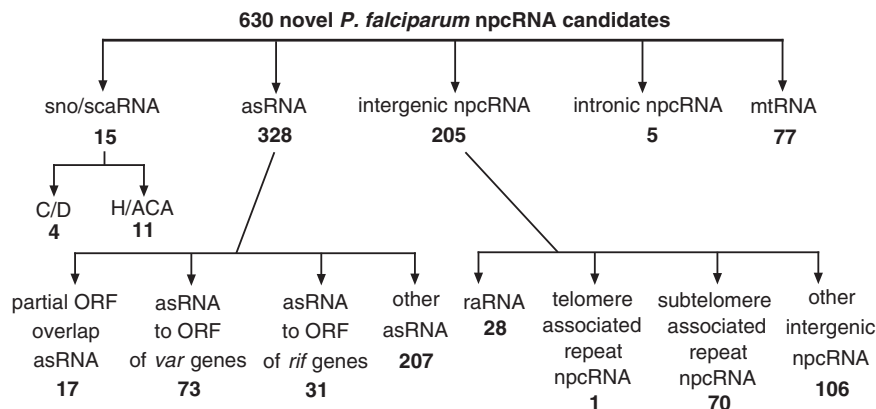


Figure 1. Categories and numbers of identified novel npcRNA candidates. Categories: snoRNA, small nucleolar RNA; scaRNA, small Cajal body-specific RNA; asRNA, *cis*-encoded antisense npcRNA to ORFs of protein coding genes; mtRNA, mitochondrial npcRNA; raRNA, repeat-associated npcRNA.

possible that some of the *P. falciparum* snoRNA candidates perform species-specific functions (see Results and Discussion in Supplementary Data).

Exonization of a *P. falciparum* snoRNA

Approximately one-third of the *P. falciparum* snoRNA genes are situated in introns, the remainders are intergenic (Supplementary Table S1; Supplementary Results and Discussion). Two of the snoRNAs, snoRNA-27 H/ACA-box and snoRNA-28 C/D-box npcRNAs (8) are located within the 3'-untranslated region (UTR) intron of the ribosomal protein gene L7a (Figure 2A; Supplementary Results and Discussion). Interestingly, following reverse transcription polymerase chain reaction (RT-PCR), cloning and sequencing, we observed alternative splicing events within the 3'-UTR of the L7a mRNA that resulted in the exonization of the 3'-portion of snoRNA-27. Structural analysis of the exonized snoRNA-27 domain revealed a K-turn motif-like structure (Figure 2B and E). Both the alternative splice site and the K-turn motif are evolutionarily conserved among all *Plasmodium* species examined, pointing toward the functional significance of the exonized snoRNA domain as an integral part of the 3'-UTR (Figure 2B and E). Our preliminary data indicated an interaction between the exonized portion of the snoRNA-27 npcRNA and the L7a protein (data not shown), suggesting a possible autogenous regulatory mechanism of gene expression. Alternatively, the identified RNA structural motif within the 3'-UTR of the L7a mRNA may be specifically recognized by K-turn-binding proteins, which, in turn, may influence mRNA transport and/or stability (18). To the best of our knowledge, this is the first report of the exonization of a snoRNA gene that contributes functional parts to a 3'-UTR. *In silico* screens of other eukaryotic genomes identified additional examples of potential exonizations of snoRNA genes into UTRs of various mRNAs (C.A.Raabe *et al.*, unpublished data); thus pointing toward a mechanism of general relevance.

Telomere- and subtelomere-associated small npcRNA candidates

Based on structural conservation, recent computational, genome-wide screens recovered 33 novel intergenic npcRNA candidates (9); sequence and structural analyses indicate that the vast majority of them belong to the class of the above discussed snoRNA candidates and many among them were also reported by Chakrabarti *et al.* (8). Our experimental approach uncovered 205 potentially novel npcRNA candidates that are supposedly transcribed within intergenic regions (excluding snoRNAs) (Supplementary Tables S3, S4a, S4b and S5). Half of them contain sequence repeats and are derived from telomeric, subtelomeric or internal satellite-like regions (Supplementary Tables S3, S4a and S4b). Thus, 71 of 205 intergenic npcRNA candidates identified originated from chromosome terminal regions.

The ends of all 14 *P. falciparum* chromosomes share a common and conserved organization; the telomere repeat units (GGGTTTA) and the nonprotein-coding

subtelomeric domain of 15–30 kb are followed by members of the *var*, *rif* and *stevor* multigene families, which encode determinants of virulence and pathology, such as the *var*-encoded immuno-variant adhesin PfEMP1 (see below) (2,19). The subtelomeric domains consist of six different telomere-associated repetitive elements (TAREs 1–6) (20). TAREs 1–5 are located toward the telomeres, whereas TARE-6, originally known as rep20, is situated upstream of the multigene family clusters. We identified one npcRNA candidate associated with the telomeres and 70 candidates associated with the six TAREs. The telomere-associated repeat npcRNA-6209 resembles the recently described eukaryal telomeric-repeat containing RNAs (Figure 3A and B; Supplementary Table S3) (21–23) and contains the characteristic UU/CAGGG *P. falciparum* telomere repeat unit (Figure 3A and B). Northern hybridizations suggested that npcRNA-6209 is processed from longer precursors that are transcribed toward the chromosomal ends (Figure 3A and B). Telomeric npcRNAs are thought to function in maintaining telomere integrity (22,23). The 70 subtelomere associated repeat npcRNA candidates identified are derived from the *P. falciparum* TARE 1–6 regions, and are transcribed from both strands (Figure 3A and C; Supplementary Table S3).

Although the functions of the identified subtelomere-associated repeat npcRNA remain elusive, it is tempting to propose, based on their origin and the striking similarities to the RNA-based machinery of *S. cerevisiae* [i.e. homologs of *S. cerevisiae* Sir2 or Rrp6 proteins, etc.; (24)], that they are involved in an RNA-mediated regulatory network associated with transcriptional repression and gene silencing through RNA-induced histone deacetylation. *Plasmodium falciparum*, like *S. cerevisiae*, lacks all key components of the siRNA machinery (11). However, it is plausible that alternative npcRNA-driven mechanisms have evolved to substitute for the missing siRNA regulatory functions. In *S. cerevisiae*, parts of the npcRNA transcriptome are associated with transcriptional repression and gene silencing through RNA-induced histone deacetylation (25). *Plasmodium falciparum* possesses histone deacetylases, and previous studies showed that a functional knock-out of the histone deacetylase PfSir2 disrupted monoallelic expression of the *var* gene family (26). PfSir2 is a telomere-binding protein that spreads from telomeres along the subtelomeric domain into the cluster of virulence factors, thereby forming a chromosomal gradient of heterochromatin structure and histone hypoacetylation, which contribute to gene silencing and allelic exclusion (27,28).

Alternatively, the subtelomere-associated repeat npcRNA may form DNA/RNA duplexes that, as shown in other systems, are highly recombinogenic (29) and which, in turn, may contribute to the high degree of ectopic recombination displayed by *P. falciparum* chromosome ends. Ectopic recombination involves the virulence gene clusters and is thought to generate genetic and, in turn, antigenic diversity among virulence factors (30,31). The 20 npcRNA candidates flanking the subtelomeric virulence gene clusters on the side facing the centromere

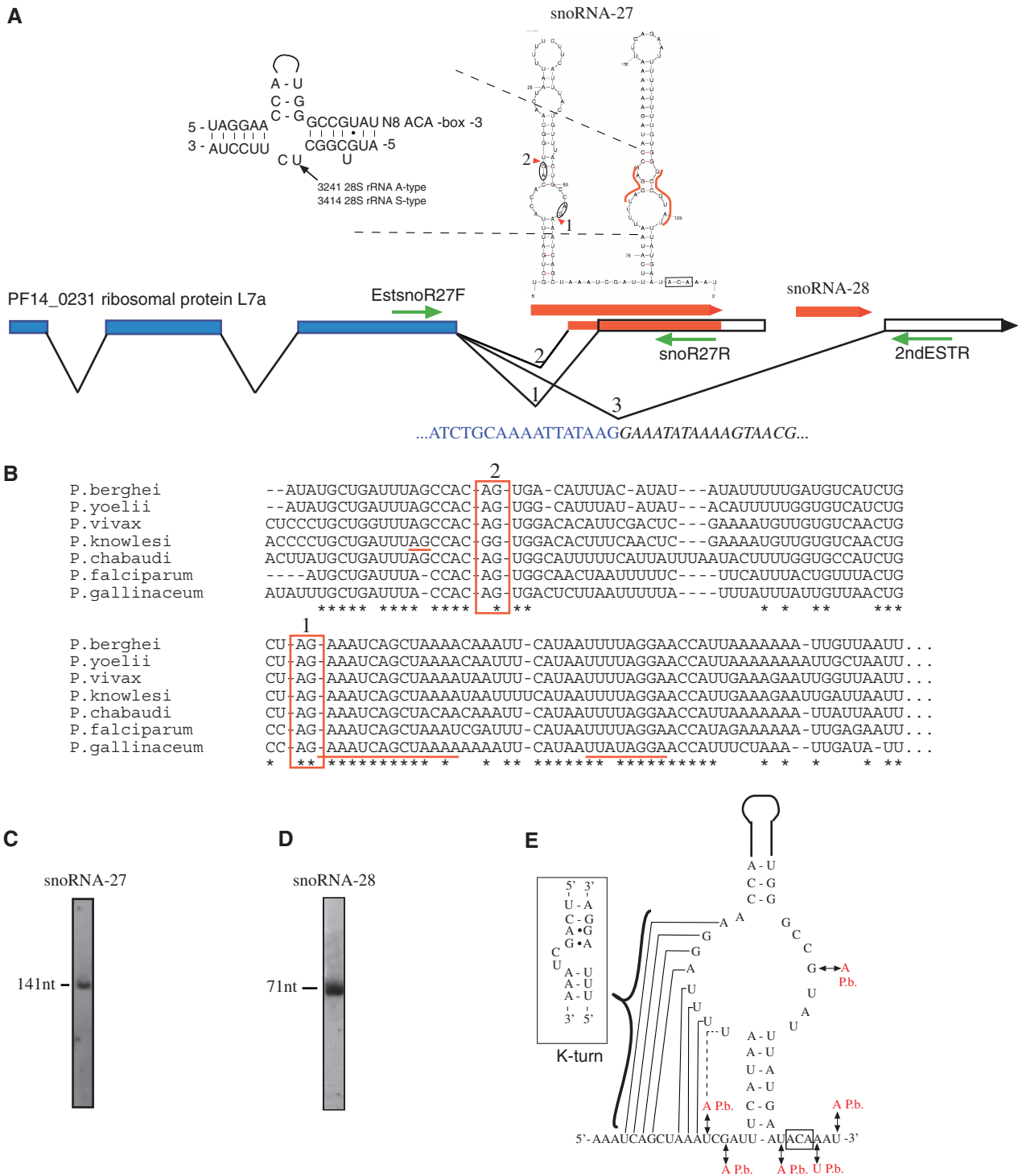


Figure 2. Exonization of the *P. falciparum* snoRNA-27H/ACA-box npcRNA. (A) Schematic representation of the ribosomal protein L7a gene. Guide element with predicted 2D structure of snoRNA-27 npcRNA is outlined in red; complementarities within 28S rRNAs target sites are shown. Red arrows indicate locations of snoRNA-27 and snoRNA-28 npcRNA genes. Blue bars represent the ORF of the L7a gene, unfilled bars represent the 3'-UTR. Alternative splicing sites for L7a 3'-UTR are indicated with 2D snoRNA-27 structure, splice variants 1 and 2 and on the scheme of the L7a gene, splice variants 1, 2 and 3. The sequence of the third splice variant is shown; nucleotides of the third exon are blue, nucleotides of the last exon are italicized. Green arrows indicate oligonucleotides used for RT-PCR analysis. Drawing is not to scale. (B) Sequence alignment of alternative 3'-splice sites of L7a mRNA located within the snoRNA-27 npcRNA gene of various *Plasmodium* species. Boxes 1 and 2 represent the alternative splice sites. (C and D) Northern blot analyses of mature snoRNA-27 (C) and snoRNA-28 (D) npcRNAs, sizes are indicated in nucleotides (nt). (E) Potential K-turn motif structure of the exonized 3'-portion of the *P. falciparum* snoRNA-27 is shown on the left panel. Nucleotide substitutions between *P. falciparum* and *P. berghei* (P.b.) are indicated by black double arrows.

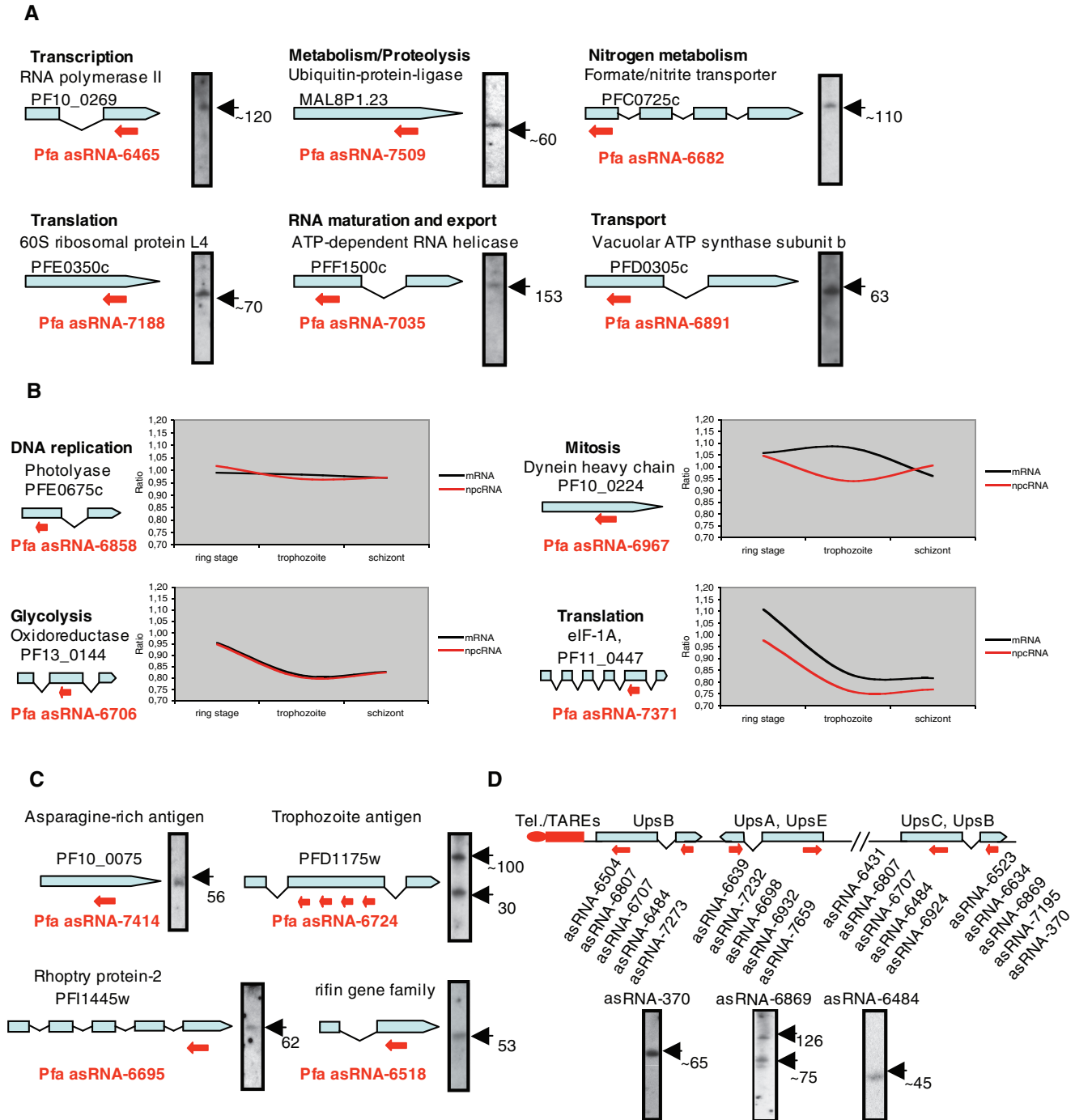


Figure 4. Expression analysis and schematic representation of examples from the *cis*-encoded asRNA transcriptome in *P. falciparum*. (A–D) asRNA candidates to protein-coding genes of different biochemical pathways. ORFs of protein-coding genes are drawn as thick light blue arrows, spliced ORFs are drawn as fused bars. Red arrows depict locations of npcRNA candidates. Drawing is not to scale. Gene ID, npcRNA and possible biochemical pathway are indicated. npcRNA sizes (nt) are indicated on each panel. (B) Differential expression profiles of *cis*-encoded asR candidates and sense *P. falciparum* protein-coding genes during the intra-erythrocyte stages of the parasite’s life cycle as determined by real-time PCR. For relative quantification, seryl-tRNA synthetase (PF07_0073) mRNA was used. Expression ratios and intra-erythrocyte stages are indicated on Y and X axes, respectively. Expression profiles of asR and sense mRNA genes are indicated by red and black lines, respectively. (C and D) Small asRNA candidates to *P. falciparum* protein-coding genes encoding known/potential virulence factors. (D) Candidate npcRNAs (asRNA) antisense to *var* genes subtypes. A number of asRNA candidates to *var* genes on chromosome 4 are indicated. Northern blot analysis verified expression and sizes of *var* asRNA candidates.

mechanism to control the expression of its small asRNA transcriptome.

In addition, we detected stable antisense transcripts to various virulence-associated genes (Figure 4C and D;

Supplementary Tables S7–9). Of particular interest are the small asRNA candidates complementary to open reading frames (ORFs) of the multigene families *var* and *rif*, the main determinants of the parasite’s virulence (Figure 4C

and D; Supplementary Tables S7 and S8). We detected 73 small *cis*-encoded asRNA candidates potentially transcribed from 137 loci complementary to the ORFs of *var* genes. We also detected 31 *rif* gene-associated asRNA candidates that might be generated from as many as 71 loci (Supplementary Results and Discussion).

The *var* gene superfamily encodes immunovariant adhesins, collectively termed *Plasmodium falciparum* erythrocyte membrane protein1 (*PfEMP1*), which mediate a broad range of cytoadhesive interactions that eventually lead to the sequestration of infected erythrocytes in the deep vasculature. This pathological property is responsible for much of the disease associated with falciparum malaria (41). Genomic analysis has so far identified a pool of about 60 different *var* genes per haploid genome (2). Irrespective of the specific upstream region (upsA-E) and the actual promoter sequence associated with a particular member, all *var* genes follow the rule of monoallelic expression, thereby providing a mechanism to escape the host's immune defense (42–44). Important questions regarding which genetic elements and epigenetic processes connect single *var* gene family members to control the transcription of all alleles collectively remain to be solved. We identified small stable asRNAs to all *var* subgroups (except for the upsE subtype). Moreover, we distinguished two main subtypes of the *var* asRNA candidates. The first comprises 56 asRNA candidates, each expressed in antisense orientation to the ORF of a single specific *var* gene. The second comprises 17 asRNA candidates that are each complementary to more than one *var* gene (i.e. probably a consequence of the aforementioned extensive degree of segmental duplication). An example is asRNA-6707, which may be derived from 20 different *var* gene loci (Supplementary Table S7). Interestingly, we also observed 41 *var* genes associated with more than one small asRNA candidate. In summary, we report asRNA candidates to almost all members of the *var* mRNAs.

Recent reports consider the *var* antisense RNAs as leaky by-products of *var* gene expression resulting from the open chromatin structure generated by active sense gene transcription (45). However, Northern blot hybridizations indicated that *var* antisense transcription results in distinct small stable asRNAs (Figure 4D and data not shown). Interestingly, for various *P. falciparum* protein-coding genes, there are many detectable ESTs, suggesting that, due to active transcription, open chromatin structures should be assumed. However, for many of those genes we failed to detect small asRNA candidates (data not shown). These observations suggest that active sense gene transcription is not a precondition or an imperative for small stable *var* asRNA expression. Interestingly, *Giardia lamblia*, the most common human intestinal parasite, expresses 190 variant-specific surface protein (VSP)-coding genes in response to the host's immune reaction. Similar to the *P. falciparum* *var* genes, expression of VSP genes is strictly regulated and only one VSP is present on the surface of each parasite at any particular time (46). Recently, small *cis*-encoded antisense RNAs corresponding to the silenced VSP genes were detected, suggesting the involvement of an RNA

interference pathway in antigenic variation (47). Although *P. falciparum* lacks components of the siRNA-based machinery, the similarity in the *cis*-encoded antisense transcriptome to genes responsible for the severe progression of infection leads one to propose that alternative npcRNA-driven mechanisms have evolved to substitute for the missing siRNA regulatory functions.

How these small asRNAs act is still largely unknown. Analysis of human antisense transcription revealed a number of evolutionarily conserved and functionally significant SAS gene pairs that tend to be co-regulated (48), and, as shown in other systems, antisense transcripts exert their regulatory functions through various means. Antisense transcripts can stall RNA polymerases (49). RNA–RNA duplex formation can further interfere with translation and can facilitate and regulate alternative splicing events (50). There is also evidence suggesting that RNA–RNA duplexes are retained within the nucleus (51). Moreover, recent reports demonstrated positively co-regulated expression of SAS pairs at the levels of both transcription and translation (52). Given that asRNAs can only exert their regulatory functions when their target RNAs are simultaneously present within the cell, the co-expression of SAS pairs in *P. falciparum* (which is also seen in other systems) fulfills a prerequisite necessary for their biological function (Figure 4B; Supplementary Figure S3). However, in case of the *P. falciparum* PF6967 gene, the negatively correlated expression of the sense gene and that of the asRNA candidate might represent the first example of transcriptional interference observed between a sense CDS-encoding gene and a small asRNA candidate in *P. falciparum* (Figure 4B). Certainly, further studies are needed to unravel mechanistic details concerning this antisense transcriptome; however, the results presented here draw a complex picture of the small stable asRNAs in the parasite. The number of asRNAs involved in such a broad range of biochemical pathways suggests a fundamental regulatory network that may compensate in part for the absence of defined transcriptional control elements within the A/T-rich genome of *P. falciparum*.

Small npcRNA species in mitochondria: fragmented rRNA

To date, the *P. falciparum* mitochondrion contains the smallest known genome of only 6 kb (14). Apart from three protein-coding genes, approximately 26 npcRNAs have been identified (53). Based on sequence and structural conservation, 20 of the identified npcRNAs previously have been reported to contribute to the formation of the mitochondrial ribosome featuring highly fragmented large rRNAs (14,53). Thus far, 12 rRNA fragments were reported to be part of the large ribosomal subunit, and eight to participate in the formation of the small ribosomal subunit (53). The actual order of the npcRNAs coding for the ribosomal fragments in the mitochondrial genome is not congruent with the arrangement within the functional ribosome. The sizes of known rRNA fragments ranged between 25 and 100 nt (54). Here, we identified 112 npcRNA candidates. Many of them

are novel, whereas the remaining are identical or overlap the previously described mitochondrial npcRNAs (Supplementary Table S11; Supplementary Figures S4 and S5). For most of the known npcRNAs, we observed ragged 5'- and 3'-ends (Supplementary Figures S4 and S5; Supplementary Table S11). In addition, we report longer RNAs that span more than one mature candidate. In fact, and in agreement with previous reports on RNA processing, most of the novel mitochondrial npcRNAs harbor 3'-terminal oligo(A) tracts (Supplementary Table S11). All novel npcRNA candidates appear to be oligoadenylated and the size of the oligo(A)-tails vary between 3 and 27 nt per candidate. In contrast to previous reports, we detected substantial differences, with respect to oligo(A)-tail length, between cDNAs of the same RNA. About 12 RNA candidates completely lack an oligo(A) tract, but mainly are represented by a single cDNA clone. Furthermore, we verified two rRNA fragments that form short helices, which, based on sequence and structural conservation, were computationally predicted to participate in the formation of the active GTP center of the large subunit of rRNA (Supplementary Figure S4) (53). Moreover, 10 npcRNA candidates not evidently associated with rRNAs are transcribed from loci intermingled with large or small subunit-coding regions (Supplementary Figure S4; Supplementary Table S11). The degree of fragmentation observed for *P. falciparum* mitochondrial rRNAs is far higher than reported earlier (Supplementary Figures S4 and S5; Supplementary Table S11) (14). This suggests a possible combinatorial model to generate different RNA structures from a given sequence, depending on where a given fragment of RNA is processed endonucleolytically (Supplementary Figures S4 and S5; Supplementary Table S11). Following this idea, the fragmentation may theoretically compensate for the small size of the *P. falciparum* mitochondrial genome. In agreement with this idea, the ladder-like Northern blot results indicate that even the longer, precursor-like RNAs display a significant degree of RNA stability, maybe implying that they are also involved in the formation of ribosomal subunits, rather than representing mere precursor-like RNAs or processing intermediates (Supplementary Figures S4 and S5; Supplementary Table S11).

CONCLUSIONS

In summary, our data provide the first global experimental survey of the *P. falciparum* small nonprotein-coding transcriptome. Indications are that the small npcRNAs form a complex regulatory network in *P. falciparum* that may participate in the control of gene expression at all levels and may play a pivotal role in antigenic variation and other pathogenic and virulence mechanisms that require the finely tuned expression of various gene families.

ACCESSION NUMBERS

GenBank accession numbers: GQ281811–GQ282500.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

We thank Marsha Bundman and Stephanie Klco-Brosius for editorial assistance.

FUNDING

Nationales Genomforschungsnetz (NGFNII-EP 0313358A and NGFNIII 01GS0808 to J.B. and T.S.R.); European Commission under the apices of the BIOMALPAR Center of Excellence (to M.L.); Scientific Advancement Grant Allocation from the Academy of Science Malaysia (SAGA: 30/INFORMM/6153005A1184 to T.H.T.); M.L. is a member of the Cellnetworks Excellence cluster. Funding for open access charge: NGFNIII 01GS0808.

Conflict of interest statement. None declared.

REFERENCES

- Stein, W.D., Sanchez, C.P. and Lanzer, M. (2009) Virulence and drug resistance in malaria parasites. *Trends Parasitol.*, **25**, 441–443.
- Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S. *et al.* (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, **419**, 498–511.
- Moazed, D. (2009) Small RNAs in transcriptional gene silencing and genome defence. *Nature*, **457**, 413–420.
- Cam, H.P., Chen, E.S. and Grewal, S.I. (2009) Transcriptional scaffolds for heterochromatin assembly. *Cell*, **136**, 610–614.
- Prasanth, K.V. and Spector, D.L. (2007) Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. *Genes Dev.*, **21**, 11–42.
- Royo, H., Bortolin, M.L., Seitz, H. and Cavaille, J. (2006) Small non-coding RNAs and genomic imprinting. *Cytogenet. Genome Res.*, **113**, 99–108.
- Brosius, J. (2005) Waste not, want not—transcript excess in multicellular eukaryotes. *Trends Genet.*, **21**, 287–288.
- Chakrabarti, K., Pearson, M., Grate, L., Sterne-Weiler, T., Deans, J., Donohue, J.P. and Ares, M. Jr. (2007) Structural RNAs of known and unknown function identified in malaria parasites by comparative genomics and RNA analysis. *RNA*, **13**, 1923–1939.
- Mourier, T., Carret, C., Kyes, S., Christodoulou, Z., Gardner, P.P., Jeffares, D.C., Pinches, R., Barrell, B., Berriman, M., Griffiths-Jones, S. *et al.* (2008) Genome-wide discovery and verification of novel structured RNAs in *Plasmodium falciparum*. *Genome Res.*, **18**, 281–292.
- Li, F., Sonbuchner, L., Kyes, S.A., Epp, C. and Deitsch, K.W. (2008) Nuclear non-coding RNAs are transcribed from the centromeres of *Plasmodium falciparum* and are associated with centromeric chromatin. *J. Biol. Chem.*, **283**, 5692–5698.
- Baum, J., Papenfuss, A.T., Mair, G.R., Janse, C.J., Vlachou, D., Waters, A.P., Cowman, A.F., Crabb, B.S. and de Koning-Ward, T.F. (2009) Molecular genetics and comparative genomics reveal RNAi is not functional in malaria parasites. *Nucleic Acids Res.*, **37**, 3788–3798.
- DeChiara, T.M. and Brosius, J. (1987) Neural BC1 RNA: cDNA clones reveal nonrepetitive sequence content. *Proc. Natl Acad. Sci. USA*, **84**, 2624–2628.
- Gillespie, D.E., Salazar, N.A., Rehkopf, D.H. and Feagin, J.E. (1999) The fragmented mitochondrial ribosomal RNAs of *Plasmodium falciparum* have short A tails. *Nucleic Acids Res.*, **27**, 2416–2422.
- Feagin, J.E. (2000) Mitochondrial genome diversity in parasites. *Int. J. Parasitol.*, **30**, 371–390.

15. Kiss, T. (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell*, **109**, 145–148.
16. Lestrade, L. and Weber, M.J. (2006) snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res.*, **34**, D158–D162.
17. Piekna-Przybylska, D., Decatur, W.A. and Fournier, M.J. (2007) New bioinformatic tools for analysis of nucleotide modifications in eukaryotic rRNA. *RNA*, **13**, 305–312.
18. Tiedge, H. (2006) K-turn motifs in spatial RNA coding. *RNA Biol.*, **3**, 133–139.
19. Oquendo, P., Goman, M., Mackay, M., Langsley, G., Walliker, D. and Scaife, J. (1986) Characterisation of a repetitive DNA sequence from the malaria parasite, *Plasmodium falciparum*. *Mol. Biochem. Parasitol.*, **18**, 89–101.
20. Figueiredo, L.M., Pirrit, L.A. and Scherf, A. (2000) Genomic organisation and chromatin structure of *Plasmodium falciparum* chromosome ends. *Mol. Biochem. Parasitol.*, **106**, 169–174.
21. Horard, B. and Gilson, E. (2008) Telomeric RNA enters the game. *Nat. Cell Biol.*, **10**, 113–115.
22. Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E. and Lingner, J. (2007) Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science*, **318**, 798–801.
23. Schoeftner, S. and Blasco, M.A. (2008) Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat. Cell Biol.*, **10**, 228–236.
24. Iyer, L.M., Anantharaman, V., Wolf, M.Y. and Aravind, L. (2008) Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int. J. Parasitol.*, **38**, 1–31.
25. Camblong, J., Iglesias, N., Fickentscher, C., Dieppl, G. and Stutz, F. (2007) Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*. *Cell*, **131**, 706–717.
26. Duraisingh, M.T., Voss, T.S., Marty, A.J., Duffy, M.F., Good, R.T., Thompson, J.K., Freitas-Junior, L.H., Scherf, A., Crabb, B.S. and Cowman, A.F. (2005) Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*. *Cell*, **121**, 13–24.
27. Freitas-Junior, L.H., Hernandez-Rivas, R., Ralph, S.A., Montiel-Condado, D., Ruvalcaba-Salazar, O.K., Rojas-Meza, A.P., Mancio-Silva, L., Leal-Silvestre, R.J., Gontijo, A.M., Shorte, S. *et al.* (2005) Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell*, **121**, 25–36.
28. Lopez-Rubio, J.J., Mancio-Silva, L. and Scherf, A. (2009) Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. *Cell Host Microbe*, **5**, 179–190.
29. Huertas, P. and Aguilera, A. (2003) Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol. Cell*, **12**, 711–721.
30. Freitas-Junior, L.H., Bottius, E., Pirrit, L.A., Deitsch, K.W., Scheidig, C., Guinet, F., Nehrbass, U., Wellem, T.E. and Scherf, A. (2000) Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature*, **407**, 1018–1022.
31. Taylor, H.M., Kyes, S.A. and Newbold, C.I. (2000) Var gene diversity in *Plasmodium falciparum* is generated by frequent recombination events. *Mol. Biochem. Parasitol.*, **110**, 391–397.
32. Werner, A. and Berdal, A. (2005) Natural antisense transcripts: sound or silence? *Physiol. Genomics*, **23**, 125–131.
33. Werner, A. (2005) Natural antisense transcripts. *RNA Biol.*, **2**, 53–62.
34. Beiter, T., Reich, E., Williams, R.W. and Simon, P. (2009) Antisense transcription: a critical look in both directions. *Cell Mol. Life Sci.*, **66**, 94–112.
35. Tang, T.H., Bachellerie, J.P., Rozhdzestvensky, T., Bortolin, M.L., Huber, H., Drungowski, M., Elge, T., Brosius, J. and Huttenhofer, A. (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl Acad. Sci. USA*, **99**, 7536–7541.
36. Brantl, S. (2007) Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr. Opin. Microbiol.*, **10**, 102–109.
37. Lapidot, M. and Pilpel, Y. (2006) Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. *EMBO Rep.*, **7**, 1216–1222.
38. Patankar, S., Munasinghe, A., Shoaibi, A., Cummings, L.M. and Wirth, D.F. (2001) Serial analysis of gene expression in *Plasmodium falciparum* reveals the global expression profile of erythrocytic stages and the presence of anti-sense transcripts in the malarial parasite. *Mol. Biol. Cell*, **12**, 3114–3125.
39. Militello, K.T., Patel, V., Chessler, A.D., Fisher, J.K., Kasper, J.M., Gunasekera, A. and Wirth, D.F. (2005) RNA polymerase II synthesizes antisense RNA in *Plasmodium falciparum*. *RNA*, **11**, 365–370.
40. Gunasekera, A.M., Patankar, S., Schug, J., Eisen, G., Kissinger, J., Roos, D. and Wirth, D.F. (2004) Widespread distribution of antisense transcripts in the *Plasmodium falciparum* genome. *Mol. Biochem. Parasitol.*, **136**, 35–42.
41. Scherf, A., Lopez-Rubio, J.J. and Riviere, L. (2008) Antigenic variation in *Plasmodium falciparum*. *Annu. Rev. Microbiol.*, **62**, 445–470.
42. Lavstsen, T., Salanti, A., Jensen, A.T., Arnot, D.E. and Theander, T.G. (2003) Sub-grouping of *Plasmodium falciparum* 3D7 var genes based on sequence analysis of coding and non-coding regions. *Malar. J.*, **2**, 27.
43. Kyes, S.A., Kraemer, S.M. and Smith, J.D. (2007) Antigenic variation in *Plasmodium falciparum*: gene organization and regulation of the var multigene family. *Eukaryot. Cell*, **6**, 1511–1520.
44. Frank, M. and Deitsch, K. (2006) Activation, silencing and mutually exclusive expression within the var gene family of *Plasmodium falciparum*. *Int. J. Parasitol.*, **36**, 975–985.
45. Ralph, S.A., Bischoff, E., Mattei, D., Sismeiro, O., Dillies, M.A., Guigon, G., Coppee, J.Y., David, P.H. and Scherf, A. (2005) Transcriptome analysis of antigenic variation in *Plasmodium falciparum*—var silencing is not dependent on antisense RNA. *Genome Biol.*, **6**, R93.
46. Adam, R.D. (2001) Biology of *Giardia lamblia*. *Clin. Microbiol. Rev.*, **14**, 447–475.
47. Prucca, C.G., Slavin, I., Quiroga, R., Elias, E.V., Rivero, F.D., Saura, A., Carranza, P.G. and Lujan, H.D. (2008) Antigenic variation in *Giardia lamblia* is regulated by RNA interference. *Nature*, **456**, 750–754.
48. Chen, J., Sun, M., Hurst, L.D., Carmichael, G.G. and Rowley, J.D. (2005) Genome-wide analysis of coordinate expression and evolution of human cis-encoded sense-antisense transcripts. *Trends Genet.*, **21**, 326–329.
49. Prescott, E.M. and Proudfoot, N.J. (2002) Transcriptional collision between convergent genes in budding yeast. *Proc. Natl Acad. Sci. USA*, **99**, 8796–8801.
50. Lazar, M.A., Hodin, R.A., Cardona, G. and Chin, W.W. (1990) Gene expression from the c-erbA alpha/Rev-ErbA alpha genomic locus. Potential regulation of alternative splicing by opposite strand transcription. *J. Biol. Chem.*, **265**, 12859–12863.
51. Tosic, M., Roach, A., de Rivaz, J.C., Dolivo, M. and Matthieu, J.M. (1990) Post-transcriptional events are responsible for low expression of myelin basic protein in myelin deficient mice: role of natural antisense RNA. *EMBO J.*, **9**, 401–406.
52. Zhang, Q., Zhang, J., Moe, O.W. and Hsia, C.C. (2008) Synergistic upregulation of erythropoietin receptor (EPO-R) expression by sense and antisense EPO-R transcripts in the canine lung. *Proc. Natl Acad. Sci. USA*, **105**, 7612–7617.
53. Feagin, J.E., Mericle, B.L., Werner, E. and Morris, M. (1997) Identification of additional rRNA fragments encoded by the *Plasmodium falciparum* 6kb element. *Nucleic Acids Res.*, **25**, 438–446.
54. Feagin, J.E., Werner, E., Gardner, M.J., Williamson, D.H. and Wilson, R.J. (1992) Homologies between the contiguous and fragmented rRNAs of the two *Plasmodium falciparum* extrachromosomal DNAs are limited to core sequences. *Nucleic Acids Res.*, **20**, 879–887.