Structural RNAs of known and unknown function identified in malaria parasites by comparative genomics and RNA analysis

KAUSIK CHAKRABARTI,[1,2] MICHAEL PEARSON,[1] LESLIE GRATE,[1] TIMOTHY STERNE-WEILER,[2] JONATHAN DEANS,[2] JOHN PAUL DONOHUE,[1] and MANUEL ARES JR.[1,2]

[1]Department of Molecular, Cell and Developmental Biology, Center for Molecular Biology of RNA, University of California at Santa Cruz, Santa Cruz, California 95064, USA
[2]Hughes Undergraduate Research Laboratory, University of California Santa Cruz, Santa Cruz, California 95064, USA

ABSTRACT

As the genomes of more eukaryotic pathogens are sequenced, understanding how molecular differences between parasite and host might be exploited to provide new therapies has become a major focus. Central to cell function are RNA-containing complexes involved in gene expression, such as the ribosome, the spliceosome, snoRNAs, RNase P, and telomerase, among others. In this article we identify by comparative genomics and validate by RNA analysis numerous previously unknown structural RNAs encoded by the *Plasmodium falciparum* genome, including the telomerase RNA, U3, 31 snoRNAs, as well as previously predicted spliceosomal snRNAs, SRP RNA, MRP RNA, and RNAse P RNA. Furthermore, we identify six new RNA coding genes of unknown function. To investigate the relationships of the RNA coding genes to other genomic features in related parasites, we developed a genome browser for *P. falciparum* (http://areslab.ucsc.edu/cgi-bin/hgGateway). Additional experiments provide evidence supporting the prediction that snoRNAs guide methylation of a specific position on U4 snRNA, as well as predicting an snRNA promoter element particular to *Plasmodium sp.* These findings should allow detailed structural comparisons between the RNA components of the gene expression machinery of the parasite and its vertebrate hosts.

Keywords: *Plasmodium*; malaria; RNA coding genes; malaria genome browser; telomerase RNA; snoRNA; snRNA; RUF

INTRODUCTION

Malaria is expanding as one of the world's most lethal human infections (Webster 2001). To date, no effective vaccine has been developed for malaria and drug resistance is a major challenge worldwide (White 2004). Given this situation and recent advances in genomics, there has been renewed interest in the fundamental biology of the *Plasmodium* species that cause malaria in humans and animals. As genomic information accumulates in raw sequence form, correct annotation of the genome becomes an important step in order to capture the biological importance of the sequences. Compared to protein coding gene annotation, less attention is usually given to the annotation of RNA coding genes. RNA coding genes typically do not present long primary sequence features like protein coding genes (Eddy 2001; McCutcheon and Eddy 2003; Washietl et al. 2005), yet structural RNAs play major roles in cells, and the *Plasmodium* parasite is no exception (Levitt 1993).

Structural RNAs and RNPs are attractive targets for therapeutic development (DeJong et al. 2002; Goodchild 2004). For example, many of the best antibacterial antibiotics target the ribosome by binding to ribosomal RNA at sites that differ slightly between prokaryotes and eukaryotes (Recht et al. 1999; Lynch et al. 2003). These slight structural differences result in exquisite selectivity of drug action (Fourmy et al. 1996; Silva and Carvalho 2007). It seems possible that apart from RNA–protein interactions, many subtle but potentially crucial differences between human and pathogens must exist in the ribosomal, spliceosomal, and other important RNA structures. Given such differences there may be compounds that selectively inhibit pathogen RNA-based processes.

P. falciparum has an extremely AT-rich genome of about 26 megabases (Mb) organized in 14 linear chromosomes (Gardner et al. 2002). Although the sequences of

Reprint requests to: Manuel Ares Jr., Molecular, Cell and Developmental Biology, University of California at Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA; e-mail: ares@biology.ucsc.edu; fax: (831) 459-3737.

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.751807.

P. falciparum and the related rodent parasite *P. yoelii yoelii* are essentially complete, those of the other *Plasmodium* species are nearing completion (Coppel et al. 2004; Kooij et al. 2006). Initial attempts have been made to predict and annotate RNA coding genes of the parasite (Upadhyay et al. 2005); however more comprehensive efforts using whole genome comparisons and experimental validation are necessary to complete this process.

To help address the question of where the differences in key RNA structures lie between vertebrate and pathogen, we have analyzed the genome of *P. falciparum* (vertebrate host: human) and compared it to those of *P. vivax* (human), *P. reichenowi* (chimpanzee), *P. knowlesi* (monkey), *P. gallinaceum* (chicken), *P. chabaudi* (rodent), *P. berghei* (rodent), and *P. yoelii yoelii* (rodent), searching for RNA coding genes. In this article we detect and map the expressed transcripts of *Plasmodium* telomerase RNA, U6 snRNA, U3 snoRNA, 27 C/D box, and 4 H/ACA box snoRNAs involved in rRNA and snRNA modifications, as well as mapping and validating the expression of the previously identified U1, U2, U4, and U5 snRNAs (Gardner et al. 2002; Upadhyay et al. 2005), RNAse MRP RNA (Piccinelli et al. 2005), RNase P RNA (Piccinelli et al. 2005), and the signal recognition particle (SRP) RNA (Zwieb et al. 2005). Additionally, we identify six novel RNAs of unknown function (RUFs), one of which is encoded by a mulitgene family whose members are often located next to cell surface protein genes. In the case of a C/D box RNA predicted to methylate U4 snRNA, we provide in vivo evidence for the methylation of its target. Comparison of the DNA near U1, U2, U3, U4, and U5, but not U6 snRNAs identified a very clear conserved sequence element we propose is a *Plasmodium* snRNA promoter element (SNPE). This study provides the first comprehensive search for structural RNAs in the genomes of *Plasmodium sp.* and lays the foundation for exploration of structural RNAs as therapeutic targets for prevention and treatment of malaria.

RESULTS AND DISCUSSION

To find structural RNAs, we used available *Plasmodium* sequences from PlasmoDB (Bahl et al. 2003) and several pattern searching methods (see Materials and Methods). Our studies were aided by the use of a genome browser (http://areslab.ucsc.edu/cgi-bin/hgGateway) for *P. falciparum* set up on our laboratory server using UCSC Genome Browser software (Karolchik et al. 2003). The browser is public and displays alignments of published genome sequences of *P. falciparum* (Gardner et al. 2002) and seven other species of malaria parasites (Bahl et al. 2003). On the browser we display conservation track (Siepel et al. 2005; this study) gene predictions (Bahl et al. 2003; Berriman and Rutherford 2003; Li et al. 2003; Hall et al. 2005), available *P. falciparum* ESTs (Chakrabarti et al. 1994; Bahl et al. 2003; Watanabe et al. 2004), microarray data (Bozdech et al. 2003; Le Roch et al. 2003), proteome data (Florens et al. 2002; Lasonder et al. 2002), and an RNA gene track (this study). As we identified potential RNA coding genes by bioinformatics, we verified their expression by Northern blots and mapped their transcripts in *P. falciparum* using primer extension or RNase protection assays. Following end mapping, we aligned the *P. falciparum* RNA gene sequence to the other *Plasmodium* sequences to obtain comparative information about the primary sequence and secondary structure of the RNA. In some cases this allowed reasonable secondary structure prediction. Below we describe this analysis for the RNAs we detected.

Telomerase RNA

Telomeres are specific nucleoprotein complexes that protect the termini of eukaryotic linear chromosomes from degradation, end-to-end fusion, and undesired recombination (de Lange 2002; Wong and Collins 2003; Blasco 2005). Subtelomeric regions of malaria parasites have been studied extensively since they harbor gene families responsible for antigenic variation (*var* genes in *P. falciparum*, *vir* genes in *Plasmodium vivax*, and *Py235* gene families in *Plasmodium yoelii yoelii*) (Smith et al. 1995; del Portillo et al. 2001; Scherf et al. 2001; Duraisingh et al. 2005). Because of its dependence on RNA primers, replication causes a progressive loss of the ends of linear DNA at telomeres (Greider and Blackburn 1987; Engelhardt and Martens 1998). This loss is compensated by the addition of repeats by a telomerase reverse transcriptase (TERT) activity, using telomerase RNA as a template and the end of the chromosome as a primer (Blackburn 1999). Telomerase activity has been reported in cell extracts of *P. falciparum* (Aldous et al. 1998; Bottius et al. 1998; Raj et al. 2003). Recently, a putative telomerase reverse transcriptase (TERT) protein was identified in *P. falciparum* (Figueiredo et al. 2005); however the telomerase RNA remained unidentified.

The telomeric repeat for *P. falciparum* is TT(T/C)AGGG (Vernick and McCutchan 1988). To find the telomerase RNA gene, we identified genomic DNA containing a predicted template sequence ACCCTGAACCC using known telomerase template-repeat relationships as a guide (Cano et al. 1999; Theimer and Feigon 2006). We eliminated telomeric sequences themselves and extracted the genomic sequence, including the several hundred nucleotides of sequence to either side. We then used each of these larger regions as BLAST query sequences against each of the other *Plasmodium* genomes. The *P. falciparum* locus that gave the highest scoring blast hits on most other *Plasmodium* genomes was on chromosome 9. These loci were found to align well to each other, with the putative template sequence embedded in a highly conserved region.

To demonstrate that this genomic region encodes telomerase RNA, we performed Northern blots on *P. falciparum* total RNA (Fig. 1a) using two oligonucleotide probes,

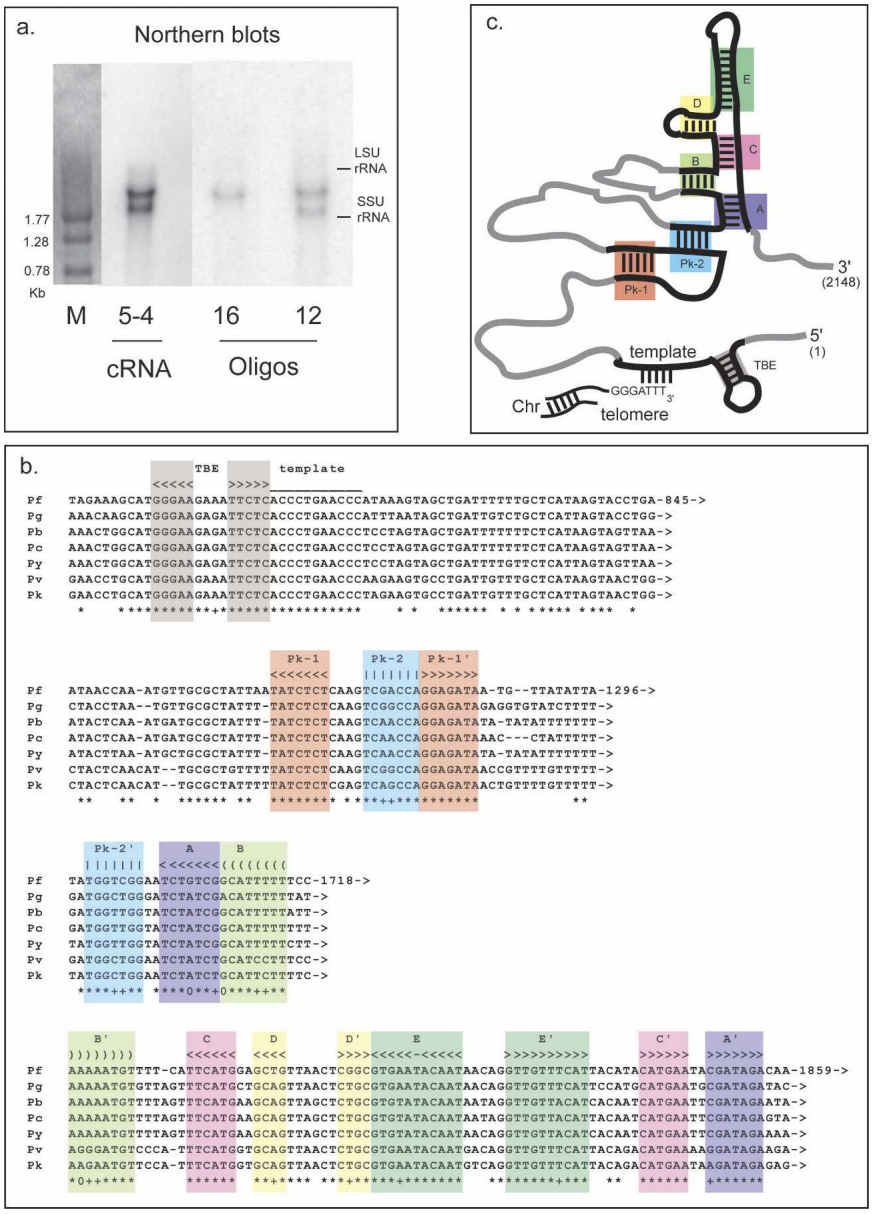

FIGURE 1. Telomerase RNA in *P. falciparum*. (*a*) Northern hybridization with *P. falciparum* telomerase RNA specific oligonucleotides (probe 12 and probe 16) and a complementary RNA (cRNA) probe (probe 5-4, representing genomic coordinates chr9:763,006–763,193). (M) RNA molecular weight marker. Positions of LSU and SSU rRNAs were determined by staining the blot with methylene blue after Northern transfer (see Materials and Methods). (*b*) Multiple sequence alignment of the highly conserved regions in *Plasmodium* telomerase RNA. (*) invariant residues; (+) pairs of compensatory changes; (0) a single base change that converts a Watson–Crick base apposition to or from a G-U apposition (this is technically not covariation). Corresponding color-coded regions are proposed to pair. (*c*) Secondary structure model of *P. falciparum* telomerase RNA. A, B, C, D, E, Pk-1 and Pk-2 are stem forming regions, color coded to match the corresponding boxes from multiple sequence alignment, described above. (Chr) chromosome; (TBE) template boundary element.

one 142 nucleotides (nt) downstream from the template region (probe 12) and one 700 nt upstream of the template sequence (Fig. 1a, probe 16; see legend for probe details). Two transcripts were detected with probe 12, whereas only a single transcript was detected with probe 16 (Fig. 1a). We also used a long complementary RNA (cRNA) probe 5-4, which corresponds to ~1000 nt downstream from the template sequence and showed a hybridization pattern identical to oligo probe 12 (Fig. 1a). To map the 5′ and 3′ ends of the telomerase RNA transcripts, we used RNase protection assays (RPAs) with cRNA probes (Supplemental Fig. 1). Based on this, the most abundant *Plasmodium* telomerase RNA was 2.2 kb long. Shorter forms with different 5′ ends were detected, and it was unclear which of the RNAs might be functional.

Telomerase RNA is very divergent across eukaryotic species (e.g., ciliates and yeasts or human) (Lingner et al. 1994; Legassie and Jarstfer 2006). We used the longest mapped RNA to generate alignments for *Plasmodium sp.* telomerase RNA (Fig. 1b). We identified patterns of sequence covariation with the help of two programs (Siebert and Backofen 2005; Washietl et al. 2005) and used these to begin modeling phylogenetically conserved RNA secondary structure. A part of the *P. falciparum* telomerase RNA sequence could be folded into three conserved structural domains and a conserved single stranded region that contains the putative template sequence (Fig. 1c). The two major *P. falciparum* telomeric repeats are 5′-TTCAGGG-3′ and 5′-TTTAGGG-3′; thus the G in the template appears to specify C or T at the third position of the telomeric repeat. Immediately 5′ of the template sequence is a 5 base-pair (bp) stem with a GNRA tetraloop that could be a template boundary element (TBE), similar to one described for ciliate telomerase RNA (Autexier and Greider 1995). In *Tetrahymena*, the TBE contains a G-C base pair essential for proper template boundary definition and required for binding of the telomerase core protein–telomerase reverse transcriptase (TERT) (Richards et al. 2006), and this G-C pair is present in the *Plasmodium* RNAs.

Potential to form a pseudoknot is a conserved feature of yeast, ciliate, and vertebrate telomerase RNA (Gilley and Blackburn 1999; Chen et al. 2000; Dandjinou et al. 2004), and disruption of the pseudoknot in ciliates prevents the stable assembly of a catalytically active telomerase in vivo

(Gilley and Blackburn 1999). Two conserved pairs of complementary regions in *Plasmodium* telomerase RNA, Pk-1 and Pk-2 (Fig. 1c), could form a pseudoknot. In our model, formation of the Pk-2 helix in *Plasmodium sp.* is supported by one pair of compensatory mutations. In contrast to mammals, the *Plasmodium* pseudoknot loop is shorter, consisting of only 4 nt, and more similar to that found in ciliates (McCormick-Graham and Romero 1995).

The third conserved structure of *Plasmodium* telomerase comprises five stems (Fig. 1c, A–E). Part of this structure (Fig. 1c, E–E′ stem) resembles the p6.1 stem of the CR4/CR5 domain in human telomerase RNA. Like the human p6.1 hairpin, this helix also has a G-U wobble base pair at the base of the loop. Studies in human have shown the G-U base pair in corresponding stem formation and also that the hairpin is essential for TERT binding and activity (Chen et al. 2002; Moriarty et al. 2004). The secondary structure model (Fig. 1c) is a conservative one that is unlikely to be complete. Additional data (Supplemental Fig. 1) suggest that there are other structured regions of the molecule, but the nature of these structures cannot be discerned reliably with the present data.

Methylation and pseudouridylation (C/D and H/ACA box) guide snoRNAs

C/D box and H/ACA box snoRNA families guide the 2′-*O*-ribose methylations and pseudouridylations, respectively, on rRNAs and spliceosomal snRNAs in vertebrates (Kiss 2001), tRNAs in Archaea (Xie et al. 2007), and even probably eukaryotic mRNAs (Cavaille et al. 2000; Kishore and Stamm 2006). Among the 31 candidate snoRNAs we identified, 27 were C/D box type and 4 were H/ACA type snoRNAs (Table 1). Twelve of the 31 snoRNAs whose

TABLE 1. Various features of C/D and H/ACA snoRNAs in malaria parasite *P. falciparum* and their orthologs in Human and Yeast

					Homologs	
Name	Type	Location	Host gene, single or cluster	Predicted modification sites	Human	Yeast
snoR01	C/D	intergenic	Cluster (1)? (with RNAseP and RUF2)	U: 3572 LSU, U: 1624 SSU		
snoR02	C/D	intergenic	Single	G: 3315 LSU, U: 1370 SSU, U: 81 U1 snRNA		
snoR03	H/ACA	intronic	PFC0290w	U: 1292 SSU		
snoR04	C/D	intergenic	PFE0543c	C: 2950 LSU		
snoR05	C/D	Intronic	PF08_0019	A: 3307 LSU	U29	snR71
snoR06[a]	C/D	intronic	PF08_0124	A: 744 LSU		
snoR07	C/D	intergenic	Single	C: 1936 SSU, U:158 U1snRNA		
snoR08	C/D	intronic	PF11_0105	A: 1264 LSU		
snoR09	C/D	intronic	PF11_0105	C: 3197 LSU, G: 3176 LSU,		snR38
snoR10	C/D	Intronic	PF11_0105	G: 1037 LSU, G: 785 SSU, G: 22 U1snRNA		
snoR11	H/ACA	intronic	PF11_0105	U: 1914 SSU/ U-5 U1snRNA	U109	
snoR12	C/D	intergenic	Single	A: 28 SSU	U27	snR74
snoR13	C/D	intergenic	Single	G: 95 SSU, C: 420 SSU	U14	
snoR14	C/D	intergenic	Single	A: 3417 LSU		
snoR15	C/D	intergenic	Single	A: 66 U4snRNA	U87	
snoR16	C/D	intergenic	Single	tRNA Ser (CGA)		
snoR17	C/D	intronic	PF13_0165	C: 1589 LSU	U24	
snoR18	C/D	intronic	PF13_0171	A: 3248 LSU		
snoR19	C/D	intronic	MAL13P1.182	A:70 U6snRNA, G: 1277 LSU		
snoR20	C/D	intronic	MAL13P1.209	C: 3320 LSU	U35	snR73
snoR21	C/D	intronic	PF13_0268	G: 2583 LSU, A: 2576 LSU		
snoR22	C/D	intronic	PF14_0027	C: 2632 LSU	U74	snR64
snoR23	C/D	intergenic	Cluster (2)	A: 1043 SSU		
snoR24	C/D	intergenic	Cluster (2)	A: 442 SSU		
snoR25	C/D	intronic	PF14_0230	G: 1674 SSU	U25	snR56
snoR26	C/D	intronic	PF14_0230	G: 2172 LSU		
snoR27	H/ACA	intergenic	Cluster (3)	U: 3187 LSU (1), U: 3241 LSU (2)	U65	snR34
snoR28	C/D	intergenic	Cluster (3)	C: 3248 LSU		
snoR29	C/D	intergenic	Cluster (3)	G: 926 LSU		snR39b
snoR30	C/D	intergenic	Cluster (3)	A: 728 LSU	U18	
snoR31	H/ACA	intergenic	Cluster (3)	U: 1617 SSU		

Each snoRNA shown has a homolog in at least one other *Plasmodium* species in addition to *P. falciparum*.
[a]High confidence prediction that has not yet been validated; see text.

targets we identified have corresponding modification sites predicted on rRNAs or snRNAs in other eukaryotes, and 19 appear so far to be species specific (Table 1). This is in line with findings in Trypanosomes (Liang et al. 2005) and *Euglena gracilis* (Russell et al. 2006), for which 40%–50% of rRNA methylations are species specific.

The organization of *Plasmodium* snoRNA genes is similar to that of other protozoa (Liang et al. 2005), plants (Brown et al. 2003), and yeast (Brown et al. 2001). They are found alone in intergenic regions as well as in the introns of protein coding genes, with intermingling of H/ACA and C/D snoRNA genes. In yeast and plants, intergenic clusters that are cotranscribed are observed (Brown et al. 2003). In *Plasmodium*, there are several examples of small intergenic clusters of snoRNA genes oriented on the same DNA strand. Although this arrangement is suggestive of a cotranscriptional means of expression, this has yet to be shown.

SnoRNA genes identified in *Plasmodium* genome and their predicted modifications sites are listed in Table 1; their characterization is described below.

Methylation gGuide (C/D box) RNAs

We confirmed expression of 26 of the 27 candidate C/D box snoRNAs during the intra-erythrocytic stages of *P. falciparum* by Northern hybridization. Representative Northern blots for selected snoRNAs are shown in Figure 2a. The 5′ ends of the transcripts were mapped by primer extension (Fig. 2b). We could not validate expression of an intronic C/D box RNA candidate, snoR06. However, RUF3 (an intronic RNA of unknown function; Table 2) was detected, arising from a different intron of the same predicted gene, Pf08_0124. Therefore, it is possible that snoR06 RNA either escaped detection by Northern hybridization or only accumulates during some other stage in the life cycle. Among the validated C/D box RNAs, snoR09 is interesting in having a predicted guide region before both the D and D′ boxes (D box: G 3176 LSU rRNA; D′ box: C 3197 LSU rRNA), which possibly modify positions separated by only 23 nt on the LSU rRNA. Many of the snoRNAs (Fig. 2c,d, snoR15 and snoR21) have extended base-paired stems between the D′ and C′ boxes. SnoR21 has two predicted targets only 6 nt apart on LSU rRNA. This suggests the two separate guide sequences on snoR21 would bind overlapping rRNA sequences (Fig. 2d). We could not identify a target sequence for snoR16 on *P. falciparum* 18S or 28S rRNA or any snRNA. However, this C/D box RNA could target tRNA Ser (CGA) on the first position of the tRNA anticodon loop (Table 1). C/D box snoRNAs direct 2′-*O*-ribose methylations at the corresponding position of tRNAs in Archea (Clouet d'Orval et al. 2001).

We identified five C/D box snoRNAs with structure consistent with the ability to guide modification of U1, U4, and U6 spliceosomal snRNAs (Table 1). Mammalian spliceosomal RNAs contain many 2′-*O*-methylated residues, some of which are modified in a snoRNA-dependent fashion (Tycowski et al. 1998, 2004). In *P. falciparum*, snoR02, snoR07, and snoR10 are predicted to target U1 snRNA at nucleotide position 81, 158, and 22, respectively (Table 1). Additionally, snoR15 could guide methylation of U4 snRNA at position 66 (Fig. 2e, Am66), whereas snoR19 is predicted to guide methylation at position 70 (Am70) on U6 snRNA (Table 1). While snoR10 and snoR19 D′ boxes appear to guide modifications on U1 and U6 snRNAs, their D box elements targets SSU and LSU rRNAs, respectively (Table 1). To provide evidence for the modification of U4 at A66, we searched by primer extension for strong stops enhanced under limiting nucleoside triphosphate conditions when native *P. falciparum* (in vivo) but not in vitro transcribed U4 snRNA was used as a template (Fig. 2e). At lower concentration of dNTPs, reverse transcriptase halts 1 nt before the modification sites (Maden 2001). The appearance of a stop after incorporation of a nucleotide opposite template (U4) nucleotide A67 in native but not synthetic, unmodified U4 RNA strongly suggests that A66 on U4 snRNA is modified in vivo (Fig. 2e). Methylation of the homologous residue in human U4 snRNA (Am65) has been demonstrated (Reddy et al. 1981a), and vertebrate U87 RNA is proposed to guide this modification (Darzacq et al. 2002; Lestrade and Weber 2006).

Pseudouridylation guide (H/ACA box) RNAs

Genomic searches for H/ACA snoRNAs, in contrast to the C/D box type, are more challenging due to their shorter primary sequence motifs and bipartite complementary sequences (Ganot et al. 1997; Vitali et al. 2003). We have identified four H/ACA box snoRNAs in *P. falciparum*. All (snoR03, snoR11, snoR27, and snoR31) are expressed in the blood stages of the parasite's life cycle. Among them snoR03, snoR11, and snoR31 are predicted to guide modifications on 18S rRNA, while snoR27 targets 28S rRNA (Fig. 2a,b; Table 1). A secondary structure model for this H/ACA RNA is shown in Figure 2f. Interestingly, the 3′ hairpin of this H/ACA box RNA contains a CAB consensus (UGAG) sequence in the terminal loop (Fig. 2f), a Cajal body-specific localization signal identified in H/ACA scaRNAs (Richard et al. 2003). In addition to rRNA modification, snoR11 is predicted to modify the U at position 5 near the 5′ end of U1 snRNA. In human, the homologous residue (U6) interacts with the 5′ splice site during pre-mRNA splicing, and is pseudouridylated (Reddy et al. 1981b), possibly by an H/ACA scaRNA, U109 (Gu et al. 2005; Lestrade and Weber 2006).

RNAs of unknown function (RUFs)

In our search for RNA coding genes in the *Plasmodium* genome, we identified six new RNAs (Table 2) that do not

readily fall into known RNA functional categories. We report them here as RNAs of unknown function (RUFs). We verified expression of six RUFs by Northern blotting (Fig. 3a–f) and mapped their transcription start sites by primer extension (data not shown). Four of them (Table 2, RUF1–4) appear to share some structural similarity with H/ACA snoRNAs but lack known (i.e., rRNA, snRNA) target complementarity (Fig. 3g, RUF1). The remaining two RUFs (Table 2, RUF5 and RUF6) were identified only in the *P. falciparum* and *P. reichenowi* genomes. RUF6 forms a single stem (Fig. 3h) and is remarkable because it appears to be encoded by multiple genes with an unusual distribution in malaria genome. Consistent with this we observed heterogeneity in RNA length (144–160 nt) for RUF6 RNA on Northern blots (Fig. 3f), although differential processing of transcripts from a single active RUF6 gene could also cause heterogeneity. Members of this gene family are usually clustered and tandemly distributed on several chromosomes (Table 2). In *P. falciparum*, these RUF6 clusters are arranged flanking malarial surface antigens, namely *RIFIN* (Cheng et al. 1998) or *VAR* (Smith et al. 1995) genes at the middle or end of the chromosomes where recombination hotspots are generally clustered (Freitas-Junior et al. 2000; Mu et al. 2005). As an example, RUF6 cluster 4A (Table 2) is flanked by subtelomeric RIFIN antigen PFD1010w, whereas the 4B is organized upstream of RIFIN gene

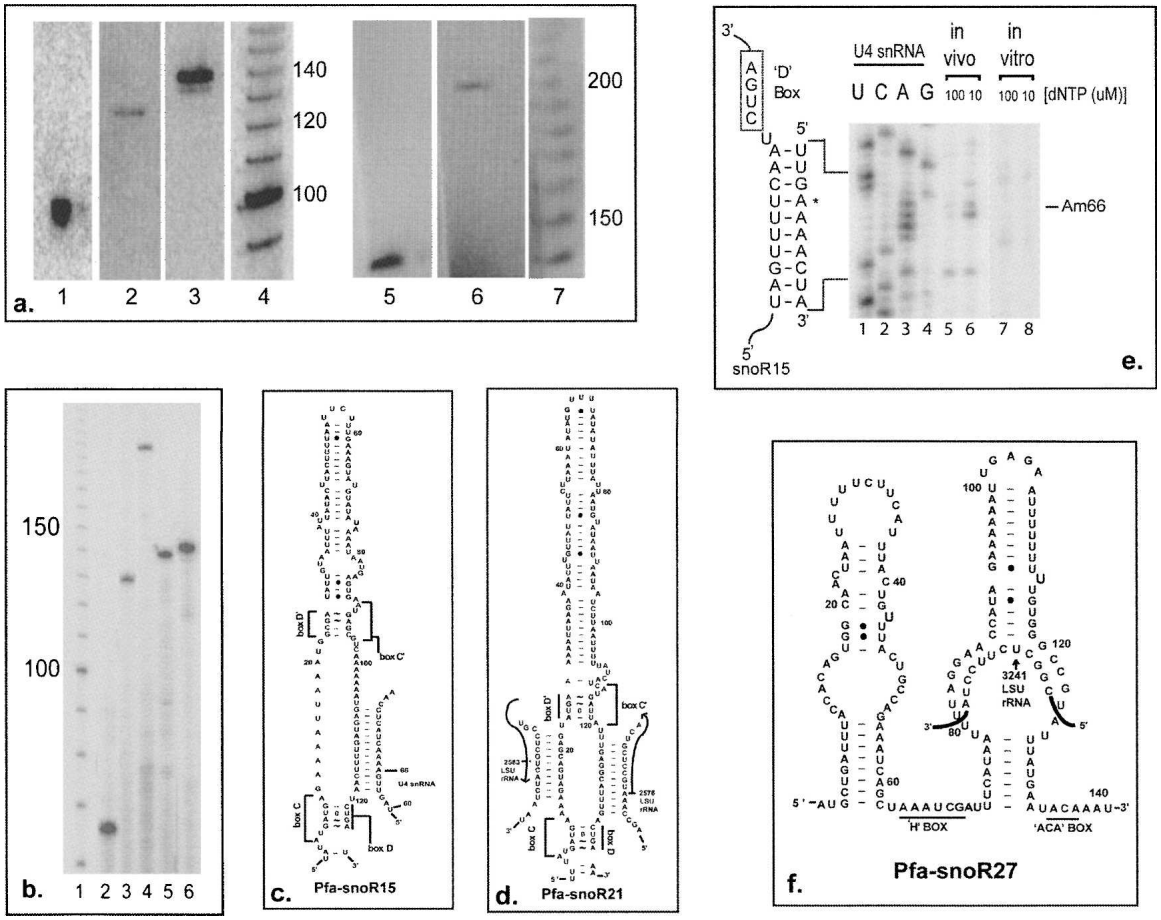

FIGURE 2. Expression, mapping, and secondary structure models of methylation and pseudouridylation guide snoRNAs from *P. falciparum*. (*a*) Northern blots of *P. falciparum* RNA run on denaturing polyacrylamide gels. (Lane *1*) snoR09; (lane *2*) snoR15; (lane *3*) snoR21; (lane *5*) snoR27; (lane *6*) snoR11; (lanes *4*,*7*) 10-bp ladder molecular weight marker. (*b*) Primer extension analysis to map the 5′ end of snoRNA transcripts: (Lane *1*) 10-bp ladder; (lane *2*) snoR09; (lane *3*) snoR15; (lane *4*) snoR11; (lane *5*) snoR21; (lane *6*) snoR27. (*c*) SnoR15 is a C/D box RNA that may guide modification on U4 spliceosomal snRNA. (*d*) SnoR21 is a C/D box snoRNA that contains two guide sequences for the large subunit (LSU) ribosomal RNA. (*e*) Mapping 2′-*O*-methylation sites on U4 snRNA . End-labeled snRNA specific primer was extended with Reverse Trascriptase (RT) in the presence of different amounts of dNTPs (100 μM and 10 μM) using either *P. falciparum* total RNA (lanes *5*,*6*, in vivo) or in vitro transcribed U4 RNA (lanes *7*,*8*). Lanes *1–4* show a sequencing ladder generated by the dideoxy termination method using *P. falciparum* U4 RNA. Modification site at A66 is marked with *. RT halts one nucleotide before this modification site in the presence of low dNTPs. The sequence of snoR15 paired with U4 snRNA is shown alongside sequencing lanes. (*f*) snoR27 is an H/ACA type *Plasmodium* snoRNAs that has an ortholog in yeast (snR34) and human (U65). H and ACA boxes are underlined. The proposed target LSU rRNA sequence is shown in the second hairpin loop near ACA box.

TABLE 2. Genomic locations and properties of RNAs of unknown function (RUFs) in *P. falciparum*

RNA	Northern size	Strand	Arrangement	Comments
RUF1	264	+	Intergenic	Upstream of PF13_0152, Sir2 homolog
RUF2	176	–	Between snoR01 and RNAseP	Conflict with gene prediction PFB0877c
RUF3	187	+	Intronic	Intron of predicted gene Pf08_0124
RUF4	275	+	Between snoR29 and snoR31	Possibly cotranscribed with other snoRNAs
RUF5	280	–	Intergenic	Not conserved, EST evidence
RUF6-1	144–160[a]	+	Intergenic, clustered	Multiple hits in the genome
RUF6-2	144–160[a]	+	Upstream of RIFIN, chr7 cluster	98.8% similar to RUF6-1
RUF6-3	144–160[a]	+	Upstream of RIFIN, chr4A cluster	97.4% similar to RUF6-1
RUF6-4	144–160[a]	–	Downstream from Asparagine-rich antigen, chr8 cluster	96.8% similar to RUF6-1
RUF6-5	144–160[a]	–	Upstream of PFL1975c, chr12 cluster	96.2% similar to RUF6-1
RUF6-6	144–160[a]	+	Upstream of RIFIN, chr4B cluster	96.8% similar to RUF6-1
RUF6-7	144–160[a]	+	Upstream of RIFIN, chr4B cluster	96.8% similar to RUF6-1
RUF6-8	144–160[a]	+	Upstream of RIFIN, chr4A cluster	94.9% similar to RUF6-1
RUF6-9	144–160[a]	+	Upstream of RIFIN, chr4A cluster	94.9% similar to RUF6-1
RUF6-10	144–160[a]	–	Downstream from Asparagine-rich antigen, chr8 cluster	94.9% similar to RUF6-1
RUF6-11	144–160[a]	+	Downstream from RIFIN, chr4B cluster	94.9% similar to RUF6-1
RUF6-12	144–160[a]	+	Upstream of RIFIN, chr7 cluster	94.2% similar to RUF6-1
RUF6-13	144–160[a]	+	Upstream of RIFIN, chr7 cluster	94.2% similar to RUF6-1
RUF6-14	144–160[a]	–	Upstream of PFL1975c, chr12 cluster	93.6% similar to RUF6-1
RUF6-15	144–160[a]	–	Upstream of PFL1975c, chr12 cluster	92.1% similar to RUF6-1

Many *PfEMP1* loci are identified upstream of RUF6 copies on Chr4 and Chr12.
[a]Note that the origin and composition of the RUF6 Northern signal with respect to the 15 RUF6 genes is uncertain.

PFD0645w, located centrally on chromosome 4. Given that RUF6 gene clusters are in close proximity to these loci, it is tempting to speculate that RUF6 transcripts may play a role in expression or maintenance of these malarial antigens. Interestingly, in vertebrates, the 5′ end of the crossover domain of a hotspot locus is known to transcribe a novel RNA transcript (Nishant et al. 2004).

Spliceosomal snRNAs

Since nearly 50% of *P. falciparum* genes are predicted to contain introns (Gardner et al. 2002), the presence of spliceosome components in this eukaryotic genome was expected. The spliceosome contains 5 U-rich small nuclear RNAs (snRNAs)—U1, U2, U4, U5, and U6 (Ares and Weiser 1995; Staley and Guthrie 1998). These five snRNAs were identified in the *P. falciparum* genome by searching the raw genomic sequence using BLASTN (Altschul et al. 1990) with snRNA sequences from other organisms as query (e.g., vertebrates, yeast, *Tetrahymena*, etc.). We also used RNABOB (http://selab.wustl.edu), a pattern searching tool that allowed searches for RNA structural motifs.

Each region of the *P. falciparum* genome identified by one of these methods was subsequently used as a query on the other *Plasmodium* sequences, and regions found to contain extended similarity among *Plasmodium* genomes were aligned and examined for conservation of primary and secondary structure characteristic of the RNA used as an initial query. In this way we found convincing candidate genes for each of the snRNAs in most of the malaria parasite genomes (see also Upadhyay et al. 2005). To verify and map the predicted RNAs, we determined the RNA length by Northern blotting (Fig. 4a), mapped the 5′ end by primer extension (Fig. 4b), and inferred the approximate 3′ end. *Plasmodium* snRNAs are similar in size to those of vertebrates and appear capable of folding into the same overall conformations (Fig. 4c–f) predicted for the metazoan and fungal snRNAs, including the ability to undergo similar kinds of dynamic rearrangements necessary for spliceosome assembly and catalytic activation (Ares and Weiser 1995; Staley and Guthrie 1998). We have compared the snRNA sequences of *P. falciparum* to seven other species of malaria parasites as well as to human and other vertebrate species (sequence alignments between *Plasmodium* species, human, and mouse snRNAs are shown in Supplemental Figs. 2–6). Differences in several key regions of the vertebrate and *Plasmodium* snRNA secondary structures are described below.

Malaria parasite U1 snRNA secondary structure (Fig. 4c) is more like metazoan U1 than the yeast U1 homolog SNR19 (Kretzner et al. 1987; Siliciano et al. 1987). The part of *Plasmodium* U1 predicted to base pair with the 5′ splice site is identical to vertebrate U1. Primary sequences of helix I of U1 are considerably different between vertebrates and malaria parasites. *Plasmodium* helix II has G54A55 opposite G82A83 stack, compared to a single AG pair found in the homologous position of vertebrate U1. The apical loop

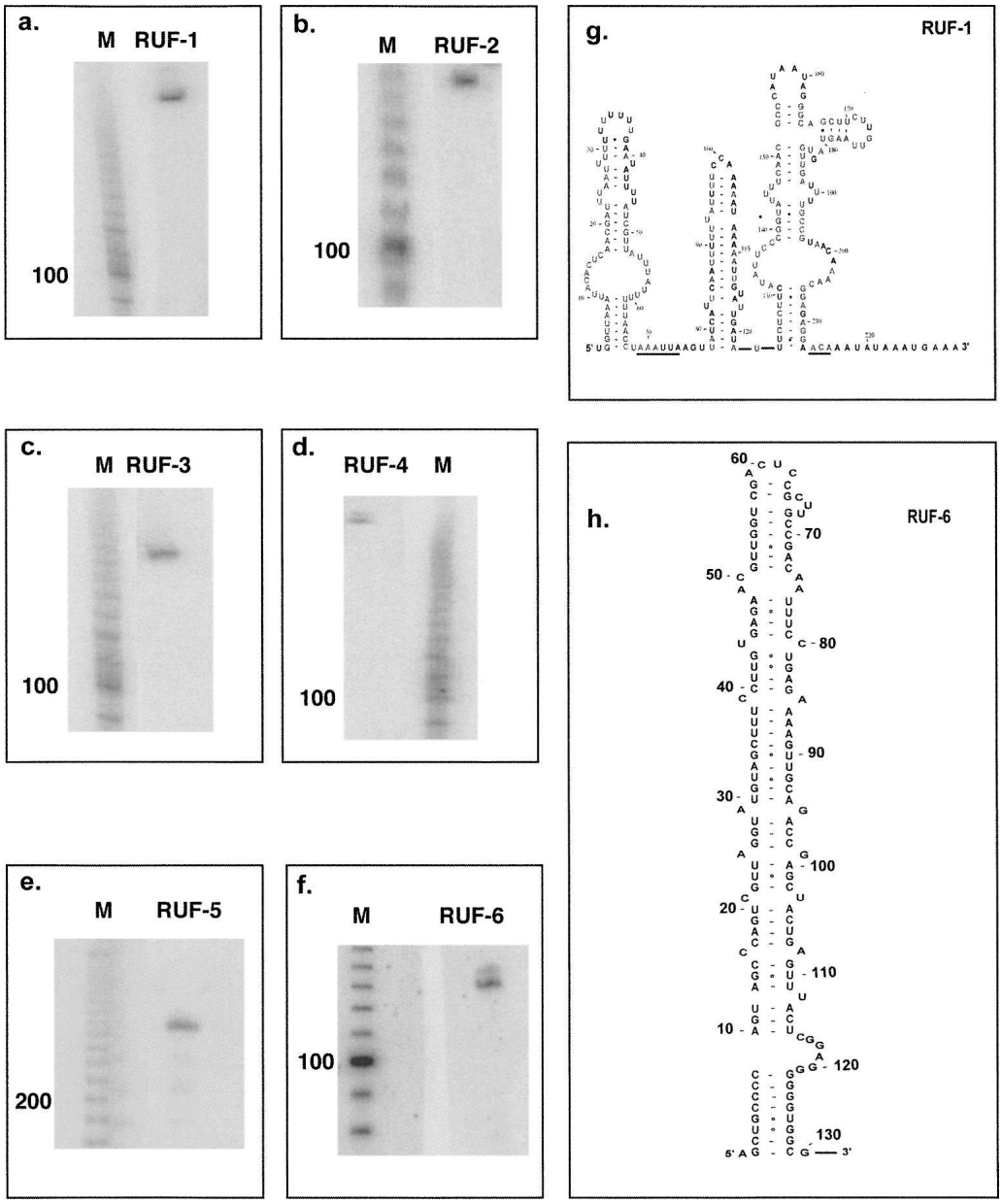

FIGURE 3. Northern analysis and secondary structure models of the RNAs of unknown function (RUFs). Northern hybridization data for (*a*) RUF1, (*b*) RUF2, (*c*) RUF3, (*d*) RUF4, (*e*) RUF5, and (*f*) RUF6. Ten-base-pair DNA ladders were used as molecular weight markers. Secondary structure models for (*g*) RUF1 (presumptive H and ACA boxes underlined, 100% conserved residues from multiple sequence alignments are shown in gray) and (*h*) putative secondary structure of RUF6 RNA.

sequence of U1 helix II is highly conserved in malaria parasites and contains an almost identical match to the human U1A protein binding site (Scherly et al. 1989). *Plasmodium* U1 helix IV is longer and does not contain the short internal loop sequence as found in human U1 snRNA (Supplemental Fig. 2). The conserved metazoan loop sequences at the termini of helix III and IV are different in malaria parasites (Fig. 4c).

U2 snRNA adopts different secondary structures during the course of its function in pre-mRNA splicing (Ares and Weiser 1995; Staley and Guthrie 1998). Although the first 76 nt of U2 snRNA are highly conserved among various *Plasmodium* species (Supplemental Fig. 3), significant difference is seen between vertebrate and *Plasmodium* in particular, in helix IIa. This helix, shown to be essential for U2 function (Ares and Igel 1990), has a terminal loop composed of 6 nt (G54 to A59) in *Plasmodium sp.* that shows conserved complementarity to a downstream single stranded region (Fig. 4f, marked with line and arc), equivalent to stem IIc (Ares and Igel 1990). This conserved region appears to

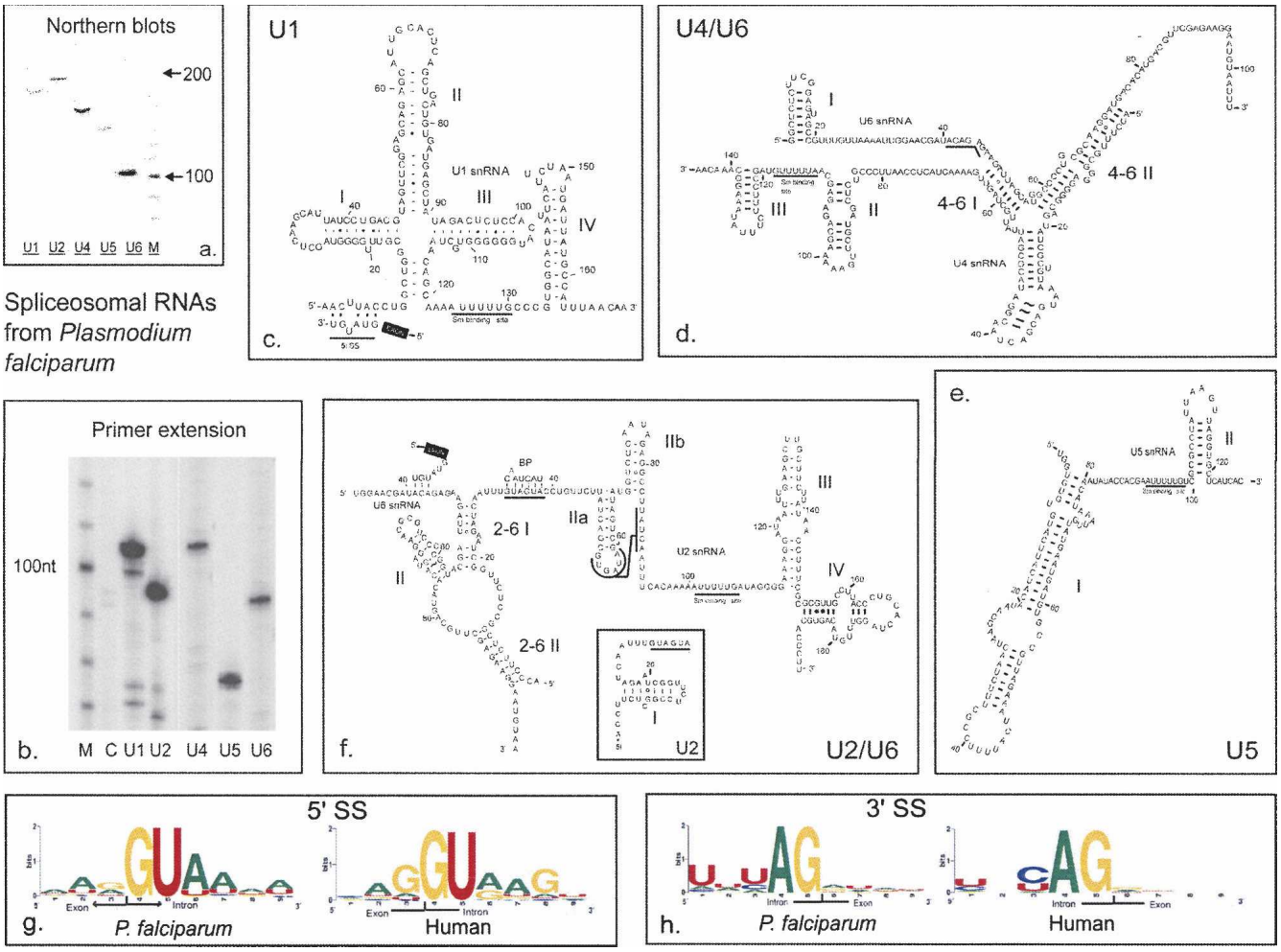

FIGURE 4. Spliceosomal RNAs from *P. falciparum*. (*a*) Northern blot showing U small nuclear RNAs in the malaria parasite. U1snRNA, 170 nt; U2snRNA, 201 nt; U4snRNA, 144 nt; U5snRNA, 127 nt; and U6snRNA, 105 nt. (M) 10-bp ladder. (*b*) Primer extension reaction to map the snRNA transcripts. Lane *C* represents a negative control reaction containing yeast total RNA. (M) 10-bp molecular weight marker. (*c*) Secondary structure models of U1 snRNA. (*d*) U4 and U6 snRNA interaction—the phylogenetically invariant U6 region A41 to A47 (*P. falciparum* numbering) is underlined. Base pairing interaction in the internal asymmetric stem of U4 helix I is shown by squiggly lines. (*e*) U5 snRNA secondary structure and (*f*) U2 and U6 snRNA interaction—native folding pattern of the 5′ region of U2 snRNA preceding branchpoint (BP) sequence is shown in *inset*. The Sm protein binding sites are underlined. (SS) Complementary region to splice site sequences. *Plasmodium* U2 stem IIa terminal loop (nucleotides U55 to A59) has conserved complementarity to downstream single stranded sequence (U85 to A89), shown by line and arc. Alternatively, G54 to A57 in this loop could also pair with downstream residues U93 to C96. (*g*) 5′ splice site and (*h*) 3′ splice site conservation in *P. falciparum* and human (2000 constitutive GT-AG introns were analyzed from human genome database, http://genome.ucsc.edu). Also see Senapathy et al. (1990).

function in spliceosomal structural transitions during splicing (Hilliker et al. 2007; Perriman and Ares 2007).

Plasmodium U4 and U6 can adopt the phylogenetically conserved Y-shaped U4–U6 interaction domain (Fig. 4d; Brow and Guthrie 1988; Bindereif et al. 1990); however, notable differences are observed between human and malaria parasite in this region and in the U4 5′ stem–loop region. The vertebrate U4–U6 stem II has a central disruption of Watson–Crick base pairing in which U4-U13 is opposite to U6-C62 (Ares and Weiser 1995). This region, which could be important in balancing the stability of this helix so that it can be unwound during splicing (Brow and Guthrie 1988, 1989), is different in the parasite, where two U4 bases, G10 and G11, oppose G65 in U6 (Fig. 4d; Supplemental Fig. 4). Like human, *Plasmodium* U4 5′ stem–loop structure (U4 helix I), originating from the Y junction, has a purine-rich internal asymmetric stem–loop, flanked by C-G base pairs and sheared G-A base pairs (Fig. 4d, noncanonical pairing shown by squiggly lines). In malaria parasites, the loop sequence is 5′-ACUAA-3′ instead of the 5′-UUUAU-3′ in vertebrates. This pentaloop sequence is invariant in all *Plasmodium sp.* shown.

The outstanding feature of the U5 snRNA is the 11-nt conserved loop sequence at the end of U5 helix I. This

sequence is necessary for 5′ splice site cleavage and the first *trans*-esterification reaction in pre-mRNA splicing (Newman and Norman 1992; O'Keefe et al. 1996). This sequence of U5 in *Plasmodium* is identical to vertebrate U5 (Fig. 4e; Supplemental Fig. 5). U5 helix II has low sequence homology among malaria parasite species and to the corresponding vertebrate sequences (Supplemental Fig. 5). Each of U1, U2, U4, and U5 snRNAs contains a sequence matching the "Sm binding site" (Fig. 4c–f). This region is important for association with core snRNP proteins (Kramer 1996).

The U6 sequence 5′-ACAGAGA-3′, required for the second step of splicing (Ares and Weiser 1995; Staley and Guthrie 1998), is invariant among *Plasmodium* species as is much of the rest of the RNA, including the U6 stem II, or the "catalytic" stem (Fig. 4f; Supplemental Fig. 6). Base pairing to form U2-U6 helix II between the 5′ end of U2 and the 3′ end of U6, shown to contribute to the efficiency of yeast splicing (Xu and Friesen 2001), is also present in *Plasmodium sp.* although the primary sequences that make up this stem are different.

The spliceosomal snRNAs make several contacts with the pre-mRNA substrate near conserved splice site sequences (Ares and Weiser 1995; Staley and Guthrie 1998). No catalog of true *Plasmodium* splice site sequences has been published yet. To identify the conservation pattern in the natural *Plasmodium* splice sites, we aligned EST sequences to the genome using GMAP (Wu and Watanabe 2005); these alignments are displayed in the public version of our browser (see Materials and Methods). We filtered alignment gaps for true introns by selecting only those gaps between EST and the genomes that end in GT and AG (on the sense strand) and that are more than 64 but less than 1000 nt long. This resulted in a set of 2804 unique GT-AG introns identified from 8397 ESTs. Of the 2804, 1921 match introns in the current set of gene predictions (Gardner et al. 2002; Bahl et al. 2003; Berriman and Rutherford 2003). We used the sequences spanning the last three bases of exon 1 and the first six bases of each intron to search for conserved 5′ splice site features (Fig. 4g), and the last five bases of the intron and the first four bases of exon 2 to search for conserved features of the 3′ splices site (Fig. 4h). Ignoring the GT and AG used to select for the set under consideration, the figure shows representation of each nucleotide at the other splice site positions for expressed sequences of *P. falciparum*.

The most surprising finding is the absence of conservation of G at position 5 of the 5′ splice sites in *P. falciparum*: Instead a majority of sites are GUAAAA or GUAAUA. Mutation of the G5 of the 5′ splice site reduces splicing efficiency in yeast (Fouser and Friesen 1986). This is thought to be due to the dual role of this base in the initial recognition of the 5′ splice site by U1 and after spliceosome rearrangement, by U6 (Seraphin et al. 1988; Lesser and Guthrie 1993). This suggests that recognition of the fifth position of the *P. falciparum* 5′ splice site is either less critical or fundamentally different in the parasite than in yeast (or vertebrates), because the U1 and U6 residues proposed to interact with that position are unchanged in the parasite. No consensus branchpoint sequence could be extracted from these introns. This indicates that parasite introns are more like metazoan introns in that they have degenerate branchpoints. Trypanosomes have an unusual U2 sequence at the branchpoint interaction region (Hartshorne and Agabian 1990; Tschudi et al. 1990), but that of *Plasmodium* is identical to yeast and human.

Less frequent splice sites found in *P. falciparum* are common in yeast and human, such as GUAUGU (75.4% in yeast, 2.1% in *P. falciparum*) (Spingola et al. 1999) and GUGAGU (15% in human, 0.48% in *P. falciparum*) or GUAAGU (13% in human, 4.8% in *P. falciparum*) sequences (see Fig. 4g; this study; Senapathy et al. 1990); they are much less frequent than *P. falciparum* GUAAAA (13%) or GUAAUA (12.5%). As in the vertebrate 5′ splice site consensus, there is a preference for AG at the end of the exon just upstream of the 5′ splice site. These bases are very near the conserved U5 loop in the active spliceosome. Most 3′ splice sites in *P. falciparum* are the canonical YAG (Y: pyrimidine), with U being most prevalent (Fig. 4h). There is also a preference for U at position −5 from the 3′ splice site. Since most of the genome annotation programs rely heavily on canonical splice site and branchpoint sequences, changes mentioned above may considerably improve the gene predictions for *Plasmodium sp.*

U3 snoRNA

U3 snoRNA plays an essential role in processing of the small subunit (SSU) rRNA (Hughes and Ares 1991; Mougin et al. 1996; Borovjagin and Gerbi 1999). We identified the U3 gene by comparative genomics and mapped its 242-nt-long transcript by Northern hybridization (Fig. 5b) and primer extension (Supplemental Fig. 7a, lane 3). Like other U3 RNAs (Parker and Steitz 1987; Hartshorne and Agabian 1994; Mougin et al. 1996), *Plasmodium* U3 has three characteristic stem–loop structures: stem I contains the box A and A′ sequence, stem II has C′/D and B/C protein interaction domains, and stem III (Fig. 5a, see also Supplemental Fig. 7b). The 3′ domain has been shown to bind the U3 snoRNP core proteins in other eukaryotes (Parker and Steitz 1987; Samarsky and Fournier 1998; Watkins et al. 2004). Both the C/D and B/C box regions predicted to bind the 15.5-kD/SNU13 protein homolog are present, and the C/D box is similar to *Trypanosoma* or *Tetrahymena* sp. U3 but different than yeast or human (Marmier-Gourrier et al. 2003). Immediately after the box C sequence, the stem–loop III structure is formed, which has U-rich internal and terminal loops. The *Plasmodium* U3 stem III is longer and lacks any sequence homology with yeast or vertebrate U3 RNAs.

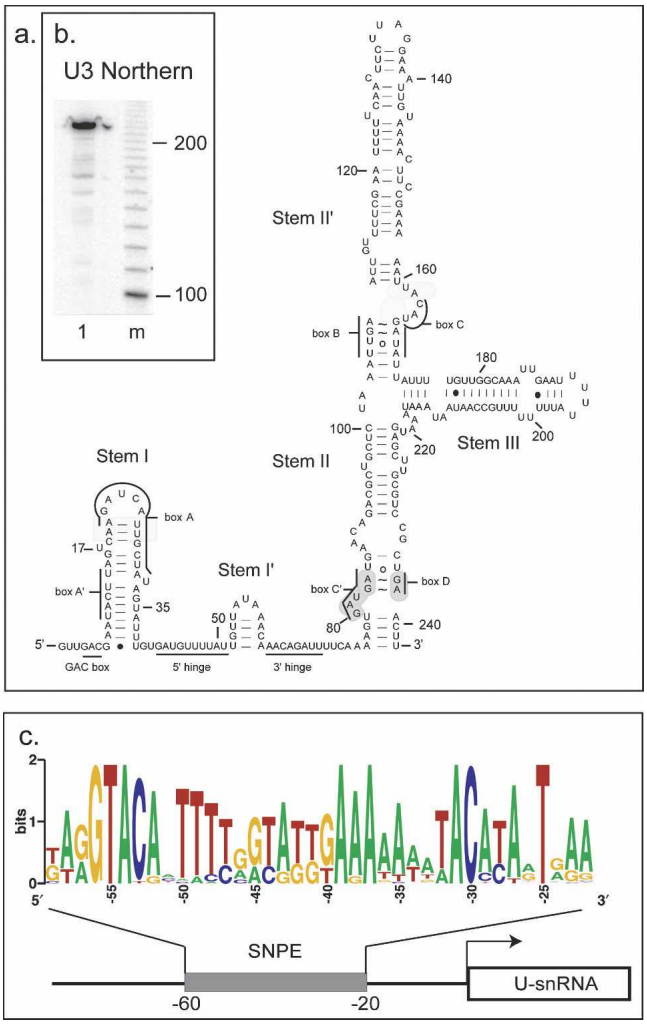

FIGURE 5. U3 snoRNA and SNPE. (*a*) Secondary structure model of *P. falciparum* U3 snoRNA. 5′ and 3′ hinge region has sequence complementarity to the 5′ upstream region of *P. falciparum* 18S rRNA. The C′/D box kink-turn motif is shaded in gray. (*b*) Northern hybridization of U3 snoRNA from *P. falciparum* asexual blood stage. One hundred- and 200-nt bands are shown in the marker lane from a 10-bp ladder marker. (*c*) Sequence logo showing snRNA promoter-like element (SNPE), identified upstream of U1, U2, U3, U4, and U5 RNAs. This logo is generated after aligning multiple sequences from available *Plasmodium sp.* for respective U snRNAs (see Supplemental Fig. 8).

The known function of U3 is its involvement in the initial cleavage within the 5′ external transcribed spacer of 18S rRNA for pre-rRNA processing (Hughes and Ares 1991). Direct contacts between the 5′ domain of U3, including conserved box A and variable hinge region bases, are implicated in this processing event (Borovjagin and Gerbi 1999, 2000; Sharma and Tollervey 1999). Base pairing is proposed to occur between the U3 5′ hinge, 3′ hinge, and complementary regions in the external transcribed spacer (ETS) of 18S rRNA (Borovjagin and Gerbi 2000). The 5′ ETS region of *Plasmodium berghei* 18S rRNA has been documented (van Spaendonk et al. 1999). We have identified complementarity of two hinge region sequences (nucleotides 42–51 and 63–70) in our proposed U3 RNA (Fig. 5a) to the homologous 5′ region of all 18S rRNA sequences of *P. falciparum*. We propose that these regions of complementarities may render base-pairing interactions with the 5′ flanking sequences of 18S rRNA to initiate cleavage and processing in *P. falciparum* SSU rRNA as shown in other eukaryotes.

SnRNA promoter-like element (SNPE) upstream of U-RNAs

The genomic regions upstream of *Plasmodium* U-rich RNAs (U1–U6) coding sequences were analyzed to search for potential snRNA promoter elements (SNPEs) with the IMPROBIZER program (http://www.soe.ucsc.edu/~kent/improbizer/improbizer.html). We identified a sequence element (Fig. 5c) located in the corresponding position of the vertebrate proximal sequence element (PSE), a specialized snRNA promoter element (Jensen et al. 1998; Hernandez 2001). In *Plasmodium*, this conserved sequence element is present upstream of the U1, U2, U3, U4, and U5 genes (Fig. 5c, Supplemental Fig. 8) but not U6 or any protein coding gene we (unpubl.) or others (Gunasekera et al. 2007) could identify. The SNPE motif appears to be composed of two sequences in inverted orientation that create an imperfect palindrome with an axis of symmetry near position −44 (Fig. 5c). This novel motif suggests specialized transcriptional control exists for U-snRNA coding genes in malaria parasites.

RNAse P RNA and MRP RNA

RNases P and MRP are ribonucleoprotein enzymes responsible for tRNA and rRNA processing, respectively. *Plasmodium* RNAse P and MRP were identified in a previous bioinformatics screen (Piccinelli et al. 2005). We have experimentally validated the expression of RNAse P and MRP RNA molecule by Northern blotting (Fig. 6c) and mapped the transcription initiation sites by primer extension (Supplemental Fig. 7a; lane 1-MRP and lane 2- RNAse P RNA). *P. falciparum* RNAseP and MRP RNA align well with the known secondary structures for these RNAs derived from other organism (Fig. 6; Walker and Avis 2005). *Plasmodium* RNAse P RNA shows extensive interspecies variation in length; for example, both *Plasmodium* RNaseP and MRP RNA have long and complex P3 helix–loop–helix structures compared to vertebrates. In other places *Plasmodium* RNAse MRP is simpler; for example, the P12-like hairpin in domain 2 (Fig. 6b) does not contain extended complex stem–loop structures as seen in yeast and metazoan species (Li et al. 2004).

SRP RNA

P. falciparum SRP RNA was identified in a bioinformatics screen (Rosenblad et al. 2004) and a secondary structure

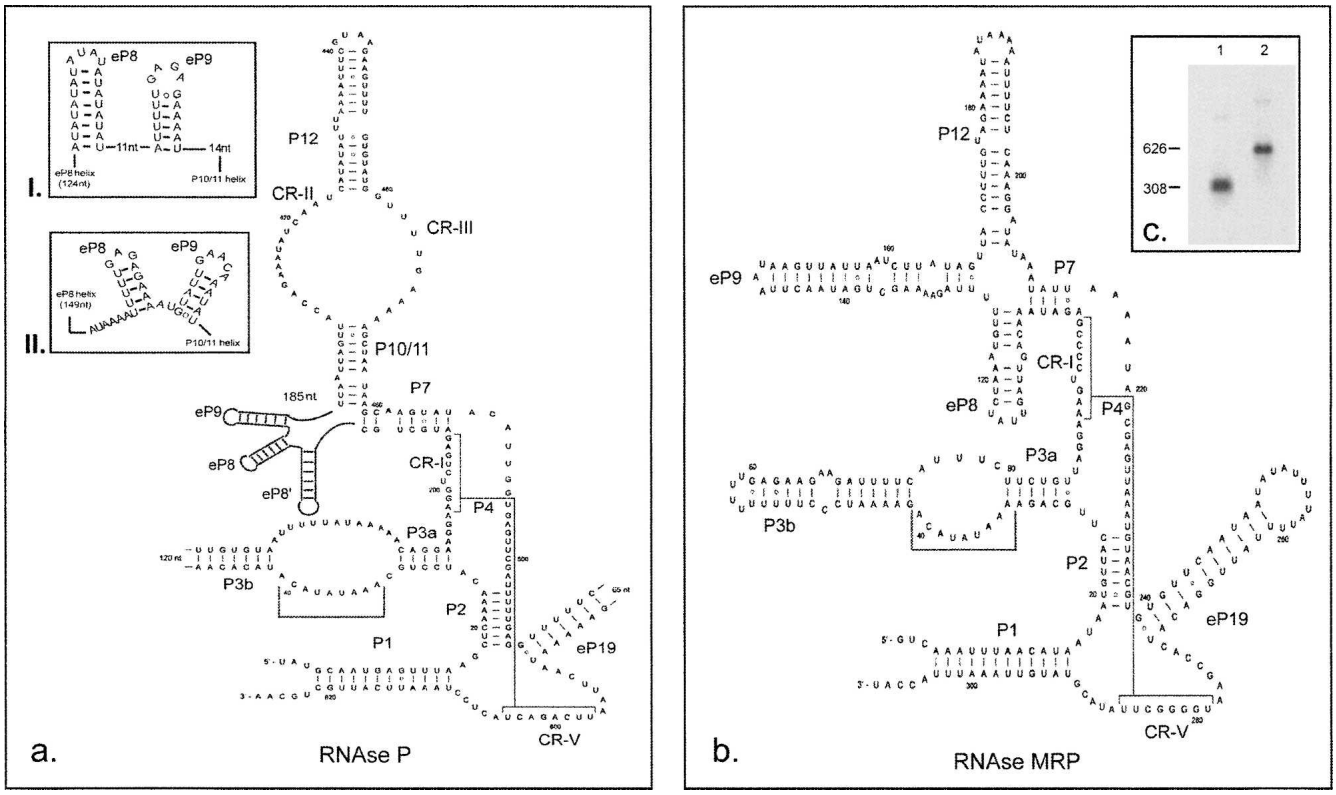

FIGURE 6. Expression data and secondary structure models for *P. falciparum* RNAse P and MRP. (*a*) Secondary structure model of *P. falciparum* RNAse P. In this structure, helices P designates paired regions that are common to archeal, bacterial, and eukaryotic RNAse P RNAs and eP for eukaryotic paired regions only. (CR) Conserved regions. Two possible arrangements of helix eP8 and eP9 are shown in *insets I* and *II*. (*b*) Secondary structure model of *P. falciparum* RNAse MRP RNA. The conserved regions of the loop internal to the P3 helix, common to RNAse P and MRP RNA, are marked with brackets. (*c*) Northern hybridization with RNAse MRP (lane *1*) and RNAse P (lane *2*) specific probes on total RNA isolated from *P. falciparum*.

was proposed based on sequence comparison from three *Plasmodium* species (Zwieb et al. 2005). To demostrate the expression of SRP and learn the limits of the RNA molecule, we mapped the SRP RNA by Northern hybridization (Supplemental Fig. 9a) and primer extension (Supplemental Fig. 7a, lane 4). The 303 nt RNA sequence from *P. falciparum* allowed the alignment of corresponding sequences from seven *Plasmodium* species (Supplemental Fig. 9b). The proposed SRP RNA secondary structure for *P. falciparum* (Zwieb et al. 2005) agrees with our working model based on an eight-species comparison (Supplemental Fig. 9b), except for the 5′ domains (H3 and H4). A predicted ORF Pf14_0183 is at the same location where we find the SRP RNA gene (Gardner et al. 2002; Bahl et al. 2003), and thus may not be an accurate prediction.

Ribosomal RNA and transfer RNA

The organization of the ribosomal RNA (rRNA) multigene family in malaria parasite is unusual in that there are only a few rRNA genes (four to eight, depending on the strain) dispersed on different chromosomes (Supplemental Table 2; McCutchan 1986; Wellems et al. 1987). Studies on *P. falciparum* and *P. berghei* have shown that there are two specialized types of nuclear rRNA genes, and these are developmentally regulated. The S-type is expressed predominantly during the sexual stage in the mosquito host, whereas the A-type is expressed in the asexual stages in the vertebrate host (Gunderson et al. 1987; Fang et al. 2004). Additionally, a third type of small subunit rRNA, type-O, has been described only in *Plasmodium vivax* in the oocyst/ookinete stage (Li et al. 1997). From the available genome sequencing data it is now clear that none of the *P. falciparum* rRNA sequences closely resemble the O-type of *Plasmodium vivax*.

So far, 44 nuclear-encoded cytoplasmic tRNAs have been identified in *P. falciparum* genome, at least one for each standard amino acid and selenocysteine (Gardner et al. 2002; Mourier et al. 2005). All the tRNA genes identified so far are between 71 and 84 nt in length, and only one tRNA-Tyr has an intron (Supplemental Table 2). This 11-nt-ong intron appears to be a standard tRNA type intron (Haugen et al. 2005). Additional tRNA genes may yet be discovered.

MATERIALS AND METHODS

Structural RNA sequence searches

Apart from *P. falciparum* complete genome sequences, sequences from contigs of seven other *Plasmodium* species, ESTs, and ORF predictions were analyzed for GC content and RNA folding potential, and all of it was organized in the UCSC malaria genome browser. With this as the base data, searches for different classes of RNAs were approached in different ways. For a class such as the splicosomal RNAs, which are well conserved, sequence search tools such as BLAST, RNABOB, and PATSEARCH (Grillo et al. 2003) were used to identify candidate sequences that matched a query sequence (e.g., Human). Resulting candidates based on primary and/or secondary sequence search were further narrowed down based on conservation across species and EST evidence. Candidate regions identified in the original search including the several hundred nucleotides of sequence to either side of the candidate were extracted from the genome. We then used each of these larger regions as BLAST query sequences against each of the other *Plasmodium* genomes. The *P. falciparum* loci that gave the highest scoring blast hits on the most other *Plasmodium* genomes were saved, and a multispecies alignment was inspected. Candidates that appeared to be conserved and had nucleotide changes consistent with RNA evolution as opposed to third position variation common to protein coding regions were tested by Northern blotting.

P. falciparum culture, synchronization, and isolation of total RNA

P. falciparum strain 3D7 was cultured in human O^+ erythrocytes at 37°C under 5% O_2, 5% CO_2, and 90% N_2 in RPMI 1640 medium (GIBCO BRL) supplemented with 27 mM sodium bicarbonate, 11 mM glucose, 10 μg/mL gentamicin, and 10% heat-inactivated human serum as described previously (Trager and Jensen 1976). RNA was isolated from either asynchronous or sorbitol synchronized cultures, as required. For parasite synchronization, two consecutive sorbitol treatments were used, and then parasites were harvested at the late trophozoite stage (18–24 h post-invasion) (Lambros and Vanderberg 1979). For total RNA isolation *P. falciparum* cultures were grown to 10% parasitaemia. Infected erythrocytes were washed once with incomplete growth medium (without serum), lysed by saponin (0.15%) and spun at 6000*g* for 5 min to pellet parasites. Pellets were either rapidly frozen and stored at −80°C or used instantly for total RNA isolation. Total RNA was prepared directly from the parasite pellets by the Trizol (Invitrogen) extraction method according to the manufacturer's instructions. Parasite RNA samples were DNAse treated for 30 min and recovered by phenol:chloroform: isoamyl alchohol extraction. RNA quality was checked by formaldehyde-agarose gel electrophoresis (Sambrook and Russell 2001) and the concentration was determined on a Nanodrop spectrophotometer (NanoDrop Technologies).

Northern hybridization and primer extension

Small RNAs were resolved on 6% polyacrylamide gel containing 7 M urea, 90 mM Tris-Borate, and 2.5 mM EDTA (TBE) (pH 8.3). Two micrograms of total RNA were loaded on each lane of the gel in 50% formamide after heating at 95°C for 3 min and instantly chilling on ice. RNA from the gel was subsequently transferred onto N^+ nylon membranes (Amersham Biosciences) electrophoretically at 300 mA constant current for 3–4 h at 4°C. Subsequently, RNA was cross-linked onto the membrane by ultraviolet irradiation. RNAs that are expected to be >300 nt long were separated on 1.5% formaldehyde-agarose gel and were capillary transferred to nylon membranes. Before hybridization, blots were stained with methylene blue (0.02% w/v methylene blue in 0.3 M sodium acetate at pH 5.5) to check the integrity of the ribosomal RNAs (LSU and SSU) and molecular weight markers. Hybridization was done in 0.5 M sodium phosphate buffer (pH 7.6) with gamma-^{32}P-ATP labeled specific oligodeoxyribonucleotides at ~5×10^6 cpm/mL. The sequences of the oligonucleotides used in this study are given in Supplemental Table 1.

Following hybridization, membranes were washed twice in 2× SSC, 0.1% SDS at hybridization temperature and once at 0.1× SSC, 0.1% SDS at room temperature for 15 min each and exposed to a PhosphorImager. Ten base-pair and 100 bp DNA markers were used to quantify the RNA sizes on denaturing PAGE. To identify RNA sizes >300 nt, either 1-kb DNA marker or commercially obtained RNA markers were used for formaldehyde agarose gels. To account for the ~10% faster migration of DNA markers on denaturing gels compared to RNA bands, Northern sizes were calculated using the following rule: RNA length = DNA length × 1.1. Primer extension reactions were performed with 2–5 μg total RNA and a molar excess of primer as described previously (Ares and Igel 1990). Annealing of oligos was done for 30 min at 37°C, and the reverse transcription reaction was done for 45 min at 42°C. Primer extension products were treated with RNAse and Proteinase K, ethanol precipitated, dissolved in 50% formamide, and analyzed on 6% urea–polyacrylamide gel, as described above.

In vitro transcription and RNAse protection assay (RPA)

Complementary RNA (cRNA) probes were made for Northern hybridization by in vitro transcription (IVT) reactions. Primary PCR amplification was done with oligonucleotides containing T7 promoter sequences at the 5′ end. Amplified product was added to the transcription reaction. The IVT reaction contained 500 ng to 2 μg of PCR product, 40 mM Tris-HCl (pH 8.0), 6 mM $MgCl_2$, 2 mM spermidine HCl, 10 mM NaCl, 10 mM DTT, 100 mg/mL BSA, 250 μM ATP, UTP, or GTP, 1 unit of RNasin, and 500–1000 units/mL of T7 RNA polymerase in a final volume of 20 μL. All reactions contained 5 μL of alpha ^{32}P-CTP (3000 Ci/mmol, 10 mCi/mL) at a final concentration of 1.5 mM. IVT reactions were incubated at 37°C for 3–4 h and then diluted with formamide loading buffer to run on a 6% polyacrylamide gel. In vitro synthesized transcripts were recovered by the crush and soak method as described previously (Sambrook and Russell 2001). Nuclease protection assay was done using RPAIII RNAse protection assay kit (Ambion Inc.), according to the manufacturer's instructions. RNAse treatment was done with RNase Digestion Buffer containing both RNase A and RNase T1 and incubation at 30°C for 30 min. After RNAse inactivation, the protected probe fragment was recovered by adding RNAse inactivation buffer followed by centrifugation. Samples were analyzed on 6% polyacrylamide/7 M urea gel.

Mapping 2′-*O*-methyl groups

Spliceosomal RNA genes from *P. falciparum* genomic DNA were PCR amplified with 3 mM $MgCl_2$ and directly cloned into Topo-TA cloning vector PCR 2.1 (Invitrogen), which contains the T7 RNA promoter for production of sense RNA transcripts. For in vitro transcription of the snRNA genes, plasmid was linearized with HindIII and transcribed with T7 RNA polymerase. RNA was ethanol precipitated and used either for sequencing reactions or in a primer extension reaction for mapping 2′-*O*-methyl groups of U4 snRNAs. For sequencing reaction, 1 μg of in vitro transcribed RNA was used with 1 mM each of ddNTPS. The positions of 2′-*O*-methyls were identified by primer extension in the presence of limiting amount of dNTPs (0.01 μM) (Maden 2001).

SUPPLEMENTAL MATERIAL

Supplemental material can be found at http://ribonode.ucsc.edu/PlasmoRNASuppl/index.html.

ACKNOWLEDGMENTS

We thank Philip Rosenthal, UCSF, and members of his laboratory for initial help with parasite strain 3D7 as well as The Malaria Research and Reference Reagent Resource Center (MR4). We gratefully acknowledge use of publicly available resources from PlasmoDB including Plasmodium sp. sequence data generated at the Wellcome Trust Sanger Institute, TIGR/NRMC, and the Stanford University Blood Center. We thank Joseph DeRisi, UCSF, for his encouragement throughout the work and Megan Hall and Melissa Cline for comments on the manuscript. We appreciate the help of Hiram Clawson from David Haussler's group in reviewing the workings of the browser. We are grateful to members of our laboratory and Hughes Undergraduate Research Laboratory (HURL) for help with experiments and discussions. This work was supported by an HHMI Professor award to M.A. by the Howard Hughes Medical Institute and by NIH grant GM040478 to M.A.

Received July 24, 2007; accepted August 15, 2007.

REFERENCES

Aldous, W.K., Martin, R.K., and Kyle, D.E. 1998. Stage specific detection and inhibition studies of *Plasmodium falciparum* telomerase. *Mol. Biochem. Parasitol.* **95:** 281–285.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Ares Jr., M. and Igel, A.H. 1990. Lethal and temperature-sensitive mutations and their suppressors identify an essential structural element in U2 small nuclear RNA. *Genes & Dev.* **4:** 2132–2145.

Ares Jr., M. and Weiser, B. 1995. Rearrangement of snRNA structure during assembly and function of the spliceosome. *Prog. Nucleic Acid Res. Mol. Biol.* **50:** 131–159.

Autexier, C. and Greider, C.W. 1995. Boundary elements of the *Tetrahymena* telomerase RNA template and alignment domains. *Genes & Dev.* **9:** 2227–2239.

Bahl, A., Brunk, B., Crabtree, J., Fraunholz, M.J., Gajria, B., Grant, G.R., Ginsburg, H., Gupta, D., Kissinger, J.C., Labo, P., et al. 2003. PlasmoDB: The *Plasmodium* genome resource. A database integrating experimental and computational data. *Nucleic Acids Res.* **31:** 212–215.

Berriman, M. and Rutherford, K. 2003. Viewing and annotating sequence data with Artemis. *Brief. Bioinform.* **4:** 124–132.

Bindereif, A., Wolff, T., and Green, M.R. 1990. Discrete domains of human U6 snRNA required for the assembly of U4/U6 snRNP and splicing complexes. *EMBO J.* **9:** 251–255.

Blackburn, E. 1999. The telomere and telomerase: How do they interact? *Mt. Sinai J. Med.* **66:** 292–300.

Blasco, M.A. 2005. Mice with bad ends: Mouse models for the study of telomeres and telomerase in cancer and aging. *EMBO J.* **24:** 1095–1103.

Borovjagin, A.V. and Gerbi, S.A. 1999. U3 small nucleolar RNA is essential for cleavage at sites 1, 2, and 3 in pre-rRNA and determines which rRNA processing pathway is taken in *Xenopus* oocytes. *J. Mol. Biol.* **286:** 1347–1363.

Borovjagin, A.V. and Gerbi, S.A. 2000. The spacing between functional *cis*-elements of U3 snoRNA is critical for rRNA processing. *J. Mol. Biol.* **300:** 57–74.

Bottius, E., Bakhsis, N., and Scherf, A. 1998. *Plasmodium falciparum* telomerase: De novo telomere addition to telomeric and non-telomeric sequences and role in chromosome healing. *Mol. Cell. Biol.* **18:** 919–925.

Bozdech, Z., Llinas, M., Pulliam, B.L., Wong, E.D., Zhu, J., and DeRisi, J.L. 2003. The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol.* **1:** E5. doi: 10.1371/journal.pbio.0000005.

Brow, D.A. and Guthrie, C. 1988. Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature* **334:** 213–218.

Brow, D.A. and Guthrie, C. 1989. Splicing a spliceosomal RNA. *Nature* **337:** 14–15.

Brown, J.W., Clark, G.P., Leader, D.J., Simpson, C.G., and Lowe, T. 2001. Multiple snoRNA gene clusters from *Arabidopsis*. *RNA* **7:** 1817–1832.

Brown, J.W., Echeverria, M., and Qu, L.H. 2003. Plant snoRNAs: Functional evolution and new modes of gene expression. *Trends Plant Sci.* **8:** 42–49.

Cano, M.I., Dungan, J.M., Agabian, N., and Blackburn, E.H. 1999. Telomerase in kinetoplastid parasitic protozoa. *Proc. Natl. Acad. Sci.* **96:** 3616–3621.

Cavaille, J., Buiting, K., Kiefmann, M., Lalande, M., Brannan, C.I., Horsthemke, B., Bachellerie, J.P., Brosius, J., and Huttenhofer, A. 2000. Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Natl. Acad. Sci.* **97:** 14311–14316.

Chakrabarti, D., Reddy, G.R., Dame, J.B., Almira, E.C., Laipis, P.J., Ferl, R.J., Yang, T.P., Rowe, T.C., and Schuster, S.M. 1994. Analysis of expressed sequence tags from *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **66:** 97–104.

Chen, J.L., Blasco, M.A., and Greider, C.W. 2000. Secondary structure of vertebrate telomerase RNA. *Cell* **100:** 503–514.

Chen, J.L., Opperman, K.K., and Greider, C.W. 2002. A critical stem-loop structure in the CR4-CR5 domain of mammalian telomerase RNA. *Nucleic Acids Res.* **30:** 592–597. doi: 10.1093/nar/30.2.581.

Cheng, Q., Cloonan, N., Fischer, K., Thompson, J., Waine, G., Lanzer, M., and Saul, A. 1998. stevor and rif are *Plasmodium falciparum* multicopy gene families which potentially encode variant antigens. *Mol. Biochem. Parasitol.* **97:** 161–176.

Clouet d'Orval, B., Bortolin, M.L., Gaspin, C., and Bachellerie, J.P. 2001. Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNATrp intron guides the formation of two ribose-methylated nucleosides in the mature tRNATrp. *Nucleic Acids Res.* **29:** 4518–4529.

Coppel, R.L., Roos, D.S., and Bozdech, Z. 2004. The genomics of malaria infection. *Trends Parasitol.* **20:** 553–557.

Dandjinou, A.T., Levesque, N., Larose, S., Lucier, J.F., Abou Elela, S., and Wellinger, R.J. 2004. A phylogenetically based secondary structure for the yeast telomerase RNA. *Curr. Biol.* **14:** 1148–1158.

Darzacq, X., Jady, B.E., Verheggen, C., Kiss, A.M., Bertrand, E., and Kiss, T. 2002. Cajal body-specific small nuclear RNAs: A novel class of 2′-*O*-methylation and pseudouridylation guide RNAs. *EMBO J.* **21:** 2746–2756.

de Lange, T. 2002. Protection of mammalian telomeres. *Oncogene* **21:** 532–540.

DeJong, E.S., Luy, B., and Marino, J.P. 2002. RNA and RNA–protein complexes as targets for therapeutic intervention. *Curr. Top. Med. Chem.* **2:** 289–302.

del Portillo, H.A., Fernandez-Becerra, C., Bowman, S., Oliver, K., Preuss, M., Sanchez, C.P., Schneider, N.K., Villalobos, J.M., Rajandream, M.A., Harris, D., et al. 2001. A superfamily of variant genes encoded in the subtelomeric region of *Plasmodium vivax*. *Nature* **410:** 839–842.

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.

Eddy, S.R. 2001. Noncoding RNA genes and the modern RNA world. *Nat. Rev. Genet.* **2:** 919–929.

Engelhardt, M. and Martens, U.M. 1998. The implication of telomerase activity and telomere stability for replicative aging and cellular immortality (Review). *Oncol. Rep.* **5:** 1043–1052.

Fang, J., Sullivan, M., and McCutchan, T.F. 2004. The effects of glucose concentration on the reciprocal regulation of rRNA promoters in *Plasmodium falciparum*. *J. Biol. Chem.* **279:** 720–725.

Figueiredo, L.M., Rocha, E.P., Mancio-Silva, L., Prevost, C., Hernandez-Verdun, D., and Scherf, A. 2005. The unusually large *Plasmodium* telomerase reverse-transcriptase localizes in a discrete compartment associated with the nucleolus. *Nucleic Acids Res.* **33:** 1111–1122. doi: 10.1093/nar/gki260.

Florens, L., Washburn, M.P., Raine, J.D., Anthony, R.M., Grainger, M., Haynes, J.D., Moch, J.K., Muster, N., Sacci, J.B., Tabb, D.L., et al. 2002. A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **419:** 520–526.

Fourmy, D., Recht, M.I., Blanchard, S.C., and Puglisi, J.D. 1996. Structure of the A site of *Escherichia coli* 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science* **274:** 1367–1371.

Fouser, L.A. and Friesen, J.D. 1986. Mutations in a yeast intron demonstrate the importance of specific conserved nucleotides for the two stages of nuclear mRNA splicing. *Cell* **45:** 81–93.

Freitas-Junior, L.H., Bottius, E., Pirrit, L.A., Deitsch, K.W., Scheidig, C., Guinet, F., Nehrbass, U., Wellems, T.E., and Scherf, A. 2000. Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature* **407:** 1018–1022.

Ganot, P., Caizergues-Ferrer, M., and Kiss, T. 1997. The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes & Dev.* **11:** 941–956.

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.

Gilley, D. and Blackburn, E.H. 1999. The telomerase RNA pseudoknot is critical for the stable assembly of a catalytically active ribonucleoprotein. *Proc. Natl. Acad. Sci.* **96:** 6621–6625.

Goodchild, J. 2004. Oligonucleotide therapeutics: 25 years agrowing. *Curr. Opin. Mol. Ther.* **6:** 120–128.

Greider, C.W. and Blackburn, E.H. 1987. The telomere terminal transferase of *Tetrahymena* is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell* **51:** 887–898.

Grillo, G., Licciulli, F., Liuni, S., Sbisa, E., and Pesole, G. 2003. PatSearch: A program for the detection of patterns and structural motifs in nucleotide sequences. *Nucleic Acids Res.* **31:** 3608–3612.

Gu, A.D., Zhou, H., Yu, C.H., and Qu, L.H. 2005. A novel experimental approach for systematic identification of box H/ACA snoRNAs from eukaryotes. *Nucleic Acids Res.* **33:** e194. doi: 10.1093/nar/gni185.

Gunasekera, A.M., Myrick, A., Militello, K.T., Sims, J.S., Dong, C.K., Gierahn, T., Le Roch, K., Winzeler, E., and Wirth, D.F. 2007. Regulatory motifs uncovered among gene expression clusters in *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **153:** 19–30.

Gunderson, J.H., Sogin, M.L., Wollett, G., Hollingdale, M., de la Cruz, V.F., Waters, A.P., and McCutchan, T.F. 1987. Structurally distinct, stage-specific ribosomes occur in *Plasmodium*. *Science* **238:** 933–937.

Hall, N., Karras, M., Raine, J.D., Carlton, J.M., Kooij, T.W., Berriman, M., Florens, L., Janssen, C.S., Pain, A., Christophides, G.K., et al. 2005. A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses. *Science* **307:** 82–86.

Hartshorne, T. and Agabian, N. 1990. A new U2 RNA secondary structure provided by phylogenetic analysis of trypanosomatid U2 RNAs. *Genes & Dev.* **4:** 2121–2131.

Hartshorne, T. and Agabian, N. 1994. A common core structure for U3 small nucleolar RNAs. *Nucleic Acids Res.* **22:** 3354–3364. doi: 10.1093/nar/22.16.3354.

Haugen, P., Simon, D.M., and Bhattacharya, D. 2005. The natural history of group I introns. *Trends Genet.* **21:** 111–119.

Hernandez, N. 2001. Small nuclear RNA genes: A model system to study fundamental mechanisms of transcription. *J. Biol. Chem.* **276:** 26733–26736.

Hilliker, A.K., Mefford, M.A., and Staley, J.P. 2007. U2 toggles iteratively between the stem IIa and stem IIc conformations to promote pre-mRNA splicing. *Genes & Dev.* **21:** 821–834.

Hughes, J.M. and Ares Jr., M. 1991. Depletion of U3 small nucleolar RNA inhibits cleavage in the 5′ external transcribed spacer of yeast pre-ribosomal RNA and impairs formation of 18S ribosomal RNA. *EMBO J.* **10:** 4231–4239.

Jensen, R.C., Wang, Y., Hardin, S.B., and Stumph, W.E. 1998. The proximal sequence element (PSE) plays a major role in establishing the RNA polymerase specificity of *Drosophila* U-snRNA genes. *Nucleic Acids Res.* **26:** 616–622. doi: 10.1093/nar/26.2.616.

Karolchik, D., Baertsch, R., Diekhans, M., Furey, T.S., Hinrichs, A., Lu, Y.T., Roskin, K.M., Schwartz, M., Sugnet, C.W., Thomas, D.J., et al. 2003. The UCSC Genome Browser Database. *Nucleic Acids Res.* **31:** 51–54. doi: 10.1093/nar/gkg129.

Kishore, S. and Stamm, S. 2006. The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science* **311:** 230–232.

Kiss, T. 2001. Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.* **20:** 3617–3622.

Kooij, T.W., Janse, C.J., and Waters, A.P. 2006. Plasmodium post-genomics: Better the bug you know? *Nat. Rev. Microbiol.* **4:** 344–357.

Kramer, A. 1996. The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu. Rev. Biochem.* **65:** 367–409.

Kretzner, L., Rymond, B.C., and Rosbash, M. 1987. *S. cerevisiae* U1 RNA is large and has limited primary sequence homology to metazoan U1 snRNA. *Cell* **50:** 593–602.

Lambros, C. and Vanderberg, J.P. 1979. Synchronization of *Plasmodium falciparum* erythrocytic stages in culture. *J. Parasitol.* **65:** 418–420.

Lasonder, E., Ishihama, Y., Andersen, J.S., Vermunt, A.M., Pain, A., Sauerwein, R.W., Eling, W.M., Hall, N., Waters, A.P., Stunnenberg, H.G., et al. 2002. Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* **419:** 537–542.

Le Roch, K.G., Zhou, Y., Blair, P.L., Grainger, M., Moch, J.K., Haynes, J.D., De La Vega, P., Holder, A.A., Batalov, S., Carucci, D.J., et al. 2003. Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* **301:** 1503–1508.

Legassie, J.D. and Jarstfer, M.B. 2006. The unmasking of telomerase. *Structure* **14:** 1603–1609.

Lesser, C.F. and Guthrie, C. 1993. Mutations in U6 snRNA that alter splice site specificity: Implications for the active site. *Science* **262:** 1982–1988.

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. doi: 10.1093/nar/gkj002.

Levitt, A. 1993. RNA processing in malarial parasites. *Parasitol. Today* **9:** 465–468.

Li, J., Gutell, R.R., Damberger, S.H., Wirtz, R.A., Kissinger, J.C., Rogers, M.J., Sattabongkot, J., and McCutchan, T.F. 1997. Regulation and trafficking of three distinct 18 S ribosomal RNAs during development of the malaria parasite. *J. Mol. Biol.* **269:** 203–213.

Li, L., Stoeckert Jr., C.J., and Roos, D.S. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13:** 2178–2189.

Li, X., Zaman, S., Langdon, Y., Zengel, J.M., and Lindahl, L. 2004. Identification of a functional core in the RNA component of RNase MRP of budding yeasts. *Nucleic Acids Res.* **32:** 3703–3711.

Liang, X.H., Uliel, S., Hury, A., Barth, S., Doniger, T., Unger, R., and Michaeli, S. 2005. A genome-wide analysis of C/D and H/ACA-like small nucleolar RNAs in *Trypanosoma brucei* reveals a trypanosome-specific pattern of rRNA modification. *RNA* **11:** 619–645.

Lingner, J., Hendrick, L.L., and Cech, T.R. 1994. Telomerase RNAs of different ciliates have a common secondary structure and a permuted template. *Genes & Dev.* **8:** 1984–1998.

Lynch, S.R., Gonzalez, R.L., and Puglisi, J.D. 2003. Comparison of X-ray crystal structure of the 30S subunit–antibiotic complex with NMR structure of decoding site oligonucleotide–paromomycin complex. *Structure* **11:** 43–53.

Maden, B.E. 2001. Mapping 2′-*O*-methyl groups in ribosomal RNA. *Methods* **25:** 374–382.

Marmier-Gourrier, N., Clery, A., Senty-Segault, V., Charpentier, B., Schlotter, F., Leclerc, F., Fournier, R., and Branlant, C. 2003. A structural, phylogenetic, and functional study of 15.5-kD/Snu13 protein binding on U3 small nucleolar RNA. *RNA* **9:** 821–838.

McCormick-Graham, M. and Romero, D.P. 1995. Ciliate telomerase RNA structural features. *Nucleic Acids Res.* **23:** 1091–1097. doi: 10.1093/nar/23.7.1091.

McCutchan, T.F. 1986. The ribosomal genes of *Plasmodium*. *Int. Rev. Cytol.* **99:** 295–309.

McCutcheon, J.P. and Eddy, S.R. 2003. Computational identification of non-coding RNAs in *Saccharomyces cerevisiae* by comparative genomics. *Nucleic Acids Res.* **31:** 4119–4128. doi: 10.1093/nar/gkg438.

Moriarty, T.J., Marie-Egyptienne, D.T., and Autexier, C. 2004. Functional organization of repeat addition processivity and DNA synthesis determinants in the human telomerase multimer. *Mol. Cell. Biol.* **24:** 3720–3733.

Mougin, A., Gregoire, A., Banroques, J., Segault, V., Fournier, R., Brule, F., Chevrier-Miller, M., and Branlant, C. 1996. Secondary structure of the yeast *Saccharomyces cerevisiae* pre-U3A snoRNA and its implication for splicing efficiency. *RNA* **2:** 1079–1093.

Mourier, T., Pain, A., Barrell, B., and Griffiths-Jones, S. 2005. A selenocysteine tRNA and SECIS element in *Plasmodium falciparum*. *RNA* **11:** 119–122.

Mu, J., Awadalla, P., Duan, J., McGee, K.M., Joy, D.A., McVean, G.A., and Su, X.Z. 2005. Recombination hotspots and population structure in *Plasmodium falciparum*. *PLoS Biol.* **3:** e335. doi: 10.1371/journal.pbio.0030335.

Newman, A.J. and Norman, C. 1992. U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. *Cell* **68:** 743–754.

Nishant, K.T., Ravishankar, H., and Rao, M.R. 2004. Characterization of a mouse recombination hot spot locus encoding a novel non-protein-coding RNA. *Mol. Cell. Biol.* **24:** 5620–5634.

O'Keefe, R.T., Norman, C., and Newman, A.J. 1996. The invariant U5 snRNA loop 1 sequence is dispensable for the first catalytic step of pre-mRNA splicing in yeast. *Cell* **86:** 679–689.

Parker, K.A. and Steitz, J.A. 1987. Structural analysis of the human U3 ribonucleoprotein particle reveal a conserved sequence available for base pairing with pre-rRNA. *Mol. Cell. Biol.* **7:** 2899–2913.

Perriman, R.J. and Ares Jr., M. 2007. Rearrangement of competing U2 RNA helices within the spliceosome promotes multiple steps in splicing. *Genes & Dev.* **21:** 811–820.

Piccinelli, P., Rosenblad, M.A., and Samuelsson, T. 2005. Identification and analysis of ribonuclease P and MRP RNA in a broad range of eukaryotes. *Nucleic Acids Res.* **33:** 4485–4495. doi: 10.1093/nar/gki756.

Raj, D.K., Das, B.R., Dash, A.P., and Supakar, P.C. 2003. Identification of telomerase activity in gametocytes of *Plasmodium falciparum*. *Biochem. Biophys. Res. Commun.* **309:** 685–688.

Recht, M.I., Douthwaite, S., Dahlquist, K.D., and Puglisi, J.D. 1999. Effect of mutations in the A site of 16 S rRNA on aminoglycoside antibiotic-ribosome interaction. *J. Mol. Biol.* **286:** 33–43.

Reddy, R., Henning, D., and Busch, H. 1981a. The primary nucleotide sequence of U4 RNA. *J. Biol. Chem.* **256:** 3532–3538.

Reddy, R., Henning, D., and Busch, H. 1981b. Pseudouridine residues in the 5′-terminus of uridine-rich nuclear RNA I (U1 RNA). *Biochem. Biophys. Res. Commun.* **98:** 1076–1083.

Richard, P., Darzacq, X., Bertrand, E., Jady, B.E., Verheggen, C., and Kiss, T. 2003. A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. *EMBO J.* **22:** 4283–4293.

Richards, R.J., Theimer, C.A., Finger, L.D., and Feigon, J. 2006. Structure of the *Tetrahymena thermophila* telomerase RNA helix II template boundary element. *Nucleic Acids Res.* **34:** 816–825. doi: 10.1093/nar/gkj481.

Rosenblad, M.A., Zwieb, C., and Samuelsson, T. 2004. Identification and comparative analysis of components from the signal recognition particle in protozoa and fungi. *BMC Genomics* **5:** 5.

Russell, A.G., Schnare, M.N., and Gray, M.W. 2006. A large collection of compact box C/D snoRNAs and their isoforms in *Euglena gracilis*: Structural, functional, and evolutionary insights. *J. Mol. Biol.* **357:** 1548–1565.

Samarsky, D.A. and Fournier, M.J. 1998. Functional mapping of the U3 small nucleolar RNA from the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **18:** 3431–3444.

Sambrook, J. and Russell, D. 2001. *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Scherf, A., Figueiredo, L.M., and Freitas-Junior, L.H. 2001. *Plasmodium* telomeres: A pathogen's perspective. *Curr. Opin. Microbiol.* **4:** 409–414.

Scherly, D., Boelens, W., van Venrooij, W.J., Dathan, N.A., Hamm, J., and Mattaj, I.W. 1989. Identification of the RNA binding segment of human U1 A protein and definition of its binding site on U1 snRNA. *EMBO J.* **8:** 4163–4170.

Senapathy, P., Shapiro, M.B., and Harris, N.L. 1990. Splice junctions, branch point sites, and exons: Sequence statistics, identification, and applications to genome project. *Methods Enzymol.* **183:** 252–278.

Seraphin, B., Kretzner, L., and Rosbash, M. 1988. A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5′ cleavage site. *EMBO J.* **7:** 2533–2538.

Sharma, K. and Tollervey, D. 1999. Base pairing between U3 small nucleolar RNA and the 5′ end of 18S rRNA is required for pre-rRNA processing. *Mol. Cell. Biol.* **19:** 6012–6019.

Siebert, S. and Backofen, R. 2005. MARNA: Multiple alignment and consensus structure prediction of RNAs based on sequence structure comparisons. *Bioinformatics* **21:** 3352–3359.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W.,

Richards, S., et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15:** 1034–1050.

Siliciano, P.G., Jones, M.H., and Guthrie, C. 1987. *Saccharomyces cerevisiae* has a U1-like small nuclear RNA with unexpected properties. *Science* **237:** 1484–1487.

Silva, J.G. and Carvalho, I. 2007. New insights into aminoglycoside antibiotics and derivatives. *Curr. Med. Chem.* **14:** 1101–1119.

Smith, J.D., Chitnis, C.E., Craig, A.G., Roberts, D.J., Hudson-Taylor, D.E., Peterson, D.S., Pinches, R., Newbold, C.I., and Miller, L.H. 1995. Switches in expression of *Plasmodium falciparum* var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* **82:** 101–110.

Spingola, M., Grate, L., Haussler, D., and Ares Jr., M. 1999. Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA* **5:** 221–234.

Staley, J.P. and Guthrie, C. 1998. Mechanical devices of the spliceosome: Motors, clocks, springs, and things. *Cell* **92:** 315–326.

Theimer, C.A. and Feigon, J. 2006. Structure and function of telomerase RNA. *Curr. Opin. Struct. Biol.* **16:** 307–318.

Trager, W. and Jensen, J.B. 1976. Human malaria parasites in continuous culture. *Science* **193:** 673–675.

Tschudi, C., Williams, S.P., and Ullu, E. 1990. Conserved sequences in the U2 snRNA-encoding genes of *Kinetoplastida* do not include the putative branchpoint recognition region. *Gene* **91:** 71–77.

Tycowski, K.T., You, Z.H., Graham, P.J., and Steitz, J.A. 1998. Modification of U6 spliceosomal RNA is guided by other small RNAs. *Mol. Cell* **2:** 629–638.

Tycowski, K.T., Aab, A., and Steitz, J.A. 2004. Guide RNAs with 5′ caps and novel box C/D snoRNA-like domains for modification of snRNAs in metazoa. *Curr. Biol.* **14:** 1985–1995.

Upadhyay, R., Bawankar, P., Malhotra, D., and Patankar, S. 2005. A screen for conserved sequences with biased base composition identifies noncoding RNAs in the A-T rich genome of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **144:** 149–158.

van Spaendonk, R.M., McConkey, G.A., Ramesar, J., Gabrielian, A., McCutchan, T.F., Janse, C.J., and Waters, A.P. 1999. Identification of the transcription initiation site of the asexually expressed rRNA genes of the malaria parasite *Plasmodium berghei*. *Mol. Biochem. Parasitol.* **99:** 193–205.

Vernick, K.D. and McCutchan, T.F. 1988. Sequence and structure of a *Plasmodium falciparum* telomere. *Mol. Biochem. Parasitol.* **28:** 85–94.

Vitali, P., Royo, H., Seitz, H., Bachellerie, J.P., Huttenhofer, A., and Cavaille, J. 2003. Identification of 13 novel human modification guide RNAs. *Nucleic Acids Res.* **31:** 6543–6551.

Walker, S.C. and Avis, J.M. 2005. Secondary structure probing of the human RNase MRP RNA reveals the potential for MRP RNA subsets. *Biochem. Biophys. Res. Commun.* **335:** 314–321.

Washietl, S., Hofacker, I.L., and Stadler, P.F. 2005. Fast and reliable prediction of noncoding RNAs. *Proc. Natl. Acad. Sci.* **102:** 2454–2459.

Watanabe, J., Suzuki, Y., Sasaki, M., and Sugano, S. 2004. Full-malaria 2004: An enlarged database for comparative studies of full-length cDNAs of malaria parasites, *Plasmodium* species. *Nucleic Acids Res.* **32:** D334–D338. doi: 10.1093/nar/gkh115.

Watkins, N.J., Lemm, I., Ingelfinger, D., Schneider, C., Hossbach, M., Urlaub, H., and Luhrmann, R. 2004. Assembly and maturation of the U3 snoRNP in the nucleoplasm in a large dynamic multiprotein complex. *Mol. Cell* **16:** 789–798.

Webster, D. 2001. Malaria kills one child every 30 seconds. *J. Public Health Policy* **22:** 23–33.

Wellems, T.E., Walliker, D., Smith, C.L., do Rosario, V.E., Maloy, W.L., Howard, R.J., Carter, R., and McCutchan, T.F. 1987. A histidine-rich protein gene marks a linkage group favored strongly in a genetic cross of *Plasmodium falciparum*. *Cell* **49:** 633–642.

White, N.J. 2004. Antimalarial drug resistance. *J. Clin. Invest.* **113:** 1084–1092.

Wong, J.M. and Collins, K. 2003. Telomere maintenance and disease. *Lancet* **362:** 983–988.

Wu, T.D. and Watanabe, C.K. 2005. GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21:** 1859–1875.

Xie, J., Zhang, M., Zhou, T., Hua, X., Tang, L., and Wu, W. 2007. Sno/scaRNAbase: A curated database for small nucleolar RNAs and cajal body-specific RNAs. *Nucleic Acids Res.* **35:** D183–187. doi: 10.1093/nar/gkl873.

Xu, D. and Friesen, J.D. 2001. Splicing factor slt11p and its involvement in formation of U2/U6 helix II in activation of the yeast spliceosome. *Mol. Cell. Biol.* **21:** 1011–1023.

Zwieb, C., van Nues, R.W., Rosenblad, M.A., Brown, J.D., and Samuelsson, T. 2005. A nomenclature for all signal recognition particle RNAs. *RNA* **11:** 7–13.

Structural RNAs of known and unknown function identified in malaria parasites by comparative genomics and RNA analysis

Kausik Chakrabarti, Michael Pearson, Leslie Grate, et al.

RNA 2007 13: 1923-1939 originally published online September 27, 2007
Access the most recent version at doi:10.1261/rna.751807