BMC Genomics

Research article

Identification and comparative analysis of components from the signal recognition particle in protozoa and fungi

Magnus Alm Rosenblad[1], Christian Zwieb[2] and Tore Samuelsson*[1]

Address: [1]Department of Medical Biochemistry, Goteborg University, Box 440, SE-405 30 Goteborg, Sweden and [2]Department of Molecular Biology, The University of Texas Health Science Center at Tyler, 11937 US Highway 271, Tyler TX 75708-3154, U.S.A

Email: Magnus Alm Rosenblad - magnus.alm@medkem.gu.se; Christian Zwieb - zwieb@uthct.edu;
Tore Samuelsson* - tore.samuelsson@medkem.gu.se

* Corresponding author

Abstract

Background: The signal recognition particle (SRP) is a ribonucleoprotein complex responsible for targeting proteins to the ER membrane. The SRP of metazoans is well characterized and composed of an RNA molecule and six polypeptides. The particle is organized into the S and Alu domains. The Alu domain has a translational arrest function and consists of the SRP9 and SRP14 proteins bound to the terminal regions of the SRP RNA. So far, our understanding of the SRP and its evolution in lower eukaryotes such as protozoa and yeasts has been limited. However, genome sequences of such organisms have recently become available, and we have now analyzed this information with respect to genes encoding SRP components.

Results: A number of SRP RNA and SRP protein genes were identified by an analysis of genomes of protozoa and fungi. The sequences and secondary structures of the Alu portion of the RNA were found to be highly variable. Furthermore, proteins SRP9/14 appeared to be absent in certain species. Comparative analysis of the SRP RNAs from different *Saccharomyces* species resulted in models which contain features shared between all SRP RNAs, but also a new secondary structure element in SRP RNA helix 5. Protein SRP21, previously thought to be present only in *Saccharomyces*, was shown to be a constituent of additional fungal genomes. Furthermore, SRP21 was found to be related to metazoan and plant SRP9, suggesting that the two proteins are functionally related.

Conclusions: Analysis of a number of not previously annotated SRP components show that the SRP Alu domain is subject to a more rapid evolution than the other parts of the molecule. For instance, the RNA portion is highly variable and the protein SRP9 seems to have evolved into the SRP21 protein in fungi. In addition, we identified a secondary structure element in the *Sacccharomyces* RNA that has been inserted close to the Alu region. Together, these results provide important clues as to the structure, function and evolution of SRP.

Background

The mammalian signal recognition particle (SRP) plays a critical role in targeting of proteins to the ER membrane. SRP first binds the N-terminal signal sequence of the nascent chain as it appears on the surface of translating ribosomes. As a result, protein synthesis is arrested and the ribosome-nascent chain-SRP complex is targeted to the ER membrane through interaction with the SRP receptor [1]. In a series of events that are accompanied by GTP hydrolysis, the SRP is released, protein synthesis is resumed and translocation of the secretory protein is initiated.

The mammalian SRP is composed of six polypeptides named SRP9, SRP14, SRP19, SRP54, SRP68 and SRP72 which form a complex with a single RNA molecule (originally referred to as 7SL RNA) of approximately 300 nucleotide residues. The S domain (Fig. 1) of SRP is responsible for signal sequence recognition and contains the central region of SRP RNA and proteins SRP19, SRP54, SRP68 and SRP72. SRP54 is a highly conserved protein which is responsible for signal sequence binding and it interacts with the helix 8 region of the RNA. The Alu domain of the SRP (Fig. 1) functions in translational arrest and is composed of proteins SRP9/14 bound to the terminal regions of SRP RNA [2,3]. High-resolution three-dimensional structures of the Alu domain and critical parts of the S domain were obtained recently [4-7] and provided considerable insight into structure and function of the mammalian SRP.

Components of the SRP have been identified in all three domains of life [8]. The genomes of the Archaea were shown to contain SRP RNAs which closely resemble the sequences and secondary structures of the SRP RNAs of metazoans, but only two SRP protein genes (SRP19 and SRP54) could be identified [9]. The bacterial SRP consists of protein SRP54 (referred to as *Ffh*) and a 4.5S RNA which corresponds in large part to SRP RNA helix 8 of mammalian SRP. Significantly larger bacterial SRP RNAs (6S RNAs) which contain an Alu-like region are present in a restricted number of taxa such as *Bacillus* [8]. A rationale for the high level of conservation of SRP54 and SRP RNA helix 8 in every SRP has been provided by the high-resolution structure of the *E. coli* SRP which suggested that the signal peptide binds within a hydrophobic groove formed by the M-domain of SRP54 as well as to SRP RNA [10].
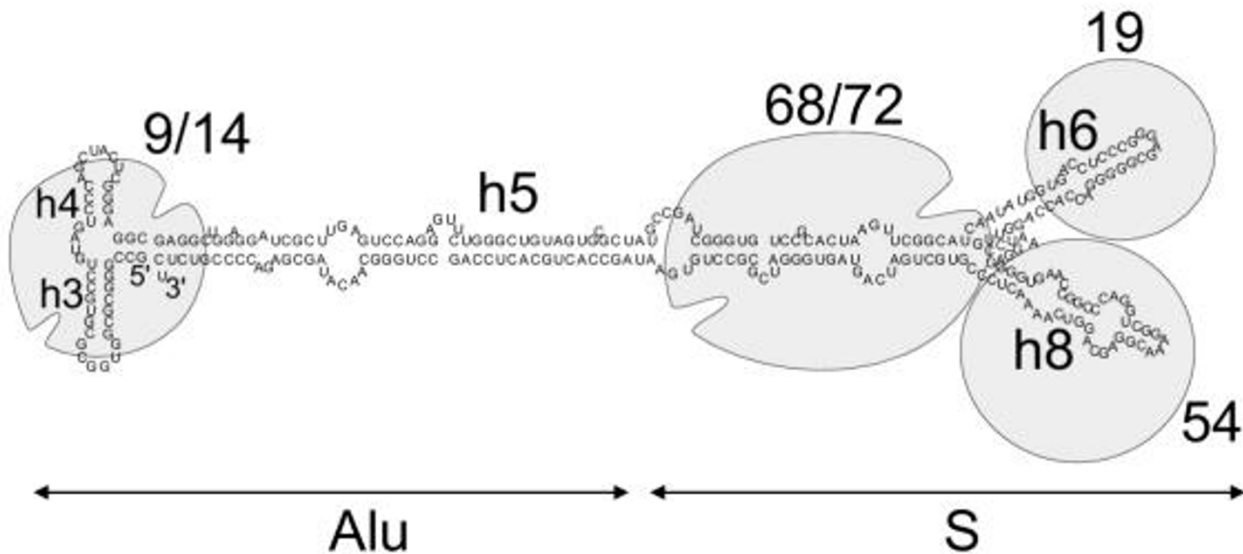
Figure 1
Schematic view of SRP in metazoans. The protein subunits with molecular weights 9, 14, 19, 54, 68 and 72 kDa are shown as well as RNA helix numbers 3, 4, 5, 6 and 8 that are referred to in the text. The human SRP RNA sequence is shown. The Alu domain contains the 5' and 3' terminal regions of the RNA as well as the SRP9 and 14 proteins. The S domain contains the central region of the RNA and the four remaining SRP proteins.

Table 1: Overview of inventory of SRP in protozoa and fungi. SRP RNAs and proteins were predicted as described in the text. Symbols are as follows: +) subunit found, -) subunit not found and genome complete, *) previously reported subunit, M) multiple SRP RNA-like sequences were found and P) only partial RNA sequence found. Empty cells are instances where subunit has not been found and where there is no complete genome assembly and preliminary gene annotation.

Taxonomic group	Organism	Genome size (MB)	Genome sequence data available in this work (MB)	SRP RNA	SRP9/21	SRP14	SRP19	SRP54	SRP68	SRP72
Euglenozoa	*Leishmania major*	34	43.7	+*			+	+	+	+
	Trypanosoma cruzi	?	43.0	P			+	+	+	+
Alveolata	*Plasmodium falciparum*	23	23.1	+	+*	+	+*	+*	+	-
	Theileria annulata	10	8.5	+			+	+	+	
	Eimeria tenella	60	52.3	+		+		+	+	
Viridiplantae	*Chlamydomonas reinhardtii*	95?	97.3	M	+	+	-	+	+	+
Diplomonadida	*Giardia lamblia*	12	120					+	+	+
Entamoebidae	*Entamoeba histolytica*	20	86.7				+		+	+
Fungi-Microsporidia	*Encephalitozoon cuniculi*	2.5	2.5	+	-	-	+	+	-	-
Fungi-Ascomycota	*S. pombe*	14	14	+*	+*	+*	+*	+*	+*	+*
	Neurospora crassa	40	38	+*	+	+	+	+	+	+
	Aspergillus nidulans	28	30	+				+	+	+
	Candida albicans	16	17.8	+*	+		+	+	+	+
	Saccharomyces	12	12	+*	+*	+*	+*	+*	+*	+*

Our understanding of the SRP components of the lower eukaryotes has suffered from a lack of data required for comparative sequence analysis. In addition, the greater diversity of this phylogenetic group has made it difficult to identify the SRP RNAs and SRP proteins even after the genome sequences became available. For instance, despite the detailed biochemical characterization of the yeast *S. cerevisiae* SRP [11], an understanding of its structure has been hampered by the fact that yeast SRP RNA is nearly twice as long (519 nucleotide residues) with no obvious homology to other known SRP RNA sequences [12].

Here we report an analysis of protozoan and fungal genomes to identify several not previously annotated SRP components. We have been able to identify several novel SRP RNAs and compare their two-dimensional structures. The analysis has lead us to propose that protein SRP21 of *S. cerevisiae* is a homolog of SRP9 and thus might form a heterodimer with SRP14 which binds to the Alu domain. These studies provide not only inroads into the comprehensive molecular characterization of the SRP but also clues as to the early evolution and origin of SRP and its Alu domain.

Results and discussion
In aiming to produce a comprehensive inventory of SRP components in protozoa and fungi we considered Euglenozoans (*Entosiphon*, *Trypanosoma*, and *Leishmania)*, Alveolata (*Plasmodium*, *Eimeria*, *Theileria*), *Chlamydomonas*, *Giardia*, *Entamoeba* and *Encephalitozoon*. Complete genome sequences and preliminary gene annotation were available for *P. falciparum* http://www.plasmodb.org, *C. reinhardtii* http://genome.jgi-psf.org/chlre1/chlre1.home.html, and *Encephalitozoon cuniculi* [13]. Significant portions of the other genomes had been sequenced as indicated in Table 1. For *Entosiphon* only a very limited amount of sequence data was available. A schematic phylogenetic tree involving the organisms discussed here is shown in Fig. 2. An overview of the results of our inventory of SRP RNA and proteins in protozoa and fungi is shown in Table 1. A significant number of these were not previously annotated.

Identification and analysis of SRP RNA genes
To predict SRP RNA genes from protozoans and fungi we used a method previously described [14]. The first step is a heuristic pattern-based search for conserved features of the helix 8 region as described under "Methods". The pattern was relatively degenerate and in many cases the result included a number of false positives. In a second step the
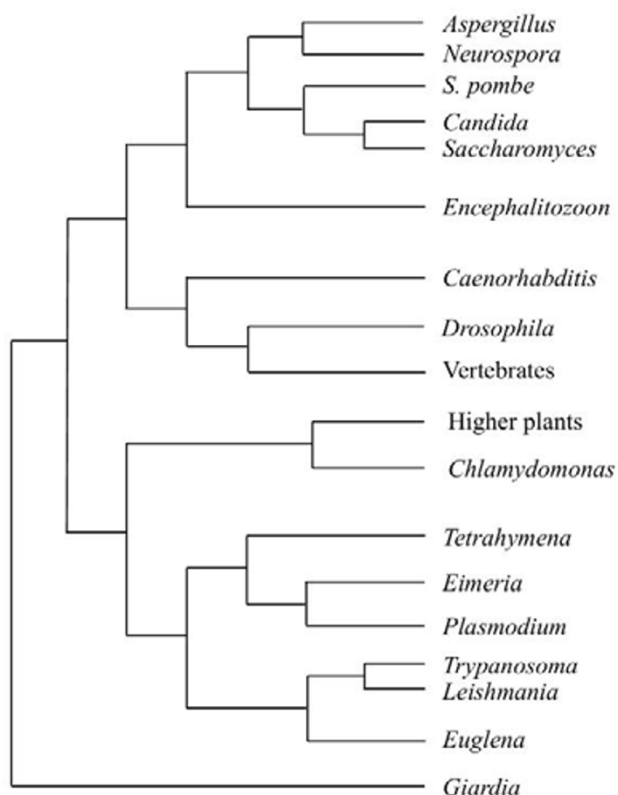
Figure 2
Phylogenetic tree. A schematic tree is shown that includes the yeasts and protozoa referred to in the text. It was based on the tree shown in Baldauf et al [18]. Branch lengths are not proportional to evolutionary distance.

candidates from the pattern-based search were screened with a COVE model of SRP RNA. COVE is an implementation of the algorithms described by Eddy & Durbin [15] that make use of probabilistic models to describe the sequence and secondary structure consensus of an RNA family. The COVE analysis is much more stringent than the first pattern-based step and we expect very few false positives among the high-scoring hits from this analysis.

The identified SRP RNA candidates aligned well to a COVE model for eukaryotic RNAs in the conserved S domain region, i.e. the part that corresponded to the helices 5, 6 and 8. The Alu domain displayed a higher degree of variation and in many instances did not align well to the COVE model. As a consequence, the prediction of the 5' and 3' ends of the molecule were in most cases unreliable.

The secondary structure of all candidates were also predicted with MFOLD [16] with default parameters or specifying constraints consistent with known conserved elements of SRP RNA. MFOLD was able to fold the S domain of the SRP RNA in a manner which was consistent with the secondary structure predicted by COVE. However, when used without constraints, MFOLD typically predicted a secondary structure for the Alu domain that was inconsistent with our identifications. As a consequence, for the prediction of Alu domains as well as their folding, we relied on consensus features, such as the presence of a conserved sequence motif UGUNR (where N is any base and R is purine, typically an A) motif and the general secondary structure outline (Fig. 1). In summary, in our prediction and folding of SRP RNA we combined pattern matching, COVE, and MFOLD, and we checked that known consensus motifs of SRP RNAs were present in the predicted RNAs. Finally, we used BLAST to show that the predicted SRP RNA genes did not overlap with predicted protein-coding regions or any other annotated features. Therefore, we believe that the final candidates presented here represent sequences that are evolutionary related to SRP RNA. Still, it should be noted that we cannot distinguish between a bona fide SRP RNA gene and pseudogenes that are known to occur in plants [8,17] and in mammals. Examples are the two SRP RNA gene candidates that we identified in the *C. reinhardtii* genome. The covariance models did not allow us to predict the 5' and 3' ends and the folding of the Alu domain of these two sequences. Therefore, it remains to be seen which of these candidates, if any, represents a functional RNA.

We were not able to identify an SRP RNA in *Giardia lamblia* and *Entamoeba histolytica*. However, as SRP proteins were identified in these organisms we expect that, as the genomes are completed, SRP RNA genes will be discovered.

SRP RNAs of Euglenozoa and Alveolates display a large variation in the Alu domain
An SRP RNA gene was identified in *Entosiphon sulcatum* (Fig. 3). Its Alu domain was found to contain a very short helix 4 and in this respect resembled the structure of the Alu domain of the trypanosomatids. This relationship was consistent with the known close evolutionary relationship between euglenids and trypanosomatids [18]). Interestingly, the *E. sulcatum* SRP RNA gene was shown to be part of a cluster which also contained genes for 5S rRNA, U1, U2 and U5 snRNAs [19,20]). This gene organization is reminiscent of that of *T. brucei* and *Leishmania* where SRP RNA genes were found to be located adjacent to other RNA genes [21,22].

In the group of the Alveolates we found an *Eimeria tenella* SRP RNA (Fig. 3) with a predicted Alu domain structure

Figure 3
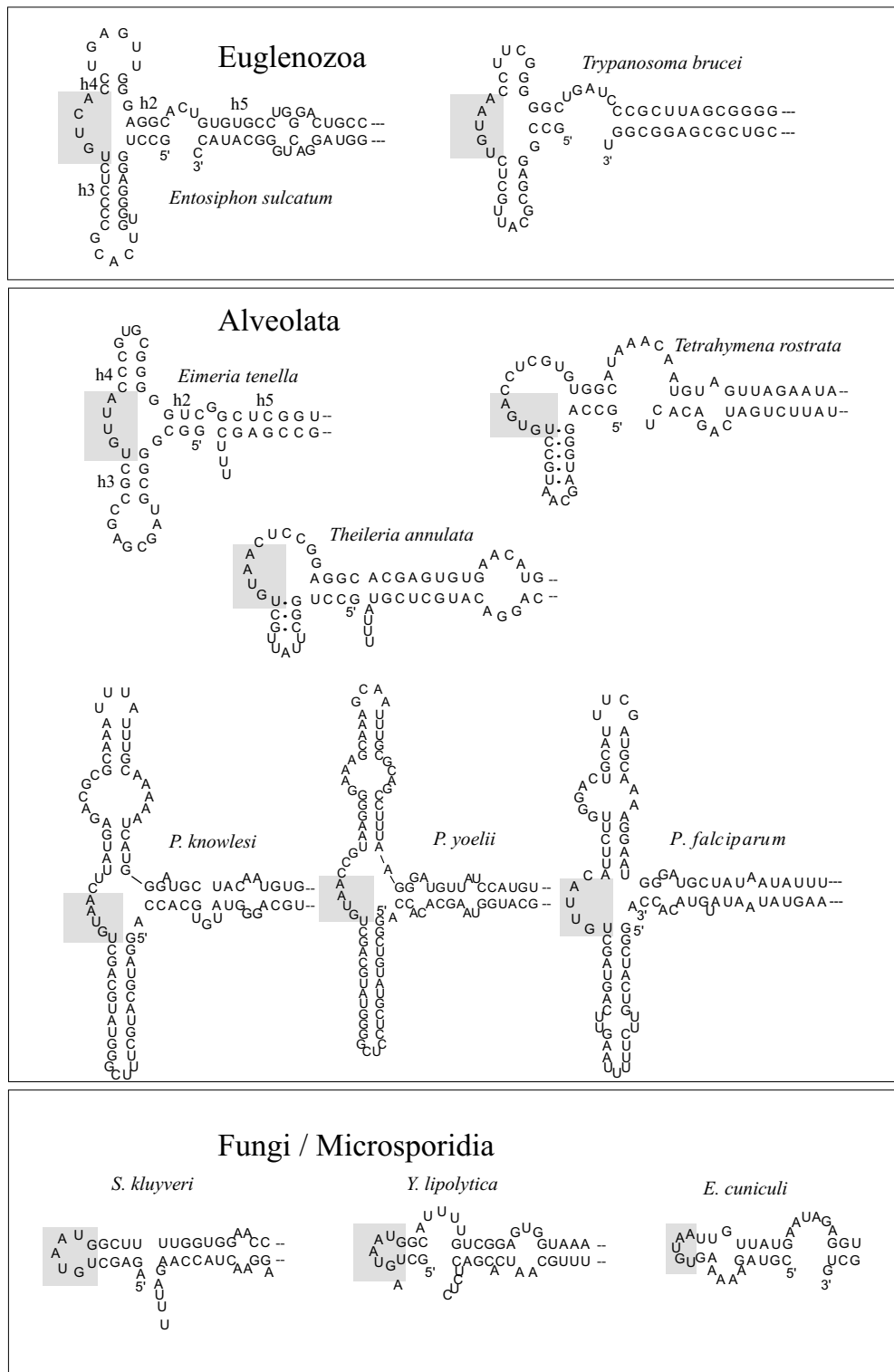**Predicted secondary structures of protozoan SRP RNA Alu domains.** Members of the groups Euglenozoa, Alveolata, and Fungi/Microsporidia are shown. The conserved UGUNR motif characteristic of the Alu domain is indicated by shading. Model of *E. cuniculi* is highly tentative. The complete structures including the S domain are shown in the web supplement at http://bio.lundberg.gu.se/srp03/.

similar to that of the metazoans. Interestingly, the Alu domain of *Theileria annulata* was reminiscent of the SRP RNA Alu domain previously identified in the Ciliophora *Tetrahymena* [23], in the respect that the helix 4 appeared to be absent.

It has previously been reported that the genome of the malaria parasite *Plasmodium falciparum* encodes several SRP proteins [24]. Here, we were able to identify the corresponding SRP RNA (Fig. 3). The secondary structure of the Alu domain of this RNA was predicted by combining COVE and MFOLD procedures. In addition, the RNA of two other Plasmodium species, *P. yoelii* and *P. knowlesi*, were predicted to form the same structure despite significant differences in their primary sequences (Fig. 3). The Alu domain of *Plasmodium* SRP RNA was different in that it possessed an internal loop in helix 4. Therefore, even within the Alveolates, a considerable variation in the predicted folding of the Alu domain was observed.

Saccharomyces SRP RNAs has an insert in helix 5 adjacent to a highly conserved Alu hairpin motif
We previously identified SRP RNA genes in *C. albicans* and *N. crassa* [14]. Here we also found an SRP RNA candidate in *Aspergillus nidulans*. As shown for *Yarrowia* SRP RNA in Fig. 4, in all of these fungi, including *S. pombe,* the SRP RNA secondary structure were shown to be very similar. For the unusually large (519 nts) SRP RNA of *S. cerevisiae*[12], the COVE model predicted an S domain with helices 5, 6 and 8 as for other eukaryotic RNAs. MFOLD also folded this part of the molecule in accordance with the consensus 2D structure of the S domain. As the 5' and 3' terminal sequences were shown to be related to Alu, we concluded that the *S. cerevisiae* SRP RNA contained at least one insert as compared to other fungi. Secondary structures for the SRP RNAs including these inserts were constructed for *S. mikatae, S. kudriavzevii, S. bayanus, S. castellii* and *S. kluyveri.* The sequences were identified by BLAST using the *S. cerevisiae* sequence as query. For the prediction of the 5' end of the RNA we took advantage of the fact that the highly conserved Alu domain was present at the very 5' end of the RNA. For prediction of the 3' end we considered a T-rich region which was conserved in all six *Saccharomyces* strains and likely is part of a transcription termination signal.

Since the number of available *Saccharomyces* SRP RNA sequences was too low for using covariation or mutual information analysis, and a COVE model that would predict the pairing in the insert regions could not be obtained, MFOLD was used first with each full-length sequence to predict the secondary structure. As expected, all *Saccharomyces* sequences folded into structures which contained helices 5, 6 and 8 as predicted for *S. cerevisiae*. Furthermore, all *Saccharomyces* RNAs possessed a hairpin

structure with the Alu UGUNR motif at the 5' end similar to what was observed in the Alu domains of the other fungi. A multiple alignment was obtained using procedures described under Methods and is available in the web supplement to this paper http://bio.lund berg.gu.se/srp03/.

As for the SRP RNA insertions specific to *Saccharomyces*, MFOLD predicted the structure shown in Fig. 4 containing the helices that we here refer to as 5c-g (Fig. 4). The helices 5h-i were also characteristic of the *Saccharomyces* RNAs. Smaller corresponding inserts reminiscent of these were found in *Yarrowia*, *Neurospora* and *Aspergillus*. The predicted secondary structure of the 5c-g region was very similar in all *Saccharomyces* species although there was significant variation in sequence. The bases involved in compensatory base changes in this part of the RNA (Fig. 4) offer support to the predicted folding.

The 5c-g insertion appeared to be specific to *Saccharomyces* and indicated that at some point during evolution this piece of the RNA was inserted near the 5' end as indicated in Fig. 4. The evolution of *Saccharomyces* also involved the enlargement of helices 5h and 5i. One may speculate that these species have developed some additional mechanisms related to SRP-mediated translational arrest or additional unknown functions.

Our findings suggested that all fungal SRP RNAs contain a hairpin structure with a Alu UGUNR motif at the 5' end. This part of the RNA was found to be conserved and there was phylogenetic support for the hairpin structure from the covariations indicated in Fig. 4. The fungal hairpin motif was distinct from all other known non-fungal Alu domains where the conserved UGUNR motif was found to be part of a more elaborate pseudoknot.

It has been shown previously that a 5' terminal 99 nt fragment of the *S. cerevisiae* RNA was able to bind *in vitro* to the SRP14 protein. A tentative model of this portion of the SRP RNA has been presented where only the 5' terminal portion formed the Alu domain [25]. However, based on the analysis presented here it is likely that also the 3' portion is part of the Alu domain.

An SRP RNA candidate in E. cuniculi
There is strong evidence that the Microsporidia, such as *E. cuniculi*, are phylogenetically related to Fungi [13]. An analysis of the *E. cuniculi* genome revealed a candidate SRP RNA in a 396 nt intergenic region (chromosome X, positions 138833–139226) which contained helices 6 and 8, and aligned well with our eukaryotic COVE model. However, we were unable to identify a typical metazoan Alu domain. Fig. 3 shows a tentative model of the Alu domain which is similar to that of other fungi.

A. nidulans	AUCUGUUAUGGAU
N. crassa	GACUGUAAUGGUC
S. pombe	CGCUGUAAUGGCU
C. albicans	GGCUGUGAUGGCU
Y. lipolytica	AGCUGUAAUGGCA
S. kluyveri	AGCUGUAAUGGCU
S. cerevisiae	GGCUGUAAUGGCU
	<<<< >>>>
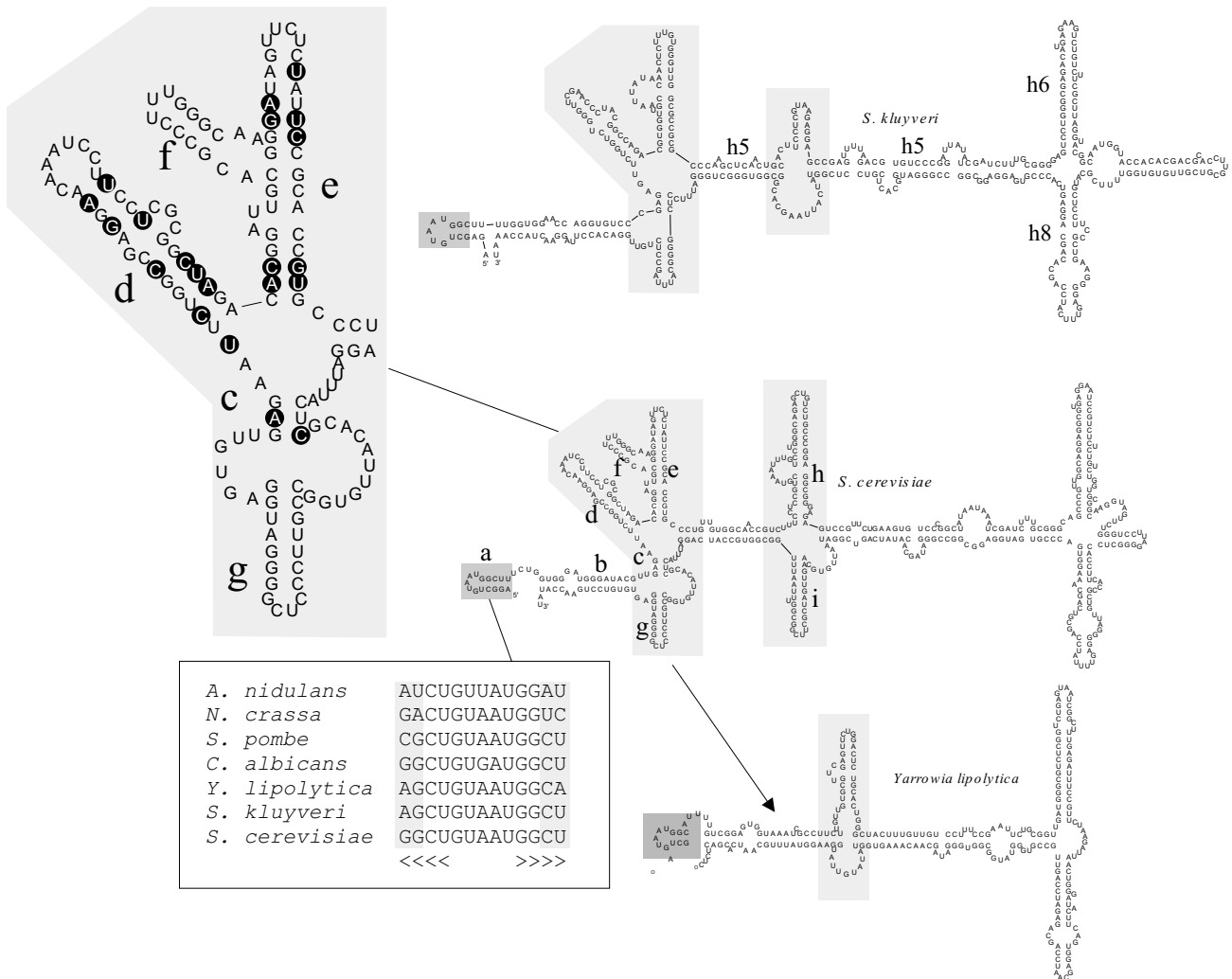
Figure 4
Proposed secondary structures of *Saccharomyces* SRP RNAs. Models of the RNAs of *S. kluyveri*, *S. cerevisiae* and *Y. lipolytica* are shown. *Saccharomyces* RNAs have regions, 5c-g and 5h-i (shaded) not found in other fungi. Inset with shaded background with helices 5c-g are showing compensatory base changes in this domain (bases with white background). Also indicated is the location of the 5c-g insert as related to *Yarrowia* (arrow) as well as conserved sequence elements of the fungal Alu domain (box). The complete models of other *Saccharomyces* species are shown in the web supplement at http://bio.lund berg.gu.se/srp03/.

Identification of SRP proteins

We used a range of tools to identify and inventory SRP proteins in protozoa and fungi. Genome sequences were analyzed for genes encoding SRP proteins using BLAST [26] or FASTA [27] using previously known eukaryotic SRP proteins as query sequences. In addition, we performed PSI-BLAST [26] searches where a SRP protein sequence, typically the human ortholog, was used to search a database with the proteins in a public protein sequence database combined with the proteins obtained by translating all possible open reading frames of the genome being analyzed. Genomes were also analyzed using Genscan [28] or GlimmerM [29] and predicted peptide sequences were used in a BLAST or PSI-BLAST procedure as above. The results of our findings are shown in Table 1. One should keep in mind that several genomes were incomplete although large portions were covered by raw sequence data or contigs. The fact that we failed to

identify a certain component of SRP was therefore not entirely conclusive in these cases (indicated by empty cells in Table 1).

As previously noted for other species, SRP54 was shown to be ubiquitous and highly conserved. Also SRP19, SRP68 and SRP72 were found in most of the genomes analyzed here (Table 1) although we were unable to identify a SRP72 homolog among the Alveolates. In *E. cuniculi* we failed to identify SRP68/72, as described further below. We obtained evidence that the SRP9/14 proteins were subject to a more rapid evolution, as discussed in the following.

Absence of SRP9/14 in certain protozoa and Microsporidia
SRP9 and 14 homologs were identified in *Plasmodium falciparum* and *Chlamydomonas reinhardtii*. The SRP14 homolog in *P. falciparum* has been identified previously but annotated as a hypothetical protein (http://www.plasmodb.org, accession PFL0160w). SRP14 was found also in *E. tenella* and *E. histolytica*. However, we were unable to identify SRP9/14 in *Leishmania major*, *Trypanosoma cruzi*, *Theileria annulata*, *Giardia lamblia* and *E. cuniculi*. Although an intriguing observation, we could not formally rule out the possibility that SRP9/14 homologs would be discovered during the completion of the genome assemblies and that the SRP9/14 protein sequences in these organisms have strongly diverged from known members of this protein family.

Evidence has been provided that *Leptomonas* and *T. brucei* possess a tRNA-like RNA that associates with the SRP [30,31]. It has been speculated that this RNA compensates for the loss of portions of the Alu domain [31]. The possibility that the tRNA-like RNA took the role of proteins SRP9/14 was considered as well. Upon completion of additional trypanosomatid genome sequences it will be interesting to determine if SRP in all these organisms carry a tRNA-like component and if they all lack SRP9/14.

The microsporidian *E. cuniculi* has a highly compact genome that seem to have been under a pressure to eliminate non-essential material. As SRP9/14, 68 and 72 seem to be missing, the evolution of this organism could have involved the loss of these genes. The lack of these proteins would suggest that they are less critical for SRP function. It is interesting to note that in this respect *E. cuniculi* resembles archaea which also appear to lack SRP9/14, 68 and 72.

Yeast SRP21 is related to metazoan and plant SRP9
Homologs of SRP14, SRP19, SRP54, SRP68, and SRP72 were identified in all the *Saccharomyces* species (Table 1). It has been reported previously that *S. cerevisiae* possess SRP21 which was thought to be a new family of SRP

proteins [11] unique to *Saccharomyces*. We have here reexamined the relationship of SRP21 to other proteins, including those of the SRP. Homologs to *S. cerevisiae* SRP21 in *S. mikatae*, *S. kudriavzevii*, *S. bayanus*, *S. castellii*, *S. kluyveri* and *S. paradoxus* sequences were initially identified using BLAST. The E-values ranged from 1e-77 (*S. paradoxus*) to 1e-29 (*S. kluyveri*). A homolog to SRP21 was found also in *Candida albicans* (E-value 8e-06). Open reading frames in the region of the BLAST hit were identified to obtain a likely full-length protein sequence for the *C. albicans* SRP21 (Fig. 5). We also identified the *S. pombe* protein YE07_SCHPO (T37873) as a distantly related homolog, with a E-value of 6.3. Interestingly, this protein displayed a distant homology to metazoan protein SRP9 and was previously listed in the SRPDB as a potential SRP9 homolog. Using the *S. pombe* protein as query we identified a possible homolog in *Neurospora crassa* (E-value 5e-6) using TBLASTN to search *N. crassa* genomic sequences (Fig. 5).

We considered that SRP21 might bind to the regions inserted specifically into *Saccharomyces* SRP RNA, i.e helices 5c-g or 5h-i. However, this possibility appeared unlikely because SRP21 homologs were identified also in those fungi that did not possess these helix 5 expansions.

To further examine the SRP21 homologs and their relationship with SRP9 we carried out profile-based searches. First, all novel putative yeast SRP21 homologs identified here were merged with the sequences of public protein sequence databases (Genpept and SWISSPROT/TREMBL). The resulting databases were used in profile-based searches including PSI-BLAST, PROFILESEARCH, and hmmsearch. For instance, when the *S. cerevisiae* SRP21 sequence was used as query in a PSI-BLAST search, after two iterations the *S. pombe* protein YE07_SCHPO (T37873) was identified above the default threshold (E-value 0.01) together with the *Candida* and *Saccharomyces* SRP21 proteins (not shown). Second, PROFILESEARCH was used with a profile based on a multiple alignment of the *Saccharomyces* sequences (Fig. 6, sequences shown in lighter gray) obtained by CLUSTALW to search SWISS-PROT/TREMBL (approximately 1 million protein sequences as of March 2003). In the result of this search the *Candida*, *Neurospora*, and *S. pombe* sequences followed immediately after the *Saccharomyces* SRP21 sequences (Fig. 6). Interestingly, the SRP9 proteins of maize, *Arabidopsis*, and *C. elegans* ranked closely to the top-scoring yeast sequences, although they had scores lower than the other SRP21-related proteins. Similar results were obtained with hmmsearch (not shown).

A multiple alignment of SRP21 and SRP9 as well as SRP14 protein sequences is shown in Fig. 5. The alignment of SRP9 to SRP14 is the structural alignment of Birse et al.
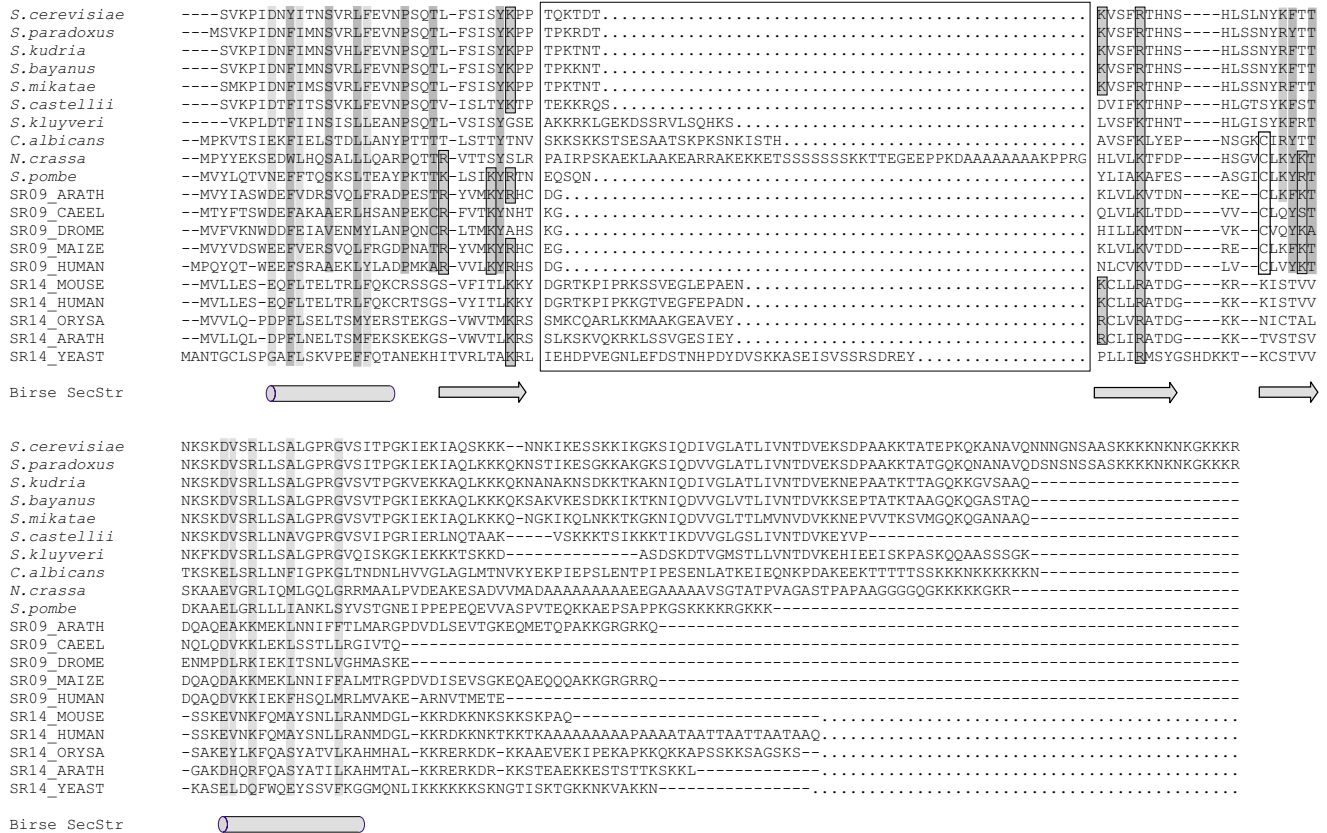
```
S.cerevisiae  ----SVKPIDNYITNSVRLFEVNPSQTL-FSISYKPP  TQKTDT.....................................................  KVSFRTHNS----HLSLNYKFTT
S.paradoxus   ---MSVKPIDNFIMNSVRLFEVNPSQTL-FSISYKPP  TPKRDT.....................................................  KVSFRTHNS----HLSSNYRYTT
S.kudria      ----SVKPIDNFIMNSVHLFEVNPSQTL-FSISYKPP  TPKTNT.....................................................  KVSFRTHNS----HLSSNYRFTT
S.bayanus     ----SVKPIDNFIMNSVRLFEVNPSQTL-FSISYKPP  TPKKNT.....................................................  KVSFRTHNS----HLSSNYKFTT
S.mikatae     ----SMKPIDNFIMSSVRLFEVNPSQTL-FSISYKPP  TPKTNT.....................................................  KVSFRTHNP----HLSSNYRFTT
S.castellii   ----SVKPIDTFITSSVKLFEVNPSQTV-ISLTYKTP  TEKKRQS....................................................  DVIFKTHNP----HLGTSYKFST
S.kluyveri    -----VKPLDTFINSISLLEANPSQTL-VSISYGSE  AKKRKLGEKDSSRVLSQHKS........................................  LVSFKTHNT----HLGISYKFRT
C.albicans    --MPKVTSIEKFIELSTDLLANYPTTTT-LSTTYTNV  SKKSKKSTSESAATSKPKSNKISTH...................................  AVSFKLYEP---NSGKCLRYTT
N.crassa      --MPYYEKSEDWLHQSALLLQARPQTTR-VTTSYSLR  PAIRPSKAEKLAAKEARRAKEKKETSSSSSSSSKKTTEGEEPPKDAAAAAAAAKPPRG  HLVLKTFDP----HSGVCLKYKT
S.pombe       --MVYLQTVNEFFTQSKSLTEAYPKTTK-LSIKYRTN  EQSQN......................................................  YLIAKAFES----ASGICLKYRT
SR09_ARATH    --MVYIASWDEFVDRSVQLFRADPESTR-YVMKYRHC  DG.........................................................  KLVLKVTDN----KE--CLKFKT
SR09_CAEEL    --MTYFTSWDEFAKAAERLHSANPEKCR-FVTKYNHT  KG.........................................................  QLVLKLTDD----VV--CLQYST
SR09_DROME    --MVFVKNWDDFEIAVENMYLANPQNCR-LTMKYAHS  KG.........................................................  HILLKMTDN----VK--CVQYKA
SR09_MAIZE    --MVYVDSWEEFVERSVQLFRGDPNATR-YVMKYRHC  EG.........................................................  KLVLKVTDD----RE--CLKFKT
SR09_HUMAN    -MPQYQT-WEEFSRAAEKLYLADPMKAR-VVLKYRHS  DG.........................................................  NLCVKVTDD----LV--CLVYKT
SR14_MOUSE    --MVLLES-EQFLTELTRLFQKCRSSGS-VFITLKKY  DGRTKPIPRKSSVEGLEPAEN......................................  KCLLRATDG----KR--KISTVV
SR14_HUMAN    --MVLLES-EQFLTELTRLFQKCRTSGS-VYITLKKY  DGRTKPIPKKGTVEGFEPADN......................................  KCLLRATDG----KK--KISTVV
SR14_ORYSA    --MVVLQ-PDPFLSELTSMYERSTEKGS-VWVTMKRS  SMKCQARLKKMAAKGEAVEY.......................................  RCLVRATDG----KK--NICTAL
SR14_ARATH    --MVLLQL-DPFLNELTSMFEKSKEKGS-VWVTLKRS  SLKSKVQKRKLSSVGESIEY.......................................  RCLIRATDG----KK--TVSTSV
SR14_YEAST    MANTGCLSPGAFLSKVPEFFQTANEKHITVRLTAKRL  IEHDPVEGNLEFDSTNHPDYDVSKKASEISVSSRSDREY...................  PLLIRMSYGSHDKKT--KCSTVV

Birse SecStr

S.cerevisiae  NKSKDVSRLLSALGPRGVSITPGKIEKIAQSKKK--NNKIKESSKKIKGKSIQDIVGLATLIVNTDVEKSDPAAKKTATEPKQKANAVQNNNGNSAASKKKKNKNKGKKKR
S.paradoxus   NKSKDVSRLLSALGPRGVSITPGKIEKIAQLKKKQKNSTIKESGKKAKGKSIQDVVGLATLIVNTDVEKSDPAAKKTATGQKQNANAVQDSNSNSSASKKKKNKNKGKKKR
S.kudria      NKSKDVSRLLSALGPRGVSVTPGKVEKKAQLKKKQKNANAKNSDKKTKAKNIQDIVGLATLIVNTDVEKNEPAATKTTAGQKKGVSAAQ----------------------
S.bayanus     NKSKDVSRLLSALGPRGVSVTPGKIEKKAQLKKKQKSAKVKESDKKIKTKNIQDVVGLVTLIVNTDVKKSEPTATKTAAGQKQGASTAQ---------------------
S.mikatae     NKSKDVSRLLSALGPRGVSVTPGKIEKIAQLKKKQ-NGKIKQLNKKTKGKNIQDVVGLTTLMVNVDVKKNEPVVTKSVMGQKQGANAAQ---------------------
S.castellii   NKSKDVSRLLNAVGPRGVSVIPGRIERLNQTAAK-----VSKKKTSIKKKTIKDVVGLGSLIVNTDVKEYVP---------------------------------------
S.kluyveri    NKFKDVSRLLSALGPRGVQISKGKIEKKKTSKKD-------------ASDSKDTVGMSTLLVNTDVKEHIEEISKPASKQQAASSSGK---------------------
C.albicans    TKSKELSRLLNFIGPKGLTNDNLHVVGLAGLMTNVKYEKPIEPSLENTPIPESENLATKEIEQNKPDAKEEKTTTTTSSKKKNKKKKKKN-----------------
N.crassa      SKAAEVGRLIQMLGQLGRRMAALPVDEAKESADVVMADAAAAAAAAAEEGAAAAAVSGTATPVAGASTPAPAAGGGGQGKKKKKGKR--------------------
S.pombe       DKAAELGRLLLIANKLSYVSTGNEIPPEPEQEVVASPVTEQKKAEPSAPPKGSKKKKRGKKK----------------------------------------
SR09_ARATH    DQAQEAKKMEKLNNIFFTLMARGPDVDLSEVTGKEQMETQPAKKGRGRKQ-------------------------------------------
SR09_CAEEL    NQLQDVKKLEKLSSTLLRGIVTQ----------------------------------------------------
SR09_DROME    ENMPDLRKIEKITSNLVGHMASKE---------------------------------------
SR09_MAIZE    DQAQDAKKMEKLNNIFFALMTRGPDVDISEVSGKEQAEQQQAKKGRGRRQ-----------------------
SR09_HUMAN    DQAQDVKKIEKFHSQLMRLMVAKE-ARNVTMETE--------------------------------------
SR14_MOUSE    -SSKEVNKFQMAYSNLLRANMDGL-KKRDKKNKSKKSKPAQ----------------------
SR14_HUMAN    -SSKEVNKFQMAYSNLLRANMDGL-KKRDKKNKTKKTKAAAAAAAAAPAAAATAATTAATTAATAAQ................
SR14_ORYSA    -SAKEYLKFQASYATVLKAHMHAL-KKRERKDK-KKAAEVEKIPEKAPKKQKKAPSSKKSAGSKS--................
SR14_ARATH    -GAKDHQRFQASYATILKAHMTAL-KKRERKDR-KKSTEAEKKESTSTTKSKKL-----------................
SR14_YEAST    -KASELDQFWQEYSSVFKGGMQNLIKKKKKKSKNGTISKTGKKNKVAKKN----------------

Birse SecStr
```

Figure 5
Alignment of *Saccharomyces* SRP21 proteins with *C. albicans*, *N. crassa* and *S. pombe* homologs, metazoan and plant SRP9 proteins as well SRP14 proteins. The alignment of SRP9 to SRP14 is that previously shown in Birse et al [32]. The alignment of SRP9 to SRP21 is based on the results of profile-based searches like the one shown in Fig. 6 and CLUSTALW. The secondary structure from *Birse et al.* is shown with gray cylinders for α-helices and arrows for β-strands. The region between β-1 and β-2 (box) is not aligned. Boxed residues are basic residues and cysteines protruding from the β-sheet into the solvent, according to *Birse et al.* Highly conserved residues are shown in dark gray, residues with conservative substitutions having the same physico-chemical properties are shown in light gray. Where organism names are given (*Saccharomyces*, *C. albicans*, *N. crassa* and *S. pombe*) they refer to the SRP21 proteins identified in this work whereas the SRP9 and SRP14 proteins are denoted by their Swissprot/TREMBL names.

[32] and the alignment of SRP9 to SRP21 was the result of a CLUSTALW analysis which was consistent with the results obtained from the profile searches described above. Interestingly, many of the positions that were conserved in the SRP9/14 structural alignment [32] were occupied by the same category of amino acids in the SRP21 proteins (Fig. 5). These data indicated that SRP21 is structurally similar to the SRP9/14 proteins and provided further evidence of the homology between SRP21 and SRP9.

In mammalian SRP, proteins SRP9 and SRP14 were shown to form a heterodimer and share a αββα topology [32]. To determine if the secondary structure predicted for fungal SRP21 was similar to the SRP9 and SRP14 structure, we made predictions with PSIPRED [33,34]. *Saccharomyces* SRP21 sequences were used as input and the results are shown in Fig. 7. For human SRP9 and SRP14 the boxed regions indicate the positions of the secondary structure elements as known from the structure of the proteins. The corresponding regions for the fungal proteins are also shown in boxes and are based on the alignment in Fig. 5. The results showed that the predicted secondary structure of SRP21 was remarkably similar to the predicted or known structures of SRP9 and SRP14. With the exception of the first α-helix of *Saccharomyces* SRP21 all the predicted secondary structure elements were consistent with those of the SRP9/14 structure. Similar

Sequence	Zscore	Orig	Length
1. SRP21 S. bayanus	84.40	349.21	146
2. SRP21 S. paradoxus	83.75	351.99	169
3. SRP21 S. kudriavzevii	83.71	346.53	146
4. SRP21 S. cerevisiae	82.41	346.02	166
5. SRP21 S. mikatae	82.06	339.92	145
6. SRP21 S. kluyveri	75.51	314.59	145
7. SRP21 S. castellii	74.91	308.02	125
8. Pot. SRP21 C. albicans	21.67	107.92	168
9. Pot. SRP21 N. crassa	12.74	74.20	197
10.YE07_SCHPO	9.61	58.63	120
11.CD3D_RAT	8.74	57.34	173
12.ARP2_PLAFA	8.52	65.04	451
13.Q23147 C. elegans	8.50	60.12	284
14.SR09_ARATH	8.45	53.56	103
15.Q8K2G5 mouse	8.45	60.59	306
16.Q8BFQ4 mouse	8.40	60.59	313
17.Q8I562 P.fal	8.39	64.65	459
18.AAH44174 D.rerio	8.28	60.11	313
19.SR09_MAIZE	8.17	52.51	103
20.CD3D_MOUSE	8.11	54.85	173
21.SR09_CAEEL	7.90	50.47	76

Figure 6
Profile-based search reveals relationship between SRP9 and SRP21. PROFILESEARCH (Wisconsin package version 10.2, Genetics Computer Group (GCG), Madison, Wisc.) with default parameters was used with a profile based on a multiple alignment of SRP21 from *Saccharomyces* species created with CLUSTALW. The database queried was one where all the SRP21-related proteins identified here were merged with Swissprot/TREMBL. Swissprot/TREMBL names are given for the sequences except for the six *Saccharomyces*, *N. crassa* and *C. albicans* sequences identified here. YE07_SCHPO is the *S. pombe* SRP21 / SRP9 homolog. The *Saccharomyces sequences* at the top of the list (lighter gray) were those used to create the profile.

results were obtained with the secondary structure predic-   tion method SAMT02 [35] (not shown).
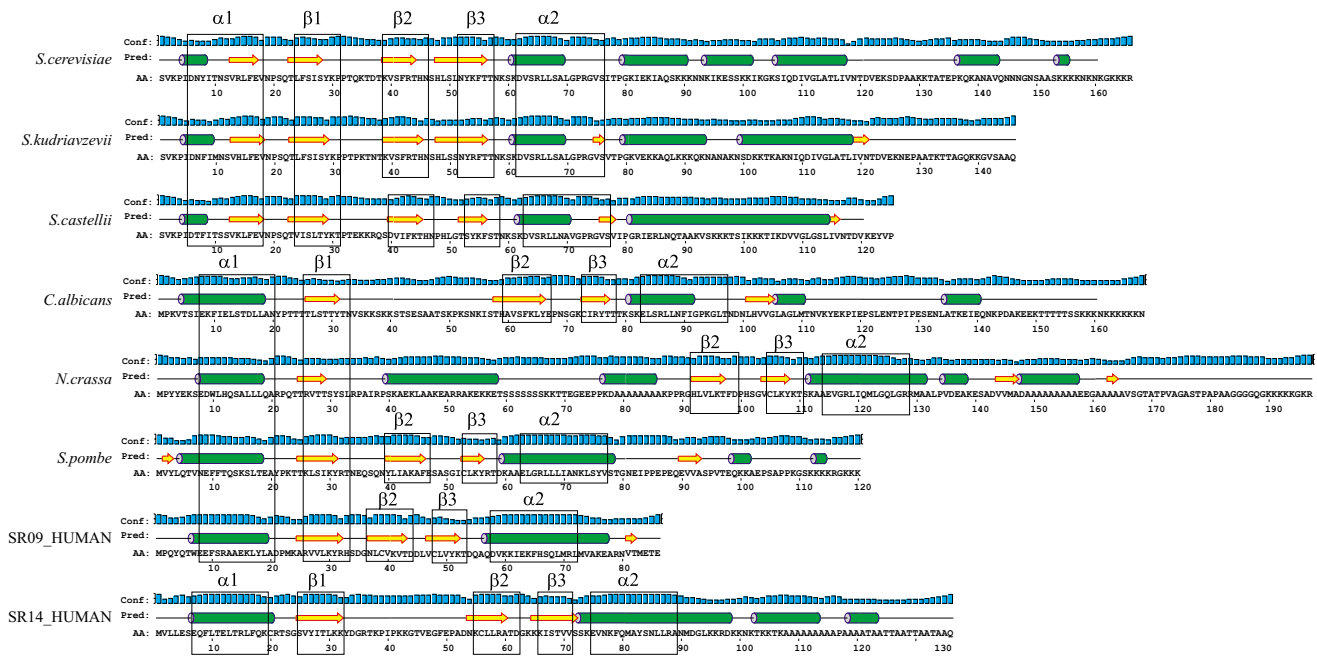
Figure 7
Protein secondary structure predictions of SRP21, SRP9 and SRP14 proteins. Protein sequences, SRP21 and poten-
tial homologs from the organisms indicated as well as human SRP9 and SRP14, were subjected to secondary structure predic-
tion by PSI-PRED [33]. Sequences and predictions are shown unaligned. α-helices and β-strands are indicated by cylinders and
arrows, respectively. For human SRP9 and SRP14 the boxed regions indicate the positions of the secondary structure elements
as known from the structure of the proteins. The corresponding regions for the fungal proteins are also shown in boxes and
are based on the alignment in Fig. 5.

Potential interactions of yeast SRP21
In metazoan cells, SRP9/14 forms a complex with the Alu
domain of SRP RNA. The structure of this complex has
been determined at high resolution [5,32]. On the other
hand, the Alu domain of yeast is less well characterized.
An analysis of yeast SRP showed that it was missing an
obvious SRP9 homolog [11]. SRP21 was not believed to
be the SRP9 equivalent in yeast because its sequence sim-
ilarity to SRP9 was not recognized and SRP21 did not
appear to be stably associated with SRP14. Furthermore,
evidence was provided that SRP14 is present in two copies
in the yeast SRP [3] and SRP14 was shown to form a
homodimer which bound to the Alu domain [25]. On the
basis of these observations it was assumed that the SRP14
homodimer was functionally equivalent to the SRP9/14
heterodimer. In contrast, we suggest that SRP21 is not
only structurally but also functionally related to SRP9. We
suggest that the protein is an integral component not only
of *Saccharomyces*, but of every yeast SRP. In the light of
these findings it will be important to reexamine experi-
mentally the role of SRP21 in yeast.

Conclusions
In the process of identifying and analyzing numerous SRP
RNAs in protozoa and fungi we have demonstrated that
the RNA portion of the Alu domain is highly variable both
in sequence and secondary structure. Although the RNAs
possess a conserved UGUNR motif, other parts of the Alu
domain show a large degree of variation. One striking
example of the plasticity of Alu is apparent in the Alveo-
lates where the *Plasmodium* Alu domain is distinct from all
other members of that group. Furthermore, in fungi the
Alu domain is a simple hairpin motif as compared to all
other species where the Alu domain is more elaborate. We
have also identified secondary structure element
insertions in the *Saccharomyces* SRP RNAs towards the ter-
minal regions which could be considered as expansions of
the Alu domain.

The Alu associated SRP9 and SRP14 proteins appear to
have been subject to rapid evolution as well. One example
is the evolution of the fungal SRP21 protein. Using sensi-
tive profile-based searches, we have presented evidence
that SRP21 is homologous to the metazoan SRP9. In addi-

tion, it seems that SRP9 and SRP14 are missing in some protozoa and fungi, and it is known from previous studies that the Bacillus type eubacteria are missing these proteins and sequence analysis has so far failed to reveal archaebacterial homologs. Based on these findings it is tempting to speculate that the ancestral Alu domain was built solely from RNA, and proteins SRP9/14 were added to adjust Alu function in subsequent evolutionary events.

Methods
Sources of genomic sequences
Some of the genomic sequences used in this work were unfinished sequences and represent as yet unpublished material. In these cases permission to present the results in this paper was obtained from the respective research groups. SRP RNA sequences from *S. mikatae*, *S. kudriavzevii*, *S. bayanus*, *S. castellii* and *S. kluyveri* were retrieved using BLAST searches against the corresponding genomes using the *S. cerevisiae* sequence as query using the BLAST server at the Genome Sequencing Center at Washington University http://www.genome.wustl.edu/blast/yeast_client.cgi. *Saccharomyces* protein sequences were identified using the same BLAST server and the Synteny viewer at the Saccharomyces Genome Database http://www.yeastgenome.org. Sequences of *Candida albicans* were from the Stanford Genome Technology Center website at http://www-sequence.stanford.edu/group/candida (Assembly 6). *Neurospora* sequences were from the Neurospora Sequencing Project, Whitehead Institute/MIT Center for Genome Research http://www-genome.wi.mit.edu. The dataset used in these studies was http://www.broad.mit.edu/ftp/pub/annotation/neurospora/assembly3/neurospora_3.fasta.gz. *Aspergillus nidulans* sequences were downloaded from the Whitehead Institute, Center for Genome Research at http://www-genome.wi.mit.edu/cgi-bin/annotation/aspergillus/download_license.cgi. *Plasmodium falciparum* sequences were from PlasmoDB at http://www.plasmodb.org.

Trypanosoma cruzi and *Entamoeba histolytica* were downloaded with permission from TIGR.

Eimeria tenella http://www.sanger.ac.uk/Projects/E_tenella/,

Theileria annulata http://www.sanger.ac.uk/Projects/T_annulata/,

Leishmania major http://www.sanger.ac.uk/Projects/L_major/ were from the Sanger Centre.

Other sources were *Chlamydomonas reinhardtii* http://genome.jgi-psf.org/chlre1/chlre1.home.html,

Giardia lamblia http://jbpc.mbl.edu/Giardia-HTML/index2.html

and *Encephalitozoon cuniculi*

ftp://ftp.ncbi.nih.gov/genomes/Encephalitozoon_cuniculi/.

Identification of SRP RNA sequences and prediction of RNA secondary structure
SRP RNA genes were predicted as described previously [14] by applying a combination of pattern searches using rnabob http://www.genetics.wustl.edu/eddy/software/#rnabob and COVE [15]. The rnabob searches made use of descriptors based on consensus features of the helix 8 region of the RNA. The search pattern was typically (XX) YYAGR (NNN) GRRA (N'N'N') AGCAR (X'X') or minor variations of it, where X pairs with X' and N with N'. RNA secondary structure predictions were carried out by COVE and MFOLD [16] and were complemented by analysis of compensatory base changes [36]. Saccharomyces RNAs were first aligned with CLUSTALW [37] using a gap extension penalty of 0, and this alignment was manually edited to be consistent with observations from MFOLD predictions to assure that bases involved in the formation of helices were properly aligned.

Identification and analysis of SRP protein sequences
The program 'sixpack' of the EMBOSS package [38] was used to obtain all possible translation products of genome sequences. PROFILESEARCH was part of the GCG package (Wisconsin package version 10.2, Genetics Computer Group (GCG), Madison, Wisc.). Hmmer package, developed by S. Eddy, was obtained from http://hmmer.wustl.edu/. Genscan was downloaded from http://genes.mit.edu/GENSCANinfo.html and GlimmerM from http://www.tigr.org/software/glimmerm/. Protein secondary structure was predicted using PSIPRED [33,34] and SAM-T02 [35].

All predicted RNA and protein sequences, secondary structures and multiple alignments are shown in a web supplement at http://bio.lundberg.gu.se/srp03/.

Authors' contributions
MAR and TS carried out bioinformatics analyses. TS and CZ conceived of the study and drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

References
1. Keenan RJ, Freymann DM, Stroud RM, Walter P: **The signal recognition particle.** *Annu Rev Biochem* 2001, **70:**755-775.

2.   Bui N, Strub K: **New insights into signal recognition and elongation arrest activities of the signal recognition particle.** *Biol Chem* 1999, **380(2):**135-145.
3.   Mason N, Ciufo LF, Brown JD: **Elongation arrest is a physiologically important function of signal recognition particle.** *Embo J* 2000, **19(15):**4164-4174.
4.   Hainzl T, Huang S, Sauer-Eriksson AE: **Structure of the SRP19 RNA complex and implications for signal recognition particle assembly.** *Nature* 2002, **417(6890):**767-771.
5.   Weichenrieder O, Wild K, Strub K, Cusack S: **Structure and assembly of the Alu domain of the mammalian signal recognition particle.** *Nature* 2000, **408(6890):**167-173.
6.   Oubridge C, Kuglstatter A, Jovine L, Nagai K: **Crystal structure of SRP19 in complex with the S domain of SRP RNA and its implication for the assembly of the signal recognition particle.** *Mol Cell* 2002, **9(6):**1251-1261.
7.   Kuglstatter A, Oubridge C, Nagai K: **Induced structural changes of 7SL RNA during the assembly of human signal recognition particle.** *Nat Struct Biol* 2002, **9(10):**740-744.
8.   Rosenblad MA, Gorodkin J, Knudsen B, Zwieb C, Samuelsson T: **SRPDB: Signal Recognition Particle Database.** *Nucleic Acids Res* 2003, **31(1):**363-364.
9.   Zwieb C, Eichler J: **Getting on target: The archaeal signal recognition particle.** *Archaea* 2002, **1:**27-34.
10.   Batey RT, Rambo RP, Lucast L, Rha B, Doudna JA: **Crystal structure of the ribonucleoprotein core of the signal recognition particle.** *Science* 2000, **287(5456):**1232-1239.
11.   Brown JD, Hann BC, Medzihradszky KF, Niwa M, Burlingame AL, Walter P: **Subunits of the Saccharomyces cerevisiae signal recognition particle required for its functional expression.** *Embo J* 1994, **13(18):**4390-4400.
12.   Felici F, Cesareni G, Hughes JM: **The most abundant small cytoplasmic RNA of Saccharomyces cerevisiae has an important function required for normal cell growth.** *Mol Cell Biol* 1989, **9(8):**3260-3268.
13.   Katinka MD, Duprat S, Cornillot E, Metenier G, Thomarat F, Prensier G, Barbe V, Peyretaillade E, Brottier P, Wincker P *et al.*: **Genome sequence and gene compaction of the eukaryote parasite Encephalitozoon cuniculi.** *Nature* 2001, **414(6862):**450-453.
14.   Regalia M, Rosenblad MA, Samuelsson T: **Prediction of signal recognition particle RNA genes.** *Nucleic Acids Res* 2002, **30(15):**3368-3377.
15.   Eddy SR, Durbin R: **RNA sequence analysis using covariance models.** *Nucleic Acids Res* 1994, **22(11):**2079-2088.
16.   Zuker M: **On finding all suboptimal foldings of an RNA molecule.** *Science* 1989, **244(4900):**48-52.
17.   Marshallsay C, Prehn S, Zwieb C: **cDNA cloning of the wheat germ SRP 7S RNAs.** *Nucleic Acids Res* 1989, **17(4):**1771.
18.   Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF: **A kingdom-level phylogeny of eukaryotes based on combined protein data.** *Science* 2000, **290(5493):**972-977.
19.   Frantz C, Ebel C, Paulus F, Imbault P: **Characterization of trans-splicing in Euglenoids.** *Curr Genet* 2000, **37(6):**349-355.
20.   Ebel C, Frantz C, Paulus F, Imbault P: **Trans-splicing and cis-splicing in the colourless Euglenoid, Entosiphon sulcatum.** *Curr Genet* 1999, **35(5):**542-550.
21.   Bell S, Nelson RG, Barry JD: **tRNAs of Trypanosoma brucei. Unusual gene organisation and mitochondrial importation.** *J Biol Chem* 1991, **266(27):**18313-18317.
22.   Naakar V, Dare AO, Hong D, Ullu E, Tscudi C: **Upstream tRNA genes are essential for expression of small nuclear and cytoplasmic RNA genes in trypanosomes.** 1994, **14(10):**6736-6742.
23.   Brennwald PJ, Siegel V, Walter P, Wise JA: **Sequence and structure of Tetrahymena SRP RNA.** *Nucleic Acids Res* 1991, **19(8):**1942.
24.   Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S *et al.*: **Genome sequence of the human malaria parasite Plasmodium falciparum.** *Nature* 2002, **419(6906):**498-511.
25.   Strub K, Fornallaz M, Bui N: **The Alu domain homolog of the yeast signal recognition particle consists of an Srp14p homodimer and a yeast-specific RNA structure.** *Rna* 1999, **5(10):**1333-1347.
26.   Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25(17):**3389-3402.
27.   Pearson WR: **Flexible sequence similarity searching with the FASTA3 program package.** *Methods Mol Biol* 2000, **132:**185-219.
28.   Burge C, Karlin S: **Prediction of complete gene structures in human genomic DNA.** *J Mol Biol* 1997, **268(1):**78-94.
29.   Salzberg SL, Pertea M, Delcher AL, Gardner MJ, Tettelin H: **Interpolated Markov models for eukaryotic gene finding.** *Genomics* 1999, **59(1):**24-31.
30.   Beja O, Ullu E, Michaeli S: **Identification of a tRNA-like molecule that copurifies with the 7SL RNA of Trypanosoma brucei.** *Mol Biochem Parasitol* 1993, **57(2):**223-229.
31.   Liu L, Ben-Shlomo H, Xu YX, Stern MZ, Goncharov I, Zhang Y, Michaeli S: **The Trypanosomatid Signal Recognition Particle Consists of Two RNA Molecules, a 7SL RNA Homologue and a Novel tRNA-like Molecule.** *J Biol Chem* 2003, **278(20):**18271-18280.
32.   Birse DE, Kapp U, Strub K, Cusack S, Aberg A: **The crystal structure of the signal recognition particle Alu RNA binding heterodimer, SRP9/14.** *Embo J* 1997, **16(13):**3757-3766.
33.   McGuffin LJ, Bryson K, Jones DT: **The PSIPRED protein structure prediction server.** *Bioinformatics* 2000, **16(4):**404-405.
34.   Jones DT: **Protein secondary structure prediction based on position-specific scoring matrices.** *J Mol Biol* 1999, **292(2):**195-202.
35.   Karplus K, Karchin R, Barrett C, Tu S, Cline M, Diekhans M, Grate L, Casper J, Hughey R: **What is the value added by human intervention in protein structure prediction?** *Proteins* 2001, **Suppl(5):**86-91.
36.   Larsen N, Zwieb C: **SRP-RNA sequence alignment and secondary structure.** *Nucleic Acids Res* 1991, **19(2):**209-215.
37.   Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22(22):**4673-4680.
38.   Olson SA: **EMBOSS opens up sequence analysis. European Molecular Biology Open Software Suite.** *Brief Bioinform* 2002, **3(1):**87-91.