Inventory and analysis of the protein subunits of the ribonucleases P and MRP provides further evidence of homology between the yeast and human enzymes

Magnus Alm Rosenblad[1], Marcela Dávila López , Paul Piccinelli and Tore Samuelsson*

Department of Medical Biochemistry, Institute of Biomedicine and [1]SWEGENE Bioinformatics, Goteborg University, Box 413, SE-405 30 Goteborg, Sweden

ABSTRACT

The RNases P and MRP are involved in tRNA and rRNA processing, respectively. Both enzymes in eukaryotes are composed of an RNA molecule and 9–12 protein subunits. Most of the protein subunits are shared between RNases P and MRP. We have here performed a computational analysis of the protein subunits in a broad range of eukaryotic organisms using profile-based searches and phylogenetic methods. A number of novel homologues were identified, giving rise to a more complete inventory of RNase P/MRP proteins. We present evidence of a relationship between fungal Pop8 and the protein subunit families Rpp14/Pop5 as well as between fungal Pop6 and metazoan Rpp25. These relationships further emphasize a structural and functional similarity between the yeast and human P/MRP complexes. We have also identified novel P and MRP RNAs and analysis of all available sequences revealed a K-turn motif in a large number of these RNAs. We suggest that this motif is a binding site for the Pop3/Rpp38 proteins and we discuss other structural features of the RNA subunit and possible relationships to the protein subunit repertoire.

INTRODUCTION

The ribonucleoprotein enzyme RNase P (ribonuclease P) is an endonuclease that processes tRNA precursors to generate the mature 5′ end. It is a nearly ubiquitous enzyme present in Archaea, Bacteria and Eukarya as well as in mitochondria and chloroplasts (1). A structurally and evolutionary related RNP, the RNase MRP (2,3), is found only in Eukarya. RNase MRP processes ribosomal RNA precursors at the A3 site allowing formation of the 5.8S pre-rRNA (4,5). RNase MRP is also known to have a role in the degradation of specific mRNAs involved in cell-cycle regulation (6) and it is

affected in the autosomal recessive disease cartilage hair hypoplasia (7).

The RNases P and MRP both have an RNA molecule and one or several protein subunits (8). The RNA molecules of P and MRP are related with respect to sequence and structure (9,10). The bacterial RNase P has a single protein subunit, but archaeal RNase P and eukaryotic nuclear RNase P/MRP enzymes contain multiple protein subunits. In eukaryotes most of the protein subunits are shared between P and MRP (11,12).

The RNA molecule in the bacterial RNase P can function as a ribozyme *in vitro*, although the cleavage rate of pre-tRNA is enhanced 20-fold by the protein moiety (13). While some archaeal RNase P RNAs show enzymatic activity under high salt conditions (14), the catalytic activity of the eukaryotic RNA subunit of RNase P requires the presence of protein subunits (15).

At least nine protein subunits are part of the nuclear RNase P of *Saccharomyces cerevisiae*; Pop1, Pop3, Pop4, Pop5, Pop6, Pop7, Pop8, Rpr2 and Rpp1 (16). Many of these subunits seem to be present also in the RNase MRP, with the exception of Rpr2 (Rpp21) which is unique to RNase P (11). MRP also contains Snm1 (17) and Rmp1 (18). Human nuclear RNase P and MRP appears to contain at least 10 protein subunits, Rpp14, Rpp20, Rpp21, Rpp25, Rpp29, Rpp30, Rpp38, Rpp40, hPop1 and hPop5 (19,20), although there is recent evidence that not all of these subunits are shared between P and MRP (21). At least six of the P/MRP subunits appear to be homologous to the subunits identified in *S.cerevisiae* (22). Comparative studies show that archaeal RNase P has at least four protein subunits homologous to eukaryotic RNase P/MRP proteins (23,24).

Models for the protein–protein and RNA–protein interactions in eukaryal RNases P and MRP have been proposed for human and yeast (16,19,20). Many of these interactions have also been found in Archaea (25–27). In the human RNase P the RNA molecule has been shown to interact with Rpp29, Rpp30, Rpp21 and Rpp38 (28). The RNA molecule in the human RNase MRP has been shown to interact with the protein subunits Pop1, Rpp29, Rpp20, Rpp25 and

---

*To whom correspondence should be addressed. Tel: +46 31 773 34 68; Fax: +46 31 41 61 08; Email: Tore.Samuelsson@medkem.gu.se

Rpp38 (20,29) and for the yeast MRP there is evidence that RNA interacts with the protein subunits Pop1 and Pop4 (16).

We have recently carried out an inventory of eukaryotic P and MRP RNAs and reported more than 100 novel sequences (10). Analysis of these sequences provided further evidence of a structural similarity between the two RNAs (10,30). The similarity between P and MRP RNA should be reflected in the set of protein subunits that are part of the RNP complexes. In order to better understand the relationship between protein and RNA subunits, RNA–protein interactions and evolution of the protein subunits in general we have systematically analyzed gene and protein sequences related to the RNase P and MRP protein subunits in all eukaryotic species where genome and protein sequences are available. Using profile-based searches we have identified several homologues that were not previously reported (24). Through a phylogenetic analysis of the protein sequences we were able to improve on classification, clarify evolutionary relationships and imply novel protein family relationships.

MATERIALS AND METHODS

Genome and protein sequences

The majority of RNase P/MRP protein sequences were retrieved from NCBI (ftp.ncbi.nih.gov/blast/db/), Swiss-Prot (http://www.expasy.ch/sprot/) and UniProt (http://www.expasy.uniprot.org/). Genomic sequences were from NCBI (ftp.ncbi.nih.gov/genomes/), EMBL (www.ebi.ac.uk), ENSEMBL (www.ensembl.org) and TraceDB (ftp.ncbi.nlm.nih.gov/pub/TraceDB). In addition the following organism specific databases were used; the *Saccharomyces* Genome Database (http://www.yeastgenome.org/), PlasmoDB (http://plasmodb.org/), CryptoDB (http://cryptodb.org/), Berkeley *Drosophila* Genome Project (http://www.fruitfly.org/), DOE Joint Genome Institute (http://www.jgi.doe.gov/), Xenbase (http://www.xenbase.org/) and WormBase (http://www.wormbase.org/). Genomic and protein sequences from *Trypanosoma cruzi*, *Entamoeba histolytica*, *Theileria parva*, *Toxoplasma gondii* and *Trichomonas vaginalis* were downloaded with permission from TIGR. The *Leishmania* sequences were from www.sanger.ac.uk; *Chlamydomonas reinhardtii* from genome.jgi-psf.org/chlre2/chlre2.home.html and *Giardia lamblia* from jbpc.mbl.edu/Giardia-HTML/index2.html. We also used fungal protein sequences not present in NCBI Genbank from the species *Clavispora lusitania*, *Pichia guilliermondi*, *Pichia stipitis*, *Kluyveromyces waltii*, *Coprinus cinereus*, *Phanerochaete chrysosporium*, and *Laccaria bicolor* as described in http://bio.lundberg.gu.se/rpp06.

Structural information and domain architectures of RNase P/MRP proteins were collected from PDB (http://www.rcsb.org/) and Pfam (http://www.sanger.ac.uk/Software/Pfam/), respectively.

Identification of protein homologues

In order to identify as many homologues as possible to previously known proteins of RNases P and MRP we made use of PSI-BLAST (Position Specific Iterative BLAST) searches (31,32). In this step all known protein subunits of RNases P and MRP were used. The proteins initially used as query

sequences are listed in Table 2 of the web supplement (http://bio.lundberg.gu.se/rpp06). An *E*-value of 0.001 was used as threshold for inclusion in PSI-BLAST iterations. In many cases multiple PSI-BLAST searches with different query sequences were carried out in order to identify as many homologues as possible belonging to a certain protein family. The database searched was the NCBI Genbank set of proteins (33), but some proteins were absent in this set and were retrieved from individual genome projects or identified from TBLASTN searches of genome sequences. Whenever relevant, these novel sequences were included in the set of sequences used as database in the PSI-BLAST search. Additional homologues were in some instances identified with Pfam models (26) and/or using models produced with the HMMER package [(34), http://hmmer.wustl.edu/].

Phylogenetic analysis

All the proteins considered to be significant homologues based on *E*-values were retrieved and multiple alignments were created using ClustalW 1.83 (35) or TCOFFEE (36). Gap columns were removed from the multiple alignments using GapStreeze (http://hiv-web.lanl.gov/content/hiv-db/GAPSTREEZE/gap.html). Phylogenetic analysis was carried out using programs of the PHYLIP package (37). For each ClustalW or TCOFFEE alignment we used parsimony (PROTPARS), maximum likelihood (PROML) and neighbour-joining (NEIGHBOR). The neighbour-joining was based on a matrix created by PROTDIST. SEQBOOT was used to generate for each of the three methods 500 bootstrapped datasets. CONSENSE was then used to compute a consensus tree by the majority-rule consensus tree method for the three different procedures. Figures 3–5 show consensus trees based on all 1500 trees and where the distances were obtained by PROML (maximum likelihood) using the consensus tree as input tree.

Identification of P and MRP RNA genes

RNase P/MRP RNA genes were predicted in the same manner as described previously (10). In brief we either used HMM models of CR-I and CR-V with hmmpfam (a tool of the HMMER package) to search genomic sequences, or FASTA34 (38) or BLAST searches in order to identify closely related homologues. Sequences were checked for conserved primary sequence motifs and the ability to fold into a secondary structure typical for P RNA or MRP RNA. Secondary structure predictions were carried out by MFOLD (39).

RESULTS AND DISCUSSION

We have inventoried the protein subunits of RNases P and MRP by analysis of available protein and genome sequences from a broad range of eukaryotic organisms. In order to identify as many homologues as possible to previously known P/MRP proteins we made use of PSI-BLAST (31,32) as described under Materials and Methods. In addition to PSI-BLAST, programs of the HMMER package [(34), http://hmmer.wustl.edu/] were used together with Pfam models (26) or models created from multiple sequence alignments whenever existing Pfam models were not adequate.

Several novel homologues were identified in these searches. Sequences were retrieved and further analyzed by creating

multiple alignments that were subjected to phylogenetic analysis as described under Materials and Methods. The phylogenetic analysis was in many cases essential for correct classification of the proteins identified in the profile-based searches.

A summary of the phylogenetic distribution of subunits, including RNA, is presented in Figure 1. Also shown in the figure are homology relationships between subunits of Fungi and Metazoa inferred from our results as will be discussed in more detail further below. A complete table of identified proteins and RNAs, the actual sequences, multiple alignments and PSI-BLAST results are presented in a web supplement to this paper, http://bio.lundberg.gu.se/rpp06/. The individual protein subunits and protein families will be discussed below, followed by a discussion of the protein composition of different phylogenetic groups.

Pop1, Pop4, Pop5 and Rpp1 are ubiquitous subunits

Pop1 is found in a large variety of eukaryotic organisms as already noted from previous studies (24) and here we identified homologues in all phylogenetic groups except green algae. All Pop1 homologues contain the Pfam domains Pop1 and Popld. These domains are conserved in sequence but all fungal sequences have an extensive insertion in the Pop1 domain as compared with other eukaryotes. (Figure 2).

A multiple alignment of full-length Pop1 sequences shows that there is a substantial length variation between species. Microsporidia and *Giardia* Pop1 are the smallest members of this family with for instance the *Nosema locustae* sequence of 375 residues as compared with the human sequence of 1024 residues. The Pop1 protein is known to interact with the P3 and P12 regions of the RNA and the small size of *Giardia* and Microsporidia Pop1 protein might be correlated to a small size in these regions of the RNA (10). Plasmodium species seem to have insertions between the Pop1 and Popld domains as compared with other eukaryotes. This is an interesting observation in the light of our finding that Plasmodium species have a P RNA with highly extended helices P3, eP8′, P12 and eP19 as compared with other organisms (10).

In PSI-BLAST searches aminomethyl transferase (AMT), an enzyme involved in the degradation of glycine, is identified as a relative to Pop1. However, the biological significance of this finding is not clear.

Pop4 is found in virtually all organisms examined, including the Microsporidia and *Giardia*, which is consistent with its important role in the RNases P and MRP and its binding to the catalytic core of the RNA. A multiple alignment of available Pop4 sequences shows that all eukaryotic Pop4 proteins has an N-terminal part (50–100 amino acids) that is missing in the Archaeal homologues, consistent with previous observations (23).

Also Pop5 is widely distributed and is found in Fungi, Viridiplantae, alveolates, *Dictyostelium*, *Giardia* and Archaea. Pop5 homologues are very similar in size but Pop5 proteins of fungi have an insertion ∼25 amino acids in length, close to the N-terminal end.

The *S.cerevisiae* Rpp1 is orthologous to human Rpp30 (40). The Rpp1/Rpp30 protein is found in all eukaryotic species as well as in Archaea. It has previously been identified in Ascomycota, Microsporidia, nematodes, insects, vertebrates,

plants and Plasmodium (24). Here we report homologues also from Basidiomycota, *Schistosoma*, *Entamoeba*, *Dictyostelium*, green and red algae, heterokonts, *Cryptosporidium*, *Eimeria* and *Theileria*.

In summary, the proteins Pop1, Pop4, Pop5 and Rpp1 are all widely distributed, consistent with essential roles of these subunits.

Orthologue and paralogue relationships within RNase P/MRP protein subunits

Pop3/Rpp38 family. Previously Pop3 homologues have been reported in *S.cerevisiae* and *C.albicans* (24,41). Here we identified homologues in a range of other Ascomycota as well in the Basidomycota *C.cinereus* and *Phanerochaete chrysosporium*, indicating that Pop3 is ubiquitous in fungi.

When Pop3 is used as query in profile-based searches Rpp38 homologues are identified as previously reported (42). Rpp38 is known to have the L7Ae/L30e domain, and other proteins with this domain are identified in PSI-BLAST searches. The relationship between proteins with the L7Ae domain, including Pop3, is illustrated by the phylogenetic analysis shown in Figure 3. The finding that Pop3 and Rpp38 are neighbours in this tree is consistent with an orthology relationship between Pop3 and Rpp38.

Rpp38 homologues are present in vertebrates, *Ciona*, sea urchin and *Branchiostoma* as well as in the plants *A.thaliana* and *O.sativa*. However, we failed to identify homologues in red and green algae, heterokonts, protozoa and the Microsporidia. Pop3 and Rpp38 will be further discussed below in the context of the K-turn motif that we have been identified in P and MRP RNAs.

Pop7/Rpp20/Rpp25 family. Human Rpp20 was proposed to be the homologue of yeast Pop7 (also named Rpp2) based on an alignment of the two protein sequences (43). In Hartmann *et al.* (24) these two proteins were not classified as homologues, but from the PSI-BLAST searches that we carried out it seems highly probable that Rpp20 is a Pop7 orthologue (data not shown). Furthermore, the Alba domain is present in both proteins and it was previously shown that in a phylogenetic analysis of different proteins containing the Alba domain Pop7 and Rpp20 are in a group distinct from other proteins with the Alba domain (44).

We identified Pop7 homologues in all Ascomycota species but not in Basidiomycota and Microsporidia. Furthermore, we found Pop7/Rpp20 homologues in all Metazoa and *Dictyostelium*, but not in other protozoa (Figure 1).

Also the Rpp25 protein is evolutionary related to Pop7/Rpp20, as previously shown (44). Here we identified novel homologues, including those of *Chlamydomonas*, heterokonts, and *Caenorhabditis*. Interestingly, the fishes *Danio rerio*, *Fugu rubripes* and *Tetraodon nigroviridis* seem to have two different proteins related to the human Rpp25 (Figure 5).

The human Rpp25 protein is evolutionarily related to a protein with unknown function referred to as C9orf23 (Chromosome 9 open reading frame 23 protein) (45). Here we have shown that Rpp25 and C9orf23 orthologues exist in all vertebrates, including fishes (Figure 5). There is a protein in *Ciona intestinalis* and *Ciona savignyi* related to these proteins but the results from the phylogenetic analysis do not allow us to conclude whether the *Ciona* proteins are

Figure 1. Phylogenetic distribution of RNase P and MRP protein subunits and inferred homologies between fungal and metazoan proteins. Boxes with shaded background represent organisms where a protein homologue was identified with profile-based searches such as PSI-BLAST. 'Sa-core' are organisms closely related to *S.cerevisiae*; 'Sa-oth' are other Saccharomycotina except *Yarrowia*; 'Sc' is *Schizosaccharomyces* and 'Pe' is Pezizomycotina. P and MRP RNA sequences new to this publication are indicated. Homologies between fungal and metazoan proteins are suggested by profile-based searches as described in the text and they are indicated by arrows on top of the table.
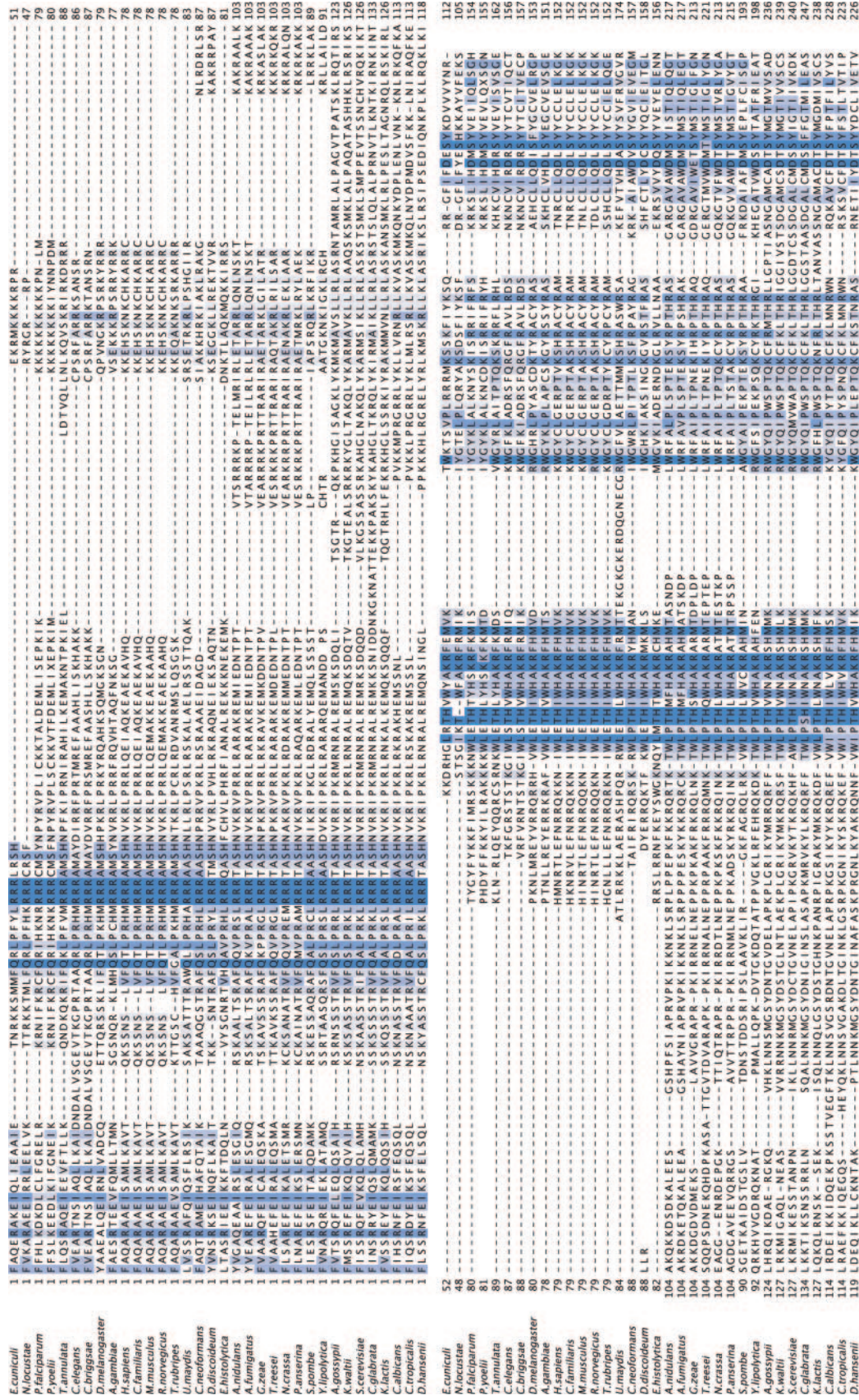
Figure 2. Alignment of Pop1 domain showing insertions in fungal sequences. The Pop1 domain of Pop1 protein sequences from Metazoa, Fungi and protozoa were aligned using ClustalW (36). Alignment is visualized with Jalview (59) and conserved residues are highlighted.
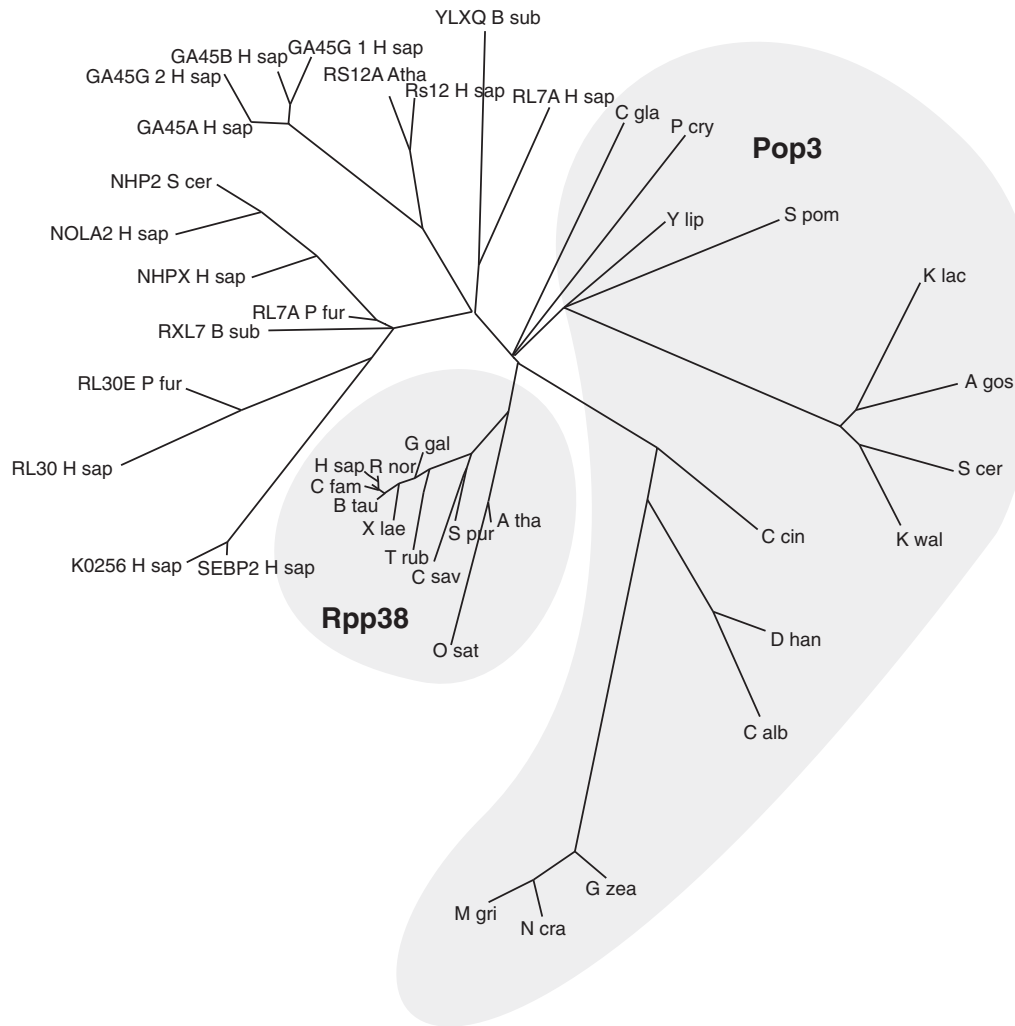
Figure 3. Phylogenetic tree of L7Ae family of proteins. Sequences of Rpp38 and Pop3 as well as other proteins with the L7Ae domain were subjected to phylogenetic analysis and a consensus tree was derived from neighbour joining, parsimony and maximum likelihood methods as described under Materials and Methods. The proximity of Rpp38 and Pop3 is consistent with an orthology relationship between these two proteins. Organisms shown for Rpp38/Pop3 are *A.gossypii, Arabidopsis thaliana, Bos taurus, Candida albicans, C.cinereus, Canis familiaris, C.glabrata, Ciona savignyi, D.hansenii, F.rubripes, Gallus gallus, G.zeae, Homo sapiens, K.lactis, K.waltii, M.grisea, N.crassa, Oryza sativa, P.chrysosporium, Rattus norvegicus, S.cerevisiae, Saccharomyces pombe, Strongylocentrotus purpuratus, Xenopus laevis and Yarrowia lipolytica*. Proteins that are not Rpp38 or Pop3 are Swiss-Prot entries with the L7Ae domain. SEBP2_HUMAN is a SECIS-binding protein and YLXQ_BACSU, RL30E_PYRFU, RL7A_HUMAN, RL30_HUMAN, RXL7_BACSU, RL7A_PYRFU, RS12A_ARATH and RS12_HUMAN are ribosomal proteins. NHPX_HUMAN is a U4 RNP and NOLA2_HUMAN and NHP2_YEAST are H/ACA RNPs. GA45A_HUMAN, GA45B_HUMAN, Q5TCA7_HUMAN and GA45G_HUMAN are growth arrest and DNA-damage-inducible proteins and K0256_HUMAN is a protein with unknown function.

most closely related to Rpp25 or C9orf23. As in *Ciona*, there seems to be a single Rpp25-like protein in insects, nematodes, plants and many protozoa. It would therefore seem that gene duplication took place at a point of evolution close to the development of Deuterostomia, giving rise to the Rpp25 and C9orf23 protein. On the other hand Plasmodium species, unlike other protozoos, also have two different Rpp25-like proteins. Possibly, gene duplications took place more than once in the evolution of the Rpp25 family of proteins.

Pop5/Rpp14 family of proteins and relationship to Pop8. It has previously been noted that Pop5 is homologous to Rpp14 (46). During the course of this work we observed that the current Pfam model named Rpp14 is not adequate as the sequences present in the seed alignment are mostly

Pop5 sequences. This is probably the reason why several archaeal proteins, such as Genbank:BAD85956, are incorrectly annotated as Rpp14. For this work, a phylogenetic analysis helped to classify Rpp14 and Pop5 homologues correctly (Figure 4). Rpp14 is found only in Metazoa, but Pop5 is more widely distributed as referred to above. In *D.melanogaster* two homologues of Rpp14 seem to be present and the coding sequences of these two proteins are located next to each other on chromosome chr3R. The two proteins are also found in other *Drosophila* species.

We also found evidence that Pop8 is related to the Pop5/Rpp14 proteins. Pop8 was previously identified in *S.cerevisiae* and *S.pombe* (24). We now identified with PSI-BLAST homologues in *Kluyveromyces lactis, Candida glabrata, P.stipitis, Debaromyces hansenii, Ashbya gossypii*
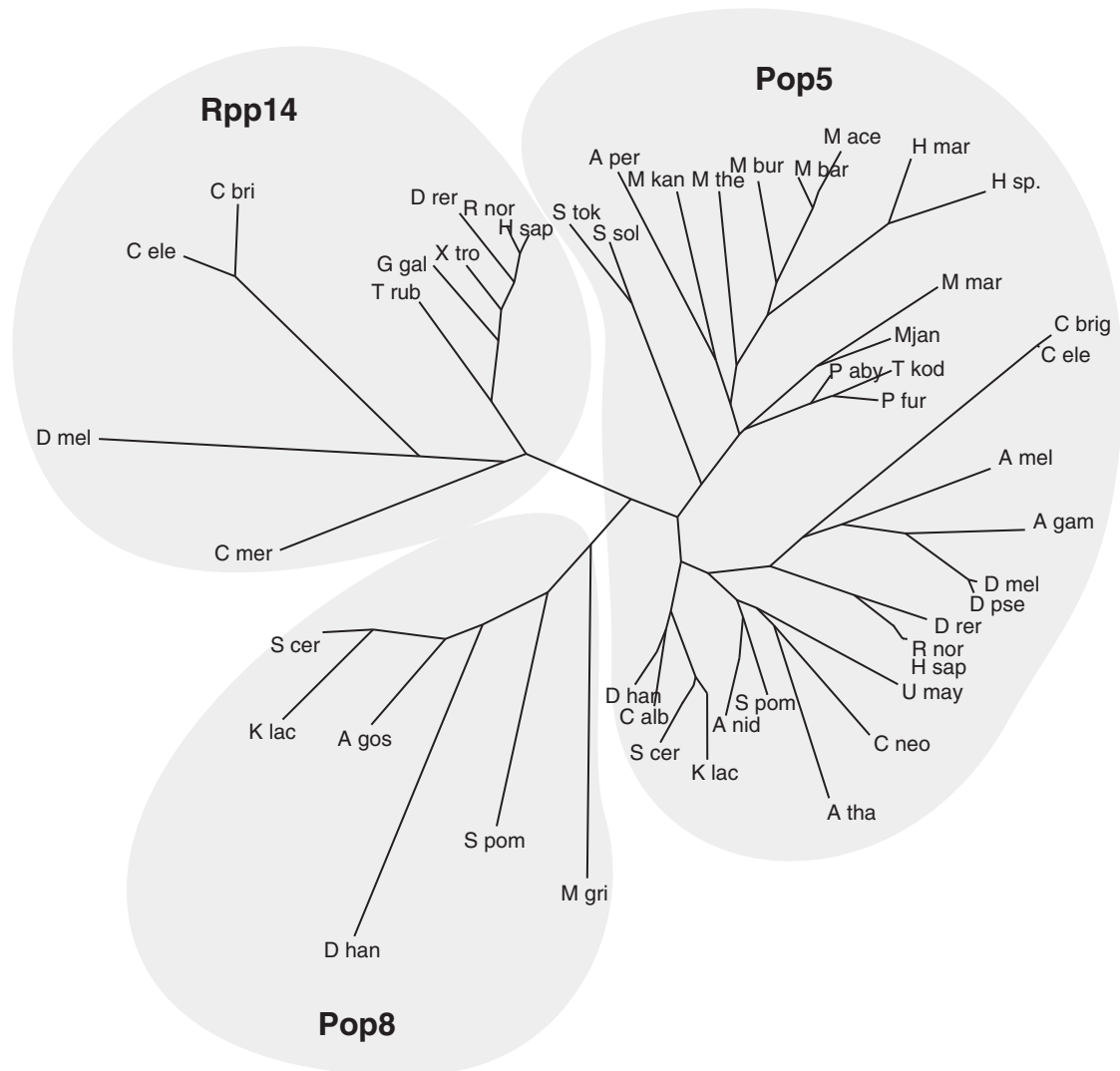
Figure 4. Classification of Rpp14/Pop5/Pop8 proteins. Rpp14, Pop5 and Pop8 protein sequences were subjected to phylogenetic analysis and a consensus tree was derived from neighbour joining, parsimony and maximum likelihood methods as described under Materials and Methods. Organisms represented in tree are *Anopheles gambiae*, *A.gossypii*, *Apis mellifera*, *A.nidulans*, *Aeropyrum pernix*, *A.thaliana*, *C.albicans*, *Caenorhabditis briggsae*, *C.elegans*, *C.merolae*, *Cryptococcus neoformans*, *D.hansenii*, *Drosophila melanogaster*, *Drosophila pseudoobscura*, *D.rerio*, *F.rubripes*, *G.gallus*, *Haloarcula marismortui*, *H.sapiens*, *Halobacterium* sp., *K.lactis*, *Methanosarcina acetivorans*, *Methanosarcina barkeri*, *Methanococcoides burtonii*, *M.grisea*, *Methanocaldococcus jannaschii*, *Methanopyrus kandleri*, *Methanococcus maripaludis*, *Methanothermobacter thermautotrophicus*, *Pyrococcus abyssi*, *Pyrococcus furiosus*, *R.norvegicus*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Sulfolobus solfataricus*, *Sulfolobus tokodaii*, *Thermococcus kodakaraensis*, *Ustilago maydis* and *Xenopus tropicalis*.

and in the Pezizomycotina *Aspergillus nidulans, Neurospora crassa, Giberella zeae* and *Magnaporthe grisea* (Figures 1 and 4). When some of these Pop8 homologues were used as query sequences Rpp14 proteins were identified. For instance, when *N.crassa* Pop8 was used as a query sequence convergence was reached after seven iterations and the sequences above threshold (E-values 10^{-37}–10^{-21}) were Pop8 sequences as well as a large number of Rpp14 homologues (for details see web supplement, http://bio.lundberg. gu.se/rpp06). In addition, when some of the Rpp14 vertebrate sequences are used as queries, fungal Pop8 proteins are identified and when using human Pop5 as query, *S.pombe* and *C.glabrata* Pop8 are identified. Hmmer searches were also carried out with profiles based on Pop8, Pop5 and Rpp14, respectively. When a profile based on Pop8 was used, profiles

of Pop5 or Rpp14 were second best, among all profiles that are part of Pfam A. Taken together, these results strongly suggest an evolutionary relationship between Pop8 and Pop5/Rpp14. An interesting possibility is that Pop8 is the fungal orthologue to Rpp14.

Fungal Pop6 and relationship to the metazoan Rpp25. Pop6 was previously identified in *S.cerevisiae* (11). Here we found homologues in a number of additional species of Saccharomycotina. No Pop6 homologues were identified in Pezizomycotina and Basidiomycota. However, in some of the PSI-BLAST searches metazoan Rpp25 homologues were identified. For instance, when *P.stipitis* Pop6 was used as query sequence convergence was reached after seven iterations and a number of Metazoan, plant and *Plasmodium*
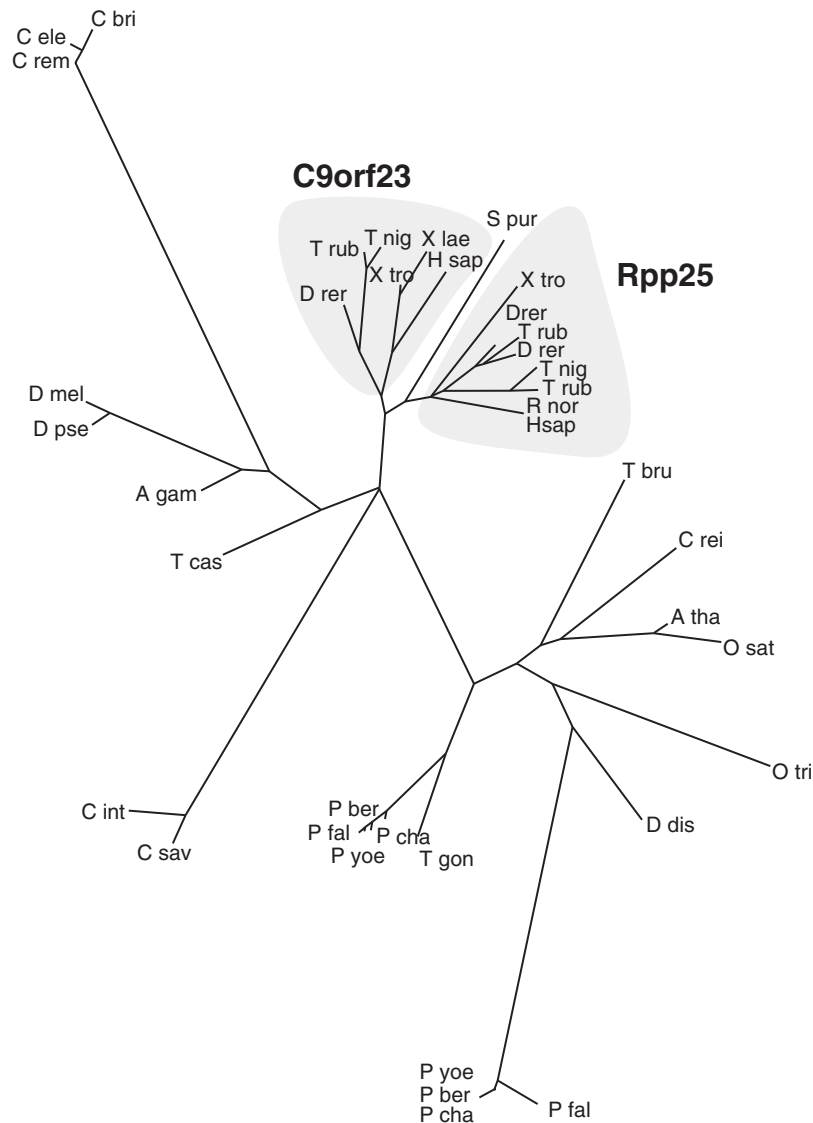
Figure 5. Relationship between Rpp25 and C9orf23 homologues. Protein sequences related to human Rpp25 and C9orf23 were subjected to phylogenetic analysis and a consensus tree was derived from neighbour joining, parsimony and maximum likelihood methods as described under Materials and Methods. Rpp25 and C9orf23 homologues are found in all vertebrates. Fishes appear to have two different Rpp25-related proteins. Organisms represented in tree are *A.gambiae, A.thaliana, C.briggsae, C.elegans, C.intestinalis, Chlamydomonas reinhardtii, Caenorhabtidis remanei, C.savignyi, Dictyostelium discoideum, D.melanogaster, D.pseudoobscura, D.rerio, F.rubripes, H.sapiens, O.sativa, Oxytricha trifallax, Plasmodium berghei, Plasmodium chabaudi, Plasmodium falciparum, Plasmodium yoelii, R.norvegicus, S.purpuratus, Trypanosoma brucei, Tribolium castaneum, Toxoplasma gondii, T.nigroviridis, X.laevis,* and *X.tropicalis.*

Rpp25 homologues were identified at that stage above threshold (*E*-values 10^{-36}–10^{-20}) (for details see web supplement, http://bio.lundberg.gu.se/rpp06). These results strongly suggest a relationship between Rpp25 and Pop6.

It is interesting to note that a relationship between Pop6 and Rpp25 is consistent with available protein–RNA and protein–protein interaction data. Thus, in the human MRP Rpp25 and Rpp20 interact with each other and Rpp25 binds to Rpp29 (Pop4) (20). In the yeast RNase P, Pop6 and Pop7 (the Rpp20 homologue) interact with each other and they are both interaction partners of Pop4 (16).

Rpp21/Rpr2/ Snm1 family. Rpp21/ Rpr2 is a universal RNase P subunit also present in Archaea (47,48). Homologues have previously been described in vertebrates, Plasmodium, Ascomycota, Microsporidia and insects (24). Here we found

the protein also in the red algae *Cyanidioshizon merolae,* Basidiomycota, sea urchin, *Giardia* and alveolates other than *Plasmodium.* However, we have failed to identify a Rpp21 (Rpr2) in *Entamoeba, Dictyostelium,* the Viridiplantae group, heterokonts and Ciliophora. Rpp21/Rpr2 from Pezizomycotina and Basidiomycota have an insertion of ∼30 amino acids as compared with other organisms.

The *S.cerevisiae* Snm1 is homologous to Rpr2 (18). Here we identified homologues in other Ascomycota, but there was no evidence of homologues in Basidiomycota or Microsporidia.

Other protein subunits

In addition to the subunits discussed above we identified Rpp40 homologues in Pezizomycotina, *Yarrowia,* Basidiomycota, Apicomplexa as well as green and red algae,

showing that Rpp40 has a much broader phylogenetic distribution than was previously known (49). We also showed that the *S.cerevisiae* Rmp1 (18) has homologues in Ascomycota and Microsporidia but not in Basidiomycota. No obvious homologues outside the fungal groups could be found and we could not find support for a *A.thaliana* homologue suggested in Salinas *et al*. (18).

Analysis of P and MRP RNA sequences reveal K-turn and K-loop motifs in P and MRP RNAs

We have recently presented an inventory of eukaryotic P and MRP RNAs with several novel homologues. We have now identified more than 40 additional RNAs (http://bio. lundberg.gu.se/rpp06/). The collection of identified RNAs now include MRP RNAs in *Caenorhabtidis elegans,* Entamoebidae and Glomeromycota. With these novel sequences we have identified P or MRP RNAs in all major phylogenetic groups. The only exceptions are Euglenozoa where neither P nor MRP RNA are identified and the plant and heterokonta groups where no P RNA is found.

We have further analyzed all available P and MRP sequences and structures and have identified a K-turn motif, a motif previously shown to be present in 23S rRNA (50), U4 snRNA (51) and snoRNAs (52). The motif has also been discussed in the context of human MRP RNA (50) but here we have found it to be present in both P and MRP RNA from a large number of organisms (for details see http://bio. lundberg.gu.se/rpp06) and they are always in the helix P12. Examples of vertebrate, insect, nematode and protozoan P and MRP RNAs with K-turn motifs are shown in Figure 6. These observations provide further evidence of a structural similarity between P and MRP RNA (10,30). We also noted that K-turn motifs were present in a few archaeal P RNAs, like that of *Thermoplasma volcanium* (Figure 6).

In addition, most fungal MRP RNAs, with the exception of Saccharomycotina, have a K-turn like motif, although it is different in the respect that the G–A base pairs are connected directly with a loop instead of a longer helical region. This type of structure has been referred to as a K-loop (53). Two alternative structures that are both compatible with all available sequences of the Pezizomycotina and Basidiomycota group of MRP RNAs are shown in Figure 6. For instance, such motifs are present in the Pezizomycotina *N.crassa*, *Potentilla anserina*, *G.zeae* and *Coccidioides immitis*, and in the Basiodiomycota *L.bicolor*, and *P.chrysosporium*. One of the structures, with only three bases in the loop, is very similar to the structure shown for P and MRP RNAs from Metazoa and Protozoa.

Does Rpp38 interact with K-turn motif of RNase P and MRP RNA?

Kink-turns were first discovered by structural studies of the ribosome (50). In the ribosome K-turns interact with a number of different ribosomal proteins, including the archaeal L7Ae. In L7Ae there are residues of asparagine and glutamic acid that are known to interact directly with the G–A pair in the K-turn. It is interesting to note that these two residues are conserved in all Metazoan Rpp38, but among fungi, they are strictly conserved only in *S.cerevisiae*, *K.lactis* and *P.chrysosporium*.

The archaeal L7Ae protein apparently has multiple roles. In addition to being a ribosomal protein, it binds to archael snoRNAs, both to K-turn (52) and K-loop (53) structures. There is evidence that L7Ae is a subunit of the archaeal RNase P (54,55). It is conceivable that L7Ae, as well as other members of this protein family, such as Rpp38/Pop3, bind to K-turns of P and MRP RNAs. Consistent with this hypothesis is the evidence from crosslinking experiments (29) as well as GST pulldown experiments (20) that Rpp38 interacts with a region of the RNA containing the putative K-turn. Although such a hypothesis is very attractive it is still not clear why not all P and MRP RNAs have an identifiable K-turn while they do have an Rpp38/Pop3 homologue. At the same time Rpp38 homologues have not been identified in some organisms, like the alveolates, heterokonts and *Giardia*, but we have identified a K-turn in the P RNAs of *Giardia* and *Tetrahymena*, as well as in the MRP RNAs of *Tetrahymena*, *Toxoplasma* and *Phytophthora*.

Phylogenetic distribution of protein subunits

Our inventory of the protein and RNA subunits of RNases P and MRP allows us to make conclusions about the composition of the P and MRP complexes in different major phylogenetic groups (Figure 1) and as described below we are able to distinguish a number of distinct groups based on their protein composition.

Archaea. In virtually all archaea that were examined, we identified homologues to Pop4, Pop5, Rpp1, Rpr2 and the multifunctional L7Ae, consistent with previous analyses (24).

Metazoa and protozoa. The Metazoa form a rather homogenous group in terms of the protein composition. Characteristic of the vertebrates are the proteins Pop1, Rpp29 (Pop4), Pop5, Rpp20 (Pop7), Rpp30 (Rpp1), Rpp21 (Rpr2), Rpp14, Rpp25, Rpp38 and Rpp40. Nematodes and insects are similar to vertebrates, but we have not identified an Rpp38 homologue in these groups. The set of proteins in the alveolates is similar to that of Metazoa, with the exception that Rpp14, Rpp38 and possibly Rpp25 are missing.

Fungi. The subunits Pop1, Pop3, Pop4, Pop5, Pop6, Pop7, Pop8, Rpp30, Rpp21, Snm1 and Rmp1 are found in *S.cerevisiae* as well as in all other Saccharomycotina except *Yarrowia*. Genetic experiments have shown that all of these proteins are required for viability in *S.cerevisiae* (SGD database, www.yeastgenome.org). The protein composition in Ascomycota is very similar to that observed in the Metazoa. Pop6 and Pop8 are the only proteins that seem to be exclusive to Ascomycota, but on the other hand we have here suggested that Pop8 is homologous to Rpp14/Pop5 and that Pop6 is a Rpp25 homologue. The Basidiomycota genomes are not as fully sequenced as genomes of the other fungal groups but based on our analysis it seems that they have a smaller set of proteins than Ascomycota as we have failed to identify Pop6, Pop7, Pop8, Snm1 and Rmp1 homologues.

Microsporidia and Giardia, small genome organisms. The Microsporidia, like Basidiomycota, are characterized by a very small set of RNase P/MRP proteins. Thus, proteins identified in Microsporidia include Pop1, Pop4, Pop5, Rpp1, Rpr2 and Rmp1. This is consistent with the fact that
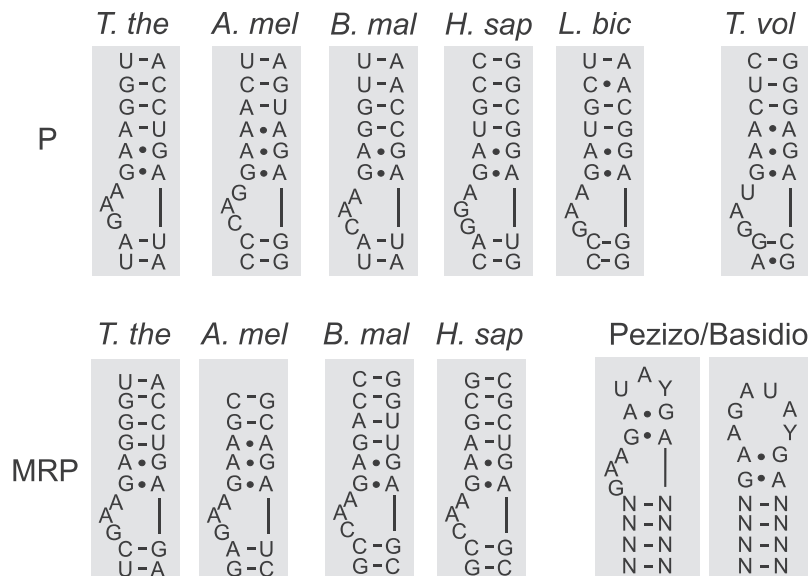
Figure 6. K-turn and K-loop motifs in P and MRP RNAs. K-turn motifs are part of helix P12 in both P and MRP RNAs. Organisms shown are *Tetrahymena thermophila*, *A.mellifera*, *Brugia malayi*, *H.sapiens*, *L.bicolor* and *Thermoplasma volcanium*. For Pezizomycotina and Basidiomycota two alternative structures are shown for the K-loop motif.

the Microsporidia have unusually small genomes (2.5 million bases) and may be considered minimal eukaryotes. They have most probably evolved from organisms related to Fungi but have lost genetic material in this process. As Microsporidia seem to have retained during their evolution only protein critical for function it seems likely that the RNase P/MRP protein subunits present in Microsporidia are particularly important for the function of the RNases P and MRP.

Giardia lamblia is another organism with a comparatively small genome, ~12 million bases. It is interesting to note that also in this organism we have identified only a small number of protein subunits.

Microsporidia have both P and MRP RNAs and a P RNA has been identified in *Giardia* (10,56). In these RNAs helix P3 is very small as well as the entire domain 2 (10). The difference in protein repertoire might be related to these differences. For instance, there is evidence that Rpp20 and Rpp25 interact with the P3 region (20,29), proteins that are absent in Microsporidia and *Giardia*. The characteristics of P3 may also be related to the structure of Pop1 as discussed above. Also in Archaea the helix P3 is very small and in this kingdom the Pop1, Rpp20 and Rpp25 proteins are all absent. We have also noted that a number of land plants have MRP RNAs with very small P3 helices (for details see web supplement), but there is no experimental information on the protein repertoire of these plants.

Red algae, green algae and plants. So far, an RNase P RNA has not been identified in plants. In our recent inventory of eukaryotic P and MRP RNAs, we failed to identify a P RNA in the plants *A.thaliana* and *O.sativa* as well as in the green algae *C.reinhardtii* and *Volvox carteri* and the red algae *C.merolae,* suggesting that RNase P is missing in these organisms. From this point of view it is interesting to compare the RNase P/MRP protein repertoire of the plant group to other phylogenetic groups. One major difference is

that Rpp21 (Rpr2) is missing in the plants and in the green algae. As Rpp21 is a subunit specific to RNase P and is not found in MRP, this observation is consistent with the idea that RNase P is absent in these organisms. There is also evidence that Rpp29 (Pop4) and Rpp14 are specific to RNase P and are not present in the human RNase MRP (21). If this applies to the plant group our failure to identify Rpp14 in plants is consistent with its absence in MRP. On the other hand, if an RNase P is missing in the plant group, the fact that an Rpp29 homologue is present seems to be in conflict with the suggestion that Rpp29 is specific to RNase P.

Euglenozoa. The Euglenozoa, with Leishmania and Trypanosoma, is the only phylogenetic group where we were not able to demonstrate a single RNase P or MRP protein subunit. The only exception is an Rpp25 homologue identified in *T.brucei* but this protein might have a function which is not related to RNase P/MRP. The absence of RNase P/MRP proteins in Euglenozoa is consistent with our failure to detect a P or MRP RNA subunit in these organisms (10). Therefore, it seems highly probable that an RNase P/MRP is missing, at least of the type found in other eukaryotes. Euglenozoa are known to be unusual in other aspects related to transcripts; they have unusual mechanisms of RNA editing (57) and very long polycistronic transcripts are produced (58). Therefore, they might be very different also with respect to tRNA and rRNA processing.

CONCLUSIONS

We have here presented a comprehensive collection of RNase P and MRP protein subunits, complementing our previous inventory of the RNA molecules of these enzymes (10). This collection of protein sequences allows us to make a number of conclusions with regard to orthologue and paralogue relationships. First, we have shown that gene duplications

took place within the group of Rpp25-related proteins and one such event gave rise to the C9orf23 protein early in the vertebrate lineage. Second, a number of observations are related to homology between fungal and metazoan P/MRP protein subunits. For instance, we have obtained evidence of homology between fungal Pop3 and metazoan Rpp38 as well as between fungal Pop7 and metazoan Rpp20, in support of previous observations (43,44). We also have evidence of relationships that were not identified previously. Thus, we suggest that the fungal Pop8 is related to the Rpp14/Pop5 protein family and that the fungal Pop6 is homologous to Rpp25. It is interesting to note that with these relationships all known fungal protein subunits that are shared between RNases P and MRP, without exception, have metazoan homologues. This implies that the fungal and metazoan RNase P/MRP complexes are structurally even more similar than previously thought.

We also present novel P and MRP RNA sequences, and thereby P and/or MRP RNAs have been identified in all phylogenetic groups. However, P RNA has not been identified in the plant and heterokonta groups and neither P nor MRP RNA is found in Euglenozoa, consistent with the absence of P/MRP protein subunits. Analysis of available P and MRP RNA sequences shows that a large number of them have a K-turn motif. This observation is further evidence of the close structural relationship between these two RNAs. The K-turn motif is probably a protein-binding site and probable candidates for binding to this site are the Pop3/Rpp38 proteins.

ACKNOWLEDGEMENTS

Conflict of interest statement. None declared.

REFERENCES

1. Frank,D.N. and Pace,N.R. (1998) Ribonuclease P: unity and diversity in a tRNA processing ribozyme. *Annu. Rev. Biochem.*, **67**, 153–180.
2. Kiss,T. and Filipowicz,W. (1992) Evidence against a mitochondrial location of the 7-2/MRP RNA in mammalian cells. *Cell*, **70**, 11–16.
3. Topper,J.N., Bennett,J.L. and Clayton,D.A. (1992) A role for RNAase MRP in mitochondrial RNA processing. *Cell*, **70**, 16–20.
4. Lygerou,Z., Allmang,C., Tollervey,D. and Seraphin,B. (1996) Accurate processing of a eukaryotic precursor ribosomal RNA by ribonuclease MRP *in vitro*. *Science*, **272**, 268–270.
5. Henry,Y., Wood,H., Morrissey,J.P., Petfalski,E., Kearsey,S. and Tollervey,D. (1994) The 5′ end of yeast 5.8S rRNA is generated by exonucleases from an upstream cleavage site. *Embo J*, **13**, 2452–2463.
6. Gill,T., Cai,T., Aulds,J., Wierzbicki,S. and Schmitt,M.E. (2004) RNase MRP cleaves the CLB2 mRNA to promote cell cycle progression: novel method of mRNA degradation. *Mol. Cell Biol.*, **24**, 945–953.
7. Ridanpaa,M., van Eenennaam,H., Pelin,K., Chadwick,R., Johnson,C., Yuan,B., vanVenrooij,W., Pruijn,G., Salmela,R., Rockas,S. *et al.* (2001) Mutations in the RNA component of RNase MRP cause a pleiotropic human disease, cartilage-hair hypoplasia. *Cell*, **104**, 195–203.
8. Walker,S.C. and Engelke,D.R. (2006) Ribonuclease P: the evolution of an ancient RNA enzyme. *Crit. Rev. Biochem. Mol. Biol.*, **41**, 77–102.
9. Forster,A.C. and Altman,S. (1990) Similar cage-shaped structures for the RNA components of all ribonuclease P and ribonuclease MRP enzymes. *Cell*, **62**, 407–409.
10. 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.
11. Chamberlain,J.R., Lee,Y., Lane,W.S. and Engelke,D.R. (1998) Purification and characterization of the nuclear RNase P holoenzyme complex reveals extensive subunit overlap with RNase MRP. *Genes Dev.*, **12**, 1678–1690.
12. Lee,B., Matera,A.G., Ward,D.C. and Craft,J. (1996) Association of RNase mitochondrial RNA processing enzyme with ribonuclease P in higher ordered structures in the nucleolus: a possible coordinate role in ribosome biogenesis. *Proc. Natl Acad. Sci. USA*, **93**, 11471–11476.
13. Guerrier-Takada,C., Gardiner,K., Marsh,T., Pace,N. and Altman,S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, **35**, 849–857.
14. Pannucci,J.A., Haas,E.S., Hall,T.A., Harris,J.K. and Brown,J.W. (1999) RNase P RNAs from some Archaea are catalytically active. *Proc. Natl Acad. Sci. USA*, **96**, 7803–7808.
15. Altman,S., Kirsebom,L. and Talbot,S. (1993) Recent studies of ribonuclease P. *Faseb J.*, **7**, 7–14.
16. Houser-Scott,F., Xiao,S., Millikin,C.E., Zengel,J.M., Lindahl,L. and Engelke,D.R. (2002) Interactions among the protein and RNA subunits of *Saccharomyces cerevisiae* nuclear RNase P. *Proc. Natl Acad. Sci. USA*, **99**, 2684–2689.
17. Schmitt,M.E. and Clayton,D.A. (1994) Characterization of a unique protein component of yeast RNase MRP: an RNA-binding protein with a zinc-cluster domain. *Genes Dev.*, **8**, 2617–2628.
18. Salinas,K., Wierzbicki,S., Zhou,L. and Schmitt,M.E. (2005) Characterization and purification of *Saccharomyces cerevisiae* RNase MRP reveals a new unique protein component. *J. Biol. Chem.*, **280**, 11352–11360.
19. Jiang,T. and Altman,S. (2001) Protein–protein interactions with subunits of human nuclear RNase P. *Proc. Natl Acad. Sci. USA*, **98**, 920–925.
20. Welting,T.J., van Venrooij,W.J. and Pruijn,G.J. (2004) Mutual interactions between subunits of the human RNase MRP ribonucleoprotein complex. *Nucleic Acids Res.*, **32**, 2138–2146.
21. Welting,T.J., Kikkert,B.J., Van Venrooij,W.J. and Pruijn,G.J. (2006) Differential association of protein subunits with the human RNase MRP and RNase P complexes. *RNA*, **12**, 1373–1382.
22. Xiao,S., Scott,F., Fierke,C.A. and Engelke,D.R. (2002) Eukaryotic ribonuclease P: a plurality of ribonucleoprotein enzymes. *Annu. Rev. Biochem.*, **71**, 165–189.
23. Hall,T.A. and Brown,J.W. (2002) Archaeal RNase P has multiple protein subunits homologous to eukaryotic nuclear RNase P proteins. *RNA*, **8**, 296–306.
24. Hartmann,E. and Hartmann,R.K. (2003) The enigma of ribonuclease P evolution. *Trends Genet.*, **19**, 561–569.
25. Kifusa,M., Fukuhara,H., Hayashi,T. and Kimura,M. (2005) Protein–protein interactions in the subunits of ribonuclease P in the hyperthermophilic archaeon Pyrococcus horikoshii OT3. *Biosci. Biotechnol. Biochem.*, **69**, 1209–1212.
26. Bateman,A., Coin,L., Durbin,R., Finn,R.D., Hollich,V., Griffiths-Jones,S., Khanna,A., Marshall,M., Moxon,S., Sonnhammer,E.L. *et al.* (2004) The Pfam protein families database. *Nucleic Acids Res.*, **32**, D138–D141.
27. Hall,T.A. and Brown,J.W. (2004) Interactions between RNase P protein subunits in archaea. *Archaea*, **1**, 247–254.
28. Jiang,T., Guerrier-Takada,C. and Altman,S. (2001) Protein–RNA interactions in the subunits of human nuclear RNase P. *RNA*, **7**, 937–941.
29. Pluk,H., van Eenennaam,H., Rutjes,S.A., Pruijn,G.J. and van Venrooij,W.J. (1999) RNA-protein interactions in the human RNase MRP ribonucleoprotein complex. *RNA*, **5**, 512–524.
30. Zhu,Y., Stribinskis,V., Ramos,K.S. and Li,Y. (2006) Sequence analysis of RNase MRP RNA reveals its origination from eukaryotic RNase P RNA. *RNA*, **12**, 699–706.
31. Altschul,S.F. and Koonin,E.V. (1998) Iterated profile searches with PSI-BLAST—a tool for discovery in protein databases. *Trends Biochem. Sci.*, **23**, 444–447.
32. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and

PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

33. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Wheeler,D.L. (2006) GenBank. *Nucleic Acids Res.*, **34**, D16–D20.

34. Eddy,S.R. (1998) Profile hidden Markov models. *Bioinformatics*, **14**, 755–763.

35. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.

36. Notredame,C., Higgins,D.G. and Heringa,J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.*, **302**, 205–217.

37. Felsenstein,J. (1996) Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol.*, **266**, 418–427.

38. Pearson,W.R. (2000) Flexible sequence similarity searching with the FASTA3 program package. *Methods Mol. Biol.*, **132**, 185–219.

39. Zuker,M. (1989) On finding all suboptimal foldings of an RNA molecule. *Science*, **244**, 48–52.

40. Stolc,V. and Altman,S. (1997) Rpp1, an essential protein subunit of nuclear RNase P required for processing of precursor tRNA and 35S precursor rRNA in *Saccharomyces cerevisiae*. *Genes Dev.*, **11**, 2414–2425.

41. Dichtl,B. and Tollervey,D. (1997) Pop3p is essential for the activity of the RNase MRP and RNase P ribonucleoproteins *in vivo*. *Embo J.*, **16**, 417–429.

42. Dlakic,M. (2005) 3D models of yeast RNase P/MRP proteins Rpp1p and Pop3p. *RNA*, **11**, 123–127.

43. Stolc,V., Katz,A. and Altman,S. (1998) Rpp2, an essential protein subunit of nuclear RNase P, is required for processing of precursor tRNAs and 35S precursor rRNA in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA*, **95**, 6716–6721.

44. Aravind,L., Iyer,L.M. and Anantharaman,V. (2003) The two faces of Alba: the evolutionary connection between proteins participating in chromatin structure and RNA metabolism. *Genome Biol.*, **4**, R64.

45. Guerrier-Takada,C., Eder,P.S., Gopalan,V. and Altman,S. (2002) Purification and characterization of Rpp25, an RNA-binding protein subunit of human ribonuclease P. *RNA*, **8**, 290–295.

46. Koonin,E.V., Wolf,Y.I. and Aravind,L. (2001) Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach. *Genome Res.*, **11**, 240–252.

47. Jarrous,N., Reiner,R., Wesolowski,D., Mann,H., Guerrier-Takada,C. and Altman,S. (2001) Function and subnuclear distribution of Rpp21, a protein subunit of the human ribonucleoprotein ribonuclease P. *RNA*, **7**, 1153–1164.

48. Kakuta,Y., Ishimatsu,I., Numata,T., Kimura,K., Yao,M., Tanaka,I. and Kimura,M. (2005) Crystal structure of a ribonuclease P protein Ph1601p from *Pyrococcus horikoshii* OT3: an Archaeal homologue of human nuclear ribonuclease P protein Rpp21(,). *Biochemistry*, **44**, 12086–12093.

49. Eder,P.S., Kekuda,R., Stolc,V. and Altman,S. (1997) Characterization of two scleroderma autoimmune antigens that copurify with human ribonuclease P. *Proc. Natl Acad. Sci. USA*, **94**, 1101–1106.

50. Klein,D.J., Schmeing,T.M., Moore,P.B. and Steitz,T.A. (2001) The kink-turn: a new RNA secondary structure motif. *Embo J.*, **20**, 4214–4221.

51. Vidovic,I., Nottrott,S., Hartmuth,K., Luhrmann,R. and Ficner,R. (2000) Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol. Cell*, **6**, 1331–1342.

52. Rozhdestvensky,T.S., Tang,T.H., Tchirkova,I.V., Brosius,J., Bachellerie,J.P. and Huttenhofer,A. (2003) Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res.*, **31**, 869–877.

53. Nolivos,S., Carpousis,A.J. and Clouet-d'Orval,B. (2005) The K-loop, a general feature of the Pyrococcus C/D guide RNAs, is an RNA structural motif related to the K-turn. *Nucleic Acids Res.*, **33**, 6507–6514.

54. Fukuhara,H., Kifusa,M., Watanabe,M., Terada,A., Honda,T., Numata,T., Kakuta,Y. and Kimura,M. (2006) A fifth protein subunit Ph1496p elevates the optimum temperature for the ribonuclease P activity from *Pyrococcus horikoshii* OT3. *Biochem. Biophys. Res. Commun.*, **343**, 956–964.

55. Terada,A., Honda,T., Fukuhara,H., Hada,K. and Kimura,M. (2006) Characterization of the Archaeal ribonuclease P proteins from *Pyrococcus horikoshii* OT3. *J. Biochem. (Tokyo)*.

56. Marquez,S.M., Harris,J.K., Kelley,S.T., Brown,J.W., Dawson,S.C., Roberts,E.C. and Pace,N.R. (2005) Structural implications of novel diversity in eucaryal RNase P RNA. *RNA*, **11**, 739–751.

57. Stuart,K.D., Schnaufer,A., Ernst,N.L. and Panigrahi,A.K. (2005) Complex management: RNA editing in trypanosomes. *Trends Biochem. Sci.*, **30**, 97–105.

58. Ivens,A.C., Peacock,C.S., Worthey,E.A., Murphy,L., Aggarwal,G., Berriman,M., Sisk,E., Rajandream,M.A., Adlem,E., Aert,R. *et al.* (2005) The genome of the kinetoplastid parasite, Leishmania major. *Science*, **309**, 436–442.

59. Clamp,M., Cuff,J., Searle,S.M. and Barton,G.J. (2004) The Jalview Java alignment editor. *Bioinformatics*, **20**, 426–427.