Research Article for BMC Evolutionary Biology

11 March 2006

RNase MRP and the RNA Processing Cascade in the Eukaryotic Ancestor

Michael D. Woodhams[1]*, Peter F. Stadler[2], David Penny[1], Lesley J. Collins[1*§]

[1] Allan Wilson Centre for Molecular Ecology and Evolution, Massey University, Palmerston North, New Zealand.

[2] Bioinformatics Group, Department of Computer Science and Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstraße 16-18, D-04107, Germany.

*These authors contributed equally to this work
[§]Corresponding author

Email addresses:
MDW: M.D.Woodhams@massey.ac.nz
PFS: Peter.Stadler@bioinf.uni-leipzig.de
DP: D.Penny@massey.ac.nz
LJC: L.J.Collins@massey.ac.nz

Abstract

Background

Within eukaryotes there is a complex 'cascade' of RNA-based macromolecules that process other RNA molecules, especially mRNA, tRNA and rRNA. A simple example is the RNase MRP processing of ribosomal RNA (rRNA) in ribosome biogenesis. One hypothesis is that this complexity was present early in eukaryotic evolution; an alternative is that an initial simplified network later gained complexity by gene duplication in lineages that led to animals, fungi and plants. Recently there has been a rapid increase in support for the complexity-early theory because the vast majority of these RNA-processing reactions are found throughout eukaryotes, and thus were likely to be present in the last common ancestor of living eukaryotes, named here as the Eukaryotic Ancestor.

Results

We present an overview of the RNA processing cascade in the Eukaryotic Ancestor and investigate in particular, RNase MRP which was previously thought to have evolved later in eukaryotes due to its apparent limited distribution in fungi and animals and plants. Recent publications, as well as our own genomic searches have uncovered previously unknown RNase MRP RNAs, indicating that RNase MRP has a wide distribution in eukaryotes. Combining secondary structure and promoter region analysis of new and previously discovered RNase MRP RNAs along with analysis of the primary substrate (rRNA), allows us to discuss this distribution in the light of eukaryotic evolution.

Conclusions

We conclude that RNase MRP can now be placed in the RNA-processing cascade present in the Eukaryotic Ancestor. This highlights the complexity of RNA-processing in early eukaryotes.

Background

There is increasing interest in investigating the expanding number of roles of RNA in modern eukaryotes . The number of putative ncRNAs (non-coding RNAs) in the mammals alone has increased about 20-fold in the last five years , thus any information on the origins and functions of well-established ncRNAs is relevant and timely. In eukaryotes a number of ncRNA-based molecules are directly involved in the cleavage and processing of other RNA molecules. A classic example is the cleavage of rRNA transcript by RNase MRP, a ribonucleoprotein complex consisting of a single RNA molecule and about 10 proteins . In at least one example there is a series of reactions, for example the snRNAs in the spliceosome release snoRNAs from introns which in turn are involved in the modification of rRNA, tRNA or snoRNAs (see Figure 1). We call the networking of these processes the eukaryotic RNA-processing cascade. This cascade is centred around the processing of three types of RNA, mRNA, tRNA and rRNA and although each of these RNAs is cleaved in separate reactions, there are linkages between these reactions as shown in Figure 1. The question we ask here is how ancient are these RNA-based processes.

Pre-mRNA contains introns that are processed by the spliceosome (consisting of 5 snRNAs and ~200 proteins) but there is also further processing such as the

addition of the 5'-cap and 3' poly-A-tail . Although the 5' capping and 3' polyT tail processes are not RNA-based reactions they include some proteins that have also been found in the spliceosome . The snRNAs within the spliceosomal complex not only direct the binding and coordination of the splice sites but are also implicated in the catalysis of the splicing reactions . Some introns contain ncRNAs such as snoRNAs (involved in modification of rRNA, tRNA and snRNAs) (reviewed in) or miRNAs involved in the degradation of mRNA . Pre-tRNA is processed by RNase P; a ribonucleoprotein consisting in eukaryotes of a single RNA and about 8-10 proteins . RNase P (abbreviated here as P) is found throughout eukaryotes and prokaryotes , and thus may date back to the RNA-world . Pre-rRNA is heavily processed by proteins; however, a specific site (the A3 site in the ITS region) is cleaved by the ribonucleoprotein RNase MRP (abbreviated here as MRP) generating the mature 5.8S rRNA. Mature rRNA along with many proteins (at least one of which is also found in the spliceosome) forms the ribosome.

MRP was originally identified as an RNA-protein endoribonuclease that in the mitochondria processes RNA primers for DNA replication and it is likely that MRP has other essential functions including roles in chromosomal segregation and control of cell division . Although named after its mitochondrial function, the majority of MRP (99%) is observed in the nucleolus where it plays its important role in pre-rRNA processing . Prior to this work, evolutionary studies used MRPs from only animals, yeasts and plants raising questions as to whether MRP was present in the last common ancestor of modern eukaryotes, named here as the Eukaryotic Ancestor . Collins et al. considered three hypotheses for the distribution of MRP (Figure 2). The first is that MRP is very ancient, occurring at least in the first eukaryotes. There are many variants on this model, and MRP could even be much older in that most catalytic roles of RNA may derive from much earlier stages in the origin of life, namely in the RNA-world . The second group of models is that MRP arose from a duplication of P within current eukaryotes. This would explain the

apparent limited distribution of MRP restricted to plants, animals and fungi, as well as the observation that P and MRP share some accessory proteins . Such duplication would be followed by specialization of the paralogous complexes, P being restricted to tRNA, and MRP to rRNA. It is unclear under this model whether MRP took a new role in an internal excision in the precursor of rRNA , or whether eukaryotes were initially different in that P initially carried out both reactions (to the precursors of tRNA and of rRNA). The third group of hypotheses is that MRP is derived from an early mitochondrial RNase P, followed by transfer of the gene to the nucleus, and co-option of MRP to a role in the nucleus in processing rRNA.

In our earlier work it was concluded that the second hypothesis was the most likely, that MRP had arisen within eukaryotes by a duplication of P, with subsequent specialization. The evidence against MRP coming from mitochondria (the third hypothesis) was that the secondary structure of the RNA component of MRP (MRP-RNA), as measured by RNA-shape comparison metrics , was more similar to the eukaryote RNase P RNA component than to the RNA from bacterial RNase P (the presumed source of the mitochondrial RNase P). Similarly, the (at that time) apparent limited distribution of MRP in eukaryotes made it seem unlikely that MRP was present in the ancestral eukaryote. This left the duplication of P within eukaryotes as the most likely explanation at that time, and it also explained why some of the proteins were shared by both P and MRP. However, this has now changed because of two developments. Firstly, it now appears that the plant lineage and the fungi and animal lineage (fungamals) are quite widely separated on the eukaryote tree . Secondly, the recent discovery of MRP outside animals, fungi and plants (as reported here and) means that our initial conclusion for the origin of MRP needs to be reconsidered.

Prior to this study and , full or partial sequences for MRP-RNA had only been published for 13 mammals, a frog, two dicotyledonous plants, 20 yeasts from the order Saccharomycetales and the fission yeast *Schizosaccharomyces pombe*. These

species came from all three multicellular kingdoms, but only a small phylogenetic range within each: land vertebrates within metazoans, Ascomycota within fungi, and the core eudicots within plants although this range was extended further in . Our study used an MRP-specific search strategy to find candidate MRP-RNA sequences in a number of eukaryotic species including some species outside of the fungamal-plant grouping. Examination of promoter regions and RNA secondary structure increased the viability of these candidates and strengthens gene and secondary structure consensus models for MRP-RNA throughout eukaryotes. In the light of these results we discuss the presence of MRP in the Eukaryotic Ancestor and re-examine the evolution of this ribonucleoprotein and the RNA-processing cascade throughout eukaryotic lineages.

Results

RNase MRP is widely distributed in Eukaryotes

Our MRP-specific search strategy found candidate MRP-RNA sequences in a range of eukaryotes (species and accessions for the new sequences are given in Table 1). the pufferfish *Takifugu rubripes*, zebrafish *Danio rerio*, the sea-squirt *Ciona intestinalis* (a non-vertebrate chordate), fruit-fly *Drosophila melanogaster* and the human malaria parasite *Plasmodium falciparum* (an apicomplexan protist). These new sequences and existing sequences were used as BLAST templates to find additional candidates (also shown in Table 1) in rabbit (*Oryctolagus cuniculus*), chimpanzee (*Pan troglodytes*), dog (*Canis familaris*), opossum (*Monodelphis domestica*), chicken (*Gallus gallus,* sequence incomplete), western clawed frog (*Xenopus tropicalis*), pufferfish (*Tetraodon nigroviridis)*, another sea-squirt (*Ciona savignyi*), the sea-urchin (*Strongylocentrotus purpuratus*), five other species of fruit-flies (*D. pseudoobscura, D. yakuba, D. mojavensis, D. virilis and D. ananassae*), the plant

Brassica oleracea (cabbage), six other species of *Plasmodium* (*P. yoelli*, *P. berghei*, *P. chabaudi*, *P. gallinaceum*, *P. knowlesi* and *P. vivax*) and one candidate each from *Cryptosporidium parvum* and *Cryptosporidium hominis* (apicomplexan protists). Search results are summarised in Table 1. RT-PCR of the candidate from *C. parvum* indicated that this sequence is expressed (M. Irimia, data not shown).

During this work a recent publication identified a large number of RNase MRP RNAs from a diverse range of eukaryotes including some species outside the fungamal and plant groups and some of the same sequences found in our study. We have included some of the overlapping species in our results as they provide validation of our search technique as well as the search method in Piccinelli et al. 2005 .

Of our new MRP-RNA candidates, only two are found in expressed sequence tag (EST) databases: *D. melanogaster* [accession CO153932] and *Plasmodium yoelli* [accessions BM161600 and BM160961]. Nevertheless, these are important in supporting our bioinformatic approach. Five of our MRP-RNA sequences are annotated in Genbank records: *D. melanogaster* is on the negative strand of an intron in the gene CG10365 [accession AE003744], *G. gallus* [accession AADN01006913], *T. nigroviridis* [accession CAAE01012081], *P. falciparum* [accession NC_004325]and *P. yoelli* [accession AABL01002665] are all between genes. With the exception of *Strongylocentrotus purpuratus*, (the sea-urchin) which appears to have five closely related sequences, only a single copy of the MRP-RNA gene was found in each organism. It is likely that at least some of the copies in the sea-urchin will turn out to be artefacts of the current genome assembly. A multiplicity of MRP-RNA genes has previously been observed in plants , but the few published results for animal sequences indicate that humans and the pufferfish *Takifugu rubripes* typically have a single true copy (although some pseudogenes were been found in humans).

Analysis of the sequences found here and in shows that MRP is distributed across a wide range of eukaryotes and is not limited to the animal, fungi and plant lineages. The characterisation of MRP in protists indicates the evolutionary relationship between MRP and P is ancient and MRP and P are likely to have been present in the last common ancestor of modern eukaryotes.

Promoter analysis of candidate MRP-RNA sequences

For MRP the genes for proteins and RNA subunits are transcribed by different RNA polymerases; the proteins by RNA Polymerase II (typical for proteins), and the RNA subunit by RNA Polymerase III (type III - typical in eukaryotes for U6 snRNA 7SK, hY4, hY5 and P-RNA) . Different organisms vary in their RNA Polymerase III promoter elements. In general, vertebrate and plant MRP-RNA promoter regions contain an upstream TATA box, Proximal Sequence Element (PSE or USE) and a Distal Promoter Element (DSE) which can contain SP1, Staf and/or Octamer motifs . In humans, the presence of the TATA box determines RNA polymerase specificity, with the other elements (e.g. PSE and DSE elements) enhancing transcription . Plants require both the TATA box and the USE promoter element (similar in sequence and position to the PSE element in vertebrates) with RNA polymerase specificity determined by the spacing between the two elements . In *Drosophila melanogaster*, specificity is determined by the presence of the TATA box and the sequence of the PSE element .

However in the yeast *Saccharomyces cerevisiae* a different RNA polymerase III promoter structure is used although it is still regarded as a type III structure because all elements are still external to the transcribed area . For example, the U6 snRNA promoter (similar to that expected in MRP-RNA) lacks PSE and DSE elements but instead includes a downstream B box ~120 nucleotides beyond the terminator .

Promoter analysis (summarised in Figure 3a) of RNA polymerase III gene candidates is useful for validating an MRP-RNA gene candidate. Analysis of MRP-RNA upstream and downstream regions indicated that MRP-RNA is likely to be transcribed using RNA polymerase III throughout eukaryotes. However, individual elements within the overall RNA polymerase III promoter structure can change even within a group of organisms. For example in fish, the MRP-RNA promoter region for *Takifugu rubripes* previously described in characterises a Staf promoter element (a binding site for the Staf transcriptional activator protein) in the DSE. However, we were unable to find any similar Staf-binding sequence in the other two fish MRP-RNAs studied here. The Zebrafish and *T. nigroviridis* MRP-RNAs have potential SP1 binding sites, but as with the *Takifugu* MRP-RNA, no Octamer sites could be determined.

In general, mammals and chicken contain a similar arrangement of their MRP-RNA promoter elements. The frog MRP-RNA promoter regions have the same individual elements with sequence motifs typical of mammals, but have a slightly different spacing between the elements within the DSE (the SP1 binding site is further upstream).

Comparisons between six species of *Drosophila* (*D. melanogaster, D. pseudoobscura, D. yakuba, D. mojavensis, D. virilis* and *D. ananassae*) show a conserved PSE element (consensus sequence gcTTAtaATTCCCAAct) 23 nucleotides upstream of a TATA box (consensus sequence taaAta) which is about 16 nucleotides upstream of the transcription start site. However, the range of RNA polymerase III promoter structures and the present lack of information about these promoter elements from the apicomplexan protists *Plasmodium* and *Cryptosporidium,* makes it difficult for us to identify promoter elements in new MRP genes. Analysis of promoter regions from apicomplexa indicate the presence of a TATA box but since we know so little about RNA Polymerase III regulation in these protists we cannot as yet predict the presence of any PSE or DSE elements.

The common features shared between MRP-RNA and P-RNA leaves little doubt that they are evolutionary related. It is interesting, however, that we find differences in the promoter regions required for their transcription. For example in humans, although both RNAs are transcribed by RNA polymerase III (type III); the P-RNA gene contains a more compact promoter with a Staf site next to the PSE element while the MRP-RNA promoter region is arranged more similar to the U6 snRNA gene but with promoter sequences closer to that of P-RNA (Figure 4).

Secondary structure analysis of MRP-RNA

Analysis of the secondary structure of MRP-RNA (this work and) reveals that the overall secondary structure is very conserved throughout eukaryotes (Figure 5). A large number of features (P1, P2, P3a, P4, P5 and P7) are found throughout all the MRP-RNA characterised to date whereas other features are nearly universal (P3b, P6 and P19). There are some features that are observed in a few organisms of limited phylogenetic range (P3c, P5a, P7a, P8 and P15).

General observations on the eukaryotic MRP-RNA are as follows. Typically P7 is long, with many internal loops (schematically represented by two internal loops in Figure 5) and occasionally bifurcations (i.e. P7a). P8 is clearly present in some *Saccharomycetes* yeasts (*Debayomyces hansenii*, *Yarrowia lipolytica*, *Pichia guilliermondi*), and present in the remainder under an alternative structure (see below). P8 is also present in some apicomplexa (*Babesia bovis*, *Eimeria tenulla*, *Toxoplasma gondii*). The P15 region is present in *Schizosaccharomyces pombe*, all *Saccharomycetes* we studied and some *Pezizomycotina* yeasts. It has significant single-strand regions on either side. However, the distinction between P8 and P15 is not always clear (e.g. *Coccidioides immitis*). The P3c feature is observed in *Cryptosporidia*, *Dictyostelium discoidium*, the mosquito *Anopheles gambiae* and the roundworm *Brugia malayi*.

Some features however, are lost in some lineages. P19 is absent from *Ciona intestinalis* and P3b and P6 are absent from microsporidia. P6 is also absent from *D. discoidium* and depending on the folding, *Cryptosporidium* (our folding has a P6 present, the secondary structures provided by for *C. parvum* and *C. hominis* do not).

One interesting structural feature is the P5 loop which has a frequently recurring, but not universal motif of GARAG, or sometimes GARA (R=G or A) on a short (3-5 pair) helix. Animals generally have GARAG, however, exceptions are the fish *Tetraodon nigroviridis* (CAAAG) and *Danio rerio* (GAGA). Within the fungi, the situation is complex. Pezizomycotina yeasts (e.g. *Aspergillus nidulans* and *Neurospora crassa*) all have GAAA, but have another helix between this one and CR-I (5'P4). Basidiomycetes (e.g. *Coprinus cinereus* and *Phanerochaete chysosporium*) and *Schizosaccharomyces pombe* do contain the GARAG motif. MRP-RNAs from *Saccharomyces* species do not contain the GARAG motif in the P5 region of published secondary structures, but display GAAAA in an alternative structure An exception in this case is *Yarrowia lipolytica* which does not contain anything resembling a GARAG motif in either structure. The alternative structure that can be drawn for Saccharomyces MRP-RNAs (supplied in supplementary data) allows for two features that are 'typical' for eukaryotic MRP-RNAs (the P8 region and the GARAG motif). However, the *Saccharomyces cerevisiae* structure was recently investigated biochemically and supports structures used previously . A possibility exits that these yeasts have changed their structure from one that may have resembled our alternative structure to the one that is seen in modern yeasts.

The microsporidian species *Nosema locustae* and *Encephalitozoon cuniculi* also contain the GARAG motif. Plants and green algae have GAGA or GAGAG, however an exception in this group is the cabbage *Brassica oleracea* (GAGG).

Among apicomplexa *Toxoplama gondii*, *Theileria annulata* conform to the motif; *Babesia bovis* (TAAAG) and *Eimeria tenulla* (GCGAG) nearly conform, however, the *Cryptosporidium* species, *Plasmodium* species and *Trichomona*

vaginalis do not contain anything resembling the GARAG motif. The other protists *Oxytricha trifallax* and *Tetrahymena thermophila* (both ciliates). *Dictyostelium discoideum,* the heterokontae *Phytophthora ramorum* and *Thalassiosira pseudonana* all contain the GARAG motif. The GARAG motif was also independently highlighted in supplementary information available from . To date it is not known as to whether this motif reflects a protein binding region or a motif required for the correct formation of the MRP-RNA tertiary structure.

Discussion

The identification of MRP across a wide distribution of eukaryotes indicates that MRP was likely to be present in the last common ancestor of modern eukaryotes (the Eukaryotic Ancestor). While there is little doubt that MRP and P are evolutionary related, there is at present no evidence to suggest that MRP arose from a duplication of P, just that they were both present in the Eukaryotic Ancestor. At this stage we cannot determine how far back beyond the Eukaryotic Ancestor that these two RNA-based complexes had a common ancestor.

The fact that we can still observe the relationship between MRP-RNA and P-RNA is extremely interesting. The high similarity between MRP and P secondary structure is indicative of an evolutionary relationship. However, this does not mean that the closeness is in evolutionary distance in time between these macromolecules: it is more likely that the closeness is maintained by the sharing of numerous proteins between the MRP and P complexes. Thus much of the large similarity in secondary structure between sections of MRP and P-RNAs (e.g. the P3-region indicated in) is likely due to the constraints placed on the RNA molecules by their interactions with their common proteins.

In the nematodes (*C. elegans* and *C. briggsae*) no MRP-RNA was found either in this study or , although MRP is present in *Brugia malayi* , another nematode

species. A recent survey for structured ncRNAs based on comparative analysis of *C. elegans* and *C. briggsae* also did not result in a plausible MRP candidate. Thus the detection of MRP (if it is present) in these species may only be possible by biochemical means.

MRP is now implicated in a number of cellular processes in eukaryotes especially in well-researched species such as humans and the yeast *S. cerevisiae*. As well as nuclear rRNA and mitochondrial primer cleavage functions, in *S. cerevisiae* at least, it has an additional function of promoting cell cycle progressing by cleaving CLB2 mRNA in its 5' UTR region at the end of mitosis to remove the 5' cap . Removal of the A3 processing site (the 'main' nuclear function of MRP) and loss of mitochondrial DNA (the 'main' mitochondrial function of MRP) are not lethal in yeast . It is possible therefore, that other functions of MRP may be found especially during study of other eukaryotes from which MRP has only recently been characterised.

The piecing together of the eukaryotic RNA-processing cascade and the investigation of the distribution of MRP has leads us to conclude that the last common ancestor of modern eukaryotes is likely to have contained an RNA-processing cascade similar to that seen today (see Figure 1). Prior to this study, MRP was decidedly the odd-man-out being seen to have arisen much later in eukaryotes unlike the other components of the cascade (e.g. spliceosomes , snoRNAs , introns , RNase P and RNAi). However, its presence in eukaryotes in most lineages of eukaryotes implies that it too was present in the RNA-processing cascade present in the Eukaryotic Ancestor. A notable exception is the protist *Giardia lamblia*. Both our searches and those of Piccinelli failed to find an MRP-RNA candidate in this species although P-RNA has been reported a number of times . To date we have also not yet recovered any MRP-RNA from a *G. lamblia* RNA library (although again, we have recovered P-RNA) (S. Chen, data not shown). This does not mean MRP is not present in *G. lamblia* because the rRNA gene arrangement is generally the same as seen in

other eukaryotes, and there is some secondary structure in the *G. lamblia* ITS1 region that suggests that an A3 site may be present (data not shown). The large evolutionary distance between *G. lamblia* and any other eukaryote, including that of the excavate from which MRP has been previously characterised (the Parabasalid, *Trichomonas vaginalis*) means that MRP may be difficult to characterise in *G. lamblia*.

One of the main conclusions in this study is that, with the placement of MRP in the RNA-processing cascade of the Eukaryotic Ancestor, we see little change in basic RNA-processing throughout eukaryotes. This has implications on rRNA processing evolution in particular. Eukaryotes and prokaryotes have fundamental differences in their processing of their rRNA transcripts; the main eukaryotic transcript contains ITS1 (between the 12S and 5.8S) and ITS2 (between the 5.8S and 28S) whereas prokaryotes have only an ITS1 with the 5'end of the prokaryotic 23S showing strong homology to the eukaryotic 5.8S sequence . Thus, there are two states in which we can find the 5.8S rRNA, either cleaved as a separate subunit or fused to the large rRNA subunit. Typically within eukaryotes we find the 5.8S rRNA cleaved but in prokaryotes they are not. However there are exceptions, for eukaryotes microsporidia do not cleave the 5.8S rRNA , and in prokaryotes RNase III cleaved IVS (intervening sequence) regions in α-proteobacteria have been found in the 23S rRNA . RNase III which is involved in cleaving the prokaryotic rRNA transcript has now been implicated in ITS1 processing in *S. pombe* . Although is likely that the cleaved 5.8S rRNA may have been present in the Eukaryotic Ancestor, we cannot as yet determine if the last universal common ancestor (of eukaryotes and prokaryotes) contained a separate 5.8S or the fused version.

Overall, it is likely that the major components of the RNA processing cascade, especially the RNA components evolved before the Eukaryotic Ancestor. The Eukaryotic Ancestor is now seen to have come after the mitochondrial endosymbiosis, and it is possible that MRP, like that found in modern eukaryotes, performed a number of functions, including functions in the nucleus and the ancient

mitochondria. It is interesting to note that MRP is still found in species that no longer contain a mitochondria as such , but contain instead reduced organelles such as mitosomes or remnant mitochondria (apicomplexa and microsporidia) and hydrogenosomes (ciliates, parabasalids and some fungi) .

The RNA-processing cascade can now been seen as a complex feature of the ancestral eukaryotic cell. Understanding ancestral RNA-processing is, of course, just the tip of the iceberg when considering eukaryotic evolution. However, once we understand which eukaryotic processes were present in the Eukaryotic Ancestor we can then look at how they evolved in the first place.

Conclusions

We present the organisation of RNA-processing in eukaryotes as a cascade of RNA-based processing reactions cleaving or modifying other RNA molecules. The main components of this cascade are seen to be conserved throughout eukaryotes and are likely to have been present in the Eukaryotic ancestor. Prior to this study evolutionary analysis of MRP was restricted to information from animals, fungi and plants and thus could not be seen as ancestral to eukaryotes. We can now place MRP in the RNA processing cascade that was likely to be present in the Eukaryotic Ancestor. This implies that basic RNA-processing has been preserved during eukaryotic evolution.

Methods

Searching genomes for RNase MRP RNA

The conserved regions around the P4 pseudoknot have been the key to our identification of candidate MRP-RNA sequences in novel organisms. We first scanned the genome for sequences similar to the conserved sequences then evaluated candidates for support of the stereotypical secondary structure. Candidates with suitable secondary structure were then evaluated for upstream promoter regions expected for a gene transcribed by RNA polymerase III. Candidate sequences were

then blasted generally against EST databases via the NCBI web page (www.ncbi.nlm.nih.gov) for any indication that the candidate was expressed.

In the scanning step we have some flexibility on how closely the candidate must match the conserved regions, and how large a separation we allow between the conserved regions. The consensus for 5'P4 and 3'P4 was set at gaaAGuCCCC and acnnnanGGGGCUnannnu respectively (paired bases in uppercase.) Any unpaired base which differs from this consensus was counted as one deviation, as was any pair that differs, so long as they remain a Watson-Crick pair (any other pairings for these bases was rejected). Two sets of search criteria was used: firstly 'tight' criteria allowed up to one deviation from the consensus, and separation of 120 to 280 bases between the conserved regions. A second 'relaxed' criteria allowed up to two deviations and a separation of 80 to 500 bases. If the tight criteria yielded no viable MRP-RNA candidates (i.e. none of the matches found can fold correctly), the search was repeated with the relaxed criteria. Secondary structure evaluation (as described below) was used to further filter potential MRP-RNA candidates.

Secondary structure analysis of MRP-RNA

General vertebrate and yeast secondary structures were obtained from the literature. Secondary structure evaluation was a semi-manual process, aided by programs such as RNAfold and Mfold . We looked for candidate P3 and P9 helices adjacent to the P4 halves and then for P2. If the number of candidates was large, we then used RNAmotif to filter out candidates that did not have suitable P2 and P3 helices.

Sequence alignments prior to structure analysis used ClustalX and DIALIGN .Secondary structure analysis was done using Alifold from the Vienna RNA package , RNAforester , RNAshapes and RNAcast .

Authors' contributions

MW carried out the search and secondary structure analysis and drafted the original manuscript. PFS contributed to the search of new and not easily available genomes. DP participated in the design of the study and contributed to the evolutionary discussions. LC carried out the promoter analysis and drafted the final manuscript. All authors read and approved the final manuscript.

Acknowledgements

Thanks to Manuel Irimia for RT-PCR work and Sylvia (Xiaowei) Chen (Allan Wilson Centre) for results from the *Giardia lamblia* RNA library. Computational analysis was carried out using the Helix Parallel Processing Facility at Massey University. This work was funded by the New Zealand Marsden Fund, the New Zealand Centres of Research Excellence Fund and the Bioinformatics Initiative of the German DFG.

References

Figure Legends

Figure 1. The eukaryotic RNA-processing cascade.
Blue arrows are cleavage reactions; Green arrows are modification reactions: Striped arrows are addition reactions and Black arrows are transitions between the cascade stages. mRNA is cleaved by the spliceosome (comprised of snRNAs and proteins) to release the processed mRNA and introns. Some introns contain snoRNAs which in turn modify snRNAs, tRNAs and rRNAs. Other introns contain miRNAs used in RNAi reactions. RNase P (P) cleaves pre-tRNA while RNase MRP (MRP) cleaves rRNA. The ribosomal complex (comprised of rRNAs) brings the tRNAs and mature mRNAs together for translation.

Figure 2. Hypothesis for the origin of RNase MRP based on [28].
The large black dots represent the point of duplication of the P-MRP ancestor. **A:** MRP was present in the last common ancestor of modern eukaryotes (the Eukaryotic Ancestor). Alternatively both MRP and P could have been present in the Last Universal Common Ancestor. **B:** MRP arose from a duplication of P after the Eukaryotic Ancestor, but before the ancestor of animals, fungi and plants. **C:** MRP arose from an early mitochondrial RNase P within the Eukaryotic Ancestor.

Figure 3: MRP-RNA gene arrangement.
Genes transcribed by RNA polymerase III (type III) usually contain a PSE (proximal sequence element) consisting of a TATA signal and PSE motif, and a DSE (distal sequence element) consisting of either a SP1, Oct or Staf binding site. Distances shown are approximate only. Key: T – TATA signal; PSE / USE – ; Oct – Octamer binding site; SP1 – SP1 binding site; Staf – Staf binding site. ? – Possible site. TT – Poly T termination signal. B-box – Downstream B-box motif.

Figure 4. Promoter regions of Human MRP , P and U6 snRNA . Although the arrangement of the MRP-RNA promoter region is similar to that of the U6 snRNA, the actual sequences within the promoter elements are closer to those found in P-RNA.

Figure 5. Summary diagram of the MRP-RNA secondary structure.
Black features (P1, P2, P3a, P4, P5, P7) are universally present. Blue features are nearly universal, red features are observed in a few organisms of limited phylogenetic range. Thick lines are paired regions while unpaired regions are shown as thin lines. Conserved sequence motifs are indicated for the P4 (5' and 3') and P5 regions.

Table Legend

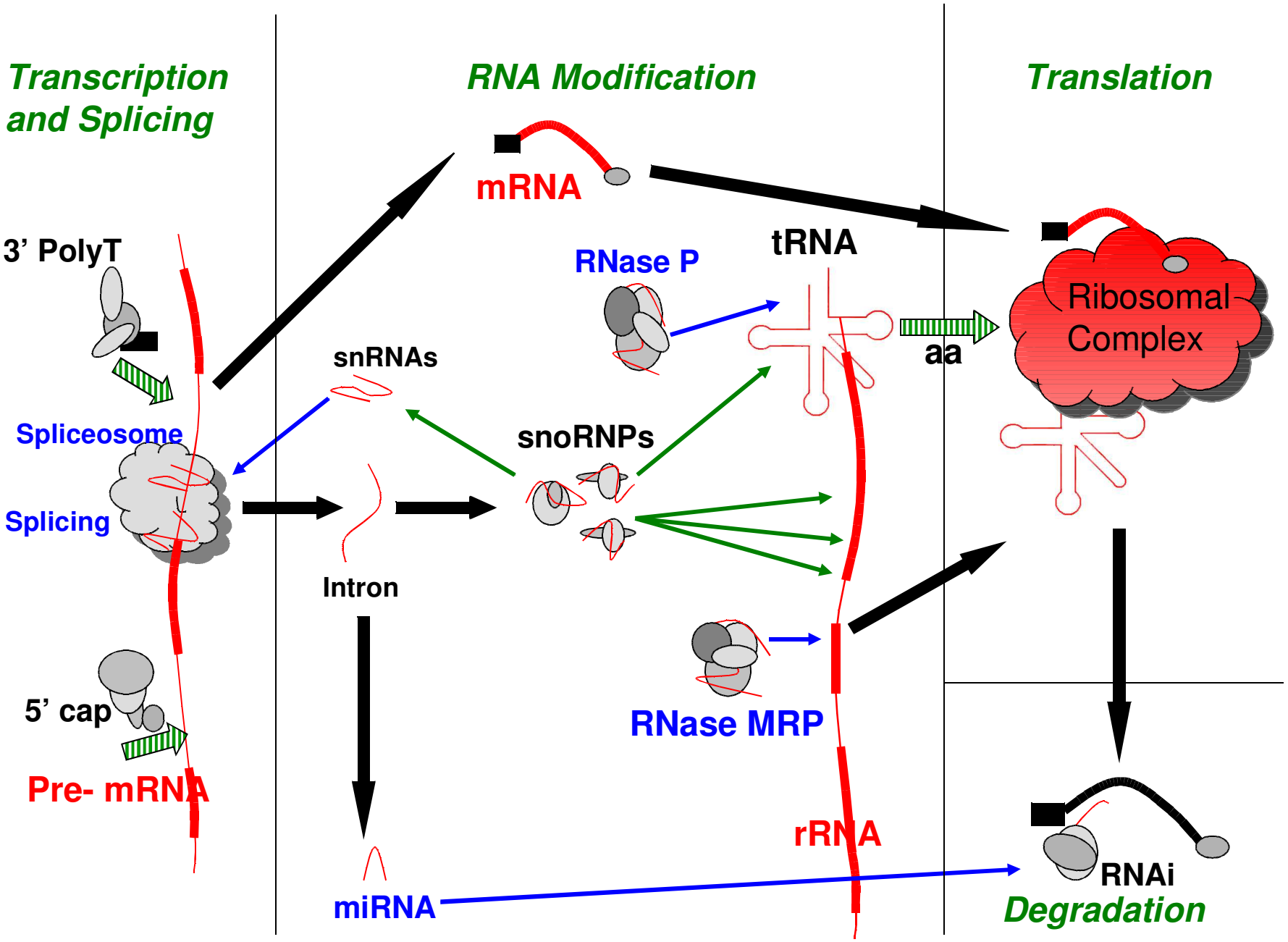
Table 1: MRP-RNA found in this study. Key: * – reported in .

Supplementary Figure 1.

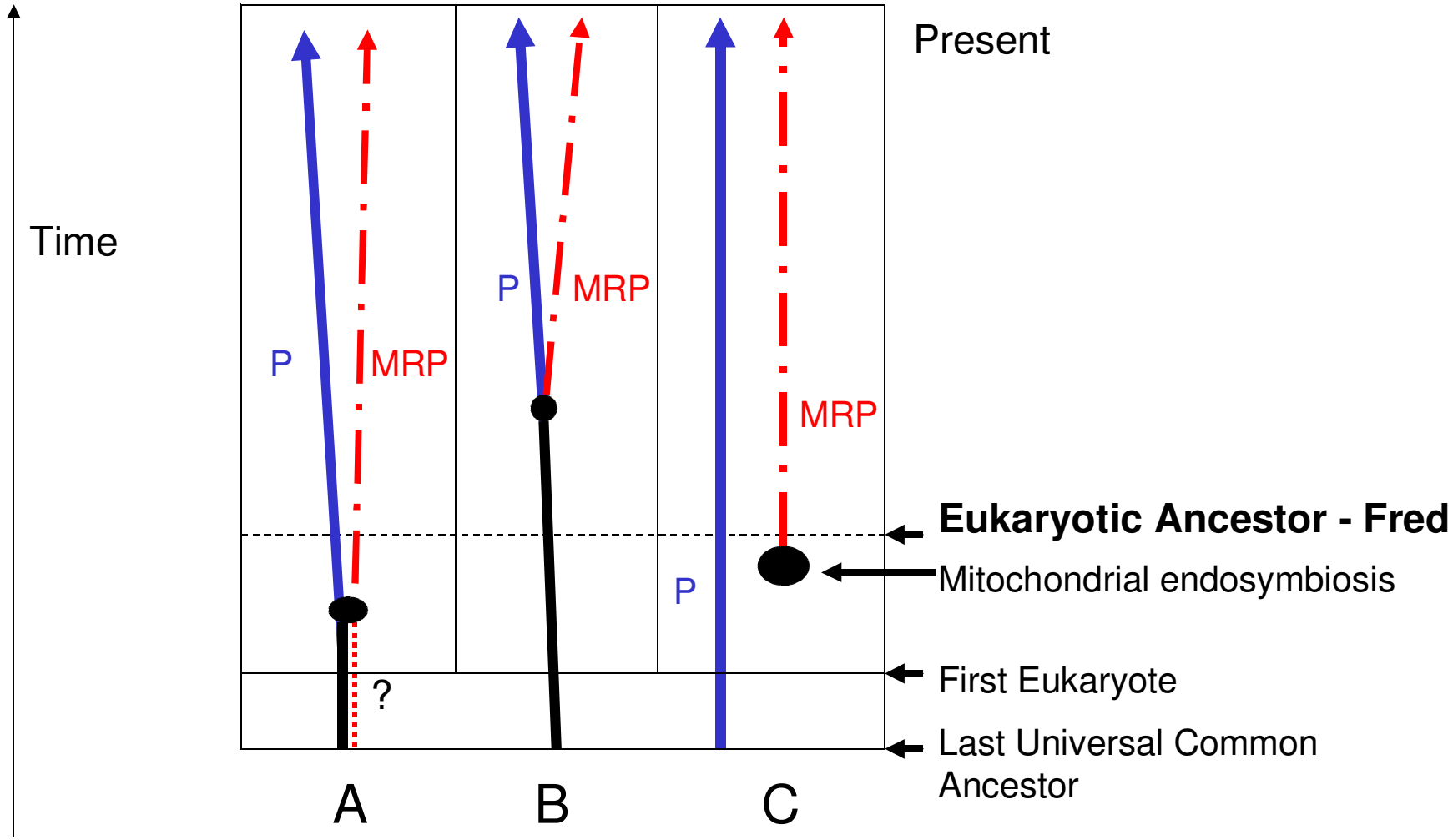
Supp Figure 1. Alternative folding for the *S. cerevisiae* MRP-RNA displaying the GARAG motif

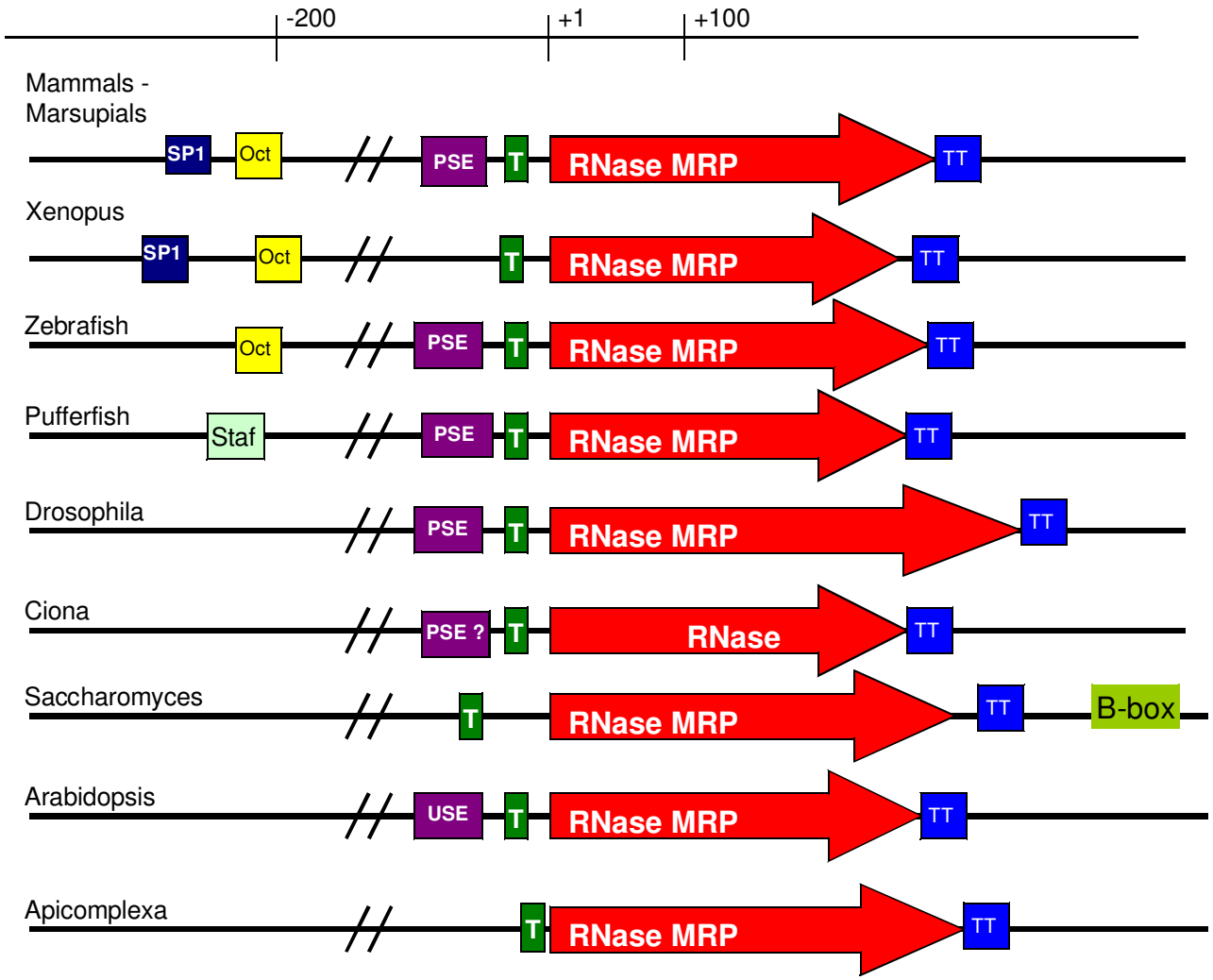
Species	Common Name (if any)	Group	Accession number	Co-ordinates
Pan troglodytes	Chimp	Animal	AADA01035511	14291-14555
Canis familaris *	Dog	Animal	AAEX01055752.1	26663-26939
Oryctolagus cuniculus	Rabbit	Animal	AAGW01261685.1	260-540
Monodelphis domestica	Opossum	Animal	Assembly 0.5, scaffold_15143	5443339-5443058
Gallus gallus	Chicken	Animal	AADN01006913.1	200-1
Xenopus tropicalis	Western clawed frog	Animal	Assembly 3.0, Scaffold 99	1471260-1471531
Danio rerio *	Zebrafish	Animal	CAAK01000119.1	2793811-2793547
Fugu rubripes *	Pufferfish	Animal	CAAB01000416.1	43232-42993
*Tetraodon nigroviridis**	Pufferfish	Animal	CAAE01012081.1	21509 - 21762
*Ciona intestinalis**	Sea-squirt	Animal	AABS01000030.1	66300-66051
*Ciona savignyi**	Sea-squirt	Animal	AACT01041809.1	6951 - 7211
Strongylocentrotus purpuratus	Purple sea-urchin	Animal	AAGJ01116184.1 AAGJ01129308.1 AAGJ01275051.1 AAGJ01178199.1 AAGJ01178201.1	71-323 7193-6941 340-88 708-451 8051-8311
*Saccharomyces mikatae**	Yeast	Fungi	gnl|ti|203281071	589-255
Brassica oleracea	Cabbage	Plant	Contig BOMBD54TR	93-316
*Plasmodium falciparum**	Human malaria parasite	Apicomplexa	NC_004325.1	971-1324
*Plasmodium berghei**	-	Apicomplexa	Pb_5607	6152-6493
*Plasmodium chabaudi**	-	Apicomplexa	Pc_6141	2512-2836
*Plasmodium knowlesi**	-	Apicomplexa	Pkn1318d01	6151-5780
*Plasmodium vivax**	-	Apicomplexa	Pv_4041	258309-257936
*Plasmodium yoelii yoelii**	Mouse malaria parasite	Apicomplexa	AABL01002665.1	3440 - 3140
*Plasmodium gallinacetia**	-	Apicomplexa	Pg_c000013117.Contig1	971-1324
*Cryptosporidium parvum**	-	Apicomplexa	Chromosome4:1,1106229	381884 -382317
*Cryptosporidium hominum**	-	Apicomplexa	AAEL01000837.1	1688-1253

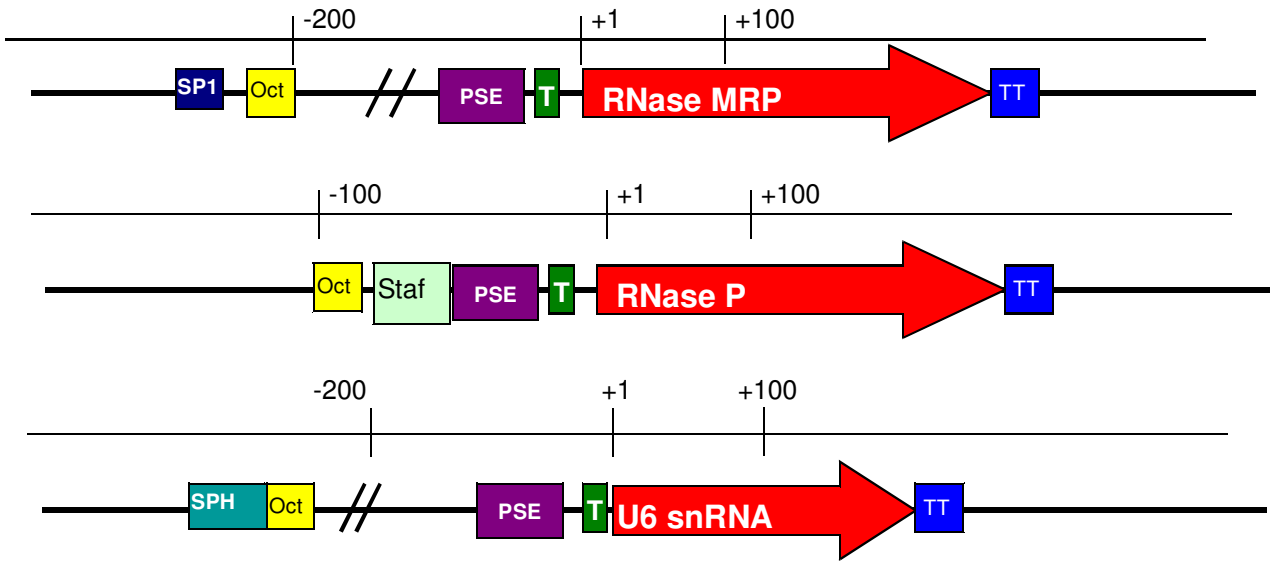
Table 1. MRP-RNAs found in this study. Key: * – reported in Piccinelli et al. 2005 .


Transcription and Splicing
RNA Modification
Translation
mRNA
3' PolyT
RNase P
tRNA
Ribosomal Complex
aa
snRNAs
Spliceosome
snoRNPs
Splicing
Intron
5' cap
RNase MRP
Pre- mRNA
rRNA
RNAi
miRNA
Degradation

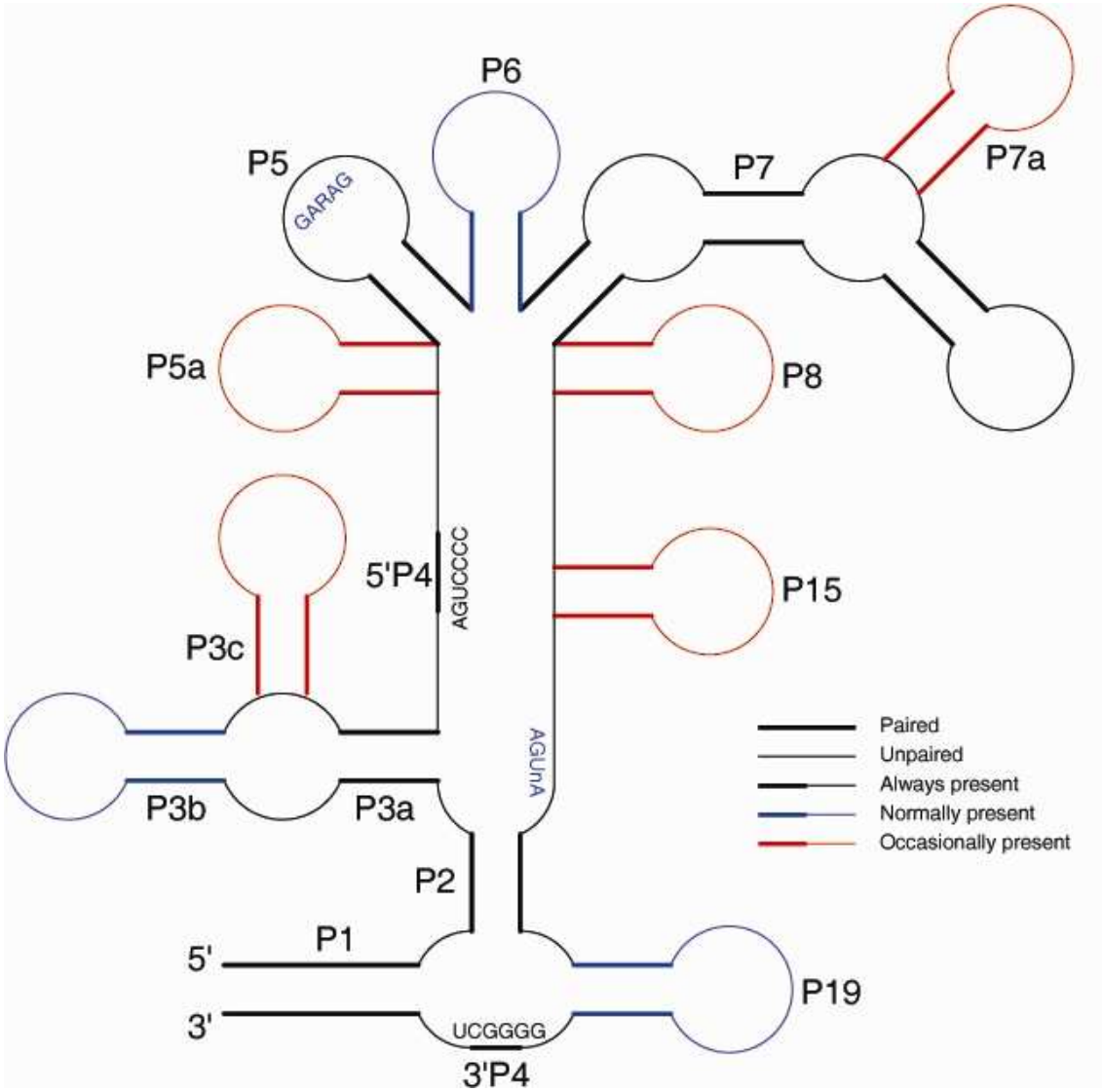
Evolution of RNase MRP and RNAse P




-200
+1
+100
Mammals -
Marsupials
SP1
Oct
PSE
T
RNase MRP
TT
Xenopus
SP1
Oct
T
RNase MRP
TT
Zebrafish
Oct
PSE
T
RNase MRP
TT
Pufferfish
Staf
PSE
T
RNase MRP
TT
Drosophila
PSE
T
RNase MRP
TT
Ciona
PSE ?
T
RNase
TT
Saccharomyces
T
RNase MRP
TT
B-box
Arabidopsis
USE
T
RNase MRP
TT
Apicomplexa
T
RNase MRP
TT



	TATA	PSE	Oct	SP1 / SPH
MRP	TATAAAA	TCACCCTAAT	ATTTGCAT	GGGCGGG
P	TATAAAA	TCACCCTAAC	ATTTGCAT	GGGCGGG
U6	TATATAT	TTACCGTAAC	ATTTGCAT	ATTTCCCATGATTCCTTCAT


P6
P5
GARAG
P7
P7a
P5a
P8
5'P4
AGUCCCC
P15
P3c
AGUnA
Paired
Unpaired
Always present
Normally present
Occasionally present
P3b
P3a
P2
P1
5'
3'
P19
UCGGGG
3'P4