J Mol Evol (2000) 51:194–204
DOI: 10.1007/s002390010081

JOURNAL OF MOLECULAR EVOLUTION

© Springer-Verlag New York Inc. 2000

Use of RNA Secondary Structure for Studying the Evolution of RNase P and RNase MRP

Lesley J. Collins, Vincent Moulton,* David Penny

Institute of Molecular BioSciences, Massey University, Private Bag 11222 Palmerston North, New Zealand

Received: 19 July 1999 / Accepted: 3 May 2000

Abstract. Secondary structure is evaluated for determining evolutionary relationships between catalytic RNA molecules that are so distantly related they are scarcely alignable. The ribonucleoproteins RNase P (P) and RNase MRP (MRP) have been suggested to be evolutionarily related because of similarities in both function and secondary structure. However, their RNA sequences cannot be aligned with any confidence, and this leads to uncertainty in any trees inferred from sequences. We report several approaches to using secondary structures for inferring evolutionary trees and emphasize quantitative tests to demonstrate that evolutionary information can be recovered. For P and MRP, three hypotheses for the relatedness are considered. The first is that MRP is derived from P in early eukaryotes. The next is that MRP is derived from P from an early endosymbiont. The third is that both P and MRP evolved in the RNA-world (and the need for MRP has since been lost in prokaryotes). Quantitative comparisons of the pRNA and mrpRNA secondary structures have found that the possibility of an organellar origin of MRP is unlikely. In addition, comparison of secondary structures support the identity of an RNase P–like sequence in the maize chloroplast genome. Overall, it is concluded that RNA secondary structure is useful for evaluating evolutionary relatedness, even with sequences that cannot be aligned with confidence.

Key words: RNase MRP — RNase P — RNA secondary structure — RNA-world — Catalytic RNA — Evolutionary trees — Covarion hypothesis

Introduction

RNase P (P) and RNase MRP (MRP) are ribonucleoproteins consisting of a catalytic RNA and at least one protein subunit. However, to date there has been little quantitative comparison of the secondary structures of their RNA (pRNA and mrpRNA). Because of the similarities in function and secondary structure pRNA and mrpRNA are suggested to be evolutionarily related (Forster and Altman 1990; Karwan 1993; Morrissey and Tollervery 1995). Despite this similarity of structure, the RNA components (pRNA and mrpRNA) have little sequence homology, and consequently there is little confidence in evolutionary relatedness inferred from sequence alignments. Qualitative comparison of the RNA secondary structures between mrpRNA and pRNA have shown similarity in shape, especially in the "cage region" of the RNA molecule, where there is a characteristic pseudoknot formation (Forster and Altman 1990). The similarities in secondary structure are possibly the direct result of conservation of tertiary (and thus functional) characteristics.

Functional similarities have also led to the conclusion that these two ribonucleoproteins (RNPs) are evolutionarily related (Morrissey and Tollervey 1995). Both the P and MRP ribozymes cleave RNAs to generate 5′ phosphate and 3′ hydroxyl termini in a reaction requiring divalent cations (Forster and Altman 1990). Both P and

**Present address:* Department of Physics and Mathematics, Mid Sweden University, Sundsvall, S 85170 Sweden
Correspondence to: D. Penny

MRP are sensitive to puromycin, an antibiotic that inhibits pre-tRNA processing (Potuschak et al. 1993), and enzymatic activities from P and MRP isolated from several organisms cofractionate through multiple stages of biochemical purification (Paluh and Clayton 1995). It has been reported that MRP and P may be involved together in a macromolecular complex within the nucleolus (Lee et al. 1996). This raises the possibility that the relationship between MRP and P may be of a functional nature, based on their sharing of many protein subunits (Sbisà et al. 1996).

The phylogenetic distributions of P and MRP are also informative. P cleaves tRNA precursors to form the mature 5′ ends of tRNA molecules, with activity being found in all cells tested, including prokaryotes, eukaryotes, and also in organelles. Prokaryotic P consists of an RNA strand and a single protein subunit, whereas the P encoded in the nucleus of eukaryotes has several protein subunits (Pace and Smith 1990). In one case it has been suggested that the RNA is lost and P activity is entirely due to proteins (Rossmanith and Karwan 1998). Yeast species such as *Saccharomyces cerevisiae* and *Aspergillus nidulans* have retained their mitochondrially encoded pRNA, whereas vertebrate and the yeast *Schizosaccharomyces pombe* mitochondria have lost their mitochondrial pRNA gene and use nuclear-encoded products. In plants, mitochondrial pRNA activity has been shown (Marchfelder and Brennicke 1993), but to date no genes have been characterized. RNase P genes have been identified in algal chloroplasts (see Turmel et al. 1999) but not yet in higher plant chloroplasts. The secondary structure of eubacterial pRNA has been reported to show characteristic features in different phylogenetic groups (Pace and Brown 1995) and consensus structures have been drawn for groups of eubacteria and of archaea (Pace and Brown 1995; Haas et al. 1996). For the purposes of this study, prokaryotic pRNA will include that from archaea, eubacteria, mitochondria, and chloroplasts.

MRP (mitochondrial ribosomal processing) has been found only in eukaryotes, initially as an endoribonuclease that cleaves RNA primers for the initiation of mitochondrial DNA replication (Morrissey and Tollervey 1995). Subsequently a nuclear function in rRNA processing was identified, consistent with its predominant localization to the nucleolus (Lygerou et al. 1996). MRP consists of an RNA moiety and multiple protein subunits with at least seven of these, Pop1p (Morrissey and Tollervey 1995), Pop3p (Dichtl and Tollervey 1997), Pop4 (Chu et al. 1997), Pop5p, Pop6p, Pop7p, and Pop8p (Chamberlain et al. 1998) being shared with P in the yeast *S. cerevisiae*. mrpRNA secondary structures (Schmitt et al. 1993) have been characterized for eight species and show great similarity with each other despite being from plant, yeast, and vertebrate species. Although the secondary structures are similar the nucleotide sequences vary greatly in length and nucleotide composition, making alignment difficult, even within the MRP group.

We consider three general hypotheses of the evolutionary relationships of MRP and P. These are as follows.

I. MRP Evolved from a Eukaryotic Nuclear P in the Nucleus of an Early Eukaryote

This could occur by gene duplication followed by divergence of function of the two homologues. This is the theory most commonly suggested in previous studies (Morrissey and Tollervey 1995; Reddy and Shimba 1996; Chamberlain et al. 1996). MRP would then have been recruited into multiple eukaryotic functions as well as into an essential function in mitochondria. Under this hypothesis MRP is found only in eukaryotes because it was never in any of the other lineages. MRP is present in animals, yeasts, and plants, indicating an early divergence from P, but would not necessarily have to be in all early eukaryotes. MRP would thus be a striking exception to the transfer of catalysis from RNA to RNP to protein (Jeffares et al. 1998) in that, even after the evolution of protein catalysts, a ribonucleoprotein evolved a new catalytic function. Under Hypothesis I, we would expect the secondary structures of the mrpRNA to be more similar to eukaryotic pRNA than to prokaryotic pRNA (archaeal or eubacterial).

II. MRP Evolved from an Endosymbiont P

There are several variants on this hypothesis. MRP could have evolved from the hypothetical endosymbiotic fusion that formed the first eukaryote (Gupta and Golding 1996; Martin and Muller 1998) or by a later event that led to the mitochondrion. This theory accounts for the essential mitochondrial function of MRP, but requires that MRP recruited additional rRNA processing functions in the nucleus. In plants it has been shown that organellar DNA can be transferred to the nucleus but retains a function in the organelle (Brennicke et al. 1993; Wischmenn and Schuster 1995; Blanchard and Schmidt 1995). It is possible that mrpRNA would retain some organellar characteristics, such as a higher A + T content in nucleotide sequence, and be more closely related in secondary structure to that of the organellar or prokaryotic pRNA. In contrast to Hypothesis I, the secondary structure of mrpRNA would be more similar to either eubacterial or archaeal pRNA than to eukaryotic pRNA, depending on the particular endosymbiotic event.

III. MRP and P Evolved in the RNA-World

The RNA-world hypothesis is that there was a stage before proteins and DNA evolved, when RNA was the

main catalytic and information storage molecule. Most of today's catalytic RNA species may be relics from this time (Jeffares et al. 1998). Three main criteria were used to evaluate the antiquity of an RNA molecule, and pRNA fits all three by being ubiquitous, catalytic, and central to metabolism. MRP on the other hand fits only the last two criteria, being present only in the eukaryotic lineage. During the transition from an RNA-world, proteins with their superior catalytic properties almost completely replaced RNA as the catalytic molecule. Conversely, no novel catalytic RNAs would be expected after the advent of efficient genetically encoded protein synthesis (Jeffares et al. 1998). It is difficult to understand how a molecule such as MRP could have evolved only in the eukaryotic lineage and then integrate itself so intimately into rRNA processing, mitochondrial genome replication, and perhaps other functions central to eukaryotic metabolism. In general it has been found that eukaryotes carry more proposed relics of the RNA-world than prokaryotes (Jeffares et al. 1998). MRP was the only widely occurring catalytic RNA not included as a relic from the RNA-world in Jeffares et al. (1998); its status was left unresolved.

There are also several variants of this third hypothesis; MRP could have evolved from P, P evolved from MRP, and MRP and P evolved independently in the RNA-world. With the possibility that MRP had a function in the RNA-world (before the advent of proteins and DNA) it is important to know more about the evolutionary relationship of P and MRP. Under this third class of hypotheses we expect that mrpRNA structures would join with the eukaryote pRNAs.

With such a large divergence expected between pRNA and mrpRNA (at least back to early eukaryotes) nucleotide sequence alignments may not be reliable enough to determine any evolutionary relationship with confidence. However, examination of the RNA secondary structure may yield the required information when the sequence data cannot. The secondary structure of the catalytic RNA molecule has fixed "motifs" (Pace and Smith 1990) that represent areas that are critical to maintaining the function, and other regions that are free to vary in presence or size. It has been shown that many sequences can fit the same secondary structure (Fontana et al. 1993), this allows the catalytic RNA sequence to vary even if the function of the molecule remains unchanged. Thus secondary structure may be useful in determining evolutionary relationships even when the sequence data cannot. Tertiary structure has been used to study evolution of proteins (Johnson et al. 1990; Bujnicki 2000).

Quantitative comparisons of secondary structures of pRNA and mrpRNA are used here to calculate distances between these molecules to assess their relatedness. As a preliminary test, we first compared trees from both sequence data and secondary structures from small rRNA subunits to test whether or not evolutionary information can be recovered. The results indicate that RNA secondary structure can be used to recover evolutionary information. A consequence of this finding is that standard evolutionary models (that assume every site always has its same rate over the entire tree) may need to be generalized to include more complex models. One such is the covarion model (Fitch 1971; Tuffley and Steel 1997), which allows individual sites to vary in rate as the secondary structure evolves.

Materials and Methods

Sequences, Alignments, and Structures. pRNA sequences and prokaryotic pRNA secondary structures were mainly obtained from the RNase P Database (Brown 1998). mrpRNA sequences were obtained from Genbank and the remaining secondary structures for pRNA and mrpRNA were obtained from the literature, references are given in Table 1. 16S rRNA sequences and secondary structures were obtained from the Ribosomal Database Project (RDP; Maidak et al. 1997). Sequence alignments were obtained for 16S (prokaryotes) and 18S (eukaryotes) rRNA using the Subalign programs at the RDP (http://rdp.life.uiuc.edu). Prior experience showed that ClustalX (Thompson et al. 1997), and especially Dialign (Morgenstern et al. 1996) and Divide and Conquer Algorithm (DCA; Stoye et al. 1997), were suitable for aligning distantly related RNA sequences (see also Hickson et al. 2000).

Distances. Genetic distances from aligned sequences were obtained using the DNAdist option from the Phylip package (Felsenstein 1989), with and without the Jukes Cantor correction for multiple substitutions. For secondary structures, two structural-distance measures were used. For 16S rRNA, homologous helices of domain I of the secondary structures (Gutell et al. 1994) were compared. The number of nucleotides within each stem, each loop, and each single-stranded region were determined for the 16 prokaryotic species in Table 1. These were the "characters" and the sum of the differences between each pair of sequences was used as the first distance measure.

The RNAdistance program in the Vienna RNA package (Hofacker et al. 1994) was used for pRNA and mrpRNA secondary structures. This second structure-distance measure computes the number of tree-edit operations required to convert one RNA structure into another (Shapiro and Zhang 1990). It was necessary to convert published structures (see Table 1) into bracket notation where a folded RNA structure is a string of parentheses and dots; ( ) for paired nucleotides and **.** for unpaired (Hofacker et al. 1994).

Trees. Subtrees for the RDP for 16S (prokaryotes) and 18S (eukaryotes) rRNA were obtained from the overall tree of life with the Subtree program (http:\\rdp.life.uiuc.edu). These were used as standards to compare with our trees based on mrpRNA and pRNA sequences or secondary structures. An advantage in using these subtrees is that they were constructed from more sequences than were available for secondary structure comparisons. A further subtree of 16S rRNA sequences from 13 prokaryotic species was used as a standard tree for the comparison of 16S rRNA secondary structure features.

Phylogenetic trees were inferred from both types of distances (from sequences and from secondary structures) using neighbor-joining in the Phylip package (Felsenstein 1989). Trees were also constructed using

Table 1. RNase P and RNase MRP RNA sequences used in this study showing length, accession details, A + T% and from where the secondary structures were obtained

	Accession number	Length of sequence	A + T%	Secondary structure reference
RNase P sequences				
Synechocystis sp. PCC6803	X65707	437	48	P
Anabaena sp. PCC 7120	X65648	465	47	P
Anacystis nidulans PCC6301	X63566	385	43	P
Pseudoanabaena sp. PCC 6903	X73135	450	52	P
Escherichia coli	M17569	377	38	P
Bacillus subtilis	M13175	401	51	P
Rhodospirillum rubrum	M59355	429	29	P
Agrobacterium tumefaciens	M59354	402	36	P
Thermotoga maritima	M64709	338	32	P
Reclinomonas americana mitochondria	AF007261	312	75	P
Porphyra purpurea chloroplast	U38804	383	63	P
Nephroselmis olivacea chloroplast	From: AF137379	408	55	Turmel et al. (1999)
Putative maize chloroplast	From: X86563 19430–19083	347	63	This paper
Archaeoglobus fulgidus	AE000782	248	35	P
Halobacterium cutirubrum	U42983	376	28	P
Methanococcus jannaschii	L77117	274	40	P
Sulfolobus acidocaldarius	L13597	315	52	P
Human (nuclear)	X15624	340	36	P
Mouse (nuclear)	L08802	288	33	P
Danio rerio (nuclear) zebrafish	U50408	308	43	—
Saccharomyces cerevisiae (nuclear)	M27035	368	48	Tranguch and Engelke (1993)
Schizosaccharomyces pombe (nuclear)	X04013	373	48	Tranguch and Engelke (1993)
RNase MRP sequences				
Human	X51867	264	36	Schmitt et al. (1993)
Bovine	Z25280	277	39	Schmitt et al. (1993)
Mouse	J03151	275	36	Schmitt et al. (1993)
Rat	J05014	273	35	Schmitt et al. (1993)
Xenopus (frog)	Z11844	277	45	Schmitt et al. (1993)
Arabidopsis thaliana	X65942	260	49	Kiss et al. (1992)
Saccharomyces cerevisiae	Z14231	339	60	Kiss et al. (1992)
Schizosaccharomyces pombe		399	57	Paluh and Clayton (1995)
16S rRNA structures	RDP sequence			
				RDP
Escherichia coli	*E.coli*	—	—	RDP
Clostridium innocuum	*C.innocuum*	—	—	RDP
Methanococcus vannielli	*Mc.vanniel*	—	—	RDP
Frankia sp.	*Fra.spORS*	—	—	RDP
Streptomyces coelicolor	*Stm.coelic*	—	—	RDP
Thermus thermophilus	*T.thermoph*	—	—	RDP
Bacillus subtilis	*B.subtilis*	—	—	RDP
Agrobacterium tumefaciens	*Ag.tumefac*	—	—	RDP
Spirochaeta aurantia	*Spi.aurant*	—	—	RDP
Thermoplasma acidophilum	*Tpl.acidop*	—	—	RDP
Mycoplasma capricolum	*M.capricol*	—	—	RDP
Methanobacterium formicicum	*Mb.formici*	—	—	RDP
Pseudomonas testosteroni	*Ps.testost*	—	—	RDP

P obtained from the RNase P Database (Brown 1998)
RDP obtained from the Ribosomal Database Project (Maidak et al. 1997)

the refined Buneman option in SplitsTree (Dress et al. 1996; Huson 1998). Refined Buneman trees have the advantage that, unlike neighbor-joining, they vary continuously on the distance matrix—that is, small changes in the matrix do not lead to large changes in the resulting tree (Moulton et al. 1997). In general the refined Buneman trees were either identical or very similar to neighbor-joining trees, so we do not present these here. All trees were drawn with TreeView (Page 1996).

Tree Comparisons. The trees from aligned sequence and from secondary structures were compared in two ways. The first is the partition metric, which counts the number of internal edges (branches) that two trees have in common. This value was then compared with the expected distribution for two random binary trees for the same number of taxa (Hendy et al. 1984), which is already known for random trees generated under different models (Steel and Penny 1993). Second, groupings of

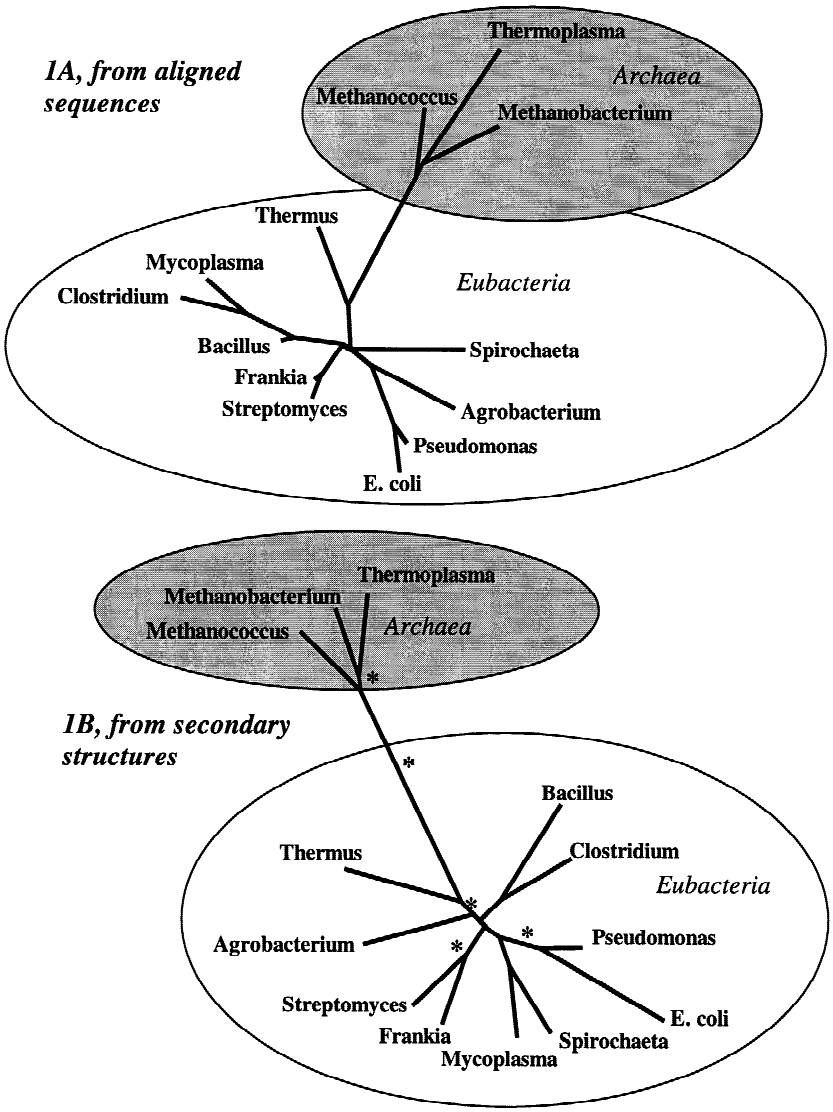

Fig. 1. Trees for 13 prokaryotes for which RNase P secondary structures have been studied. **A** Subtree of 16S rRNA eubacterial and archaeal sequences from the Ribosomal Database Project. **B** Neighbor-joining tree from secondary structure of domain 1 of the same 16S rRNA data. Identical internal branches in A and B are indicated by *.

predefined taxa (for example, archaea) were identified, and the significance of finding such groups calculated using theorem 1 in Carter et al. (1990).

Results

16S rRNA

We began by testing if evolutionary pattern could be detected quantitatively from characteristics of secondary structure. We used 16S rRNA because good-quality alignments and well-documented secondary structures are available from the RDP. Some confidence can be placed in trees from these aligned sequences, which are then used as a baseline for comparison with trees constructed only from secondary structures.

Figure 1 consists of a subtree taken from the RDP database (1A) and a tree for the same set of sequences calculated from distances between secondary structures (1B). About half of the internal branches are the same for these two trees, and this is significantly higher than the number expected for two random binary trees. In particular, for 13 taxa the probability of five identical branches on two randomly selected binary trees is $\approx 1.7 \times 10^{-4}$ (Hendy et al. 1984). Moreover, only two taxa are placed in different positions on the tree; *Agrobacterium tumefaciens* and *Mycoplasma capricolum*. If these two taxa are removed the two trees, apart from one internal branch, become identical. Another measure is the probability of a tree having the three archaea species together. From Carter et al. (1990) the probability with 13 taxa of getting 3 taxa correctly together is 0.0075, again a highly significant result. This is an important aspect of the tree to get correct because our interest is in recovering the oldest divergences; recent divergences should be better from sequences directly.

Thus from three measures we conclude that the simple

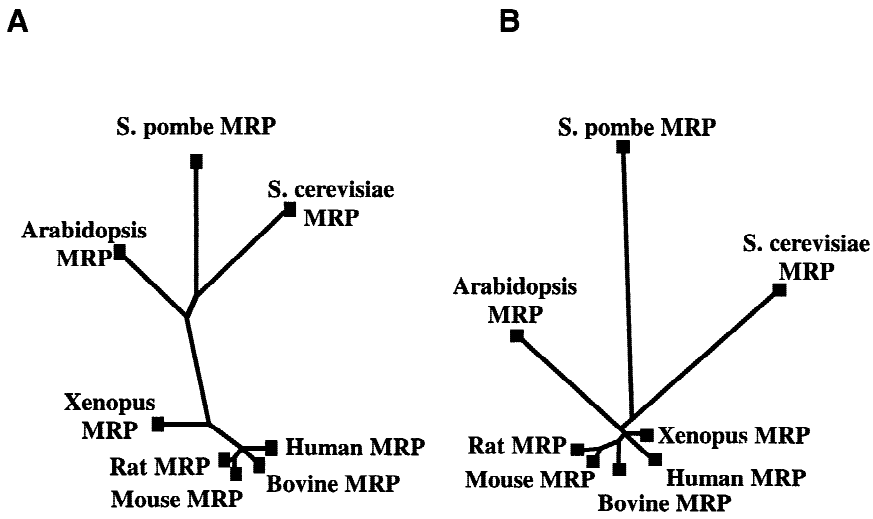

Fig. 2. Neighbor-joining trees for eight mrpRNA sequences whose secondary structure has been studied. **A** From a Divide and Conquer Algorithm (DCA) alignment. **B** From secondary structure distances calculated by RNAdistance. Deep divergence are recovered from structure distances, the poor resolution within vertebrates is of less concern because secondary structure is not expected to be informative for more recent divergences.

distances used to compare secondary structures for domain I of 16S rRNA has demonstrated that evolutionary information has been recovered, even by the relatively simple methods used to compare structures. For domain III of 16S rRNA the tree from secondary structures had 22% of internal branches the same as the sequence data tree though significantly more taxa were placed differently (trees not shown).

mrpRNA

Before studying the relationship between pRNA and mrpRNA we used just mrpRNA data to determine if secondary structure could detect deep divergences within eukaryotes, such as between the plant, yeast, and vertebrate species. All eight available mrpRNA sequences (five vertebrates, one plant, and two yeast) were aligned using the DCA and Dialign alignment methods. In this case, ClustalX did not cope well with the large number of internal gaps that were required to align the longer yeast sequences. The DCA alignment obtained was reasonable with some manual adjustment required. Neighbor-joining trees for both distance matrices (aligned sequences, and secondary-structure) are shown in Fig. 2. The trees are similar: They have 60% of internal branches in common. In both trees the deep divergence between the vertebrates, plants, and yeasts is apparent. However, the interrelationships within vertebrates differ but this is of less concern, we did not expect secondary structure to be so informative for more recent divergences.

pRNA and mrpRNA

Given that information can in principle be recovered from secondary structures, the final step was to estimate the evolutionary relationships between pRNA and mrpRNA. None of the available multiple alignment methods provided results in which we had any confidence. Not only were there problems aligning the mrpRNA and pRNA sequences, but even combining the eukaryotic nuclear and prokaryotic pRNA sequences did not give satisfactory alignments. DCA gave a reasonable alignment for just one set of eight sequences and a neighbor-joining tree was constructed from it, see Fig. 3. The alignment is at http://imbs.massey.ac.nz/Research/MolEvol/dcamrp.htm, and although it is reasonable, the dataset did not have enough sequences to test the relationships between pRNA and mrpRNAs.

In contrast to the difficulty in aligning these sequences, structural distances could be calculated for 29 eubacterial, mitochondrial, chloroplast, archael, and eukaryotic secondary structures. This includes a P-like sequence that had been identified in the maize chloroplast genome (Collins et al. 1999). This was first found from a very distant homology with the *Porphyra* chloroplast sequence and was in an unassigned region (ORF29) of the maize chloroplast genome (Table 1). The availability of the *Nephroselmis* chloroplast genome (Turmel et al. 1999) allowed the proposed secondary structure for maize to be refined further, and it is shown in Fig. 4. There is no biochemical evidence yet for this identification, so an additional test is whether this proposed secondary structure is similar to eubacterial structures. Given these 29 p and mrpRNA secondary structures, a neighbor-joining tree was calculated and shown in Fig. 5.

The first significant observation is that the tree has four separate groupings of the eubacterial (including mitochondrial and chloroplast), archaeal, eukaryotic pRNA, and mrpRNA structures. The only qualification is that the mrpRNA structures are within the eukaryotic structures, but the tree still has the minimum of three changes for four groups. It is highly significant that the tree has these four predefined groups separated. The four groups have 13, 4, 4, and 8 sequences, respectively, and the

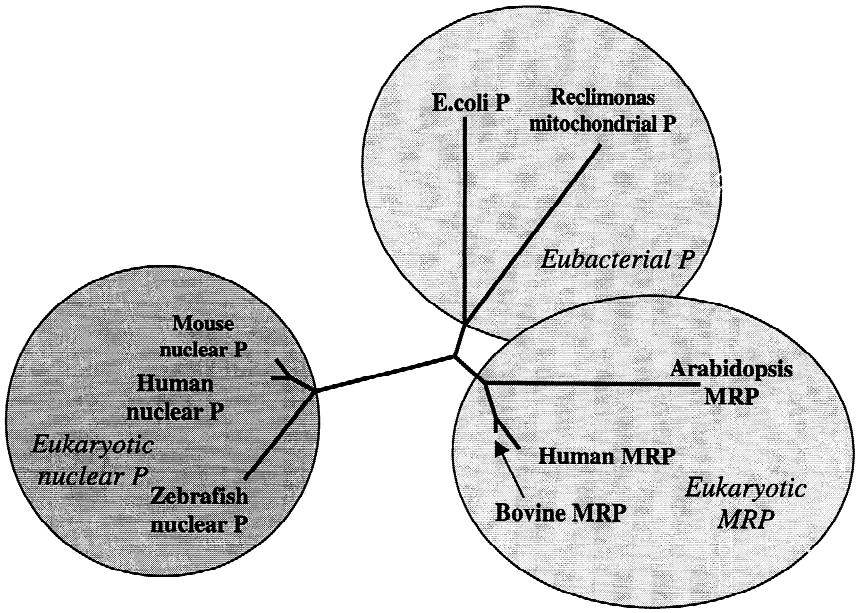

Fig. 3. Neighbor-joining tree for eight mrpRNA and pRNA sequences from a Divide and Conquer Algorithm (DCA) alignment (Stoye et al. 1997).

probability of a random tree for 30 taxa having a perfect fit (with the three minimum changes between categories) is $\approx 10^{-13}$ (Carter et al. 1990). Given this result, it is clear that distances between secondary structures contain evolutionary information, and that current methods recover some of it. There may be better methods and some errors on the tree (see later), but the main point for the present is that evolutionary information can be recovered from secondary structures.

Some other features of the tree merit comment. The subtree labeled by the sequences that we aligned with DCA (Fig. 3), but omitting the zebrafish, for which no secondary structure was available, gives the same groupings as the tree in Fig. 5. The *Bacillus* secondary structure is a little different from the consensus bacterial pRNA structure and is shown as such in the RNase P Database (Brown 1998). Thus it is not surprising that *Bacillus* is grouped away from other eubacteria. It is pleasing that the hypothetical P-like RNA from maize fits well within the eubacteria. The internal branching of the tree in Fig. 5 places the mrpRNA group closer to the eukaryotic pRNA group than to the prokaryotic pRNA group. It is certainly not clear from this data where the root should be positioned, the tree in Fig. 5 must be considered an unrooted tree.

Broadly speaking, the results contradict hypothesis II of an endosymbiotic origin of mrpRNA. There is good support from secondary structure for an evolutionary relationship between pRNA and mrpRNA and this contradicts the most extreme version of hypothesis III (that they have independent origins in an RNA-world). But whether MRP arose only in early eukaryotes (hypothesis I) or in the RNA-world (hypothesis III) cannot be decided just from this analysis of secondary structure. Our three hypotheses for the origin of MRP are considered below.

Discussion

In general, we find that quantitative analysis of secondary structures gives similar trees to those from easily aligned sequences, but allows the possibility for estimating trees when alignments are poor or unobtainable. Several quantitative tests demonstrate that the similarities of trees from sequence and structural data were highly significant, and that trees from structural data did recover expected groups such as archaea and eubacteria. Analysis of 16S rRNA Domain I (Fig. 1) indicates that a relatively simple analysis of RNA secondary structures can be useful in the analysis of ancient evolutionary relationships. Even a simple characteristic (such as the length of stems and loops) gave a good phylogenetic signal. A limitation is that structures (for example in the ribosomal database) are not normally available in bracket notation. Thus, quantitative analysis of the structure of whole 16S rRNA is not yet possible. Thus we cannot yet check whether complete ribosome structures give similar trees to those from sequences directly.

Because of problems of alignment, sequence analysis of pRNA and mrpRNA cannot find an evolutionary relationship between the two. However, in the trees constructed from secondary structures for mrpRNA and pRNA the prokaryotic pRNAs, the eukaryotic pRNAs, and mrpRNAs formed well-defined groups. This analysis failed to show a close relationship in secondary structure between mrpRNA and any of the prokaryotic pRNAs. On our analysis it is unlikely that that MRP is of an

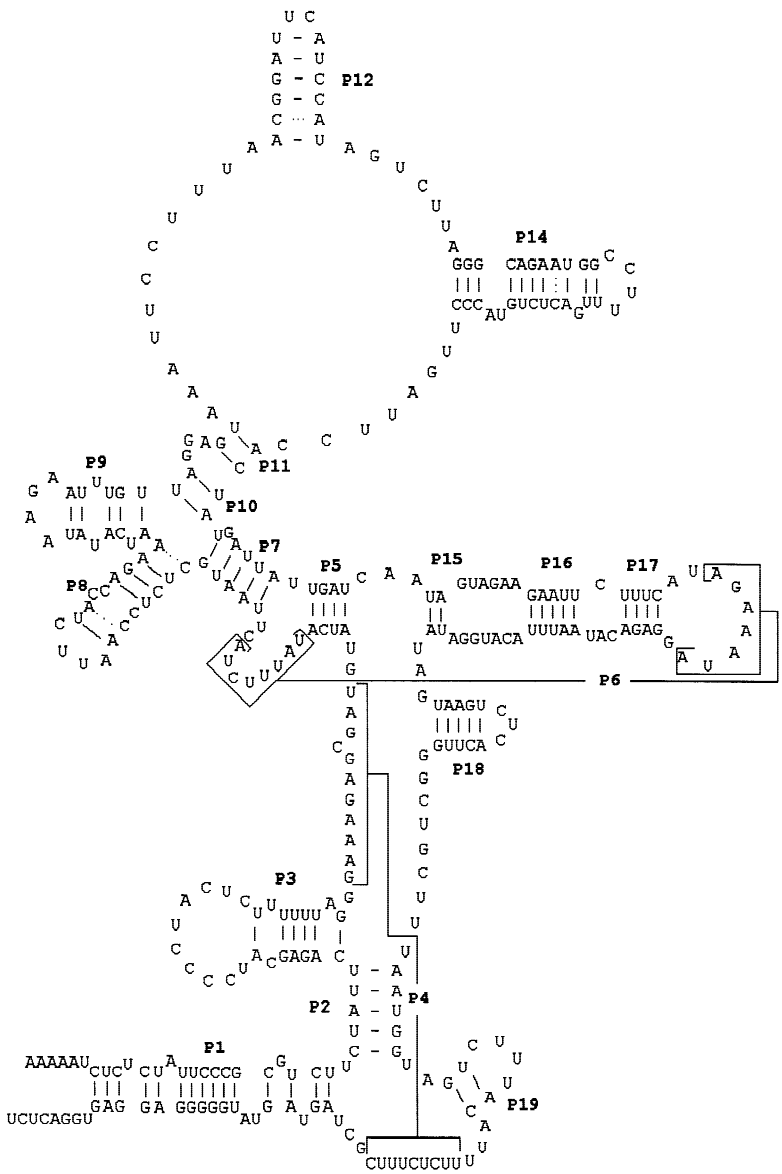

Fig. 4. Inferred secondary structure for an RNase P–like sequence from the maize chloroplast genome. The identification of this sequence as an RNase P is supported further by the position of its secondary structure in Fig. 5.

organellar/prokaryotic origin—our hypothesis II. Nor do the mrpRNA genes show any higher A+T% content that might possibly be expected with an organellar origin.

MRP is found in vertebrates, plants, and yeast; thus if MRP evolved from the nuclear P in the eukaryotic lineage it must be before the divergence of these three groups. In this context it will be interesting to determine which catalytic RNAs are found in the genomes, such as *Giardia* (http://evol3.mbl.edu/Giardia-HTML/giardia_data.html). Another indication of an early origin of MRP is its essential nature within the eukaryotic lineage, especially in the nucleus, where it is part of a cascade of RNA molecules processing other RNA (Poole et al. 1999). MRP must have evolved early enough to have a very different role from P in ribosome assembly. It also had to have evolved early enough to gain roles in the mitochondria, which in some species (e.g., vertebrates and *S. pombe*) are reliant on the nuclear pRNA and mrpRNA. Nevertheless, analysis of RNA secondary structure and functional data supports the relatedness of the RNases P and MRP, and they form a family of catalytic RNAs. For proteins, many families and superfamilies have been isolated, but there do not appear to be any other potential families of catalytic RNA molecules. Others are possible and may not be identifiable just on sequence homology data.

At present, sequence alignment and structural data alone are insufficient to determine whether hypothesis I or III is more likely—evolution early in the eukaryotic lineage or evolution in the RNA-world. If the tree constructed from structural data (Fig. 5) was rooted on the eubacterial lineage then MRP arose in eukaryotes; but recent results contradict this rooting (Lopez et al. 1999). However, if the RNA-world is used to root the tree of life (Poole et al. 1998) then the answer is unclear. But by taking functional data into account there are some factors favouring the RNA-world hypothesis (Poole et al. 1999). Under this hypothesis, and given their similarity in struc-

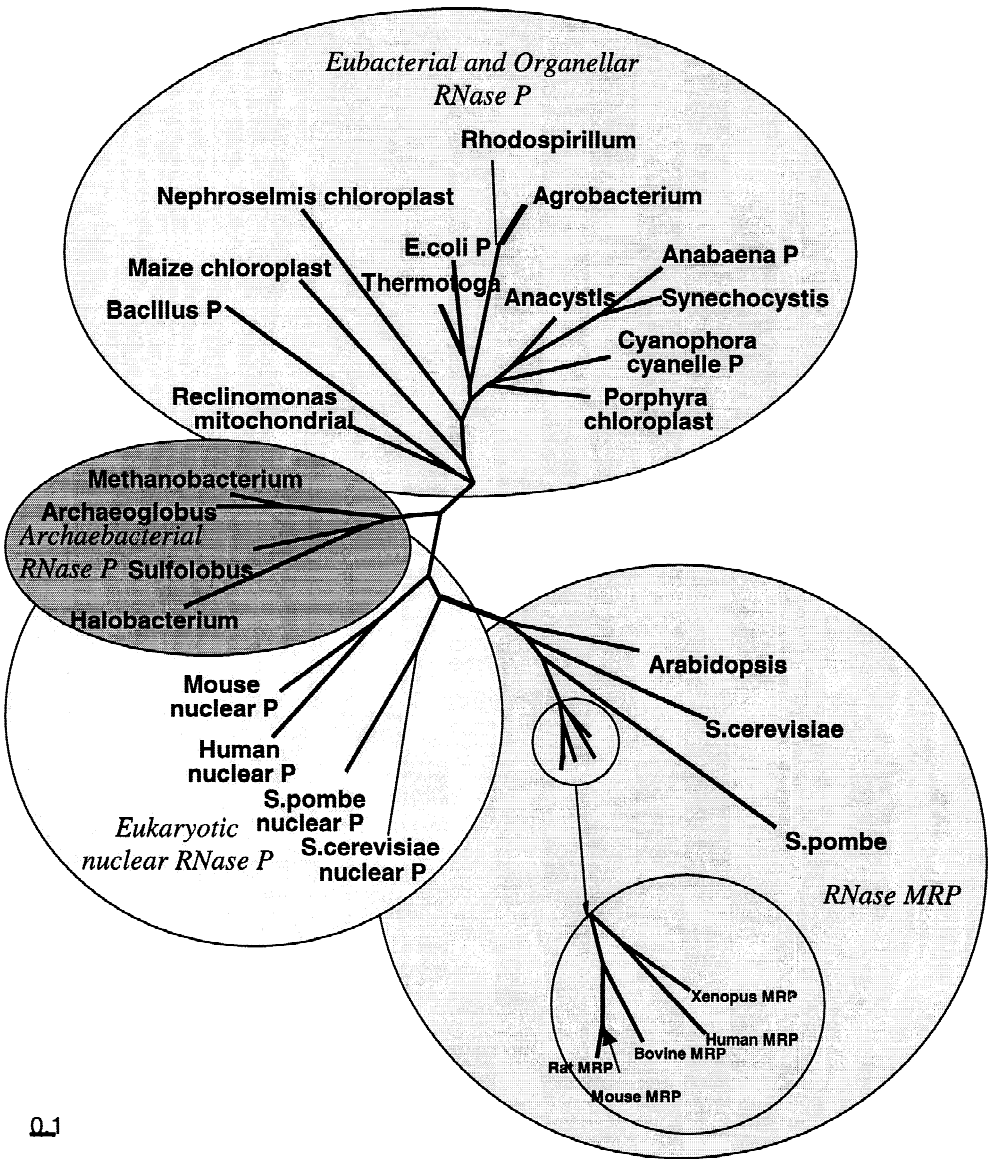

Fig. 5. Neighbor-joining tree constructed from tree-edit distances (from the program RNAdistance) between 29 pRNA and mrpRNA secondary structures. The four main groups of eubacterial, archaeal, and eukaryote P, plus eukaryote MRP are recovered.

tures, MRP could have evolved from P or vice versa. It is also possible that some of the similarities in secondary structure could have arisen from common protein binding sites. That multiple proteins are shared between MRP and the eukaryotic P (e.g., in *S. cerevisiae*—Pop1, Pop3, and Pop4) could be an indication of a relationship from the time when the RNA and protein moieties of MRP and P were first assembled. On the RNA-world hypothesis, the first proteins are expected to be RNA binding proteins with chaperone-like activity. These would increase stability of ribozyme tertiary structure (Poole et al. 1998). It is possible that MRP and P picked up chaperone-like proteins that had a function that was required by both (e.g., stability and transport) when there were few proteins.

Ribosomal proteins, for example, have been found to be multifunctional, with most of these proteins having functions additional to their role in the ribosome (Wool 1996). This co-opting of a few ancestral proteins by multiple processes may have occurred with P and MRP. Prokaryotic P is specialised for one function, the maturation of tRNA, whereas eukaryotic P and MRP are involved in many functions in the cell. Each interaction between the RNA and protein subunits (and also between the RNA and each substrate), would have been optimized to such a point that even with high Mg^{2+} concentrations, the RNA is no longer stable without chaperone-like proteins. Prokaryotic RNase P, and eukaryotic P and MRP, have evolved considerably since the first ribonucleoprotein complex. Under the RNA-world hypothesis these ancient P and MRP molecules would have had a stable RNA and then gained multiple protein subunits as protein synthesis evolved.

The present results show that secondary structures are

a valuable source of phylogenetic signal. Future studies should improve techniques by increasing our knowledge of methods for accurately comparing structures. For example, the RNAdistance program might have to be modified so that it takes the tree-like nature of secondary structure fully into account (R. Giegerich, personal communication), and new metrics for comparing RNA structures will be useful (see Moulton et al. 2000). Tertiary structures may have an even greater potential for the determination of evolutionary relationships. They have an even closer relationship to the function of the molecule than secondary structure does, and tertiary structure has been used with proteins (see Bujnicki 2000). At present there are two models for the tertiary structure of pRNA and none yet for mrpRNA (Pace and Brown 1995). It is expected that analysis of tertiary structure will be more revealing for distantly related structures, and primary sequence analysis more useful for closer, less diverged structures (Gutell 1992).

It is clear from the present results (and many earlier results) that secondary structure of RNAs diverge with time. It is now clear that this divergence in structure allows evolutionary information to be recovered. An important consequence is that this conclusion contradicts current mathematical methods for inferring (correcting) the number of multiple changes to sequences over time. Current methods assume that a site always has the same rate of change—even though there may be a distribution of rates, perhaps described by a Gamma distribution (see Swofford et al. 1996). Clearly, with rRNA and p and mrpRNA there is a change in secondary structure with time, consequently the constraints on the evolution of a particular site will vary between lineages.

The covarion model of Fitch (1971) handles cases where sites vary on their constraints over time, especially an implementation using a hidden Markov model requiring only two additional parameters (Tuffley and Steel 1997). Similarly, protein evolution shows a marked divergence in tertiary structure, especially at high sequence divergences (Chothia and Lesk 1986). Lockhart et al. (1998) report a quantitative test that rejects any model that assumes each site always has the same constraints over the whole tree. In the present context our interest is simply that secondary structure can be used to recover evolutionary information, but our results imply that the models of molecular evolution need to take into account the observation that constraints on particular sites vary with time.

In summary, although sequence alignments of pRNA and mrpRNA were sometimes obtained, because of the low homology in the sequences little confidence was placed in the trees inferred from them. However, quantitative analysis of secondary structure data offers an alternative for evaluating these trees, as well as for studying deep divergences when alignment was not possible. If MRP evolved in eukaryotes then it seems that a RNP took on a catalytic function in preference to a protein, an exception to the general process of the transfer of catalysis. An RNA-world origin of MRP, however, allows a new perspective in the analysis of this molecule.

Acknowledgments. This work was supported by the New Zealand Marsden Fund and by the Swedish National Research Council (NFR, grant M 12342-300).

References

Blanchard J, Schmidt G (1995) Pervasive migration of organellar DNA to the nucleus in plants. J Mol Evol 41:397–406

Brennicke A, Grohmann L, Hiesel R, Knoop V, Schuster W (1993) The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants. FEBS Lett 325:140–145

Brown J (1998) The ribonuclease P database. Nucleic Acids Res 26: 351–352

Bujnicki JM (2000) Phylogeny of the restriction endonuclease-like superfamily inferred from comparison of protein structures. J Mol Evol 50:39–44

Carter M, Hendy MD, Penny D, Székely LA, Wormald NC (1990) On the distribution of lengths of evolutionary trees. SIAM J Disc Math 3:38–47

Chamberlain J, Pagan-Rámos E, Kindelberger D, Engelke D (1996) An RNase P RNA subunit mutation affects ribosomal RNA processing. Nucleic Acids Res 24:3158–3166

Chamberlain J, Lee Y, Lane W, Engelke D (1998) Purification and characterization of the nuclear RNase P holoenzyme complex reveals extensive subunit overlap with RNase MRP. Genes Dev 12: 1678–1690

Chothia C, Lesk AM (1986) The relation between the divergence of sequence and structure in proteins. EMBO J 5:823–826

Chu S, Zengel J, Lindahl L (1997) A novel protein shared by RNase MRP and RNase P. RNA 3:382–391

Collins LJ, Moulton V, Penny D (1999) RNA secondary structure as an identification tool in the identification of putative pRNA sequences in the chloroplasts of four green plant species. MidSweden University, Department of Mathematics, no 3

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

Dress A, Huson D, Moulton V (1996) Analyzing and visualizing sequence and distance data using SplitsTree. Discr Appl Math 71: 95–110

Felsenstein J (1989) PHYLIP—Phylogeny inference package (version 3.2). Cladistics 5:164–166

Fitch WM (1971) Rate of change of concomitantly variable codons. J Mol Evol 1:84–96

Fontana W, Konings D, Stadler P, Schuster P (1993) Statistics of RNA secondary structures. Biopolymers 33:1389–1404

Forster A, Altman S (1990) Similar cage-shaped structures for the RNA components of all ribonuclease P and ribonuclease MRP enzymes. Cell 62:407–409

Gupta R, Golding G (1996) The origin of the eukaryotic cell. TIBS 21:166–171

Gutell R (1992) Evolutionary characteristics of 16S and 23S rRNA structures. In: Hartman H, Matsumo K (eds) The origin and evolution of prokaryotic and eukaryotic cells. World Scientific Publishing Co, New York, NY, p 243

Gutell R, Larsen N, Woese C (1994) Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. Microbiol Rev 58:10–26

Haas E, Banta A, Harris J, Pace N, Brown J (1996) Structure and

evolution of ribonuclease P RNA in Gram-positive bacteria. Nucleic Acids Res 24:4775–4782

Hendy MD, Little CHC, Penny D (1984) Comparing trees with pendant vertices labeled. SIAM J Appl Math 44:1054–1067

Hickson RE, Simon C, Perrey SW (2000) The performance of several multiple-sequence alignment programs in relation to secondary-structure features for an rRNA sequence. Mol Biol Evol 17:530–539

Hofacker I, Fontana W, Stadler P, Bonhoeffer L, Tacker M, Schuster P (1994) Fast folding and comparison of RNA secondary structures. Monatshefte fur Chemie 125:167–188

Huson D (1998) Splitstree—a program for analysing and visualizing evolutionary data. Bioinformatics 14:68–73

Jeffares D, Poole A, Penny D (1998) Relics from the RNA world. J Mol Evol 46:18–36

Johnson MS, Sutcliff MJ, Blundell TL (1990) Molecular anatomy: phyletic relationships derived from three-dimensional structures of proteins. J Mol Evol 30:43–59

Karwan R (1993) RNase MRP/RNase P: a structure function relationship conserved in evolution. FEBS Lett 39:1–4

Kiss T, Marshallsay C, Fillpowicz W (1992) 7-2/MRP RNAs in plant and mammalian cells: association with higher order structures in the nucleus. EMBO J 11:3737–3746

Lee B, Matera A, Ward D, 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

Lockhart PJ, Steel MA, Barbrook AC, Huson D, Charleston MA, Howe CJ (1998) A covariotide model explains apparent phylogenetic structure of oxygenic photosynthetic lineages. Mol Biol Evol 15: 1183–1188

Lopez P, Forterre P, Philippe H (1999) A method for extracting ancient phylogenetic signal: the rooting of the universal tree of life based on elongation factors. J Mol Evol 49:496–508

Lygerou Z, Allmang C, Tollervey D, Seraphin B (1996) Accurate processing of a eukaryotic precursor ribosomal RNA by Ribonuclease MRP in vitro. Science 272:268–270

Maidak B, Olsen G, Larsen N, Overbeek R, McCaughey M, Woese C (1997) The RDP (Ribosomal Database Project). Nucleic Acids Res 25:109–111

Marchfelder A, Brennicke A (1993) Plant mitochondrial RNase P and *E. coli* RNase P have different substrate specificities. Biochem Mol Biol Intern 29:621–633

Martin W, Muller M (1998) The hydrogen hypothesis for the first eukaryote. Nature 392:37–41

Morgenstern B, Dress A, Werner T (1996) Multiple DNA and protein sequence alignment based on segment-to-segment comparison. Proc Natl Acad Sci USA 93:12098–12103

Morrissey JP, Tollervey D (1995) Birth of the snoRNPs: the evolution of RNase MRP and the eukaryotic pre-rRNA-processing system. TIBS 20:78–82

Moulton V, Steel M, Tuffley C (1997) Dissimilarity maps and substitution models. Proc DIMACS Workshop Math Hierarchies 37:111–131

Moulton V, Zuker M, Steel M, Pointon M, Penny D (2000) Metrics on RNA secondary structure. J Comput Biol (in press)

Pace N, Brown J (1995) Evolutionary perspective on the structure and function of Ribonuclease P, a ribozyme. J Bact 177:1919–1928

Pace N, Smith D (1990) Ribonuclease P: function and variation. J Biol Chem 265:3587–3590

Page R (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. CABIOS 12:357–358

Paluh J, Clayton D (1995) *Schizosaccharomyces pombe* RNase MRP RNA is homologous to metazoan RNase MRP RNAs and may provide clues to interrelationships between RNase MRP and RNase P. Yeast 11:1249–1264

Poole A, Jeffares D, Penny D (1998) The path from the RNA world. J Mol Evol 46:1–17

Poole AM, Jeffares DC, Penny D (1999) Prokaryotes, the new kids on the block. BioEssays 21:880–889

Potuschak T, Rossmanith W, Karwan R (1993) RNase MRP and RNase P share a common substrate. Nucl Acid Res 21:3239–3243

Reddy R, Shimba S (1996) Structural and functional similarities between MRP and RNase P. Mol Biol Rep 22:81–85

Rossmanith W, Karwan R (1998) Characterization of human mitochondrial RNase P: novel aspects in tRNA processing. Biochem Biophys Res Commun 247:234–241

Sbisà E, Pesole G, Tullo A, Saccone C (1996) The evolution of the RNase P- and RNase MRP-associated RNAs: phylogenetic analysis and nucleotide substitution rate. J Mol Evol 43:46–57

Schmitt M, Bennett J, Dairaghi D, Clayton D (1993) Secondary structure of RNase MRP RNA as predicted by phylogenetic comparison. FASEB J 7:208–213

Shapiro B, Zhang K (1990) Comparing multiple secondary structures using tree comparison. CABIOS 6:309–318

Steel MA, Penny D (1993) Distributions of tree comparison metrics—some new results. Syst Biol 42:126–141

Stoye J, Dress A, Moulton V (1997) DCA: an efficient implementation of the divide-and-conquer approach to simultaneous multiple sequence alignment. CABIOS 13:625–626

Swofford DL, Olsen GJ, Waddell PJ, Hillis DM (1996) Phylogenetic inference. In: Hillis DM, Moritz C, Mable B (eds) Molecular systematics. Sinauer Associates, Sunderland, MA, pp 407–514

Thompson J, Gibson T, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl Acids Res 25:4876–4882

Tranguch A, Engelke D (1993) Comparative structural analysis of nuclear RNase P RNAs from yeast. J Biol Chem 268:14045–14053

Tuffley C, Steel MA (1997) Modeling the covarion hypothesis of nucleotide substitution. Math BioSci 147:63–91

Turmel M, Otis C, Lemieux C (1999) The complete chloroplast DNA sequence of the green alga *Nephroselmis olivacea:* insights into the architecture of ancestral chloroplast genomes. Proc Natl Acad Sci USA 96:10248–10253

Wischmann C, Schuster W (1995) Transfer of *rps10* from the mitochondrion to the nucleus in *Arabidopsis thaliana:* evidence for RNA-mediated transfer and exon shuffling at the integration site. FEBS Lett 374:152–156

Wool IG (1996) Extraribosomal functions of ribosomal proteins. TIBS 21:164–165