Structural implications of novel diversity in eucaryal RNase P RNA

STEVEN M. MARQUEZ,[1] J. KIRK HARRIS,[2] SCOTT T. KELLEY,[4] JAMES W. BROWN,[5] SCOTT C. DAWSON,[3] ELISABETH C. ROBERTS,[1] and NORMAN R. PACE[1]

[1]Department of Molecular, Cellular and Developmental Biology, University of Colorado at Boulder, Boulder, Colorado 80309-0347, USA
[2]Graduate Group in Microbiology and [3]Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, California 94720, USA
[4]Department of Biology, San Diego State University, San Diego, California 92182, USA
[5]Department of Microbiology, North Carolina State University, Raleigh, North Carolina 27695-7615, USA

ABSTRACT

Previous eucaryotic RNase P RNA secondary structural models have been based on limited diversity, representing only two of the ~30 phylogenetic kingdoms of the domain Eucarya. To elucidate a more generally applicable structure, we used biochemical, bioinformatic, and molecular approaches to obtain RNase P RNA sequences from diverse organisms including representatives of six additional kingdoms of eucaryotes. Novel sequences were from acanthamoeba (*Acathamoeba castellanii, Balamuthia mandrillaris, Filamoeba nolandi*), animals (*Caenorhabditis elegans, Drosophila melanogaster*), alveolates (*Theileria annulata, Babesia bovis*), conosids (*Dictyostelium discoideum, Physarum polycephalum*), trichomonads (*Trichomonas vaginalis*), microsporidia (*Encephalitozoon cuniculi*), and diplomonads (*Giardia intestinalis*). An improved alignment of eucaryal RNase P RNA sequences was assembled and used for statistical and comparative structural analysis. The analysis identifies a conserved core structure of eucaryal RNase P RNA that has been maintained throughout evolution and indicates that covariation in size occurs between some structural elements of the RNA. Eucaryal RNase P RNA contains regions of highly variable length and structure reminiscent of expansion segments found in rRNA. The eucaryal RNA has been remodeled through evolution as a simplified version of the structure found in bacterial and archaeal RNase P RNAs.

Keywords: eucaryal RNase P; comparative analysis; RNA structure

INTRODUCTION

Ribonuclease P (RNase P) is the ribonucleoprotein enzyme that catalyzes the removal of 5′ leader sequences from precursor-tRNAs (pre-tRNA) to produce the mature 5′-ends of the tRNAs. The biological significance of RNase P is underscored by its universal presence in all three phylogenetic domains, Bacteria, Archaea, and Eucarya, as well as the two major organelles, mitochondria and chloroplasts (Dang and Martin 1993; Pace and Brown 1995; Frank and Pace 1998; Pitulle et al. 1998; Cordier and Schon 1999). RNase P comprises a single RNA moiety and one or more associated proteins, ranging from a single protein in Bacteria, to at least four proteins in Archaea (Brown and Pace 1992; Hall and Brown 2002) and as many as 10 proteins in Eucarya (Chamberlain et al. 1998; Xiao et al. 2002). Bacterial and several archaeal RNase P RNAs are able to process pre-tRNA in vitro in the absence of any protein under conditions of high ionic strength (Guerrier-Takada et al. 1983; Pannucci et al. 1999). In contrast, none of the eucaryal RNase P RNAs is catalytically active in the absence of protein. Yet, the sensitivity of the eucaryal holoenzymes to micrococcal nuclease digestion suggests that the RNA subunit is vital for enzymatic activity (Krupp et al. 1986; Jayanthi and Van 1992). Moreover, clear homology with the bacterial and archaeal RNAs indicates that, although catalytically inactive in vitro, the eucaryal RNA is the catalytic subunit of the holoenzyme.

To determine what sequences or structures of the eucaryal RNase P RNA distinguish it from the bacterial and archaeal versions, a generally applicable secondary structure model of the eucaryal RNase P RNA is necessary. A general secondary structure model could identify structural features that might explain the catalytic inactivity of the eucaryal

RNA. An accurate secondary structure is also essential to guide study of the interplay between the RNA and protein components of the eucaryal holoenzyme.

A general approach for the determination of higher-order structure of large RNAs is phylogenetic-comparative sequence analysis (Fox and Woese 1975; Woese and Pace 1993). The accuracy of this technique has been confirmed by the correspondence of structures obtained through comparative analysis and crystallographic studies of a variety of RNAs, including tRNA and ribosomal RNAs (RajBhandary et al. 1966; Fox and Woese 1975; Noller et al. 1981; Ban et al. 2000). Phylogenetic comparative analyses of many diverse archaeal and bacterial RNase P RNA sequences have resulted in detailed models of their secondary structures (Harris et al. 2001). Derivation of a comparably satisfactory structure for the eucaryal RNase P RNA has not been possible, however, because sequences of only limited diversity, two of the ~30 eucaryotic kingdoms, were available. Previous comparative analyses of eucaryal RNase P RNAs successfully refined the secondary structures within the animal (vertebrates only) and fungal kingdoms (Tranguch and Engelke 1993; Chen and Pace 1997; Pitulle et al. 1998; Frank et al. 2000). Nevertheless, comparisons of such a phylogenetically limited representation of sequences could not yield a general model for the eucaryal RNase P RNA. Sequences from a more diverse array of organisms, from other phylogenetic kingdoms, are required.

We report a substantial expansion in the known phylogenetic diversity of eucaryal RNase P RNA. We purified, cloned, and sequenced the genes from *Acanthamoeba castellanii* and *Giardia intestinalis*. Several other new eucaryal RNase P RNA sequences were identified by genomic database searches. This diverse collection of sequences allowed us to identify widely conserved sequences, which were then used to design general PCR primers for amplification of RNase P RNA genes from representatives of the most evolutionarily diverse eucaryotes. The new sequences were aligned with those of other known eucaryal RNase P RNAs on the basis of structure predictions, and statistical covariation analyses were conducted. The results of the comparative analyses have significant implications with regard to eucaryotic RNase P RNA structure and evolution.

RESULTS

Isolation and sequence analysis of *A. castellanii* and *G. intestinalis* RNase P RNA

Attempts to obtain diverse RNase P RNA genes by PCR using primers complementary to conserved animal and fungal sequences were unsuccessful outside of those phylogenetic groups (data not shown). Consequently, to obtain a more diverse RNase P RNA sequence from another phylogenetic group, we biochemically isolated and determined the sequence of the RNA from *A. castellanii*, a representative

of the acanthamoebae kingdom-level phylogenetic group (Amaral Zettler et al. 2000). Three chromatographic steps were used: two consecutive ion-exchange columns (DEAE-cellulose with step elution, followed by DEAE-Toyopearl with gradient elution) and gel-filtration (Sephacryl S-500) (Materials and Methods).

Samples from the DEAE-Toyopearl column were assayed for RNase P activity. RNAs from aliquots of these same fractions then were end-labeled and the RNAs that cofractionated with activity were identified by gel electrophoresis (Fig. 1). Three RNAs, with approximate sizes of 290, 250, and 170 nucleotides (nt), were identified as possible *A. castellanii* RNase P RNAs. RNAs that coincided with activity from the DEAE-Toyopearl column, as well as the control RNAs that did not coincide with activity, were purified by gel filtration and isolated by denaturing gel electrophoresis. The RNA was then used for primer extension and sequence analysis (Materials and Methods).

Based on an alignment of animal and fungal Conserved Region V (CR V) sequences (Chen and Pace 1997), several different complementary oligonucleotides were designed and tested by primer extension analysis on the isolated RNAs (Materials and Methods). The 290-, 250-, and 170-nt RNAs all served as templates for reverse transcription with primer CRV*, consistent with RNase P RNA identity of the abundant RNAs in chromatographic fractions with RNase P activity. Sequencing revealed that the entire 170-nt sequence was contained in the 290-nt RNA. The 250-nt RNA was not sequenced by primer extension. Presumably the RNase P RNA was fragmented in vivo or during purification, and the 250- and 170-nt RNAs represent 3′-terminal fragments.

Using the sequences of the primer extension products, we designed primers to clone and sequence the genomic *A. castellanii* RNase P RNA gene using ligation-mediated PCR (Materials and Methods). The putative *A. castellanii* RNase P RNA contains sequences consistent with RNase P, including helices P1-P4 and eP5/7-P10/11, which are signature helices of RNase P RNAs, and sequences that correspond to CRs I–V. Finally, the overall secondary structure predicted for the *A. castellanii* RNA was similar to those of known eucaryotic RNase P RNAs (Fig. 2 will be discussed fully below). All of these data, taken together with chromatographic association with the biochemical activity, indicate that the RNA characterized is the RNase P RNA from *A. castellanii*. An oligonucleotide complementary to CR IV, when used as a probe in a hybridization analysis with *A. castellanii* genomic DNA, hybridized to a single genomic DNA fragment, indicating that a single-copy gene encodes RNase P RNA in this organism (data not shown).

To further substantiate the secondary structure of *A. castellanii* RNase P RNA, the corresponding genes from other members of the acanthamoebae kingdom, *Balamuthia mandrillaris*, *Filamoeba nolandi*, and an unnamed *Filamoeba* isolate (*Filamoeba* sp.) were cloned by PCR using
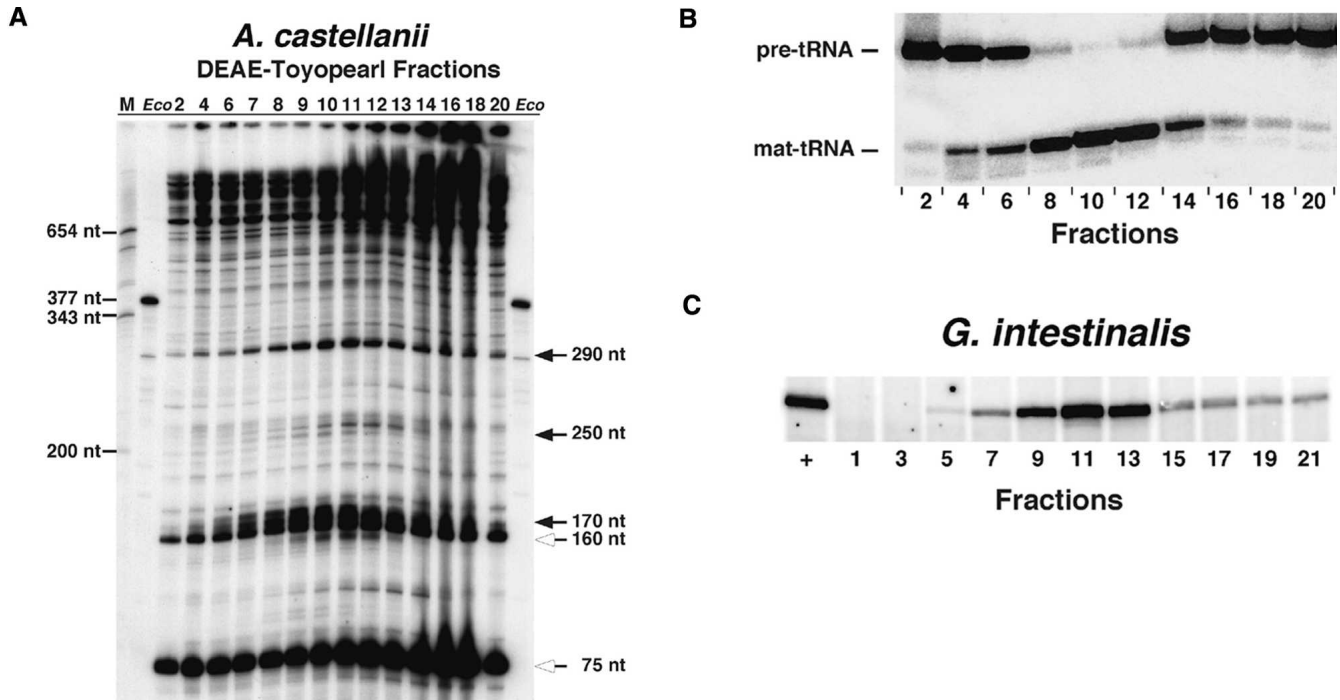
FIGURE 1. RNA molecules isolated from individual gradient DEAE-Toyopearl fractions. (*A*) RNAs isolated from DEAE-Toyopearl fractions were 3′-end labeled with [5′-^{32}P]pCp by T4 RNA ligase, and separated on a 6% denaturing polyacrylamide gel. M (RNA markers); (*Eco*) 3′-end labeled *E. coli* RNase P RNA. Black arrows identify *A. castellanii* RNAs that roughly coincide with enzyme activity. Approximate molecular sizes are indicated. White arrows identify RNAs that did not correlate with activity and were used as negative controls in primer extension experiments. (*B*) *A. castellanii* RNase P activity distribution from the column shown in panel *A*. Samples from the column shown in panel *A* were assayed with ^{32}P-pre-tRNA as described in Materials and Methods. The products were resolved on a 6% sequencing gel. The *upper* band in the autoradiogram is the substrate (pre-tRNA) and the *lower* band is the product (mature-tRNA). (*C*) Primer extension analysis of putative *G. intestinalis* RNase P RNA. Primer extension by avian myeloblastosis virus reverse transcriptase was carried out with ^{32}P-labeled oligonucleotide G120R and purified *G. intestinalis* RNAs as described in Materials and Methods, and the products were resolved on a 6% sequencing gel. RNAs that served as templates are as follows: lanes, (+) in vitro transcribed putative *G. intestinalis* RNase P RNA; primer extension products using RNAs purified from DEAE-Toyopearl fractions 1–21.

primers based on the *A. castellanii* RNase P RNA and sequenced. Pair-wise sequence identities between the *A. castellanii* RNase P RNA and those of *B. mandrillaris*, *F. nolandi*, and *F. sp.* sequences were 60%–68%, a good level of variation for testing pairing possibilities in the secondary structure (Pace et al. 1989). We obtained two different RNase P RNA sequences from a nominally pure culture of *F. nolandi* and designated these sequences (87% sequence identity) *F. nolandi* 1 and *F. nolandi* 2. Identical structure and highly similar sequences between the two RNAs would be consistent with the occurrence of both genes in the *F. nolandi* genome, which would be unprecedented. However, the result is also consistent with heterogeneity in the *F. nolandi* stock culture. We have not investigated this further. Comparisons of all the new sequences, upon alignment with the *A. castellanii* RNA, provided further covariation evidence and allowed refinement of the secondary structure model of the *A. castellanii* RNA sequence shown in Figure 2.

Analysis of RNase P RNAs from the most phylogenetically diverse organisms can provide perspective on the possible extent of variation in the RNA. We used degenerate PCR primers (below) based on known CR I and CR V RNase P RNA sequences to amplify a partial gene from the genomic DNA of *G. intestinalis*. From the sequence of the PCR product, full-length sequence was identified from the *G. intestinalis* genomic database (Materials and Methods). To corroborate that the *G. intestinalis* PCR product indeed corresponded to the RNase P RNA gene, we biochemically purified *G. intestinalis* RNase P holoenzyme, as described above, and performed primer extension analysis on RNAs in DEAE-Toyopearl column fractions with RNase P activity. RNA in fractions that correlated with activity served as template for primer extension with a primer derived from the sequence of the *G. intestinalis* PCR product. The lengths of these reverse transcription products were the same length as the reverse transcription product of the putative full-length *G. intestinalis* RNase P RNA gene, transcribed in vitro (Fig. 1C). Together, these results indicate that the PCR product identified as the gene for *G. intestinalis* RNase P RNA on the basis of secondary structure and CR sequences correlates with activity and therefore likely is the *G. intestinalis* RNase P RNA gene (Discussion). To gather further evidence for the secondary structure model of the RNA, the correspond-
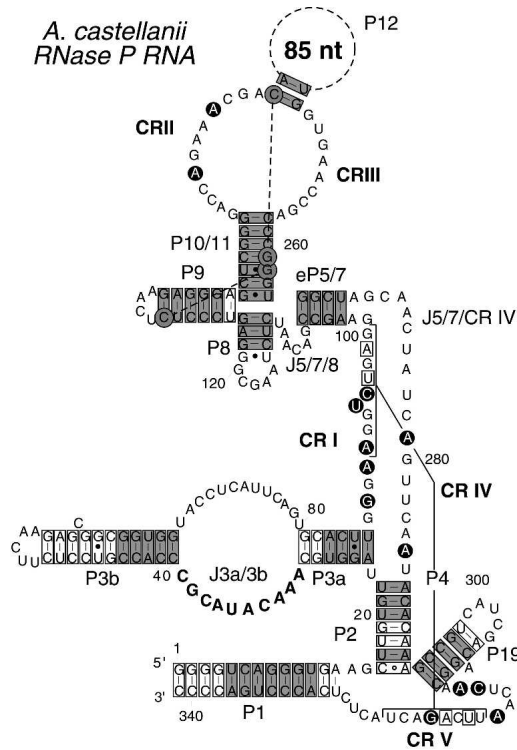
FIGURE 2. Secondary structure of *A. castellanii* RNase P RNA. Helices are labeled 5′ to 3′, P1–P19 as defined for bacterial RNase P RNA (Haas et al. 1994). Nucleotides circled in black are universally conserved among all three domains of life and constitute the five Conserved Regions I–V. Base pairs indicated by closed dots indicate a conserved noncanonical (G·U or A·C) interaction. The *A. castellanii* RNase P RNA secondary structure is based on evidence provided by an alignment of 63 diverse eucaryal RNase P RNA sequences. Shaded boxes indicate base pairs supported by a H_{ij} score ≥ 32 (χ^2 test, $P < 0.0002$; 9 d.f.). Outlined base pairs indicate base pairs confirmed by covariation among eucaryal RNase P RNAs. Bold nucleotides indicate the location of the weak consensus sequence identified in J3a3b of eucaryal RNase P RNAs. Lines connecting circled bases indicate tertiary interactions supported by statistical covariation analysis. The structure of the P12 helix is not established by comparative results and so is represented by a dashed circle. The 5′ and 3′ ends of the *A. castellanii* RNase P RNA, as well as all new sequences in this study, have not been experimentally determined. Instead, the terminus of helix P1 is arbitrarily established as the end of complementarity.

ing gene from the *Giardia* S/M human isolate, a close relative of *G. intestinalis*, was cloned and sequenced using primers based on the *G. intestinalis* sequence. The degree of sequence identity between the *G. intestinalis* and *Giardia* S/M RNAs was high (92%) and so the additional sequence was only marginally helpful with the alignment of the *G. intestinalis* sequence.

New eucaryal RNase P RNAs from genomic DNA sequence databases

In a second approach to expand the diversity represented in the eucaryal RNase P RNA sequence alignment, we used different variations of the CRs I–V consensus sequences

based on animal, fungal, and acanthamoebae RNase P RNAs to search eucaryal genomic databases (Materials and Methods). Putative RNase P RNA sequences were identified in the completed genome sequences of two animals, *Caenorhabditis elegans* and *Drosophila melanogaster*, two alveolates, *Babesia bovis* and *Theileria annulata*, and one microsporidian, *Encephalitozoon cuniculi*. These genes have not been annotated previously. In addition, using primers based on *C. elegans* RNase P RNA, we obtained by PCR the gene from a close relative, *Caenorhabditis briggsae*, and determined that sequence. The conserved structure and sequence elements in these six additional sequences are characteristic of RNase P RNAs, indicating that we identified the RNase P RNA genes for these organisms.

Degenerate PCR amplification of RNase P RNA genes

As we collected more sequences from diverse examples of eucaryal RNase P RNA, we were able to design more generally applicable PCR primers. Degenerate primers based on CR I and CR V successfully amplified RNase P genes from genomic DNAs of *Dictyostelium discoideum*, *Physarum polycephalum*, *Trichomonas vaginalis* and *G. intestinalis* (above) (Materials and Methods). These genes have conserved structure and sequence elements characteristic of RNase P RNA. Based on sequences obtained by PCR, full-length sequences for *D. discoideum* and *G. intestinalis* were identified from the corresponding genomic databases (Materials and Methods). The full-length gene for the *T. vaginalis* RNA was obtained by inverse PCR and sequenced.

We are confident that all of the RNase P RNAs identified in this study are nuclear RNase P RNAs. They are not AU rich and degenerate in structure and therefore are unlikely to be mitochondrial RNase P RNAs (Wise and Martin 1991; Lee et al. 1996). Moreover, these putative RNase P RNAs contain CR IV, and so are unlikely to represent the RNA subunit of RNase MRP, a paralog of RNase P RNA, that lacks the CR IV sequence (Schmitt et al. 1993). The newly identified RNAs are probably not expressed isoforms of RNase P RNA because RNase P RNA isoforms have no evolutionary constraints and tend to have mutations in CRs and large sequence deletions (Li and Williams 1995). As is the case with other eucaryal RNase P RNAs, those of *A. castellanii*, *C. elegans*, and *G. intestinalis* are not active in the absence of protein subunits when tested under a variety of reaction conditions (Material and Methods).

Sequence alignments

A stringent sequence alignment was required to refine the secondary structure model of the eucaryal RNase P RNA. Previous eucaryal RNase P RNA alignments were rudimentary and founded on sequences of limited phylogenetic diversity. Our alignment process included consideration of possible secondary structures supported by covariation and

ongoing refinement as more sequences are added. Sixty-three diverse eucaryal RNase P RNA sequences were aligned manually. Eight of the ~30 eucaryotic phylogenetic kingdoms were represented in the alignment, including 27 species of animals, 24 species of fungi, five species of acanthamoebae, two species of alveolates, two species of conosids, and one species each of trichomonads, microsporidia, and diplomonads. To maximize variation in the data set, most sequences chosen for inclusion in the alignment were <80% identical to one another.

The alignment reveals commonalities and differences between eucaryal RNase P RNAs. An example of a eucaryal RNase P RNA structure based on the alignment is that of *A. castellanii*, shown in Figure 2. All eucaryotic RNase P RNAs include helical elements P1, P2, and P3 with or without an internal bulge-loop; CR I; P4 helix (formed from an interaction of CR I and CR V); a potential four-way junction (eP5/7, P8, P9, P10/11); a loop of 21–24 nt containing CRs II and III (J11/12-J12/11); P12 and a number of nucleotides connecting helices eP5/7 and P2; CR IV; and CR V. The helix between P4 and P8 is designated eP5/7. This helix is homologous among eucaryal RNase P RNAs but the details of homology with bacteria is unclear. Only the *E. cuniculi* and *G. intestinalis* RNAs do not have a bulge-loop in helix P3. The alignment (Ribonuclease P Database http://jwbrown.mbio.ncsu.edu/RNaseP/home.html) shows that nucleotides previously considered universally conserved in CRs I and II continue to appear conserved in the broader context of further diversity, while a few bases in CRs III, IV, and V are less conserved than previously thought. The fifth position of CR III (ugNN**A**) can change to a C, as seen in the *A. castellanii* RNA. The eighth position of CR IV (AgNNNNN**A**U) can change to a C as seen in *E. cuniculi*. The 14th position of CR V (ACNNNANNNNGNN**U**a) can change to a C as seen in *Saccharomyces globosus*. Several regions in the eucaryal RNase P RNAs are particularly variable in length and sequence, and are not alignable. These include, for instance sequences between helices eP5/7 and P8, sequences in helix P12, and between eP5/7 and CR IV. Even though helix P12 is variable in length and sequence, its presence at the same location in most RNase P RNA structures indicates homology with P12 helices in other RNase P RNAs at least at the base of the helix.

Statistical covariation analysis

To test the validity of the secondary structure model and to seek tertiary interactions, the phylogenetically based statistical programs R_{ij} and H_{ij} (R_{ij}/H_{ij}) (Akmaev et al. 2000) were used to analyze bases that covary in the sequence alignment and thereby indicate base-specific interactions. The R_{ij}/H_{ij} programs identify statistical correlations between a position in a multiple sequence alignment and every other position in that alignment (Kelley et al. 2000). First, the R_{ij} program is used to identify correlated positions

in the alignment. Then, the correlated positions are tested using the more robust but more computationally intensive program H_{ij}. The probability that an association occurs between the two positions is calculated and assigned.

The R_{ij}/H_{ij} analysis of aligned eucaryal RNase P RNAs provides strong support for most of the base pairs in the proposed secondary structure. As shown for the *A. castellanii* RNase P RNA (Fig. 2, gray boxes), these base pairing interactions are significant at the $P < 0.0002$ level (H_{ij} value >24; χ^2 test, 9 degrees of freedom). Thus, the probability of random occurrence of a particular base correlation is 0.02% or, conversely, the probability that a correlation exits between two bases is 99.98%. Correlations with $P < 0.0002$ are not expected to occur by chance, and this value was chosen as a limit of confidence. Several of the sequences in the alignment were obtained as PCR products using internal primers, so those sequences did not include the termini. Consequently, the results of the R_{ij}/H_{ij} analysis give little support for terminal portions of P1 and P3. However, substantial covariation in full-length sequences supports the occurrence of these helices (Noller and Woese 1981).

The R_{ij}/H_{ij} statistical analysis of the sequence alignment identifies two correlations between positions that do not involve canonical base pairing. One correlation is between P9 and P10/11, nucleotide positions 130 and 262 (*A. castellanii* numbering, Fig. 2). This interaction was significant at the $P < 0.0002$ level (H_{ij} value >32; χ^2 test, 9 degrees of freedom). The other is between the base of P12 and P10/11, nucleotide positions 160 and 261, significant at the $P < 0.0001$ level (H_{ij} value >34; χ^2 test, 9 degrees of freedom).

In addition to sequence covariations, we considered covariations between structural elements of the RNAs. Helix P12, J5/7/8 (joining region between eP5/7 and P8), and J5/7/CR IV may abut on the same three-dimensional space in the eucaryotic RNase P RNAs. We observe an inverse correlation between the length of P12 and the length of J5/7/8 and J5/7/CR IV (Fig. 3). The inverse relationship suggests that there may be a crowding effect between P12 and J5/7/8 and J5/7/CR IV. If they shared potentially overlapping space, an increase in the size of one region would require the reduction of the other. Although the comparative data suggest that the structures can occupy the same space, the nature of the arrangement of the nucleotides and helices in this space is unknown.

A feature that is present in most eucaryal RNase P RNAs and absent in all bacterial and archaeal RNAs is a bulge-loop (J3a/3b) between helices P3a and P3b. This bulge-loop in helix P3 is conserved in presence, but not in size, in most known eucaryal RNase P RNAs. A weak consensus sequence seems to occur in the 5' sequence of the bulge-loop, as shown in Figure 4. The statistical analysis (above) offered no support for covariation between this weak consensus sequence and other sequences within the RNA. Size variation in J3a/3b is 8–16 nt, with the majority of sequences
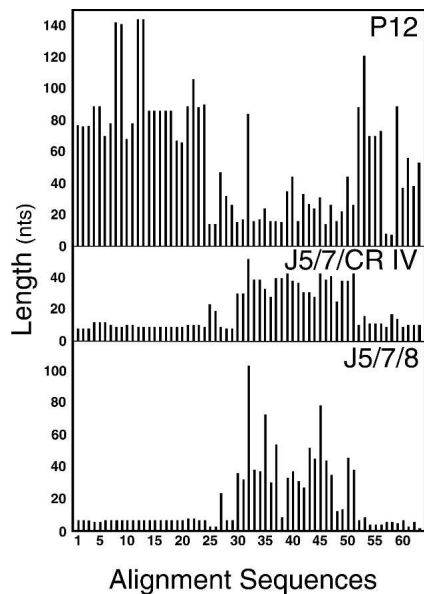
FIGURE 3. Compensatory length variation. The *X*-axis represents eucaryal RNase P RNA sequences as numbered in the eucaryal RNase P 63 alignment (http://pacelab.colorado.edu/publications.html/). The *Y*-axis is the length, in nucleotides, of the J5/7/8 and J5/7/CR IV regions and helix P12 as indicated.

containing 9 nt. The opposite side of the loop, J3b/3a, varies from 2 nt to 25 nt in size and no conservation of this sequence was observed. Notably, the P3 bulge-loop is absent from the *E. cuniculi* and *G. intestinalis* RNAs (Discussion).

DISCUSSION

Our study expands the phylogenetic representation of eucaryotic RNase P RNA sequences from two to eight kingdoms, to include examples of sequences from organisms that span the extremes of evolutionary divergence among eucaryotes (Fig. 5). Acanthamoebae (*A. castellanii*, *B. mandrillaris*, *F. nolandi*), animals (*C. elegans*, *D. melanogaster*), and the alveolates (*T. annulata*, *B. bovis*) branch throughout the "crown-radiation" of eucaryotes (Sogin 1991). Conosids (*D. discoideum*, *P. polycephalum*) branch more deeply than the "crown-radiation". Trichomonads (*T. vaginalis*), microsporidia (*E. cuniculi*), and the diplomonads (*G. intestinalis*) are thought to represent some of the most deeply divergent eucaryotic lineages known, based on the ribosomal RNA tree. Thus, a general perspective on eucaryal RNase P RNA structure emerges from the alignment of these sequences.

Evidence for the general secondary structure model of the eucaryal RNase P RNA was significantly refined and enhanced by inclusion of the new sequences in comparative analysis. Highly diverse novel sequences can be particularly useful for inference of structure because the differences in sequence and structure tend to increase with greater evolutionary distance between organisms. Elements of the RNA that are not universally present are not likely to be a core

component of the RNA subunit but still may be essential for optimal function of the RNA. The main structural differences among eucaryal RNase P RNAs so far determined occur in regions of sequence length variation and the presence or absence of the bulge-loop in P3.

The nonhomologous regions that occur between most of the conserved elements of eucaryal RNase P RNAs can be highly variable in sequence and length. They are found in eucaryal RNase P RNAs but generally do not occur in bacterial or archaeal RNAs. Of the eight eucaryotic kingdoms analyzed here, variation in interhelix sequence length in RNase P RNAs is most prominently found in fungi, with as many as 103 nt between helices eP5/7 and P8 of *Arxiozyma telluris* (Frank et al. 2000). The largest and most frequent sites of insertions are between the helices eP5/7 and P8, P12 and between eP5/7 and CR IV. Presumably these locations are preferred because "extra sequences" placed there are likely to occur on the surface of the RNA molecule and can expand without distortion of the active elements. Thus, the covariation observed between the lengths of J5/7/8, helix P12, and J5/7/CR IV suggests that these three structures share the same space on the surface of eucaryal RNase P RNA.

Helix P12 is highly variable in length and structure, even among close relatives. The length ranges from 7 to 144 nt among eucaryal RNase P RNAs analyzed here, and generally phylogenetically disparate P12 sequences cannot be aligned. Only the base of P12 can be aligned. Nevertheless, the occurrence of P12 in all eucaryal RNase P RNAs indicates an important structural role, perhaps to position CR II and CR III sequences. The J11/12-J12/11 module likely forms a specific structure in all RNase P RNAs, with P12 forming one boundary of this structure (Krasilnikov et al. 2003).

Regions of highly variable length and structure found in eucaryal RNase P RNA are reminiscent of rRNA "expansion segments" (Clark et al. 1984). The functions of expansion

FIGURE 4. J3a/3b consensus sequence. Nucleotide conservation is measured in bits of information. The information contained in a nucleotide position ranges from zero to 2 bits. One hundred percent conserved positions contain two bits of information while completely random positions contain zero bits of information. Horizontal-axis numbers indicate nucleotide position in J3a/3b (*A. castellanii* numbering). Sequence logo generated by using "WebLogo" at http://www.weblogo.berkeley.edu/.
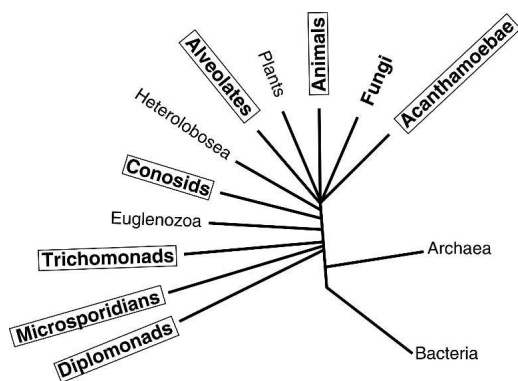
FIGURE 5. Phylogenetic distribution of eucaryal RNase P RNA sequences. Representative organism groups from the major eucaryotic kingdoms are shown as a diagrammatic phylogenetic tree. The phylogenetic tree is based on small subunit ribosomal RNA sequences and is modified from Dawson and Pace (2002). Bold names indicate phylogenetic kingdoms represented by RNase P RNA sequences. Boxed names represent taxa from which eucaryal RNase P RNA sequences were determined in this study.

segments in rRNA seem to vary (Houge et al. 1995; van Nues et al. 1997). Like some expansion segments found in rRNAs, it is possible that these regions are important for normal biogenesis and the stability of eucaryal RNase P RNAs (Jeeninga et al. 1997). When the variable sequences are omitted, all eucaryal RNase P RNAs have the same core secondary structure, as illustrated in Figure 6 for two of the RNAs.

The bulge-loop in helix P3 occurs in all eucaryotic RNase P RNAs examined except those of *E. cuniculi* and *G. intestinalis*. The presence or absence of this loop may shed light on the controversy of the placement of the kingdom microsporidia, to which *E. cuniculi* belongs, in the eucaryotic phylogenetic tree. In small subunit ribosomal RNA phylogenetic trees, diplomonads, trichomonads, and microsporidia consistently branch deeply in the eucaryotic lineage (Fig. 5; Sogin 1997; Dawson and Pace 2002). In some sequence comparisons, however, microsporidia are seen to branch with fungi (Baldauf et al. 2000; Van de Peer et al. 2000). The most parsimonious explanation for the similarities between the *G. intestinalis* and *E. cuniculi* helix P3 structures is that the microsporida are more closely related to the diplomonads than to fungi, consistent with a deep branching of the microsporidia lineage. In this case, the ancestral eucaryal RNase P RNA would have lacked an internal loop in P3, consistent with the absence of a bulge-loop in the bacterial and archaeal RNAs. The bulge-loop in most eucaryal RNAs would derive from an event in the main eucaryal line of descent subsequent to the divergence of the evolutionary lines that led to modern-day diplomonads and microsporida. The possible weak consensus sequence seen in the bulge-loop of P3 in those RNAs that contain the bulge-loop may indicate that an interaction with a protein was gained at the time of acquisition of the P3 bulge-loop.

These events could signify that a new function was imparted to the holoenzyme. Previous studies have shown the protein Pop1p to interact with the P3 domain and to be essential for *S. cerevisiae* RNase P RNA maturation (Jacobson et al. 1997; Ziehler et al. 1997). Eucaryal P3, with a bulge-loop, is also required for localization of RNase P RNA to the nucleolus of HeLa and rat kidney epithelial cells (Jacobson et al. 1997).

The essential elements of the eucaryal RNase P RNA can be identified in the context of the collection of structural elements and sequences that are common to all eucaryal RNAs. This "minimum consensus core" structure is highly similar to that of the bacterial minimum consensus core structure (Fig. 7), which has been shown sufficient for catalysis (Siegel et al. 1996). The similarities of the core structures of the bacterial and eucaryotic RNAs suggest that the eucaryotic version of RNase P RNA, even though catalytically inactive without protein, contains many of the elements required for substrate binding and catalysis. The general similarity of the core structures of all the eucaryotic RNAs indicates that the basic secondary structure of eucaryal RNase P RNA was established early in evolution and has persisted throughout time. Although the secondary structure of the core is highly conserved evolutionarily, the overall sequence conservation is low, even within phylogenetic kingdoms. For instance, the core RNase P RNAs of the animals *C. elegans*, an invertebrate, and *Homo sapiens*, a vertebrate, are only 49% identical in sequence. This compares with 82% identity between the corresponding small subunit rRNA sequences. Despite otherwise extensive sequence variation in the RNase P RNA core, however, the universally conserved nucleotides and secondary structures are unchanged between these organisms.

Some features are not conserved between the bacterial and eucaryal RNAs, including structures and sequences important in the bacterial RNAs for tRNA binding and structural stability. Indeed, the typical eucaryotic RNA is roughly two-thirds the sequence length of the typical bacterial RNase P RNA. For example, the eucaryal RNAs all lack a homolog of the bacterial J15/16 loop, shown in Figure 7. Loop J15/16 specifically interacts with the 3′-end of tRNA (CCA) and contributes to the binding and assembly of the catalytic site (Kirsebom and Svard 1994; Oh and Pace 1994). It is possible that in eucaryotes a protein assumes this role (Jarrous et al. 2001). Additionally, eucaryal RNase P RNAs lack structurally stabilizing helices found in bacterial and archaeal RNase P RNAs. In bacterial RNAs, noncore helices are attached to the conserved core and interact with other regions of the RNA to stabilize structure important for function (Darr et al. 1992; Brown et al. 1996). Deletion of stabilizing helices not present in the eucaryotic RNAs leads to global destabilization of the bacterial RNAs (Darr et al. 1992). It is possible that the multiple proteins of the eucaryotic holoenzyme help stabilize the global RNA structure in the natural absence of these stabilizing RNA helices.
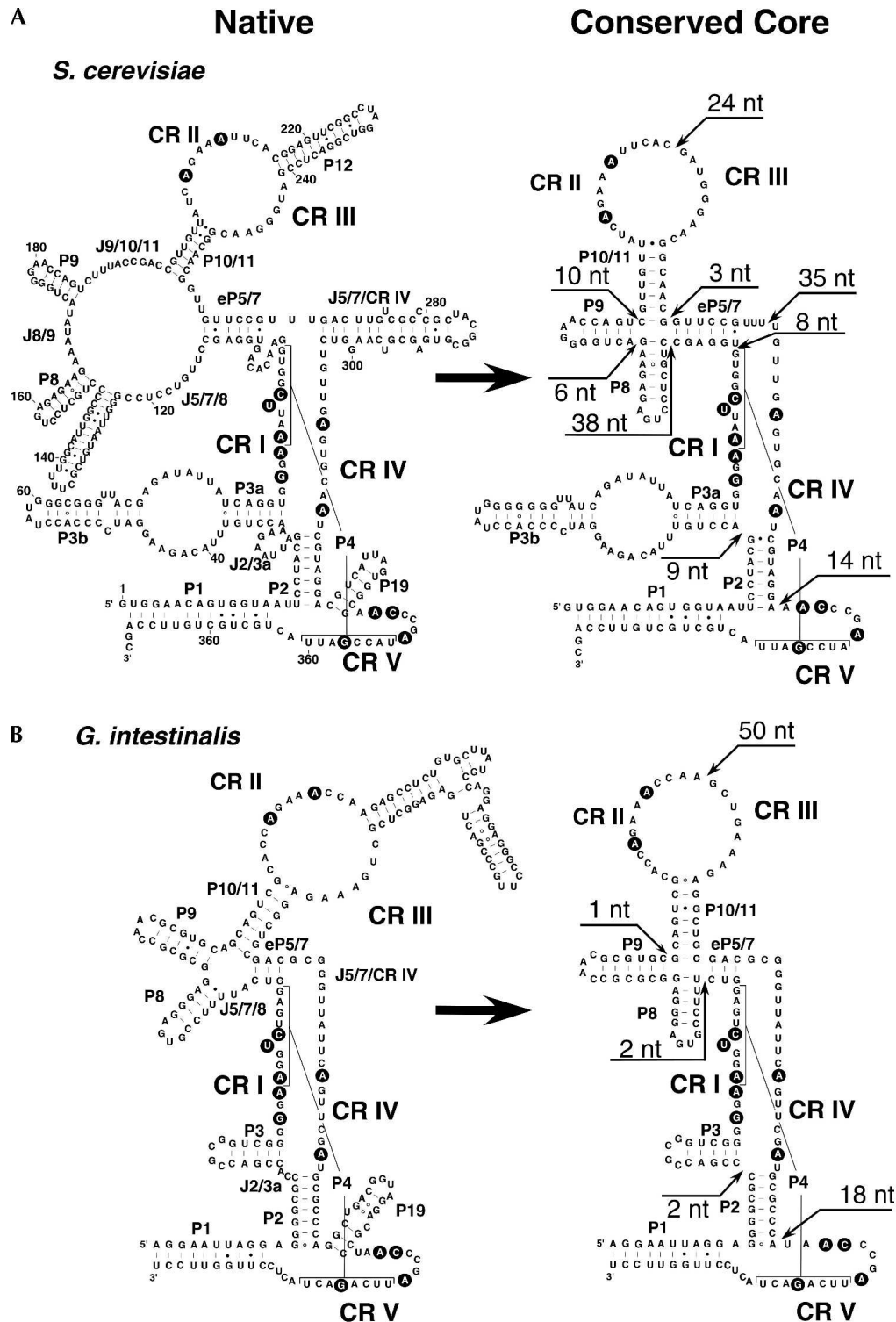
FIGURE 6. Secondary structures of native eucaryal RNase P RNAs and with nonhomologous sequences omitted. The secondary structures of (*A*) *S. cerevisiae* and (*B*) *G. intestinalis* with and without J5/7/8, J8/9, J9/10,11, P12, and P19. Arrows and numbers indicate positions and size, respectively, deleted to form minimal structures.

The statistical covariation analysis identified two possible correlations between positions that do not involve canonical base pairing. Significant covariation occurs between nucleotide positions 130 and 262 and between nucleotide positions 159 and 261 (*A. castellanii* numbering). The statistical analysis identifies correlations between nucleotide
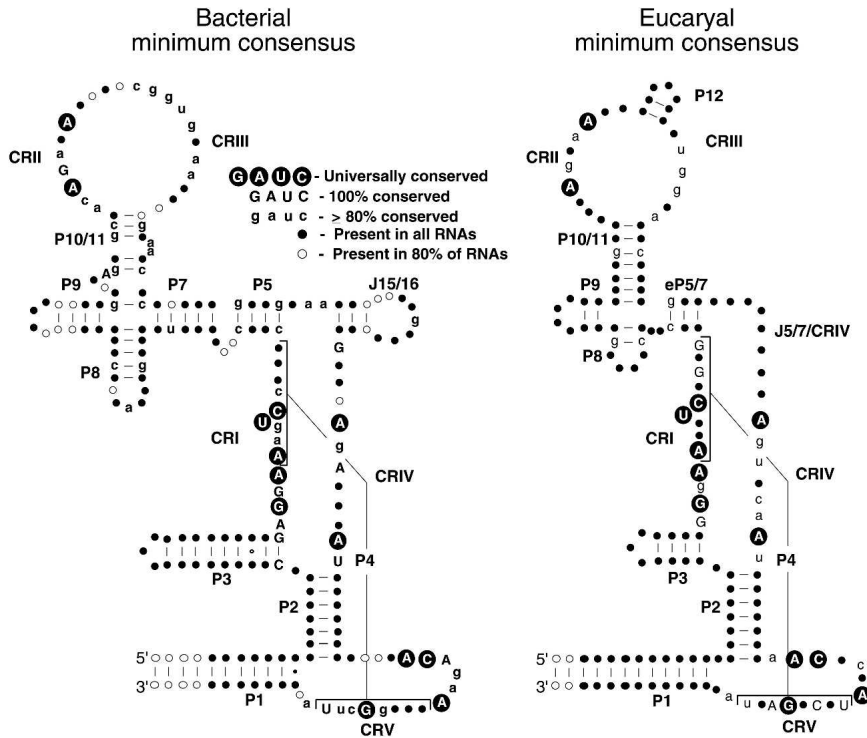
FIGURE 7. Phylogenetic minimum consensus RNase P RNA secondary structures. Updated phylogenetic bacterial minimum consensus RNase P RNA secondary structure (Siegel et al. 1996) and the eucaryal minimum consensus RNase P RNA secondary structure are shown.

tion from RNA-based function to protein and RNA-based function.

MATERIALS AND METHODS

For a list of oligonucleotides used in this study, see Table 1.

Cell cultures

A. castellanii cells were grown as described in Detke and Paule (1975) with modifications. *A. castellanii* cells were grown in 2.8-L baffled flasks to a density of 2.0×10^6 cells/mL, late log phase, and harvested by centrifugation in a GSA rotor (Sorvall), at 3300g for 15 min at 4°C. The cells were washed once with two volumes of 10 mM Tris-HCl, pH 7.6, 15 mM KCl, and recentrifuged. Cells not used immediately were frozen in liquid nitrogen and stored at −80°C. A typical yield of *A. castellanii* was 4.5 g wet cell paste/liter medium. Axenic cultures of *G. intestinalis* (clone WBC6, ATCC 30957) trophozoites were maintained in a modified, carbonate-buffered (25 mM) TYI-S-33 medium, and supplemented with bovine bile (Keister 1983) and 10% adult bovine serum. Ten liters of trophozoites were grown in a 10-chambered cell factory (Nalge Nunc International). Trophozoites were harvested on ice, centrifuged at 900g, and washed twice in 1× phosphate-buffered saline.

positions in the alignment and may not necessarily reflect direct physical interactions between these two positions. In the crystal structure of a portion of the *Bacillus subtilis* and *Escherichia coli* RNase P RNAs, helices P9 and P10/11 are in close proximity to one another, consistent with a possible interaction. Although the correlation between the base of P12 and P10/11 is supported by a high H_{ij} value, these two positions are not proximal to one another in the crystal structures of the fragments of the *B. subtilis* and *E. coli* RNAs (Krasilnikov et al. 2003, 2004). It may be that a physical interaction does not exist. Since comparative analysis is interpreted as structure in vivo, it is also possible that the two bases are near one another or interact via an intermediary in vivo. We used the R_{ij}/H_{ij} statistic on an alignment of 331 type A bacterial RNase P RNAs (Brown 1999) and did not detect the same correlation. The relevant sequences in the bacterial P10/11 are highly conserved, however, so any variation that would indicate structure may be rare or not occur. The R_{ij}/H_{ij} statistical analysis of the type A RNase P RNA confirmed the interaction between P8 and P14 (Krasilnikov et al. 2004).

The eucaryal RNA evidently was remodeled through evolution as a simplified version of the structure found in bacterial and archaeal RNase P RNAs. This simplification apparently was accompanied by the acquisition of multiple proteins by the eucaryotic holoenzyme. Thus, the eucaryal RNase P RNA may be a useful tool to examine the transi-

Purification of *A. castellanii* RNase P RNA

All purification steps were conducted at 4°C. Thirty-seven grams of wet cell paste were resuspended in DE 52 equilibration buffer, DE52EB (40 mM Tris-HCl, pH 8.0, at 4°C, 0.15 mM spermine, 0.5 mM spermidine, 0.1 mM PMSF, 0.1 mM DTT, 0.2 mM EDTA, 0.2 mM EGTA, 5 mM MgCl$_2$, 0.05% NP-40, 10% glycerol) + 20 mM KCl at 2 mL/g cell paste. A Dura-Grind stainless steel dounce tissue grinder (Wheaton) was used (30 strokes) to lyse cells without disrupting the nuclei. The supernatant, which contains all RNase P activity, was separated from the nuclear pellet by centrifugation in a SS-34 (Sorvall) rotor at 5100g for 15 min (Rimm and Pollard 1989). The nuclei were frozen in liquid nitrogen and stored at −80°C for subsequent isolation of genomic DNA. One hundred seventy-one milliliters of nuclei-free supernatant was applied to a 180-mL bed volume DEAE-cellulose (DE52), (Whatman) column (4.0 cm × 11.5 cm, 4 mL/min). The column was equilibrated with DE52EB + 10 mM KCl. The column was washed with a one-column volume of DE52EB, followed by a stepwise elution using a one-column volume of DE52EB + 80 mM KCl, and subsequently a one-column volume of DE52EB + 200 mM KCl. Five-milliliter fractions were collected and assayed for RNase P activity. Fractions with activity were pooled (85 mL) and dialyzed 3× against 1 L of DE52EB + 10 mM KCl. The dialyzed pool (76 mL) was applied at 1 mL/min to a Toyopearl DEAE-650M (TosoHaas) column (1.6 cm × 15 cm), equilibrated in the same

TABLE 1. Oligonucleotides

CRIV#2	5′-CATAACTTGACTATCAACGATCG-3′
CRV*	5′-GGAATTCTAGTCIGAATIGNGT-3′
ACGSP1	5′-GTCGATGACGGCTTTGTGTATTGAACT-3′
ACGSP2	5′-GATAGTTGCTAGCCACCCGGGTCGGTT-3′
ACGSPREV1	5′-CAAGAGGGAGCTCGGGGACCAGAAACG ACA-3′
ACGSPREV2	5′-TGAACCGACCCGGGTGGCTAGCAACTA-3′
ACLOCF2	5′-ACAATAGTATGTGCGGGCTGCGCGAGG TGA-3′
ACLOCR2	5′-ATAGGCCAAGTTTTCAGCATTCCYTGC TTTC-3′
ACF	5′-GGGTGAAGCTTGTGTTAGGTGC-3′
ACR	5′-GTGAGAGTAGTCTGAATTGAGT-3′
ACT7	5′-CGCGGTACCTAATACGACTCACTATAGGGG TCAGGGTGAAGCTTGTGTTAGGTGC-3′
ACT7R	5′-GGGGTCAGGGTGAGAGTAGTCTGAATTG AGTTG-3′
CET7	5′-CGCGGTACCTAATACGACTCACTATAGAG GTTGGCCGGAAGAAGACGG-3′
CER	5′-AAGGTTGGCCGAAGACTAGTCTGAATTGT GTTGG-3′
G1F	5′-TTTGGGAAGGTCTGA-3′
G1R	5′-TTTTAGTCTGAATYDNGT-3′
G3F	5′-ATTAGGAGGGGCGCCACCGA-3′
G3R	5′-AGGAGTAGTCTGAATCGGGT-3′
G4R	5′-AITAGTCTGAATYDNGTT-3′
TrF	5′-GGCTCGTTACTGGTTCCAAGC-3′
TrR	5′-CACGCTGCGTTTGTGAGTGAG-3′
G120R	5′-GTTTCTGGTGCGACTGCTGCACGCGTTG-3′
T7GF	5′-CGCGGTACCTAATACGACTCACTATAGGAA TTAGGAGGGGCGCCACCGACCGCG-3′
T7GR	5′-CGCGGTACCAGAACCAAAGGAGTAG TCTGAATCGGGTTAGGCGTCCTACC-3′

Where Y is pyrimidine; I is inosine; D is A, G, or T; and N is any base.

buffer. The RNA was eluted using a 200-mL linear gradient of 80–500 mM KCl in DE52EB, and 3-mL fractions were collected. Two hundred fifty microliters of each fraction from the Toyopearl DEAE-650M column was phenol extracted and the RNA precipitated with ethanol. The RNAs (1 μL) from each fraction were analyzed following labeling with [5′-^{32}P]pCp and T4 RNA ligase (England et al. 1980). The first DEAE-cellulose column with step elution provides a rapid method to process crude extract and the second DEAE-Toyopearl column with gradient elution fractionates the extract. In addition, large losses of RNase P do not occur with DEAE-based chromatography. Labeled RNAs were resolved by electrophoresis in a 6% polyacrylamide gel containing 8 M Urea in TBE (Sambrook et al. 1989) and detected by autoradiography. Fractions containing RNase P activity were pooled (23 mL). EDTA was added to the pooled enzyme-containing fractions to a final concentration of 10 mM, incubated at room temperature for 2 min, then the pool was extracted with an equal volume of buffer-saturated phenol (Gibco-BRL) and centrifuged at 2000*g* in a HB-4 (Sorvall) for 10 min. The phenol phase was re-extracted with 3 mL TE, pH 8.0 (10 mM Tris-HCl, pH 8.0, 1 mM EDTA), and centrifuged at 16,000*g*. The aqueous phases were combined, phenol extracted, ethanol precipitated, and centrifuged at 5000*g* for 10 min in a SS-34 rotor. The precipitate was washed by resuspending

the RNA in 8 mL TE and reprecipitated with ethanol. RNA (18 mg) purified from enzyme fractions was resuspended in 5 mL of gel filtration buffer (10 mM Tris-HCl, pH 8.0, 2 mM EDTA, 100 mM NaCl, 0.1% SDS) and applied to a Sephacryl S-500 (Amersham Biosciences) gel filtration column (0.8 cm × 100 cm, 1 mL/min) and 1.8-mL fractions were collected. Ribosomal RNAs constitute the majority of RNA after the Toyopearl DEAE-650M column and in order to purify large quantities of potential RNase P RNA with preparative gel electrophoresis, we needed to remove large molecular weight RNA by gel filtration. To visualize RNAs associated with activity, 40 μL from each gel filtration fraction was removed, precipitated with ethanol, 3′ end labeled, resolved by 6% denaturing PAGE, and detected by autoradiography. Fractions containing RNAs that coincided with RNase P RNA activity in the Toyopearl DEAE-650M column were ethanol precipitated. Approximately 100 μg of RNA from the gel filtration column, potentially containing RNase P RNA, was fractionated by electrophoresis through a 1.5-mm thick 4% polyacrylamide 8 M Urea gel in TBE. To identify potential RNase P RNA bands in the preparative gel, labeled RNA from the peak of enzymatic activity from the Toyopearl DEAE-650M column was run in adjacent lanes. Following electrophoresis, previously identified RNAs, which are associated with RNase P activity, from the Toyopearl DEAE-650M column were detected by autoradiography and the film aligned with the ethidium bromide stained preparative gel. Putative RNase P RNAs were excised and eluted passively in 40 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.1% SDS at 4°C overnight followed by ethanol precipitation and resuspension in TE.

Preparation of RNAs

Uniformly labeled *Bacillus subtilis* pre-tRNA^{asp} was prepared as previously described (Reich et al. 1986). Eucaryal RNase P RNAs were prepared by run-off in vitro transcription with T7 RNA polymerase and Fok I-digested pJLC101 containing respective RNAs (Chen 1997) or PCR amplification of the RNase P RNA gene from respective genomic DNA using a unique forward primer containing the T7 promoter sequence and a unique reverse primer. Primer pairs for PCR templates are *A. castellanii* (ACT7, ACTR), *C. elegans* (CET7, CETR), and *G. intestinalis* (T7GF, T7GR).

RNase P activity assays

The 20-μL holoenzyme assay included 30 nM pre-tRNA^{asp}, 50 mM NH$_4$OAc, 10 mM MgOAc, 1 mM DTT, 0.1% Nonidet P-40 (NP-40), 10% glycerol, 50 mM Tris-HCl, pH 8.0, and 2 μL from the purification fraction. Reactions were incubated at 37°C for 20 min, then terminated by adding EDTA and SDS to a final concentration of 10 mM and 0.1%, respectively. After adding loading buffer (8 M urea, 10 mM Tris-HCl, pH 8.0, 2 mM EDTA, 0.1% SDS, 0.1% bromophenol blue, and 0.1% xylene cyanol), the products were resolved by 6% denaturing PAGE and detected by autoradiography. The activity assays of eucaryal RNase P RNAs (17–300 nM) were carried out in 50 mM Tris-HCl, pH 8.0, 0.05% NP-40, and 0.01% SDS containing 1–4 M NH$_4$OAc and 25–300 mM MgOAc. Reactions were incubated for 3 h at 37°C. The reactions were stopped by adding three volumes of ethanol and 1/10 volume of 3 M NaOAc, pH 5.2. Precipitated RNAs were recovered by centrifugation and suspended in 20 μL of loading buffer. The products were analyzed as above.

Primer extension sequence analysis of RNAs

Gel-purified, putative *A. castellanii* RNase P RNAs were sequenced by primer extension as previously described (Chen et al. 1998) using the primer CRV*. The *G. intestinalis* specific primer G120R was annealed to ~30% of purified RNA from each Toyopearl DEAE-650 column fraction and primer extended. For a control, end-labeled G120R was annealed to ~2 ng of in vitro transcribed putative *G. intestinalis* RNase P RNA and also primer extended.

Isolation of genomic DNA

A. castellanii (nuclei), *G. intestinalis* (4.0×10^7 cells) and *G*. SM (4.8×10^7 cells) genomic DNAs were isolated using DNAzol according to the manufacturer's protocol (Molecular Research Center).

Ligation-mediated PCR: Sequence determination of the *Acanthamoebae* RNase P RNA genes

To sequence the full-length *A. castellanii* RNase P RNA gene from genomic DNA, an adaptor-ligated genomic library was constructed and PCR amplification of target sequences was carried out using the Universal Genome Walker Kit (Clonetech). The primers ACGSP1 and ACGSP2 were used to amplify the sequence 5′ of CR V, and the primers ACGSPREV1 and ACGSPREV2 were used to amplify the sequence 3′ of CR II. These primers were designed on the basis of *A. castellanii* sequence obtained through primer extension analysis (above). To confirm the DNA sequence of the *A. castellanii* RNase P gene, PCR-generated clone sequence-specific primers, ACLOCF2 and ACLOCR2, were made upstream and downstream of the *A. castellanii* gene. Because of its high GC content, routine PCR failed to amplify the *A. castellanii* gene from genomic DNA, so PCR amplifications were carried out using the Advantage-GC Genomic PCR kit (Clonetech). Amplification products were purified with the QIAquick PCR Purification Kit (QIAGEN). PCR product was sequenced in both directions. The primers ACF and ACR were used to amplify putative RNase P RNA genes from the genomic DNAs, of *B. mandrillaris*, *F. nolandi*, and *F. sp.,* provided by Linda Amaral Zettler (The Marine Biological Laboratory, MA), using Advantage-GC Genomic PCR.

Genomic database searches

RNase P RNA-encoding sequences from *C. elegans*, *D. melanogaster*, *B. bovis*, *T. annulata*, and *E. cuniculi* were identified in their genome sequences using various combinations of the Conserved Region V (CR V) sequence, based on vertebrate and fungi sequences, to query GenBank, The Sanger Institute, and using their respective BLAST servers (Altschul et al. 1990). Resulting sequences were inspected for the presence of CRs I–IV at their respective positions in the overall sequence. A secondary structure was inferred from the sequences to check for the presence of potential helices known to exist in other eucaryal RNase P RNAs.

Degenerate PCR amplification of RNase P RNA genes

RNase P RNA sequences were amplified by PCR using the following general protocol: 1× PCR buffer: 20 mM Tris-HCl, pH 8.3, 50 mM KCl, 4 mM MgCl$_2$; 0.05 mg/mL BSA, 2 mM of each dNTP, 0.05% NP-40, 1 pM of each forward and reverse primers, 300 ng of genomic DNA, 3 units/100 μL reaction of *Taq* polymerase. Reactions were incubated in a Mastercycler Gradient thermal cycler (Eppendorf [Brinkmann Instruments, Inc.]). The profile for amplification was 94°C for 2 min, 30 cycles of 94°C for 30 sec, an annealing temperature gradient across the block of 35°C–55°C for 1 min, 72°C for 1.5 min, and a final incubation at 72°C for 5 min. Amplifications were done with the following primers: G1F/G1R (*G. intestinalis* genomic DNA provided by Fran Gillin, University of California-San Diego, School of Medicine, CA); G3F/G3R (*G*. SM isolate cells were provided by Fran Gillin); G1F/G1R (*D. discoideum* genomic DNA provided by William Loomis, University of California-San Diego, CA); G1F/G4R (*T. vaginalis* genomic DNA provided by Patricia Johnson, University of California-Los Angeles, CA); and G1F/G2R (*P. polycephalum* provided by Jonatha M. Gott, Case Western Reserve University, OH). PCR amplification products were examined by agarose electrophoresis and those of appropriate size were excised from the gel and purified by the QIAquick gel extraction kit (QIAGEN). PCR products were then cloned with the TOPO TA cloning kit (Invitrogen) and sequenced. Full length sequences for *D. discoideum* and *G. intestinalis* RNase P RNA genes were identified in their genome sequences using their PCR sequences to query the *Dictyostelium* (http://dicty.sdsc.edu) and *Giardia* (http://www.mbl.edu/Giardia) genomic databases using their respective BLAST servers. The full-length *G. intestinalis* RNase P RNA gene was cloned using the primers T7GF/T7GR. The putative RNase P RNA genes were amplified from 300 ng of genomic DNA using *Pfu* DNA polymerase (Stratagene). The PCR product was digested with Kpn I and cloned into pJLC101. Inverse PCR was used to obtain upstream and downstream sequences flanking the *T. vaginalis* RNase P RNA gene. *T. vaginalis* genomic DNA, 100 ng, was digested with Mse I in a 20-μL reaction. The enzyme was heat inactivated at 65°C for 20 min. The DNA fragments (entire reaction) were circularized in a 400-μL ligation reaction containing 10 units of T4 DNA ligase (Invitrogen) for 13 h at 4°C. Circularized DNA was precipitated with ethanol and resuspended in 20 μL TE. Sequences were amplified by PCR using "touchdown PCR" with 1 pM each of the primers TrF/TrR. The profile for amplification was 94°C for 2 min; 10 cycles of 94°C, 30 s; 64°C, 30 s, −2°C/cycle; 25 cycles of 94°C, 30 s; 45°C, 30 s; 72°C, and 72°C for 5 min.

Alignment

Eucaryal RNase P RNA sequences were aligned manually using SeqApp v.1.9 (Indiana University, Bloomington), a biological sequence editor and analysis program for Macintosh computers. Conserved regions CRI–V were identified and served as markers on which other sequences were aligned. Sequences were moved until putative homologous residues occupied the same column in the alignment. To achieve this, the alignment process included drawing and refining possible secondary structures as more sequences were added, on the basis of support for homologous base pairing in the RNase P RNA sequences of different organisms. Secondary structures not supported by covariation were not aligned. Previously determined RNase P RNA sequences were obtained from the RNase P Database (http://www.mbio.ncsu.edu/RNaseP) (Brown 1999). All eucaryal alignments are available at http://www.pacelab.colorado.edu. All sequences, secondary

structures, and the alignments are available in the RNase P RNA Database.

Statistical analysis of eucaryal RNase P RNAs

The eucaryal 63, 51, and 28 Mask alignments (http://www.pacelab.colorado.edu) were analyzed for covariation using the phylogenetically based statistical programs R_{ij} and H_{ij} (R_{ij}/H_{ij}) (Akmaev et al. 2000). Because the Rij/Hij method ignores nucleotide sequence when positions being compared are absent >50% of the time, sequences were grouped into three different alignments so that as many of the sequences as possible could be analyzed. For each data set a phylogenetic tree was calculated using the neighbor-joining algorithm in the Phylip phylogeny package (University of Washington, Seattle). In this analysis, the computationally faster R_{ij} method was used to identify the set of initial pairs with high correlation values and then the computationally intensive H_{ij} statistic was used to further evaluate the high scoring pairs. The H_{ij} statistic approximates a χ^2 distribution with nine degrees of freedom allowing an assessment of statistical confidence in specific base-pairings.

Accession numbers

New RNase P RNA sequences were deposited into GenBank: *A. castellanii* (AF440361), *B. mandrillaris* (AF440362), *F. nolandi* 1 (MMM-2004a), *F. nolandi* 2 (MMM-2004b), *F. sp.* (AF440364), *C. elegans* (AF434764), *D. melanogaster* (AF434763), *T. vaginalis* (AY627758), *G. intestinalis* (AY627759), *Giardia* sp. SM (AY627760), *D. discoideum* (AY627761), and *P. polycephalum* (AY627762). *E. cuniculi* RNase P RNA gene can be found in AL590447, nucleotides 87178–86929. The *T. annulata* and *B. bovis* RNase P RNA gene sequence data were produced by the Pathogen Sequencing Unit at The Wellcome Trust Sanger Institute and can be obtained from ftp://ftp.sanger.ac.uk/pub/pathogens/T_annulata/and ftp://ftp.sanger.ac.uk/pub/pathogens/babesia/.

ACKNOWLEDGMENTS

REFERENCES

Akmaev, V.R., Kelley, S.T., and Stormo, G.D. 2000. Phylogenetically enhanced statistical tools for RNA structure prediction. *Bioinformatics* **16:** 501–512.

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

Amaral Zettler, L.A., Nerad, T.A., O'Kelly, C.J., Peglar, M.T., Gillevet, P.M., Silberman, J.D., and Sogin, M.L. 2000. A molecular reassessment of the Leptomyxid amoebae. *Protist* **151:** 275–282.

Baldauf, S.L., Roger, A.J., Wenk-Siefert, I., and Doolittle, W.F. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* **290:** 972–977.

Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289:** 905–920.

Brown, J.W. 1999. The Ribonuclease P Database. *Nucleic Acids Res.* **27:** 314.

Brown, J.W. and Pace, N.R. 1992. Ribonuclease P RNA and protein subunits from Bacteria. *Nucleic Acids Res.* **20:** 1451–1456.

Brown, J.W., Nolan, J.M., Haas, E.S., Rubio, M.A.T., Major, F., and Pace, N.R. 1996. Comparative analysis of ribonuclease P RNA using gene sequences from natural microbial populations reveals tertiary structural elements. *Proc. Natl. Acad. Sci.* **93:** 3001–3006.

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.

Chen, J.L. 1997. "Structure of ribonuclease P RNA." Ph.D. thesis, Department of Biology, Indiana University, Bloomington.

Chen, J.-L. and Pace, N.R. 1997. Identification of the universally conserved core of ribonuclease P RNA. *RNA* **3:** 557–560.

Chen, J.L., Nolan, J.M., Harris, M.E., and Pace, N.R. 1998. Comparative photocross-linking analysis of the tertiary structures of *Escherichia coli* and *Bacillus subtilis* RNase P RNAs. *EMBO J.* **17:** 1515–1525.

Clark, C.G., Tague, B.W., Ware, V.C., and Gerbi, S.A. 1984. *Xenopus laevis* 28S ribosomal RNA: A secondary structure model and its evolutionary and functional implications. *Nucleic Acids Res.* **12:** 6197–6220.

Cordier, A. and Schon, A. 1999. Cyanelle RNase P: RNA structure analysis and holoenzyme properties of an organellar ribonucleoprotein enzyme. *J. Mol. Biol.* **289:** 9–20.

Dang, Y.L. and Martin, N.C. 1993. Yeast mitochondrial RNase P. Sequence of the RPM2 gene and demonstration that its product is a protein subunit of the enzyme. *J. Biol. Chem.* **268:** 19791–19796.

Darr, S.C., Zito, K., Smith, D., and Pace, N.R. 1992. Contributions of phylogenetically variable structural elements to the function of the ribozyme ribonuclease P. *Biochemistry* **31:** 328–333.

Dawson, S.C. and Pace, N.R. 2002. Novel kingdom-level eukaryotic diversity in anoxic environments. *Proc. Natl. Acad. Sci.* **99:** 8324–8329.

Detke, S. and Paule, M.R. 1975. DNA-dependent RNA polymerases from *Acanthamoeba castellanii*: Properties and levels of activity during encystment. *Biochim. Biophys. Acta* **383:** 67–77.

England, T.E., Bruce, A.G., and Uhlenbeck, O.C. 1980. Specific labeling of 3′ termini of RNA with T4 RNA ligase. *Methods Enzymol.* **65:** 65–74.

Fox, G.W. and Woese, C.R. 1975. 5S RNA secondary structure. *Nature* **256:** 505–507.

Frank, D.N. and Pace, N.R. 1998. Ribonuclease P: Unity and diversity in a tRNA processing ribozyme. *Annu. Rev. Biochem.* **67:** 153–180.

Frank, D.N., Adamidi, C., Ehringer, M.A., Pitulle, C., and Pace, N.R. 2000. Phylogenetic-comparative analysis of the eukaryal ribonuclease P RNA. *RNA* **6:** 1895–1904.

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.

Haas, E.S., Brown, J.W., Pitulle, C., and Pace, N.R. 1994. Further perspective on the catalytic core and secondary structure of ribonuclease P RNA. *Proc. Natl. Acad. Sci.* **91:** 2527–2531.

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.

Harris, J.K., Haas, E.S., Williams, D., Frank, D.N., and Brown, J.W. 2001. New insight into RNase P RNA structure from comparative analysis of the archaeal RNA. *RNA* **7:** 220–232.

Houge, G., Robaye, B., Eikhom, T.S., Golstein, J., Mellgren, G., Gjert-

sen, B.T., Lanotte, M., and Doskeland, S.O. 1995. Fine mapping of 28S rRNA sites specifically cleaved in cells undergoing apoptosis. *Mol. Cell. Biol.* **15:** 2051–2062.

Jacobson, M.R., Cao, L.G., Taneja, K., Singer, R.H., Wang, Y.L., and Pederson, T. 1997. Nuclear domains of the RNA subunit of RNase P. *J. Cell Sci.* **110:** 829–837.

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.

Jayanthi, G.P. and Van, T.G. 1992. Characterization of ribonuclease P isolated from rat liver cytosol. *Arch. Biochem. Biophys.* **296:** 264–270.

Jeeninga, R.E., Van Delft, Y., de Graaff-Vincent, M., Dirks-Mulder, A., Venema, J., and Raue, H.A. 1997. Variable regions V13 and V3 of *Saccharomyces cerevisiae* contain structural features essential for normal biogenesis and stability of 5.8S and 25S rRNA. *RNA* **3:** 476–488.

Keister, D.B. 1983. Axenic culture of *Giardia lamblia* in TYI-S-33 medium supplemented with bile. *Trans. R. Soc. Trop. Med. Hyg.* **77:** 487–488.

Kelley, S.T., Akmaev, V.R., and Stormo, G.D. 2000. Improved statistical methods reveal direct interactions between 16S and 23S rRNA. *Nucleic Acids Res.* **28:** 4938–4943.

Kirsebom, L.A. and Svard, S.G. 1994. Base pairing between *Escherichia coli* RNase P RNA and its substrate. *EMBO J.* **13:** 4870–4876.

Krasilnikov, A.S., Yang, X., Pan, T., and Mondragon, A. 2003. Crystal structure of the specificity domain of ribonuclease P. *Nature* **421:** 760–764.

Krasilnikov, A.S., Xiao, Y., Pan, T., and Mondragon, A. 2004. Basis for structural diversity in homologous RNAs. *Science* **306:** 104–107.

Krupp, G., Cherayil, B., Frendewey, D., Nishikawa, S., and Söll, D. 1986. Two RNA species co-purify with RNase P from the fission yeast *Schizosaccharomyces pombe*. *EMBO J.* **5:** 1697–1703.

Lee, Y.C., Lee, B.J., and Kang, H.S. 1996. The RNA component of mitochondrial ribonuclease P from *Aspergillus nidulans*. *Eur. J. Biochem.* **235:** 297–303.

Li, K. and Williams, R.S. 1995. Cloning and characterization of three new murine genes encoding short homologues of RNase P RNA. *J. Biol. Chem.* **270:** 25281–25285.

Noller, H.F. and Woese, C.R. 1981. Secondary structure of 16S ribosomal RNA. *Science* **212:** 403–411.

Noller, H.F., Kop, J., Wheaton, V., Brosius, J., Gutell, R.R., Kopylov, A.M., Dohme, F., Herr, W., Stahl, D.A., Gupta, R., et al. 1981. Secondary structure model for 23S ribosomal RNA. *Nucleic Acids Res.* **9:** 6167–6189.

Oh, B.-K. and Pace, N.R. 1994. Interaction of the 3′-end of tRNA with ribonuclease P RNA. *Nucleic Acids Res.* **22:** 4087–4094.

Pace, N.R. and Brown, J.W. 1995. Evolutionary perspective on the structure and function of ribonuclease P, a ribozyme. *J. Bacteriol.* **177:** 1919–1928.

Pace, N.R., Smith, D.K., Olsen, G.J., and James, B.D. 1989. Phylogenetic comparative analysis and the secondary structure of ribonuclease P RNA—A review. *Gene* **82:** 65–75.

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.* **96:** 7803–7808.

Pitulle, C., Garcia-Paris, M., Zamudio, K.R., and Pace, N.R. 1998. Comparative structure analysis of vertebrate ribonuclease P RNA. *Nucleic Acids Res.* **26:** 3333–3339.

RajBhandary, U.L., Stuart, A., Faulkner, R.D., Chang, S.H., and Khorana, H.G. 1966. Nucleotide sequence studies on yeast phenylalanine sRNA. *Cold Spring Harb. Symp. Quant. Biol.* **31:** 425–434.

Reich, C., Gardiner, K.J., Olsen, G.J., Pace, B., Marsh, T.L., and Pace, N.R. 1986. The RNA component of the *Bacillus subtilis* RNase P: Sequence, activity, and partial secondary structure. *J. Biol. Chem.* **261:** 7888–7893.

Rimm, D.L. and Pollard, T.D. 1989. Purification and characterization of an Acanthamoeba nuclear actin-binding protein. *J. Cell Biol.* **109:** 585–591.

Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. *Molecular cloning: A laboratory manual*, 2d ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schmitt, M.E., Bennett, J.L., Dairaghi, D.J., and Clayton, D.A. 1993. Secondary structure of RNase MRP RNA as predicted by phylogenetic comparison. *FASEB J.* **7:** 208–213.

Siegel, R.W., Banta, A.B., Haas, E.S., Brown, J.W., and Pace, N.R. 1996. *Mycoplasma fermentans* simplifies our view of the catalytic core of ribnuclease P RNA. *RNA* **2:** 452–462.

Sogin, M.L. 1991. Early evolution and the origin of eukaryotes. *Curr. Opin. Genet. Dev.* **1:** 457–463.

Sogin, M. 1997. History assignment: When was the mitochondrion founded? *Curr. Opin. Genet. Dev.* **7:** 792–799.

Tranguch, A.J. and Engelke, D.R. 1993. Comparative structural analysis of nuclear RNase P RNAs from yeast. *J. Biol. Chem.* **268:** 14045–14055.

Van de Peer, Y., Ben Ali, A., and Meyer, A. 2000. Microsporidia: Accumulating molecular evidence that a group of amitochondriate and suspectedly primitive eukaryotes are just curious fungi. *Gene* **246:** 1–8.

van Nues, R.W., Venema, J., Planta, R.J., and Raue, H.A. 1997. Variable region V1 of *Saccharomyces cerevisiae* 18S rRNA participates in biogenesis and function of the small ribosomal subunit. *Chromosoma* **105:** 523–531.

Wise, C.A. and Martin, N.C. 1991. Dramatic size variation of yeast mitochondrial RNAs suggests that RNase P RNAs can be quite small. *J. Biol. Chem.* **266:** 19154–19157.

Woese, C.R. and Pace, N.R. 1993. Probing RNA structure, function, and history by comparative analysis. In *The RNA world* (eds. R.F.G. Atkins and J.F. Gesteland), pp 91–117. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

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.

Ziehler, W.A., Lee, Y., Pagan-Ramos, E., and Engelke, D.R. 1997. An active domain of the nuclear RNase P RNA. *Nucleic Acids Symp. Ser.* **36:** 45–48.