Phylogenetic-comparative analysis of the eukaryal ribonuclease P RNA

D. N. Frank, C. Adamidi, M. A. Ehringer, C. Pitulle and N. R. Pace

RNA 2000 6: 1895-1904

RNA (2000), *6*:1895–1904. Cambridge University Press. Printed in the USA.

Phylogenetic-comparative analysis of the eukaryal ribonuclease P RNA

DANIEL N. FRANK,[1] CATHERINE ADAMIDI,[2] MARISSA A. EHRINGER,[3] CHRISTIAN PITULLE,[4] and NORMAN R. PACE[1]

[1]Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, Colorado 80309-0347, USA
[2]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032, USA
[3]Department of Pharmacology, University of Colorado Health Sciences Center, Denver, Colorado 80262, USA
[4]Department of Clinical Sciences, College of Veterinary Medicine, North Carolina State University, Raleigh, North Carolina 27606, USA

ABSTRACT

Ribonuclease P (RNase P) is the ribonucleoprotein enzyme that cleaves 5′-leader sequences from precursor-tRNAs. Bacterial and eukaryal RNase P RNAs differ fundamentally in that the former, but not the latter, are capable of catalyzing pre-tRNA maturation in vitro in the absence of proteins. An explanation of these functional differences will be assisted by a detailed comparison of bacterial and eukaryal RNase P RNA structures. However, the structures of eukaryal RNase P RNAs remain poorly characterized, compared to their bacterial and archaeal homologs. Hence, we have taken a phylogenetic-comparative approach to refine the secondary structures of eukaryal RNase P RNAs. To this end, 20 new RNase P RNA sequences have been determined from species of ascomycetous fungi representative of the genera *Arxiozyma*, *Clavispora*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Saccharomycopsis*, *Torulaspora*, *Wickerhamia*, and *Zygosaccharomyces*. Phylogenetic-comparative analysis of these and other sequences refines previous eukaryal RNase P RNA secondary structure models. Patterns of sequence conservation and length variation refine the minimum-consensus model of the core eukaryal RNA structure. In comparison to bacterial RNase P RNAs, the eukaryal homologs lack RNA structural elements thought to be critical for both substrate binding and catalysis. Nonetheless, the eukaryal RNA retains the main features of the catalytic core of the bacterial RNase P. This indicates that the eukaryal RNA remains intrinsically a ribozyme.

Keywords: ribozyme; RNA processing; tRNA

INTRODUCTION

Ribonuclease P (RNase P) is the ribonucleoprotein endonuclease that functions in pre-tRNA maturation by removing 5′-precursor sequences (Frank & Pace, 1998). As a key tRNA processing enzyme, RNase P is found in all three phylogenetic domains (Darr et al., 1992). In Bacteria and a subset of Archaea, the RNA component of RNase P is catalytically active in vitro in the absence of protein subunits (Guerrier-Takada et al., 1983; Pannucci et al., 1999). In contrast, among members of the Eukarya, the RNase P RNAs are not catalytically active by themselves; rather, eukaryal RNase P activity requires cooperation between RNA and protein subunits.

Numerous RNase P RNAs have been characterized from Bacteria and Archaea, and the global architecture of these RNAs is becoming increasingly well characterized (Harris et al., 1997; Chen et al., 1998). Moreover, significant progress has been made in mapping functional aspects of the enzyme with respect to its structure in several bacterial RNase P RNA model systems (e.g., *Escherichia coli*, *Bacillus subtilis*, and *Thermus thermophilus*; reviewed in Kirsebom, 1995; Harris et al., 1998). Because the functional differences between catalytically active and inactive RNase P RNAs undoubtedly arise from variations in RNA structure, elucidation of the structural features that are present in bacterial RNase P RNAs, but absent in those of the Eukarya, can provide meaningful insight into why the latter RNAs are not enzymatically active independently of protein. The analysis of eukaryal RNase P has been limited, however, to a few animals and fungi, with the RNAs proving to differ substantially from one another and from their bacterial and archaeal homologs (Baer et al., 1990; Doria et al., 1991; Tranguch & Engelke, 1993; Eder et al., 1996). Nevertheless, a few highly

Reprint requests to: Norman R. Pace, Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, Colorado 80309-0347, USA; e-mail: nrpace@colorado.edu.

conserved sequence elements can be identified within eukaryal RNase P RNAs that are similar to bacterial RNA sequences (Forster & Altman, 1990; Chen & Pace, 1997; Pitulle et al., 1998). Intriguingly, several of these conserved regions are thought to form a part of the active site of the catalytically active bacterial RNase P RNA, suggesting that the eukaryal RNase P holoenzyme retains vestiges of RNA-mediated catalysis (Forster & Altman, 1990; Chen & Pace, 1997; Pitulle et al., 1998).

Because an understanding of eukaryal RNase P structure and function is limited by the paucity of structural information concerning the RNA subunits of these enzymes, we have taken a phylogenetic approach to refine the current models of eukaryal RNase P RNA secondary structure. Based on conserved regions within these molecules, PCR primers were designed to amplify the genes encoding nuclear RNase P RNA components from a variety of fungal species that span several genera (Fig. 1); 20 new full-length and partial RNase P RNA sequences were determined in this manner. Phylogenetic-comparative analysis of these and additional sequences proved fruitful in refining the secondary structure of the eukaryal RNase P RNA and in identifying conserved sequence elements that are likely to comprise the functional core of the enzyme. Comparisons to bacterial RNase P RNA sequences and enzymatic properties suggest that eukaryal RNase P RNAs may be deficient in activities required for both substrate binding and chemical catalysis.

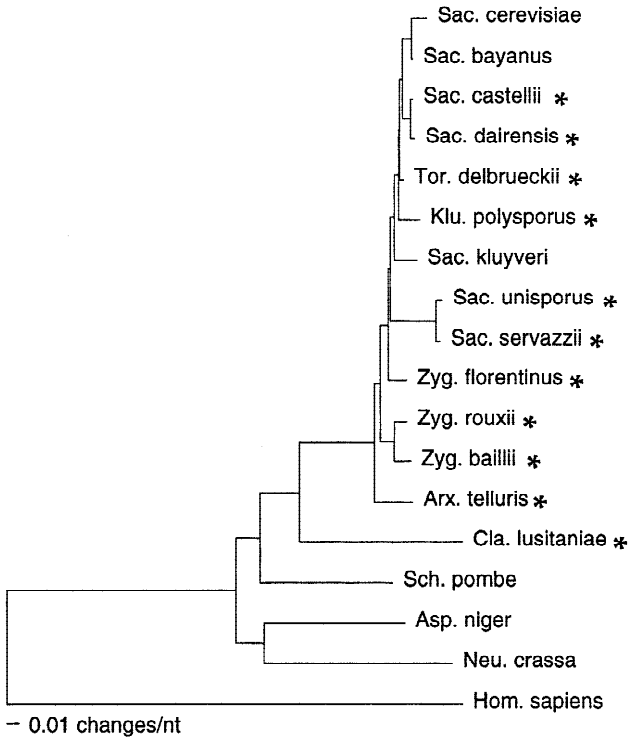

FIGURE 1. Phylogenetic distribution of yeast species from which RNase P RNA sequences have been isolated. Available 18S rRNA sequences were subjected to distance-matrix phylogenetic analysis, using PAUP* (Swofford, 1999). Yeast species from which RNase P RNA sequences were isolated in this study are marked with asterisks. The 18S sequences of *H. sapiens*, *A. niger* and *N. crassus* are included as outgroups.

RESULTS AND DISCUSSION

Based on an alignment of known *Saccharomyces* and *Schizosaccharomyces* nucleus-encoded RNase P RNA genes, we designed PCR primers complementary to conserved sequence elements within these genes. Reverse PCR primers were designed to anneal to the highly-conserved P4 helix and its flanking sequence (Conserved Region- [CR] V; Fig. 2). Forward PCR primers were complementary to either the P4 region (CR-I) or the more phylogenetically heterogenous P1 helix. Partial RNase P RNA genes were amplified, via polymerase chain reaction, from 20 fungal species, including representatives of *Arxiozyma*, *Clavispora*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Saccharomycopsis*, *Torulaspora*, *Wickerhamia*, and *Zygosaccharomyces* (Table 1). These genera span the phylogenetic distance between *Saccharomyces* and *Schizosaccharomyces* (Fig. 1). The RNase P RNA sequence of *Saccharomyces pastorianus* was identical to that of *Saccharomyces carlsbergensis* (Tranguch & Engelke, 1993).

In several instances, full-length sequences were obtained by a strategy in which annealed oligonucleotides of defined sequence (adaptors) were ligated to endonucleolytically restricted genomic DNA. Ligation products were then subjected to PCR with RNase P-

TABLE 1. Yeast strains and RNase P RNA accession numbers.

Organism	Strain	Accession No.
Arxiozyma telluris	NRRL YB-4302	AF186214
Clavispora lusitaniae	CG78-75	AF186215
Clavispora opuntia	CG28-540	AF186216
Kluyveromyces polysporus	NRRL Y-8283	AF186217
Kluyveromyces thermotolerans	NRRL Y-2233	AF186218
Pichia canadensis	NRRL Y-2340	AF186219
Pichia guillermondii	NRRL Y-2075	AF186220
Pichia mississippiensis	NRRL YB-1294-7	AF186221
Pichia strasburgensis	NRRL Y-11980	AF186222
Saccharomyces castellii	NRRL Y-12630	AF186223
Saccharomyces dairensis	NRRL Y-12639	AF186224
Saccharomyces pastorianus	NRRL Y-1525	AF186225
Saccharomyces servazzii	NRRL Y-12661	AF186226
Saccharomyces unisporus	NRRL Y-1556	AF186227
Saccharomycopsis fibuligera	NRRL Y-2388	AF186228
Torulaspora delbrueckii	NRRL Y-866	AF186229
Wickerhamia fluorescens	NRRL Y-4819	AF186230
Zygosaccharomyces bailii	NRRL Y-2227	AF186231
Zygosaccharomyces florentinus	NRRL Y-1560	AF186232
Zygosaccharomyces rouxii	NRRL Y-229	AF186233

specific and adaptor-specific primers. Positive PCR products were detected by either ethidium staining of agarose gels or by Southern blotting with probes generated by nick-translation of cloned, partial-length PCR products (see Materials and Methods). Presumably, amplification with the gene-specific primer produced an enrichment in the specific product that was sufficient to out-compete the expected nonspecific amplification products. Unlike more traditional blotting and cloning methodologies, this strategy permits more facile cloning, because PCR products can be cloned directly into TA-vectors. In addition, because the gene of interest is located at one end of the PCR product, further subcloning steps that might be required to localize the gene are obviated; gene sequences can be generated by vector-specific priming of sequencing reactions. In practice, the best results were obtained when several adaptor libraries were constructed and screened in parallel, each using a different restriction enzyme. Although we have not optimized this protocol, positive PCR clones were isolated in 5 of 12 species tested; the use of Southern blotting to identify clones was successful in 2 of 3 species tested.

Each of the putative RNase P RNA genes that we isolated encodes several regions of sequence that are highly similar to previously characterized fungal nuclear RNase P RNA genes (Tranguch & Engelke, 1993). Furthermore, each of the newly identified genes contains all of the signature nucleotides (i.e., CR-I, CR-II, CR-III, CR-IV, and CR-V) that are present in all known cellular RNase P RNA genes, including those of the Bacteria and Archaea (Chen & Pace, 1997). The presence of these conserved sequence elements at hallmark sites in the new genes unambiguously identifies them as templates for nucleus-encoded eukaryal RNase P RNA.

All available eukaryal RNase P RNA sequences were aligned with each other based on conserved elements of primary and secondary RNA structure (this alignment is available at http://pacelab.colorado.edu/publications.html/). The secondary structures of a subset of the RNase P RNA species analyzed are shown in Figure 2. To generalize the discussion, we designate as "P" (for paired region) those eukaryal helices that are present in all species and are similar in structure and location to bacterial and archaeal counterparts. These helices are taken to be homologs of the corresponding bacterial and archaeal helices (Haas et al., 1994). Helices designated "eP", for eukaryal paired region, seem for reasons of structure or variable occurrence to be eukaryote specific, not necessarily homologs of bacterial or archaeal helices that might occupy the same position in the universal secondary structure. Helices eP8 and eP9 may be homologs of the corresponding bacterial and archaeal helices, but the variability in spacing of helices in this region of eukaryal RNase P RNA sequences leaves the homology uncertain. Table 2 summarizes criteria by which we relate some commonly occurring eukaryal RNase P RNA helices to bacterial and archaeal features.

The new fungal RNase P RNA structures conform, in general, to previously proposed models of eukaryal RNase P RNA structure (Tranguch & Engelke, 1993;

TABLE 2. Phylogenetic occurrence of RNase P RNA substructures.

Helix[a]	Bacteria	Archaea	Eukarya	Criteria for assigning eukaryal structure
P1	+	+	+	Position relative to CR-V, P2
P2	+	+	+	Position relative to CR-IV, CR-V, P1, P3
P3	+	+	+	Position relative to CR-I, P2, P4
P4	+	+	+	Portion of CR-I, CR-V
P5	+	+	−	Complementary pairing absent in Eucarya
P6	+	+/−[b]	−	Complementary pairing absent in Eucarya
P7	+	+	+	Position relative to CR-IV, P10/P11
P8	+	+/−[c]	−	n.a.[d]
eP8	−	−	+	Proximity to P10/P11
P9	+	+	−	n.a.
eP9	−	−	+	Proximity to eP8
P10	+	+	+	Position relative to CR-II, CR-III, P11
P11	+	+	+	Position relative to CR-II, CR-III, P10
P12	+	+	+	Position relative to CR-II, CR-III
P15	+	+[e]	−	n.a.
eP15	−	−	+	Position relative to CR-IV, P7
P19	+	+	−	n.a.
eP19	−	−	+	Position relative to CR-V, P2

[a]Helix numbering corresponds to the *E. coli* structure (Haas et al., 1994).
[b]Absent in *Methanothermus fervidus* (Haas et al., 1996), *Methanococcus jannaschii* (Bult et al., 1996).
[c]Absent in *Methanococcus jannaschii* (Bult et al., 1996).
[d]n.a.: not applicable.
[e]Homology with bacterial structure uncertain.

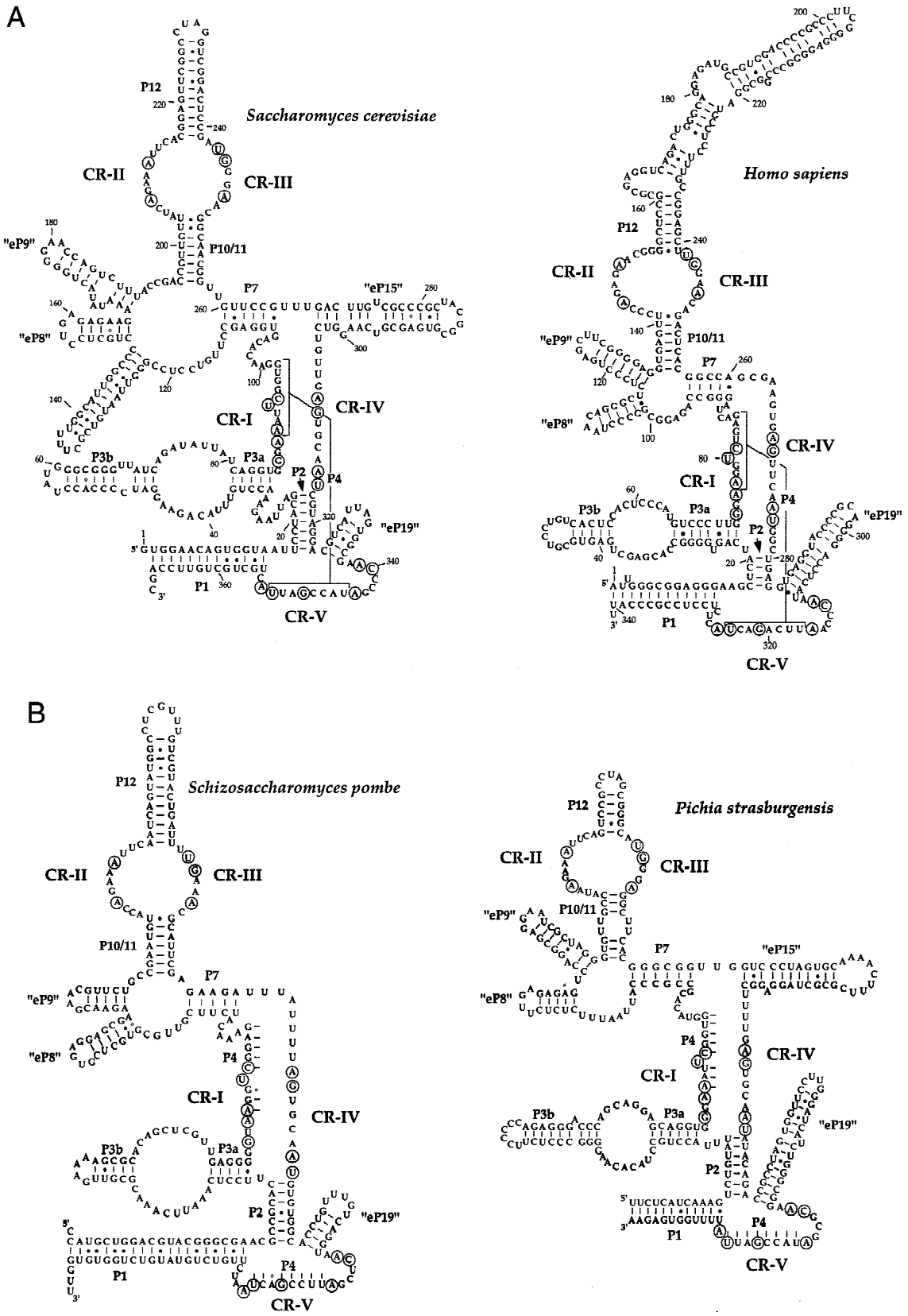

FIGURE 2. Secondary structures of eukaryal RNase P RNAs. Nucleotides that are absolutely conserved in archaeal, bacterial, and eukaryal RNase P RNAs are circled. The nomenclature for these conserved regions (CRs) is that of Chen and Pace (1997). Helices are numbered based on their putative homology to bacterial structures (Haas et al., 1994). Eukaryal helices of uncertain homology to particular bacterial or archaeal structures, but which occur at the same positions in the structures are designated "eP", for eukaryal paired region. Lower-case nucleotides represent primer sequences. (*Figure continues on facing page.*)

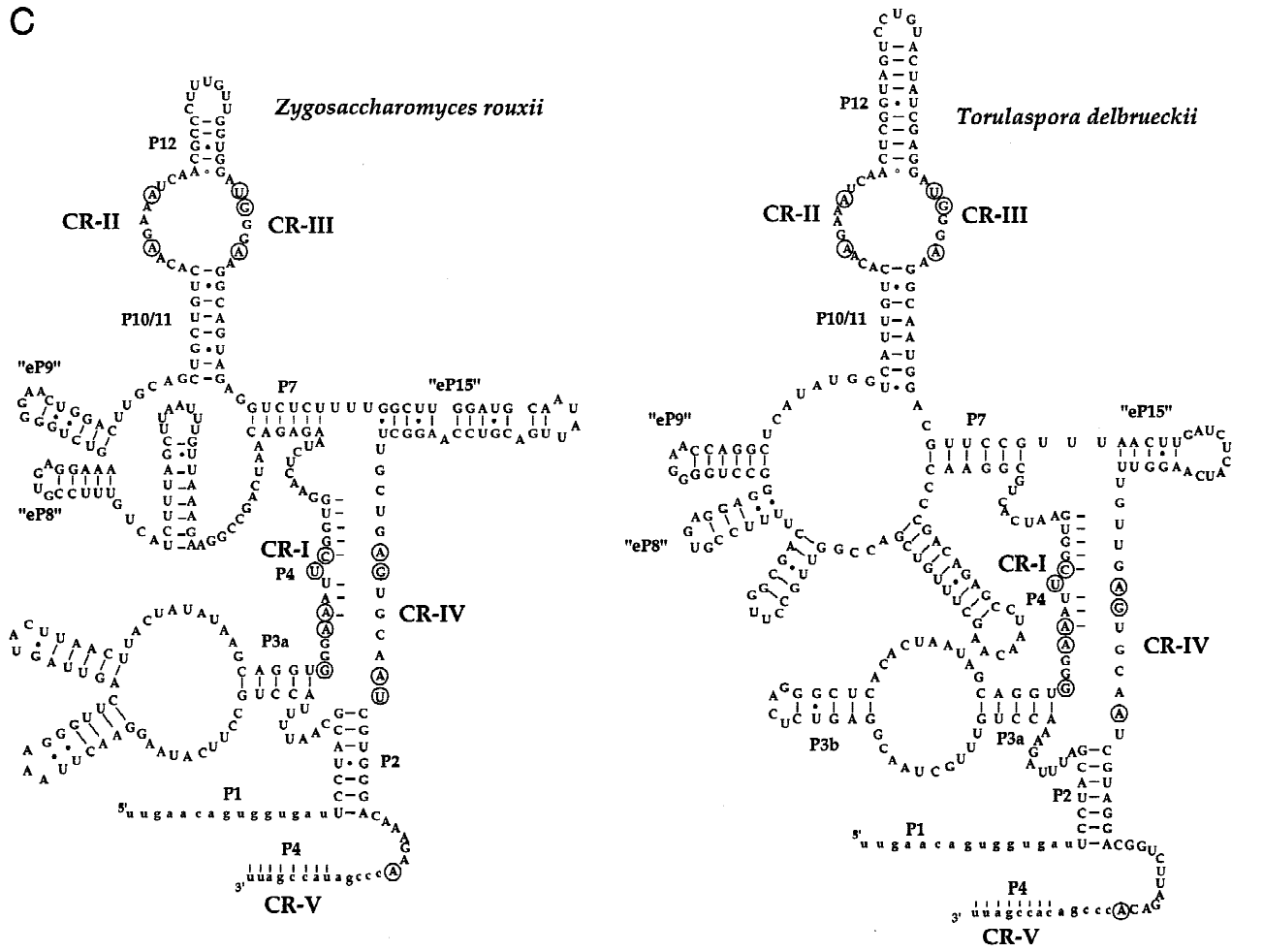

FIGURE 2. *Continued.*

Chen & Pace, 1997; Pitulle et al., 1998) and provide numerous examples of sequence covariations that support the existence of the conserved helices depicted in Figure 2 (i.e., P3, P4, eP9, P10/11, P12). The revised fungal secondary structure models differ from previously proposed fungal RNase P RNA structures (Tranguch & Engelke, 1993) in predicting the existence of helix eP8. Sequence alignments of both eP8 and eP9 (Fig. 3) reveal many covarying base pairs that provide evidence for these hairpins. All fungal eP8 hairpins, with the exception of those of *Schizosaccharomyces spp.*, contain NUGA tetraloops (the loops of *Schizosaccharomyces pombe* and *Schizosaccharomyces octosporus*, in contrast, have the sequence GUGAG). With few exceptions, fungal eP9 hairpin loops are of the GNRA type. In most instances, eP8 is found less than 2 nt 5′ of eP9. Although the spacing between eP8 and eP9 in *Saccharomyces dairensis* is 23 nt, the sequence alignments unambiguously identify these structures in *S. dairensis*. Fungal eP8 and eP9 structures cannot be aligned readily at the primary sequence level with the proposed vertebrate structures. However, the overall arrangement of helices in the P7/eP8/eP9/P10/P11 regions of several fungal RNase P RNAs (e.g., *S. pombe*, *Pichia strasburgensis*) is strikingly similar to that of most vertebrate RNase P RNAs (e.g. *Homo sapiens*, *Danio rerio*). In these instances, eP8 and eP9, the only helices found in the short span of nucleotides that separates P7 from P10/11, are situated adjacent to P10/P11. Consequently, we propose that the fungal and vertebrate examples of eP8 and eP9 are homologs, despite the absence of shared sequence identity.

The new sequence data do not support the existence of the previously proposed eukaryal pairing that corresponds to the bacterial helix P5 (Tranguch & Engelke, 1993; Chen & Pace, 1997). In bacterial and archaeal RNase P RNAs, helix P5 lies immediately adjacent to the 5′ end of helix P4. The complementarity required to form the corresponding helix generally is absent from the eukaryal sequences. In *H. sapiens*, for instance (Fig. 2), formation of P5 would require pairing of the trinucleotide A86-C87-U88 with G263-C262-G261. Similarly, formation of P5 would re-

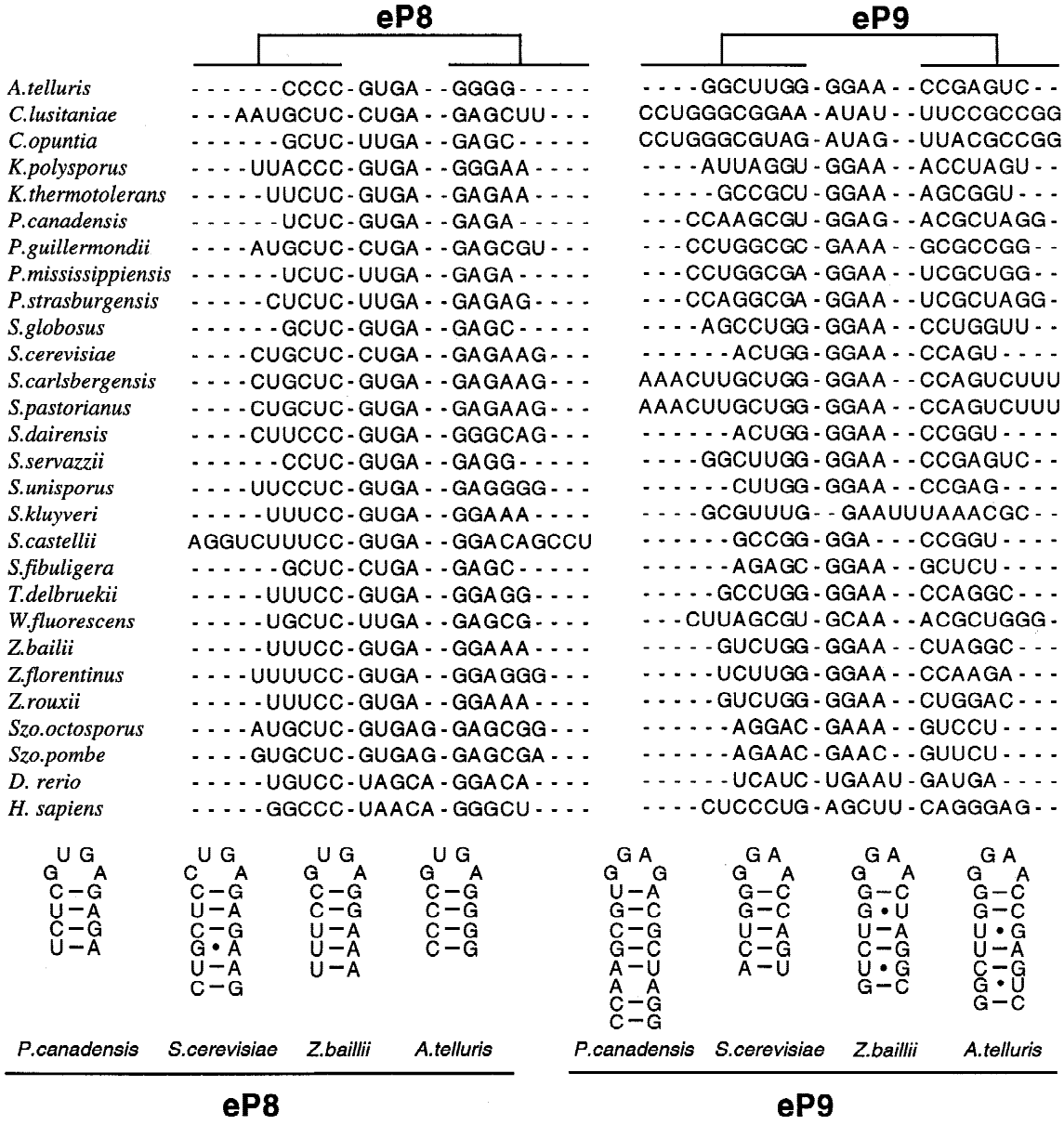

FIGURE 3. Multiple alignment of eP8 and eP9 sequences. A subset of the eukaryal RNase P RNA sequence alignment that corresponds to the hairpin structures eP8 and eP9 is shown in the upper portion of the figure. Overlining denotes nucleotides proposed to form helices. Several examples of these structures are shown in the lower portion of the figure. Note covariations in sequence identity that maintain base pairing, thus providing support for the existence of the helices.

quire U-U and G-G appositions in *P. strasburgensis* (Fig. 2). Like P5, there is no sequence complementarity that would indicate a homolog of the bacterial and archaeal P6. The apparent absence in the Eukarya of homologs of the bacterial/archaeal helix P5 and P6 may suggest a lack of structural rigidity between the core of eukaryal RNase P RNA (i.e., helices P1, P2, P3, and P4, the catalytic center of the bacterial RNA) and the remainder of the RNA (i.e., helices eP7, eP8, eP9, P10/11, and P12).

In contrast to the known examples of vertebrate RNase P RNA, many of the fungal RNAs contain regions of extensive length variation, particularly in the region between helices eP7 and eP8 (e.g., *Torulaspora delbrueckii*, *Zygosaccharomyces rouxii*). In many cases, elements of secondary structure in these regions can be inferred from base complementarity, but are not strictly proven by the phylogenetic criteria of covariations in the data set. All bacterial and most archaeal RNase P RNAs, on the other hand, maintain a cruciform structure in this region, formed by helices P7, P8, P9, and P10/P11; no additional helices are observed within this structural domain. The fungal RNase P RNAs, with the exception of those of *Schizosaccharomyces* species, also differ from the vertebrate RNAs in possessing a helix 5′ of the CR-IV

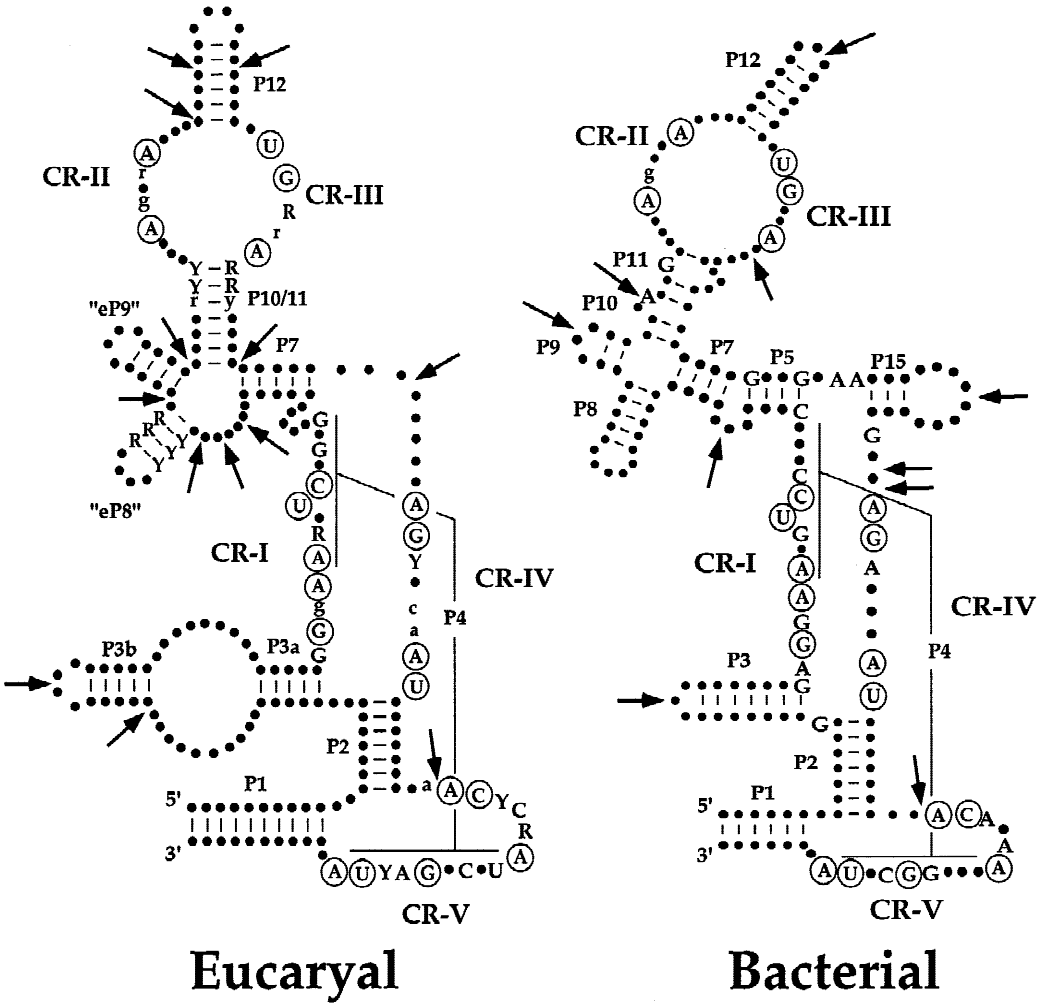

FIGURE 4. Phylogenetic minimum-consensus RNase P RNA secondary structures. Nucleotides that are absolutely conserved in archaeal, bacterial, and eukaryal RNase P RNAs are circled. The nomenclature for these conserved regions (CRs) is that of Chen and Pace (1997). Helices are numbered based on their putative homology to bacterial structures (Haas et al., 1994). Eukaryal helices of uncertain homology to bacterial or archaeal structures are designated "eP", for eukaryal paired region. Nucleotides that are invariant within the Eukarya are shown in upper case and those conserved at the 90% level are shown in lower case (R: purine, Y: pyrimidine). Nucleotide positions that are found in all eukaryal RNAs, but are not conserved in sequence identity, are shown as black circles. Arrows represent sites at which one or more phylogenetically variable helices are inserted in selected species.

region (designated eP15; Tranguch & Engelke, 1993; Fig. 2). Although the fungal helix occurs in the same position in the structure as the bacterial and archaeal P15, as indicated by its conserved sequence length from CR-IV, there is no sequence identity between the eukaryal and archaeal/bacterial structures. This finding, coupled with the absence of P15 in the vertebrate and *Schizosaccharomyces* RNAs, suggests a model of RNase P evolution in which the ancestral P15 structure (i.e., that which was retained in the Archaea and Bacteria) was lost during the evolution of eukaryal RNase P RNA and then replaced by a new and idiosyncratic structure in most yeasts, subsequent to their split from *Schizosaccharomyces*. Thus, the structure of the *Saccharomyces cerevisiae* RNA can be seen as basically the structure of the *S. pombe* RNA with a few added helical elements.

The conserved features of the known eukaryal RNase P RNAs, the sequences and structures that are present in all known examples of the eukaryal RNA, define a minimum-consensus RNA secondary structure (Fig. 4, Table 2). These phylogenetically conserved structures, most of which are also conserved in archaeal and bacterial RNAs (Fig. 4), are likely to comprise the functional core of the eukaryal RNase P RNA. Although the minimum-consensus RNA is a hypothetical construct, it bears striking resemblance to several native RNase P RNAs, including species of the *Pichia*, *Clavispora*, and *Schizosaccharomyces* (Fig. 2), showing that these RNAs have been significantly pared down

during their evolution, compared to the bacterial and archaeal versions, yet have retained functionality in the in vivo context, in the presence of protein subunits.

Comparison of the eukaryal and bacterial RNase P RNAs provides some clues as to the catalytic inactivity of the eukaryal, in contrast to the bacterial, version. The eukaryal RNAs lack features of the bacterial RNA that are thought to be important for both substrate binding and catalysis. For example, the lack of the bacterial-type helix P15, which participates in the binding of the 3′-CCA of precursor-tRNA by bacterial RNase P RNAs (Kirsebom & Svard, 1994; Oh & Pace, 1994; Oh et al., 1998), indicates that the eukaryal RNAs may be defective in binding to pre-tRNA-CCA substrates. Although it has yet to be conclusively determined whether, in fact, 3′-CCA is a binding determinant of the eukaryal RNase P holoenzyme, this sequence is not absolutely necessary for cleavage (Lee et al., 1997). Additionally, eukaryal RNase P RNAs so far examined, especially those of the fungi, appear to have much less constrained structures than the Bacteria and Archaea in the P7–P10/11 region. This is evidenced by the occurrence of a multitude of nonconserved, single-stranded, and helical regions in the fungal RNAs (Fig. 2 and Fig. 4). In the bacterial RNA, this region is thought to contact a portion of the pre-tRNA substrate (Nolan et al., 1993; Pan et al., 1995; Loria & Pan, 1997). The evolutionarily volatile nature of this region in the eukaryal RNase P RNAs also may denote a diminished capacity of the RNA to bind substrate pre-tRNA, at least in the absence of auxiliary proteins. Indeed, tRNAs containing photoaffinity crosslinking agents and which readily crosslink to bacterial RNase P RNAs have not been reported to crosslink to the eukaryal RNA.

Although not catalytically active independently of protein, the eukaryal RNase P RNAs all contain homologs of most of the structures and sequences that are thought to be critical for catalysis by bacterial RNase P RNAs. These structures include the highly conserved P4 helix and its flanking sequences CR-I and CR-V (Harris & Pace, 1995; Frank et al., 1996; Frank & Pace, 1997; Kazantsev & Pace, 1998), which are thought to constitute the catalytic center of the RNase P RNA. We conjecture, therefore, that RNase P in eukaryotes, although not catalytically active without protein subunits, remains intrinsically an RNA enzyme. Some features of the bacterial RNA thought to be important for catalysis are nonetheless absent from the eukaryotic version. These include the P15 loop, into which the substrate 3′-CCA apparently docks during the reaction (Kirsebom & Svard, 1994; Oh & Pace, 1994). Disruption of the docked structure by chemical modification or mutagenesis additionally and significantly reduces the rate of the chemical step of the reaction (Perreault & Altman, 1992, 1993; Oh et al., 1998), so the P15 loop-3′-CCA interaction somehow contributes to formation of the active site. This is not unexpected considering the proximity of the 3′-CCA to the scissile bond, on the opposite strand of the acceptor stem double helix. By analogy, one might predict that the eukaryal RNase P catalytic reaction would also require stabilization of its substrate's 3′ end (either the immature 3′ trailer sequence or mature 3′ end). However, the absence of a bacterial-type P15 structure in eukaryal RNase P RNAs indicates that other structures within the holoenzyme must participate in such an interaction. Additionally, some base identities that are conserved in all Bacteria and are known from mutagenesis (M.A. Rubio, D.N. Frank, & N.R. Pace, unpublished) or crosslinking (Burgin & Pace, 1990) to interact with, or occur in the vicinity of, the active site are altered in eukaryal RNAs. For instance, the AA dinucleotide immediately 5′ of P15 in all bacterial RNAs (nt 248 and 249 in *E. coli*) is not present in the eukaryal RNAs (Figs. 2 and 4). Mutation of either of these A residues in the *E. coli* RNA to pyrimidines drastically reduces activity, and both crosslink with high efficiency to a crosslinking agent attached to the 5′ phosphate of mature tRNA, the substrate phosphate in pre-tRNA. Other examples with similar properties include adenosine residues in CR-IV and CR-V, which are universally present in bacterial RNase P RNAs, but replaced with other bases in eukaryotes.

It is possible that protein constituents of the eukaryal RNase P holoenzyme have usurped some roles of the RNA in substrate binding, such as recognition of the immature 3′ trailers of pre-tRNA by protein–RNA contacts, or provision of a catalytic chemical group. Proteins also may play essential roles in folding or stabilizing eukaryal RNase P RNA structures, for example, as might be required by the P7–P10/11 region that contacts pre-tRNA substrates. Some helical elements that are present in the bacterial, but not the eukaryal, RNase P RNA are known to contribute to the stability of global folding. Thus, the overall pared-down nature of the eukaryal RNA structure, compared to those of the bacterial or archaeal versions, may indicate a general instability of global packing of substructures in the eukaryal RNA (Waugh et al., 1989; Siegel et al., 1996). In the eukaryal holoenzyme, then, the proteins in essence might serve to squeeze and thereby compact the RNase P RNA into a functional structure. Additional comparative and functional analyses will be required to establish the validity of the hypothesis that the eukaryal RNase P in essence is a ribozyme.

MATERIALS AND METHODS

Yeast strains and genomic DNA isolation

The fungal strains used in this study, listed in Table 1, were obtained from Prof. C. Guthrie (University of California, San Francisco), Dr. C. Kurtzman (National Center for Agricultural Utilization Research), and Prof. R. Mortimer (Berkeley Yeast Genetic Strain Collection). Cultures were grown at 30 °C in

YEPD media (Sherman, 1991). Genomic DNA was isolated from fungal strains by either the protocol of Philippsen et al. (1991) or by use of the QIAmp® kit (Qiagen).

Oligonucleotides

PStras1F: 5′-TTCTCATCAAAGTCTGTA;
Pic1R: 5′-ATATTGCACTCAAYAGCC;
Sac3F: 5′-GGTGGGAAATTCGGTG;
Sac5F: 5′-GAACAGTGGTRATTCCTACG;
Sac1R: 5′-GCCTGCAGCGGCCGCAITRATCGGTATCGGG;
Sac3R: 5′-GTAATCGGTATCGGGTT;
Yea1R: 5′-GCCTGCAGCGGCCGCAITGGTCGGIATCGGGT;
Bam.1: 5′-GATCGGTACGCAGTC;
Hin.1: 5′-AGCTGGTACGCAGTC;
Xba.1: 5′-CTAGGGTACGCAGTC;
Uni.1: 5′-CTCGTAGACTGCGTACC.

Consensus PCR amplification of RNase P RNA genes

The core of each RNase P RNA gene was amplified with the helix P4-specific primers Sac3F (forward) and Yea1R (reverse). Additional sequence information was obtained for three species by amplification with the following primers: Sac5F/Sac2R (*T. delbrueckii*); Sac5F/Sac3R (*Z. rouxii*); and PStras1F/*Pic1R* (*Pichia mississippiensis*). Each 100-μL PCR reaction contained 30 mM Tris (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM each dNTP, 0.05% NP40, 1 U Taq polymerase, 200–500 ng of each primer, and 100–1,000 ng genomic DNA. For some reactions, amplification was more efficient in the presence of 5% acetamide. PCR amplification of RNase P RNA genes was carried out by either a standard regimen (e.g., 30 cycles, 92 °C, 1 min; 50 °C, 2 min; 72 °C, 3 min) or by "touchdown PCR" (20 cycles, 92 °C, 30 s; 65 °C, 30 s, −1 °C/cycle; 72 °C, 90 s; 20 cycles 92 °C, 30 s; 45 °C, 30 s; 72 °C, 90 s). For standard PCR reactions, multiple annealing temperatures were assayed between 45–55 °C and products were cloned from the highest temperature that produced DNA fragments of the expected length upon agarose gel electrophoresis. If several bands were present, the band of correct size was excised and its DNA purified by the QIAquick gel extraction kit (Qiagen) following the manufacturer's instructions. PCR products were cloned into either the TA Cloning® or TOPO Cloning® kits (Invitrogen) following the manufacturer's instructions. Genes were amplified from pure cultures of colony-purified yeast strains and multiple clones were analyzed for homogeneity.

Adaptor PCR isolation of 5′ and 3′ ends of RNase P RNA genes

Samples of genomic DNA (~500 ng) were digested in parallel with one of the following restriction enzymes: *Bam*HI, *Bgl*II, *Bcl*I, *Hin*dIII, *Nhe*I, *Spe*I, or *Xba*I. After a 3-h incubation (37 °C or 50 °C), the enzymes were inactivated by either heat treatment (75 °C, 10 min) or phenol/chloroform extraction. The DNA samples were ethanol precipitated and then resuspended in 50 μL of 10 mM Tris (pH 8.3). Adaptors were annealed by mixing 2.5 μg of the oligonucleotides Bam.1, Hin.1, or Xba.1 with 2.5 μg of Uni.1 in 100 μL of H_2O, heating 5 min at 85 °C, and then slowly cooling to room temperature. One hundred nanograms of digested genomic DNA were then ligated to 50 ng of the appropriate adaptor (e.g., *Bam*HI, *Bgl*II, and *Bcl*I digested DNAs were ligated to the Bam.1 adaptor) by overnight incubation at 16 °C [20 μL volume; 10 U T4 DNA ligase (New England Biolabs), 1× NEB ligase buffer, 0.4 mM rATP]. PCR amplification of 0.5 μL of each ligation mix was carried out as described above using primer pairs Uni.1/Sac3F to amplify 3′ ends of RNase P RNA genes or Uni.1/Sac1R to amplify 5′ ends. PCR products of *S. dairensis* and *Saccharomyces castellii* ligation reactions were identified by ethidium bromide staining following agarose gel electrophoresis. In the cases of *Pichia canadensis* and *P. strasburgensis*, PCR products were identified by Southern blots of agarose gels using the Phototope-Star Chemiluminescent Detection Kit (New England Biolabs). Nonradioactive probes were made by nick-translation (NEBlot Phototope Kit; New England Biolabs) of cloned, partial RNase P RNA genes of *P. canadensis* and *P. strasburgensis*. To clone the DNA fragments detected by Southern blots, 300 μL PCR reactions were performed and the DNA products precipitated and separated by agarose gel electrophoresis. Gel slices corresponding in size to the products that had been detected by Southern blot were isolated and the DNA purified by the QIAquick® kit (Qiagen). PCR products were cloned into either the TA Cloning® or TOPO Cloningrm kits (Invitrogen) following the manufacturer's instructions.

Sequence alignment and analysis

RNase P RNA genes were sequenced on an ABI 373A automated DNA sequencer, following the manufacturer's instructions. Newly determined sequences were deposited in GenBank and assigned accession numbers (Table 1). Previously identified RNase P RNA sequences were obtained from Pitulle et al. (1998) or GenBank: *D. rerio* (U50408), *H. sapiens* (X16612), *S. cerevisiae* (M27035), *S. carlsbergensis* (L12746), *Saccharomyces globosus* (L12748), *Saccharomyces kluyveri* (L12750), *S octosporus* (X52531), and *S. pombe* (X52530). Sequences were manually aligned using the application SeqApp (courtesy of Dr. Don Gilbert). RNA secondary structures were drawn in Canvas 5.0 (Deneba) or blueStem 1.0 (D. Frank). Phylogenetic analysis of 18S sequences used the application PAUP* (Swofford, 1999). 18S rRNA sequences were obtained from GenBank: *Saccharomyces unisporus* (Z75582), *Saccharomyces bayanus* (X97777), *S. castellii* (Z75577), *S. dairensis* (Z75579), *S. kluyveri* (Z75580), *Saccharomyces servazzii* (Z75581), *Arxiozyma telluris* (Y15809), *Zygosaccharomyces bailii* (X91083), *Zygosaccharomyces florentinus* (X91086), *S. cerevisiae* (Z75582*), Kluyveromyces polysporus* (X69845), *T. delbrueckii* (X53496), *Z. rouxii* (D01174), *Clavispora lusitaniae* (M60306), *S. pombe* (X58056), *Aspergillus niger* (X78538), *Neurospora crassa* (X04971), *H. sapiens* (M10098).

ACKNOWLEDGMENTS

We wish to thank Drs. C. Guthrie (University of California, San Francisco [UCSF]), C.P. Kurtzman (National Center for Agricultural Utilization Research), R. Mortimer (Berkeley Yeast

Genetic Stock Center), and H. Roiha (UCSF) for providing fungal strains. Kirk Harris and Scott Dawson provided excellent assistance in building phylogenetic trees. This work was supported by National Institutes of Health Grant GM 34527 (N.R.P.).

Received July 25, 2000; returned for revision August 16, 2000; revised manuscript received September 8, 2000

REFERENCES

Baer M, Nilsen TW, Costigan C, Altman S. 1990. Structure and transcription of a human gene for H1 RNA, the RNA component of human RNase P. *Nucleic Acids Res 18*:97–103.

Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb JF, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghagen NSM, Weidman JF, Fuhrmann JL, Nguyen D, Utterback TR, Kelley JM, Peterson JD, Sadow PW, Hanna MC, Cotton MD, Roberts KM, Hurst MA, Kaine BP, Borodovsky M, Klenk HP, Fraser CM, Smith HO, Woese CR, Venter JC. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science 27*:1058–1073.

Burgin AB, Pace NR. 1990. Mapping the active site of ribonuclease P RNA using a substrate containing a photoaffinity agent. *EMBO J 9*:4111–4118.

Chen JL, Nolan JM, Harris ME, Pace NR. 1998. Comparative photocross-linking analysis of the tertiary structures of *Escherichia coli* and *Bacillus subtilis* RNase P RNAs. *EMBO J 17*:1515–1525.

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

Darr SC, Brown JW, Pace NR. 1992. The varieties of ribonuclease P. *Trends Bio Sci 17*:178–182.

Doria M, Carrara G, Calandra P, Tocchini-Valentini GP. 1991. An RNA molecule copurifies with RNase P activity from *Xenopus laevis* oocytes. *Nucleic Acids Res 19*:2315–2320.

Eder PS, Srinivasan A, Fishman MC, Altman S. 1996. The RNA subunit of ribonuclease P from the zebrafish, *Danio rerio*. *J Biol Chem 271*:21031–21036.

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

Frank DN, Ellington AE, Pace NR. 1996. In vitro selection of RNase P RNA reveals optimized catalytic activity in a highly conserved structural domain. *RNA 2*:1179–1188.

Frank DN, Pace NR. 1997. In vitro selection for altered divalent metal specificity in the RNase P RNA. *Proc Natl Acad Sci USA 94*:14355–14360.

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

Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S. 1983. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell 35*:849–857.

Haas ES, Armbruster DW, Vucson BM, Daniels CJ, Brown JW. 1996. Comparative analysis of ribonuclease P RNA structure in Archaea. *Nucleic Acids Res 24*:1252–1259.

Haas ES, Brown JW, Pitulle C, Pace NR. 1994. Further perspective on the catalytic core and secondary structure of ribonuclease P RNA. *Proc Natl Acad Sci USA 91*:2527–2531.

Harris ME, Frank DN, Pace NR. 1998. Structure and catalytic function of the bacterial ribonuclease P ribozyme. In: Simons RW Grunberg-Manago M, eds., *RNA structure and function*. Plainview, New York: Cold Spring Harbor Laboratory Press. pp 309–337.

Harris ME, Kazantsev AV, Chen J-L, Pace NR. 1997. Analysis of the tertiary structure of the ribonuclease P ribozyme-substrate complex by site-specific photoaffinity crosslinking. *RNA 3*:561–576.

Harris ME, Pace NR. 1995. Identification of phosphates involved in catalysis by the ribozyme RNase P RNA. *RNA 1*:210–218.

Kazantsev AV, Pace NR. 1998. Identification by modification-interference of purine N-7 and ribose 2′-OH groups critical for catalysis by bacterial ribonuclease P. *RNA 4*:937–947.

Kirsebom LA. 1995. RNase P: A Scarlet Pimpernel. *Mol Microbiol 17*:411–420.

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

Lee Y, Kindelberger DW, Lee J-Y, McClennen S, Chamberlain J, Engelke DR. 1997. Nuclear pre-tRNA terminal structure and RNase P recognition. *RNA 3*:175–185.

Loria A, Pan T. 1997. Recognition of the T stem-loop of a pre-tRNA substrate by the ribozyme from *Bacillus subtilis* ribonuclease P. *Biochemistry 36*:6317–6325.

Nolan JM, Burke DH, Pace NR. 1993. Circularly permuted tRNAs as specific photoaffinity probes of ribonuclease P RNA structure. *Science 261*:762–765.

Oh BK, Frank DN, Pace NR. 1998. Participation of the 3′-CCA of tRNA in the binding of catalytic Mg^{2+} ions by ribonuclease P. *Biochemistry 37*:7277–7283.

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

Pan T, Loria A, Zhong K. 1995. Probing of tertiary interactions in RNA: 2′-hydroxyl-base contacts between the RNase P RNA and pre-tRNA. *Proc Natl Acad Sci USA 92*:12510–12514.

Pannucci JA, Haas ES, Hall TA, Harris JK, Brown JW. 1999. RNase P RNAs from some archaea are catalytically active. *Proc Natl Acad Sci USA 96*:7803–7808.

Perreault JP, Altman S. 1992. Important 2′-hydroxyl groups in model substrates for M1 RNA, the catalytic RNA subunit of RNase-P from *Escherichia coli*. *J Mol Biol 226*:399–409.

Perreault JP, Altman S. 1993. Pathway of activation by magnesium ions of substrates for the catalytic subunit of RNase P from *Escherichia coli*. *J Mol Biol 230*:750–756.

Philippsen P, Stotz A, Scherf C. 1991. DNA of *Saccharomyces cerevisiae*. In: Guthrie C, Fink GR, eds., *Guide to yeast genetics and molecular biology*. San Diego, California: Academic Press. pp 169–181.

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

Sherman F. 1991. Getting started with yeast. In: Guthrie C, Fink GR, eds., *Guide to yeast genetics and molecular biology*. San Diego, California: Academic Press. pp 3–20.

Siegel RW, Banta AB, Haas ES, Brown JW, Pace NR. 1996. *Mycoplasma fermentans* simplifies our view of the catalytic core of ribonuclease P RNA. *RNA 2*:452–462.

Swofford DL. 1999. *Phylogenetic analysis using parsimony (*and other methods)*, 4th ed. Sunderland, Massachusetts: Sinauer Associates.

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

Waugh DS, Green CJ, Pace NR. 1989. The design and catalytic properties of a simplified ribonuclease P RNA. *Science 244*:1569–1570.