Discovery of a minimal form of RNase P in *Pyrobaculum*

Lien B. Lai[a,1], Patricia P. Chan[b,1], Aaron E. Cozen[b], David L. Bernick[b], James W. Brown[c], Venkat Gopalan[a,2], and Todd M. Lowe[b,2]

[a]Department of Biochemistry and Center for RNA Biology, Ohio State University, 484 West Twelfth Avenue, Columbus, OH 43210; [b]Department of Biomolecular Engineering, University of California, 1156 High Street, Santa Cruz, CA 95064; and [c]Department of Microbiology, North Carolina State University, Raleigh, NC 27695-7615

Edited* by Sidney Altman, Yale University, New Haven, CT, and approved October 29, 2010 (received for review September 20, 2010)

RNase P RNA is an ancient, nearly universal feature of life. As part of the ribonucleoprotein RNase P complex, the RNA component catalyzes essential removal of 5′ leaders in pre-tRNAs. In 2004, Li and Altman computationally identified the RNase P RNA gene in all but three sequenced microbes: *Nanoarchaeum equitans*, *Pyrobaculum aerophilum*, and *Aquifex aeolicus* (all hyperthermophiles) [Li Y, Altman S (2004) *RNA* 10:1533–1540]. A recent study concluded that *N. equitans* does not have or require RNase P activity because it lacks 5′ tRNA leaders. The "missing" RNase P RNAs in the other two species is perplexing given evidence or predictions that tRNAs are trimmed in both, prompting speculation that they may have developed novel alternatives to 5′ pre-tRNA processing. Using comparative genomics and improved computational methods, we have now identified a radically minimized form of the RNase P RNA in five *Pyrobaculum* species and the related crenarchaea *Caldivirga maquilingensis* and *Vulcanisaeta distributa*, all retaining a conventional catalytic domain, but lacking a recognizable specificity domain. We confirmed 5′ tRNA processing activity by high-throughput RNA sequencing and in vitro biochemical assays. The *Pyrobaculum* and *Caldivirga* RNase P RNAs are the smallest naturally occurring form yet discovered to function as *trans*-acting precursor tRNA-processing ribozymes. Loss of the specificity domain in these RNAs suggests altered substrate specificity and could be a useful model for finding other potential roles of RNase P. This study illustrates an effective combination of next-generation RNA sequencing, computational genomics, and biochemistry to identify a divergent, formerly undetectable variant of an essential noncoding RNA gene.

catalytic RNA | gene finding | RNA processing

RNase P is best known for its role in removing the 5′ leaders of pre-tRNAs, an essential step in tRNA maturation. It also processes other RNAs in bacteria and eukaryotes, but these roles are less understood (1–3). RNase P typically functions as an RNA–protein complex, comprised of one conserved RNA and a varying number of protein subunits, depending on the domain of life: one in Bacteria, at least four in Archaea, and nine or more in the eukaryotic nucleus (4, 5). A precedent in which the RNA component is missing entirely is found in human and *Arabidopsis* organellar RNase P (6, 7), although a recent study suggests the possible coexistence of protein-only and RNA–protein-based RNase P complexes in human mitochondria (8).

The inability to identify RNase P in some organisms has sown doubts about whether it is a universal feature of life. Studies of the hyperthermophilic bacterium *Aquifex aeolicus* showed that it exhibits RNase P-like trimming of tRNAs (9, 10), yet a gene for the expected protein component is absent and the RNA has remained elusive (11), prompting speculation that it may have developed a unique solution for pre-tRNA processing (9). Perhaps most surprisingly, Söll and coworkers (12) demonstrated that the archaeal symbiont *Nanoarchaeum equitans* does not contain any identifiable RNase P genes or detectable RNase P activity, and it appears to lack tRNA leaders entirely. These findings leave RNase P conspicuously absent in just one other studied microbial species: *Pyrobaculum aerophilum*, a hyperthermophilic crenarchaeon that has been refractory to prior biochemical (12) and computational (13, 14) identification efforts. Now, with the advent of new genome and RNA sequencing, augmented by improved computational search methods, we were able to uncover a unique form of RNase P in multiple *Pyrobaculum* species and related genera.

Results and Discussion

Pre-tRNAs in *Pyrobaculum* Have 5′ Leaders. We first obtained evidence for RNase P activity in *Pyrobaculum* using comparative genomics and RNA sequencing. The genomes of four *Pyrobaculum* species [*Pyrobaculum arsenaticum, Pyrobaculum calidifontis, Pyrobaculum islandicum*, and *Thermoproteus neutrophilus* (to be reclassified as a *Pyrobaculum* species)] were recently sequenced in collaboration with the Joint Genome Institute, providing extensive comparative information. As in *P. aerophilum*, the RNase P RNA genes could not be identified in these genomes using existing computational methods (13, 14) (see *Materials and Methods*). However, alignment of orthologous tRNA loci and upstream promoter regions from all these *Pyrobaculum* species provided compelling evolutionary evidence for pre-tRNA leaders, thus hinting at a requirement for RNase P. If no RNase P activity is present to remove 5′ leaders, then one should expect very little or no variation in the distance between the TATA sequence and the 5′ end of the mature tRNA gene, especially among orthologs. In fact, we counted more than 12 tRNA ortholog groups where the distance between the promoter and tRNA gene varies by at least two nucleotides among species (Fig. S1). Next, we examined native transcripts from tRNA loci for four of these species to confirm the computational observations. High-throughput RNA sequencing reads of small RNAs identified many tRNA transcripts with 5′ leaders: 15 in *P. aerophilum*, 17 in *P. arsenaticum*, 18 in *P. calidifontis*, and 21 in *P. islandicum* have 1- to 6-nt leaders (Table 1 shows counts by leader length). All of the sequenced pre-tRNAs with 5′ leaders were also found in their

Author contributions: L.B.L. performed cloning, and designed and executed RNase P purification and pre-tRNA processing assays; P.P.C. performed promoter/sequencing analyses, identified Caldivirga and Vulcanisaeta RNase P RNAs, and created protein alignments; A.E.C. and D.L.B. grew Pyrobaculum cultures; A.E.C. purified total RNA, and performed northern analyses; D.L.B. designed and carried out RNA sequencing, and performed initial bioinformatic sequencing analyses; J.W.B. created secondary structure predictions for Pyrobaculum RNase P RNAs; V.G. guided biochemical studies; T.M.L. identified the Pyrobaculum RNase P RNAs, and analyzed the operonic context of RNase P proteins; L.B.L., V.G., and T.M.L. identified Pop5 and Rpp30 candidates; and L.B.L., P.P.C., V.G., and T.M.L. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

See Commentary on page 22371.

[1]L.B.L. and P.P.C. contributed equally to this work

[2]To whom correspondence may be addressed. E-mail: lowe@soe.ucsc.edu and gopalan.5@osu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1013969107/-/DCSupplemental.

Table 1. tRNA genes found to have a transcribed 5′ leader by high-throughput RNA sequencing

Genome	Total no. of tRNA genes in genome	No. of tRNA genes with 5′ leaders indicated by RNA sequencing results						
		1-nt	2-nt	3-nt	4-nt	5-nt	6-nt	Total
Pyrobaculum aerophilum	46	8	4	2	1	—	—	15
Pyrobaculum arsenaticum	46	8	3	—	6	—	—	17
Pyrobaculum calidifontis	46	13	3	1	—	1	—	18
Pyrobaculum islandicum	46	9	3	5	3	—	1	21

For all counts, the existence of a 5′ leader is consistent with its prediction by computational promoter detection.

mature form. These data strongly suggest that *Pyrobaculum* has some form of pre-tRNA 5′-processing activity.

***Pyrobaculum aerophilum* Cell Extract Processes 5′ Leader from pre-tRNA.** Because RNase P activity was previously not found in a *P. aerophilum* crude extract (12), we opted to assay this activity after chromatographic fractionation. When a cell-free extract of *P. aerophilum* was subjected to successive weak cation (CM)- and anion (DEAE)-exchange matrices, a peak of 5′-processing activity (Fig. S2) was observed with a *P. aerophilum* pre-tRNA^{Phe} (Fig. 1*A*) as the substrate. The products generated were identical in size to those obtained with in vitro reconstituted *Escherichia coli* RNase P (15) when pre-tRNA^{Phe} containing uridine at the −1 position (U_{-1}) was used (Fig. 1*B*); cleavage took place as expected between nucleotides −1 and +1 (Fig. S3 provides additional data characterizing cleavage-site selection). Another hallmark of RNase P-mediated processing is the presence of a 5′ phosphate in the mature tRNA product (15). TLC of mature tRNA^{Phe} generated by *P. aerophilum* RNase P and subsequently digested with RNase T2 showed the presence of a 5′ phosphate on G_{+1} (pGp in Fig. 1*C*), as observed with *E. coli* RNase P. Collectively, these results establish that the partially purified pre-tRNA 5′-processing activity in *P. aerophilum* is indeed RNase P.

Evidence for Three Out of Four Known Archaeal RNase P Proteins in *Pyrobaculum*. All known archaeal RNase P complexes require both RNA and protein components, so we tested *P. aerophilum* RNase P for these constituents. Treatment with proteinase K, which degrades proteins, eliminated the *P. aerophilum* RNase P activity (Fig. 2*A*), raising the question of the protein component identities. Of the four established archaeal RNase P proteins

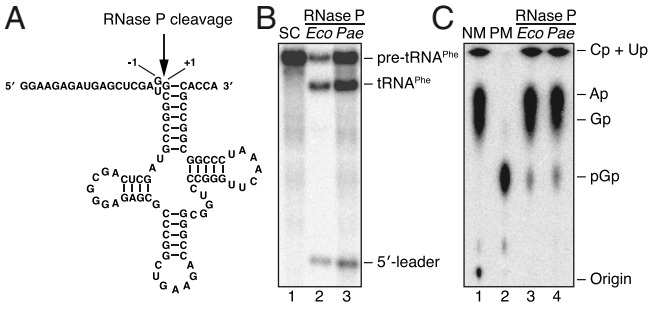

Fig. 1. The pre-tRNA 5′-processing activity from *P. aerophilum (Pae)* cell extract has all the cleavage properties of RNase P. (*A*) A *P. aerophilum* pre-tRNA^{Phe} with an 18-nt leader used for RNase P assays; pre-tRNA^{Phe} (G_{-1}) and pre-tRNA^{Phe} (U_{-1}) differ only in the identity of the -1 nucleotide. (*B*) *Pae* RNase P cleaves internally labeled pre-tRNA^{Phe} (U_{-1}) at the canonical cleavage site (lane 3), similar to *E. coli (Eco)* RNase P (lane 2; see also Figs. S2 and S3*A*). SC, a substrate control incubated without RNase P. (*C*) TLC analysis of tRNA^{Phe} containing a 5′ pGp, produced by *Eco* and *Pae* RNase P (lanes 3 and 4). NM and PM, a negative marker without and a positive marker with pGp.

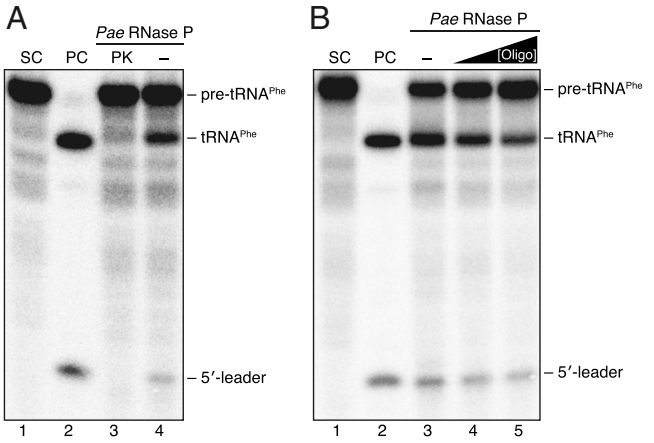

Fig. 2. RNase P activity from *P. aerophilum (Pae)* cell extract requires both protein and RNA subunits. (*A*) Treatment with proteinase K (PK) eliminated *Pae* RNase P activity. *Pae* RNase P was preincubated with PK (lane 3) or without PK (lane 4) before assaying for RNase P activity. (*B*) Preincubation of *Pae* RNase P with the oligo *Pae*RPR-L15 (complementary to nucleotides 157–170 in *Pae* RNase P RNA) resulted in decreased activity. *Pae* RNase P was preincubated without (lane 3) or with 50 and 150 μM *Pae*RPR-L15 (lanes 4 and 5) before assaying for RNase P activity. All assays were performed with internally labeled pre-tRNA^{Phe} (U_{-1}). SC, a substrate control incubated without RNase P; PC, a positive control with pre-tRNA^{Phe} processed by *E. coli* RNase P.

(Rpp29, Rpp30, Pop5, Rpp21), Pfam (16) and other sequence profile searches identified *Pyrobaculum* orthologs for Rpp29 (first noted by Hartmann and Hartmann; ref. 17), gave weak support for Rpp30 candidates, but had no predictions for Pop5 or Rpp21 orthologs. Koonin et al. (18) previously noted that the Rpp30 and Pop5 genes are sometimes adjacent to each other and are often in the same operon as the gene for a predicted exosome protein (DUF54/COG1325) (18). In *Pyrobaculum* species, this exosome gene clearly identifies a well-conserved operon that contains both the candidate Rpp30 genes and plausible Pop5 homologs (Table S1). Alignment of the *Pyrobaculum* candidates (PAE1830 for Rpp30; PAE1829 for Pop5) with the known archaeal protein homologs reveals potential homology: The *Pyrobaculum* candidates retain most of the highly conserved residues, but have lost some conserved segments, making the proteins 10–25% shorter than known archaeal family members. Alignment of the predicted Rpp29 ortholog (PAE1777) to archaeal Rpp29 proteins provides similar positive support. Further operon and sequence analyses, including a highly sensitive structure-based protein search, failed to identify the Rpp21 ortholog in any *Pyrobaculum* genome. These results suggest that only three of the four known archaeal RNase P proteins exist in *Pyrobaculum*.

Discovery and in Vitro Activity of the Minimized *Pyrobaculum* RNase P RNA. Given these apparent changes in the RNase P proteins, it was still unclear how the missing *Pyrobaculum* RNA might have changed to elude prior detection. Therefore, we made no assumptions about the RNA's primary sequence length or secondary structure features. Using the University of California Santa Cruz Archaeal Genome Browser (19) (archaea.ucsc.edu), we aligned the four new *Pyrobaculum* genomes with *P. aerophilum*, allowing us to focus on the most highly conserved regions common to all five species. After removing known RNA genes (rRNAs, tRNAs, C/D box sRNAs, signal-recognition particle RNA), only a few regions were conserved >90%, forming an enriched group to study. In one of these conserved regions, we detected a very weak partial hit to archaeal RNase P RNA using the Infernal RNA search program (20), but we also observed properties expected in good candidates: high (73–77%) guanine/cytosine content typical for a hyperthermophilic structural RNA, a

strong transcription factor B recognition element (BRE)-TATA promoter sequence upstream, and extremely high conservation (Fig. 3). Although the encoded RNA is much shorter than any identified nonorganellar RNase P RNAs (21), the *Pyrobaculum* candidates can be folded into an RNase P-like consensus secondary structure (Fig. 4*A* and Fig. S4) with one surprising difference. All known bacterial and archaeal RNase P RNAs consist of both a substrate-specificity (S) domain that aids substrate recognition, and a catalytic (C) domain essential for phosphodiester cleavage (22, 23). The *Pyrobaculum* RNase P RNA candidates have lost most of their S domain, but retain an intact C domain that includes all 11 universally conserved nucleotides (24) (Fig. 4*A* and Fig. S4).

To verify the candidate *P. aerophilum* RNase P RNA, we first confirmed expression by northern analysis of total RNA (Fig. 4*B*). Second, to assess association of this RNA with the partially purified, native *P. aerophilum* RNase P holoenzyme, we designed *Pae*RPR-L15, an antisense 14-mer RNA oligonucleotide which we expected would invade an essential loop region in the RNA moiety (nucleotides 157–170; Fig. 4*A*) and thereby interfere with activity. As shown for bacterial RNase P (25), we observed a progressive decrease in native *P. aerophilum* RNase P activity when the enzyme was preincubated with increasing concentrations of this oligonucleotide (Fig. 2*B*) and that the extent of this inhibition was more pronounced compared to that observed with a nonspecific oligonucleotide (see *SI Text*). We did not expect complete inhibition because *Pae*RPR-L15's access to the RNA moiety of *P. aerophilum* RNase P could be hampered by enveloping protein subunits. Such a premise is consistent with our findings that *P. aerophilum* RNase P could not be inactivated with micrococcal nuclease, for which there is a precedent in another archaeal hyperthermophile (26). Last, but most importantly, an in vitro transcript of the *P. aerophilum* candidate RNase P RNA was able to process pre-tRNA^{Phe} (Fig. 4*C* and Fig. S5), confirming it is indeed the missing RNase P ribozyme. As a negative control, deletion of the universally conserved, bulged U51 (Fig. 4*A*) in the same transcript eliminated the RNA-alone activity (Fig. 4*C*, lane 2); this bulged U in bacterial RNase P RNA appears to be required for maintaining a unique geometry essential for substrate positioning and binding of catalytically important Mg^{2+} ions (27). As with the partially purified native enzyme, preincubation of the in vitro RNase P RNA transcript with *Pae*RPR-L15 drastically decreased its activity as expected.

Phylogenetic Distribution of the Minimized Form of RNase P RNA. Equipped with a broader view of RNase P RNA, we surveyed all 88 currently available, complete archaeal genomes to determine which have traditional, *Pyrobaculum*-like, or possibly other

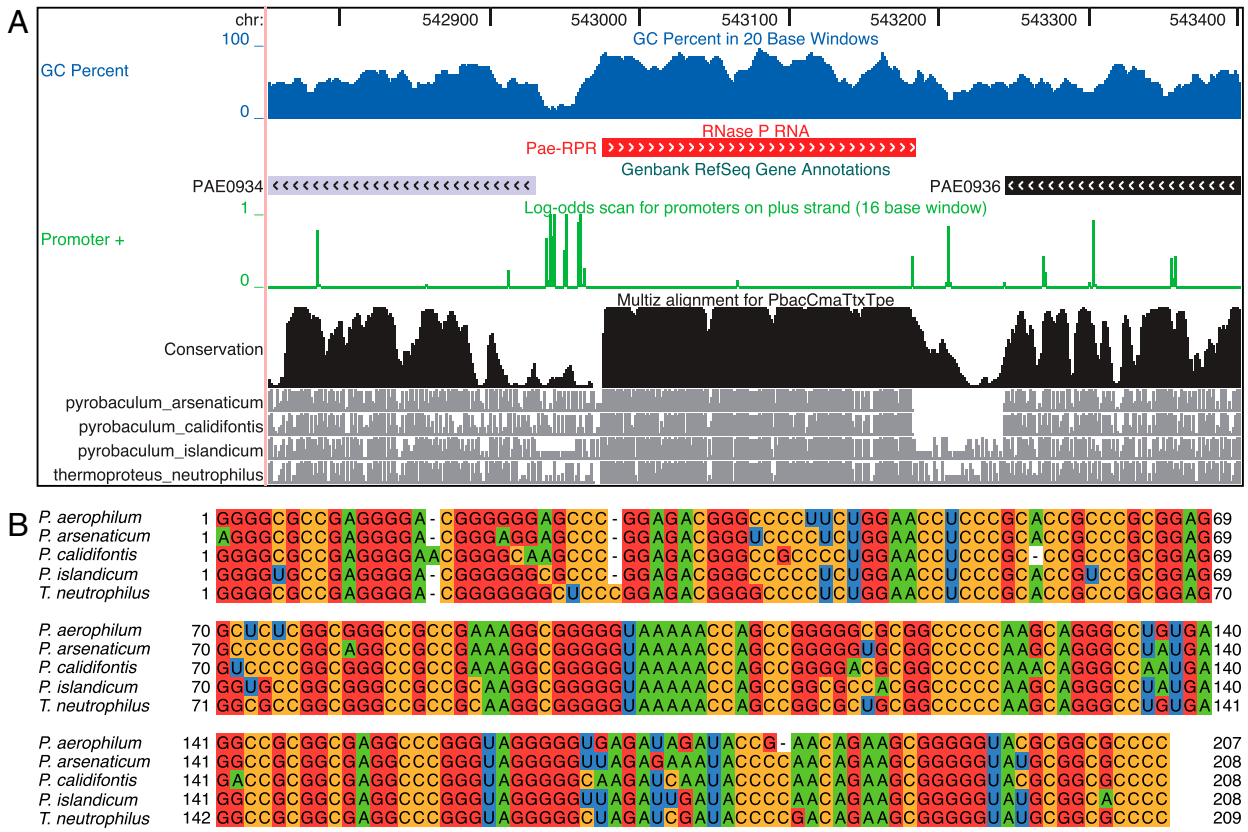

Fig. 3. *P. aerophilum* RNase P RNA displayed on the Archaeal Genome Browser (19) and alignment of the RNase P RNA sequences from five *Pyrobaculum* species. (*A*) The red segment located between protein-coding genes PAE0934 and PAE0936 corresponds to the *P. aerophilum* RNase P RNA. The arrows on the genes indicate the 5′-to-3′ expression direction. The blue track above the genes represents the guanine/cytosine (G/C) content computed with a 20-base sliding window. Compared to its neighboring protein-coding genes, *P. aerophilum* RNase P RNA has a high G/C content that is required for structural RNA stability in hyperthermophiles. Green lines in "Promoter +" indicate BRE-TATA promoter signals on the positive strand. A strong promoter signal just upstream of the *P. aerophilum* RNase P RNA locus is consistent with start of transcription precisely at the predicted RNase P RNA gene. The black "Conservation" track at the bottom displays a graph of the nucleotide conservation level of the alignment with multiple genome sequences (including *P. aerophilum* and the other four *Pyrobaculum* genomes). The most highly conserved region within the sequence window corresponds to *P. aerophilum* RNase P RNA, consistent with a structural noncoding RNA gene (as opposed to the weaker nucleotide conservation of the flanking protein-coding genes). (*B*) Multiple RNA sequence alignment shows high conservation and a few single-nucleotide insertions/deletions, which are common in noncoding RNA genes, but unlikely in protein-coding genes due to frameshifts.

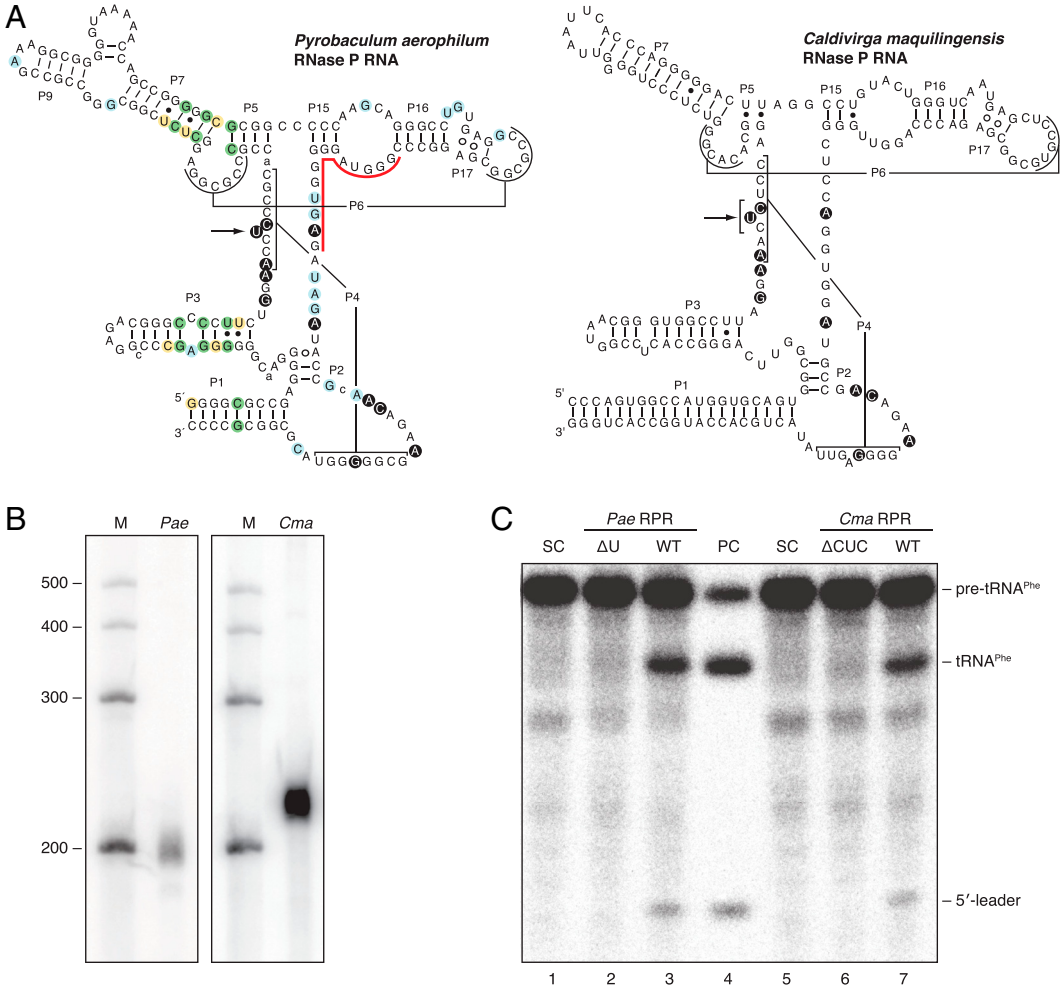

Fig. 4. Predicted secondary structure, native expression, and in vitro activity of *P. aerophilum (Pae)* and *Caldivirga maquilingensis (Cma)* RNase P RNAs (RPRs). (*A*) The *Pae* RPR structure shared by four other *Pyrobaculum* spp., with unlabeled nucleotides identical in all five species, and others highlighted as follows: universally conserved nucleotides (black circles), pairs showing covariation among different *Pyrobaculum* RPRs (green), conservative G-C to G-U changes (yellow), nonconserved insertions in some *Pyrobaculum* species (lowercase), differences in unpaired regions (blue), and complementary region of *Pae*RPR-L15 (red line). The RPR from *C. maquilingensis (Cma)* shows low sequence identity (<50%) but high secondary structure similarity. Arrow, deletion of this catalytically important bulge abolishes activity. (*B*) *Pae* and *Cma* RPRs are expressed in vivo, as shown by northern analysis of total RNA. M, size markers. (*C*) RNase P assay of in vitro transcribed, WT *Pae* (100 μM, lane 3) or *Cma* (40 μM, lane 7) RPRs with ~1 nM internally labeled pre-tRNA^{Phe} (G_{-1}). Deletion of the bulged U51 (Fig. 4*A*, *Left*, arrow) in the P4 helix rendered *Pae* RPR inactive (lane 2). Similarly, deletion of C62-U63-C64 (Fig. 4*A*, *Right*, arrow) in the P4 helix inactivated *Cma* RPR (lane 6). SC, substrate control, and PC, positive control, as in Fig. 2.

missing and/or undetectable versions. To search for traditional RNase P RNA genes, we used the Rfam archaeal RNase P covariance model (14) and the rule-based method by Li and Altman (13). Although slower, the covariance model was much more sensitive, finding matches in every species except *N. equitans, Pyrobaculum* spp., and unexpectedly, *Caldivirga maquilingensis* and *Vulcanisaeta distributa*, two crenarchaea in the same family (Thermoproteaceae) as *Pyrobaculum* (Table S2). By employing a covariance model based on the unusual *Pyrobaculum* RNAs, we uncovered a similarly shortened form of RNase P RNA in *C. maquilingensis* and *V. distributa* (Fig. 4*A*, Fig. S4, and Table S2), now accounting for an RNase P RNA gene in all complete archaeal genomes except *N. equitans* (12). We established expression of the *C. maquilingensis* candidate RNase P RNA in vivo (Fig. 4*B*) and that an in vitro transcript of this RNA can process pre-tRNA^{Phe} (Fig. 4*C* and Fig. S5). Like its *Pyrobaculum* counterpart, deletion of the bulged U63 together with C62 and C64 in the *C. maquilingensis* transcript eliminated the RNA-alone activity (Fig. 4*C*).

Both of these RNAs display all the hallmarks of other RNase P RNAs: They cleave at the expected site (Fig. S3*B*), they generate 5′-phosphate and 3′-hydroxyl termini (Fig. 1*C* and Fig. S5*C*), and they are capable of cleaving pre-tRNA^{Phe} with a short and long leader (Fig. S5*B*). Taken together, these data suggest that our *Pyrobaculum*-based covariance search model is complementary to the existing Rfam RNase P model and should facilitate detection of other shortened forms of RNase P (dubbed "type T" for the phylogenetic family Thermoproteaceae).

A search for the RNase P proteins in *Caldivirga* and *Vulcanisaeta* revealed likely homologs for Pop5, Rpp30, and Rpp29 (see *SI Text*), but, as in *Pyrobaculum*, no apparent ortholog of Rpp21. Because *Pyrobaculum*, *Caldivirga*, and *Vulcanisaeta* are closely related genera, it is most parsimonious to assume that the type T RNase P RNA was a feature of the common ancestor of this family. Outside of Thermoproteaceae, the next most closely related species with a sequenced genome is *Thermofilum pendens*, which seems to possess a typical RNase P RNA (14) with a traditional S domain and all four traditional archaeal RNase P proteins (Tables S1 and S2). Thus, the conspicuous absence of

Rpp21 in the Thermoproteaceae may reflect fewer cognate RNase P proteins for the smaller type T RNA, or that Rpp21 has diverged too extensively to be detected by sequence similarity searches. Reconstitution studies have revealed that the archaeal RNase P proteins function as two binary complexes: Pop5•Rpp30 and Rpp21•Rpp29 (23). Footprinting studies indicate that Pop5•Rpp30 interacts with the C domain, and Rpp21•Rpp29 interacts with the S domain of archaeal RNase P RNA (23, 28). Thus, the loss or radical change of Rpp21 would be consistent with loss of the traditional S domain in type T RNase P RNAs.

Search for RNase P RNA in *Aquifex* and Related Species. With the identification of the smaller RNase P RNA in *Pyrobaculum* and related archaeal species, we decided to apply existing and our newly developed covariance search models to four bacterial species in which RNase P RNA appears to be absent. These species include the well-studied *Aquifex aeolicus*, as well as three other species also in the Aquificaceae family (*Hydrogenivirga* sp. 128-5-R1-1, *Hydrogenobacter thermophilus*, and *Hydrogenobaculum* sp. Y04AAS1). We searched these genomes using all existing bacterial, archaeal, and eukaryotic RNase P RNA models in Rfam (14), as well as our *Pyrobaculum*-specific model, producing no good candidates. Because there are significant structural differences between bacterial and archaeal RNase P RNAs (4), we also created a model based on the RNase P RNAs from the closest related bacterial species that have known RNase P RNAs and artificially removed the specificity domain. The shortened bacterial search model produced no good candidates in any of the four genomes.

Conclusions

The reduced size of the type T form of RNase P RNA may correlate with unique features of this group of organisms. *Pyrobaculum* and *Vulcanisaeta* species are unusual for the large number of tRNAs with more than one intron, many at "noncanonical" positions (29, 30). *Caldivirga* does not have as many atypical tRNA introns, but does contain a number of *trans*-spliced split tRNAs (31), a rare trait shared only with *N. equitans* (32). It is unclear if changes in the pre-tRNA structures are related to the altered substrate recognition domain, but a link would not be surprising.

Our discovery of the drastically smaller type T RNase P emphasizes the striking and surprising plasticity in the subunit composition of an essential and ubiquitous enzyme. This flexibility may be an adaptive trait related to the variability in the RNAs that are substrates for RNase P. In bacteria, these assorted substrates include 4.5S, tmRNA, viral RNAs, riboswitches, and some mRNAs (2). In Thermoproteaceae, type T RNase P may play a similarly expanded role in small RNA processing.

Collectively, our findings highlight the rich evolutionary story of RNase P and offer opportunities for structural, comparative, and functional studies. As the cost of genome and transcriptome sequencing continues to decrease, we expect the combination of comparative genomics and improved RNA search models will continue to reveal exceptional cases advancing RNA biology.

Materials and Methods

Archaeal RNase P RNA Sequence Search. Infernal v1.0 (20) was used to search for RNase P RNA candidates in archaeal genomes. The program was initially run in the global search mode using the Rfam (14) archaeal RNase P RNA covariance model (RF00373). All hits with a score >0 bits were manually examined. Local search mode was also employed, which provided better sensitivity but decreased selectivity. Using local search mode with a threshold of 0 bits (necessary to initially detect the *Pyrobaculum* RNase P RNA with the RF00373 model) was only feasible when searching a small set of short candidate regions (i.e., not entire genomes) because low-scoring partial hits required close manual inspection and further structural analyses.

The rule-based program by Li and Altman (13) was used to search for RNase P RNA candidates in all available archaeal genomes retrieved from ftp://ftp.ncbi.nih.gov/genomes/Bacteria/. The genome of *Vulcanisaeta distributa* was downloaded from the publicly accessible Integrated Microbial Genomes with the Genomic Encyclopedia of Bacteria and Archaea Genomes Web site (http://img.jgi.doe.gov/cgi-bin/geba/main.cgi?section=TaxonDetail&taxon_oid=2502790013). Because hits were not found in a number of genomes, Infernal (global search mode) was applied using the Rfam RF00373 model. A universal primary sequence pattern for RNase P RNA, modified from the original Li and Altman pattern, was used to prescreen candidates that were then analyzed with Infernal. For archaeal genomes producing no hits using this fast prescreening method, whole genome searches were conducted with Infernal alone (greatly slowing search speed, but improving search sensitivity). The *Pyrobaculum* RNase P RNA covariance model was built with Infernal using the five *Pyrobaculum* RNA sequences and a manually predicted secondary structure (Fig. 4*A*).

RNase P Purification. *P. aerophilum* cells were grown as described in *SI Materials and Methods*. One gram of cells was resuspended in 5 mL cold extraction buffer (EB; 20 mM Tris•HCl, pH 8 at 25 °C and 5 mM $MgCl_2$) containing 50 mM NaCl, 10 mM DTT, and 1 mM PMSF. Cell-free extract was generated by sonication and clarified by centrifugation at 14,000 × *g* and 4 °C for 30 min. The crude extract was then loaded on a 1-mL HiTrap CM FF Sepharose column (GE Healthcare). EB supplemented with various NaCl concentrations, 10% (vol/vol) glycerol, 2 mM DTT, and 0.2 mM PMSF was used in all the following steps. With an FPLC apparatus, fractions were eluted using a 50–2000 mM NaCl gradient and subsequently assayed for RNase P activity (33). The active CM fractions were pooled, dialyzed, and loaded on a 1-mL HiTrap DEAE FF Sepharose column (GE Healthcare). The activity was eluted with three sequential NaCl gradients: 50–400, 400–1,000, and 1,000–1,250 mM. The peak of RNase P activity eluted at ~700 mM NaCl.

RNase P Assay. The *P. aerophilum* pre-tRNA^{Phe} was in vitro transcribed and radioactively labeled either internally using [α-^{32}P]-GTP during transcription or at the 5′ end using T4 polynucleotide kinase (New England Biolabs, NEB) and [γ-^{32}P]-ATP after dephosphorylation of the transcript by calf intestinal alkaline phosphatase (NEB). RNase P activity was assayed at 55 °C in 50 mM Tris•HCl (pH 7.5 at 25 °C), 400 mM NH_4OAc, 10 mM $Mg(OAc)_2$, 5 mM DTT, 0.1 unit/μL RiboLock (Fermentas), 100 nM unlabeled pre-tRNA^{Phe}, and trace amounts of labeled pre-tRNA^{Phe} (~1 nM). Activity of the in vitro transcribed *P. aerophilum* and *C. maquilingensis* RNase P RNA was assayed, without a prefolding incubation, at 55 °C for 20 h in 50 mM Tris•HCl (pH 7.5 at 25 °C), 1.5 M NH_4OAc, trace amounts (~1 nM) of labeled pre-tRNA^{Phe}, and $Mg(OAc)_2$ at 75 and 100 mM, respectively. The cleavage products were then separated on 12% (wt/vol) polyacrylamide/7 M urea gels (or 15% for the 4-nt leadered pre-tRNA^{Phe}) and radiographed on a PhosphorImager.

TLC. Subsequent to cleavage of internally labeled pre-tRNA^{Phe} by *P. aerophilum* and *E. coli* RNase P or *P. aerophilum*, *C. maquilingensis* and *E. coli* RNase P RNA, the mature tRNA^{Phe} products were purified using an 8% (wt/vol) polyacrylamide/7 M urea gel. Uncleaved internally labeled pre-tRNA^{Phe} was used as a negative marker and uncleaved 5′ end-labeled pre-tRNA^{Phe} (with the 5′ monophosphorylated G being the only labeled residue) as the positive marker of pGp. These RNAs were then completely digested with 12 units of RNase T2 (Gibco BRL) at 50 °C for 1 h in 20 mM NaOAc (pH 5.2 at 25 °C), 1 mM EDTA, and 0.4 μg/μL yeast tRNAs. Each sample was spotted on a TLC PEI cellulose F plate (EMD Chemicals), separated with 0.5 M potassium phosphate (pH 6.3):methanol in an 80:20 ratio, and radiographed with a PhosphorImager.

ACKNOWLEDGMENTS. We thank J. Murphy [University of California, Santa Cruz (UCSC)] for her contribution of *Pyrobaculum* cell material, L. Lui (UCSC) for her contribution of *Caldivirga* cell material for northern analysis, and K. Karplus (UCSC) for his Rpp21 structural protein search and expert advice on protein alignment interpretation. We are grateful to members of the Joint Genome Institute for making 454 sequencing possible (P. Richardson and J. Bristow for providing resources, and E. Lindquist and N. Zvenigorodsky for sample preparation and analysis). We thank J. Jackman [Ohio State University (OSU)] for advice and reagents for the TLC analysis, and E. J. Behrman (OSU) for helpful discussions. We are indebted to S. Eddy (Howard Hughes Medical Institute Janelia Farm) for insightful comments that helped improve the manuscript. This work was supported by National Science Foundation Grants MCB0238233 and MCB0843543 (to V.G.) and EF-0827055 (to T.L.), and National Institutes of Health GM067807 (to Mark P. Foster and V.G.) and HG004002-02 subaward (to T.L.).

1. Liu F, Altman S (2010) *Ribonuclease P* (Springer, New York).
2. Kazantsev AV, Pace NR (2006) Bacterial RNase P: A new view of an ancient enzyme. *Nat Rev Microbiol* 4:729–740.
3. Coughlin DJ, Pleiss JA, Walker SC, Whitworth GB, Engelke DR (2008) Genome-wide search for yeast RNase P substrates reveals role in maturation of intron-encoded box C/D small nucleolar RNAs. *Proc Natl Acad Sci USA* 105:12218–12223.
4. Hall TA, Brown JW (2001) The ribonuclease P family. *Methods Enzymol* 341:56–77.
5. Lai LB, Vioque A, Kirsebom LA, Gopalan V (2010) Unexpected diversity of RNase P, an ancient tRNA processing enzyme: Challenges and prospects. *FEBS Lett* 584:287–296.
6. Holzmann J, et al. (2008) RNase P without RNA: Identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. *Cell* 135:462–474.
7. Gobert A, et al. (2010) A single *Arabidopsis* organellar protein has RNase P activity. *Nat Struct Mol Biol* 17:740–744.
8. Wang G, et al. (2010) PNPASE regulates RNA import into mitochondria. *Cell* 142:456–467.
9. Marszalkowski M, Willkomm DK, Hartmann RK (2008) 5'-end maturation of tRNA in *Aquifex aeolicus*. *Biol Chem* 389:395–403.
10. Lombo TB, Kaberdin VR (2008) RNA processing in *Aquifex* aeolicus involves RNase E/G and an RNase P-like activity. *Biochem Biophys Res Commun* 366:457–463.
11. Willkomm DK, Minnerup J, Huttenhofer A, Hartmann RK (2005) Experimental RNomics in *Aquifex* aeolicus: Identification of small non-coding RNAs and the putative 6S RNA homolog. *Nucleic Acids Res* 33:1949–1960.
12. Randau L, Schroder I, Soll D (2008) Life without RNase P. *Nature* 453:120–123.
13. Li Y, Altman S (2004) In search of RNase P RNA from microbial genomes. *RNA* 10:1533–1540.
14. Gardner PP, et al. (2009) Rfam: updates to the RNA families database. *Nucleic Acids Res* 37:D136–140.
15. 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.
16. Finn RD, et al. (2008) The Pfam protein families database. *Nucleic Acids Res* 36: D281–288.
17. Hartmann E, Hartmann RK (2003) The enigma of ribonuclease P evolution. *Trends Genet* 19:561–569.
18. Koonin EV, Wolf YI, Aravind L (2001) Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach. *Genome Res* 11:240–252.
19. Schneider KL, Pollard KS, Baertsch R, Pohl A, Lowe TM (2006) The UCSC archaeal genome browser. *Nucleic Acids Res* 34:D407–410.
20. Nawrocki EP, Kolbe DL, Eddy SR (2009) Infernal 1.0: Inference of RNA alignments. *Bioinformatics* 25:1335–1337.
21. Brown JW (1999) The Ribonuclease P Database. *Nucleic Acids Res* 27:314.
22. Loria A, Pan T (1996) Domain structure of the ribozyme from eubacterial ribonuclease P. *RNA* 2:551–563.
23. Tsai HY, Pulukkunat DK, Woznick WK, Gopalan V (2006) Functional reconstitution and characterization of *Pyrococcus furiosus* RNase P. *Proc Natl Acad Sci USA* 103:16147–16152.
24. Marquez SM, et al. (2005) Structural implications of novel diversity in eucaryal RNase P RNA. *RNA* 11:739–751.
25. Gruegelsiepe H, Willkomm DK, Goudinakis O, Hartmann RK (2003) Antisense inhibition of *Escherichia coli* RNase P RNA: Mechanistic aspects. *ChemBioChem* 4:1049–1056.
26. Darr SC, Pace B, Pace NR (1990) Characterization of ribonuclease P from the archaebacterium *Sulfolobus solfataricus*. *J Biol Chem* 265:12927–12932.
27. Kaye NM, Zahler NH, Christian EL, Harris ME (2002) Conservation of helical structure contributes to functional metal ion interactions in the catalytic domain of ribonuclease P RNA. *J Mol Biol* 324:429–442.
28. Xu Y, Amero CD, Pulukkunat DK, Gopalan V, Foster MP (2009) Solution structure of an archaeal RNase P binary protein complex: Formation of the 30-kDa complex between *Pyrococcus furiosus* RPP21 and RPP29 is accompanied by coupled protein folding and highlights critical features for protein-protein and protein-RNA interactions. *J Mol Biol* 393:1043–1055.
29. Chan PP, Lowe TM (2009) GtRNAdb: A database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res* 37:D93–97.
30. Marck C, Grosjean H (2003) Identification of BHB splicing motifs in intron-containing tRNAs from 18 archaea: Evolutionary implications. *RNA* 9:1516–1531.
31. Fujishima K, et al. (2009) Tri-split tRNA is a transfer RNA made from 3 transcripts that provides insight into the evolution of fragmented tRNAs in archaea. *Proc Natl Acad Sci USA* 106:2683–2687.
32. Randau L, Munch R, Hohn MJ, Jahn D, Soll D (2005) *Nanoarchaeum equitans* creates functional tRNAs from separate genes for their 5'- and 3'-halves. *Nature* 433:537–541.
33. Vioque A, Arnez J, Altman S (1988) Protein-RNA interactions in the RNase P holoenzyme from *Escherichia coli*. *J Mol Biol* 202:835–848.