Supporting Information

Lai et al. 10.1073/pnas.1013969107

SI Materials and Methods

Culture Conditions for *Pyrobaculum aerophilum*. *P. aerophilum* cultures derived from Deutsche SammLung von Mikroorganismen und Zellkulturen (DSMZ) strain DSM7523 were generously provided by Christopher House (Penn State University, University Park, PA). *P. aerophilum* cultures used for RNase P purification were grown in a slightly modified version of DSMZ DSM390 medium (1), amended with 1% tryptone and 0.1% yeast extract to produce higher cell yields. These cultures were grown microaerobically at 95 °C under a gas headspace of N_2 plus 2% or 3% O_2 until early log phase, then shifted to fully aerobic growth under atmospheric air and collected at mid- to late-log phase.

Culture Conditions for *Caldivirga maquilingensis*. Cultures derived from DSMZ 13496/IC-167 were provided by Christopher House (Penn State University, University Park, PA). *C. maquilingensis* cultures were grown in modified DSM883 media containing (per liter) 2.94 g Na_3-citrate•$2H_2O$, 0.5 mg resazurin, 0.5 g yeast extract, 10 g tryptone, 0.02 g $FeCl_3$•$6H_2O$, 1 mL of a vitamin solution (DSM141), 100 mL of a 10× salts solution, 10 mL of a 100× trace elements solution, and 2 mL of a polysulfide solution. The vitamin solution was filter-sterilized and stored at 4 °C. The 10× salts solution derived from DSM88 was prepared by combining (per liter) 13 g $(NH_4)_2SO_4$, 2.8 g KH_2PO_4, 2.5 g $MgSO_4$•$7H_2O$, and 0.7 g $CaCl_2$•$2H_2O$. The 100× trace elements solution derived from DSM88 was prepared by combining (per liter) 0.18 g $MnCl_2$•$4H_2O$, 0.45 g $Na_2B_4O_7$•$10H_2O$, 0.022 g $ZnSO_4$•$7H_2O$, 0.005 g $CuCl_2$•$2H_2O$, 0.003 g Na_2MoO_4•$2H_2O$, 0.003 g $VOSO_4$•$2H_2O$, and 0.001 g $CoSO_4$. The polysulfide solution was prepared using 15 g Na_2S•$9H_2O$, 3 g elemental sulfur in 100 mL anoxic water under N_2. The complete medium preparation was adjusted to pH 4 using H_2SO_4 for creating colloidal sulfur and sterilized by autoclaving. All *C. maquilingensis* cultures were incubated anaerobically at 85 °C, typically with 200 mL of medium in sealed 0.5 L Wheaton bottles under a gas headspace of N_2. Cells were harvested at midlog phase by centrifugation at 10,000 × *g* for 10 min. Cell pellets were frozen in liquid N_2 and stored at −80 °C.

Total RNA Preparation. Total RNA was extracted from the frozen cell pellets using a Polytron tissue homogenizer and TRI Reagent (Sigma-Aldrich). RNA samples were treated with TURBO DNase (Ambion) to remove any residual DNA, reextracted with TRI Reagent, and normalized to 1 μg/μL.

High-Throughput RNA Sequencing. Total RNA from *P. aerophilum*, *Pyrobaculum arsenaticum, Pyrobaculum calidifontis*, and *Pyrobaculum islandicum* (100 μg each) was denatured and resolved by electrophoresis in separate lanes on a 15% (wt/vol) polyacrylamide-urea gel. RNA with a size of 15–70 nt (migrating just ahead of the tRNA band, down to and including 75% of the region between the xylene cyanol and bromophenol blue loading dye bands) was excised from the gel. Samples were eluted and precipitated with ethanol. A 3′ linker with 5′ adenylation and a 3′-terminal dideoxy-C base [Linker-1 from Integrated DNA Technologies (IDT)] was ligated to the RNA as described by Lau et al. (2). A second gel purification was performed to remove excess 3′ linker by extracting the ligated RNA which migrated above the xylene cyanol dye band. Resulting linked molecules were reverse transcribed using Superscript III (Invitrogen) and a complementary DNA primer. Exonuclease I (Thermo) followed by treatment with EDTA and sodium hydroxide were used to degrade excess primer, inactivate reverse transcriptase, and selectively hydrolyze RNA, respectively. The resulting cDNA was purified using a NucAway spin column (Ambion) before the addition of a 5′-adenylated linker (Linker-2 from IDT) using T4 RNA ligase (Ambion). cDNA was amplified by a 16-cycle PCR which was followed by a second 16-cycle PCR reamplification using Roche/454 specific hybrid primers based on the method described by Hannon (3). A four-base barcode, as described by Ambros (4), was included in the 5′ hybrid primer. The final reaction was purified using the DNA Clean and Concentration kit (Zymo Research). Samples were then sent to the Joint Genome Institute, where they were pooled in equal quantities, amplified using the manufacturer's protocol, and analyzed on a Roche GS-FLX sequencer.

Cloning and Transcription of *P. aerophilum* and *C. maquilingensis* RNase P RNAs, and *P. aerophilum* pre-tRNA^{Phe}. PCR was used to amplify the coding sequence of *P. aerophilum (Pae)* RNase P RNA using the *P. aerophilum* genomic DNA as the template, and the primers *Pae*RPR-F (5′-GGCGCCGAGGGGACG-3′) and *Pae*RPR-R (5′-GGGGCGCCGCGTACC-3′).

Cloning of *C. maquilingensis (Cma)* RNase P RNA was also performed with PCR amplification. Because the sequence of the P1 helix in *C. maquilingensis* RNase P RNA is long and completely complementary (Fig. 4*A*), only one primer (*Cma*RPR-FR, 5′-CCCAGTGGCCATGGTGC-3′) was used to amplify the sequence from a BAC clone harboring the *C. maquilingensis* RNase P RNA gene.

The *P. aerophilum* pre-tRNA^{Phe} with an 18-nt leader was amplified from *P. aerophilum* genomic DNA using the primers *Pae*Phe-F (5′-AAGAGATGAGCTCGAKGCGGCCGTAGCTCAGC-3′) and *Pae*Phe-R (5′-GC*GAATT***CCTGG**TGCGGCCGCC-3′); the degenerate nucleotide K (G or T) allowed for cloning both the pre-tRNA^{Phe} (G_{-1}) and (U_{-1}), discriminated by the presence and absence of an XhoI site, respectively. The 18-nt leader facilitated electrophoretic separation of pre-tRNA^{Phe} and its cleavage products after the RNase P assay. For cloning the 4-nt-leadered pre-tRNA^{Phe}, *Pae*Phe(4)-F (5′-AGGCGGCCGTAGCTCAGC-3′) was used with *Pae*Phe-R. The underlined sequences in *Pae*Phe-F and *Pae*Phe(4)-F correspond to the 18-nt and 4-nt 5′ leaders, respectively. Note that the 5′-GG of each leader sequence was encoded by the host vector on account of the cloning strategy employed.

Subsequent to amplification, all PCR products were cloned downstream of the T7 promoter in pBT7 (5). Both the *P. aerophilum* and the *C. maquilingensis* RNase P RNA fragments were ligated blunt-ended into the StuI site, and the pre-tRNA^{Phe} fragments were digested with EcoRI (the italicized sequence introduced in *Pae*Phe-R) before ligating into the StuI and EcoRI sites. Deletion mutants of *P. aerophilum* and *C. maquilingensis* RNase P RNA were generated by PCR using these clones as the respective templates; pairs of primers were designed to flank the deleted nucleotide(s) and orient outward such that the complete plasmids would be amplified and then circularized. All clones were confirmed by DNA sequencing.

Transcripts were generated from runoff in vitro transcription with T7 RNA polymerase using these clones as the templates after linearizing pBT7-*Pae*RPR (RNase P RNA) and pBT7-*Cma*RPR with BstBI (present in the vector), and pBT7-pre-tRNA^{Phe} with BstNI (depicted in bold in *Pae*Phe-R).

Inactivation of *P. aerophilum* RNase P Activity by Treatment with Proteinase K. An aliquot of partially purified *P. aerophilum* RNase P

(DEAE fraction 9 in Fig. S2*B*) was first incubated with proteinase K (8 μg/μL; Roche) in 50 mM Tris•HCl (pH 7.9 at 25 °C) and 5 mM $CaCl_2$ for 30 min at 55 °C, before incubating at 55 °C for 20 h to assay for pre-tRNA^{Phe}-cleavage activity.

Inactivation of *P. aerophilum* RNase P Activity Using an Antisense RNA Oligonucleotide Complementary to the *P. aerophilum* RNase P RNA. Aliquots of partially purified *P. aerophilum (Pae)* RNase P (DEAE fraction 9 in Fig. S2*B*) were preincubated at 55 °C for 30 min in the presence of all components needed for RNase P assay, except for pre-tRNA^{Phe} (U_{-1}). Where indicated, *Pae*RPR-L15, an antisense RNA oligonucleotide (Sigma Genosys; 5′-CUUGCCCCCUACCC-3′), designed to invade L15 and the downstream region (nucleotides 157–170 in the *Pae* RNase P RNA) was also present in the preincubation at either 50 or 150 mM (Fig. 2*B*). Subsequently, assay for RNase P activity was initiated with addition of pre-tRNA^{Phe} (U_{-1}) and incubated at 55 °C for 20 h. In independent experiments undertaken to assess the specificity of *Pae*RPRL-15, we employed as a negative control the RNA oligonucleotide pGln-16 (IDT, Inc., 5′-UGGG-GUGUAGCCAAGC-3′), which has a similar GC content as *Pae*RPRL-15. In contrast to *Pae*RPRL-15, at concentrations below 120 μM, pGln-16 fails to inhibit *Pae* RNase P (in fact, was weakly stimulatory); even at 180 μM where pGln-16 does inhibit, it is at least twofold lower in potency relative to *Pae*RPR-L15. This latter finding is consistent with high concentrations of GC-rich oligonucleotides exhibiting nonspecific binding to GC-rich RNAs, such as the *Pae* RNase P RNA. Nevertheless, the inhibition of *Pae* RNase P activity observed at lower concentrations of *Pae*RPR-L15 lends support for its specificity.

Northern Analysis. Three microgram samples of total RNA extracted from *P. aerophilum* and *C. maquilingensis* cell cultures were combined with two parts (vol/vol) load buffer (95% formamide, 18 mM EDTA, and 0.025% each of SDS, xylene cyanol, and bromophenol blue), and denatured for 5 min at 70 °C. Denatured RNA samples were resolved in a 6% (vol/vol) acrylamide, 8 M urea gel, and transferred to Hybond N^+ membranes (GE Healthcare). Probes for detection of the *P. aerophilum* and *C. maquilingensis* RNase P RNAs were prepared by PCR of genomic DNA, followed by asymmetric amplification of the PCR product in the presence of [α-^{32}P]-CTP using only the antisense (reverse) primers to generate strand-specific radiolabeled probes. To prepare template for the *P. aerophilum* probe, a 200-bp segment of the *P. aerophilum* RNase P RNA gene (corresponding to nucleotides 10-199; Fig. 4*A*) was amplified from *P. aerophilum* genomic DNA using the forward primer (5′-GGCCCCTTCTGGAA-CCTC-3′) and the reverse primer (5′-CCTCACAGGCCC-TGCTTG-3′). To prepare template for the *C. maquilingensis* probe, a 150-bp segment of the *C. maquilingensis* RNase P RNA gene (corresponding to nucleotides 48–197; Fig. 4*A*) was amplified from *C. maquilingensis* genomic DNA using the forward primer (5′-GGCCCCTTCTGGAACCTC-3′) and the reverse primer (5′-CCTCACAGGCCCTGCTTG-3′). Hybridizations were performed at 42 °C using UltraHyb buffer (Ambion).

tRNA Gene and Promoter Identification. tRNAs were predicted using tRNAscan-SE (6). Manual inspection and adjustments were made due to difficulty in identifying tRNAs with introns placed outside of the canonical position (between nucleotides 37 and 38). Annotations and sequences were deposited into the Genomic tRNA Database (7) (http://gtrnadb.ucsc.edu).

To generate a training set for promoter identification, potential operons were predicted genome-wide with the requirement of a minimum intergenic separation of at least 100 bp (on the same strand). A 16-mer motif search of the 90 bp upstream of known genes (not annotated as putative or hypothetical genes) using MEME (8) was conducted to identify the consensus promoter, including the transcription factor B response element (transcription factor B recognition element—1 to 3 adenosine nucleotides) plus the TATA box. A position-specific scoring matrix (PSSM) was generated from the alignments of the MEME results after manual inspection. Each organism's PSSM was used to scan the 150-bp upstream region of all noncoding and protein-coding genes to identify potential promoter regions. Ten virtual genomes for each target genome were generated with the use of a fifth-order Markov chain to retain the base frequency of the target genome. The PSSM was applied to the same features in the target genome for null hypothesis formation. The promoter candidates previously identified were filtered according to expected position (9) and a threshold p value equivalent to that of the lowest-scoring known gene was established.

RNase P RNA Sequence Search in Aquificaceae. Infernal v1.0 (10) was used to search for RNase P RNA candidates in *Aquifex aeolicus*, *Hydrogenivirga* sp. 128-5-R1-1, *Hydrogenobacter thermophilus*, and *Hydrogenobaculum* sp. Y04AAS1. The program was run in both global and local search modes using the Rfam (11) bacterial (RF00010 and RF00011), archaeal (RF00373), and eukaryotic (RF00009) RNase P RNA covariance models. All hits with a score >0 bits were manually examined. A covariance model built with *Pyrobaculum* RNase P RNA sequences was also applied in global and local search modes. Finally, a shortened bacterial covariance model was developed to search more sensitively against the target Aquificaceae genomes by aligning the RNase P RNA sequences of *Persephonella marina*, *Sulfurihydrogenibium* sp. YO3AOP1, and *Sulfurihydrogenibium azorense* against the Rfam (11) bacterial type A RNase P RNA covariance model (RF00010) using Infernal (10) and removing the specificity domain from the alignments manually.

RNase P Protein Database Searches and Sequence Alignments. Protein sequences of Pop5, Rpp30, Rpp29, and Rpp21 for all archaeal genomes were retrieved from GenBank. PSI-BLAST (12), Pfam (13) domain searches (RNase_P_Rpp14 [Pop5]: PF01900; RNase_P_p30 [Rpp30]: PF01876; UPF0086 [Rpp29]: PF01868; and Rpr2 [Rpp21]: PF04032), and phylogenetic hidden Markov model-guided (14) multiple alignments in the Archaeal Genome Browser (15) were used to predict homology. Default scoring thresholds for PSI-BLAST (E-value, 10; word size, 3) and Pfam (trusted cutoff for Pop5, 23.4 bits; Rpp30, 20.3 bits; Rpp29, 21.1 bits; Rpp21, 23.2 bits) searches were initially adopted. Thresholds were further adjusted (E-value, 100 and word size, 2 for PSI-BLAST; trusted cutoff as −80 bits for Pfam) to search for proteins not identified with the default scan. Multiple sequence alignments across all archaeal genomes for each RNase P protein were generated using MUSCLE v3.7 (16) with default options. Alignments were visualized with the ClustalX color scheme within Jalview v2.4 (17), a multiple alignment editor freely available at http://www.jalview.org/. The RNase P protein sequence alignments in FASTA format are provided in a separate supplementary dataset file.

1. Cozen AE, et al. (2009) Transcriptional map of respiratory versatility in the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. *J Bacteriol* 191: 782–794.
2. Lau NC, Lim LP, Weinstein EG, Bartel DP (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294:858–862.
3. Hannon G. *Cloning Small RNAs for Sequencing with 454 Technology,* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) (http://www.454.com/downloads/hannon_smallRNA-cloning_protocol2.pdf).
4. Ambros V. *MicroRNA Cloning Protocol* (http://146.189.76.171/lab/MicroRNAs/Ambros_microRNAcloning.htm).

5. Tsai HY, Lai LB, Gopalan V (2002) A modified pBluescript-based vector for facile cloning and transcription of RNAs. *Anal Biochem* 303:214–217.
6. Lowe TM, Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964.
7. Chan PP, Lowe TM (2009) GtRNAdb: A database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res* 37:D93–97.
8. Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2:28–36.
9. Slupska MM, et al. (2001) Leaderless transcripts of the crenarchaeal hyperthermophile *Pyrobaculum aerophilum*. *J Mol Biol* 309:347–360.
10. Nawrocki EP, Kolbe DL, Eddy SR (2009) Infernal 1.0: Inference of RNA alignments. *Bioinformatics* 25:1335–1337.
11. Gardner PP, et al. (2009) Rfam: Updates to the RNA families database. *Nucleic Acids Res* 37:D136–140.
12. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
13. Finn RD, et al. (2010) The Pfam protein families database. *Nucleic Acids Res* 38:D211–222.
14. Siepel A, Haussler D (2004) Combining phylogenetic and hidden Markov models in biosequence analysis. *J Comput Biol* 11:413–428.
15. Schneider KL, Pollard KS, Baertsch R, Pohl A, Lowe TM (2006) The UCSC archaeal genome browser. *Nucleic Acids Res* 34:D407–410.
16. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.
17. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189–1191.

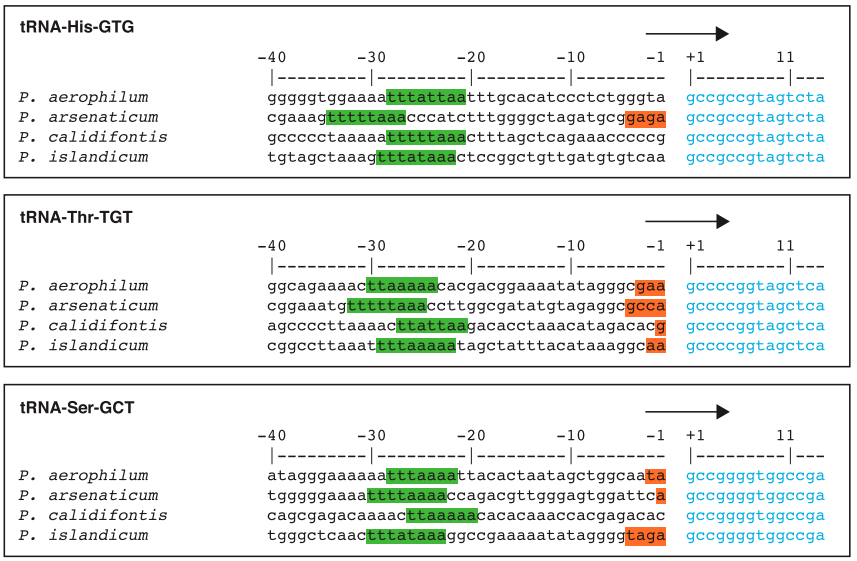

Fig. S1. Alignment of tRNA promoters and 5′-leader sequences across four *Pyrobaculum* species for three sets of tRNA orthologs. The predicted TATA box of each tRNA gene is highlighted in green, mature tRNA-encoding sequence in blue, and the 5′-leader sequence that is supported by RNA sequencing in orange. Scales above sequences are positions relative to the 5′ end of mature tRNAs. The black arrows indicate the direction of transcription.

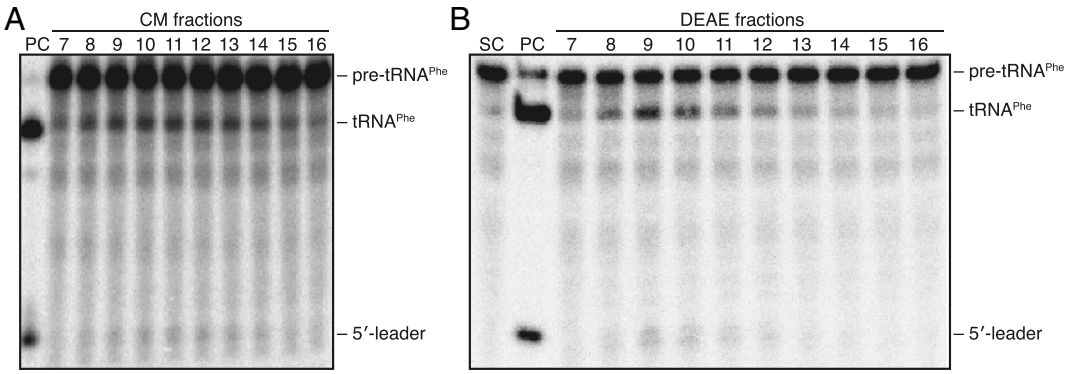

Fig. S2. Partial purification of *P. aerophilum* RNase P by ion-exchange chromatography. (*A*) *P. aerophilum* RNase P was first purified from crude extract using CM-Sepharose, and the peak of activity (~900 mM NaCl) was then dialyzed to remove NaCl before loading on to the DEAE column. (*B*) Peak of RNase P activity from DEAE-Sepharose. Fractions were eluted with three sequential NaCl gradients: 50–400 mM (fractions 1–6), 400–1,000 mM (fractions 7–12), and 1,000–1,250 mM (fractions 13–16). The RNase P activity in eluted fractions was assayed with internally labeled pre-tRNA^{Phe} (G_{-1}). Fraction 9 eluted from DEAE-Sepharose above was used in all subsequent characterization assays. PC, the positive control obtained from processing of pre-tRNA^{Phe} by *Escherichia coli* RNase P; SC, the substrate control incubated without RNase P.

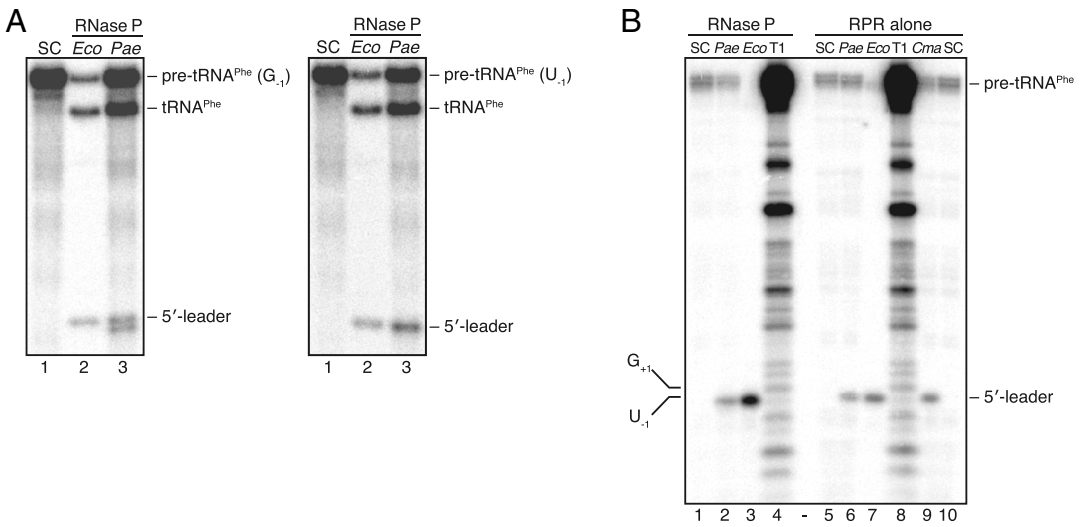

Fig. S3. Analysis of the site of cleavage in pre-tRNA^{Phe} (G_{-1}) and (U_{-1}) by partially purified native *P. aerophilum (Pae)* RNase P and in vitro transcribed RNase P RNAs (RPRs). (*A*, *Left*) Internally labeled *P. aerophilum* pre-tRNA^{Phe} (G_{-1}) with an 18-nt leader was used for these assays. *Pae* RNase P cleaves with equal frequency at the canonical cleavage site and one nucleotide upstream (lane 3) unlike E. coli (*Eco*) RNase P (lane 2) which does not miscleave. (*A*, *Right*) Fig. 1*B* is provided here for comparison to contrast the correct cleavage of pre-tRNA^{Phe} (U_{-1}) by partially purified *Pae* RNase P. SC, a substrate control incubated without RNase P. (*B*) Both partially purified *Pae* and *E. coli* (*Eco*) RNase P cleave the 5′ end-labeled pre-tRNA^{Phe} (U_{-1}) between U_{-1} and G_{+1} as expected for bona fide RNase P (lanes 2 and 3). Similarly, in vitro transcribed *Pae* and *C. maquilingensis* (*Cma*) RPRs exhibit correct processing of the substrate (lanes 6 and 9). T1, a G ladder generated by partial digestion of denatured 5′ end-labeled pre-tRNA^{Phe} (U_{-1}) with RNase T1 (Ambion) that cleaves 3′ to the G residues; subsequent to RNase T1 digestion, the cleaved RNA fragments were treated with T4 polynucleotide kinase to remove their 3′ phosphate (lanes 4 and 8) and normalize their migration in the gel with the 5′ leader of pre-tRNA^{Phe}, which is not 3′ phosphorylated. SC, a substrate control without RNase P; the three SC reactions (lanes 1, 5, and 10) correspond, respectively, to controls which mimic the assay conditions employed for the native *Pae* holoenzyme (lane 2), *Pae* RPR (lane 6), and *Cma* RPR (lane 9).

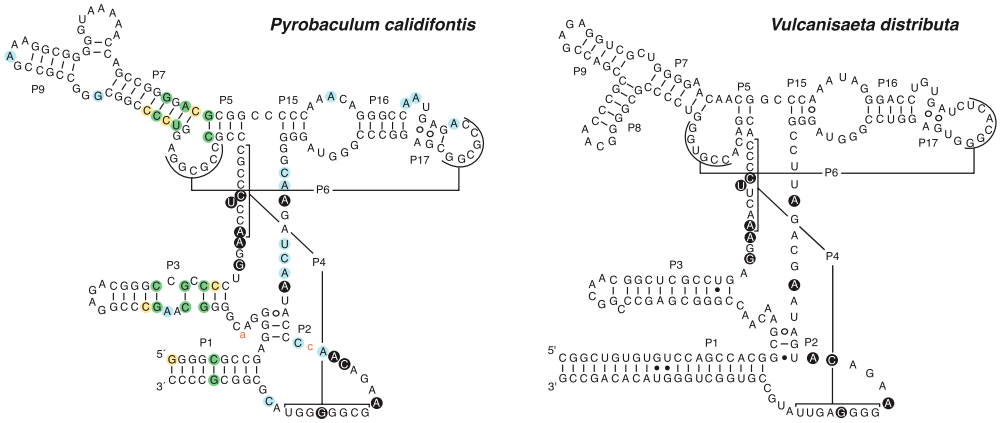

Fig. S4. Predicted secondary structure of *P. calidifontis* and *Vulcanisaeta distributa* RNase P RNAs. Black circles indicate universally conserved nucleotides, and all others highlighted relative to *P. aerophilum* (Fig. 4*A*) as follows: pairs showing covariation (green), conservative G-C to G-U changes (yellow), and differences in unpaired regions (blue).

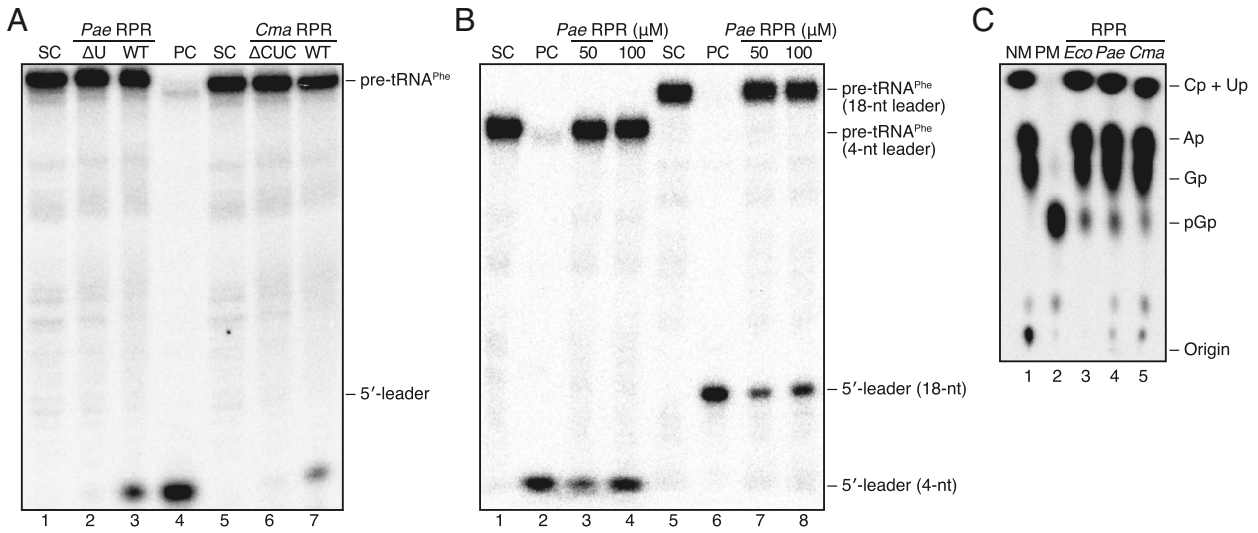

Fig. S5. *P. aerophilum* (*Pae*) and *C. maquilingensis* (*Cma*) RNase P RNAs (RPRs) can process pre-tRNA^{Phe} (G-1) with a 4-nt leader. (*A*) Cleavage of a 5′-labeled substrate was assessed using WT *Pae* (lane 3) or *Cma* (lane 7) RPRs, and their corresponding inactive mutant derivatives (lanes 2 and 6; Fig. 4). (*B*) *Pae* RPR cleaves pre-tRNA^{Phe} (G-1) with either a 4- or 18-nt leader with equal efficiency. SC, a substrate control incubated without the RNase P RNA; PC, a positive control with pre-tRNA^{Phe} processed by *E. coli* RNase P. (*C*) TLC analysis of mature tRNA^{Phe} containing a 5′-pGp, produced by *E. coli* (*Eco*), *Pae*, and *Cma* RPRs (lanes 3, 4, and 5). NM and PM, a negative marker without and a positive marker with pGp.

Table S1. Annotated or computationally identified RNase P proteins and associated DUF54 protein

Genome	Pop5	Rpp30	DUF54/COG1325	Rpp29	Rpp21
Aeropyrum pernix	APE_1161.1 (NP_147751)	**APE_1450.1 (NP_147952)**	**APE_1451.1 (NP_147953)**	APE_0362 (NP_147179)	APE_1092 (NP_147712)
Sulfolobus acidocaldarius	**Saci_0614 (YP_255299)**	**Saci_0615 (YP_255300)**	**Saci_0616 (YP_255301)**	Saci_0589 (YP_255274)	Saci_0033 (YP_254756)
Sulfolobus solfataricus	**SSO0739 (NP_342245)**	**SSO0740 (NP_342246)**	**SSO0741 (NP_342247)**	SSO0710 (P60833)	SSO2225 (NP_343605)
Pyrobaculum aerophilum	**PAE1829 (NP_559579)**	**PAE1830 (NP_559580)**	**PAE1832 (NP_559581)**	PAE1777 (NP_559538)	?
Pyrobaculum calidifontis	**Pcal_0504 (YP_001055398)**	**Pcal_0505 (YP_001055399)**	**Pcal_0503 (YP_001055397)**	Pcal_0551 (YP_001055444)	?
Pyrobaculum arsenaticum	**Pars_0716 (YP_001152954)**	**Pars_0713 (YP_001152951)**	**Pars_0712 (YP_001152950)**	Pars_0767 (YP_001153005)	?
Pyrobaculum islandicum	**Pisl_1748 (YP_931242)**	**Pisl_1749 (YP_931243)**	**Pisl_1750 (YP_931244)**	Pisl_1727 (YP_931221)	?
Thermoproteus neutrophilus	**Tneu_0764 (YP_001794149)**	**Tneu_0765 (YP_001794150)**	**Tneu_0766 (YP_001794151)**	Tneu_0745 (YP_001794130)	?
Vulcanisaeta distributa	**Vdis_2044**	**Vdis_2043**	**Vdis_2042**	Vdis_0362	?
Caldivirga maquilingensis	Cmaq_0774 (YP_001540599)	Cmaq_1319 (YP_001541135)	Cmaq_1852 (YP_001541659)	Cmaq_0514 (YP_001540349)	?
Thermofilum pendens	**Tpen_0387 (YP_919799)**	**Tpen_0388 (YP_919800)**	**Tpen_0389 (YP_919801)**	Tpen_0486 (YP_919897)	Tpen_0560 (YP_919969)
Pyrococcus horikoshii	PH1481 (NP_143344)	**PH1877 (NP_143706)**	**PH1876 (NP_143705)**	PH1771 (NP_143607)	PH1601 (NP_143456)
Pyrococcus furiosus	PF1378 (NP_579107)	**PF1914 (NP_579643)**	**PF1913 (NP_579642)**	PF1816 (NP_579545)	PF1613 (NP_579342)
Archaeglobus fulgidus	AF0489 (NP_069325)	**AF2317 (NP_071142)**	**AF2318 (NP_071143)**	AF1917 (NP_070742)	AF0109 (NP_068950)
Methanothermobacter thermautotrophicus	**MTH687 (NP_275830)**	**MTH688 (NP_275831)**	**MTH689 (NP_275832)**	MTH11 (NP_275156)	MTH1618 (NP_276730)
Methanopyrus kandleri	**MK0386 (NP_613671)**	**MK0387 (NP_613672)**	**MK0388 (NP_613673)**	MK1216 (NP_614499)	MK1622 (NP_614905)

Each protein is represented by the gene locus tag with the protein accession number in parenthesis. Proteins located in the same neighborhood (within three genes) are highlighted in bold.

Table S2. RNase P RNA search using Infernal v1.0 (10)

Genome	Covariance Model Search Score (bits)	
	Archaeal RNase P RNA (RF00373) Model	*Pyrobaculum* RNase P RNA Model
Crenarchaeota		
Aeropyrum pernix K1	125.72	5.07
Caldivirga maquilingensis IC-167	Not Found	52.95
Desulfurococcus kamchatkensis 1221n	166.33	7.44
Hyperthermus butylicus	124.17	Not Found
Ignicoccus hospitalis KIN4-I	117.03	2.32
Ignisphaera aggregans AQ1.S1	174.72	4.79
Metallosphaera sedula	121.51	7.21
Pyrobaculum aerophilum	Not Found	206.22
Pyrobaculum arsenaticum	1.34	206.63
Pyrobaculum calidifontis	Not Found	192.45
Pyrobaculum islandicum	0.05	217.18
Thermoproteus neutrophilus V24Sta	Not Found	196.85
[to be reclassified as a *Pyrobaculum* species]		
Pyrobaculum oguniense	Not Found	210.93
Staphylothermus hellenicus	139.85	Not Found
Staphylothermus marinus F1	165.76	3.05
Sulfolobus acidocaldarius	184.72	Not Found
Sulfolobus islandicus L.D.8.5	169.11	Not Found
Sulfolobus islandicus L.S.2.15	169.11	Not Found
Sulfolobus islandicus M.14.25	169.11	Not Found
Sulfolobus islandicus M.16.4	169.11	Not Found
Sulfolobus islandicus M.16.27	169.11	Not Found
Sulfolobus islandicus Y.G.57.14	169.15	Not Found
Sulfolobus islandicus Y.N.15.51	164.82	Not Found
Sulfolobus solfataricus P2	180.39	4.08
Sulfolobus tokodaii str. 7	162.43	8.65
Thermofilum pendens Hrk 5	72.24	6.24
Thermosphaera aggregans M11TL	143.48	Not Found
Vulcanisaeta distributa IC-017	5.32	48.24
Euryarchaeota		
Aciduliprofundum boonei T469	175.42	Not Found
Archaeoglobus fulgidus	124.98	Not Found
Archaeoglobus profundus Av18	178.91	4.65
Ferroglobus placidus AEDII12DO	162.95	5.37
Haloarcula marismortui	89.22	5.35
Halobacterium salinarum R1	136.86	5.19
Halobacterium sp. NRC-1	136.86	5.19
Haloferax volcanii DS2	133.15	9.89
Halogeometricum borinquense PR3	60.64	8.00
Halomicrobium mukohataei	106.3	4.40
Haloquadratum walsbyi	172.81	Not Found
Halorhabdus utahensis	125.24	6.72
Halorubrum lacusprofundi	162.14	7.91
Haloterrigena turkmenica	112.04	5.92
Methanothermobacter thermautotrophicus str. Delta H	228.58	13.49
Methanobrevibacter ruminantium M1	193.70	Not Found
Methanobrevibacter smithii ATCC 35061	206.29	Not Found
Methanocaldococcus sp. FS406-22	147.62	Not Found
Methanocaldococcus fervens AG86	144.16	Not Found
Methanocaldococcus infernus ME	126.83	Not Found
Methanocaldococcus vulcanius M7	144.13	Not Found
Methanocella paludicola SANAE	165.96	9.18
Methanococcoides burtonii	184.98	Not Found
Methanococcus aeolicus	102.82	Not Found
Methanocaldococcus jannaschii	149.37	Not Found
Methanococcus maripaludis C5	117.63	Not Found
Methanococcus maripaludis C6	113.98	Not Found
Methanococcus maripaludis C7	115.95	Not Found
Methanococcus maripaludis S2	120.92	Not Found
Methanococcus vannielii SB	113.31	Not Found
Methanocorpusculum labreanum Z	173.66	Not Found
Methanoculleus marisnigri JR1	196.35	12.39
Methanohalophilus mahii	155.97	Not Found
Methanopyrus kandleri AV19	53.55	5.29
Candidatus Methanoregula boonei	199.44	3.83
Methanosaeta thermophila PT	192.62	Not Found
Methanosarcina acetivorans C2A	181.08	6.52

Genome	Covariance Model Search Score (bits)	
	Archaeal RNase P RNA (RF00373) Model	*Pyrobaculum* RNase P RNA Model
Methanosarcina barkeri str. Fusaro	187.82	Not Found
Methanosarcina mazei Go1	191.12	8.56
Methanosphaera stadtmanae	137.68	Not Found
Candidatus Methanosphaerula palustris E1-9c	201.92	5.85
Methanospirillum hungatei JF-1	197.20	9.90
Natrialba magadii	96.05	8.78
Natronomonas pharaonis	153.40	5.56
Picrophilus torridus	205.24	3.44
Pyrococcus abyssi GE5	196.36	4.10
Pyrococcus furiosus	196.16	4.77
Pyrococcus horikoshii OT3	190.46	Not Found
Thermococcus barophilus	171.24	Not Found
Thermococcus gammatolerans EJ3	175.13	8.20
Thermococcus kodakarensis KOD1	171.24	4.66
Thermococcus onnurineus NA1	175.74	4.99
Thermococcus sibiricus MM 739	199.28	3.91
Thermoplasma acidophilum	206.76	Not Found
Thermoplasma volcanium GSS1	186.71	2.59
Uncultured methanogenic archaeon RC-I	191.28	6.72
Nanoarchaeota		
Nanoarchaeum equitans Kin4-M	Not Found	Not Found
Korarchaeota		
Candidatus Korarchaeum cryptofilum OPF8	151.76	8.67
Thaumarchaeota		
Cenarchaeum symbiosum	173.78	7.99
Nitrosopumilus maritimus SCM1	184.41	Not Found

Archaeal RNase P RNA Rfam covariance model RF00373 and *Pyrobaculum* RNase P RNA covariance model with Infernal v1.0 (10) were used to search 88 archaeal genomes. Identified RNase P RNA candidates with highest bit scores were reported.