Chemistry & Biology *17*

Supplemental Information

A Eubacterial Riboswitch Class That Senses the Coenzyme Tetrahydrofolate

Tyler D. Ames, Dmitry A. Rodionov, Zasha Weinberg, and Ronald R. Breaker

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Discovery and refinement of the THF motif

The THF aptamer motif was detected using a computational pipeline described previously[1]. Briefly, an early version of a method to cluster intergenic regions by sequence similarity[2] was applied to genomes of lactic acid bacteria from RefSeq[3] version 25. Conserved RNA structures were predicted in these clusters using CMfinder[4] and automated homology searches performed using RAVENNA[5,6]. Also, the conserved structure of the RNA element was refined by manual covariance analysis using multiple alignment of upstream regions of *folT* genes from Firmicutes and the Mfold RNA folding algorithm[7]. Manual homology searches also using RAVENNA and improvements to the structural model were then conducted. Ultimately, we searched the microbial subsets of RefSeq version 32, and metagenome DNA sequences isolated from acid mine drainage[8], soil and whale fall[9], human gut[10,11], mouse gut[12], gutless sea worms[13], sludge[14], Global Ocean Survey scaffolds[15,16], other marine sequences[17], and termite hindgut[18]. The degree of conservation and covariation at different positions on the structure shown in Figure 1A and Figure S1 was calculated using the Infernal[19] implementation of the GSC algorithm[20], with previously established protocols[1].

Chemicals and DNA oligonucleotides

All folate, dihydrofolate (DHF), tetrahydrofolate (THF), and modified THF derivatives were provided as a gift from Merck Eprova. An additional tetrahydrofolate sample was purchased from Sigma-Aldrich, which was used as noted. Other compounds used as ligands in in-line probing assays were supplied by Sigma-Aldrich. DNA oligonucleotides were prepared by Sigma-Genosys.

RNA Preparation

The following synthetic DNAs were used to generate double-stranded (ds) DNA templates for *in vitro* transcription.

1: 5′-
TAATACGACTCACTATAGGGTATGAAGGCAGAGTAGGTGTTATGCGTTAAGTGTTAGGGGAT
GGGAAGTT

2: 5′-
TAATACGACTCACTATAGGGTATGAAGGCAGAGTAGGACATATGCGTTAAGTGTTAGGGGA
TGGGAAGTT

3: 5′-
TAATACGACTCACTATAGGGTATGAAGGCAGAGTAGGTGTTATGCGTTAAGTGTTTCCGGAT
GGGAAGTT

4: 5′-TAATACGACTCACTATAGGGGGTGTTATGCGTTAAGTGTTAGGGGATGGGAAGTT

5: 5′-
CTGTGACAGTGGACGCGGTGTTATATCGCGAAATGAGTTCTTCGTTTAGGAGCAACTTCCCA
TCCCCT

6: 5′-
CTGTGACAGTGGACGCGGTGTTATATCGCGAAATGAGTTCTTCGTTTAGGAGCAACTTCCCA
TCCGGA

7: 5′-
CTGTGACAGTGGACGCGGACATATATCGCGAAATGAGTTCTTCGTTTAGGAGCAACTTCCCA
TCCCCT

8: 5′-
CTGTGACAGTGGACGCGGTGTTATATCGCGAAATGAGTTCTTCGTTTTCCAGCAACTTCCCAT
CCGGA

9: 5′-GGTGTTATATCGCGAAATGAGTTCTTCGTTTAGGAGCAACTTCCCATCCCCT

DNAs 1 and 5 were used to make dsDNA template for the 106 *folC* RNA, 2 and 5 for M1, 3 and 6 for M2, 2 and 7 for M3, 3 and 8 for M4, and 4 and 9 for M5. To generate full length dsDNA temples, 300 pmoles of each primer were added to a 100 µL reaction containing 50 mM Tris-HCl (pH 8.3 at 23°C), 75 mM KCl, 3 mM $MgCl_2$, 10 µM dithiothreitol (DTT), 1 mM of each of the four deoxynucleoside 5′-triphosphates (dNTPs), and 800 units of SuperScript II Reverse Transcriptase (RT) (Invitrogen). The solution was incubated at 42°C for 2 hours.

A 15 µL aliquot of the from the RT reaction was used to supply template for a 100 µL transcription reaction which contained 80 mM HEPES-KOH (pH 7.5 at 23°C), 24 mM $MgCl_2$, 2 mM spermidine, 40 mM DTT, 2.5 mM each of the four ribonucleoside 5′-triphosphates (NTPs), and T7 RNA polymerase (final concentration of 10 units μL^{-1}). Samples were heated for 2 hours at 37°C. RNA was purified using the QIAquick Nucleotide removal Kit (Qiagen).

In-line probing assays

In-line probing analyses were conducted using methods similar to those described previously[21]. RNA transcripts were dephosphorylated using alkaline phosphatase (Roche) according to the manufacturer's protocol. The RNAs were subsequently 5′ labeled with [γ-^{32}P] ATP using T4 polynucleotide kinase (New England Biolabs) according the to the manufacturer's instructions. Radiolabeled RNAs were then purified via denaturing (8 M urea) 6% polyacrylamide gel electrophoresis (PAGE), and desired products were eluted from the gel using 10 mM Tris-HCl (pH 7.5 at 23°C), 200 mM NaCl and 1 mM EDTA (pH 8.0 at 23°C). The purified radiolabeled RNAs were then incubated for approximately 40 hours at 23°C in a solution composed of 50 mM Tris-HCl (pH 8.3 at 23°C), 20 mM $MgCl_2$, 100 mM KCl, and varying concentrations of THF or other potential ligands (folate, DHF, etc). In some in-line probing reactions, such as those used in Figure 2 of the main text, 1 mM DTT was added to avoid excessive breakdown of some of the compounds examined as potential ligands[22].

The resulting RNA degradation products were separated via denaturing 10% PAGE and imaged using a Molecular Dynamics PhosphorImager (GE Healthcare). Band intensities were quantified using SAFA v1.1 software[23]. Groupings of consecutive positions of the twenty bands which modulated the greatest were used to make K_D estimates. Other positions of decreasing, increasing, or constant cleavage indicated in Figure 2A on the structural model were visually identified from the gel image.

Figure S1. Multiple sequence alignment of the THF aptamer

Individual sequences are putative THF aptamers. Accession numbers for bacterial species and environmental sequences are shown in the left column, followed by the nucleotide coordinates of the aptamer. Colored positions of the sequences indicate areas involved in predicted base paired elements. Gray positions represent nucleotides that are mismatched. Based paired areas are additionally indicated in the fifth line from the bottom with angle brackets. In the fourth to the last line, a "2" denotes a positions containing a covarying mutation, and a "?" indicates a positions where the predicted base pair is broken > 10% of the time with non-Watson-Crick or G-U pairings. The last line contains the consensus sequence. Annotations for this line are as follows: "R" = "A" or "G", "Y" = "C" or "U". Red nucleotides are conserved in at least 97% of the representatives; black nucleotides: 90%; gray nucleotides: 75%. Red, black and gray shading indicate that a nucleotide of any identity (n) is present in at least 97%, 90%, or 75% of the representatives, respectively. Duplicate representatives (see below) with identical sequences to another THF aptamer hit are excluded in the alignment and in the analysis of nucleotide representation.

Duplicates are listed in pairs, where the hit (or hits) not represented in the alignment is (are) listed after its identical representative. Hits are separated by "=".

NC_009442.1/125748-125890 = NC_009443.1/125651-125793

NC_008526.1/2247840-2247748 = NZ_ABQV01000046.1/83746-83654

BABB01000157.1/63-158 = BABD01002599.1/316-411

BAAX01002679.1/2103-2006 = BABG01001398.1/1123-1026 = BABG01002228.1/1277-1180

NC_004668.1/903576-903739 = NZ_ABPI01000001.1/702486-702649

NZ_AAAK03000102.1/3116-3216 = BABC01004059.1/195-295 = BABC01009447.1/651-751

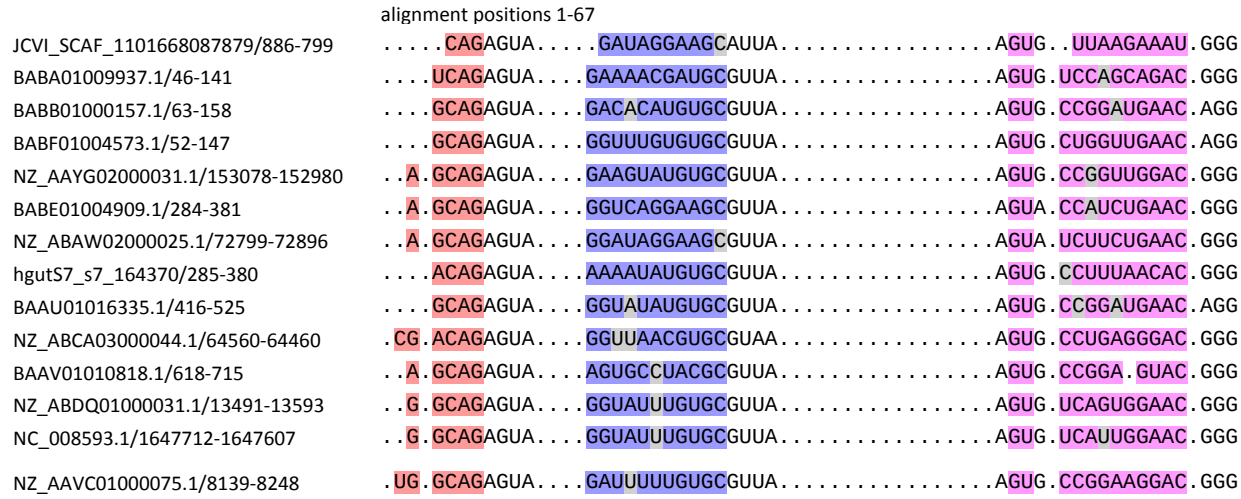
NZ_ABCA03000044.1/64560-64460 = BABG01026547.1/439-539 = BABG01026548.1/657-557

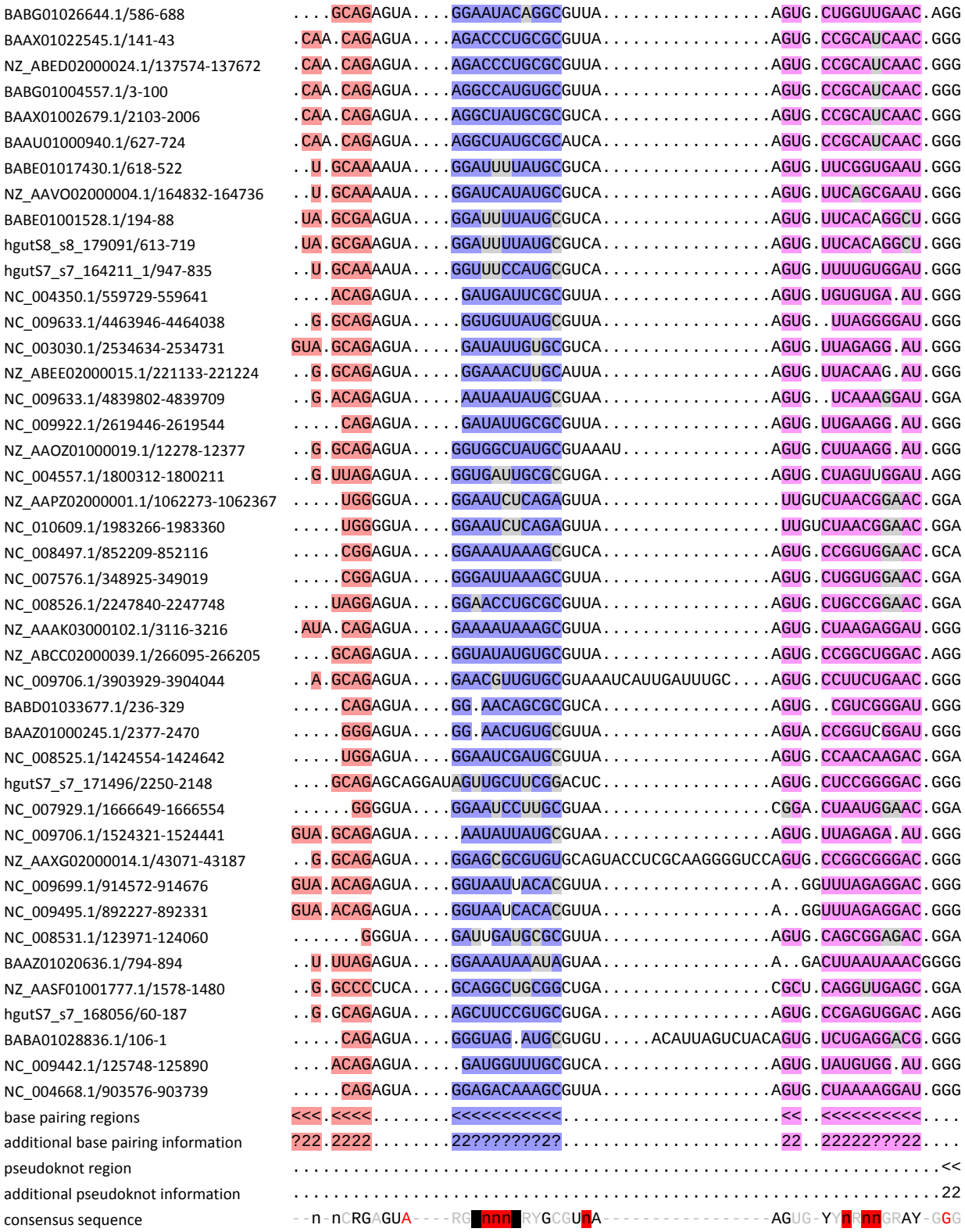
NC_009699.1/914572-914676 = NC_010516.1/904741-904845 = NC_010520.1/924118-924222 = NZ_ABDO02000001.1/854207-854311 = NZ_ABDP01000015.1/2757-2861

NC_009495.1/892227-892331 = NC_009697.1/864554-864658 = NC_009698.1/864721-864825

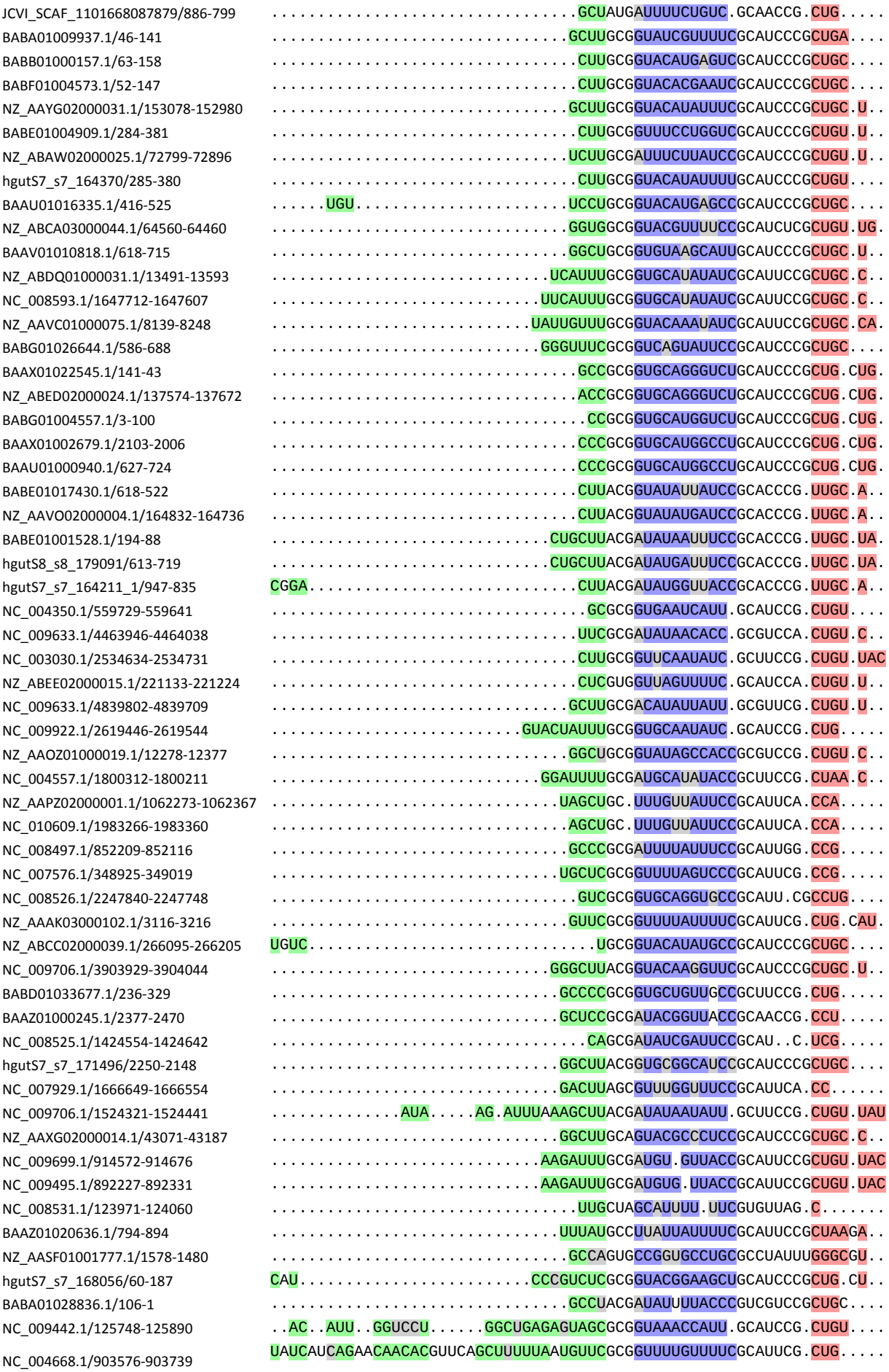
NZ_ABED02000024.1/137574-137672 = hgutS7_s7_175872/176-78

NC_010609.1/1983266-1983360 = NC_009513.1/1943850-1943944




BABG01026644.1/586-688
BAAX01022545.1/141-43
NZ_ABED02000024.1/137574-137672
BABG01004557.1/3-100
BAAX01002679.1/2103-2006
BAAU01000940.1/627-724
BABE01017430.1/618-522
NZ_AAVO02000004.1/164832-164736
BABE01001528.1/194-88
hgutS8_s8_179091/613-719
hgutS7_s7_164211_1/947-835
NC_004350.1/559729-559641
NC_009633.1/4463946-4464038
NC_003030.1/2534634-2534731
NZ_ABEE02000015.1/221133-221224
NC_009633.1/4839802-4839709
NC_009922.1/2619446-2619544
NZ_AAOZ01000019.1/12278-12377
NC_004557.1/1800312-1800211
NZ_AAPZ02000001.1/1062273-1062367
NC_010609.1/1983266-1983360
NC_008497.1/852209-852116
NC_007576.1/348925-349019
NC_008526.1/2247840-2247748
NZ_AAAK03000102.1/3116-3216
NZ_ABCC02000039.1/266095-266205
NC_009706.1/3903929-3904044
BABD01033677.1/236-329
BAAZ01000245.1/2377-2470
NC_008525.1/1424554-1424642
hgutS7_s7_171496/2250-2148
NC_007929.1/1666649-1666554
NC_009706.1/1524321-1524441
NZ_AAXG02000014.1/43071-43187
NC_009699.1/914572-914676
NC_009495.1/892227-892331
NC_008531.1/123971-124060
BAAZ01020636.1/794-894
NZ_AASF01001777.1/1578-1480
hgutS7_s7_168056/60-187
BABA01028836.1/106-1
NC_009442.1/125748-125890
NC_004668.1/903576-903739
base pairing regions
additional base pairing information
pseudoknot region
additional pseudoknot information
consensus sequence
alignment positions 68-134
JCVI_SCAF_1101668087879/886-799
BABA01009937.1/46-141
BABB01000157.1/63-158
BABF01004573.1/52-147
NZ_AAYG02000031.1/153078-152980
BABE01004909.1/284-381
NZ_ABAW02000025.1/72799-72896

alignment positions 135-200


JCVI_SCAF_1101668087879/886-799
BABA01009937.1/46-141
BABB01000157.1/63-158
BABF01004573.1/52-147
NZ_AAYG02000031.1/153078-152980
BABE01004909.1/284-381
NZ_ABAW02000025.1/72799-72896
hgutS7_s7_164370/285-380
BAAU01016335.1/416-525
NZ_ABCA03000044.1/64560-64460
BAAV01010818.1/618-715
NZ_ABDQ01000031.1/13491-13593
NC_008593.1/1647712-1647607
NZ_AAVC01000075.1/8139-8248
BABG01026644.1/586-688
BAAX01022545.1/141-43
NZ_ABED02000024.1/137574-137672
BABG01004557.1/3-100
BAAX01002679.1/2103-2006
BAAU01000940.1/627-724
BABE01017430.1/618-522
NZ_AAVO02000004.1/164832-164736
BABE01001528.1/194-88
hgutS8_s8_179091/613-719
hgutS7_s7_164211_1/947-835
NC_004350.1/559729-559641
NC_009633.1/4463946-4464038
NC_003030.1/2534634-2534731
NZ_ABEE02000015.1/221133-221224
NC_009633.1/4839802-4839709
NC_009922.1/2619446-2619544
NZ_AAOZ01000019.1/12278-12377
NC_004557.1/1800312-1800211
NZ_AAPZ02000001.1/1062273-1062367
NC_010609.1/1983266-1983360
NC_008497.1/852209-852116
NC_007576.1/348925-349019
NC_008526.1/2247840-2247748
NZ_AAAK03000102.1/3116-3216
NZ_ABCC02000039.1/266095-266205
NC_009706.1/3903929-3904044
BABD01033677.1/236-329
BAAZ01000245.1/2377-2470
NC_008525.1/1424554-1424642
hgutS7_s7_171496/2250-2148
NC_007929.1/1666649-1666554
NC_009706.1/1524321-1524441
NZ_AAXG02000014.1/43071-43187
NC_009699.1/914572-914676
NC_009495.1/892227-892331
NC_008531.1/123971-124060
BAAZ01020636.1/794-894
NZ_AASF01001777.1/1578-1480
hgutS7_s7_168056/60-187
BABA01028836.1/106-1
NC_009442.1/125748-125890
NC_004668.1/903576-903739

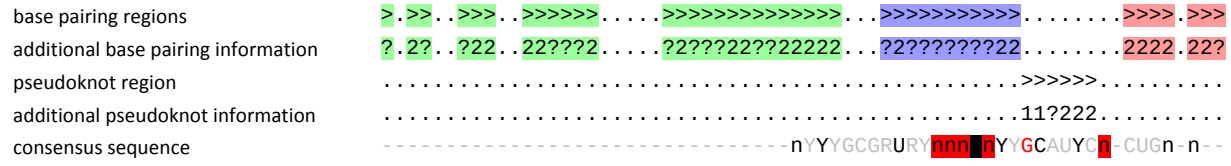


Figure S2. Phylogenetic tree of FolT family of folate transporters

folT genes under regulation of THF riboswitch are marked by yellow. *folT* genes in chromosomal clusters with the folate metabolism gene *folC* are in red. Species abbreviations are listed in Table S1. Phylogenetic tree was constructed using maximum likelihood approach implemented in the Phylip software package[24].

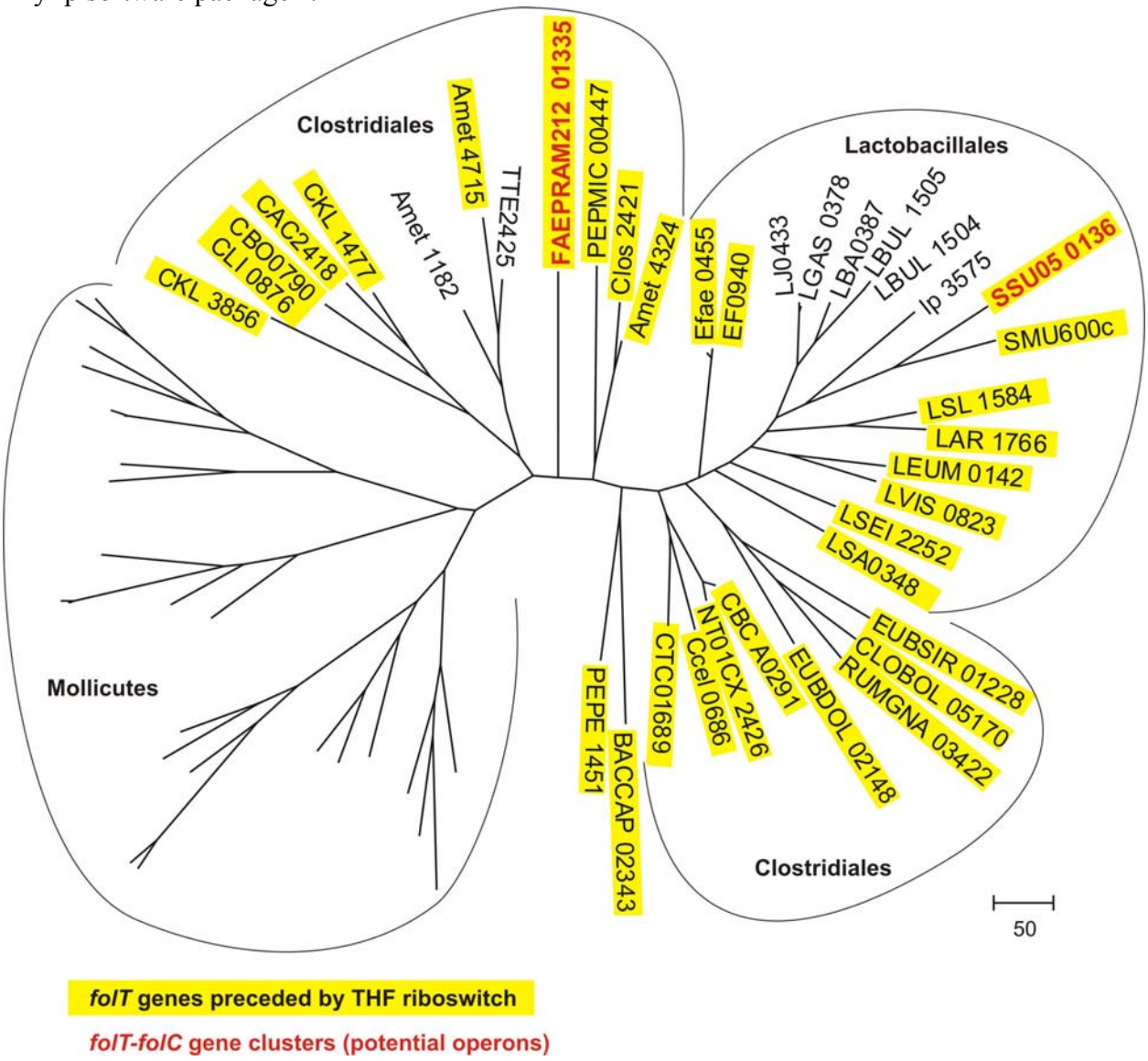


Figure S3. The THF riboswitch aptamer can bind dihydrofolate (DHF) and tetrahydrobiopterin (THBP)

(**a**) In-line probing analysis of the *A. metalliredigens* 106 *folT* RNA using the compounds DHF and THBP as ligands. The NR, T1 and $^{-}$OH lanes were loaded, respectively, with 5′ ^{32}P-labeled precursor (Pre) RNAs subjected to no reaction, partial digestion with RNase T1 (cleaves after G residues), and partial digestion under alkaline conditions (cleaves at every position). Other lanes include Pre RNAs subjected to in-line probing reaction conditions without any additional compound (–) or with increasing concentrations of DHF or THBP (1 nM through 10 μM from left to right). Vertical lines identify bands that were quantitated and used to estimate K_D values in **b**.

(**b**) Plot of the normalized band intensities (normalized to the intensity measured at 1 nM ligand) versus the logarithm of the concentration of ligand. The solid line represents a theoretical binding curve for a one-to-one interaction with a K_D of 300 nM.

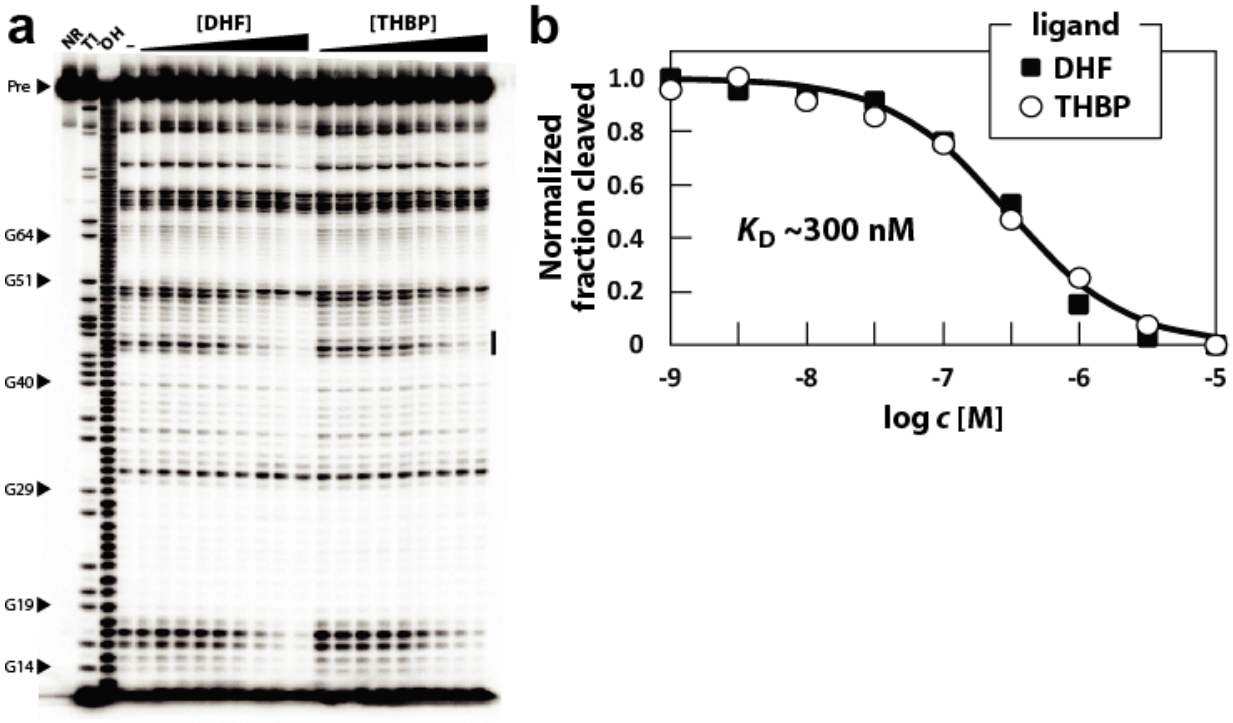


Figure S4. Disruption of predicted base-paired regions of the THF aptamer interferes with ligand binding

In-line probing analyses using THF as a ligand with the RNA constructs (**a**) M1, (**b**) M2, (**c**) M3, and (**d**) M4. THF concentrations range from 100 nM to 1 mM. (**e**) Plot of the normalized band intensities versus the logarithm of the concentration of THF. The solid lines represent a theoretical binding curve for a one-to-one interaction with K_D values of 1 μM and 80 μM. Other details are as described in the legend to Figure S3. M2 is not represented on this graph as there was no evidence of it being able to bind to THF.

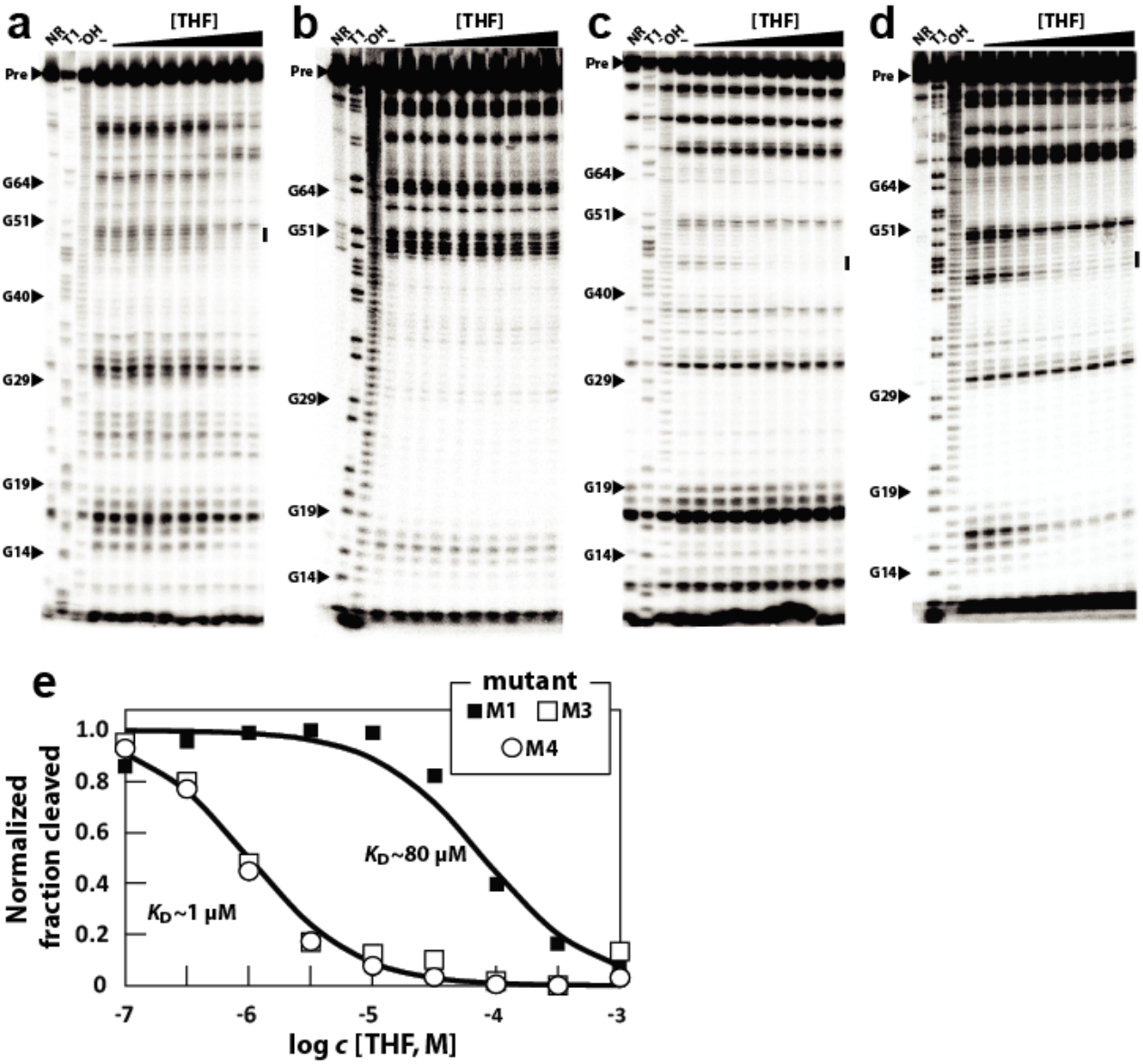


Table S1. THF riboswitch-containing genomic loci and predicted target genes in microbial genomes and metagenomes

Nucleotide ID in Genbank	Organism	Target gene (operon)	Locus tag for target gene
NC_009633.1/4463946-4464038	Alkaliphilus metalliredigens QYMF	*folT*	Amet_4324
NC_009633.1/4839802-4839709	Alkaliphilus metalliredigens QYMF	*folT*	Amet_4715
NC_009922.1/2619446-2619544	Alkaliphilus oremlandii OhILAs	*folT*	Clos_2421
NZ_AAXG02000014.1/43071-43187	Bacteroides capillosus ATCC 29799	*folT*	BACCAP_02343
NC_003030.1/2534634-2534731	Clostridium acetobutylicum ATCC 824	*folT*	CAC_2418
NZ_ABCC02000039.1/266095-266205	Clostridium bolteae ATCC BAA-613	*folT*	CLOBOL_05170
NC_009495.1/892227-892331	Clostridium botulinum A str. ATCC 3502	*folT*	CBO0790
NZ_ABDQ01000031.1/13491-13593	Clostridium botulinum C str. Eklund	*folT*	CBC_A0291
NC_009699.1/914572-914676	Clostridium botulinum F str. Langeland	*folT*	CLI_0876
NC_009706.1/3903929-3904044	Clostridium kluyveri DSM 555	*folT*	CKL_3856
NC_009706.1/1524321-1524441	Clostridium kluyveri DSM 555	*folT*	CKL_1477
NC_008593.1/1647712-1647607	Clostridium novyi NT	*folT*	NT01CX_2426
NC_004557.1/1800312-1800211	Clostridium tetani E88	*folT*	CTC01689
NC_004668.1/903576-903739	Enterococcus faecalis V583	*folT*	EF0940
NZ_AAAK03000102.1/3116-3216	Enterococcus faecium DO	*folT*	Efae_0455
NZ_ABAW02000025.1/72799-72896	Eubacterium dolichum DSM 3991	*folT*	EUBDOL_02148
NZ_ABCA03000044.1/64560-64460	Eubacterium siraeum DSM	*folT*	EUBSIR_01228
NZ_ABED02000024.1/137574-137672	Faecalibacterium prausnitzii M21/2	*folT-folC*	FAEPRAM212_01335
NC_008497.1/852209-852116	Lactobacillus brevis ATCC 367	*folT*	LVIS_0823
NC_008526.1/2247840-2247748	Lactobacillus casei ATCC 334	*folT*	LSEI_2252
NC_010609.1/1983266-1983360	Lactobacillus reuteri JCM 1112	*folT*	LAR_1766
NC_007576.1/348925-349019	Lactobacillus sakei subsp. sakei 23K	*folT*	LSA0348
NC_007929.1/1666649-1666554	Lactobacillus salivarius UCC118	*folT*	LSL_1584
NC_008531.1/123971-124060	Leuconostoc mesenteroides subsp. mesenteroides ATCC 8293	*folT*	LEUM_0142
NZ_ABEE02000015.1/221133-221224	Parvimonas micra ATCC 33270	*folT*	PEPMIC_00447
NC_008525.1/1424554-1424642	Pediococcus pentosaceus ATCC 25745	*folT*	PEPE_1451
NZ_AAYG02000031.1/153078-152980	Ruminococcus gnavus ATCC 29149	*folT*	RUMGNA_03422
NZ_AAVO02000004.1/164832-164736	Ruminococcus obeum ATCC 29174	*folE-folQ-folP-folBK*	RUMOBE_01496
NC_004350.1/559729-559641	Streptococcus mutans UA159	*folT*	SMU.600c
NC_009442.1/125748-125890	Streptococcus suis 05ZYH33	*folT-folC*	SSU05_0136
hgutS7_s7_164370/285-380	human distal gut microbiome	*folT*	
hgutS8_s8_179091/613-719	human distal gut microbiome	*folT*	
hgutS7_s7_164211_1/947-835	human distal gut microbiome	*folE*	
hgutS7_s7_171496/2250-2148	human distal gut microbiome	*folT*	
hgutS7_s7_168056/60-187	human distal gut microbiome	*folT*	
BABA01009937.1/46-141	human gut metagenome	*folT*	
BABB01000157.1/63-158	human gut metagenome	*folT*	
BABF01004573.1/52-147	human gut metagenome	*folT*	
BABE01004909.1/284-381	human gut metagenome	*folT*	
BAAU01016335.1/416-525	human gut metagenome	-	
BAAV01010818.1/618-715	human gut metagenome	*folT*	
BABG01026644.1/586-688	human gut metagenome	*folT*	
BAAX01022545.1/141-43	human gut metagenome	-	
BABG01004557.1/3-100	human gut metagenome	*folT-folC*	
BAAX01002679.1/2103-2006	human gut metagenome	*folT*	
BAAU01000940.1/627-724	human gut metagenome	*folT-folC*	
BABE01017430.1/618-522	human gut metagenome	*folE*	
BABE01001528.1/194-88	human gut metagenome	-	
BABD01033677.1/236-329	human gut metagenome	-	
BAAZ01000245.1/2377-2470	human gut metagenome	*folT*	

BAAZ01020636.1/794-894	human gut metagenome	-
BABA01028836.1/106-1	human gut metagenome	-
JCVI_SCAF_1101668087879/886-799	marine metagenome	*folT*

1. Weinberg, Z., Wang, J.X., Bogue, J., Yang, J., Corbino, K.A., Moy, R.H., & Breaker, R.R. *Genome Biol.* (2010). (submitted)
2. Tseng, H., Weinberg, Z., Gore, J., Breaker, R.R. & Ruzzo, W.L. *J. Bioinform. Comput. Biol.* **7**, 373-388 (2009).
3. Pruitt, K.D., Tatusova, T. & Maglott, D.R. *Nucleic Acids Res.* **33**, D501-504 (2005).
4. Yao, Z., Weinberg, Z. & Ruzzo, W.L. *Bioinformatics* **22**, 445-452 (2006).
5. Weinberg, Z. & Ruzzo, W.L. *Bioinformatics* **22**, 35-39 (2006).
6. Eddy, S.R. & Durbin, R. *Nucleic Acids Res.* **22**, 2079-2088 (1994).
7. Zuker, M. *Nucleic Acids Res.* **31**, 3406-3415 (2003).
8. Tyson, G.W. *et al. Nature* **428**, 37-43 (2004).
9. Tringe, S.G. *et al. Science* **308**, 554-557 (2005).
10. Gill, S.R. *et al. Science* **312**, 1355-1359 (2006).
11. Kurokawa, K. *et al. DNA Res.* **14**, 169-181 (2007).
12. Turnbaugh, P.J. *et al. Nature* **444**, 1027-131 (2006).
13. Woyke, T. *et al. Nature* **443**, 950-955 (2006).
14. García Martín, H. *et al. Nat. Biotechnol.* **24**, 1263-1269 (2006).
15. Rusch, D.B. *et al. PLoS Biol.* **5**, e77 (2007).
16. Venter, J.C. *et al. Science* **304**, 66-74 (2004).
17. Konstantinidis, K.T., Braff, J., Karl, D.M. & DeLong, E.F. *Appl. Environ. Microbiol.* **75**, 5345-5355 (2009).
18. Warnecke, F. *et al. Nature* **450**, 560-565 (2007).
19. Eddy,S.R. ftp://ftp.genetics.wustl.edu/pub/eddy/software/infernal/Userguide.pdf (2005).
20. Gerstein, M., Sonnhammer, E.L. & Chothia, C. *J. Mol. Biol.* **236**, 1067-1078 (1994).
21. Soukup, G.A. & Breaker, R.R. *RNA* **5**, 1308-1325 (1999).
22. Lucock, M.D., Green, M., Hartley, R. & Levene, M.I. *Food Chemistry* **47**, 79-86 (1993).
23. Das, R., Laederach, A., Pearlman, S.M., Herschlag, D. & Altman, R.B. *RNA* **11**, 344-354 (2005).
24. Felsenstein, J. *Cladistics* **5,** 164-166 (1989).