The CMfinder comparative genomics pipeline discovers 22 novel putative RNAs in bacteria
Supplementary Data: additional analysis and data on motifs

Zasha Weinberg, Jeffrey E. Barrick, Zizhen Yao, Adam Roth, Jane N. Kim, Jeremy Gore, Joy Xin Wang, Elaine R. Lee,
Kirsten Block, Narasimhan Sudarsan, Shane Neph, Martin Tompa, Walter L. Ruzzo and Ronald R. Breaker

May 29, 2007

Contents

1 Explanation of figures in this supplement

In Section 2, alignments and gene context are given for all motifs. In some cases, additional analysis is described that is of a specialized nature, and was therefore not included in the main text.

All motifs are shown with the following data (explained in detail below):

- Notes on specific instances of the motif.
- Taxonomy of all motif instances, with codes assigned to each instance
- Gene context of all motif instances, i.e., what genes surround each

abbrev. of hits	taxonomy of species
Van-1-1 to Van-1-1	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio angustum* S14
Vch-1-1 to Vch-1-2	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio cholerae* MO10
Vch-2-1 to Vch-2-2	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio cholerae* O1 biovar eltor str. N16961
Vch-3-1 to Vch-3-2	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio cholerae* O395

Figure 1: Example of taxonomy of motif instances
(See text for explanation.)

abbrev	RefSeq accession		$5'$ at	$3'$ at	genes
[1]Gme-1-13	NC_007517.1	+	1911272	1911497	**RNA**→ possible cytochrome c→
[2]Gsu-1-6	NC_002939.4	-	198072	197847	**RNA**→ **RNA**→ LysM (cd00118)ErfK_YbiS_YhnG (pfam03734)→ lipoprotein, putative→

Figure 2: Example of part 1 of gene context diagrams
(See text for explanation.)

homolog.

- Multiple sequence alignment of motif with a prediction of consensus structure.

1.1 Notes

Notes are numbered starting at 1. Notes indicate such cases as motif instances that may overlap putative ORFs, or have other features. Whenever the motif instance appears in the gene context or in the alignment, the superscript note number is given.

Example: note #1 for GEMM (see Page 6) reads " [1] The $5'$ side of a predicted rho-independent transcription terminator stem overlaps the $3'$ side of the P2 stem..." Motif instances in which predicted terminator stems overlap their P2 stem will be indicated with a superscript "1".

1.2 Taxonomy and motif instance codes

The taxonomy of all motif instances is listed, and abbrevations are assigned to motif instances. An example is presented in Fig.1. In this example, several species in the genus *Vibrio* are shown. Motif instances are abbreviated. For example, the *Vibrio angustum* instance is abbreviated "Van-1-1". The abbreviation "Van" is derived from the species name.

Several species of *Vibrio cholerae* are given, each having two instances of the motif. In the last line, where the abbreviations are "Vch-3-1 to Vch-3-2", the number "3" indicates that this is the third *Vibrio cholerae* strain listed. The two instances will be written as "Vch-3-1" and "Vch-3-2".

Species names and taxonomy are extracted from RefSeq annotations.

1.3 Gene context

The gene context is divided into two parts: (1) a list of motif instances, with genome coordinates, and diagrams of neighboring genes, (2) descriptions of gene families into which these genes are assigned via the Conserved Domain Database.

1.3.1 Gene context: part 1

An example of part 1 is in Fig. 1.

The first line in the example labels the columns. The next line is an instance in *Geobacter metallireducens*. Its abbreviation is "Gme-1-13". The

RefSeq accession number is given, as are nucleotide coordinates of the 5′ and 3′ ends of the sequence included in the alignment.

Note: we often include extra 3′ sequence to show start codons and possible expression platforms; therefore the coordinates given may be larger than the putative conserved motif. Moreover, it is likely in those cases that the coordinates will contains sequence that is not part of the functional element.

If the RNA is on the minus strand, the 5′ coordinate will be greater than the 3′ one. The strand (plus or minus) is shown redundantly as "+" or "-". The genes are shown next.

The putative motif instance is always shown using the word "RNA" with an arrow pointing in its direction. Genes forming a putative transcription unit are then shown. In this case, there is just one "possible cytochrome c" gene in the transcription unit. At the beginning of the line, the superscript "1" indicates that note #1 applies, so this has a predicted transcription terminator overlapping P2.

The third line in the example has a more complex gene context. The text "**RNA**→" is listed twice, indicating that there are tandem RNAs. Two genes (each with arrows pointing in the same direction as the RNA) appear in the putative transcriptional unit. The first gene, written as "LysM (cd00118)ErfK_YbiS_YhnG (pfam03734)→" has two domains, each of which were assigned to a different conserved domain (cd00118) and (pfam03734). Conserved domains that appear in more than one line of the gene context are given a color to help differentiate them from other domains. (However, there are a limited number of distinct colors, so sometimes the same color is assigned to more than one domain.)

Note: "hypo" = "hypothetical protein". "hypo (cons)" = "conserved hypothetical protein". Genes that were not assigned to a conserved domain are always colored gray. Due to limited genome annotation, genes from environmental sequences are typically of "unknown" type.

Motif instances are listed in an order that tends to put those with similar sequences together (which makes the alignment easier to analyze). The one exception is that GEMM RNAs are first grouped into broad taxa.

1.3.2 Gene context: part 2

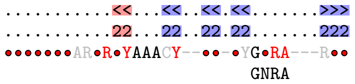
The second part of the gene context describes each conserved domain that was associated with a gene instance. The descriptions are extracted from the Conserved Domain Database, but only the first part of the description is used. (The full descriptions are too long.) The conserved domains are sorted by code (e.g., "cdd00118" comes before "pfam03734").

1.4 Multiple sequence alignment

The multiple sequence alignment of GEMM RNA begins on Page 21. The alignment principally consists of the nucleotides of each motif instance, with dots used to force corresponding nucleotides to line up. The top line says "alignment positions 1 · · · 160"; wide alignments are split into multiple blocks so that they fit on the page.

Each row is labeled with the abbreviation of the motif instance, e.g., "Sth-1-1".

The nucleotides are then shown. At the bottom of each block of the alignment, alignment columns are annotated. The column annotations of GEMM start off like this:



```
...........<<...<<..<<.<<.......>>>
...........22...22..22.22.......222
●●●●●●●AR●R●YAAACY--●●-●YG●RA---R●●
                         GNRA
```

The top line specifies which alignment columns typically base pair; matching angle brackets ("<" and ">") indicate pairing. Different stems and different parts of stems are drawn in different colors to assist in matching up the angle brackets.

The second line shows which base-paired columns covary, etc. A period (".") is printed when the column is not base paired. A question mark ("?") indicates that the pairing is either non-canonical or missing a nucleotide more than 5% of the time. (See Methods in the main paper.) Otherwise, "0" means the nucleotides in the pair are perfectly conserved (so, no evidence for or against covariation), "1" means that there is a compatible mutation and "2" means that there is a compensatory mutation.

The third line is the consensus nucleotide sequence, and is related to the consensus diagrams: the same red/black/gray letters are given for conserved nucleotides, while circles indicate nucleotides whose identity is not consevred. The colors and meanings of letters and circles is the same as in the main paper. A gray dash ("-") indicates that there is usually no nucleotide in that position, which is what happens when a column cannot be assigned a letter or circle. (Note: consensus diagram drawings differ from the alignment consensus sequence, because diagrams are drawn by hand to highlight conserved features and variations, whereas the consensus sequence is computed automatically.)

The last line of per-column annotation indicates where the GNRA tetraloop is. Abbreviations used in this line are as follows:

- `AUG` : predicted start codon.
- `SD+++` : predicted Shine-Dalgarno sequence (including columns annotated with the '`+`' symbols). In all cases our annotated Shine-Dalgarno sequences are based on manual inspection, since these are rarely annotated in genome sequences.
- `w` : wobble positions. This is used for the ATPC motif, where we wish to show that the pattern of mutations after our re-annotated start codons is expected for a coding sequence.
- `T` : predicted start of transcription, using [11] in prokaryote mode.
- `GNRA` : GNRA tetraloop
- `5-TR` and `3TR` : $5'$ and $3'$ parts of GNRA tetraloop receptor

(The yellow nucleotides in GEMM RNA correspond to the stems of predicted transcription terminators, as described on Page 21.)

1.5 Notes on display of gene contexts and alignments

In alignments, RNAs within motifs that have exactly the same sequence as an existing member are not shown. RNAs with duplicate sequences are, however, shown in the gene context.

Except as noted, start codons were taken according to the annotation of the given genome in GenBank. Shine-Dalgarno sequences, being infrequently annotated, were estimated manually.

Gene information is automatically extracted from RefSeq genome annotations, and automatically assigned to Conserved Domain Database clusters. We have modified this data in the following contexts:

- Annotated genes that overlapped motifs are not shown in the gene context when we have reason to believe that either the gene is a false positive, or both RNA and gene are real and transcribed on opposite strands; these cases are always noted. We generally adopted the strategy of using BLAST to see if the putative gene is conserved, and assumed that unconserved genes are plausible false annotations. We acknowledge that BLAST may not find homologs with significant E-values when the query gene is very short, even if the gene is conserved.
- The Conserved Domain Database is comprised of models from COG, Pfam and the SMART database, as well as additional CDD-specific models. For many gene families, multiple source databases have essentially the same family, and genes associated with RNA motifs are arbitrarily assigned to one or the other. In these cases, we have chosen one model, and mapped all genes onto that. For example, in the case of the GEMM RNA motif, we mapped genes that were classified as pfam00672 ("HAMP domain") to smart00304 (also "HAMP domain"), so that all predicted HAMP domains are shown under the same code.
- Similarly, gene names were unified. For example, in the case of SAH RNA, we changed "MTHFR" to the more standard "metF"; some genome annotations had used MTHFR, while others used metF.
- Gene names were also assigned based on the literature. We added the annotations of OmcA, OmcG, OmcH, OmcS and OmcT to GEMM-associated genes based on papers [8, 6].
- Transcriptional unit predictions are modified where there was evidence to do so.

For each motif, we decided what to show in the gene context. For motifs that strongly appear to be *cis*-regulatory elements, we show only the downstream operon. In some cases, we also show the first gene, if it points the wrong way. For other non-*cis* motifs, we typically show the first 2 genes in either direction. Operons were inferred by adding genes to an operon until the next gene is on the opposite strand, or is more than 100 nucleotides beyond the previous gene. In some cases, we overrode these predictions, either because genes in a putative operon were not conserved in related organisms, or based on literature. Our main goal in depicting the gene context is to help readers to identify possible patterns in the data, but our operon predictions must be viewed as rough.

2 Motif data and further analysis

2.1 The GEMM motif

2.1.1 Additional analysis

Intriguingly, GEMM is present in several species that can reduce metal ions [7]: in addition to *Geobacter*, *Pelobacter* and the related *Desulfuromonas acetoxidans*, GEMM is found in *Syntrophobacter fumaroxidans*, *Rhodoferax ferrireducens*, *Desulfotomaculum reducens* and *Alkaliphilus metalliredigenes*. However, clearly many unrelated bacteria also use GEMM RNA.

The GEMM RNA upstream of *omcS* is close to the gene's Shine-Dalgarno sequence and start codon. We had therefore hypothesized that this gene is controlled by sequestering of the SD sequence, and not at the transcriptional level. However, studies show that OmcS mRNAs are more abundant in certain conditions [5]. This transcriptional change may be a side effect of reduced translation, or arise from a parallel system of *OmcS* regulation.

GEMM RNA is frequently in front of two gene families that are poorly understood, at least in bacteria:

- Fibronectin type III (in some cases, these genes have regions with BLAST homology to subtilase, E-value $< 10^{-35}$; also sometimes homology to cyd-5 cytochrome c family, E-value $< 10^{-30}$)
- ErfK/YbiS/YcfS/YnhG (often fused with lysin domain, a domain mentioned in the main paper [2])

$tfoX^{\mathrm{GEMM}}$ expression is slightly increased (1.7-2.1 fold) by GlcNAc, but this was below the threshold used in the expression analysis [9].

GEMM RNA was not found upstream of cytochrome c genes in the genus *Pelobacter*, which is closely related to *Geobacter*. *P. carbinolicus* has only 14 predicted cytochrome c genes and no known homologs of OmcG, OmcH or OmcS [4].

2.1.2 Genome context and alignment

Notes:

[1] The 5′ side of a predicted rho-independent transcription terminator stem overlaps the 3′ side of the P2 stem. If GEMM is a riboswitch, and its conserved structure forms when the ligand is present (as is the case for other riboswitches), we predict that this instance is an ON switch, i.e., gene expression turns on when the ligand is present. In this case, the terminator was predicted by Rnall.

[2] The 5′ side of a predicted rho-independent transcription terminator stem

overlaps the $3'$ side of the P2 stem. If GEMM is a riboswitch, and its conserved structure forms when the ligand is present (as is the case for other riboswitches), we predict that this instance is an ON switch, i.e., gene expression turns on when the ligand is present. In this case the terminator was predicted by manual inspection, and may have an unusually high number of base pairs or bulges in the stem.

[3] The putative Shine-Dalgarno sequence is very close to the conserved sequence on the $3'$ part of the GEMM motif. If GEMM is a riboswitch, and its conserved structure forms when the ligand is present (as is the case for other riboswitches), we predict that this is an OFF switch, whereby ligand binding causes the GEMM structure to block ribosomal binding. However, in all cases, we were unable to find homologs that conserved the $5'$-most sequence, so it is possible that the start codon is merely annotated as earlier than the true start codon. On the other hand, we are encouraged by the COG1305-associated hits, which at least conserve the type of gene controlled. In all cases where GEMM RNA is upstream of COG1305, the COG1305 start codon overlaps the conserved part of the GEMM motif.

[4] The hypothetical ORF corresponding to the annotated start codon (which lacks a likely Shine-Dalgarno sequence) is not conserved in any bacteria.

[5] Overlaps gene on the opposite strand (NP_866542.1). This gene is not conserved in any bacteria. However, it may be that both gene and RNA are genuinely on opposite strands. We find the homology of the RNA compelling.

[6] Homology is unclear.

[7] No genes are annotated nearby, probably because it is an environmental sequence fragment.

[8] No downstream gene is annotated because this homolog is near the end of a sequence fragment.

[9] No gene is annotated in this sequence (the RNA occurs roughly 2 Kb from one end of the sequence fragment). However, translated BLAST shows clear homology to a gene annotated as "Methyl-accepting chemotaxis protein" (E-value 1e-58), and other related genes.

[10] Overlaps extremely short ORF, but there is no evidence from BLAST that this ORF is conserved.

[11] The extremely long P2 loop in this putative homolog has no obvious low-energy fold. However, based on its other features, we think it reasonably likely that this is a true homolog.

[12] The homology here is borderline, in terms of extra insertions in P1, and mispairings in P2. However, there is substantial basis for hypothesizing homology, and the downstream gene is signalling-related, so a plausible one for this motif. Ultimately, given previous work with cell-cell communication in *Myxococcus*, we felt that failing to mention a plausible GEMM RNA would be unwise. We also note that this putative GEMM overlaps the $5'$-end of the downstream gene, but this overlap region is not conserved in any bacteria.

[13] This RNA overlaps a putative ORF, but there is no evidence that this ORF is conserved in any bacteria.

[14] *omcG* is annotated as being ≈ 180 nucleotides from *omcH*, but the correct start codon is ≈ 60 upstream of the annotated one [6], so there is good reason to suspect that they are part of the same transcript.

[15] Coding sequence at annotated start codon is conserved; this is probably the correct start codon. However, Mfold shows no convincing stem that would sequester the putative Shine-Dalgarno sequence.

[16] Experimental and computational evidence suggest that OmcT is in the same operon as OmcS [8], so is shown here.

[17] This putative homolog overlaps the start codon of a gene (ZP_01140822.1). However, this gene is not conserved in any bacteria.

[18] Although the downstream gene is annotated as “hypothetical”, BLASTP finds strong homology to genes annotated as “type IV pilus assembly PilZ”.

[19] This gene is 1.5 Kb from the nearest annotated gene. However, using translated BLAST, we find that the intervening region has strong (E-value=0) homology to putative chitin-binding protein genes.

[20] Overlaps 5′ region of downstream gene, but the overlapped part of the gene is not conserved in any bacteria, even though the gene as a whole is conserved.

[21] Overlaps 5′ region of downstream gene, but the entire downstream gene has no clear homologs in known bacteria, so it’s unclear whether the overlap region is correct.

[22] Overlaps ≈ 300 nucleotides of downstream gene, but the overlapped part of the gene is not conserved in any bacteria, although the gene as a whole is conserved.

[23] An annotated start codon overlaps the putative transcription terminator, but (1) this start codon does not have an obvious Shine-Dalgarno, and (2) the 5′ part of the predicted coding region is not conserved in any bacteria.

[24] BLAST results support that the coding region is conserved from the annotated start codon, and therefore that the annotated start codon is correct.

abbrev. of hits	**taxonomy of species**
Sth-1-1 to Sth-1-1	Actinobacteria *Symbiobacterium thermophilum* IAM 14863
Bha-1-1 to Bha-1-1	Firmicutes Bacillales Bacillaceae *Bacillus halodurans* C-125
Bli-1-1 to Bli-1-2	Firmicutes Bacillales Bacillaceae *Bacillus licheniformis* ATCC 14580 (DSM 13)
Bsp-1-1 to Bsp-1-9	Firmicutes Bacillales Bacillaceae *Bacillus sp.* NRRL B-14911
Ban-1-1 to Ban-1-1	Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* ’Ames Ancestor’
Ban-2-1 to Ban-2-1	Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* A1055
Ban-3-1 to Ban-3-1	Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* A2012
Ban-4-1 to Ban-4-1	Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Ames
Ban-5-1 to Ban-5-1	Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Australia 94
Ban-6-1 to Ban-6-1	Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* CNEVA-9066
Ban-7-1 to Ban-7-1	Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Kruger B
Ban-8-1 to Ban-8-1	Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Sterne
Ban-9-1 to Ban-9-1	Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Vollum
Ban-10-1 to Ban-10-1	Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus anthracis str.* Western North America USA6153
Bce-1-1 to Bce-1-1	Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* ATCC 10987
Bce-2-1 to Bce-2-2	Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* ATCC 14579
Bce-3-1 to Bce-3-1	Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* E33L
Bce-4-1 to Bce-4-3	Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus* G9241
Bce-5-1 to Bce-5-1	Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus cereus subsp. cytotoxis* NVH 391-98
Bth-1-1 to Bth-1-5	Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus thuringiensis serovar israelensis* ATCC 35646
Bth-2-1 to Bth-2-1	Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus thuringiensis serovar konkukian str.* 97-27
Bwe-1-1 to Bwe-1-2	Firmicutes Bacillales Bacillaceae Bacillus Bacilluscereus group *Bacillus weihenstephanensis* KBAB4
Oih-1-1 to Oih-1-1	Firmicutes Bacillales Bacillaceae *Oceanobacillus iheyensis* HTE831
Ame-1-1 to Ame-1-11	Firmicutes Clostridia Clostridiales Clostridiaceae *Alkaliphilus metalliredigenes* QYMF
Cbe-1-1 to Cbe-1-4	Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium beijerincki* NCIMB 8052
Cdi-1-1 to Cdi-1-8	Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium difficile* QCD-32g58
Cpe-1-1 to Cpe-1-2	Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium perfringens* ATCC 13124
Cpe-2-1 to Cpe-2-1	Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium perfringens* SM101
Cpe-3-1 to Cpe-3-1	Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium perfringens str.* 13
Cph-1-1 to Cph-1-2	Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium phytofermentans* ISDg
Csp-1-1 to Csp-1-1	Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium sp.* OhILAs
Cte-1-1 to Cte-1-1	Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium tetani* E88

Cth-1-1 to Cth-1-4	Firmicutes Clostridia Clostridiales Clostridiaceae *Clostridium thermocellum* ATCC 27405
Chy-1-1 to Chy-1-7	Firmicutes Clostridia Clostridiales Peptococcaceae *Carboxydothermus hydrogenoformans* Z-2901
Dha-1-1 to Dha-1-5	Firmicutes Clostridia Clostridiales Peptococcaceae *Desulfitobacterium hafniense* DCB-2
Dha-2-1 to Dha-2-6	Firmicutes Clostridia Clostridiales Peptococcaceae *Desulfitobacterium hafniense* Y51
Dre-1-1 to Dre-1-6	Firmicutes Clostridia Clostridiales Peptococcaceae *Desulfotomaculum reducens* MI-1
Csa-1-1 to Csa-1-4	Firmicutes Clostridia Clostridiales Syntrophomonadaceae *Caldicellulosiruptor saccharolyticus* DSM 8903
Hor-1-1 to Hor-1-2	Firmicutes Clostridia Halanaerobiales Halanaerobiaceae *Halothermothrix orenii* H 168
Mth-1-1 to Mth-1-7	Firmicutes Clostridia Thermoanaerobacteriales Thermoanaerobacteriaceae Moorella group *Moorella thermoacetica* ATCC 39073
Tet-1-1 to Tet-1-1	Firmicutes Clostridia Thermoanaerobacteriales Thermoanaerobacteriaceae *Thermoanaerobacter ethanolicus* ATCC 33223
Tte-1-1 to Tte-1-1	Firmicutes Clostridia Thermoanaerobacteriales Thermoanaerobacteriaceae *Thermoanaerobacter tengcongensis* MB4
Bma-1-1 to Bma-1-1	Planctomycetes Planctomycetacia Planctomycetales Planctomycetaceae *Blastopirellula marina* DSM 3645
Rba-1-1 to Rba-1-1	Planctomycetes Planctomycetacia Planctomycetales Planctomycetaceae *Rhodopirellula baltica* SH 1
Rso-1-1 to Rso-1-1	β-proteobacteria Burkholderiales Burkholderiaceae *Ralstonia solanacearum* UW551
Rfe-1-1 to Rfe-1-1	β-proteobacteria Burkholderiales Comamonadaceae *Rhodoferax ferrireducens* T118
Tde-1-1 to Tde-1-1	β-proteobacteria Hydrogenophilales Hydrogenophilaceae *Thiobacillus denitrificans* ATCC 25259
Mfl-1-1 to Mfl-1-1	β-proteobacteria Methylophilales Methylophilaceae *Methylobacillus flagellatus* KT
Neu-1-1 to Neu-1-1	β-proteobacteria Nitrosomonadales Nitrosomonadaceae *Nitrosomonas europaea* ATCC 19718
Neu-2-1 to Neu-2-1	β-proteobacteria Nitrosomonadales Nitrosomonadaceae *Nitrosomonas eutropha* C71
Nmu-1-1 to Nmu-1-3	β-proteobacteria Nitrosomonadales Nitrosomonadaceae *Nitrosospira multiformis* ATCC 25196
Asp-1-1 to Asp-1-1	β-proteobacteria Rhodocyclales Rhodocyclaceae *Azoarcus sp.* EbN1
Bba-1-1 to Bba-1-1	δ-proteobacteria Bdellovibrionales Bdellovibrionaceae *Bdellovibrio bacteriovorus* HD100
Dde-1-1 to Dde-1-1	δ-proteobacteria Desulfovibrionales Desulfovibrionaceae *Desulfovibrio desulfuricans* G20
Dac-1-1 to Dac-1-2	δ-proteobacteria Desulfuromonadales Desulfuromonadaceae *Desulfuromonas acetoxidans* DSM 684
Gme-1-1 to Gme-1-19	δ-proteobacteria Desulfuromonadales Geobacteraceae *Geobacter metallireducens* GS-15
Gsp-1-1 to Gsp-1-7	δ-proteobacteria Desulfuromonadales Geobacteraceae *Geobacter sp.* FRC-32
Gsu-1-1 to Gsu-1-17	δ-proteobacteria Desulfuromonadales Geobacteraceae *Geobacter sulfurreducens* PCA
Gur-1-1 to Gur-1-30	δ-proteobacteria Desulfuromonadales Geobacteraceae *Geobacter uraniumreducens* Rf4
Pca-1-1 to Pca-1-9	δ-proteobacteria Desulfuromonadales Pelobacteraceae *Pelobacter carbinolicus* DSM 2380
Ppr-1-1 to Ppr-1-17	δ-proteobacteria Desulfuromonadales Pelobacteraceae *Pelobacter propionicus* DSM 2379
Ade-1-1 to Ade-1-1	δ-proteobacteria Myxococcales Cystobacterineae Myxococcaceae *Anaeromyxobacter dehalogenans* 2CP-C
Mxa-1-1 to Mxa-1-1	δ-proteobacteria Myxococcales Cystobacterineae Myxococcaceae *Myxococcus xanthus* DK 1622
Sfu-1-1 to Sfu-1-7	δ-proteobacteria Syntrophobacterales Syntrophobacteraceae *Syntrophobacter fumaroxidans* MPOB
Sde-1-1 to Sde-1-2	γ-proteobacteria Alteromonadales Alteromonadaceae *Saccharophagus degradans* 2-40
Cps-1-1 to Cps-1-3	γ-proteobacteria Alteromonadales Colwelliaceae *Colwellia psychrerythraea* 34H
Iba-1-1 to Iba-1-2	γ-proteobacteria Alteromonadales Idiomarinaceae *Idiomarina baltica* OS145
Ptu-1-1 to Ptu-1-3	γ-proteobacteria Alteromonadales Pseudoalteromonadaceae *Pseudoalteromonas tunicata* D2
Pin-1-1 to Pin-1-5	γ-proteobacteria Alteromonadales Psychromonadaceae *Psychromonas ingrahamii* 37
Sba-1-1 to Sba-1-2	γ-proteobacteria Alteromonadales Shewanellaceae *Shewanella baltica* OS155
Sba-2-1 to Sba-2-2	γ-proteobacteria Alteromonadales Shewanellaceae *Shewanella baltica* OS195
Sde-2-1 to Sde-2-1	γ-proteobacteria Alteromonadales Shewanellaceae *Shewanella denitrificans* OS217
Son-1-1 to Son-1-1	γ-proteobacteria Alteromonadales Shewanellaceae *Shewanella oneidensis* MR-1
Ssp-1-1 to Ssp-1-2	γ-proteobacteria Alteromonadales Shewanellaceae *Shewanella sp.* ANA-3
Ssp-2-1 to Ssp-2-6	γ-proteobacteria Alteromonadales Shewanellaceae *Shewanella sp.* MR-4
Ssp-3-1 to Ssp-3-3	γ-proteobacteria Alteromonadales Shewanellaceae *Shewanella sp.* MR-7
Ssp-4-1 to Ssp-4-1	γ-proteobacteria Alteromonadales Shewanellaceae *Shewanella sp.* PV-4
Ssp-5-1 to Ssp-5-1	γ-proteobacteria Alteromonadales Shewanellaceae *Shewanella sp.* W3-18-1
Osp-1-1 to Osp-1-3	γ-proteobacteria Oceanospirillales *Oceanospirillum sp.* MED92
Tcr-1-1 to Tcr-1-4	γ-proteobacteria Thiotrichales Piscirickettsiaceae *Thiomicrospira crunogena* XCL-2
Ppr-2-1 to Ppr-2-3	γ-proteobacteria Vibrionales Vibrionaceae *Photobacterium profundum* 3TCK
Ppr-3-1 to Ppr-3-3	γ-proteobacteria Vibrionales Vibrionaceae *Photobacterium profundum* SS9
Psp-1-1 to Psp-1-1	γ-proteobacteria Vibrionales Vibrionaceae *Photobacterium sp.* SKA34
Val-1-1 to Val-1-2	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio alginolyticus* 12G01
Van-1-1 to Van-1-1	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio angustum* S14
Vch-1-1 to Vch-1-2	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio cholerae* MO10

Vch-2-1 to Vch-2-2	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio cholerae* O1 biovar eltor str. N16961
Vch-3-1 to Vch-3-2	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio cholerae* O395
Vch-4-1 to Vch-4-1	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio cholerae* RC385
Vch-5-1 to Vch-5-2	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio cholerae* V51
Vch-6-1 to Vch-6-2	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio cholerae* V52
Vfi-1-1 to Vfi-1-1	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio fischeri* ES114
Vpa-1-1 to Vpa-1-2	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio parahaemolyticus* RIMD 2210633
Vsp-1-1 to Vsp-1-1	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio sp.* Ex25
Vsp-2-1 to Vsp-2-1	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio sp.* MED222
Vsp-3-1 to Vsp-3-2	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio splendidus* 12B01
Vvu-1-1 to Vvu-1-2	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio vulnificus* CMCP6
Vvu-2-1 to Vvu-2-2	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio vulnificus* YJ016
env-1 to env-11	environmental samples

abbrev	**RefSeq accession**		**5′ at**	**3′ at**	**genes**
Actinobacteria					
Sth-1-1	NC_006177.1	-	3227495	3227236	**RNA**→ FliM (COG1868)→ COG1406 (COG1406)→ FlgG (COG4786)→ FlgB (COG1815)→ FlgC (COG1558)→ FliE (COG1677)→ FliF (COG1766)→ FliG (COG1536)→ flagellar assembly protein→ FliI (COG1157)→ flagellar basal-body protein→ hypo→ Flg_hook (pfam02120)→
Firmicutes Bacillales					
Bsp-1-1	NZ_AAOX01000005.1	+	210212	210443	**RNA**→ AtoS (COG2202)RocR (COG3829)COG5001 (COG5001)→
Bce-5-1	NZ_AALL01000002.1	-	59384	59148	**RNA**→ HAMP (smart00304)MA (COG0840)→ EntF (COG1020)COG3320 (COG3320)PP-binding (pfam00550)→
Bce-1-1	NC_003909.8	+	504284	504517	**RNA**→ HAMP (smart00304)MA (COG0840)→
Bce-4-1	NZ_AAEK01000015.1	-	90981	90748	**RNA**→ HAMP (smart00304)MA (COG0840)→
Bth-2-1	NC_005957.1	+	421273	421506	**RNA**→ HAMP (smart00304)MA (COG0840)→
Ban-4-1	NC_003997.3	+	397292	397525	**RNA**→ HAMP (smart00304)MA (COG0840)→
Ban-8-1	NC_005945.1	+	397305	397538	**RNA**→ HAMP (smart00304)MA (COG0840)→
Ban-1-1	NC_007530.2	+	397292	397525	**RNA**→ HAMP (smart00304)MA (COG0840)→
Ban-3-1	NZ_AAAC02000001.1	+	914180	914413	**RNA**→ HAMP (smart00304)MA (COG0840)→
Ban-6-1	NZ_AAEN01000023.1	+	138415	138648	**RNA**→
Ban-2-1	NZ_AAEO01000030.1	+	151854	152087	**RNA**→
Ban-9-1	NZ_AAEP01000046.1	+	103744	103977	**RNA**→
Ban-7-1	NZ_AAEQ01000043.1	-	90673	90440	**RNA**→
Ban-10-1	NZ_AAER01000042.1	-	136402	136169	**RNA**→
Ban-5-1	NZ_AAES01000043.1	+	103414	103647	**RNA**→
Bce-3-1	NC_006274.1	+	414665	414898	**RNA**→ HAMP (smart00304)MA (COG0840)→
Bce-2-1	NC_004722.1	+	408156	408389	**RNA**→ HAMP (smart00304)MA (COG0840)→
Bth-1-1	NZ_AAJM01000023.1	+	10970	11203	**RNA**→ HAMP (smart00304)MA (COG0840)→ EntF (COG1020)COG3320 (COG3320)PP-binding (pfam00550)→
Bwe-1-1	NZ_AAOY01000004.1	-	100864	100631	**RNA**→ HAMP (smart00304)MA (COG0840)→
Bsp-1-2	NZ_AAOX01000005.1	-	65987	65754	**RNA**→ GGDEF (cd01949)PAC (pfam00785)→ DmpA (COG3191)→
[1]Bce-2-2	NC_004722.1	+	1036786	1037027	**RNA**→ COG4932 (COG4932)Cna_B (pfam05738)→
[1]Bth-1-2	NZ_AAJM01000012.1	-	18087	17846	**RNA**→ COG4932 (COG4932)Cna_B (pfam05738)→

[1]Bwe-1-2	NZ_AAOY01000011.1	+	28331	28570	**RNA**→ Collagen_bind (pfam05737)Cna_B (pfam05738)→
[2]Bth-1-5	NZ_AAJM01000107.1	-	5968	5731	**RNA**→ **RNA**→ **RNA**→ Cell surface protein→ Cell surface protein→ Hypothetical exported protein→ Hypothetical exported protein→
[2]Bth-1-4	NZ_AAJM01000107.1	-	5811	5574	(shown above)
[2]Bth-1-3	NZ_AAJM01000107.1	-	5655	5418	(shown above)
[2]Bce-4-3	NZ_AAEK01000003.1	-	195343	195105	**RNA**→ **RNA**→ hypothetical protein protein→ cell wall surface anchor family protein→ conserved hypothetical protein protein→ conserved hypothetical protein protein→
[2]Bce-4-2	NZ_AAEK01000003.1	-	195166	194929	(shown above)
Bsp-1-3	NZ_AAOX01000001.1	+	400824	401055	**RNA**→ hypo→
Bsp-1-4	NZ_AAOX01000001.1	+	401431	401666	**RNA**→ hypo→
Oih-1-1	NC_004193.1	-	2580327	2580094	**RNA**→ HTH_XRE (cd00093)→
Bsp-1-5	NZ_AAOX01000001.1	+	434573	434807	**RNA**→ ←LytR (COG1316)
[21]Bsp-1-6	NZ_AAOX01000001.1	+	422555	422791	**RNA**→ hypo→
Bli-1-1	NC_006270.2	+	378603	378838	**RNA**→ EntF (COG1020)PP-binding (pfam00550)Condensation (pfam00668)→
Bli-1-2	NC_006322.1	+	378413	378648	**RNA**→ EntF (COG1020)PP-binding (pfam00550)Condensation (pfam00668)→
Bsp-1-7	NZ_AAOX01000007.1	-	21294	21058	**RNA**→ hypo→
[20]Bsp-1-8	NZ_AAOX01000025.1	-	55956	55723	**RNA**→ amidase→
Bha-1-1	NC_002570.2	+	597582	597816	**RNA**→ hypo→ hypo→ MA (COG0840)→
Bsp-1-9	NZ_AAOX01000005.1	-	221613	221358	**RNA**→ hypo→
Firmicutes Clostridia					
Tet-1-1	NZ_AAKQ01000003.1	-	58667	58431	**RNA**→ hypo→
Tte-1-1	NC_003869.1	+	706140	706376	**RNA**→ CheW (COG0835)→ HD (pfam01966)→
Mth-1-1	NC_007644.1	+	774359	774594	**RNA**→ Tar (COG0840)→
Mth-1-2	NC_007644.1	+	543022	543258	**RNA**→ Copper amine oxidase-like→
[1]Mth-1-3	NC_007644.1	-	1536251	1536015	**RNA**→ Cache (pfam02743)Tar (COG0840)→
Mth-1-4	NC_007644.1	-	1595675	1595443	**RNA**→ PulF (COG1459)→
Mth-1-5	NC_007644.1	+	394734	394968	**RNA**→ HATPase_c (pfam02518)→ Transposase_8 (pfam01527)→ rve (pfam00665)→
Mth-1-6	NC_007644.1	-	2598937	2598700	**RNA**→ COG2206 (COG2206)→
[1]Mth-1-7	NC_007644.1	+	805514	805760	**RNA**→ FlgB (COG1815)→ FlgC (COG1558)→
Hor-1-2	NZ_AAOZ01000009.1	+	24934	25185	**RNA**→ **RNA**→ hypo→
[3]Hor-1-1	NZ_AAOZ01000009.1	+	25095	25337	(shown above)
Firmicutes Clostridia Clostridiales					
[3]Cph-1-1	NZ_AAQT01000008.1	-	50840	50604	**RNA**→ COG1305 (COG1305)→
Cdi-1-1	NZ_AAML02000042.1	-	9725	9483	**RNA**→ hypo→
Cpe-3-1	NC_003366.1	-	1510292	1510060	**RNA**→ hypo→
Cpe-1-1	NC_008261.1	-	1695696	1695464	**RNA**→ F5/8 type C domain protein→
Cpe-2-1	NC_008262.1	-	1445209	1444977	**RNA**→ Leucine Rich Repeat domain protein→
Cdi-1-2	NZ_AAML02000007.1	+	159539	159778	**RNA**→ FlgB (COG1815)→ FlgC (COG1558)→ FliE (COG1677)→ FliF (COG1766)→ FliG (COG1536)→ FliH (COG1317)→ FliI (COG1157)→ hypo→ Flg_hook (pfam02120)→ FlgD (pfam03963)→ FlgG (COG4786)→ FlbD (pfam06289)→ MotA (COG1291)→
Cdi-1-3	NZ_AAML02000041.1	+	5723	5961	**RNA**→ hypo→
Cdi-1-4	NZ_AAML02000042.1	+	10693	10939	**RNA**→ hypo→

Cdi-1-5	NZ_AAML02000047.1	+	1065	1303	**RNA**→ hypo→
Cdi-1-6	NZ_AAML02000055.1	-	108482	108248	**RNA**→ hypo→
Cdi-1-7	NZ_AAML02000057.1	-	5330	5096	**RNA**→ hypo→
Cpe-1-2	NC_008261.1	+	1218268	1218504	**RNA**→ discoidin domain protein→
Cth-1-1	NZ_AABG04000032.1	+	32829	33061	**RNA**→ hypo→
[3]Cte-1-1	NC_004557.1	+	477013	477237	**RNA**→ TGc (COG1305)→
Ame-1-1	NZ_AAKU01000077.1	+	12384	12618	**RNA**→ Tar (COG0840)Cache (pfam02743)→
[13]Chy-1-1	NC_007503.1	+	777679	777912	**RNA**→ hypo→
Chy-1-2	NC_007503.1	+	785822	786055	**RNA**→ MA (COG0840)→ Tautomerase (pfam01361)→ hypo→ BioB (COG0502)→
Chy-1-3	NC_007503.1	+	851296	851533	**RNA**→ COG2206 (COG2206)→ MotA (COG1291)→ MotB (COG1360)→ Tar (COG0840)→ CheW (COG0835)→ CheA (COG0643)→ CheB (COG2201)→ negative regulator of flagellin synthesis FlgM→ hypo→ FlgK (COG1256)DUF1078 (pfam06429)→
Chy-1-4	NC_007503.1	+	900392	900625	**RNA**→ COG1315 (COG1315)→ DUF517 (pfam04460)→ Tar (COG0840)Cache (pfam02743)→
[13]Chy-1-5	NC_007503.1	+	875560	875795	**RNA**→ FlgB (COG1815)→ FlgC (COG1558)→ FliE (COG1677)→ FliF (COG1766)→ FliG (COG1536)→ flagellar protein→ FliI (COG1157)→ flagellar protein→ hypo→ FlgD (COG1843)→ flagellar operon protein→ FlgG (COG4786)→
Chy-1-6	NC_007503.1	-	90445	90210	**RNA**→ COG2206 (COG2206)→ Nonheme_Ferritin (cd01055)→
Chy-1-7	NC_007503.1	+	906921	907153	**RNA**→ COG1413 (COG1413)→ MeTrc (smart00138)→ MA (COG0840)→ CheW (COG0835)CheW (COG0835)→ CheB (COG2201)→ REC (cd00156)→ CheA (COG0643)Hpt (pfam01627)→ hypo→
Cbe-1-1	NZ_AALO01000004.1	+	136897	137136	**RNA**→ WbbJ (COG0110)→ Cache (pfam02743)Tar (COG0840)→
Cbe-1-2	NZ_AALO01000002.1	+	236750	236987	**RNA**→ MotA (COG1291)→ MotB (COG1360)→
Cbe-1-3	NZ_AALO01000001.1	+	198488	198724	**RNA**→ hypo→ KAS_III (cd00830)→
Cph-1-2	NZ_AAQT01000020.1	-	55929	55692	**RNA**→ PAS (cd00130)EAL (cd01948)GGDEF (cd01949)HAMP (smart00304)→
Ame-1-2	NZ_AAKU01000027.1	-	37722	37484	**RNA**→ HAMP (smart00304)MA (COG0840)→
Ame-1-3	NZ_AAKU01000041.1	-	3355	3117	**RNA**→ MA (COG0840)→
Ame-1-4	NZ_AAKU01000109.1	-	8750	8512	**RNA**→ MA (COG0840)→
Ame-1-5	NZ_AAKU01000048.1	-	2875	2638	**RNA**→ NTP_transf_2 (pfam01909)→ HEPN (smart00748)→
Ame-1-6	NZ_AAKU01000072.1	+	19443	19680	**RNA**→ Tar (COG0840)Cache (pfam02743)→
Ame-1-7	NZ_AAKU01000018.1	-	30739	30501	**RNA**→ Tar (COG0840)→
Dha-2-1	NC_007907.1	+	4879337	4879573	**RNA**→ COG2206 (COG2206)→
Dha-1-1	NZ_AAAW04000004.1	+	324653	324889	**RNA**→ COG2206 (COG2206)→
Csa-1-1	NZ_AALW01000029.1	+	1770	2008	**RNA**→ Glyco_hydro_9 (pfam00759)CBM_3 (pfam00942)Glyco_hydro_48 (pfam02011)→
Csa-1-2	NZ_AALW01000025.1	-	10447	10209	**RNA**→ hypo→
Dre-1-1	NZ_AAOP01000011.1	+	80934	81165	**RNA**→ REC (cd00156)→
Dha-2-2	NC_007907.1	-	15633	15395	**RNA**→ ←hypo
Dha-1-2	NZ_AAAW04000002.1	-	132552	132315	**RNA**→ RpoE (COG1595)→
Dha-1-3	NZ_AAAW04000002.1	-	149642	149405	**RNA**→ ←BaeS (COG0642)
Dha-2-3	NC_007907.1	+	5437254	5437487	**RNA**→ HAMP (smart00304)Cache (pfam02743)MA (COG0840)→
Dha-1-4	NZ_AAAW04000003.1	-	899692	899459	**RNA**→ HAMP (smart00304)Cache (pfam02743)MA (COG0840)→
Csa-1-3	NZ_AALW01000051.1	+	6424	6669	**RNA**→ hypo→
Csa-1-4	NZ_AALW01000072.1	+	611	859	**RNA**→ COG2206 (COG2206)→ CheA (COG0643)→ CheW (COG0835)→ methyl-accepting chemotaxis protein→
Dha-2-4	NC_007907.1	-	3900342	3900103	**RNA**→ hypo→
Cbe-1-4	NZ_AALO01000011.1	+	48837	49071	**RNA**→ REC (cd00156)→
Dre-1-2	NZ_AAOP01000061.1	-	15039	14807	**RNA**→ PulE (COG2804)→ PilT (COG2805)→ PulF (COG1459)→ PilM (COG4972)→ hypo→ PilO (COG3167)→ hypo→ hypo→ Pilin (pfam00114)→

Dha-2-5	NC_007907.1	-	5220217	5219978	**RNA**→ hypo→
[9]Dre-1-3	NZ_AAOP01000054.1	-	2020	1789	**RNA**→
Ame-1-8	NZ_AAKU01000005.1	+	13685	13919	**RNA**→ COG5001 (COG5001)HAMP (smart00304)→
Ame-1-9	NZ_AAKU01000041.1	-	13084	12850	**RNA**→ AtoS (COG2202)COG5001 (COG5001)→
Dre-1-4	NZ_AAOP01000052.1	-	4703	4474	**RNA**→ chemotaxis sensory transducer→
Cth-1-2	NZ_AABG04000035.1	-	25415	25181	**RNA**→ **RNA**→ COG1305 (COG1305)→
[3]Cth-1-3	NZ_AABG04000035.1	-	25154	24915	(shown above)
Dre-1-5	NZ_AAOP01000044.1	+	11953	12183	**RNA**→ hypo→
Csp-1-1	NZ_AAQV01000007.1	-	78184	77955	**RNA**→ N-terminal methylation→ CpaA (COG4960)DiS_P_DiS (pfam06750)→
[1]Dre-1-6	NZ_AAOP01000033.1	+	12044	12284	**RNA**→ MA (COG0840)→
[8]Ame-1-10	NZ_AAKU01000035.1	+	35923	36135	**RNA**→
[8]Cdi-1-8	NZ_AAML02000062.1	+	200600	200719	**RNA**→
[1]Ame-1-11	NZ_AAKU01000081.1	+	17391	17635	**RNA**→ ←DUF6 (pfam00892)
Cth-1-4	NZ_AABG04000005.1	+	29518	29764	**RNA**→ ←RecO (COG1381)
Dha-2-6	NC_007907.1	-	2621237	2620994	**RNA**→ Na_H_antiport_1 (pfam06965)→
Dha-1-5	NZ_AAAW04000001.1	-	760692	760449	**RNA**→ Na_H_antiport_1 (pfam06965)→
Planctomycetes					
Bma-1-1	NZ_AANZ01000050.1	-	14174	13950	**RNA**→ Flp (COG3847)→ hypo→ CpaA (COG4960)→
[5]Rba-1-1	NC_005027.1	+	2657899	2658159	**RNA**→ hypothetical protein-transmembrane prediction→
β-proteobacteria					
Asp-1-1	NC_006513.1	+	718866	719099	**RNA**→ Peptidase_S8 (pfam00082)→
Nmu-1-1	NC_007614.1	-	2965168	2964938	**RNA**→ hypo→ hypo→ FhaC (COG2831)→
Nmu-1-2	NC_007614.1	-	2175552	2175316	**RNA**→ Peptidase_S8 (pfam00082)→
Tde-1-1	NC_007404.1	+	532900	533124	**RNA**→ CA (cd00031)Peptidase_S8 (pfam00082)→
Neu-1-1	NC_004757.1	+	1707540	1707771	**RNA**→ Peptidase_S8 (pfam00082)→
Neu-2-1	NZ_AAJE01000004.1	+	105506	105735	**RNA**→ Peptidase_S8 (pfam00082)→
Rfe-1-1	NC_007908.1	+	4400341	4400577	**RNA**→ MCPsignal (pfam00015)→
Rso-1-1	NZ_AAKL01000070.1	+	1534	1763	**RNA**→ Spherulin 4 precursor→
Mfl-1-1	NC_007947.1	-	503076	502835	**RNA**→ FhaC (COG2831)→ FhaB (COG3210)Haemagg_act (pfam05860)→ putative transmembrane protein→ hypo→ putative transmembrane protein→
Nmu-1-3	NC_007614.1	+	3131016	3131257	**RNA**→ DUF839 (pfam05787)→
δ-proteobacteria					
[13]Sfu-1-1	NZ_AAJF01000025.1	-	15131	14894	**RNA**→ hypo→
Sfu-1-2	NZ_AAJF01000012.1	-	40314	40081	**RNA**→ hypo (cons)→
[1]Sfu-1-3	NZ_AAJF01000032.1	-	3002	2779	**RNA**→ hypo→
[1]Sfu-1-4	NZ_AAJF01000007.1	-	1839	1616	**RNA**→ hypo (cons)→
[8]Sfu-1-5	NZ_AAJF01000003.1	-	516	286	**RNA**→
Bba-1-1	NC_005363.1	-	1073951	1073719	**RNA**→ ←BCAT_beta_family (cd01557)
Dde-1-1	NC_007519.1	+	841128	841346	**RNA**→ AtoC (COG2204)→
Sfu-1-6	NZ_AAJF01000089.1	+	7303	7541	**RNA**→ hypo→
Sfu-1-7	NZ_AAJF01000020.1	-	67248	67012	**RNA**→ COG1639 (COG1639)→

Ade-1-1	NC_007760.1	-	2466924	2466692	**RNA**→ COG2421 (COG2421)→ Tar (COG0840)→ CheW (COG0835)→
[12]Mxa-1-1	NC_008095.1	-	5500568	5500339	**RNA**→ TolB (COG0823)→ hypo→ hypo→ hypo→
δ-proteobacteria Desulfuromonadales					
Gsu-1-1	NC_002939.4	+	1081543	1081768	**RNA**→ hypo→
[2]Gsu-1-2	NC_002939.4	-	2225994	2225769	**RNA**→ HTH_XRE (pfam01381)→ PilM (COG4972)→ PilN (COG3166)→ PilO (COG3167)→ PilP (pfam04351)→ HofQ (COG4796)→
[1]Gsu-1-3	NC_002939.4	-	1711433	1711208	**RNA**→ hypo→
[1]Gme-1-1	NC_007517.1	-	271494	271269	**RNA**→ hypo→ COG3357 (COG3357)→
[2]Gsu-1-4	NC_002939.4	+	1922655	1922878	**RNA**→ cytochrome c family protein→
[2]Gme-1-2	NC_007517.1	+	3913064	3913289	**RNA**→ **RNA**→ ErfK_YbiS_YhnG (pfam03734)→ lipoprotein, putative→
[2,20]Gme-1-3	NC_007517.1	+	3913272	3913497	(shown above)
[2]Gur-1-2	NZ_AAON01000112.1	+	3208	3432	**RNA**→ **RNA**→ ErfK_YbiS_YhnG (pfam03734)→
[2]Gur-1-1	NZ_AAON01000112.1	+	3392	3617	(shown above)
[2]Ppr-1-1	NZ_AAJH01000005.1	-	49446	49222	**RNA**→ **RNA**→ LysM (cd00118)ErfK_YbiS_YhnG (pfam03734)FtsN (COG3087)→
[1]Ppr-1-2	NZ_AAJH01000005.1	-	49227	49002	(shown above)
[2]Gsp-1-1	NZ_AASH01000001.1	-	38765	38540	**RNA**→ ErfK_YbiS_YhnG (pfam03734)→ lipoprotein, putative→ COG3034 (COG3034)UPF0169 (pfam03696)→
[2]Gsp-1-2	NZ_AASH01000108.1	-	3993	3768	**RNA**→ hypo (cons)→
[2]Gur-1-3	NZ_AAON01000070.1	+	16014	16239	**RNA**→ hypo (cons)→
[2]Gme-1-4	NC_007517.1	+	1334239	1334463	**RNA**→ Peptidase_S8 (pfam00082)→
[1]Gur-1-4	NZ_AAON01000003.1	-	136025	135799	**RNA**→ OmpW (pfam03922)→
[1]Gur-1-5	NZ_AAON01000006.1	+	111134	111359	**RNA**→ **RNA**→ fn3 (cd00063)→
[1]Gur-1-10	NZ_AAON01000006.1	+	111360	111586	(shown above)
[2]Gsu-1-6	NC_002939.4	-	198072	197847	**RNA**→ **RNA**→ LysM (cd00118)ErfK_YbiS_YhnG (pfam03734)→ lipoprotein, putative→
[2]Gsu-1-5	NC_002939.4	-	197873	197648	(shown above)
[2]Gur-1-7	NZ_AAON01000029.1	+	31644	31869	**RNA**→ **RNA**→ possible cytochrome c→
[2]Gur-1-6	NZ_AAON01000029.1	+	31872	32097	(shown above)
[1]Gsp-1-3	NZ_AASH01000018.1	-	17206	16981	**RNA**→ HTH_3 (pfam01381)→ PilM (COG4972)→ PilN (COG3166)→ PilO (COG3167)→ PilP (pfam04351)→ HofQ (COG4796)→
[2]Gur-1-8	NZ_AAON01000042.1	-	17722	17497	**RNA**→ HTH_3 (pfam01381)→ PilM (COG4972)→ PilN (COG3166)→ PilO (COG3167)→ PilP (pfam04351)→ HofQ (COG4796)→
[2]Gur-1-9	NZ_AAON01000012.1	-	22183	21958	**RNA**→ hypo (cons)→
[2]Gsu-1-7	NC_002939.4	+	1348999	1349223	**RNA**→ hypo→
[2]Gme-1-6	NC_007517.1	+	2280692	2280917	**RNA**→ LasI (COG3916)→
[2]Gme-1-7	NC_007517.1	+	325755	325980	**RNA**→ fibronectin type III domain protein→
[1]Gur-1-11	NZ_AAON01000001.1	+	329163	329388	**RNA**→ hypo→
[2,13]Gsu-1-8	NC_002939.4	-	2133695	2133470	**RNA**→ **RNA**→ fibronectin type III domain protein→
[2]Gsu-1-15	NC_002939.4	-	2133429	2133190	(shown above)
[2]Gme-1-8	NC_007517.1	+	2447619	2447844	**RNA**→ Integrins alpha chain→
[2]Gsu-1-9	NC_002939.4	-	2136377	2136152	**RNA**→ **RNA**→ hypo→ hypo→
Gsu-1-16	NC_002939.4	-	2136167	2135927	(shown above)
[2]Gme-1-9	NC_007517.1	+	1977793	1978018	**RNA**→ hypo→
Gme-1-10	NC_007517.1	-	1324617	1324392	**RNA**→ **RNA**→ cytochrome c→
[1]Gme-1-5	NC_007517.1	-	1324478	1324253	(shown above)

[2]Gsu-1-10	NC_002939.4	-	1147037	1146813	**RNA**→ hypo→
[2]Gur-1-12	NZ_AAON01000007.1	-	118125	117901	**RNA**→ hypo (cons)→
[2]Gsp-1-4	NZ_AASH01000010.1	-	61352	61127	**RNA**→ COG2340 (COG2340)→
[1]Gme-1-11	NC_007517.1	-	1212229	1212004	**RNA**→ **RNA**→ cytochrome c family protein, putative→
[1]Gme-1-15	NC_007517.1	-	1212049	1211824	(shown above)
Gur-1-13	NZ_AAON01000001.1	-	323790	323566	**RNA**→ FN3 (cd00063)→
[2]Ppr-1-3	NZ_AAJH01000033.1	-	4348	4124	**RNA**→ **RNA**→ FN3 (cd00063)→
[22]Ppr-1-9	NZ_AAJH01000033.1	-	4090	3851	(shown above)
[2]Gme-1-12	NC_007517.1	+	158170	158394	**RNA**→ hypo→
[1]Gme-1-13	NC_007517.1	+	1911272	1911497	**RNA**→ possible cytochrome c→
[1]Gme-1-14	NC_007517.1	-	265065	264840	**RNA**→ hypo→
[1,13]Gsu-1-12	NC_002939.4	+	2773389	2773614	**RNA**→ cytochrome c family protein, putative→
[2]Gsp-1-5	NZ_AASH01000089.1	-	6415	6191	**RNA**→ hypo (cons)→
[1]Gsp-1-6	NZ_AASH01000068.1	-	14654	14430	**RNA**→ NLPC_P60 (pfam00877)→
[2]Gur-1-14	NZ_AAON01000036.1	+	34196	34421	**RNA**→ Peptidase_S8 (pfam00082)He_PIG (pfam05345)→
[2]Gme-1-16	NC_007517.1	+	1079465	1079690	**RNA**→ HTH_XRE (pfam01381)→ PilM (COG4972)→ PilN (COG3166)→ PilO (COG3167)→ PilP (pfam04351)→ HofQ (COG4796)→
[2]Gur-1-15	NZ_AAON01000052.1	-	20759	20534	**RNA**→ hypo (cons)→
[1]Gur-1-17	NZ_AAON01000016.1	-	26009	25785	**RNA**→ **RNA**→ hypo→
[1]Gur-1-16	NZ_AAON01000016.1	-	25805	25581	(shown above)
[2,14]Gsu-1-13	NC_002939.4	-	3169147	3168924	**RNA**→ **RNA**→ NHL repeat domain protein→ OmcA cytochrome c→ OmcH cytochrome c→ OmcG cytochrome c→
[2]Gsu-1-11	NC_002939.4	-	3168962	3168739	(shown above)
[1]Ppr-1-4	NZ_AAJH01000009.1	+	15928	16152	**RNA**→ lipoprotein, putative→
[2]Gur-1-18	NZ_AAON01000061.1	-	21952	21726	**RNA**→ COG4728 (COG4728)→
[1]Gur-1-19	NZ_AAON01000001.1	-	326323	326099	**RNA**→ FN3 (cd00063)→
[1]Ppr-1-5	NZ_AAJH01000009.1	+	90120	90344	**RNA**→ hypo→ hypo→ ErfK (COG1376)→
[6]Gur-1-20	NZ_AAON01000044.1	-	2314	2088	**RNA**→ fibronectin type III domain protein→
[2]Gur-1-21	NZ_AAON01000009.1	+	860	1088	**RNA**→ Rfe (COG0472)FkpA (COG0545)→ hypo→
[1,6,18]Ppr-1-6	NZ_AAJH01000007.1	+	70238	70462	**RNA**→ hypo→
[1]Pca-1-1	NC_007498.2	+	1987037	1987265	**RNA**→ PaaK (COG1541)→
[2]Gur-1-22	NZ_AAON01000045.1	+	6858	7084	**RNA**→ ElaC (COG1234)→
Gur-1-23	NZ_AAON01000001.1	+	248116	248343	**RNA**→ conserved hypothetical transmembrane protein→
[3]Gur-1-24	NZ_AAON01000012.1	-	27568	27340	**RNA**→ hypo→
[1]Ppr-1-7	NZ_AAJH01000013.1	-	36477	36247	**RNA**→ hypo→ arch_bact_SO_family_Moco (cd02109)→
[2]Gur-1-25	NZ_AAON01000009.1	-	57295	57067	**RNA**→ DUF11 (pfam01345)→
Gur-1-27	NZ_AAON01000009.1	-	58079	57851	**RNA**→ **RNA**→ hypo→
[4,17]Gur-1-26	NZ_AAON01000009.1	-	57835	57608	(shown above)
Pca-1-2	NC_007498.2	+	1774619	1774846	**RNA**→ Calx_beta (smart00237)Amb_all (smart00656)→
Pca-1-3	NC_007498.2	+	1141626	1141851	**RNA**→ **RNA**→ Tar (COG0840)Cache (pfam02743)HAMP (smart00304)→
Pca-1-4	NC_007498.2	+	1141853	1142078	(shown above)
Dac-1-1	NZ_AAEW02000001.1	-	146764	146537	**RNA**→ PAS (cd00130)GGDEF (cd01949)→
Dac-1-2	NZ_AAEW02000011.1	+	10503	10727	**RNA**→ BaeS (COG0642)→ AtoC (COG2204)→
Ppr-1-8	NZ_AAJH01000005.1	-	168865	168639	**RNA**→ COG3287 (COG3287)→

Pca-1-5	NC_007498.2	+	2380546	2380772	**RNA**→ PAS (cd00130)COG5001 (COG5001)→
Gur-1-28	NZ_AAON01000004.1	+	8881	9106	**RNA**→ hypo→
[2]Gsu-1-14	NC_002939.4	-	1100143	1099911	**RNA**→ hypo→
[2]Ppr-1-10	NZ_AAJH01000008.1	+	50456	50692	**RNA**→ FN3 (cd00063)→
Gme-1-17	NC_007517.1	-	824437	824205	**RNA**→ Peptidase_S8 (pfam00082)→
Gme-1-18	NC_007517.1	-	827292	827060	**RNA**→ Peptidase_S8 (pfam00082)→
[15,16]Gsu-1-17	NC_002939.4	-	2760060	2759831	**RNA**→ OmcS cytochrome c→ OmcT cytochrome c→
Ppr-1-11	NZ_AAJH01000006.1	-	37060	36835	**RNA**→ COG4748 (COG4748)→
[13]Ppr-1-12	NZ_AAJH01000014.1	+	103351	103576	**RNA**→ RVT (pfam00078)→ COG3440 (COG3440)→ hypo→ Hit (COG0537)→ HNH nuclease→
Ppr-1-13	NZ_AAJH01000020.1	-	5337	5113	**RNA**→ hypo→ hypo→ hypo→ KU (cd00109)→ hypo→ hypo→ CumB (COG0590)→
[13]Ppr-1-14	NZ_AAJH01000008.1	+	119644	119871	**RNA**→ ←hypo
Ppr-1-15	NZ_AAJH01000013.1	-	52189	51962	**RNA**→ DUF898 (pfam05987)→ COG4783 (COG4783)→
Pca-1-7	NC_007498.2	-	2648195	2647968	**RNA**→ LprI (COG4461)→
Gur-1-29	NZ_AAON01000062.1	-	19084	18856	**RNA**→ hypo→ COG0714 (COG0714)→ COG1721 (COG1721)→ COG1305 (COG1305)→
Ppr-1-16	NZ_AAJH01000030.1	+	12235	12457	**RNA**→ hypo (cons)→
[13]Ppr-1-17	NZ_AAJH01000006.1	+	28150	28373	**RNA**→ hypo→ COG2206 (COG2206)→
Pca-1-8	NC_007498.2	-	668521	668294	**RNA**→ CheW (COG0835)→
Pca-1-9	NC_007498.2	-	2400104	2399873	**RNA**→ **RNA**→ hypo→ hypo→ FimC (COG3121)→ FimD (COG3188)→ COG5430 (COG5430)→ hypo→
Pca-1-6	NC_007498.2	-	2399816	2399586	(shown above)
Gme-1-19	NC_007517.1	-	1623888	1623647	**RNA**→ hypo→ PstS (COG0226)→ hypo→ Peptidase_S8 (pfam00082)→
Gsp-1-7	NZ_AASH01000154.1	+	1294	1555	**RNA**→ OtsB (COG1877)→
[2,11]Gur-1-30	NZ_AAON01000055.1	-	19768	19456	**RNA**→ hypo→
γ-proteobacteria					
Osp-1-1	NZ_AAOW01000025.1	+	38985	39208	**RNA**→ ←hypo
Osp-1-2	NZ_AAOW01000010.1	-	54987	54764	**RNA**→ hypo→
Tcr-1-1	NC_007520.1	+	1048719	1048942	**RNA**→ PrfC (COG4108)→
Tcr-1-2	NC_007520.1	+	326011	326238	**RNA**→ **RNA**→ hypo→
[21]Tcr-1-3	NC_007520.1	+	326188	326416	(shown above)
Tcr-1-4	NC_007520.1	+	1051134	1051366	**RNA**→ ←MgtE (COG2239)
Osp-1-3	NZ_AAOW01000017.1	-	31244	31010	**RNA**→ hypo→
γ-proteobacteria Alteromonadales					
Ptu-1-1	NZ_AAOH01000007.1	+	198689	198925	**RNA**→ hypo (COG3397)→
Pin-1-1	NZ_AAQS01000002.1	+	382853	383077	**RNA**→ ←hypo
Pin-1-3	NZ_AAQS01000002.1	+	375851	376075	**RNA**→ **RNA**→ FimC (COG3121)→ hypo→ hypo→ hypo→
[23]Pin-1-2	NZ_AAQS01000002.1	+	376323	376547	(shown above)
Ptu-1-2	NZ_AAOH01000001.1	-	581796	581564	**RNA**→ hypo→ Cdd (COG0295)→
Cps-1-1	NC_003910.7	+	4612611	4612842	**RNA**→ ←DUF885 (pfam05960)
Ptu-1-3	NZ_AAOH01000011.1	-	15155	14922	**RNA**→ ←DUF885 (pfam05960)
Iba-1-1	NZ_AAMX01000014.1	-	21303	21068	**RNA**→ **RNA**→ hypo→ FimC (COG3121)→ FimD (COG3188)→ hypo→ hypo→
[3]Iba-1-2	NZ_AAMX01000014.1	-	21204	20963	(shown above)
Cps-1-2	NC_003910.7	-	1642525	1642281	**RNA**→ hypo→

Cps-1-3	NC_003910.7	+	807375	807608	**RNA**→ Peptidase_S8 (pfam00082)→
Pin-1-5	NZ_AAQS01000001.1	-	448948	448709	**RNA**→ **RNA**→ FN3 (cd00063)→
Pin-1-4	NZ_AAQS01000001.1	-	448943	448714	(shown above)
Sde-1-1	NC_007912.1	-	787494	787252	**RNA**→ COG2206 (COG2206)→
Sde-1-2	NC_007912.1	-	989437	989199	**RNA**→ VCBS→
γ-proteobacteria Alteromonadales Shewanellaceae					
Ssp-2-1	NC_008321.1	-	3621627	3621383	**RNA**→ COG3397 (COG3397)→
Ssp-2-2	NZ_AALX01000002.1	-	189810	189566	**RNA**→ COG3397 (COG3397)→
Ssp-3-1	NC_008322.1	+	1084468	1084712	**RNA**→ COG3397 (COG3397)→
Ssp-1-1	NZ_AALH01000037.1	-	36492	36248	**RNA**→ COG3397 (COG3397)→ chitin-binding protein, putative (COG3397)→
Son-1-1	NC_004347.1	+	1112481	1112723	**RNA**→ COG3397 (COG3397)→
Ssp-1-2	NZ_AALH01000007.1	+	38142	38379	**RNA**→ ChtBD3 (cd00036)Chitinase_N_term (cd02848)CBM_5_12 (pfam02839)Glyco_18 (smart00636)→
Sde-2-1	NC_007954.1	+	2534228	2534466	**RNA**→ ChtBD3 (cd00036)Chitinase_N_term (cd02848)CBM_5_12 (pfam02839)Glyco_18 (smart00636)→
Ssp-2-3	NC_008321.1	-	3357377	3357141	**RNA**→ Chitinase_N_term (cd02848)Glyco_18 (smart00636)CBM_5_12 (pfam02839)ChtBD3 (cd00036)→
Ssp-2-4	NZ_AALX01000009.1	-	110338	110102	**RNA**→ ChtBD3 (cd00036)Chitinase_N_term (cd02848)CBM_5_12 (pfam02839)Glyco_18 (smart00636)→
Ssp-3-2	NC_008322.1	-	3427742	3427506	**RNA**→ Chitinase_N_term (cd02848)Glyco_18 (smart00636)CBM_5_12 (pfam02839)ChtBD3 (cd00036)→
Sba-1-1	NZ_AAIO01000021.1	+	39081	39318	**RNA**→ Chitinase_N_term (cd02848)Glyco_18 (smart00636)CBM_5_12 (pfam02839)ChtBD3 (cd00036)→
Sba-2-1	NZ_AATK01000022.1	-	37835	37598	**RNA**→ ChtBD3 (cd00036)Chitinase_N_term (cd02848)CBM_5_12 (pfam02839)Glyco_18 (smart00636)→
Sba-1-2	NZ_AAIO01000008.1	+	90948	91196	**RNA**→ COG3397 (COG3397)→
Sba-2-2	NZ_AATK01000001.1	-	353245	352997	**RNA**→ COG3397 (COG3397)→
Ssp-5-1	NZ_AALN01000129.1	+	3287	3518	**RNA**→ hypo→
Ssp-2-5	NC_008321.1	-	4029417	4029179	**RNA**→ hypo→
Ssp-2-6	NZ_AALX01000001.1	-	135065	134827	**RNA**→ hypo→
Ssp-3-3	NC_008322.1	+	675036	675274	**RNA**→ hypo→
Ssp-4-1	NZ_AALS01000036.1	+	17674	17908	**RNA**→ Peptidase_M11 (pfam05548)→
γ-proteobacteria Vibrionales Vibrionaceae					
Ppr-3-1	NC_006370.1	+	2634481	2634717	**RNA**→ hypothetical ToxR-regulated lipoprotein→
Ppr-2-1	NZ_AAPH01000029.1	-	24025	23789	**RNA**→ hypothetical ToxR-regulated lipoprotein→
Vsp-3-1	NZ_AAMR01000015.1	-	100102	99864	**RNA**→ COG3397 (COG3397)→
Ppr-3-2	NC_006371.1	+	355541	355774	**RNA**→ COG3397 (COG3397)→ CybB (COG3038)→
Ppr-2-2	NZ_AAPH01000007.1	-	119243	119010	**RNA**→ CybB (COG3038)→
[24]Vch-2-1	NC_002505.1	+	1860690	1860925	**RNA**→ TfoX (COG3070)TfoX_C (pfam04994)→
Vch-1-1	NZ_AAKF01000010.1	+	66156	66391	**RNA**→ TfoX (COG3070)TfoX_C (pfam04994)→
Vch-3-1	NZ_AAKG01000001.1	+	1413803	1414038	**RNA**→ TfoX (COG3070)TfoX_C (pfam04994)→
Vch-5-1	NZ_AAKI01000205.1	+	250	485	**RNA**→ TfoX (COG3070)TfoX_C (pfam04994)→
Vch-6-1	NZ_AAKJ01000049.1	-	15093	14858	**RNA**→ TfoX (COG3070)TfoX_C (pfam04994)→
[13]Vpa-1-1	NC_004603.1	-	1078742	1078504	**RNA**→ TfoX (COG3070)TfoX_C (pfam04994)→
Vsp-1-1	NZ_AAKK01000069.1	+	16343	16581	**RNA**→ TfoX (COG3070)TfoX_C (pfam04994)→
Val-1-1	NZ_AAPS01000039.1	+	21459	21697	**RNA**→ TfoX (COG3070)TfoX_C (pfam04994)→
Vvu-2-1	NC_005139.1	+	2327906	2328143	**RNA**→ TfoX (COG3070)TfoX_C (pfam04994)→ hypo→ COG0398 (COG0398)→ hypo→
Vvu-1-1	NC_004459.1	-	2138128	2137891	**RNA**→ TfoX (COG3070)TfoX_C (pfam04994)→

Vsp-3-2	NZ_AAMR01000021.1	-	11847	11613	**RNA**→ TfoX (COG3070)TfoX_C (pfam04994)→
Vsp-2-1	NZ_AAND01000010.1	-	47376	47142	**RNA**→ TfoX (COG3070)TfoX_C (pfam04994)→
Vfi-1-1	NC_006840.1	+	1765122	1765353	**RNA**→ TfoX (COG3070)TfoX_C (pfam04994)→
Van-1-1	NZ_AAOJ01000002.1	+	319461	319700	**RNA**→ TfoX (COG3070)TfoX_C (pfam04994)→
Psp-1-1	NZ_AAOU01000011.1	-	39728	39489	**RNA**→ TfoX (COG3070)TfoX_C (pfam04994)→
Vvu-2-2	NC_005140.1	-	624584	624344	**RNA**→ COG3397 (COG3397)→
Vvu-1-2	NC_004460.1	-	45973	45733	**RNA**→ COG3397 (COG3397)→
Val-1-2	NZ_AAPS01000001.1	-	220352	220103	**RNA**→ COG3397 (COG3397)→
Vpa-1-2	NC_004605.1	+	1695926	1696179	**RNA**→ COG3397 (COG3397)→
[10]Vch-2-2	NC_002506.1	+	755256	755511	**RNA**→ gbpA (COG3397)→
Vch-1-2	NZ_AAKF01000075.1	-	10547	10292	**RNA**→ COG3397 (COG3397)→
Vch-3-2	NZ_AAKG01000002.1	-	469864	469609	**RNA**→ COG3397 (COG3397)→
Vch-6-2	NZ_AAKJ01000108.1	+	431	686	**RNA**→ COG3397 (COG3397)→
Vch-5-2	NZ_AAKI01000159.1	+	2735	2990	**RNA**→ COG3397 (COG3397)→
Vch-4-1	NZ_AAKH01000394.1	+	795	1050	**RNA**→ COG3397 (COG3397)→
Ppr-3-3	NC_006370.1	+	2427572	2427826	**RNA**→ TfoX (COG3070)TfoX_C (pfam04994)→
Ppr-2-3	NZ_AAPH01000002.1	+	334491	334746	**RNA**→ TfoX (COG3070)TfoX_C (pfam04994)→
environmental					
[19]env-1	AACY01000627.1	-	67366	67122	**RNA**→ unknown→
env-2	AACY01284074.1	-	497	253	**RNA**→
env-3	AACY01051455.1	-	8816	8574	**RNA**→ unknown→
env-4	AACY01015263.1	-	990	756	**RNA**→ unknown→
env-5	AACY01087074.1	+	8438	8674	**RNA**→ unknown→
env-6	AACY01120393.1	+	33821	34057	**RNA**→ unknown→
env-7	AACY01051687.1	-	29054	28817	**RNA**→ unknown→
[7]env-8	AACY01072018.1	+	1679	1913	**RNA**→
env-9	AACY01017015.1	-	1336	1099	**RNA**→ unknown→
env-10	AACY01032628.1	+	900	1132	**RNA**→ unknown→
[8]env-11	AACY01085957.1	-	232	1	**RNA**→

cd00031 Cadherin repeat domain; Cadherins are glycoproteins involved in Ca2+-mediated cell-cell adhesion; these domains occur as repeats in the extracellular regions which are thought to mediate cell-cell contact when bound to calcium; plays a role in cell fate, signalling, proliferation, differentiation, and migration; members include E-, N-, P-, T-, VE-,CNR-,proto-,and FAT-family cadherin, desmocollin, and desmoglein, exists as monomers or dimers (hetero- and homo-); two copies of the repeat are present here

cd00036 Chitin/cellulose binding domain.

cd00063 Fibronectin type III domain.

cd00093 Helix-turn-helix XRE-family like proteins.

cd00109 BPTI/Kunitz family of serine protease inhibitors; Structure is a disulfide rich alpha+beta fold.

cd00118 Lysin motif;

cd00130 PAS domain; PAS motifs appear in archaea, eubacteria and eukarya.

cd00156 Signal receiver domain; originally thought to be unique to bacteria (CheY, OmpR, NtrC, and PhoB), now recently identified in eukaroytes ETR1 Arabidopsis thaliana; this domain receives the signal from the sensor partner in a two-component systems; contains a phosphoacceptor site that is phosphorylated by histidine kinase homologs; usually found N-terminal to a DNA binding effector domain; forms homodimers

cd00830 Ketoacyl-acyl carrier protein synthase III (KASIII) initiates the elongation in type II fatty acid synthase systems.

cd01055 Nonheme Ferritin domain, found in Archaea and Bacteria, is a member of a broad superfamily of ferritin-like diiron-carboxylate proteins.

cd01557 BCAT_beta_family: Branched-chain aminotransferase catalyses the transamination of the branched-chain amino acids leusine, isoleucine and valine to their respective alpha-keto acids, alpha-ketoisocaproate, alpha-keto-beta-methylvalerate and alpha-ketoisovalerate.

cd01948 EAL domain.

cd01949 Diguanylate-cyclase (DGC) or GGDEF domain: Originally named after a conserved residue pattern, and initially described as domain of unknown function 1 (DUF1).

cd02109 bacterial and archael members of the sulfite oxidase (SO) family of molybdopterin binding domains.

cd02848 Chitinase N-terminus domain.

COG0110 Acetyltransferase (isoleucine patch superfamily) [General function prediction only]

COG0226 ABC-type phosphate transport system, periplasmic component [Inorganic ion transport and metabolism]

COG0295 Cytidine deaminase [Nucleotide transport and metabolism]

COG0398 Uncharacterized conserved protein [Function unknown]

COG0472 UDP-N-acetylmuramyl pentapeptide phosphotransferase/UDP-N-acetylglucosamine-1-phosphate transferase [Cell envelope biogenesis, outer membrane]

COG0502 Biotin synthase and related enzymes [Coenzyme metabolism]

COG0537 Diadenosine tetraphosphate (Ap4A) hydrolase and other HIT family hydrolases [Nucleotide transport and metabolism / Carbohydrate transport and metabolism / General function prediction only]

COG0545 FKBP-type peptidyl-prolyl cis-trans isomerases 1 [Posttranslational modification, protein turnover, chaperones]

COG0590 Cytosine/adenosine deaminases [Nucleotide transport and metabolism / Translation, ribosomal structure and biogenesis]

COG0642 Signal transduction histidine kinase [Signal transduction mechanisms]

COG0643 Chemotaxis protein histidine kinase and related kinases [Cell motility and secretion / Signal transduction mechanisms]

COG0714 MoxR-like ATPases [General function prediction only]

COG0823 Periplasmic component of the Tol biopolymer transport system [Intracellular trafficking and secretion]

COG0835 CheW, a small regulator protein, unique to the chemotaxis signalling in prokaryotes and archea.

COG0840 Methyl-accepting chemotaxis-like domains (chemotaxis sensory transducer).

COG1020 Non-ribosomal peptide synthetase modules and related proteins [Secondary metabolites biosynthesis, transport, and catabolism]

COG1157 Flagellar biosynthesis/type III secretory pathway ATPase [Cell motility and secretion / Intracellular trafficking and secretion]

COG1234 Metal-dependent hydrolases of the beta-lactamase superfamily III [General function prediction only]

COG1256 Flagellar hook-associated protein [Cell motility and secretion]

COG1291 Flagellar motor component [Cell motility and secretion]

COG1305 Transglutaminase-like enzymes, putative cysteine proteases [Amino acid transport and metabolism]

COG1315 Predicted polymerase, most proteins contain PALM domain, HD hydrolase domain and Zn-ribbon domain [DNA replication, recombination, and repair]

COG1316 Transcriptional regulator [Transcription]

COG1317 Flagellar biosynthesis/type III secretory pathway protein [Cell motility and secretion / Intracellular trafficking and secretion]

COG1360 Flagellar motor protein [Cell motility and secretion]

COG1376 Uncharacterized protein conserved in bacteria [Function unknown]

COG1381 Recombinational DNA repair protein (RecF pathway) [DNA replication, recombination, and repair]

COG1406 Predicted inhibitor of MCP methylation, homolog of CheC [Cell motility and secretion]

COG1413 FOG: HEAT repeat [Energy production and conversion]

COG1459 Type II secretory pathway, component PulF [Cell motility and secretion / Intracellular trafficking and secretion]

COG1536 Flagellar motor switch protein [Cell motility and secretion]

COG1541 Coenzyme F390 synthetase [Coenzyme metabolism]

COG1558 Flagellar basal body rod protein [Cell motility and secretion]

COG1595 DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog [Transcription]

COG1639 Predicted signal transduction protein [Signal transduction mechanisms]

COG1677 Flagellar hook-basal body protein [Cell motility and secretion / Intracellular trafficking and secretion]

COG1721 Uncharacterized conserved protein (some members contain a von Willebrand factor type A (vWA) domain) [General function prediction only]

COG1766 Flagellar biosynthesis/type III secretory pathway lipoprotein [Cell motility and secretion / Intracellular trafficking and secretion]

COG1815 Flagellar basal body protein [Cell motility and secretion]

COG1843 Flagellar hook capping protein [Cell motility and secretion]

COG1868 Flagellar motor switch protein [Cell motility and secretion]

COG1877 Trehalose-6-phosphatase [Carbohydrate transport and metabolism]

COG2201 Chemotaxis response regulator containing a CheY-like receiver domain and a methylesterase domain [Cell motility and secretion / Signal transduction mechanisms]

COG2202 FOG: PAS/PAC domain [Signal transduction mechanisms]

COG2204 Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains [Signal transduction mechanisms]
COG2206 HD-GYP domain [Signal transduction mechanisms]
COG2239 Mg/Co/Ni transporter MgtE (contains CBS domain) [Inorganic ion transport and metabolism]
COG2340 Uncharacterized protein with SCP/PR1 domains [Function unknown]
COG2421 Predicted acetamidase/formamidase [Energy production and conversion]
COG2804 Type II secretory pathway, ATPase PulE/Tfp pilus assembly pathway, ATPase PilB [Cell motility and secretion / Intracellular trafficking and secretion]
COG2805 Tfp pilus assembly protein, pilus retraction ATPase PilT [Cell motility and secretion / Intracellular trafficking and secretion]
COG2831 Hemolysin activation/secretion protein [Intracellular trafficking and secretion]
COG3034 Uncharacterized protein conserved in bacteria [Function unknown]
COG3038 Cytochrome B561 [Energy production and conversion]
COG3070 Regulator of competence-specific genes [Transcription]
COG3087 Cell division protein [Cell division and chromosome partitioning]
COG3121 P pilus assembly protein, chaperone PapD [Cell motility and secretion / Intracellular trafficking and secretion]
COG3166 Tfp pilus assembly protein PilN [Cell motility and secretion / Intracellular trafficking and secretion]
COG3167 Tfp pilus assembly protein PilO [Cell motility and secretion / Intracellular trafficking and secretion]
COG3188 P pilus assembly protein, porin PapC [Cell motility and secretion / Intracellular trafficking and secretion]
COG3191 L-aminopeptidase/D-esterase [Amino acid transport and metabolism / Secondary metabolites biosynthesis, transport, and catabolism]
COG3210 Large exoproteins involved in heme utilization or adhesion [Intracellular trafficking and secretion]
COG3287 Uncharacterized conserved protein [Function unknown]
COG3320 Putative dehydrogenase domain of multifunctional non-ribosomal peptide synthetases and related enzymes [Secondary metabolites biosynthesis, transport, and catabolism]
COG3357 Predicted transcriptional regulator containing an HTH domain fused to a Zn-ribbon [Transcription]
COG3397 Uncharacterized protein conserved in bacteria [Function unknown]
COG3440 Predicted restriction endonuclease [Defense mechanisms]
COG3829 Transcriptional regulator containing PAS, AAA-type ATPase, and DNA-binding domains [Transcription / Signal transduction mechanisms]
COG3847 Flp pilus assembly protein, pilin Flp [Intracellular trafficking and secretion]
COG3916 N-acyl-L-homoserine lactone synthetase [Signal transduction mechanisms / Secondary metabolites biosynthesis, transport, and catabolism]
COG4108 Peptide chain release factor RF-3 [Translation, ribosomal structure and biogenesis]
COG4461 Uncharacterized protein conserved in bacteria, putative lipoprotein [Function unknown]
COG4728 Uncharacterized protein conserved in bacteria [Function unknown]
COG4748 Uncharacterized conserved protein [Function unknown]
COG4783 Putative Zn-dependent protease, contains TPR repeats [General function prediction only]
COG4786 Flagellar basal body rod protein [Cell motility and secretion]
COG4796 Type II secretory pathway, component HofQ [Intracellular trafficking and secretion]
COG4932 Predicted outer membrane protein [Cell envelope biogenesis, outer membrane]
COG4960 Flp pilus assembly protein, protease CpaA [Posttranslational modification, protein turnover, chaperones / Intracellular trafficking and secretion]
COG4972 Tfp pilus assembly protein, ATPase PilM [Cell motility and secretion / Intracellular trafficking and secretion]
COG5001 Predicted signal transduction protein containing a membrane domain, an EAL and a GGDEF domain [Signal transduction mechanisms]
COG5430 Uncharacterized secreted protein [Function unknown]
pfam00015 Methyl-accepting chemotaxis protein (MCP) signaling domain.
pfam00078 Reverse transcriptase (RNA-dependent DNA polymerase).
pfam00082 Subtilase family.
pfam00114 Pilin (bacterial filament).
pfam00550 Phosphopantetheine attachment site.
pfam00665 Integrase core domain.
pfam00668 Condensation domain.
pfam00759 Glycosyl hydrolase family 9.
pfam00785 PAC motif.
pfam00877 NlpC/P60 family.
pfam00892 Integral membrane protein DUF6.
pfam00942 Cellulose binding domain.
pfam01345 Domain of unknown function DUF11.
pfam01361 Tautomerase enzyme.
pfam01381 Helix-turn-helix XRE-family like proteins ;
pfam01527 Transposase.
pfam01627 Hpt domain.
pfam01909 Nucleotidyltransferase domain.
pfam01966 HD domain.
pfam02011 Glycosyl hydrolase family 48.

pfam02120	Flagellar hook-length control protein.
pfam02518	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase.
pfam02743	Cache domain.
pfam02839	Carbohydrate binding domain.
pfam03696	Uncharacterized protein family (UPF0169).
pfam03734	ErfK/YbiS/YcfS/YnhG.
pfam03922	OmpW family.
pfam03963	Flagellar hook capping protein.
pfam04351	Pilus assembly protein, PilQ.
pfam04460	Protein of unknown function (DUF517).
pfam04994	TfoX C-terminal domain.
pfam05345	Putative Ig domain.
pfam05548	Gametolysin peptidase M11.
pfam05737	Collagen binding domain.
pfam05738	Cna protein B-type domain.
pfam05787	Bacterial protein of unknown function (DUF839).
pfam05860	haemagglutination activity domain.
pfam05960	Bacterial protein of unknown function (DUF885).
pfam05987	Bacterial protein of unknown function (DUF898).
pfam06289	Flagellar protein (FlbD).
pfam06429	Domain of unknown function (DUF1078).
pfam06750	Bacterial Peptidase A24 N-terminal domain.
pfam06965	Na+/H+ antiporter 1.
smart00138	Methyltransferase, chemotaxis proteins ; Methylates methyl-accepting chemotaxis proteins to form gamma-glutamyl methyl ester residues.
smart00237	Domains in Na-Ca exchangers and integrin-beta4; Domain in Na-Ca exchangers and integrin subunit beta4 (and some cyanobacterial proteins)
smart00304	HAMP domain.
smart00636	Glyco_18 domain.
smart00656	Amb_all domain.
smart00748	Higher Eukarytoes and Prokaryotes Nucleotide-binding domain;

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Ban-1-1, Ban-10-1, Ban-2-1, Ban-3-1, Ban-5-1, Ban-6-1, Ban-7-1, Ban-8-1, Ban-9-1, Bli-1-2, Cpe-1-1, Dha-1-5, Ssp-2-2, Ssp-2-4, Ssp-2-6, Vch-1-1, Vch-1-2, Vch-5-1, Vch-6-1, Vch-6-2, Vvu-1-1

.

Text like this represents the stem of a rho-independent transcription terminator that was predicted by Rnall.

Text like this represents the stem of a rho-independent transcription terminator that was predicted by manual inspection. Note that many of these have features that are not typical of these terminators, primarily long stems and many bulges.

Text like this represents start codons annotated from RefSeq genome annotations, and Shine-Dalgarno sequences annotated by manual inspection based on the position of the start codon.

alignment positions 1 · · · 160

```
Sth-1-1   CAGCGAUAAAGGCACACC..CG.CCGAAA...GAC..GG.G.GCCGC.AAAGCCA..C.GGGGCUACAGGAA........GCGGG..........CGCCGCC(11 nucs)..........CCC.GC.CUCCCAUGCU....AGCCG.GGCUGCCGCUCG
Bsp-1-1   CUAUAUAAGUGGCAAAUU..UU.AUGAAA...GUA..AA.A.GACGC.AAAGCUG..C.GGGCCUAAGGC.........................UUUU...( 0 nucs)...................GCUAUGGC....AGCCG.GGCUGCUUGCUA
Bce-5-1   GGGCGAGAAAGGCAAACU..CG.UGGAAA...CAU..AA.G.GCUGC.AAAACUG..UAGGGCUUAAGGUC........................AAGAGAA( 0 nucs)..................GGCGA.AGC...UAGUC..AGUUACCGAAUA
Bce-1-1   CAGCGAGAAAGGCAAACU..GG.UGGAAA...CAU..UA.G.GACGC.AAAACUA..CAGGAGCUAAGGCC........................GAAA...( 0 nucs)..................GGCUACGCU....AGCC..AGUUACCGGACG
Bce-4-1   CAGCGAGAAAGGCAAACU..GA.UGGAAA...CAU..UA.G.GACGC.AAAACUA..CAGGAGCUAAGGCC........................GAGA...( 0 nucs)..................GGCUACGCU....AGCC..AGUUACCGGACG
Bth-2-1   CAGCGAGAAAGGCAAACU..GA.UGGAAA...CAU..UA.G.GACGC.AAAACUA..CAGGAGCUAAGGCC........................GAAA...( 0 nucs)..................GGCUACGCU....AGCC..AGUUACCGGACG
Ban-4-1   CAGCGAGAAAGGCAAACU..GA.UGGAAA...CAU..UA.G.GACGC.AAAACUA..CAGGAGCUAAGGCC........................GAAA...( 0 nucs)..................GGCUACGCU....AGCC..AGUUACCGGACG
Bce-3-1   CAGCGAGAAAGGCAAACU..GA.UGGAAA...CAU..UA.G.GACGC.AAAACUA..CAGGAGCUAAGGCC........................GAAA...( 0 nucs)..................GGCUACGCU....AGCC..AGUUACCGGACG
Bce-2-1   CAGCGAGAAAGGCAAACU..GA.UGGAAA...CAU..GA.G.GACGC.AAAACUA..CAGGAGCUAAGGUC........................GAAA...( 0 nucs)..................GGCUACGCU....AGCC..AGUUACCGGACG
Bth-1-1   CAGCGAGAAAGGCAAACU..GA.UGGAAA...CAU..GA.G.GACGC.AAAACUA..CAGGAGCUAAGGUC........................GAAA...( 0 nucs)..................GGCUAUGCU....AGCC..AGUUACCGGACG
Bwe-1-1   CAGCGAGAAAGGCAAACU..UA.UGGAAA...CAU..GA.G.GACGC.AAAACUA..CAGGAGCUAAGGUC........................GAAG...( 0 nucs)..................GGCUAAGCU....AGCC..AGUUACCGGACG
Bsp-1-2   UAAAUAUACAGGCAAAUU..CA.UCGAAA...GGU..GA.A.GACGC.AAAACUA..UAGGGGCUAAAGCC........................AAAA...( 0 nucs)..................GGCUAUGUC....UGCC..AGUUACCGAUUC

          ...........<<...<<..<<.<<.......>>>..>>.>....>>...<<<<...<.<<<<<....<<<<........<<<<<<<<<<<<<<<.......         >>>>>>>>>>>>>.>>.>>>>...>>>......>>>.>>>>........
          ...........22...22..22.22.......222..22.2....22...??22...?.22?22....222?........???2??2?2?21000.......         00012?2?2??2?.??.?222...22?......22?.22??........
          ●●●●●●●AR●R●YAAACY--●●-●YG●RA---R●●--●G-G-GRCG●-AAAGCY●--Y-RGGYC○○○○●●○○-----------------------●●●○---         -----------------○○●●○○○GRY----RGCYR-GGYUGCCRRR●●
                                   GNRA                                   5-TR                                                                   3TR
```

```
 1 Bce-2-2   AGAUAUUUUUAGCACACU..AU.UCGAAA...GGA..UA.G.GCCGC.AAAGCUU..A.GAGUCUACGGUAA........U..............ACAUAUU( 1 nucs)...............G.UUACUAAGAU....CGUCU.GGUUGCACAUUU
 1 Bth-1-2   AGAUAUUUUUAGCACACU..AU.UCGAAA...GGA..UA.G.GCCGC.AAAGCUU..A.GAGUCUACGGUAA........U..............ACAUAUU( 1 nucs)...............G.UUACUAAGAU....CGUCU.GGUUGCACAUUU
 1 Bwe-1-2   AAAUAUUUUUAGCACACU..AU.UCGAAA...GGA..UA.G.GCCGC.AAAGCUU..A.GAGUCUACGGUAA........U..............AUAUUG.( 0 nucs)...............G.UUACUAAGAU....CGUCU.GGUUGCACAUUU
 2 Bce-4-2   GAAAUAAAUCGGCACACU..AU.UCGAAA...GGA..UA.G.GUCGC.AAAGCUA..A.GAGUCUAAGGUAA.......................UACA...( 0 nucs)..............UA.UUACUAUGAU....AGUCU.GGUUGCAAUUUU
 2 Bth-1-3   AAAAUAAAUAGGCACACU..GU.UCGAAA...GGA..UA.G.GUCGC.AAAGCUA..A.GAGUCUAAGGUAA........U..............GAAA...( 0 nucs)...............A.UUACUAUGAU....AGUCU.GGUUGCAGUUUG
 2 Bth-1-4   AAAAUAAAUGGGCACACU..GU.UCGAAA...GAA..UA.G.GUCGC.AAAGCUA..A.GAGUCUAAGGUAA........U..............GAAA...( 0 nucs)...............A.UUACUAUGAU....AGUCU.GGUUGCGGUUUG
 2 Bth-1-5   UCCUAUAUUUGGCACACU..AU.UCGAAA...GGA..UA.G.GUCGC.AAAGCUA..A.GAGUCUAAAGUAA........U..............GAAA...( 0 nucs)...............A.UUACUAUGAU....AGUCU.GGUUGCAGUUUG
 2 Bce-4-3   AAAUGUAAUUGGCACACU..AU.UCGAAA...GGA..UA.G.GUCGC.AAAGCUU..AAGAGUCUAAGGUAA........U..............ACAU...( 0 nucs)...............A.UUACUAUGAU....AGUCU.GGUUGCAGUUUG
   Bsp-1-3   UAUUGAAAAAGGCAAACU..UA.UUGAAA...AAU..AA.G.GACGC.AAAGCCG..C.GAGUCUAAGGC.........................GUAA...( 0 nucs)...................GCUAUGAU....AGUCG.GGCUGCCAGGGU
   Bsp-1-4   UUCUGUGAAACGCAAACU..UA.UUGAAA...GAU..AA.G.GGCGC.AAAGCCU..C.AGGUCUAAGGCCA.......................UGUA...( 0 nucs).................UGGUUAAGAC....GGCUG.GGCUGCCGCAAU
   Oih-1-1   CUCAGAAAAAGGCAAACC..UA.UUGAAA...GAU..GG.G.GACGC.AAAGUCA..C.AGAUCUAAGGUA........................UUUU...( 0 nucs)..................UACUAAGAU....GGCUG.GACUAUCUGGAU
   Bsp-1-5   AAAUACCAAAGGCAAAAU..CA.UUGAAA...AAU..GA.U.GACGC.AAAGCUA..AAGGGGCCUCUAUUC.......................AGUC...( 0 nucs).................GA.UAAGGCC....AGCC..AGCUGCCAUCAC
21 Bsp-1-6   GAAUAUCAAGGGCAAAAU..CA.CUGAAA...AGU..GA.U.GACGC.AAAGCUA..UAGGGGCUUAUGUGG.......................AAAAA..( 0 nucs).................AAACAAUGCC....AGCC..AGCUGCCAUCAA
   Bli-1-1   AGCAGAUAAAGGCAAACC..UG.CGGAAA...CGC..AG.G.GACGC.AAAGCCA..U.G.GCCUAAGGUGC.......................UGACG..( 0 nucs).................GUGCUACGGU....UGACA.GGUUGCCGAAUA
   Bsp-1-7   UCUUACUAAGGGCAAAUU..UU.ACGAAA...GUA..AA.A.GACGC.AAAGCCA..C.GGGCCUAAAACAA.......................GUUAU..( 0 nucs).................UUGUCAUGGC....AGCCG.GGCCGCCAAGAA
20 Bsp-1-8   UACUGAUAAAGGCAAACC..UG.UCGAAA...GGC..AG.G.GACGC.AAAGCUA..U.GGGCCUU.CCCGC.......................AUCA...( 0 nucs).................GCGG.UUGGC....UGCCA.GGCUGCCGGGCU
   Bha-1-1   CCUUGAAAAAGGCGAGUG..UU.UG.AAA...UGG..AA.A.GACGC.AAAGCUG..C.GAGUCUGAAAUCC.......................UUCU...( 0 nucs).................GAAUAGGGAU....AGUCG.AGUUGUCAAAAA
   Bsp-1-9   AGCAGAAAAAGGCAAACC..UG.UUGAAA...GAU..AG.G.GACGC.AAAGUCA..C.AGGUCUACCGCGC........CU.............AUUGAGA(13 nucs)..............AG.GCGCCAGGAC....GGCUG.GAUUGCCUGGAA
   Tet-1-1   UGUUGAAAAAGGCAAACU..CG.UCGAAA...GGC..GG.G.GACGC.AAAGCCA..C.GGGUCUAAAGCGA.......................UAGUU..( 0 nucs).................UUGCUAUGAU....AGCCG.GGCUGCCAACAU
   Tte-1-1   CUUCGAUAAGGGCAAACC..CG.UCGAAA...GGC..GG.G.GGCGC.AAAGCCA..C.GGGCCUAAAUCCC.......................UGAUU..( 0 nucs).................GGGACAUGGU....AGCCG.GGCUGCCGAAGG
   Mth-1-1   UCCCCGAAAGGGCAAACC..UG.CUGAAA...GGC..AG.G.GGCGC.AAAACUA..GAGGGUCUACAGCCU.......................UCAC...( 0 nucs).................GGGCCAUGAC....AGCC..AGUUACCACGGG
   Mth-1-2   AACCGCUAAAGGCGAACC..UGACCGAAA...GGC..GG.G.GGCGC.AAAGCCA..C.GGGCCUAAAGCCU.......................CAAC...( 0 nucs).................AGGCCAUGGC....AGCCG.GGUUGCCGGAUA
 1 Mth-1-3   AUCAACAAAGGGCAAACC..UG.UCGAAA...GGC..AG.G.GACGC.AAAGCCA..U.GGGCCUAAAGCCC.......................GACAA..( 0 nucs).................GGGCUAUGGU....UGCCA.GGUUGCCGGUAG
   Mth-1-4   CCUUGUUAAGGGCAAACC..UG.GUGAAA...GCC..AG.G.GGCGC.AAAGCCA..C.GGGCCUGCAGAC........................GCAA...( 0 nucs)..................GUC.AUGGC....CGCCG.GGCUGCCGGGGA
   Mth-1-5   UAGUGUAAAAGGCAAACC..AC.CUGAAA...GGG..UG.G.GACGC.AAAGCCA..U.GGGUCUACAGCCG.......................UUAC...( 0 nucs).................CGGU.AUGAC....UGCCA.GGCUGCCGCUAG
 3 Hor-1-1   CUCCGGAAAUGGUAAACC..CA.UCGAAA...GGU..GG.G.GACAC.AAAGUCA..C.GGGCCUGUAGCCG........GUUA...........UUUU...( 0 nucs)...........UA.AC.CGGU.AUGGC....AGCCG.GACUGCCGAAGG
   Mth-1-6   UGCUCAAAAGGGUAAACC..CA.UCGAAA...GGU..GG.G.GACAC.AAAGCCA..U.GGGUCUAAGGUCC........U..............UAAA...( 0 nucs)...............A.GGGCUAUGAU....CGCCA.GGCUGCCGGCUC
 1 Mth-1-7   GGGCGUCAAAGGCAAGCU..CU.CCGAAA...GGG..GAUG.CGCGC.AAAGCCA..U.GGGUCUAAGGCCU........GGU............AGAUACA( 1 nucs)............A.CC.GGGCUAUGGC....UGCCA.GGCUACCGUAAC
   Hor-1-2   GACCGAUAAAGGCAAACC..UG.UCGAAA...GGC..AG.G.GACGC.AAAACCA..C.GGGCCUGUAGCCU........UAUAUAA........AAAAAAC( 0 nucs)........UUAUA.UA.AGGU.AUGGC....AGCCG.GGUUACCGAAGG
 3 Cph-1-1   CAACGAUAAUAGCAAACC..UA.GUGAAA...ACU..AG.C.GACGC.AAAACUA..UAGGGUCUUCCUUAG........A..............UAU....( 0 nucs)...............U.CUAAGAUGAU....AGCC..AGUUACCGAAAG
   Cdi-1-1   UUCAGAUAUAGGCAAUAU..UA.GAUAAA...UCU..AA.U.GACGC.AAAACUA..UAGGGACUAAAAUCU........UUAU...........GUA....( 0 nucs)...........AU.AA.AGGUCAUGUC....AGCC..AGUUGCCAAAAA
   Cpe-3-1   UAAAAAAAUGGGCAAAAU..UA.GAGAAA...UCU..AA.U.GACGC.AAAGCUA..UAGGGACUAAGGUU........................UAU....( 0 nucs)..................AACUAUGUC....AGCC..AGUUGCCAAAGA
   Cpe-2-1   UUUAAAAUUCGGCAAAAU..UA.GAGAAA...UCU..AA.U.GACGC.AAAGCUA..UAGGGACUAAGGUU........................UAU....( 0 nucs)..................AACUAUGUC....AGCC..AGUUGCCAAAGA
   Cdi-1-2   AAACUAAAUAGGCAAAUC..UA.GAGAAA...UCU..AG.U.GACGC.AAAGCUA..UAGGGGCUAAGGUUU........GU.............AUAC...( 0 nucs)..............AC.AAACUAUGUC....AGCC..AGUUGCCAAAAA
   Cdi-1-3   AAACUAAAUAGGCAAAAC..UA.GAGAAA...UUU..AG.U.GACGC.AAAGCUA..UAGGGACUAAGACUU........AU.............GAA....( 0 nucs)..............AU.GAGUUAUGUC....AGCC..AGUUGCCAAAGA
   Cdi-1-4   AAACUAAAUCGGCAAAAC..UA.GAGAAA...UUU..AG.U.GACGC.AAAGCUA..UAGGGACUAAGACUU........AUA............UAAAUAU( 2 nucs)............U.AU.GAGUUAUGUC....AGCC..AGUUGCCAAAGA
   Cdi-1-5   AAACUAAAUCGGCAAAAC..UA.GAGAAA...UUU..AG.U.GACGC.AAAGCUA..UAGGGACUAACGCUU........AU.............AAA....( 0 nucs)..............AU.AAGUUAUGUC....AGCC..AGUUGCCAAAAA
   Cdi-1-6   AAAAUUAAAUGGCAAAAU..UA.GGGAAA...UCU..AA.U.GACGC.AAAACUA..UAGGGAUUAAAGCUU.......................UUA....( 0 nucs).................AAGCCAUAUC....AGCC..AGUUGCUAAAAA
   Cdi-1-7   AAUUAAAUGAGGCAAAAC..UA.GGGAAA...CCU..AG.U.CACGC.AAAGCUA..UAGGGGCUAAAGUUU.......................AGU....( 0 nucs).................AAACCAUGCC....AGCC..AGUUGCCGAAAA
   Cpe-1-2   AUAAUAUAUUGGCAAACC..UA.GGGAAA...UCU..AG.A.GACGC.AAAGCUA..AAGGGGCUAAGGUUA.......................UAUG...( 0 nucs)...............A.UAACUAUGCU....AGCC..AGUUGCCGAAGA
   Cth-1-1   ACAUUACAAAGGCAAACC..CA.UCAAAA...GAU..GG.C.GACGC.AAAGCUA..AAGGGGCUAAAGCAU.......................U......( 0 nucs).................AUGCCAUGCC....AGCC..AGUUGCAAUGCA
 3 Cte-1-1   CCACGAUAAAGGCAAAAC..UA.UUGAAA...GAU..AG.U.GACGC.AAAGCUA..UAGGGUCUGUA...........................AA.....( 0 nucs)......................AUGAC....AGCC..AGUUAUCGAAAG
   Ame-1-1   GAAUACAGAGGGCAAACU..UA.UCGAAA...GGU..AA.G.GACGC.AAAGCCA..U.GGGCCUACAUUCU.......................AAAC...( 0 nucs).................AGAAGAUGGU....AGCC..GGUUGCAGAACU
13 Chy-1-1   UACCGAUAAAGGCAAACC..CG.GCGAAA...GCC..GG.C.GACGC.AAAGCUA..CAGGGGCUAAGGCC........................GUAA...( 0 nucs)..................GGCUAUGCC....AGCC..AGCUGCCGAAGG
   Chy-1-2   CUUUGUCAAUAGCAAACC..CG.GCGAAA...GCC..GG.C.GACGC.AAAGCUA..CAGGGGCUAAAGCC........................GUAA...( 0 nucs)..................GGCUAUGCC....AGCC..AGCUGCCGAAGG
   Chy-1-3   UUCCGAAAAAGGCAAACC..CG.GCGAAA...GCC..GG.G.GACGC.AAAGCUA..CAGGGGCUAAAGCAG........G..............UUGU...( 0 nucs)...............C.CUGCUAUGCC....AGCC..AGCUGCCGAAGG
   Chy-1-4   AAACGUUAAGGGCAAACC..CG.GCGAAA...GCC..GG.C.GACGC.AAAGCUA..CAGGGGCUAAAGCC........................GUAA...( 0 nucs)..................GGCUAUGCC....AGCC..AGCUGCCAAAGG
13 Chy-1-5   CGAAAAGUUUGGCAAAAC..CG.GCGAAA...GCC..GG.C.GACGC.AAAGCUA..AAGGGGCUAAGGCGG.......................GUAA...( 0 nucs).................CCGCUAUGCC....AGCC..AGCUGCCGGGGG
   Chy-1-6   UGUCGAUAAAGGCAAAAC..CG.GCGAAA...GCC..GG.U.GACGC.AAAGCUA..CAGGGGCUAAAGCGA.......................AUUU...( 0 nucs).................UCGCCAUGCC....AGCC..AGCUGCCGAAGG
   Chy-1-7   CUUUGUCAAAGGCAAACC..CG.GUGAAA...GCC..GG.G.GGCGC.AAAGCUA..CAGGGGCUAAAGCUU.......................A......( 0 nucs).................AAGCUAAGCC....AGCC..AGCUGCCAAAGG

             ...........<<...<<..<<.<<.......>>>..>>.>....>>...<<<<...<.<<<<<....<<<<........<<<<<<<<<<<<<<<.......         >>>>>>>>>>>>>.>>.>>>>...>>>......>>>.>>>>........
             ...........22...22..22.22.......222..22.2....22...??22...?.22?22....222?........???2??2?2?21000.......         00012?2?2??2?.??.?222...22?......22?.22??........
             •••••••AR•R•YAAACY--••-•YG•RA---R••--•G-G-GRCG•-AAAGCY•--Y-RGGYC○○○○•●○○------------------------••●○---         -----------------○○●•○○○GRY----RGCYR-GGYUGCCRRR••
                                         GNRA                                  5-TR                                                                     3TR
```

```
Cbe-1-1    AAUAUCAAUAGGCACACU..UA.UUGAAA...AAU..AG.G.GUCGC.AAAGCUA..U.GAGUCUAAGGAAA........A..............AUUUAA.( 0 nucs)...............U.UUUCUAUGAU....UGUCA.GGUUGCCAAAAU
Cbe-1-2    AAAUGUUACUAGCACACU..UA.UUGAAA...GAU..AA.G.GCCGC.AAAGCCA..U.GAGUCUAAGGGAA........U..............UUAC...( 0 nucs)...............A.UUUCUAUGAU....UGUCA.GGUUGCCAAAAU
Cbe-1-3    UAUUGAUAAUAGCACACU..UG.CCGAAA...GGU..AG.G.GCCGC.AAAGCUA..U.GGGUCUAAAGAAA.......................UUCA...( 0 nucs)...............U.UUUCAAUGAU....UGCCA.GGUUGCCAAAAU
Cph-1-2    UUAUAAAUUUAGCAAACU..AA.UUGAAA...GAU..UA.G.GACGC.AAAGCUA..U.GGGUCUAAGGAUA........UC.............AUAA...( 0 nucs).................UAUCAACGAU....UGCCA.GGUUACAGAAAA
Ame-1-2    UGUUGAUAAAAGCAAACU..UA.UUGAAA...AAU..AA.G.GGCGC.AAAGCUA..U.GGGUCUAAGGGAA.......................GACACUA( 0 nucs).................UUCCUAUGAU....UGCCA.GGUUGCCGGAAU
Ame-1-3    UAUUGAUAAGGGCAAACU..UA.UUGAAA...AAU..AA.G.GACGC.AAAGCUA..U.GGAUAUAAGGGAA.......................AACAUCA( 0 nucs).................UUCCUAUGAU....UGCCA.GGUUGCCGAAAU
Ame-1-4    UGUUGGUAAGGGCAAACU..UA.UUGAAA...AAU..AA.G.GGCGC.AAAGCUA..U.GGAUCUAAGGGAA........UA.............UUA....( 0 nucs)...............UA.UUCCUAUGAU....UGCCA.GGUUGCCGAAAU
Ame-1-5    UGUUGAUAAAGGCAAACU..UA.UUGAAA...AAU..AA.G.GACGC.AAAGCUA..U.GGAUCUAAGGGAA.......................GUUG...( 0 nucs)...............UA.UUCCUAUGAU....UGCCA.GGUUGCCGAAAU
Ame-1-6    ACUUGAUAAAGGCAAACU..UA.UUGAAA...AAU..AA.G.GACGC.AAAGCUA..U.GGAUCUAAGGGGA........UC.............UUUA...( 0 nucs).................UUCCUAUGAU....UGCCA.GGUUGCCAAAAU
Ame-1-7    CGUUGAUAAGGGCAAACU..UA.CUGAAA...AGU..AA.G.GACGC.AAAGCCA..U.GAGUCUAAGGGAA.......................GUUUUUA( 0 nucs).................UUCCUAUGAU....UGUCA.GGUUGCAAAUCU
Dha-2-1    AUUCGACAAUAGCAAACC..UA.GCGAAA...GCC..AG.G.GACGC.AAAGCUA..U.GGGUCUUCCUGGA.......................UACAC..( 0 nucs).................UCCUGAUGAU....CGCCA.AGCUACCGAAGC
Dha-1-1    AUUCGACAAUAGCAAACC..UG.GCGAAA...GCC..AG.G.GACGC.AAAGCUA..U.GGGUCUUCCUGGA.......................UACA...( 0 nucs)...............C.UCCUGAUGAU....CGCCA.AGCUACCGAAGC
Csa-1-1    UGGUCGAAAGGGCAAACC..UG.UCGAAA...GGC..AG.G.GGCGC.AAAGCCA..U.GGGCCUGUCGGGA........GU.............AAA....( 0 nucs)...............AC.UUCCUAUGGU....UGCCA.GGCUGCCGAAAC
Csa-1-2    AUAUCGAAAGGGCAAACC..UG.UCGAAA...GGC..AG.G.GGCGC.AAAGCCA..U.GGGCCUGUCGGAA........GU.............AAA....( 0 nucs)...............AC.UUCCUAUGGU....UGCCA.GGCUGCCGAAAA
Dre-1-1    AAUAAAGAAAGGCAAACC..UA.CUGAAA...GGU..AG.G.GGCGC.AAAGCUA..GAGGGUCUAAGG.........................GGAA...( 0 nucs)...................CCUAUGAC....AGCC..AGUUGCUCCGGG
Dha-2-2    UAGAGCCAAGGGCAAACC..CG.GUGAAA...GCC..GG.G.GGCGC.AAAGCUAC.A.GGGUCUACGGCAU........U..............UUCG...( 0 nucs)...............G.AUGCUAUGAU....AGCCU.GGCUGCCAAAAA
Dha-1-2    UAUAGCUAUGGGCAAACC..CG.GUUAAA...GCC..GG.G.GACGC.AAAGCCG..A.GGGUCUAAGGUGU........C..............UUAG...( 0 nucs)...............G.ACGCUAUGAU....AGUCU.GGCUGCUGACAG
Dha-1-3    UAAAGCUAUGGGCAAACC..CG.GCUAAA...GCC..GG.G.GACGC.AAAGCCG..A.GGGUCUAAGGUGU........C..............UUAG...( 0 nucs)...............G.ACGCUAUGAU....AGUCU.GGCUGCUGACAG
Dha-2-3    UUGCGACAAAGGCAAACU..UA.GUGAAA...GCU..AA.G.GACGC.AAAGCUA..U.GGGCCUUCCUGAU.......................CAAG...( 0 nucs)...................CUGAUGGC....CGCCA.AGCUACCGAAGU
Dha-1-4    UUGCGACAAAGGCAAACU..UA.GUGAAA...GCU..AA.G.GACGC.AAAGCUA..U.GGGCCUUCCUGAU.......................CAAG...( 0 nucs)...................CUGAUGGC....CGCCA.AGCUACCGAAGU
Csa-1-3    UGUCGAUAAAGGCAAACC..UG.UCGAAA...GGC..AG.G.GGCGC.AAAGCCA..U.GGGUCUGUCGGAA........AAGG...........GUUUUUA( 0 nucs)...........CC.UU.UUCCUAUGAU....CGCCA.GGCUGCCGAAAC
Csa-1-4    UGUCGACAAAGGCAAACC..UG.UCGAAA...GGC..AG.G.GGCGC.AAAGCCA..U.GGGUCUGUCGGAA........AAGGU..........AUAGAGG( 0 nucs)..........ACC.UU.UUCCUAUGAU....CGCCA.GGCUGCCGAAAC
Dha-2-4    UAAAAUCAAGGGCAAACC..GG.CCUAAA...GGU..AG.G.GACGC.AAAGCCGU.A.GGAUUUAAGGUGC.......................UCAAAAA( 0 nucs).................GUGCUAUGAU....AGCCU.GGUUGCCAAACA
Cbe-1-4    AUAUUGAAUUGGCAAAGC..AG.UUGAAA...AAC..UG.U.GACGC.AAAGCUA..GAGGGCCUUUUCCUU.......................UUGA...( 0 nucs).................AAGG.AUGGC....AGCC..AGUUGCAAGUCU
Dre-1-2    AAAAUUAAAAGGCAAACU..UU.CCGAAA...GGA..AA.G.GACGC.AAAGCCA..C.GGGCCUAAGGACA.......................U......( 0 nucs).................UGUCUAUGGU....AGCCG.GGCCGCCUUAGG
Dha-2-5    UAAAGUCAAGGGCAAACC..UG.CCUAAA...GGC..AG.G.GACGC.AAAACCGU.A.GGAUCUAAGGUGC........UU.............AAA....( 0 nucs)...............AG.GUGCUAUGAU....AGCCU.GGUUGCCAAUGU
9Dre-1-3   UAACAAACAGGGCAAACC..UG.GUGAAA...GUC..AG.G.GACGC.AAAGCCA..U.GGGUCUAAAGC.........................GAAU...( 0 nucs)...................GCCAGGAC....GGCUA.GGCUACUAAAAU
Ame-1-8    AAUAAGUAAAGGCAAAAC..CA.UCGAAA...GAU..GG.U.GACGC.AAAGCCA..C.GAAUCUAAGUCGU.......................UUG....( 0 nucs).................ACGAUAGGAU....AGUCG.GGUUGCCUAUUA
Ame-1-9    UAUUAGUAAAGGCAAAAC..CA.UUGAAA...GAU..GG.U.GACGC.AAAGCCA..C.GAGUCUAAGUCGU.......................UAU....( 0 nucs).................ACGAUAUGAU....AGUCG.GGUUGCCAAUUG
Dre-1-4    UUCAAAAAUGGGUAAAUC..UU.G.AAAG...GU...AG.G.GACGC.AAAGCCA..U.GGGUCUAAUGCA........................UU.....( 0 nucs)..................UGCAAUGAU....AGCCA.GGCAGCAUAAUU
Cth-1-2    UUUUGAAAAGGGCAAAUC..UA.UCGAAA...GAU..AG.A.GACGC.AAAGCCAA.C.GAGUCUAAGGUG........................GAAA...( 0 nucs)..................CACUAUGAC....GGUCG.GGUUGCCAAAAG
Dre-1-5    UUCCGACAAAGGCAAACC..CU.CCGAAA...GGG..UG.G.GGCGC.AAAACCA..C.GGGCCUAAGGU.........................GUA....( 0 nucs)...................GCUAGGGU....AGCCG.GGUUGCCGAGAA
Csp-1-1    ACAAUAAAUGGGCAAACC.......UA...........G.G.GACGC.AAAGCUU....GAGUCUACGGUUA........UCU............CAA....( 0 nucs)............A.GA.UAAUUACGAU....CGUUC.AGCUGCAUAGAU
3Cth-1-3   UUUCGACAAAGGCAAACU..CA.AUGAAA...AUU..GA.G.GACGC.AAAGCUC..C.AGGUCUUCGGUUU........G..............GUAU...( 0 nucs)............U.CC.AAACCAUGAC....GGCUG.AGUUACCGAAAA
1Dre-1-6   GAAACUGAAGGGCAAACC..CGAACGAAA...GAC..GG.G.GACGC.AAAACCU..U.GGGUCUAAG.GUU........C..............UAGGAAU( 0 nucs)...............G.AACUAUUGAU....UGCCA.GGCUGCAAAAGA
8Ame-1-10  UAAAUAAAAGGGCAAACU..UA.UCGAAA...GAU..AA.G.GGCGC.AAAGCCA..A.GGGUCUUCAUCAA.......................UUGA...( 0 nucs).................UUGAUAUGAC....AGCC..GGUUGCAGAGUA
8Cdi-1-8   AAACUAAAUCGGCAAAAC..UA.GAGAAA...UUU..AG.U.GACGC.AAAGCUA..UAGGGACUAACGCUU........AUA............AAAUAAA( 1 nucs)............U.AU.GAGUUAUGUC....AGCC..AGUUGCCAAAAA
1Ame-1-11  AGAAUGUAAAGGCAAACC..AU.UGCAAA...CAA..UG.G.GACGC.AAAGCCA..G.GAACCUAAAGUGU........GUU............AUAAAAA( 0 nucs)............A.AU.AUACCAAGAU....CGUCC.GACUGCCAUAGA
Cth-1-4    UAUGAAACAGGGCAAAAU..CA.UCGAAA...GGU..GA.U.GACGC.AAAGCCA..U.GGGUCUACUGUUU........UAAAAC.........AAU....( 0 nucs).........GUUU.UA.AAGCUAUGAU....CGCCA.GGCUGCCAUUAA
Dha-2-6    UGAUGGUUGGAGCAAACU..UU.GUGAAA...GCA..AA.G.GACGCUAAAGCUA..U.GGGUCUAAGGCCG.......................UCGCGCA( 4 nucs).................CGGCUAGGAU....CGCCA.GGCUACAAGUAG
Bma-1-1    CCUCGAAAAAGGCAAACC..GA.UCGUGA...GGU..CG.G.GACGC.AAAGCCA..A.GGGCC....AC.........................GUAA...( 0 nucs)...................GU...GGA....CGCCA.GGCUGCCGAAGG
5Rba-1-1   CCCUGACAAAGGCAAACC..GU.UCGAGA...GGA..CG.G.CACGC.AAAACCA..C.GGGUCCGU.GAGU........CGGUCGAUCGCA...UUGUU..( 0 nucs)...UGCGUUCGGCUUG.GCUUCAGGAA....AGCCG.AGUUGCCGUGGA
Asp-1-1    CUCUGAAAAAGGCAAACC..CG.UCGAAA...GGC..GG.G.GACGC.AAAAUCA..C.CGGUCUACAGGG........................GUCG...( 0 nucs)..................CCCCAUGAC....AGCGG.GAUUGCCAGGCG
Nmu-1-1    CAGGAACAAAGCCAAACC..CG.UUGCGA...AGC..GG.G.GACGG.AAAGCUG..C.GGGUC....UCC........................AAGCAAG( 1 nucs)..................GGA...GGU....AGCCG.GGUUGCUCCAGG
Nmu-1-2    AUCUGUAAAAGGCAAAUC..UC.GCGAAA...GCG..GG.G.GACGC.AAAGUCA..U.CGGUCUACCGGGC.......................UCAAG..( 0 nucs).................GCCCCAGGAC....GGCAG.GACUGCCAGAUA
Tde-1-1    AUCCCGAAAGGGCAAACC..CG.CCGCGA...GGC..GG.G.GGCGC.AAAGCGA..C.CGGUC....UC.........................GAAA...( 0 nucs)...................GA...GAU....AGCGG.CGCUGCCGUGGA
Neu-1-1    AACCGACAAAGGCUAAUC..UG.UCGUAA...GGC..AG.G.GGCGC.AAAGUUU..C.CGGUC....CUGA........UC.............CAU....( 0 nucs)...............GA.UCAG...GAA....AGCGG.GACUGCCGAAGU
Neu-2-1    AACCGACAAAGGCCAACC..UG.UCGUGA...GGC..AG.G.GGCGC.AAAGUUU..C.CGAUC....CUGA........U..............UGC....( 0 nucs)................A.UCAG...GAA....AGCGG.GACUGCCGAAGU
Rfe-1-1    CAAGUUCAAAAGCACACC..CGAUCGAAA...GAC..GG.G.GUCGC.AAAGCUU..C.CGGCCUGA..CGC........A..............UUCG...( 0 nucs)...............U.GCGAAGUGGU....AGCGG.GGUUGCCCAAAA
Rso-1-1    UUCAGUUUUGGGCACGGC..CG.GAGCGA...UCC..GG.U.GGCGU.AAAGCUA..GAGGGAC....UUC........................GAUGUUG( 0 nucs)..................GAA...GUC....AGCCA.GGUAUCCGCCGA

           ...........<<...<<..<<.<<.......>>>..>>.>....>>...<<<<...<.<<<<<....<<<<........<<<<<<<<<<<<<<<.......         >>>>>>>>>>>>>.>>.>>>>...>>>......>>>.>>>>........
           ...........22...22..22.22.......222..22.2....22...??22...?.22?22....222?........???2??2?2?21000.......         00012?2?2??2?.??.?222...22?......22?.22??........
           ●●●●●●●AR●R●YAAACY--●●-●YG●RA---R●●--●G-G-GRCG●-AAAGCY●--Y-RGGYC○○○○●●○○-----------------------●●●○---         -----------------○○●●○○○GRY----RGCYR-GGYUGCCRRR●●
                                    GNRA                                   5-TR                                                               3TR
```

```
     Mfl-1-1    AAAUGUUAAGGGCAACCU..UGGUUGAAA...AAC.CAA.G.GACGC.AAAGUCA..C.CGGUCUAAGGUA.A.......GCA............GCU....( 0 nucs)............U.GC..UACUAUGAC....AGCGG.AACCGCCAUAUU
     Nmu-1-3    UACCGUAAAAGCCACACC..CG.UCGCGA...GGC..GG.G.GCCGG.AAAGUCC..C.GGAUU....UGCG........C..............AAUUGAA( 8 nucs)...............G.CGCA...GGU....AGCCG.GACUGCCAGUAC
  13 Sfu-1-1    ACCUGUAAAGGGCAAACC..AU.CCGAAA...GGG..UG.G.GACGC.AAAGCCG..C.UGGCCUAAGUCCC.......................CCGCGU.( 0 nucs).................GGGAUAUGGC....AGCAG.GGUUGCCAGGGG
     Sfu-1-2    GUUUGGAAAAGGCAAACU..AU.UCGAAA...GAC..UG.G.GACGC.AAAGCUU..C.UGGCCUAAGUCC........................AAAA...( 0 nucs)..................GGAUAUGGC....AGCAG.AGCCGCCGAGCG
   1 Sfu-1-3    UUGUGAAAAAGCCAAACC..UC.CCGCGA...GGG..GG.G.GACGG.AAAACCG..C.GGGUC....U..........................CACGU..( 0 nucs)....................A...GAC....GGCCG.GGUUGCCUGAAA
   1 Sfu-1-4    GUCAGAAAACGCCAACCC..UG.UUGCGA...GAC..GG.G.GACGG.AAAGCCG..C.GGGUC....UC.........................AUG....( 0 nucs)...................GA...GAU....GGCCG.GGUUGCCUGACA
   8 Sfu-1-5    GUUAGAAAAAGCCAAAUC..CA.UCGCAA...GGU..GG.A.GACGG.AAAGCCG..C.GGGUC....UC.........................CUUCCCA( 3 nucs)...................GA...GAC....GGCCG.GGUUGCCUGAUC
     Bba-1-1    UUUGCCUCAAGGCAAACC..GU.CGGAAA...CGG..CG.G.GACGC.AAAGCUA..AAGGGGU....GUCG.......................GUGCAAA( 1 nucs).................CCAC...GCC....AGCC..AGCUGCCAAAGG
     Dde-1-1    CGAGAAUCAAGCCAACCC..GC.C.UCA.....GG..CG.G.GACGG.AAAGCCA..C.GGGUC....U..........................UUC....( 0 nucs)....................A...GAC....AGCCG.GGUUGCCUCGGU
     Sfu-1-6    ACCCGAAAAAGGCAAACC..AU.UCGAAA...GGG..UG.G.GACGC.AAAGCCA..C.UGGCCUACGUCUU........CC.............GAG....( 0 nucs)..............GG.AGGAUAGGGC....GGCAG.GGUUGCCGGAGG
     Sfu-1-7    GUUCGAUAAUGGCAAACU..GG.UCGAAA...GAC..CG.G.GACGC.AAAGCCA..C.UGGCCUAAGUCC........................CCUCUGA( 0 nucs)..................GGAUACGGU....AGCAG.GGUCGCCGGGCG
     Ade-1-1    CACGCUUCAGGGCAAUUC..AC.CCGUACCCAGGG..UG.A.AGCGC.AAAGCCG..C.GGGUC....CGGU.......................GAAC...( 0 nucs).................GCCG...GAC....GGCCGUGGCCGCCGUGGC
  12 Mxa-1-1    CCUGUUCAAGGGCAAUCCC.GG.GCGAAA...GCCC.CG.G.UGCGC.AAAGCCG..C.GGGGC....CUGA.......................AG.....( 0 nucs).................UCCG...GCC....GGCCGUGACCGCCGUGGG
     Gsu-1-1    ACACGACAAUACUAAACC..AU.CCGCGA...GGG..UG.G.GACGG.AAAGCCC.ACAGGGUC....UC.........................UCU....( 0 nucs)...................GA...GAC....AGCCG.GGAUGCCGAAAU
   2 Gsu-1-2    UAUCGAUAAUACUAAACC..AU.CCGUGA...GGA..UG.G.GGCGG.AAAGCCU.ACAGGGUC....UC.........................ACC....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAAAU
   1 Gsu-1-3    ACACGAUACUACUAAACC..AU.CCGCGA...GGA..UG.G.GGCGG.AAAGCCC.ACAGGGUC....UC.........................ACC....( 0 nucs)...................GA...GAC....AGCCG.GGAUGCCGAACU
   1 Gme-1-1    ACACGACAAUACUAAACC..AU.CCGCGA...GGA..UG.G.GGCGG.AAAGCCC.ACAGGGUC....UC.........................ACC....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAAAU
   2 Gsu-1-4    ACACGACAAUACUAAACC..AU.CCGCGA...GGA..UG.G.GGCGG.AAAGCCU..AAGGGUC....UC.........................CCU....( 0 nucs)...................GA...GAC....AGCCG.GGUCGCCGAAAU
   2 Gsu-1-5    AAACGAUAAUACUAAACC..AU.CCGCGA...GGA..UG.G.GGCGG.AAAGCCU.ACAGGGUC....UC.........................ACC....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAACU
   2 Gur-1-1    AAACGAUAAUACUCAACC..AU.CCGCGA...GGG..UG.G.GGCGG.AAAGCCU.ACAGGGUC....UC.........................ACU....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAAAU
   2 Gme-1-2    AGACGACAAUACUUAACC..AU.UCGCGA...GAA..UG.G.GGCGG.AAAGCCU.AUAGGGUC....UC.........................AUG....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAAAU
   2 Gur-1-2    AGACGAUAAUACUUAACC..AU.UCGCAA...GAA..UG.G.GGCGG.AAAGCCUA.C.GGGUC....UU.........................ACU....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAAAU
   2 Ppr-1-1    AGACGAUACUACUUAACC..AU.UCGCAA...GAA..UG.G.GGCGG.AAAGCCU..AAGGGUC....UUA........................C......( 0 nucs)..................UGA...GAC....AGCCG.GGUUGCCGAAAU
   2 Gsp-1-1    AGACGAUAAUACUCAACC..AU.UUGCAA...GAA..UG.G.GGCGG.AAAGCCU.AUAGGGUC....UU.........................ACU....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAAAU
   2 Gsp-1-2    AGACGAUAAUACUCAACC..AU.CCGCGA...GGA..UG.G.GGCGG.AAAGCCU.AAAGGGUC....UC.........................UCU....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAAAU
   2 Gur-1-3    AGCCGAUAAUACUCAACC..AU.CCGCGA...GGA..UG.G.GGCGG.AAAGCCU.ACUGGGUC....UC.........................ACC....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAAAU
2,20 Gme-1-3    AGACGAUAAUACUAAACC..AU.UCGCGA...GAA..UG.G.GGCGG.AAAGCCU.AUAGGGUC....UC.........................CCU....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAAAU
   2 Gme-1-4    AGACGACAAUACUAAACC..AU.CCGCGA...GGG..UG.G.GACGG.AAAGCCU.ACAGGGUC....UC.........................CCC....( 0 nucs)...................GA...GAC....AGCCG.GGUCGCCGAAAU
   1 Gur-1-4    AGACGAUAAUACUAAACC..AU.CCGUGA...GGA..UG.G.GGCGG.AAAGCCU.AAAGGGUC....U..........................CAAGCC.( 0 nucs)....................A...GAC....AGCCG.GGUUGCCGAAAU
   1 Gur-1-5    AGACGAUAAUACUUAACC..AU.CCGCGA...GGG..UG.G.GGCGG.AAAGCCU.AAAGGGUC....UC.........................CAU....( 0 nucs)...................GA...GAU....AGCCG.GGUUGCCGAAAU
   2 Gsu-1-6    AGACGACAAUACUCAACC..AU.UCGCGA...GAA..UG.G.GGCGG.AAAGCCU.ACAGGGUC....UC.........................ACC....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAAAU
   2 Gur-1-6    AGACGAUAAUACUAAACC..AU.UCGUGA...GAA..UG.G.GGCGG.AAAGCCU.AUAGGGUC....UC.........................GAA....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAAAU
   2 Gur-1-7    AGACGAUAAUACUAAACC..AU.UCGCGA...GAA..UG.G.GGCGG.AAAGCCUA.AUGGGUC....UC.........................GAA....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAAAU
   1 Gsp-1-3    UAACGAUAAUACUAAACC..AU.UCGCGA...GAA..UG.G.GGCGG.AAAGCCU.AUAGGGUC....U..........................CAAGC..( 0 nucs)....................A...GAC....AGCCG.GGUUGCCGAAAU
   2 Gur-1-8    UAACGAUAAUACUAAACC..AU.CCGCGA...GGA..UG.G.GGCGG.AAAGCCU.AUAGGGUC....U..........................CACGC..( 0 nucs)....................A...GAC....AGCCG.GGUUGCCGAAAU
   2 Gur-1-9    ACGCGAUACUACUAAACC..AC.CCGCGA...GGG..UG.G.GGCGG.AAAGCCC.ACAGGGUC....UU.........................UAA....( 0 nucs)...................AA...GAC....AGCCG.GGUUGCCGAACU
   2 Gsu-1-7    AGACGAUAAGACUAAACC..AC.CCGCGA...GGG..UG.G.GGCGG.AAAGCCC.AC.GGGUC....UC.........................UCU....( 0 nucs)...................GA...GAC....AGCCG.GGAUGCCGAAAU
   1 Gme-1-5    AGUCGAUAAUACUAAACC..AU.UCGUGA...GAA..UG.G.GGCGG.AAAGCCU.ACAGGGUC....UU.........................ACU....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAAAU
   1 Gur-1-10   ACACGAUAAUACUAAACC..AU.CCGUGA...GGG..UG.G.GGCGG.AAAGCCU.AUAGGGUC....U..........................CAAGUC.( 0 nucs)....................A...GAC....AGCCG.GGUUGCCGAAAU
   2 Gme-1-6    AGACGACAAUACUCAACC..AU.CCGUGA...GGA..UG.G.GGCGG.AAAGCCU.AUUGGGUC....UC.........................ACC....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAAAU
   2 Gme-1-7    AGACGAUAAUACUCAACC..AU.CCGCGA...GGA..UG.G.GGCGG.AAAGCCU.ACAGGGUC....UC.........................UUC....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAAAU
   1 Gur-1-11   AGACGAUAAUACUAAACC..AU.CCGCAA...GGA..UG.G.GGCGG.AAAGCCUA.AAGGGUC....UC.........................ACU....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAAAU
2,13 Gsu-1-8    ACACGAUAAUACUCAACC..AU.CCGCGA...GGA..UG.G.GGCGG.AAAGCCU.ACAGGGUC....UC.........................ACA....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAAAU
   1 Ppr-1-2    ACACGAUAAUACUCAACC..AU.CCGCGA...GGA..UG.G.GGCGG.AAAGCCU.ACAGGGUC....UC.........................ACC....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAAAU
   2 Gme-1-8    ACACGACAAUACUCAACC..AU.CCGCGA...GGA..UG.G.GGCGG.AAAGCCU.ACAGGGUC....U..........................CCAGC..( 0 nucs)....................A...GAC....AGCCG.GGUUGCCGAAAU
   2 Gsu-1-9    AGACGAUAAUACUAAACC..AU.UCGCGA...GAA..UG.G.GGCGG.AAAGCCU.AUAGGGUC....UC.........................CAC....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAAAU
   2 Gme-1-9    AGACGAUAAUACUCAACC..AU.UCGCGA...GAA..UG.G.GGCGG.AAAGCCU.AUAGGGUC....U..........................CAUGC..( 0 nucs)....................A...GAC....AGCCG.GGUUGCCGAAAU
     Gme-1-10   AGACGACAAUACUAAACC..AU.UCGCGA...GAA..UG.G.GGCGG.AAAGCCU.ACAGGGUC....UU.........................AUU....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAAAU
   2 Gsu-1-10   ACACGACAAUACUAAACC..AU.CCGCGA...GGA..UG.G.GACGG.AAAGCCC.ACAGGGUC....UC.........................ACC....( 0 nucs)...................GA...GAC....AGCCG.GGUCGCCGAAUU
   2 Gur-1-12   AAACGAUAAUACCAAACC..AU.CCGCGA...GGA..UG.G.GACGG.AAAGCCUA.CAGGGUC....UC.........................AA.....( 0 nucs)...................GA..AGAC....AGCCG.GGUCGCCGAAAU

                ...........<<...<<..<<.<<.......>>>..>>.>....>>...<<<<...<.<<<<<....<<<<........<<<<<<<<<<<<<<<.......         >>>>>>>>>>>>>.>>.>>>>...>>>......>>>.>>>>........
                ...........22...22..22.22.......222..22.2....22...??22...?.22?22....222?........???2??2?2?21000.......         00012?2?2??2?.??.?222...22?......22?.22??........
                ●●●●●●●AR●R●YAAACY--●●-●YG●RA---R●●--●G-G-GRCG●-AAAGCY●--Y-RGGYC○○○○●●○○-----------------------●●●○---         -----------------○○●●○○○GRY----RGCYR-GGYUGCCRRR●●
                                         GNRA                                   5-TR                                                                3TR
```

```
[2]Gsp-1-4        AGACGACAAUACUAAACC..AU.UCGCAA...GGG..UG.G.GACGG.AAAGCCU.ACAGGGUC....UU.........................ACU....( 0 nucs)...................GA...GAC....AGCCG.GGACGCCGAAAU
[2]Gsu-1-11       ACACGAUAAGACUAAACC..GU.CCGCGA...GGG..CG.G.GGCGG.AAAGCCU..A.GGGUC....UC.........................CUA....( 0 nucs)...................GA...GAC....AGCCG.GGAUGCCGAAUU
[1]Gme-1-11       UAGCGAUACUACUAAACC..AU.CCGCGA...GGA..UG.G.GGCGG.AAAGCCU.ACAGGGUC....UC.........................AUC....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAACU
Gur-1-13          AGACGAAAAUACUAAACC..AU.UCGCGA...GAA..UG.G.GACGG.AAAGCCU.AAAGGGUC....UC.........................ACC....( 0 nucs)...................GA...GAC....AGCCG.GGUCGCCGAAAU
[2]Ppr-1-3        AGACGACAAUACUAAACC..AU.UCGCGA...GAA..UG.G.GACGG.AAAGCCU.ACAGGGUC....UC.........................ACC....( 0 nucs)...................GA...GAC....AGCCG.GGUCGCCGAAAU
[2]Gme-1-12       AGACGAUAAUACUAAACC..AU.CCGCGA...GGG..UG.G.GACGG.AAAGCCU.ACAGGGUC....UC.........................CCC....( 0 nucs)...................GA...GAC....AGCCG.GGUCGCCAAAUA
[1]Gme-1-13       AGACGACAAUACUAAACC..AU.CCGCGA...GGA..UG.G.GGCGG.AAAACCC.ACAGGGUC....UC.........................CCU....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAAAU
[1]Gme-1-14       AGACGACAAUACUAAACC..AU.CCGCGA...GGA..UG.G.GGCGG.AAAGCCC.AUAGGGUC....UC.........................ACC....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAAAU
[1,13]Gsu-1-12    UAGCGAUACUACUAAACC..AU.CCGCGA...GGA..UG.G.GGCGG.AAAGCCC.ACAGGGUC....U..........................CACGA..( 0 nucs)....................A...GAC....AGCCG.GGUUGCCGAACU
[1]Gme-1-15       UAGCGAUACUACUAAACC..AU.CCGCGA...GGA..UG.G.GGCGG.AAAGCCU.ACAGGGUC....UC.........................AUC....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAACU
[2]Gsp-1-5        AGACGAUAAUACUAAACC..AU.CCGCGA...GGG..UG.G.GACGG.AAAGCCC.AUUGGGUC....UC.........................AUC....( 0 nucs)...................GA...GAU....AGCCG.GGUCGCCGAAAU
[1]Gsp-1-6        AGACGACAAUACUAAACC..AU.UCGCGA...GAA..UG.G.GGCGG.AAAACCC..AUGGGUC....U..........................CCUGU..( 0 nucs)....................A...GAC....AGCCG.GGCUGCCGAAAU
[2]Gur-1-14       AGACGACAAGGCUCAACC..AU.UCGUGA...GAA..UG.G.GGCGG.AAAGCCCAUC.GGGUC....UC.........................CAC....( 0 nucs)...................GA...GAC....AGCCG.GGUUACCGAAAU
[2]Gme-1-16       UAUCGACAAUACUAAACC..AU.CCGCGA...GGG..UG.G.GACGG.AAAGCCU.ACAGGGUC....UC.........................UCU....( 0 nucs)...................GA...GAC....AGCCG.GGAUGCCGAAAU
[2]Gur-1-15       AGACGAUAAUACUAAACC..AC.CUGUGA...AGG..UG.G.GGCGG.AAAGCCU.AUUGGGUC....UC.........................CAG....( 0 nucs)...................GA...GAC....AGCCG.GGUUGCCGAAAU
[1]Gur-1-16       AAACGAUAAUACUAAACC..AU.CCGUGA...GGA..UG.G.GACGG.AAAGCCU.AUAGGGUC....U..........................CAUGA..( 0 nucs)....................A...GAC....AGCCG.GGUCGCCGAAUA
[1]Gur-1-17       AGACGAUAAUACUAAACC..AU.CCGCGA...GGA..UG.G.GACGG.AAAGCCU.ACAGGGUC....U..........................CAUGA..( 0 nucs)....................A...GAC....AGCCG.GGUCGCCGAAUA
[2,14]Gsu-1-13    AGCCGAUAAGACUAAACC..AU.CCGCGA...GGG..UG.G.GGCGG.AAAGCCC..A.GGGUC....UC.........................AAG....( 0 nucs)...................GA...GAU....GGCCG.GGAUGCCGAAGU
[1]Ppr-1-4        ACACGACACUACUCAACC..AU.CCGCGA...GGA..UG.G.GGCGG.AAAGCCU..CAGGAUC....U..........................CCAGA..( 0 nucs)....................A...GAC....AGCCG.GGUUGCCGAAUU
[2]Gur-1-18       AGACGAUACUACUAAACC..AU.CCGCGA...GGA..UG.G.GGACGGAAAAGCC.ACAGGAUC....U..........................CCAGA..( 0 nucs)....................A...GAC....AGCCG.GGUCGCCGAAAU
[1]Gur-1-19       AGACGAUAAUACUAAACC..AU.UCGCGA...GAA..UG.G.GACGG.AAAGCCU.AAAGGGUC....UC.........................ACC....( 0 nucs)...................GA...GAC....AGCCA.GGUCGCCGAAAU
[1]Ppr-1-5        AAUCGAUACUACUAAACC..AU.CCGCGA...GGA..UG.G.GACGG.AAAGCCC.ACAGGGUC....U..........................CCAGA..( 0 nucs)....................A...GAC....AGCCG.GGUCGCCGAAAU
[6]Gur-1-20       AGGCAAUAAUGCUAAACC..UU.CCGUGA...GGA..GG...AGCGG.AAAGCUA..CAGG.UC....UC.........................CAG....( 0 nucs)...................GA...GACAGGCAGCCC.GGUUGCCGAAAU
[2]Gur-1-21       UCCCGACAAAGCCAACCC..CU.UCGUGA...GGG..GG.GGAACGG.AAAGCCA..C.GGGCC....UCCG.......................GAA....( 0 nucs).................UGGA...GAC....AGCCG.AGCCGCCAAAGG
[1,6,18]Ppr-1-6   AAACGUUACACCCAAGCC..AU.CCGCGA...GGA..UG.G.GGCGC.AACGCCU..CAGGAUC....UU.........................ACG....( 0 nucs)...................AA...AAC....AGCCG.GGUUGCCGAACC
[1]Pca-1-1        UCCCGAUAAAACCAAACC..UG.UCGCAA...GAC..AG.G.GACGG.AAAGCAA..C.GGAUC....UUC........................GAACCA.( 0 nucs)..................GAA...GAU....GGCCG.AGCUGCCGAAGG
[2]Gur-1-22       UACCGAAAAAGCCAAACC..UU.UCGUGA...GAG..GG.G.GGCGG.AAAGCAA..C.GGGUC....UUCA.......................GA.....( 0 nucs).................UGAA...GAU....GGCCG.GGCUGCCGAAGG
Gur-1-23          UUCCGACAAAGCCAAACC..CU.UCGCGA...GGA..GG.G.GGCGG.AAAGCGG..C.GGGUC....UUCG.......................CGA....( 0 nucs).................CGAA...GAU....GGCCG.AGCUGCCGAGGG
[3]Gur-1-24       UCACGAUAAAGCCGAACC..CU.CCGCAA...GAA..GG.G.GGCGG.AAAGUCG..C.GGGUC....UUC........................GCAGUG.( 0 nucs)..................GAA...GAC....GGCCG.GACUGCCGAAGG
[1]Ppr-1-7        UUUUGAAAAGGCCAAACC..CU.UCGCGA...GGA..GG.G.GGCGG.AAAGCGA..G.GGAUC....UUC.........C..............CUGAA..( 0 nucs)...............G.AGAA...GAU....GGCCG.AGCUGCCAAAGG
[2]Gur-1-25       UCUCCGAACGGGCAAAAC..CA.GAGCAA...UCU..GG.U.GACGC.AAAGCCA..C.GGGUC....CU.........................GACCUAA( 1 nucs)...................AG...GAU....AGCCG.GGUUGCCGAAGA
[4,17]Gur-1-26    UCUCCGAACGGGCAAAAC..CA.GAGUAA...UCU..GG.U.GACGC.AAAGCCA..C.GGGUC....CU.........................GACCUAA( 0 nucs)...................AG...GAU....AGCCG.GGUUGCCGAAGA
Gur-1-27          UCUCCGAACGGGCAAAAC..CA.GAGUAA...UCU..GG.U.GACGC.AAAGCUA..C.GGGUC....CU.........................GACCUAA( 1 nucs)...................AG...GAU....AGCCG.GGUUGCCGAAGA
Pca-1-2           UUUCGUUCAAGGCAAAGU..CA.GAGUAA...UCU..GG.C.CACGC.AAAACCA..C.GGGUC....CAUG.......................GUU....( 0 nucs).................CAUG...GAU....AGCCG.GGUUGCCGAAAA
Pca-1-3           UUUAACAUAGGGCAAAAC..UG.GAGUAA...UCC..AG.U.GACGC.AAAGCCA..C.GGGUC....CUU........................GCC....( 0 nucs)..................GAG...GAU....AGCCG.GGUUGCUGAAGA
Pca-1-4           AGGAUAUAUCGGCAAAGC..CG.GAGUAA...UCC..GG.U.GACGC.AAAGCCA..C.GGGUC....CUC........................AUU....( 0 nucs)..................GAG...GAU....AGCCG.GGUUGCUGAAGA
Dac-1-1           CUCCGAUAAAGGCAAAAC..CG.GAGUAA...UCC..GG.U.GACGC.AAAGCCA..C.GGGUC....CUGU.......................UUG....( 0 nucs).................ACAG...GAU....CGCCG.GGUUCCCGAAGA
Dac-1-2           GCAAGAUAAAGGCAAAAC..CG.AAGUAA...UUC..GG.U.GACGC.AAAGCCA..C.GGGUC....CA.........................UAAC...( 0 nucs)...................UG...GAU....CGCCG.GGUUGCUCAACU
Ppr-1-8           UCUAUGAAAAGGCAAACU..CA.GGGCAA...CCU..GG.G.GACGC.AAAGCCA..C.GGGUC....CGCA..............................( 0 nucs)...............GCAGCG...GAC....AGCCG.GGACAUCAAGAG
Pca-1-5           UUCCGACAAUGGCAAAAU..CG.GGGUGA...CCC..GA.U.GACGC.AAAGCCA..C.GGGUC....C..........................CAUUGCA( 1 nucs)....................G...GAC....AGCCG.GGUUAUCGAAGA
[22]Ppr-1-9       GCCUGAAAAAGGCAAACC..AU.UCGAAA...GGA..UG.G.GGCGC.AAAACCA..C.UGGUCUAAAUCCC........GG.............AAAU...( 0 nucs)..............UC.GGGACAUGAC...AGCAG.GGCUGCCGGGCU
Pca-1-6           UUCCGAUAAGGGCAAAAC..UG.GAGUAA...UCC..AG.U.GACGC.AAAGCUGC.G.GGUGC....GUGG.......................GAAA...( 0 nucs).................CCAC..GGAU....AGCCG.GGCUACCGAAGA
Gur-1-28          UCUAGAUAAAGCCAAAAC..CC.GGGUAA...CCG..GG.U.GACGG.AAAGCCA..C.GGGCC....C..........................UGCGAAU( 0 nucs)....................G...GGC....AGCCG.GGUUGCCUGGGG
[2]Gsu-1-14       UUUUCGAAGAGGUAAACC..CC.GGGUAA...CCG..GG.G.GGCAC.AAAGCUG..C.GGGUC....CCGCC.......UG.............AAA....( 0 nucs)..............CA.GCGG...GAC....AGCCG.GGUUGGCGAAGA
[2]Ppr-1-10       UUUCCGAAGGGGUAAACC..CA.GGGUAA...CCU..GG.G.GACAC.AAAGCCG..C.GGGUC....CUGA........UUGG...........UGGA...( 0 nucs)...........UC.GG.UCAG...GAU....AGACG.GGUUGGCGAGAA
Gme-1-17          UUUCCGAAAGGGCAAACU..UC.GGGUAA...CCG..GG.G.GACGC.AAAGCCA..C.GGGUC....CUGA........CU.............CCAU...( 0 nucs)..............GG.UCAG...GAU....AGCCG.GGUUGUCGAAGA
Gme-1-18          UUUCCGAGAGGGCAAACU..UC.GGGUAA...CCG..GA.G.GACGC.AAAGCCA..C.GGGUC....CUGA.......................CUCCAUG( 1 nucs).................UCAG...GAU....AGCCG.GGUUGUCGAAGA
[2]Gsu-1-15       UUUCCGAAGGGGUAAACC..CA.GGGUAA...CCU..GG.G.GACAC.AAAGCUA..C.GGGUC....CUGAA.......A..............UGUCCUA( 5 nucs)...............U.UCAG...GAU....AGCCG.GGUUGGCGAGAA
Gsu-1-16          UUUCCGAAGAGGUAAAAC..CA.GGGUAA...CCU..GG.U.GACAC.AAAGCCA..C.GGGUC....CUGA........AAU............CGUCCUA( 3 nucs)............A.UU.UCAG...GAU....GGCCG.GGUUGGCGAAGA
[15,16]Gsu-1-17   CUCCGAAAAGAGUAAACC..CA.UCGCAA...GGU..GG.G.GACAC.AAAGCCGA.C.GGGUG....CCGC.......................UGGA...( 0 nucs).................GCGG...GAC....GGCCG.GGUUGCCGGAGA

                  ...........<<...<<..<<.<<.......>>>..>>.>....>>...<<<<...<.<<<<<....<<<<........<<<<<<<<<<<<<<<.......         >>>>>>>>>>>>>.>>.>>>>...>>>......>>>.>>>>........
                  ...........22...22..22.22.......222..22.2....22...??22...?.22?22....222?........???2??2?2?21000.......         00012?2?2??2?.??.?222...22?......22?.22??........
                  ●●●●●●●AR●R●YAAACY--●●-●YG●RA---R●●--●G-G-GRCG●-AAAGCY●--Y-RGGYC○○○○●●○○-----------------------●●●○---         -----------------○○●●○○○GRY----RGCYR-GGYUGCCRRR●●
                                           GNRA                                   5-TR                                                                3TR
```

```
     Ppr-1-11  UCACCGAAAGGGCACAGC..AG.GGGCAA...CCC..UG.U.GUCGC.AAAGCUA..C.C.AUC....CCU........................GAGC...( 0 nucs)..................GGG...GAU....AGAGG.GGUUAUCGAAGU
   13Ppr-1-12  UCACCGAUAGGGCACAGC..AG.GGGCAA...CCC..UG.U.GUCGC.AAAGCUA..C.C.AUC....CUC........................GGGA...( 0 nucs)..................GGG...GAU....AGAGG.GGUUGUCGAAGU
     Ppr-1-13  CCACCGACGGUGCACAGC..AC.GGGUAA...CCG..UG.U.GUCGC.AAAACUA..C.C.AUC....CC.........................GAAGA..( 0 nucs)...................GG...GAU....AGAGG.GGUUGUCGAAGU
   13Ppr-1-14  UCGCCGAAAGGGUAAAGC..AC.GGGUAA...CCG..UG.U.GGCAC.AAAGCCA..C.C.AUC....CAGA.......................AUGG...( 0 nucs).................UCUG...GAU....AGCGG.GGCUGUCGAAGU
     Ppr-1-15  UCGCCGAAAGGGUAAAGC..AC.GGGUAA...CCG..UG.U.GGCAC.AAAGCCA..C.CACCC....AGC........................ACAA...( 0 nucs)..................GCU...GGG...CAGAGG.AGCUGCCGAAGU
     Pca-1-7   CCGACGAAAAGGUAAACC..CC.GGGCAA...CCG..GG.G.GGCGC.AAAGCCG..C.UCAUC....CGGA.......................UCU....( 0 nucs).................UCCG...GAA....AGGAG.GGCUGUCGAACC
     Gur-1-29  CCGACGAAAAGGCAAAGC..AC.GGGCAA...CCG..UG.U.GACGC.AAAGCUA..C.CUGUC....CAGU.......................UGCA...( 0 nucs).................ACUG...GAU....AGAGG.GGCUGCCGAAUA
     Ppr-1-16  AAACCGAAAAGGCAAAAU..CC.GGGGAA...CCG..GA.U.GACGC.AAAGCCA..C.CUCCC...............................AGGCCG.( 0 nucs)........................GGC....AGAGG.GGCCGCCGAAGU
   13Ppr-1-17  UCACCGAAAAGGCAAAAU..CC.GGGGGA...CCG..GA.U.GACGC.AAAGCCA..C.CUACC....C..........................AGAAC..( 0 nucs)....................G...GGA....AGAGG.GGUUGUCGAAGC
     Pca-1-8   UUCCGAAAAUAGCAAACU..CG.GGGUAA...CUC..GG.G.GACGC.AAAGCCA..U.GGACC....CGUU.......................GCA....( 0 nucs).................GGCG...GGC....CGCCA.GGCUACCGAAGA
     Pca-1-9   CUCCGAUAAGAGCACAGC..CA.GAUUCU...UCU..GG.U.GUCGC.AAAGUCA..C.GGGUU....CGCU........UU.............UAA....( 0 nucs)..............AG.AGCG...GAU....GGCCG.GUCUGCCGAAGA
     Gme-1-19  ACCAGUAAAGGCCAAAAC..AC.GGGUAA...CCG..UG.U.GACGG.AAAGCCU.AC.GGGUCCAC.GAGA.......................CGGAUUC( 3 nucs).................UUUUGACGGA....GACCAAGGCGGCGAGGAG
     Gsp-1-7   UUUCCGAAAAGGUAAGCUUCGA.GAGAAA...UCUCCUG.G.CGCAC.AAAGUCA..C.GGGUU....UAUA........AUCUUUGAUCAGAUCUCC....( 0 nucs)GAUUUAAUCUCAG.GU.UAUC...GAC....AGCCG.GGCUACCGAAGA
 2,11Gur-1-30  GACUUGAAAUGCCAAACC..UU.GAGCAA...UCA..GG.G.GACGG.AAAGCCA.AC.GGGAC....UUGGGUUCCAAUGAAU...........UCCAAUG(63 nucs)...........AU.UCUCCAA...GUU....GGCCG.GGUUGCCGUUUG
     Osp-1-1   UCUCGAUAAUGGCAAACC..AG.AAGCGA...UUC..UG.G.GGCGC.AAAGUCA..C.AGGUC....C..........................CUUGU..( 0 nucs)....................G...GAU....GGCUG.GGCUACCGAAAG
     Osp-1-2   CUCUCGAAAAGGCAAAAC..CC.GAGUAA...UCG..GG.U.GACGC.AAAGUCA..C.GGGUC....CC.........................AAU....( 0 nucs)...................GG...GAC....AGCCG.GACCGCCACAGG
     Tcr-1-1   AUUUGAUAAAGCUAAAUU..CG.UCGCGA...GGC..GA.A.GACAG.AAAGCCA..C.GGAUC....UC.........................CCU....( 0 nucs)...................GA...GAU....GGCCG.GGUUACCUUUUU
     Tcr-1-2   AGAAAAUUUUGCCAAACU..UG.UUGUGA...AGC..AA.G.GACGG.AAAGCCA..U.GGGUC....UCAA.......................UGG....( 0 nucs).................UUGA...GAU....CGCCA.GGUUGCCAAAUU
   21Tcr-1-3   GUUUUUAAUUGCCAAACU..UG.UUGUGA...AGC..AG.G.GACGG.AAAGCCA..U.GGGUC....UCAA.......................UGUU...( 0 nucs).................UUGA...GAU....CGCCA.GGUUGCAUACCA
     Tcr-1-4   AUGUAGAUUUGCCAAACU..UA.CUGUGA...AGU..AA.G.GACGG.AAAGCCA..C.GGGUC....UUUG........U..............CAUCAA.( 0 nucs)...............A.UAAA...GAU....AGCCG.GGUUGCGUAGAG
     Osp-1-3   AACCCUAAUGGCCAAACC..UG.CCGUGA...GAC..AG.G.GACGG.AAAGCUA..C.AGGUC....UA.U........G..............UUUUGAU( 2 nucs)...............C.AUUA...GAU....GGCUG.AGCUGCGCAGAA
     Ptu-1-1   GCACCUAUUUGGCAAACU..AA.CCGAAA...GGU..UA.G.GACGC.AAAGCUU..C.CGGUCUACCAUCA........U..............UUU....( 0 nucs)...............A.UGAUCAUGAU....AGCGG.GGUUGCCGCAAC
     Pin-1-1   GAAAGUCAUUCGCAAACU..UG.GGGAAA...UUC..AA.G.GACGC.AAAACUA..UAGGGCC...............................UGUAAAA( 1 nucs)........................GGC....AGCC..AGUUGCCAAAAG
   23Pin-1-2   GAAGUUGAUUCGCAAACC..UG.GGGAAA...CCU..UG.G.GGCGC.AAAGCUA..UAGGGCC...............................UGUAAAA( 1 nucs)........................GGU....AGCC..AGUUGCCGAAAG
     Pin-1-3   AUAACUAAUUCGCAAACC..UG.GGGAAA...CCC..UG.G.GGCGC.AAAGCUA..UAGGGCC...............................UGUAAAA( 1 nucs)........................GGU....AGCC..AGUUGCCAAGAG
     Ptu-1-2   GCAAGUCAAAGGCAAGCC..AC.UUGAAA...AAG..UA.GUCACGC.AAAGUUA..C.AGAUCUAAGGG.........................GAAA...( 0 nucs)...................CCUAUGAU....GGCUG.AACUACCAAAGC
     Cps-1-1   GCUUGAUAAGGGCAAAUC..UA.AUGAAA...GUU..AG.A.GACGC.AAAGUUA..C.CGGUCUAAUGG.........................GAAA...( 0 nucs)...................CCUAUGAU....AGCGG.GACUGCAAAUUU
     Ptu-1-3   GUUCGUAAAGGGCAAACU..UA.GUGAAA...AUU..AA.G.GACGC.AAAGCUA..U.UGGUCUAAGGGUA.......................UU.....( 0 nucs).................UACCUAUGAU....UGCAA.AGCUGCCAACGA
     Iba-1-1   ACUUGAGUCGGCCAAACC..CU.CCGUGA...GGC..GG.G.GACGG.AAAGCUU..C.GGGUC....UUC........................UAAAAAU( 6 nucs)..................GAA...GAU....AGCCG.GGUUGCAGCUAA
     Cps-1-2   ACAAAGAAAAGGCAAAGC..AC.GCGAAA...ACG..UG.U.UACGC.AAAACCA..C.CGGUCUAAGGGUA........AA.............GUUUAAU( 2 nucs)..............UU.UAUCUAUGAC....AGCGG.AGUCACCUCGUU
     Cps-1-3   AAUUGAAAAGAGCAAAAC..UU.UCGAAA...GAA..AG.U.GGCGC.AAAGUCA..C.CGGUCUACGGGA........................AUAA...( 0 nucs)..................UUCUACGAC....AGCGG.AACUGCCAACUA
     Pin-1-4   AAAAA.....AGCAAAUU..AA.GGGAAA...CCU..UA.A.UACGC.AAAACCA..C.CGGCCUGUUGAUU........GA.............CUAU...( 0 nucs)..............UU.AGUCUAUGGC....AGCGG.AGUUACCAAGGC
     Pin-1-5   GCUGGAAAAAAGCAAAUU..AA.GGGAAA...CCU..UA.A.UACGC.AAAACCA..C.CGGCCUGUUGAUU........GA.............CUAU...( 0 nucs)..............UU.AGUCUAUGGC....AGCGG.AGUUACCAAGGC
     Sde-1-1   CACUCUAAAUGGCUAACC..GUCGCGAAA...GCG..CG.GCCACGC.AAAGCUU..C.CGGUCUAAGGGUA........UG.............UAAAC..( 0 nucs)..............CA.UAUCUACGAU....AGCGG.GGUUGCUGGAAU
    3Iba-1-2   UUGUUUAACUGCCAAACU..CA.UCGCGA...GAU..GA.G.GACGG.AAAGCCA..C.GGGUCUCAGAGUG........UGUU...........GA.....( 0 nucs)...........GA.UG.UAUCUCAGAU....AGCCG.GGUUGCGGAACU
     Sde-1-2   CAACGAUAAAGGCAAACC..CG.CGGUAA...CGU..GG.G.GACGC.AAAGCCA..G.U.GCC....CAA.CUA.....CUUG...........UAAUU..( 0 nucs)...........UA.AGA.UUG...GGU....CGAAC.GGUUACCGAAUG
     Ssp-2-1   CACCGUAUUUGGCAAACU..AA.UUGAAA...AAU..UA.G.GACGC.AAAGCCU..C.CGGUCUAAGGUUG........UAA............AGGGAUG( 0 nucs)............U.UG.AAACUAAGAU....AGCGG.GGUUGCCUCAGA
     Ssp-3-1   CACCGUAUUUGGCAAACU..AA.UUGAAA...AAU..UA.G.GACGC.AAAGCCU..C.CGGUCUAAGGUUG........UAA............AGGGAUG( 0 nucs)............U.UG.AAACUAAGAU....AGCGG.GGUUGCCUGAGA
     Ssp-1-1   CACCGUAUUUGGCAAACU..AA.UUGAAA...AAU..UA.G.GACGC.AAAGCCU..C.CGGUCUAAGGUUG.......................UAAAGGG( 6 nucs).................UAACUAAGAU....AGCGG.GGUUGCCUCUGA
     Son-1-1   CACCGUAUUGGGCAAACU..AA.UUGAAA...AAU..UA.G.GACGC.AAAGCCU..C.CGGUCUAAGGUUG.......................UAAAGGG( 4 nucs).................UAACUAAGAU....AGCGG.GGUUGCUUAAGG
     Ssp-1-2   UAGUGUAAAAGGCAAACC..AA.UCGAAA...GAU..UG.G.GACGC.AAAGCCU..C.CGGUCUAAGGGUU........CG.............UU.....( 0 nucs)..............UG.UACCUAUGAU....AGCGG.GGAUACCACAGG
     Sde-2-1   UGCUGUGGAAGGCAAACC..AGUUUUAAA...GAC..UG.G.GACGC.AAAGCCU..C.CGAUCUAAAGGU........................UUGCUUG( 1 nucs)..................ACCUAUGAU....AGCGG.GGAUCCCACAGG
     Ssp-2-3   UAUUGUGAAAGGCAAACC..AA.UUGAAA...GAU..UG.G.GACGC.AAAGCCU..C.CGGUCUAAGGGU........................CGUUUGU( 0 nucs)..................ACCUAUGAU....AGCGG.GGAUACCGCAGG
     Ssp-3-2   UAUUGUGAAAGGCAAACC..AA.UUGAAA...GAU..UG.G.GACGC.AAAGCCU..C.CGGUCUAAGGGU........................CGUUUGU( 0 nucs)..................ACCUAUGAU....AGCGG.GGAUACCGCAGG
     Sba-1-1   GCGUGUAAAAGGCAAACC..AG.UUGAAA...AAC..UG.G.GACGC.AAAGCCU..C.CGGUCUAAGGGUU........CG.............UC.....( 0 nucs)..............UG.UACCUAUGAU....AGCGG.GGAUACCACAGG
     Sba-2-1   GCGUGUAAAAGGCAAACC..AG.UUGAAA...AAC..UG.G.GACGC.AAAGCCU..C.CGGUCUAAGGGUU........CG.............UC.....( 0 nucs)..............UG.UACCUAUGAU....AGCGG.GGAUACCACAGG
     Sba-1-2   UACCGUAUUUGGCAAACU..AA.UUGAAA...AGU..UA.G.GACGC.AAAGCUU..C.CGGUCUAAGGUU........................GUAAAUG(12 nucs)..................AACUAAGAU....AGCGG.AGUUGCCGCAGA
     Sba-2-2   UACCGUAUUUGGCAAACU..AA.UUGAAA...AGU..UA.G.GACGC.AAAGCUU..C.CGGUCUAAGGUU........................GUAAAUG(12 nucs)..................AACUAAGAU....AGCGG.AGUUGCCGCAGA
     Ssp-5-1   CUGACGAACAGGUACACC..CG.UUGUGA...GAC..GG.G.GUCGC.AAAGCUU..C.GGGUC....UUGA........U..............UUUUU..( 0 nucs)...............A.UUAA...GAU....AGCCG.GGUUGCCGAAUC
     Ssp-2-5   CUGACGAACAGGCACACC..CG.UUGUGA...AAC..GG.G.GUCGC.AAAGCUU..C.GGAUC....UUAU........UGUACA.........GGC....( 0 nucs).........UGU..UA.UUAA...GAU....GGCUG.AGUUGCCGAAUC

               ...........<<...<<..<<.<<.......>>>..>>.>....>>...<<<<...<.<<<<<....<<<<........<<<<<<<<<<<<<<<.......         >>>>>>>>>>>>>.>>.>>>>...>>>......>>>.>>>>........
               ...........22...22..22.22.......222..22.2....22...??22...?.22?22....222?........???2??2?2?21000.......         00012?2?2??2?.??.?222...22?......22?.22??........
               •••••••AR•R•YAAACY--••-•YG•RA---R••--•G-G-GRCG•-AAAGCY•--Y-RGGYC○○○○•●○○-----------------------••●○---         -----------------○○●•○○○GRY----RGCYR-GGYUGCCRRR••
                                        GNRA                                   5-TR                                                                      3TR
```

```
   Ssp-3-3  CUGACGAACAGGCACACC..CG.UUGUGA...AAC..GG.G.GUCGC.AAAGCUU..C.GGAUC....UUAU........UGUACA.........GGC....( 0 nucs).........UGU..UA.UUAA...GAU....GGCUG.AGUUGCCGAAUC
   Ssp-4-1  CUCUCGAAACGGCAAGCA..UU.AAGUAA...UUA..AU.G.UACGC.AAAGCCA..C.AGGCCCUAAUUGC.......................AUA....( 0 nucs).................GCAAAAGGGU....GGCUG.GGUUACCGAAUG
   Ppr-3-1  GUAUGUAAAGGGCAAACU..AG.UCGAAA...GGC..UA.G.GACGC.AAAGCUA..C.CGGUUUAAGGGUU.......................AUUAA..( 0 nucs).................GACCUAUGAC....AGCGG.AGCCGCCAUAGC
   Ppr-2-1  GUAUGUAAAGGGCAAACU..AG.UCGAAA...GGC..UA.G.GACGC.AAAGCUA..C.CGGUUUAAGGGUU.......................ACACA..( 0 nucs).................GACCUAUGAC....AGCGG.AGCCGCCAUAGC
   Vsp-3-1  ACGCACUUCUGGCAAACU..AU.UUGAAA...GAA..UA.G.GACGC.AAAGCCA..C.CGGUCUGUCGACA.......................AUUAUUC( 0 nucs).................UGUCUAUGAU....AGCGG.GGCCGCCACUUA
   Ppr-3-2  CCACAUUGUUGGCAAACU..AU.CCUAAA...GGG..UA.G.GACGC.AAAGCUU..C.CGGUCUGUCGAC........................GUCA...( 0 nucs)..................GUCUAUGAU....AGCGG.GGUUACCGUCUU
   Ppr-2-2  CCACAUUAUUGGCAAACU..AU.CCUAAA...GGG..UA.G.GACGC.AAAGCUU..C.CGGUCUGUCGAC........................GCCA...( 0 nucs)..................GUCUAUGAU....AGCGG.GGUUACCGUCUU
24 Vch-2-1  UCACGCACAGGGCAAACC..AU.UCGAAA...GAG..UG.G.GACGC.AAAGCCU..C.CGGCCUAAACCA........................GAAGACA( 0 nucs)..................UGGUA.GGU....AGCGG.GGUUACCGAUGG
   Vch-3-1  UCACGCACAGGGCAAACC..AU.UCGAAA...GAG..UG.G.GACGC.AAAGCCU..C.CGGCCUAAACCA........................GAAGACG( 0 nucs)..................UGGUA.GGU....AGCGG.GGUUACCGAUGG
13 Vpa-1-1  UCACGCACAGGGCAAACC..AU.CCGAAA...GGG..UG.G.GACGC.AAAGCCU..C.CGGCCUGAACCAU........A..............GUAGUA.( 0 nucs)...............U.AUGGAA.GGU....AGCGG.GGUUACCGAUGG
   Vsp-1-1  UCACGCACAGGGCAAACC..AU.UCGAAA...GGG..UG.G.GACGC.AAAGCCU..C.CGGCCUGAACCAU........A..............AUACUA.( 0 nucs)...............U.AUGGAA.GGU....AGCGG.GGUUACCGAUGG
   Val-1-1  UCACGCACAGGGCAAACC..AU.UCGAAA...GAG..UG.G.GACGC.AAAGCCU..C.CGGCCUGAACCAU........AUU............AC.....( 0 nucs)............G.AU.AUGGAA.GGU....AGCGG.GGUUACCGAUGG
   Vvu-2-1  UCACGCACAGGGCAAACC..AU.UUGAAA...AAG..UG.G.GACGC.AAAGCCU..C.CGGCCUAAACCAG........A..............CGUUU..( 0 nucs)...............U.UUGGUA.GGU....AGCGG.GGUUACCGAUGG
   Vsp-3-2  UCACGCACAGGGCAAAUC..AC.UUGAAA...GAG..UG.A.GACGC.AAAGCUU..C.CGGCCUAAACCA........................CUCGU..( 0 nucs)..................UGGUAUGGU....AGCGG.GGUUACCGAUGG
   Vsp-2-1  UCACGCACAGGGCAAACC..AC.UUGAAA...AAG..UG.G.GACGC.AAAGCUU..C.CGGCCUAAACCA........................CUCGU..( 0 nucs)..................UGGUAUGGU....AGCGG.GGUUACCGAUGG
   Vfi-1-1  UCACGAACAUGGCAAACU..AC.AUGAAA...AUG..UA.G.GACGC.AAAGCUU..C.CGGCCUAAGUAU........................UU.....( 0 nucs)..................AUAUAAGGU....AGCGG.GGUUGCCGACGG
   Van-1-1  GUCACGAACUGGCAAAUU..AA.UUGAAA...GAU..UA.A.GACGC.AAAACCA..C.CGGCCUAAAAAUU........AU.............GUUA...( 0 nucs)..............AU.AAUUCAAGGU....AGCGG.GGUUACCGAAGG
   Psp-1-1  GUCACGAACUGGCAAAUU..AA.UUGAAA...GAU..UA.A.GACGC.AAAACCA..C.CGGCCUAAAAAUU........AU.............GUUA...( 0 nucs)..............AU.AAUUCAAGGU....AGCGG.GGUUACCGAAGG
   Vvu-2-2  AUCACACAUAGGCAAACC..UU.CUGAAA...AGA..UG.G.GACGC.AAAGCUU..C.CGGUCUGUCGGUA........U..............UUUUAUC( 0 nucs)...............A.UACCCAAGAU....AGCGG.GGCCGCCGCUCU
   Vvu-1-2  AUCACACAUAGGCAAACC..UU.CUGAAA...AGA..UG.G.GACGC.AAAGCUU..C.CGGUCUGUCGGUA........U..............UUUUAUC( 0 nucs)...............A.UACCCAAGAU....AGCGG.GGCCGCCGCUCU
   Val-1-2  AUCACAUUUCGGCAAACC..AU.UCGAAA...GGG..UG.G.GACGC.AAAGCUU..C.CGGUCUGUCAGCA........UUU............AAAAAUC( 5 nucs)............G.AA.UGCUCAUGAU....AGCGG.GGUUACCAUGAC
   Vpa-1-2  AUCACAUUUCGGCAAACC..AU.UCGAAA...GGA..UG.G.GACGC.AAAGCUU..C.CAGUCUGCCAGCA........CUG............CUUAACU( 9 nucs)............U.AG.UGCUUAAGAU....AACGG.GGUUGCCACUUU
10 Vch-2-2  GUCACACUUUGGCAAACC..CU.UUGAAA...AAA..UG.G.GACGC.AAAGCUU..C.CGGUCUGUCGGUC........AA.............GAAAGCU(12 nucs)..............UUACACCUAGGAU....AGCGG.GGUUACCAAUUC
   Vch-3-2  GUCACACUUUGGCAAACC..CU.UUGGAA...AAA..UG.G.GACGC.AAAGCUU..C.CGGUCUGUCGGUC........AAGA...........AAGCUCA( 8 nucs)...........UC.UUACACCUAGGAU....AGCGG.GGUUACCAAUUC
   Vch-5-2  GUCACACUUUGGCAAACC..CU.UUGAAA...AAA..UG.G.GACGC.AAAGCUU..C.CGGUCUGUCGGUC........AAGA...........AAGCUCA( 8 nucs)...........UC.UUACACUUAGGAU....AGCGG.GGUUACCAAUUC
   Vch-4-1  GUCACACUUUGGCAAACC..CU.UUGAAA...AAA..UG.G.GACGC.AAAGCCU..C.CGGUCUGUCGGUC........AAGA...........AAGCUCA( 8 nucs)...........UC.UUACACCUAGGAU....AGCGG.GGUUACCAAUUC
   Ppr-3-3  CACGAAACAAGGCAAACU..AU.UCGAAA...GAA..UA.G.GACGC.AAAGCUU..C.CGACCUAAAAGGUA.......UUAUU..........UGUGUCU( 5 nucs)..........AAU.AA.ACCUCAUGGU....AGCGG.GGUUGCCGAUGG
   Ppr-2-3  CACGAAACAAGGCAAACU..AU.UCGAAA...GAA..UA.G.GACGC.AAAGCUU..C.CGGCCUAAAAAGG........UAUUAUUU.......GUAUCUU( 0 nucs).......AAAUAA.UAACCCUCAUGGU....AGCGG.GGUUGCCGAUGG
19 env-1    CACCGUAUUUGGCAAACU..AA.UUGAAA...AAU..UA.G.GACGC.AAAGCCU..C.CGGUCUAAGGUUG........UAA............AGGGAUG( 0 nucs)............U.UG.AAACUAAGAU....AGCGG.GGUUGCCUAAGA
   env-2    CACCGUAUUUGGCAAACU..AA.UUGAAA...AAU..UA.G.GACGC.AAAGCCU..C.CGGUCUAAGGUUG.......................UAAAGGG( 6 nucs).................UAACUAAGAU....AGCGG.GGUUGCCUCUGA
   env-3    CACCGUAUUGGGCAAACU..AA.UUGAAA...AAU..UA.G.GACGC.AAAGCCU..C.CGGUCUAAGGUUG.......................UAAAAGG( 4 nucs).................UAACUAAGAU....AGCGG.GGUUGCUGUAGA
   env-4    CCAGGUAAAGGGCAAACC..GG.UCGAGA...UGC..CG.G.GACGC.AAUGCCU..C.CGGUCUUCAGAUC.......................GUGC...( 0 nucs).................GAUCCA.GAU....AGCGG.GGCUUCCCAGGG
   env-5    UUGUGCAAAAGGCAAACC..GG.AUGAAA...GUC..CG.G.GACGC.AAAGCUU..C.CGGUCUAAGGGU........................CAGCAGU( 0 nucs)..................ACCUAAGAU....AGCGG.GGCCACCACAGG
   env-6    UUGUGUGAAAGGCAAACC..AA.UCGAAA...GAU..UG.G.GACGC.AAAGCCU..C.CGGUCUAAGGGU........................UGUUAGU( 0 nucs)..................ACCUAUGAU....AGCGG.GGAUACCGCAGG
   env-7    UAGUGUAAAAGGCAAACC..AA.UCGAAA...GAU..UG.G.GACGC.AAAGCCU..C.CGGUCUAAGGGUU........CG.............UU.....( 0 nucs)..............UG.UACCUAUGAU....AGCGG.GGAUACCACAGG
 7 env-8    UAGAGUAAAGGGCAAACC..GG.UCGAAA...GGU..CG.G.GACGC.AAAGCCA..C.CGGUCUAAGGGUU.......................CAU....( 0 nucs).................AACCUAAGAC....AGCGG.AGCUGCCAUCUG
   env-9    GAAAUUAAAAGGCAAACU..GG.CCUAAA...GGU..CA.G.GACGC.AAAGCCU..C.CGGUCUAAGGAUC........C..............AAAA...( 0 nucs)...............G.GAUCUACGAU....AGCGG.GGUUGCAUUAGC
   env-10   GCGCUCGAAAGGCAAGUC..UC.CCGAAA...GGG..AGUA.CACGC.AAAGCUG..C.AGAUCUAAGGG.........................GAAA...( 0 nucs)...................CCUAUGAU....GGCUG.GGCUGCCGUAAU
 8 env-11   GUAGUCAUUUGGCAAACU..GG.UUGAAA...GGC..CA.G.GACGC.AAAGCCU..C.CGGUCUAAAGAC........................AUGUC..( 0 nucs)..................GUCCAGGAU....AGCGG.GGUUGCCACAUC
            ...........<<...<<..<<.<<.......>>>..>>.>....>>...<<<<...<.<<<<<....<<<<........<<<<<<<<<<<<<<<.......         >>>>>>>>>>>>>.>>.>>>>...>>>......>>>.>>>>........
            ...........22...22..22.22.......222..22.2....22...??22...?.22?22....222?........???2??2?2?21000.......         00012?2?2??2?.??.?222...22?......22?.22??........
            •••••••AR•R•YAAACY--••-•YG•RA---R••--•G-G-GRCG•-AAAGCY•--Y-RGGYC○○○○••○○-----------------------•••○---         -----------------○○••○○○GRY----RGCYR-GGYUGCCRRR••
                                     GNRA                                   5-TR                                                                3TR
```

alignment positions 161 · · · 307

```
   Sth-1-1  GGACGCCGCAGGUGAUGUACCUCGGCGUCUGCUGGAGGGUGCUUUUUCAUGGUCAAAUUUCGCAGUUAAAUGUAUGGGGUCCCGAUCCUUAUGCAACUCCAGAAGGUCUAUUGCGUUUAGGCCUAGAAGUUCCCGUGCAUGGAAUAA
   Bsp-1-1  UUUUUUUAGCAUGACACCUUGCCAGAUUACUAGAAAGGUGAAGAAAUCAUUGAAAAAAAUAGAAAAUCAUAAAUUGAAAACAACUAUUCAAGAACCAGGUGCCCCCCUGGUUGAAAAUAUACAUAUCUAUGAGAGCUUAUUCCUGAA
   Bce-5-1  GAUUGGAUUGGUUUUGACCAAUUUUUGAUAUUGGUCUUUUUUUAUUUACGUACCUUUAAAUGAGAAAUAAUAUCAAUUUAGGGGCUAAAACAUUAUAUUAUUUAACUAUACAAAAAAUAAGGAGGAAGUGCUUUUAUGAGGAAAUAU

            ...................................................................................................................................................
            ...................................................................................................................................................
            •••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••••
```

```
   Bce-1-1   GAUAGGGUUGGUCUUGACCAAUGCUUUUUCAUUGGUCUUUUUUUUAUUUGCGAACAAAUGAGAAACCUUAUCAGCUGUUGAUAAGGUUUCUAUUUUAAUAGUUUUCCUUUAUAUUAAUUUAAUUAAGAAAGGUGAUGUUGAUGAAAA
   Bce-4-1   GAUAGGGUUGGUCUUGACCAAUGCUUUUUCAUUGGUCUUUUUUUUAUUUGCAAAUAAAUGAGAAUUGUUAUCAAUUGUUAAUAAGGCUUAUAUUUUAAUAGUUUUUCUUUAUAUUGAGUUAAUUAAGAAAGGUGAUGUUAAUGAAAA
   Bth-2-1   GAUAGGGUUGGUCUUGACCAAUGCUUUUUCAUUGGUCUUUUUUUAUUUGUAAACAAAUGAGAAACCUUAUCAGUUGUUGAUAAGGUUUCUAUUUUAAUAGUUUUCCUUUAUAUUAAGUUAAUUAAGAAAGGUGAUGUUGAUGAAAAA
   Ban-4-1   GAUAGGGUUGGUCUUGACCAAUGCUUUUUCAUUGGUCUUUUUUUAUUUGCAAACAAAUGAGAAACCUUAUCAGUUGUUGAUAAGGUUUCUAUUUUAAUAGUUUUCCUUUAUAUUAAGUUAAUUAAGAAAGGUGAUGUUGAUGAAAAA
   Bce-3-1   GAUAGGGUUGGUCUUGACCAAUGCUUUUUCAUUGGUCUUUUUUUAUUUGCAAACAAAUGAGAAACCUUAUCAGUUGUUGAUAAGGUUUCUAUUUUAAUAGCUUUCCUUUAUAUUAAGUUAAUUAAGAAAGGUGAUGUUGAUGAAAAA
   Bce-2-1   GAUAGGGUUGGUCUUGGCCAAUGCUUUUACAUUGGCUUUUUUUUAUUCUCGAACAUAUUGAGAAACAUUUUCAAACAAAAUAUCCAUAUAUUUUGUUUUUUUGUUUGUGUUGAUUAAGGAAGGUGGAUAUUAAUGAAAAAAUAUUGG
   Bth-1-1   GAUAGGGUUGGUCUUGGCCAAUGCUUUUACAUUGGCUUUUUUUUAUUCUCGAACAUAUUGAGAAACAUCUUUAAACAAAAUAUUCAUAUAUUUUGUUUUUUUGUUUAGAAUCAUUAAGGAAGGUGGAUAUUAAUGAAAAAAUAUUGG
   Bwe-1-1   GAUAGGGUUGGUCUUGGCCAAUGCUUUUUCAUUGGUCUUUUUUUUUGGUUUAUAUUUGAGAAUUGUUAUCAAUUAAAGGUGUGUUUUGAAAUUAUUAUUUUGUAUUAACUGAAUUAAGGAAGGUAAUGUGAAUGAAAAAAUAUUGGC
   Bsp-1-2   AUCUUUAUGGACGAUAGGCUAAUCGGUUUUGGUUGGUCUAUUUUUUAUGCUGAAGGAUUGAGGAAAAUGAAGAGGGACAGCAAUGAUGAACAUACAAUUUGCACAAGGUUUUGGAUUAGGGGCACUGACUGUUUUAUUGAGUGUGAU
1  Bce-2-2   UUGAUGCAACCGGACGGUCUUUUUUUGUUAUAUAAAUAUAAAAAAUAGGAGGGAACUAUGCAAAUGAAAAGUACUAAAAGACAUAUUUCAAUGGUUUCAUUAGUUUUUCUAGUAUUAUUUUCAGUGCUGAAUGUGUGGGCUCCAGUU
1  Bth-1-2   UUGAUGCAACCGGACGGUCUUUUUUUGUUAUAUAAAUAUAAAAAAAUAGGAGGGAACUAUGCAAAUGAAAAGUACUAAAAGGCAUAUUUCAAUGGUUUCAUUAGUUUUUCUAGUAUUAUUUUCAGUGCUGAAUGUGUGGGCUCCAGU
1  Bwe-1-2   UUGAUGCAACCGGACGGUCUUUUUUUGUUAUAUAAAUAUAAAAAAAUUAGGAGGAAAUUAUGCAAACGAAAAGUACUAAAAGGCACAUUUCAAUGAUUUCAUUAGUUUUUCUAGUAUUGUUUUCAGUGCUGAAUGUAUGGGCUCCAGU
2  Bce-4-2   GAUGCUCGCACAUAUUGUUAUGUGUGGAAAUUGAAGAGGCAACCAGACUUUUUGUUUUUUUAUAAAGUGUUAUAAACGUUCCAUCUAUAUGGAAGUGGAUUACUAUACGAUUUAUAUACAGAAUAUUCAAAAAUAAGGAGGGAGCGA
2  Bth-1-3   GAUUUUCACACAUGUUGUGAUAUAUGAAAAUCGAAAAGGCAACCAGGCUUUUUAUUUUUUCAGAAUACACAUUAUGAUUGUAAAUUUAGUCGAUUAAUACAUUUAAUGUUCGGAAUAUUCGGUGUAUAAAAAGCGAGGGAGGGAAGU
2  Bth-1-4   GAUUUUCACACAUAGUUGUGUGUAUGAAAAUCGAAGAGGCAACUGGACUUUUUGUUGUCUCAAAAAAAAUAAAUAGGCACACUGUUCGAAAGGAUAGGUCGCAAAGCUAAGAGUCUAAGGUAAUGAAAAUUACUAUGAUAGUCUGGU
2  Bth-1-5   GAUUUUCACACAUAGUUGUAUGUAUGAAAAUCGAAGAGGCAACUGGACUUUUUGUUUUCUUAAAAAAAAAAUAAAUGGGCACACUGUUCGAAAGAAUAGGUCGCAAAGCUAAGAGUCUAAGGUAAUGAAAAUUACUAUGAUAGUCUGG
2  Bce-4-3   GAUUUUCACCCAUAUUAGUAUGGACGAAAAUCCAAGAAGCAACCAGACUUUUUAUUUUUUAAAAGAAAAAUAGCUAUACUACAACGAAAUAAAUCGGCACACUAUUCGAAAGGAUAGGUCGCAAAGCUAAGAGUCUAAGGUAAUACA
   Bsp-1-3   GAUAACGCGGAUAGCUUUGGAUUUUAUGCUUAGUCCUUUCAGAGCUAAUCGGCCGAACCUUUUGGCAUAGUGACUGAAAGGAAAGGGUGCAUAAUCAUGUAUUUUUUUGGUACGGAAGAAUUUUUUGAGGAGCAGUUAAGUAAAGCU
   Bsp-1-4   AGGGGAAACGGUUAUGGGUAUGAGUAAUGAAGUAGAAAAAGAGUGGAUCGAAUUAUUGCUGGAAGCAAGAAAUGUUGGUUUGACCUUGGGAGAAGUAGAGAGCUUUUUUAGAAGAGAACAUGAUUUUGGAUCAGGCAGGAAAAUAU
   Oih-1-1   GUAAGGUAGGGGACUAAAUCGUGGAUAAAACAUUACAUUCUGUUAAGGUUGAUCAAGAGUGGGUACGUUUAAUCAAACAAGCGAGAGAAUUGGGACUUACAAUCGAAGAGAUAAGACAGUUCAUAAUGGUUACAGGUCAGAAGGUGU
   Bsp-1-5   UUGUAAAGGAGAUGUCAAACCUUGAAAAGAAAUUGGACCACACCGGAAGAAAUCAAGAGGAAGAAACAGCUGAAAAAGCGACUGCUCUACAUUUCAAUCCCUGUAGUAACCGUACUGGCCAGUGCAGUUGUAACGAUACUCGCCAAC
21 Bsp-1-6   ACAGAAAAGGGGACUAUUUGAUGUCAAAAAUGAAGUGGGACACGCCUGAAGUUAUAAAACGGAGAAAAAAAAAGAAACAAAAUUAUAUUGUACAUAUCAAUUUUUGUAUUUACUAUUGUUUCAAGUGUGCUGGCAACAGUUUUAAUGA
   Bli-1-1   AACAGGGAGUUCGCCCGUUUUUAUUCGGGCGGGCUCUUUUCUUUUUAUUUCCAAUAUAAUGUUUUAUUGGAAACGACAAAUCUGUGACAGCGUUUUUCGCUCAUCGCAAAACCGCAACAUUGCAUUGCGGCUUGGCUGUUCGCAUCG
   Bsp-1-7   UAUCAUUCUUGUAAACUUUUAUAAAUCAUACACCCUUUCUAAUUUGCCUAAUCCAUUUAUUAAAGAGAGGGAGAUCAAUGUUUUACCCGACAUUUUUAUGCAAAGAUACUAAUCAGAGGCAUCUGGUCAAAGACUCGAAUUUCUGUG
20 Bsp-1-8   CCUUUUUAUGGAUGUCAUGGCAUCAUACAAAGGAGAGUAUACAAUGACAAUCAAAUUUUUUAAGACAGCUCUUUUGGCUGUACUGCUCGUGUUUGCAGCCUGCCCCGUGAAAGGAGAGGCUGCUGAUAUGAACCGCUCAACCUGGCU
   Bha-1-1   GGAUCACAAACGUUUAAUUGGCUAUCCUCAUUCACGUCUGGGGUAGCUUUUUAUUUAUUAGGCUUCGGUGGAAUAUGAUGAGUAUCGCAAACGAUAAGAAAAAGGAGAUGGAGAGAUGAUCGUAGUAGAUAAACGAAAAUAUAAAAU
   Bsp-1-9   UACGAAAUGUAUUUUAGGGGGAAGAGAAGGGUUAUGAAAAAUAAUAUGCAGGAAAUGGACAUGGAAUGGGUCGCACUAAUUACUGAGGCGAAGGAAAUAGGCCUGAAAAAAGAAGAUAUACUUAACUUUCUUGCCGGCCAGCCGGCA
   Tet-1-1   UUAUAUAUAAUCUCCUGCCUUUUGACAGGAGGUUUAUUAAUUGAGGAAGUGAUGUUUUAUAUGACUAAUAAAGAAAAAAUUUACAAUUUCAUUAUAGAAAAGAGUAAGGAAUUUAAUCUUCAGGAUUUUUUGAUAAAUAGAAAAGGU
   Tte-1-1   AUGUGUAAAAAUAAAAAUUGAUUUUCUGCGGCAGCCUUGGGUAUGGGCUGUCGCUUUUUAUGUUUUUUUCAAAACCACCAUGAGAAAAGGAGGAAAGUAAUGCCUAAAAAAAUAGUAGUUUUUAGCCUGGCAGAAGAACUUUAUGGA
   Mth-1-1   UGGAACGGGGCUACCCAGGGGGUAGCCUUUUAAAUUUUGUCCAGGCAGGAGGAGGGUUCUUAUGAGGUUUAAACUAUUUAAAAAAAAGGGACUGCAGGUAGAAGGGCCGGACCCCCGGGCUCAAAAGAAGGCCCGGGCCGGAGGGAA
   Mth-1-2   AGCCAAUUUGAAUAUUGCUGGCAAUUAGGGGCAACUACGGUUUUUGCGGGUUGUCUUAAAAUUUUAUAGGGGGUAUGCAGGAUGAAGUACAAAAAAUUAAUCAGUAUGCUGGCCAUAGCGAGCCUGCUUCUGCCGGCGGGGAUGGCU
1  Mth-1-3   CCGGGGUACUUCUUUUUGCUACCGUAACGGAAAGGAGGGAGCUUCCGGUGCGGAGGCUUUUAUUAUUUGGUUCUGAUACACGACCUGCUAAGGAACCAUGCGGGACAAUGCCAGAUGACAGUCAUGAUUAAUUCGGCUUUAAUUACA
   Mth-1-4   AGAGUUUUGCCUUUCCCUGCUACCGGUGGCUAUUUAUUUAUAUGUUUGGAGGGGAAGCAUGAACUUUGCCUACCGGGGAAGGGAUGCCAGGGGGAAUGCAGUUAGCGGGAAUCUAGAAGCGGAAAACCAGGAAGUGGCCAUCCAGGA
   Mth-1-5   UAAAGUCUUGAUUGAUAGGUUUUUAUAAAGGCAGCCUGGUUAUGUUAAAUGGCUGCCUUUUCUAUUUUUAUUCAGCCUCCGGGGGUUGACUGGCUUUUGGAGGUGAUUGUUCUUUGAAAGCCCAGGGUUCCUAGCCUUAACACCGGC
3  Hor-1-1   AGGUGUCAUUAUGAAGAGAAGGACAGCUCUUACUCUCAUGGUAGCUGUAUCUCUUUUACUAAUCAUUACGGUCUCAAUCCAGGCAGGUGGAAAAUCUGGACAAGUAGUCAGGGUAACAUCAACAGUUAAAGCCUAUCAGGAAUUAAC
   Mth-1-6   CAGGAAGCGGCCUAGUCUUGUUGAGGAGCAAGGCUUUUUUUAUUACCGGAUUUAGCAGUGCUGGGAGAGGAUGACUAUGGUUUGGAAUGGCAGGCGAUACAUAACUGGUUGGGUCGGCUGGCGGCCCAUUCUCCUUUUACCUAUCGC
1  Mth-1-7   CGGUUAAAUUGUUAUGCCCGGAAGAGGGCCUUUAUUUUUAGGCAUGGGACUUUAGUCCCGGGGCAAUUGGGCCUUUUGUCCGAUACCAAGAUUACCACGAAAGUGAUAAAGUGGUAACUGUUGGAGAAUAGCAUCGAUUUUUCAGGG
   Hor-1-2   UAGGAGAUAGUGGAAAAACCUGCCGGGUCUGGGCAGGUUUUUUCUAUUCCUUCUACCUCCGGAAAUGGUAAACCCAUCGAAAGGUGGGGACACAAAGUCACGGGCCUGUAGCCGGUUAUUUUUAACCGGUAUGGCAGCCGGACUGCC
3  Cph-1-1   GAGAGUUUUUAUGAGCUCAAACCAAAUUAUAAAGAAAUUAUUUCUCUUCCUGUUUUCUUUUUUAUUAUUAUUUAGCUCUGUGAAAGGACAAGCAAGUGUCAUCGUCGAGUUAGAUGAAAGCCAUUUGUCAGAAGGCUAUGUAUCAAC
   Cdi-1-1   GAUAAUAAUUCUUUUUGUUAAAAGUAUUUUUAGAGUUAUAUAUUUUAAUAACUUAAUAUUUGGCAAGAGGAGGGCUUAAGAUGAUAUCUGAAUUAAUGUUAAAAAAUUUUGGCAAAUCUAGCAUCAAUUUUAAAACUAAAAUUAAAU
   Cpe-3-1   GUGCAUUAAGCUUUUUGUUUUAUAUUAAUUAUGGAGGAGGGUGUAGGUUAAAAAUUUAAAAUUUAUUUUGUACAAGCUUAUAUUAAAUUGUAAGUAGAAAUGCACUUCUGAGGGGAAAGUCAAUCAGGAGGAGAGAAAAAAUGAAUA
   Cpe-2-1   GUGCAUUAAGCUUUUUGUUUUAUAUUAAUUAUUGGAGGAGGGUGUAGGUUAAAAAUUUAUUUUGUACAAGCUUAUAUUAAAUUAUAAGUAGAAAUGCACUUCUGAGGGGAAAGUCAAUCAGGAGGAUAGAAAAAAUGAAUAAGAGAA
   Cdi-1-2   GAGUUCUAGGUAUAUUGUAUACGUAGGAAGAGUCUAGGAUGACUGCCUUUUUGGUGGUCAUUUUGUUAUGGAAAAAAUCACAUUAUGUAGUAAAAAACACCUAUAAUAACAUAUUAUACCAAAUAUAAUUUAAAAUAAAGAUAAAUU
   Cdi-1-3   GAUAUUUUCUUUUUGUUUUUAUGAAUGUUUUUUAGGGGGAUAUAUGAUAUAUGUUUAUGGAUAAAAUGGAUAGAUGUACACAUAUACUAACUGCAUAUAUAUGCAGUUCACGCGAUUACUGCAAUUUUAUAGAUACACAAUUAAAUG
   Cdi-1-4   GAUAGUGUUCUUUUUGUUUUUAUGAAAGUUUUUUAGGGGGAUAAUAAUUUAUAUGUUUGGAAAUAAAAUGGAUAAAUGUACACAUGUAUUGACUGCUUAUAUUAGCGGUUCGUAUGAUUAUUGUAAUUUUUUAGAUACACAAUUAGA
   Cdi-1-5   GAUAUUUUCUUUUUGUUUUUCAUGAAAGUUUUUUAGGGGGAUAAAUACAUUAUGUUUAGAGAUAAAAUGGAUAAAUGUACACAUAUGCUAACUGCUUAUAUUGGUAGUUCAUAUGAUUAUUGUGAUUUUAUAGAUACACAGUUAGAU
   Cdi-1-6   GAGUAAAGUCUUUUUAUUUUAUAAUUAAGAUUCUCUUUGAGUGUCUUUUUUUAUUGUCUGAAACUUAUAGUAUUUAAAUUCAUGAAAAAUUAGAGAUUUAUUAAGGGAGGAUUGUUGUUUCAUGAAAAAGGCAAUAUCUUGUGUACU
   Cdi-1-7   GAGUAAAAAUUUCUUUUUGUAUAUACAAAGAUACUCUUGGGUAUCUUUUUUAUUUUGAAAUUUAUUACUCCCGGCCUAACUCAAAGGAGGGAAUAAUAUGAGACCAAGUAAAAAAUUAUUAAUAGCUAUAAUAUCAAUAUUCUUAAU
   Cpe-1-2   AUUAAUAUUCUUUGUUUUAAUAGAGUUAAUUUAUAAAUAUAUACUUAAGUAUAUAUUUAUUUUUAGAACUGUUUUUAUGCAACUAUCCUUUAAUUAAUAUUAAGGAUAGUUUUUUUUAGUAAAAAUACUUUAUUAAAUGGUAUAGGC
   Cth-1-1   AUAAUACUUACCCUUUUAGUAAAAAGAGGGUAAGUUUUUUAUUUUGCCUUAACGUUGCAACUGUAUUGAAGCACACCAUUUUAUAAUAAUUUACAAAGGAAGUGGAAAUAUGAAAAAACAUUUGCUAAUCAUCGUAUCAGCACUCGU

             ......................................................................................................................................................
             ......................................................................................................................................................
             ●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●
```

```
3Cte-1-1    GAGAAUUUUUAUGAAGAAGUUUUUUAGAGGAUUUCUAACUAUCGUUUUUUUAUUUUCAUAUUCAAACAUUUGUUAUGCUAAGAAUAUAAGUUCCAAUAUAGAAUUUAUUAAUAAAAAUAUUUCAGCAGUUAAUAGCUUAUGUGCUUC
 Ame-1-1    AGGAUUUCAAAGCCUUUUGUUUUCGAUUUUUUGAAAAUACAGAGGUUUUUUAUUUGGAAAAAAAUGAGAGGGAGAUAAAGUAUUUAGUAAAAUUAAGACAAUUGUGAGGCAUUAAUAAGAGAGACUAAAAGGGGGAUUAAAAUGAAA
13Chy-1-1   UGUUAAGGGGAUUAUUAACCCUCCGCUUUCUGGCAGGAGGGUUUAAAAUUUUAAAGGGGGUAUUUUAAUGAAGAAGAUAGCGGUUUUAGUUUUAACGGUAUUUUUACUUGGAUUACUUUACCUGCGUUGGCCGAUACUGCUACCAC
 Chy-1-2    UGUGGAUUUAGAGUUAAGCCUCCGCCUUUGGGCAGGAGGUUUUUAUAUUAAUUUAAAGGGAGGGGAUGGAGAAAUGAAAAUAAAUAAAAUCCUGGAAACUGUUUCUGAUUAUCAAAAAUUAUUUACUGAAGACGCAACGUUUAUAGU
 Chy-1-3    UGGAAAUUAAAAAAAAAACCCUCCGCCCGACCUUACGGCAGGAGGGUUUUAGUAUUUUCCAAAGUUCAAUGAGGAAAAGUUCGAAAUAUCUUUUGAAAGUAAGAACUAUGUUGUACAAUAAAAAAAUAAUGGCGAUAGAGGUGGAAA
13Chy-1-4   AUAAAAAACCUCCUCUGGCAGGAGGUUUUUAAUUUUAAAAAAGGUGGGGGGCACGGUGGUAAACUGGGAAUUAAAGUUAACCAACAAUAAUCUGGAGGCUUAUGUCAAAUAUUCUUUGCUGGAAGGGGAGAAAGGGGGGAAGUUCC
 Chy-1-5    UAGGAAUUAAAGCCUCCUUGGGCAGGAGGUUUUUAUUUUUUCUGAAAGCAAAGAUAUCUUGCUGAAUUUUUUUUUGGAAAAAGUUAAAAUAAAAUCGAGAAGGUAAAGGAAUUUAGCUAAUUUUGUCAAAUAUAUUAUAGAAAGUUA
 Chy-1-6    AUGAACUAAAAACCUCCUUGGGCAGGAGGUUUUUUGUUUUUGAGGAGGGAGAAUUGCUUGGCCCGAAAAGCCAAUAUCUUUGAAAAAAGUUUUCAUGCAUCAAGUCGCGCCUUAUUAUUAACCGACUGGGAAGGAAAUUUAAUUGCA
 Chy-1-7    GCUUGGGAGGCUUUAAGCUCUCUUCCGUGGGGGUAGAGAGCUUUUAUGUUUCAAUAGAGGAACGAAUUUCAAAGAUAAUGGAGGAAGGCAAGGUGUCUGCAAGCGAAAAAGAAAUUCAGGAACUUUUUUCUCAGUUGGUGAGUACA
 Cbe-1-1    GUAUCUAUGAAUAAUAAAUUUUAUAUUUUAAUUUUGGUGUCCACAAUUAUGUAGACACCAUUUUUGUGCUCUUAAAACUUACAAUAAGAAAGCUAUUUUUAUUAUUUUUUCAUGUGCAAAGUAUUUGCUGCUAGAGAUUCGAAGUUU
 Cbe-1-2    AUAAUUAUAAUAAAUUUUUUAUUUAUUUAAUAAUGGUUUUAUUUUGGUGCCUAUAAUUUUUAUAGACACUUUUUAUAUUGGGAAAAUAAUGUAGAAUCGGGUGAUAAUUAGUUUAUUUUAAAGGAUGAAGUUGCAGAAAAAUUUCUG
 Cbe-1-3    GUAGUUAUAUAAACAUUAGUCAUUUGUUUAUUUUUGGUGCCUAUUAAUAAUUAAUAGGUACCUUUUUAUUUUGAUACAAAAUCAUUGAAAUAUGUGUAUAUGGUAACAAUUUAAAUAAAGAAAGGUUAGGUGUAGUAAAUGAAAAAG
 Cph-1-2    GAAUCCAGGUGAAAGAUCUGGAUUUUUUUAUUAUUAAUAAGGAAAAGUUCGCAAAGUGAGCAUUUUCCUGUUUAUAGAAAUUGCUUUACCGAUUAUUAAUAAAUAUUUGUAAUGGAAAAGGACUUGCUUGGAGAAAUAUAUGCUAGA
 Ame-1-2    GCUAAUUAGUAAGUGUAAAAGCCAAUAUGCUUACUGAUUAGUUGUGUAUUAAGACAUCGGGCCCCUUUAAUAGAGGCCCAUUUUUUAUGAAGAGAUAAAGAUUUUUUAUUGAUAGUAAGUUCUCACUUACAUUCCGCAUAAAUAUUC
 Ame-1-3    ACAAACUCGAAAGCAUGAAAACCAAUAUGCUUUGUGAGUAUUUAUGUAUUGAGACACCUGGCCUCCCUUUAGAGAUCGGGUACUUUUUAAUGAAAAAAUAUAAUUGAGUCUAGUGAAGAUAAGGUCUUUAAAUUAUGAUGCACUCCA
 Ame-1-4    GCAAACUUGUAAGUAUAAAGAUUAAUAUGCUUUCUGAGUAUUUAUGUAUUGCGACACCUGGCCUUUUCCUAGAGGCCAGUUUUUUGUGGAAAAAAGGAAGAAAGUAAAUCAGUUUUCAGCAUAGAAGCAUAGAAGCAUAGAAGCAU
 Ame-1-5    GUGAAUUAGAAAGUAUGGAAACUAAGUAUGUUUUCUGAAAAUGCACACAUUAAGACACCUGUCUCCUUUAUCUAGAGACGGGUGUCUUUUGUGUGAAAAUCAGUGGAAGUAUGUAUAAAUUCAGUUUGUUGUUAGGAGAAAAUUUU
 Ame-1-6    AAAACUAGAAAUCAUAGAACAAGUAUGAUUUGUGGGAUUUGUUUAUGCAGGCACCUGGGUUUUUACAUAAAAAGUGGGUGUCUUUUUAUGUGGUUAAAUAAGAGUAAAGUAAGGAAUUUUAUUUAAAAUAAAAUAUGGGGGGGUUA
 Ame-1-7    UAGACUAAUGUAUCUAUUAGGAGAAGGCACCUUUAUCAAGGGGUCUUCUUUUUUAUAUGUUUUUUUAGGGUAUCAGCAAGUAAAACACUGUAAUCAUGAGGGAGAAGACAAUAUAAAGGGAGGUAUGUAUUACUUCGUGGUCUCAAA
 Dha-2-1    CUUAAGCUUUAAACACCUUUGUCGCCAGUUCACAUGGCGACUUUAAUUUUAUACUGUUUUUAUAAUUUUUAUUAAGAUGAUUUUGAAAGGAAGAACGUUCAAUGCGUCAGUUAAGCAUAAACAGGCUUAAGGAAGGGGAUAUUCUAG
 Dha-1-1    CUUAAGCUUUAAACACCUUUGUCGCCAGUUCACAUGGCGACUUUAAUUUUAUACUGUUUUUAUAAUUUUUAUUAAGAUGAUUUUGAAAGGAAGAACGUUCAAUGCGUCAGUUAAGCAUAAACAGGCUUAAGGAAGGGGAUAUUCUAG
 Csa-1-1    CUACUCGGGCUUAAUAGAGGGUUGAGAAAAGGUUAAAACGGCGUCUGGCGUGACGCCGUUUUAUUUUUGGAAUGAAAUUAAAAAAUGGUUCAUAGGUAUAUGUUAAAGUAUAAAGAAGAUAGUAAUCUAAAAUUUAUAUAGAUUCUU
 Csa-1-2    UAUGCCAAGAGAUGAGUGAGAUAUUUAAUUAUUGGCAAAUGGCAUCUGGCGACCGCCAGAUGCUUUUAUUUUUAUAUCCAUUGUUGUUUUAGUAGCAAAGAAUUAAUGAACAAGUGGGAAGUACGGAGAGAGUUAAGAGGAAAUAGA
 Dre-1-1    UGCUGAUUAACACUCCAUGACCCUCCUUUUAUAAGGAGGAUUUGCAUUUAUGACAAACGCUAGUCUCAUCAUUGAUAAAAAUGCAGCAACCCGAUACGACUUGAAGCUUCUUCUUUGGCAGUUUAAUUUUCUUACCUAUGAGUGUGG
 Dha-2-2    CCGCGCAUGGGCCUGCCCUCUCCAAUAGAGGGCAGGCCCAUGCCAUCUUUUUACGGCUUUUGACGGAUCAUUUGGAUAUCGGGAUACCGGCUUUCGCUCCGCACUGAGGCGAAAGGGGCAUCCCUAUGGGUAAUAACAGCCGCCCUC
 Dha-1-2    UUAAGCAAAGUCUGCCCUUUAUACUCGGAGGGCAGACUUUUGCGGCUUUUAACGGAUGCUUCAGGUAUCAGGAUACCGGCUUUCGCUGUGCUAUAGCGAGAGAGGUAUCCGCAUGAAGAUAUUUAGGUUUUGUCCCUUUCCGAGGUG
 Dha-1-3    UUAAGCAAAGUCUGCCCUUUAUACUCGGAGGGCAGACUUUUGCGCAAUUUUACAACUAUUAUCGGUUGUGUUGGGUGGGUGGGCGCUAUGAGAGAUGGCGAAAAAACUCAAUGUGAGGAUUUGAUAGAUAAUCUCCUAUGCGGUAGC
 Dha-2-3    CUUCUUUUUGCUGUCGCUAAAUGAAUGCGGCUUUUUUUAUUGUUUUUUUCAGCCAACUUUUAUCCUUGGUGAGCCGGUUUAGUAUAGUCAGAAACUCUAUGGAAAAGUAAGCUAUUCAAAUAAAAUAAUAAGAUAUUAUAGGGAGGA
 Dha-1-4    CUUCUUUUUGCUGUCGCUAAAUGAAUGCGGCUUUUUUUAUUGUUUUUUUCAGCCAACUUUUAUCCUUGGUGAGCCGGUUUAGUAUAGUCAGAAACUCUAUGGAAAAGCAAGCUAUUCAAAUAAAAUAAUAAGAUAUUAUAGGGAGGA
 Csa-1-3    AUAUGUCAAUGGUAAAGUGAGAAAAGGAGUAUUUACCUUUGGCAGAGAGGCACCUGGCUAAAAUGUCAGGUGCCUUUUGAUAUAUUAAUGAGCAAGGCUUUGCUUGCUCAACUUAAAGGGGGUGUAAAAAGUUAUGAAGCUUCAUG.
 Csa-1-4    AUAUAUCAAGGUGUCUGACAAUUAUAAAGUUAAUAAGCUGCUAAGACAAAAAGUUUCAGUGAAAAUGGUGAUAUAGCUUCUUUGACGGAGUAUCAUGUAAAAAUUUUACUUUGUCGGCACCAAUUUGAUAUUGAAAGGCACCUGGCG
 Dha-2-4    UUGCAAAGCUUGUUCUUUUUAGAACAGGUUUUUGUUAUAUAAGGCUCCGCAACGUUUGAACAAUGUUUCUUAGGCCCGGCUUUUGCCGCCUUUUGUUCUUGGUGAAAGGGGGCCUUUGUAUUCGUUGUCUCGGCUUGUGUUACCAAC
 Cbe-1-4    UUGCCUAUUUGAAAUGAAUAGGAGGAAAAUGAAAUGAAUGACAAAAUUUUAAUUUGUGAUGAUGCUAGAUUUAUGAGAGUUACUUUAAAGAAUUCCUUAGAAAAAGAGGGAUAUCAAGUAGUUGACGAGGCAGAAAAUGGAGAAGUG
 Dre-1-2    AAGAUCCAAAGCCCUAAGUUUUCAAACUUGUGUUUUGGAUUUUUUGCAUUUAAUAAUAAGUUUUAUAGUAGGGGAUUUGUUAUGAAAUUUACCCGUGGUUUUUUAGGUACCUGGUUAGUAAAUGAAGGUAUUAUUAGCCAGAAACAG
 Dha-2-5    UGCAAGCCUGUCUUUUAGACAGGCUUUUGUUGUAUAUGCUUUACAACAUUGAUUAAUGCUUCUCAGGUCUGCCUUUUGCUGACAUUUGUACUUGGCGAAAGGGGGCCUCAGGAUUCGUUGCCGCGCUUACCGUUAUCAAUUCAUCAG
9Dre-1-3    ACACCAAAUUAUUUUAUGGUGUAUUUUUUUUAUUAUUAAUUUUUGCCAGACAAAAUGAAGGAUUAAAAUUAAGGGGAGGAUAUCUUGCAUGUCUAUGGUUCCGAGUUCCAGAAUUAAACCAUCGGGAAGUCUUAAAUUUCAAGUGAG
 Ame-1-8    AAGUUAUAUUUUGUCAAAAUAGGUGCCCUUUAAGGCACUUAUUUUUUUGAAUAAAAUCUAUUGAAUUGAGGGGAGGCACCCCAUGCCAAAGUCAAAGCAUAAAGUGAAAAACCACAUUGUUAAUAAGUACAAAUUAGUAUACCUGCC
 Ame-1-9    AGUUUAUAUUUGAUAAAAAGUCGGCAACCACAGGUUUGCCGAUUUUAUUUUUUGUAAUGAUAGAGUAGUAAUGGUGACUAUAUAACGUAAGGCUUUUCAUCAAAUAUGAAAAGGGGGAUAACAAAUGGCUACAUUAGGGAUGGGGAU
 Dre-1-4    ACACCAAGCAGUGAGGGUAUGUUUUAUUUUUUACAGGGGAGAUGGGGCUAUGUACGGACAGGAUUGGCAAGAAAUCAUUAAUACAGCAAAGAGCUUCAAAAAUCUUUUAGGUCCAGAGGUAGUGGUUGUUGUGAGUGACAAUGAGAA
 Cth-1-2    UAAAAGUUGUAAGAUCAUUUUGAUCUUAAAGGUCACUUUGGUCUUAAUAGCUCUGUUUGGUAUACCCAGACAGAGCUUUAAUUUUUCUUAACACCUGCUUUCGUGUCGACUUAAAUUUACUCGCUUCAUUGUUACAUUUUCAAAACU
 Dre-1-5    AAAAUUUUAAUAGUCGGGGCUUUCGGGAGGUUUAAGGAAUGACUUAUUAUCAAAAGGCCCUUAAAGGCACAUUAGUAUCCGACAAAGGUGAGCAACGUACUAGAUUACAGAAGAUAGAAGAGUUAAAACAAAAAAUAGUUGGCGAGG
 Csp-1-1    AAUAUCUUUAUUAUUUAUGCAGUUUUUUUAUUAUAAAUGUUUCAAUUGAAAGAUUGGGUCUAUAAAUUUUAUGGCUACAUAAUUUCGUAAUCAAUAAUAAGCAGAUUCAAUGCUAUUAUUUUUUAAUAACCUAAAACACAAAAAGAA
3Cth-1-3    AGGGAGGAAUUGAUUUAUGAAAAAAAUAGUAAGCAGUAUUACUAUGCUCAUCUUAACUAUUGCCAUUUUUACAAAUCCGGGAUUUGCACAAACUACCACCGUCACAGAGUACAUAAUGUCCAAUGGAGUCUCUCUCCAAAAAAUCCC
1Dre-1-6    AAAAGGGCAGCCAUGCCCUUUUUCUUUGUUUAGCAAAUCAGAAAGCAUAAAACUUUCCGACUUGCAUAAGGGCAACUAAGGCUUACGCCAUCGCCUAAAGGCUUUCGUGCGGCGCAAGCCAAGUUUUCUUUAGGAAAUUAUGCUUUU
8Ame-1-10   UAAUUGAAGCCUUUAUUAUCAGAACAUUCUGUUGAUAGGGGUUCGUUUUUUUAAAAGAACUGGGGGGAUACACAUGAAGAGUAUUAAAACAAAACUAGUUAUCUUUUCUCUAUUUUGCUAUUGGU......................
8Cdi-1-8    GAUAUUGUUCUUUUUGUUUUA................................................................................................................................
1Ame-1-11   UAUACUUUAUGGCAGUUUUUUUAUAUGGAUUUUAGUAUUAAACUGUAGAACUCAAUCUUAAGCUUAAUGGAGAAUUUAAAAAUGUACUGUAAGGGCACACAAACAAACCCUUUAGAUUCAAAUAUAAUGAUUCUAACUAUAAUAAGA
 Cth-1-4    AAGUCCUGUUUAUACUCUUCAAAAUAAAUGUAAAGGAAAUAAGUCAGGUAUAUUUCCUUAUACCUACUUUUUUUUAGCAUUUUAUCCUGUCAAGAGUCUUGAUAAAAUCCAGCUUGGUGUAGUUUUUUUCAAGUCUUUCUUUCAUAU
 Dha-2-6    UUAGAACUUCUUAGGGGAGAGUAGUUCCUUUUUGGACUACUCUCUUUUUGUAUGCUUUUUUUGCCCUGUUGGGAAGAAUAAAGGGUUGAACAGCAGUAUAAGGCAGCCGGCAGUGUCGCCUUUUGUGGGGCGUGAAAAUAUAAGGAA
 Bma-1-1    GUGCGCUCGCGAAUGUUCGCGGCGGUAAGAACCACGGAGGCGAGAGUGAGGGACAUGGCGAAAUUCGUCGUAGGUUGAUUUUCAGUUGCAUUUUGCAGGCCGGCGUCGGAGACAACGUCGCGGCUGAAAAUCAAAAUCGCAUCCUUG

            ......................................................................................................................................................
            ......................................................................................................................................................
            ●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●
```

^{5}Rba-1-1	AAGAACUCUACGUCGUUUCGAUGUGAAACACGUUUGAGUUUGAACCGAGGCUUGUCAGUUUGAUUUCCAUUGACGAAGGCUGUUGUCACCUGGGUCGGUUCACAUCGAAUCGACUGACACUUUGUCGCCGGGUUAACGAGACAUAAU
Asp-1-1	AUGCAGUCGCUCUCCUCUCGGGUGGGGCUGUUUAUCUGCUGUCCGCAUUUCCUCGCGUGUUCCACGGAUCGGCCCGGAGAAACGCGAGGAAUCGAAGAUGAACGACACGCCCAUGAAUACCCGCCCCAACCGGUCCCGGCGGUUUC
Nmu-1-1	ACGGUUGAUCAGAUGAAGCCAGGCUUCAAGGAUCAGUCAGCCGGAUUUUUUUGGUGCAACCAUUAAUUACCUACGGAGUAUCUCGCAUGAAAAUAAAUUUCAGACUCAAGGCAUUCGCUGUUGCUGCCAUCGCGACUCUUUCCGGUC
Nmu-1-2	AUGCCAUUUUCUUUCUUGUCGAUUCUUUGGAUGACGAGAAGCUAUUAGAACGUGUUGAGUGGAUCUCGAUGCUCCGGCUCAGUACGUUUUUCGGCACCCGGCCUAUCUAGGCGGUUCCGCCUCCCCAAGAGGAGGAACUACGUGAAU
Tde-1-1	UGACCCGUGCUGGGUCCCCAGCGUGUGCAUGCCUGUCGGGAGUUCUUCGAACGCCAGGAGCCCCCAUGUCACGUCACAGCCCAUUCCCUGUCCACCUUUCCGCCACGCACCGCGCCGAGCGCGACGGCUCCGCCCGCCUCGGCCUAA
Neu-1-1	UCGCAUCCGAUUUUCCAUUCGGGCAUGCGUGCUGAUGAGUCGGCUUUCUAUAACCCAUCUUCUCAAGGAGCCGUGCAUGGCAAAUGCAAAACCAGUUUGCUCACUCUUUCACUCUACUUUUCAUGAUGGUAUUGCCAGUGUUUUUCA
Neu-2-1	UCGCAUCCACUCCCUUUCUUCAUACAGGUAUGCGUACUGCGGGUCGGCUUCGUUAUAACCUAUCUUUCCAAGGAGCCGCGCAUGUCAAAUGCAAGAUCAAUUCAUUCGCCAUUUUCAGAUCACUUAUAUCAUUGCUUCAGUUUUUCA
Rfe-1-1	CGGGUCCAGUCCAACUGGCGCCCUCUAACUUCGCCGACCUGGCGAAGUCGGCAACCAUUGCCGCAUACGCUUUGGUCCGAAGCCCUGCCCCCCUUCUUUCUGCCUGAUGCGCCCUUGACUGUGAGCCGUCGAACCCGUGACCGUCAU
Rso-1-1	UUCGAUCAUCGACGGCGGGCCGAUUGACGGACGCAGACCAUCGCGUUCAUGUCAGGAGGCACACACUAUGAAAAAUAAGUUUGUCGGGGGGCUUGUGCUUACAUCUUCUACCGCGCUCGGCGCGGAAUAUCUUGUCGCGCGCGCUC.
Mfl-1-1	UGAGGAUCCCGCCUCGCCGAUCCGGCCUGGUGCGUCAAUACUAGAUAUAUCCCAGCGUCUCGCGUGAUAUACGCCCGGCACUUGCCCCUUAGCAAUAAAAAUACCGAGUACAUAUUGUGAUAGGGGAACCAUACUUGCUGCCAUCCC
Nmu-1-3	AUCAAAAAGCAGACUUUUCCUCAAUUUUCCGGAGGUUUUCAAUGUUCAAGCAGAAAAAAAUAGGCGCUCUUGUUUCAGGUGCGCUCUUGAUGGCGACCGGCACGACCGGAGCCAUUGCGGCGAACACGCAUUUCAGUAGCUUUACCC
^{13}Sfu-1-1	AGGUUUCCCAUGUUCAGAACUUUCAACUCCGGAACGUUGUUUCCCAAGAGAACCUUUCAGUCUUCAUGUCAGCCCAUUCGGGGCUCCACUGAAACCGGCCGUGAGAAGUGACUCGUAGAAUCUCAACCGAUUCAUUUGAGAGAAGAU
Sfu-1-2	AGCACGGUAUGAACGACUUCUUCAGCGCAUCGCACAUCAUUAUCGGAAAAAGCGGAACGAUGUGCCCUGAUGCCGUCCUCGCCAUUCGGUCAUUUUUCUCAUCGGAAUACAGCACCACGACCUUGCGGCGGUCUUUGCAUCCCGGGA
^{1}Sfu-1-3	GACGGCAGCCCGGCUUUUCUGUUUGUGGACGGGGUGCUUCAAGCGUCGAACAGGCGGAGAGGUUACAAGGAUGCGAACAAGGAAUCUGUUAGGUCUGAAACGGUGAGAACACCCAUAAUGCAAGGAGGUGGACAGAGAUGAAAAAGC
^{1}Sfu-1-4	AAGGGUAGCCCGGCUUUUUGUUUCUUCGUCAUCGAGUCGAGAAGAUUUUGUGUCAAGAAGCUUUUGAUGCUGAUCCCACAAACUCCGGACGGAUUCAACAGGCAAUGAAUCCCAGCCGCAACGCGUCACGGACGCGACAAAAGGAGA
^{8}Sfu-1-5	AAGCCACCACACAUUUGCUUGUAAUACAGUAUUGUCUGAGACCUUCGCCGCAGUCAUUUGAGACUGUGCGUUGCGAGGAUGUGCAGAGGCGGAUUCUGUCCGGCGCGGAGGGGAGCCGGGCCUGUACAUGUUUGCAUGGCGGAGCUU
Bba-1-1	UCAUGAGUCACCUGAGGAAUCAGAGGUGUGCAUAUGGAAGGAACAAAUACAUACAAGGAUUGGUUUCAUUCCCCUUAGCUAGCGAGUUAUCACUGACGAUUCCUCGCUAGCACUUUUUUUGGCGAACGAAUUUCUUAAGACAAAAUC
Dde-1-1	UCAUGCCGCUGCCGUUUCUCAGGGGGAAACUAUGGCCAACGUCAUGAUUGUGGACAGGCAAUCCGAUUUUAUACGGAAUAUGGAGCGGACGCUUCUUUCCGAAGGGCACACGUUGUCCAUCGCCGGUUCGGUGACGGAGUGCCUUGA
Sfu-1-6	UGCCGUAUGACCGAUCCAAAUGUGUCCUGAAACAGGUGAACCAGCCCCGUUGACAUGCCAAUUGUGUGUUGGAGCGCGACUCCGCAGUGUUAUGAGAGCGUGGGUUGCUCCGCAAGUUCUUCGGAGCGCGACCCAAUUCCUGGGUAC
Sfu-1-7	CUCUUAUAUAAUGAGCGGUGGGAAGGUCUGUCCUCGUAACCGCACGCUCUGCCAUGUGCUCCACCCGCCAUCUCUCGCUCCGACCUCUCUGUUGUACAGGCGGUUACCCGACGUCCGGGAAGAAAGCUGGAUAGGGUUCGCGUCCGG
Ade-1-1	GAGACCGGUCCUCCGGCCCAGCCCACGUCUGCCCUGAGCACGCGUGGGCUUCGUCGUUUCGACGGGGCGAUCGCGGUGAGGACGGACCAGCCAGGAGAGGCGUCAAUGAGCGUGAAGCACGACAAGGGGCACGAGGCCGGGGCGACG
^{12}Mxa-1-1	AAGUCAGCACCGGACAGGUUCCGGGCCCACUGCGCCUUGGGCGGGCAUGGGUCUUCAUUUCGGAACCCCCGGAGGUCCUAUCCGAUGCUGCGGACACUGGCCCUGAGUUGCUGCCUGUUUUCCCCGCUCGUCUUCGCCUCCACACC
Gsu-1-1	GUCAGCACGACGACGUUCGGCCCCGGCUUGCUCUUUAUCCGGUGCCGGAAGGAGACGAGCCAUGGGAGUUUCUCGGGAGUGUUGUCGCAGAUGGGGCAAGGCCGUUGUUGCGCAGGUUGUACUGUUCGUCGUUGCCGUUGUGGCGAU
^{2}Gsu-1-2	AUCACCACCGUGAUAGUCGGCCCCGGCUUUUCUUUUGCCCGGGUCCGAUGCCGGAGCGAGGCAGCGACGCAUGACCACCGUGAAAUUAUCCUCAAGUUUUCCUGUUGUUCCGCCGAUAGAUACACUAUUGAUGUCAAUUUCGCAUAA
^{1}Gsu-1-3	AUCACCACAAAUUCGGCAUCCCGUUUUUUUUUGCGUGGAUGUCCAGACGAAACGGGGAAAGGAGUGGACAAUGGUACGCGCGAAAAAGUCAGCUAUGGGAGUACUGGGGAUUCUUGCCGCCGCGCUUGCGGCACUGAUGCUGAUGCA
^{1}Gme-1-1	GUCACCACAUUUUCGACAGCCCGGUUUUUUUGUGCUAUGGCACUGAAAAAUCGGGGAAAGGAGAUUCAUGAUGGUACAGAUGAGAUCCGCUCUGAAAACAAUGGUGUUCUUUGCUGUCGCGACGCUGGCUCUGGUCCAGUUCCAGGG
^{2}Gsu-1-4	AUCUGAACGAUAUCAGGCCCCGGCUUUUUUGUGCCCGGGCUGCCGGAUCGUUCUGAUCCGGUACACUACCACGCGCUCAGGGAGGAAACUUAAUGACCGCACGUAAAGGGCUAAUUUCGCUCCAGACACAACAGGGCACGCUGCCC.
^{2}Gsu-1-5	AUCACAGAAUAUUCGGCCCCGGCUUGUUUAUUUUCAUGAGACAGAAAGGUACCCGUGUAACCCAUGCGCAGAAUCGUUACGCUGCCAACCGUGGCCUACUGGCUUCGUUCCGCUGUCCUGGGCAUCUGCGUCAUUGCCCUCACGGGC
^{2}Gur-1-1	AUCACACGAUAUUCGGUCCCGGCUUUUUUGUUGUCAAAACAUAAAAAGGAGGUCGAGUCAGCAAGAGCUGCGCUUUUAAAGACGCAUAGAGCUGAGGGGGGGCCAUUUGUCCCGUGUCAAGGGCUGACUCUUUUAUUUCUUCUUUUA
^{2}Gme-1-2	AUCACACGAUAUUUGGUCCCGGCUUUUUAUUGUGCAAACCAUGACUGGAGGCCGAGCCGACAAGGAGUGCGUCUCUGACGUGGCGGUGCCUGUGGGCGCCGGAAAACUGCAACAAAAUAGUUGGUGGUAGACGAUAAUACUAAACC
^{2}Gur-1-2	AUCACACGAUAUUCGGUCCCGGCUUUUUUGUUGUCAAAACAUAAAAGGAGGUCGAGUCAGCAAGAGGUGCGCCUUUAAAGGGCACCGUAUUAUGACGGUUAGUGAUAAACGAUAAUACUCAACCAUCCGCGAGGGUGGGGCGGAAAG
^{2}Ppr-1-1	AUCACUGGAUAUUCGGGCCCGGCUUUUUUGUUGUCAAGACACCAUGGGGUUCGAGGCAGGACGAAAAGCACCGUAACAGGCGCAUCGCACUAUGCCGUUUCUCGUCGAUGACGAUAAACGGCCCGACAAUUGUUGGUGGUACACGA
^{2}Gsp-1-1	AUCACAAGAUAUUAGGUCCCGGCUUUUUUGUGGUCUAAACAUUGUUGAAAGGUGCCUCAUUUAACCCAUGUGCCUAACCGUUGACCUGUCAUUUUCAUCCAAUUUCCCGCUUCGCCUGAUGGCCCUGUGCAUGGUCAUCCUAUCUCU
^{2}Gsp-1-2	GUCAUCGAUAUUUCUGCCCCGGCUUUUCUUUUUCCGAAAGGCCGGGAACAAGGGGAAAUUCGCUGCUAGGAGGUUGUCAUGCUGAAGAAGGUAAGAAGUUUAUGGGGAUUGACAUUGGCGGUUUUGCUGGCACUGGCGGUUUCGCAG
^{2}Gur-1-3	AUCAAUCGUAUUUCUGCCCCGGUUUUUUUGUGUCCGCAACACCGGGCAGAGGUGGGGAACUUUCACCAGGAGGUAGUAGAUAUGAAAAGAAGGAUGCAGAAUGUAUUCGCCCUGGCAGCCCUCGUACUGCUGGGUGCCCUCGGCGUG
^{2,20}Gme-1-3	AUCACGCGAUAUUUGGUCCCGGCUUUUUGUGUCUUGACGCGGAACUUACCCAUUCUCGAGAAAGGUACCGAACUCAAUCCAUGUACCGAAUCGCUGACCUGCCCCUCUCUACCGCCUGUUUUCGCCUGAUCGCUCUAGGGGUAUGU
^{2}Gme-1-4	AUCGCAUCGAUAUUCCGACCCGGCUUUUUUGUGCCGGACCGGACGAAAAAGCGGCUCCAGUAGCCGGGACACGUCGAAAGGGGAGGAUGGAUGAUGAAGAAAAGGUUGGAACGCUUCGUUUGCUCGCUGUUUAUGGUCGUGCUCA.
^{1}Gur-1-4	AUCAGACGAUAUUCGGCAACCCGGCUUUUUUUUGUGGAAAAUCGGCAGAAGGGAGGUGGUGAAGCGGUACACGUACACGGUUACAUCUGUCUGCGGUGGAUGCAGGAAUCGCCGGUUUUCGCGGCAGACGGACAGGGUGAACUUCCG
^{1}Gur-1-5	AUCAUACGAUAUUCGGCGACCCGGCUUUUUUGUGGUCAAAAUUGGAUAAAGGGGGGUGAUAAAGCCGUAAGAUGUGACGCGCCUGCGAGGGCGUAAGGUUAUGUCGAGUUGUGAAUUACUGUUACUCGGUAAUAGUCCUUUAGUGGU
^{2}Gsu-1-6	AUCACAGGAUAUCGGCCCCGGCUUUUUGUUGCGCGAGUCAGUGAGAGGUGCGCCCGAACAGGCACCGGAUCGUGCCGGUUCCCAUCGACAACGAACACGGCACGACAACUGCUGGUGAUAAACGAUAAUACUAAACCAUCCGCGAG
^{2}Gur-1-6	AUCCCUCAGAUAUCCGGCCCCGGCUUUUUUGUAUUCAAAGCGGUACGAAUUCCGGGGCAGUAAAGGGGGUGAUAAGGAAAUAAUAAAGCUUGACGACGAAUUAGAGAGGAAGGAUUCAUGCAAAUCCACAUUCCAUCAACAUGUAUU
^{2}Gur-1-7	AUCCCUCAGAUAUUCGGCCCCGGCUUUUUUUAUUGCAAAAAUCACAGGGAGGUCAUAGACGCAGCAGGUGCGGUAAAGCUCACUGGAGCUGACAAUGAAUACCGGUUCGCAUCAUGAUGCCGUUGAAUCGGUACGAACAGUCAGGCG
^{1}Gsp-1-3	AUCAUGAAUAGGCGAUCCGGUUUUUUUGCGUCAAAAGUAGUCAAAUGACAUAAUUAUAUAAUUGUAUUUAUUUUUAUAGAAAAGAGAUUAAUUUCUUGACUCUACUUUCUUUUGUGUUAUUUUUUGUUUAAAAUUCACCGGAAGU
^{2}Gur-1-8	AUCACAAUAUUUCCGGCGCCCGGUUUUUUUUAUUUUGUUAACUAAAAAUAUAAAAGACUAUAAUUUUAUAUAAUUUAACAUAAUUAGCGUUAUUUGCUUGACAAUUCUUUUAUUUAUUGUAAUUUCAACAAAAAAUAAGUAAAAGCA
^{2}Gur-1-9	AUCUCAGGAUAUUCUGCCCGGCUUUUUUGCGUCCCCGUUCAUCGGGAAACCCUGUGUCAAAGAAAGGAGAAGCGUAUGCGAAGCAUUCUUGUAAGCAUGGUAGUUGCGGCAGUCGCGGCAGCAGGCCCUGCCCAGGCAGCAAACGU
^{2}Gsu-1-7	GUCGAUACGAUGUUUCCGUCCCGUUUUUUUGUGCCUGCGUGCUGCCGGUCGGGACGGCAAGACCCCUCUGUUCCGAAAGGAGAUCGUCAUGAAGAGAGCAGUCCUCGCGUUGUCCCUGGCCUUGACGGCCGUACCCUGCUUCGCCGC
^{1}Gme-1-5	AUCACCAGAUAUUCGCAACCCGGCUUUUUUGUUGACUGCAGCCUCAAAAGAGCCGGGAUGCGAAGAACCAGGAAAGGGGGAACCUUCUGUCACCAUGCUUCUGAUACUUAUAACCAUGGGACAGACACAGUUAAUGACGAGCAGGAC
^{1}Gur-1-10	AUCACGAGAUAUUCGGCAACCCGGUUUUUAUUUGCCCAUUUACCGGAUGUGGCCGAAAAAAGUAACUCGUGAACAAGGUUCGGAUCACCCCAUGACGGGCAGUGGGAUUGAGCCGGAAGGACAACAAAAAUUGCUGUGCCGUCUGGC
^{2}Gme-1-6	AUCUGACCAAAGAUAUAGGCCCCGGCUUUUUUUGUCCAUUUUGGGGGCCGCACAAAUUCAAGGAAGAACCGUCAAGGAGGACACCAAUAAUGACAGGAAGACUUUCAAGGCUCUCUCCCGAACCGGGGCUCGUCGAUUACUUCAAC
^{2}Gme-1-7	AUCUCACAGAUAUGCGGCCCCGGCUUUUUGUGUCCAAAAUCCUGGGUCGGAACCACAACCGGGAUAAAAGGAGAUCAUGCGCCAGAAAAGAAAAAUUCCCAGUACCGAUCGAUUGAGCGAAAGCAGAACGAGAACAUUUUUUGAGA
	....
	....
	●●

```
[1]Gur-1-11     GCCAGAUUGGUUUCGGCCCCGGCUAUUUUUGUUUUGGCGAUUCGAAUCACAUAAAACCAUCAUCACAUCAGCCUCUUCAUCUUUUUGAAGCGUUGGACGCAGCACCAUGCCAUCAAUGGUGAAAAGGAGUACAUAUGAAGAAACGUU
[2,13]Gsu-1-8   GCCUGAAAUGGUAUACGGCCCCGGCUUUUUUUUGUUUUUUUUCGGGAAGCCAACCGCAACGAAAAUAGACGGUAAAUCAUGUGGUUGAAUCUUUUUUGCGACAUGGCCGGUUUUGGUGAUACGAGUCACAACGGUUCUGCAAAACAG
[1]Ppr-1-2      GUCACCUUUUCUUCGGCCCCGGCUUUUUUGUGUCCGCCGCCCGUUUCAUUUUCGAUAGUAAAGAAAGGUACCGUUUUUCGCCCAUGCGCCGAACUGUUCACCACACCUUCGCACACAAUUGCUUUUGUCUCUUAGCAGUAUGGGGUU
[2]Gme-1-8      AUCACUCGAUGAUAGCGGUCCCGGCUUUUUUGUGCGCAAAAAUCCGGGCAACAACCAGCACAGGGGAGGAAUUAUUUCAAAAGAUGUCACAACGAGGAGAGCAGUUGUUUUCGGACGAAGAAACGUAUCUUCCAAGGAAUGGAAAAG
[2]Gsu-1-9      ACCAGUUCAUGGGUAGCGGCUCCCGGCAUUUUUUUUGUCUGAUAGGUCAAAUGGGGCUUUGGUGACACGAAUCACAUUUAAUUUUUAGCAGGUGGACGAAAGUAUAAACAAAUAAAAAAAUAUUCGCUCUGUUUCCGAAGAGGUAAA
[2]Gme-1-9      AUCAUUUUAGAAUGAUGCGGCCCCGGCUUUUUUGCGUCCAGCCACCUUGGGUGCUAUAGUGGUCAAGGUAAUCUUAUCGAGAAAGGAAAUUCUUUAUGAAAAAAUUUCUUACGGUCCUACCUUUGUUUUUUGCAACGGUGUCCCCGU
Gme-1-10        AUCAUUAUGAUAAUCGGCCCCGGCUGUUCUGUUUUCCCCGUCUCUAUCUUCAAUAUCGAUAGUCGAUAAUACUAAACCAUUCGUGAGAAUGGGGCGGAAAGCCUACAGGGUCUUACUGAGACAGCCGGGUUGCCGAAAUAUCACCAG
[2]Gsu-1-10     AUCGAGAUCGAUAUUCAGGCGACCCGGCUUUUUUUGUGGCGACGAUUAAGGAGGAAACACCAUGAAUAUGGGAAAAAAGAUACAGGAAAUGGCCGUGUGGCUCGAAUUUCUCGUGGGAUCGGGACUGGCGAUCUUCUUCCACCUGG.
[2]Gur-1-12     AUCUAGAUGGAUAUUCGGAGACCCGGCUUUUUUGCGUUCGAAAAAACUUGUAAAGUUAUUGGAGGAUUCUCAAACAUGAAUAAAAAGAUCAACAAUAUCGCAGUAUUAAUCGAAUUUAUCGCAGGCUCAGGGCUUGCCAUCUUUUU.
[2]Gsp-1-4      AUCUCUCUGGAUAUUGAAGCCCCGGCUUUUUUAUGCCAUGGGCCAACAUGAAUCAAGGAGAAGGAGAAAGCUUAUGAGAUUGGUGGCAUGGUUUACGGCGAUGGUACUGGUAGGUACUCUGAACGGUUGCGGAAGCGGCAGUGGCG
[2]Gsu-1-11     AUCCGAAGGAUAAGUGGCCCCGGCUUUUCUGCGCCCGCUGGACGGCGGGAAGCGGCAGACGCGGGGCGCCGGGAGGCUCGGGUGAUGGUUGGUACGGUCCGGAAGCUGUUGGCGGCGGUAGUUGUCAUUGUUGGUGCCCAAGCGGUC
[1]Gme-1-11     AUCACAAUAGAUAGGCGGCGACCCGGCUUUUUUGUUUUUCCCGGAUUGCCUUGAAAAUUUCCGUGGACGCAGACUUGUGCCGCGGCAAUAGCUUCAGGCGAUAGCGAUACUACUAAACCAUCCGCGAGGAUGGGGCGGAAAGCCUAC
Gur-1-13        ACCAACCAGGUAUAAAUGGCCCGGCAGACUUUAUUAUUCAUGAUGAAGCGCACGCCGGGUUGUGGGGAAAUCACAACACGGCUUUUUAUUGCCAGUAAGAUGGGAGAUAAAGAUGAAUAUUGGAAACGGAUCAAGAGGGCAACUUA.
[2]Ppr-1-3      AUCCAACUCAGAUAUUCGGCCCCGGCUUUUUAUUGCCUGCGACACAAUGUCGCGGCUCUUAUGACAACAGCAUACAGGAGGGUAUGAUGACGGGAAUAUGCUCAAUAGAAUCAAGUACUAAGAAUAUAUCGCGUGAUUCGGGAGG.
[2]Gme-1-12     UCUCAACAGGAUAUGCCGGCCCCGGCUUUUUCAUGACCGAAAGGGGCGUAACAUGAAAGGAGGAAACACCAUGACCAGCAAAAUUAUCGCAUCUAUCGCCUGUCUCGCCGCCCUGACAGGAACGGCCCACGCCGGAAACGUCGAC.
[1]Gme-1-13     AUCAAUUGAUAUAUAUGGCAACCCGGCUGUUUUAUUACCAUCAGGGAGGUAGAAGUAUGCAAAAGAAAAAGAUUUUGGUUGUAAGUCUCGUGGUCUGCGCAGUUGUAACUCUUGUAUUCUUGUCGCUACCUAAAAAAGGUGUGACCG
[1]Gme-1-14     AUUUUUUCGGCGACCCGGCUUUUUCGUUUUCUGGAACCACGCAGGAGGUACAAAUCAUGAACAAGAAAAUCCACAACAUGGCGGUAUUUAUCGAAUUCCUCGUCGGGUCCGGCUUGGCGAUCUUUUUCCACUGGGUACUCCGCCACG
[1,13]Gsu-1-12  AUCACACCACGAUAGGGCGGCGGCCCGGCUUUUUUUGUGCCGUGGCCGCGCCGGCCGGUGACGGAAGGGAGGUGACCACUAACCGAUAGAGACCGGCCAUGAUCUACACACGACACGGAUUCAGGCAUCCUAUUCAGGGAGGAAACU
[1]Gme-1-15     AUCACACCACAGAUAGGCGGCGACCCGGCUUUUUAUUUUGACAGGCUGGCAGCGGAAGGAGGUGAUGACAUCUCACAGAGAUCAAACUUCAUGGCUAUCUCCCUUAACGACAAACAUUAACUGAAUUCACGGAAACAACGGAGGACA
[2]Gsp-1-5      AUCACACCGAUAUUAUGGCGAACCGGCUUUUCUUUUGCCUUCGGAACUGGUGCCGGUUAUCGCAACGAAAGGAGGAUCAAUGAAAGAUUUAAGUAAAAAAAAUGAGUGCAAAGCAAAGGCGCACGAACUCAUUGCCGCCACCAUGAA.
[1]Gsp-1-6      AUCACAUUUCCGGCACCCGGCUUUCUUUUUUACAGCGGAUUGAAAAGCGGGCAAGGAGGUGAUUAAAGAGUUCUGGCCGCUGGAAUAACAAAUGGAGGUGUUGCACCAUGUCGAUUCGUCAUCGCUUGCUUCUGCUCUUCUCCCUGU
[2]Gur-1-14     AUCCCACAGGAUACGAGGCCCCGGUUUUUUUGCGUUCAGAUGCAAAAACCGGGAACGGCAUCUCACGGGCAAGAGGGCAAAGGAGAGGAGAAAAUGAGCGAGAAAUGGGCACGUUUUGUUGUUGUUCUGCUGCUGGUAUUGCUCGUU
[2]Gme-1-16     AUCACAAUUCGUGAUGCUCGGUCCCGGCAUUUCUUUUUUGGACCCAGAAAUACAGGUUGUCAAAAUAUCGGCAAGGAGUGGACGGAAAUUGUCAGUCAUUUGCUUCAAUUCAUGGAGUUUCGUUGCCAGAACGAUCGAAAUGUUUAC
[2]Gur-1-15     AUCUCUAAAAAAGAUAUCCGGUCCCGGCUUUUUUUUAUUUAACCAACCGGGCGGAAAUGAAAGGAGAAGAUUGAAUGAAAAAAGUGUUUUUAGUGACGGUUCUGACCAUGCUUGUUGGCUCCCAGCCGCUCCUGGCCGCAGAUGUCG
[1]Gur-1-16     UCACCUGAUAUCGGCGACCCGGCUUUUUUUUGUGCCCAGGGAGCUGCGACACGAAGAUCACUCAAAGGAGGUACACAAGCAGAUGGAAUCGGCAGGAUUGCAUUGAUAUCCGGUCUCGUGAAUUCACUUGGAGGGCAAACAGGCAGC.
[1]Gur-1-17     UCACCUGAUAUCGGCGACCCGGCUUUUUUUUGUGCCCAGGGAGCUGCGACACGAAGAUCACGCAAAGGAGGUGCACAAAGGCAGUAUUGCGUUAGGGUAACGAUCAGCAACACAUGGAAACAGGGAAAACGAUAAUACUAAACCAUC.
[2,14]Gsu-1-13  AUCUGAUGAUAUGAGGCUCCCGGCUUUUUAUUUUUUCUGUGCCAUUUCUGCUGGUUGGCUUGUUUGUUGCAUAAAUAAAACGCAAAAGUAACAAAAUAAAUACUGGAUACACGAUAAGACUAAACCGUCCGCGAGGGCGGGGCGGAA
[1]Ppr-1-4      AUCUCAGGAUAGGCAACCCGGCUUUUAUUUUCCUUUUACAGACCAACCGCGAACAUUCAGGCAUGAACUGCUACGACAGGGGGUAAAUGUCCAUGCUGUCCAGAAUAACCAUCAUCCUGCUCUGUUCCUUAUUGCUGACAUCUCCC
[2]Gur-1-18     AUCUCUCCUGAUAUUCGGUCCCGGCUUUUUCGUUUGUUGACACGCUAAACAGAGAAAGAUUAAAAACAAGAGGUGACAGUAUGAAUAUAUUUAAAGUAUUGAGCUCCCAGAAAGUUUGUCGUACCCGCAUGAUCAGUCAGGAAUUU
[1]Gur-1-19     GCUUAGAUGGCCCGGCCUUCUUUUUUUAUUCAUAAAAUAUCGCAAGCCGGGUUGUUGAGAAAAUCACAACGCGGCUUUUUAUUGCCAAGAGAAGGAGAGAGAAAGAUGAGUAUUGGAAAAGCAACAAGAAAAGGGCAUCUCGUGUC.
[1]Ppr-1-5      ACCUGUUGCAGGGAGUCGGCCCCGGCUUGUUUUUUGGAGGGGGUAUCGCUCCAUCCGCCGAACCCGCUUCAGGUUGUAUCCGUACCUUCGGGCGGUGUCACCCACAGGACGGAGGUUUCUCGCACGGCAGGAGCGGUCCUGUCGAC.
[6]Gur-1-20     AUCCCACGUAUACAACCUCCACAUUUUUUGUGCACAAAAUCUGCUGUGCGCACCCACGUAACAAGAAGCCAUUCACAAGCGUUAAAGCAUGCAGUGCUCACGACGAUCAAGGUAUGUUCCGAGCAAACAACGAAAGGAGUCAUGACA
[2]Gur-1-21     AAUGAAAACCUUUUUGCGCACUCGGUUUUUUUUAUUUCGUAGAGAAGUCAAACUGCCGAGGAAACGUUCAACAGCAAGUUAUUAAAGGAGUUCGCACAUGAAAUUCAUUUUUCCCUGUGUCGCUUCGUUCAUGCUUACUGUCGUUAU
[1,6,18]Ppr-1-6 ACUUCCAGGUGAUUGAUCGGCAGCCUGUUUCGUGUCCACCUAUCAGGCGACAGCAACGAUGCCGGCAUGCUGUCAUCUUCCCCAUGAUCACCGGUCAGGCGAGGUUACGGCCAUGAGUGACUCCCACGAACAAAAAGUCGAUAUGAG
[1]Pca-1-1      AUAAUUCAUUUCGGUAUGCUCGGUCUUUUUGUUUGCUUUAUAAAACCAGGUCAUUUGUGGAUGUGGGAUUUCUCCCCUCAAUGAGUCUGAGACAAGGAUAAUAGGGUAGAAAUUGUGUUUAAGGCGUCAUGAAGAAUGCGGGGUAU
[2]Gur-1-22     UUGUUCCUUUGGUUUGCUCGGCCUUUGUUUUUUUAAACGAGAAUACGGCAAGCCAAGACUCUUGAUACACAGCGCGAGUGCUGAAAGAGAAAACGAGUCGGGUUUACAAUCGGGAUGUGUUUAGAUUUAAAAACACAAAGGAGGAC
Gur-1-23        ACGGGUCCUUCAGGUGGGCUCGGCCUUUUUAUUUGAUGCACAAACUUGGAAUGGUUCAAGGUGCAUAAAAAGUUUGAGCUGAAGAAAAGAUUACCUUCAAAGCGCAGCCUCAAAGUUCAAAUGGGCAAGGCCGCUGAACGCAUACGC
[3]Gur-1-24     AGAACCUUCGGUAUGUGCGGCCGUUUGCGUUUCCGGGGUGCGUGGAUCGUCCCGAUGCCUCCGAACAGAGAAGACGGCCGUGGAGAAAAUAACAUGGAACCAAAGAACUCCAUUGCAGAGGUCCGUGGGAUCAUGCUGGACUAUGCG
[1]Ppr-1-7      GUUAUUUCCUUUUGGCGUGCUCGGUUUUUUUAUGGGUUGGCGCCUACUGGAACCACGGCCGUACCGUGCGGUCACGAAAGGAGGUGACGGAACAGCGCAACAUCCAUGACAGUGGAGUGCAUUUGAUCAACAGCUCCCCUGGGGUGA
[2]Gur-1-25     GAUGAGUUUGCCUCUUGGUACCCGGUUUUUACCGACAACAAGGCGCUCUCAUCUUCGGAAGGUGAGAGCGCUUAUUUAUUUUUGGAUGUCUGUGCAAUUCUUGUCGAGGGCUGAGUGCGAAUCGUUAUGAAAAUUUACUCACUUCGC
[4,17]Gur-1-26  GAUGAGAUAUGCAUCCUGGCACCCGGUAUGUUACCGACACAAAGGAGCACUCUUAUCUUCGGAAGAUGAGAGUGCUCCUUUUUUUAUCGGUCUACGAGGUAUAGCCAUGGAGAAGAUCAUACGAAAACCGGGCCUUGACCCGGAGUG
Gur-1-27        GAUGAGAUAUGCAUCUUGGUACCCGGUAUAUUACCGACACAAGGAGCGCUCUUAUCUUCGGAAGAUAAGAGUGCUCCUUUUUUUAUUGAAUGUGCAGGUAAUUUUUUUGUUCAAAAGCGCAGGAGCGAGGUGUUCAUCGAUUUGAUC
Pca-1-2         UCCUUAAUUCUUCAAAACACUUCAAGGUCAGUAGGUACAUCUGCGGCGCUUUGCUCUGGAAUUAUAACGGCAGCGCCAUUUUUUUGUGCCCGCACAGGUACAAAAAAUUGAGAGACCCGGUUGAUUCGGUGUCUAAAAAAAAACAUU
Pca-1-3         CGCGAUUGUUACUCAAUUUACGGCCGUAUUUUUAGACCUGUAAAGCAGUCCACCUCUUGCCCGGAGGUGGAUUUUUUUAUUUGCUGACUGCCCGGUUCAAUCGAAACAAGGAUGGGAGAGCCAACGGACCUUUUUUAUUAUAAGAAG
Pca-1-4         CGCGAUGGUUCGUCAUUGAUGGCCUUAUUUAAGGCCUGUAAAGCAGUCCACCUUUGCCCGGGGUGGAUUUUUUUUAUUUCAUAGCCCGGAUAAGAAAAGAAAAAGAGGAGGUUCUUAUUUGUCCGAUUUCAUUCAAUUGGUAAAAAG
Dac-1-1         GGCUGGUUCCCGUCCUUUGUCAAUGAAGGACUACAGAGGUGCUCUUCUUCGCGUAAAGGAGGGCACUUUUUGUUUUUGGGCGGUGAACUGCUUUUUCUGAGUGGGGGGUGUUUUCUCUGAGAAGGAAACCUGAUGCGGCUGAAAUUG
Dac-1-2         GGUGGAGAUCUCUCAUCACCGUGAUAAACAGACACUGAUUUCAACGGCCGGCGUCCUGUUUCUUUCACCCGCCCUUUGCGACUUGCCCUAUCAACGAUACACACGUGUUACCUUUGACAGGAGCAGUGACCAUGCCUCUUCAUCCAU
Ppr-1-8         AGGGCAAACAUGGUUUUCUUAGCCCUUCUUCGCAUCAUCCAAAGAAGGGCUUUUUUGUGGCAUUCACAAUUCGCGACCAUUACUGUCAACCCUGAACAAAAUACUGCAUGCCCUGACCGCAUCGGUGACGUACUGAUGCCGCUGCCA
Pca-1-5         AAUGACGGCAGCAACGAUUUUACUUCUGCACGUGAAGUCAUGGCACCCGGUUUUUCCGCACCACAGGAAUGAUUGGGUGCUUUUUUAUUGAGGAUAAGCAAAAACACAACGCUUCAUAUUGGAUAAAACCAUCAAUCGGCACUAAAC
[22]Ppr-1-9     UAAGCAGUACCUCCUUGUGAGGUGCGGUUUGCUCAGGCACAGACAACAAGGAGAAAUCUGCAUGAAAUCAAUUGGUUCAGUAGUUGCGGUGCGAAGGGCAGUAGGCACAUCGAAAUCAGGCUUCUUGCGGGCCACAGGACUAUUGCG

                ......................................................................................................................................................
                ......................................................................................................................................................
                ●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●
```

```
Pca-1-6        GCUGAUGGAGGAGGGGCAACUCGAAUAUUUUCGGAUAUUUGAUUUAUAAGCGUUCAUCUCCAUUGUCGGGAUGGACGUUUUUUUUUACGCAAUGGACAUAGACAGCGGUGGCGUAGCGGUUUGCCGAUGGGGCAUGGAGUAAAACUG
Gur-1-28       GGAUCGGCAGCACACGUUAAUUCAAAGCAAGAAAUUAACAAUCACCCAAGGAAGCAUUAUUAAAAAGGAGAUGAAAAAAUGAAGAAAAAAAUGAUUACCCUUCUCGCAGCAGCCAUGCUCACCAUGAGCGCCGCAAGCAGCGCCAUG
[2]Gsu-1-14    AACAGUACCGCGUCCCACAGGGCAGGCUGUUUCUUUUUUUUGCCCGCACCCACCACACACACGAAGGAGGAACCAUGAAAGCGGCAGUACUGAGUCUCGUUAUCCUUUUCGCGGCAUCGGCAUCGGCCCUUGCCCAACCGUUGCA
[2]Ppr-1-10    ACAGCUCCGCGAGCAGCAUGCUCGACGGACUGUUUCUUUUUUUGCCGCCCCCGGCUGGUGAACAGGGACCAGAUGACAGGAAAGGGACAGGCUAUGAAAACCAAUCUGCUCUCGUUGAUUACCGCACUGGCGCUCUCCCUGGCGGC
Gme-1-17       GCAACGCCCGCCUUCCAUAGGUACGGGGGUCUGUUUCUGAGGUAGGAAGCAAACUCAACCCAUGGAGGUGGCGCAUGCGCUGCACGAAACGAGUUUGUGGAAAGAUCACCCUUCACCUUGGCACCCUCGGCAUCCUCUGCCUGGCAG
Gme-1-18       GCAACGCCCGCCUUCCAUAGGUACGGGGGUCUGUUUCUGAGGUAGGAAGCAAACUCAACCCAUGGAGGUGGCGCAUGCGCUGCACGAAACGAGUUUGUGGGAAGAUCACCCUUCACCUUGGCACCCUCGGUAUCCUCUGCCUGGCAG
[2]Gsu-1-15    ACAGGUUCCUCAUUAGUCCAUCAAAAGACUAAGGCGGGCUGUUUCUCUUUUUUGUGCUCGCGACUCGAAAGGCGCGGGCAGGAAAAGGAGAAGCGGACGCCGAAAGAAGUACAAAGGAGCCCGUUUAUGAAAACCCGCACGCUAACA
Gsu-1-16       AACAGCCCAUCUGUCGACAAGUCGUACCAUUGCCGCGGGGGGCUGUUUUUUGUUGGGAAAAGUCUCGGUGUACGCCGGGAAAACGAGUAAACUAUUGAGUAGCAGGGAGGUAGUUAAACAUGAAAACCAAGAGGACUCUUCAGGCGU
[15,16]Gsu-1-17 AACCAACCAAAAUGGAGGAAAUGAUGAAAAAGGGGAUGAAAGUAAGUCUUUCCGUGGCAGCAGCAGCACUUCUCAUGAGCGCCCCCGCGGCGUUCGCUUUCCACUCCGGCGGCGUUGCCGAGUGCGAAGGGUGCCACACGAUGCACA
Ppr-1-11       GAGUGGAUAUGGUUUUUACGAGAACAGGCCACGUUUCCUUUGAUGCACGGGAGGCGUGGCUCUUUUAUGCAUAGGUGGGAGGGUUAAUAGGGGAUAGUUCCCUGUUUAGUGAACACAAAAAAUCUGGGAAGUGUCCCACGUUUUUCC
[13]Ppr-1-12   GAGUGUAGAAAAUAUUUUGUGAUAAGAGCCACGCUUCCGUUUGUUGAAUGGGGGGCGUGGUUUUUUUGUGCUUGAAAUUGUUGAGGCGGGAUCACAGGGAGUUGGUCUUUCCGGUGGGCAGAAGAAGCGGAACAACGUGACGGUGCC
Ppr-1-13       GAGGGAAGAAUAUUUUGCAUAAGAGCCACGCUUUCGGUUUAUACACGGGAAGUGUGGUUUUUUUUGCCUGUUUUGGAAGCUAUGCGGAUCAUGUCGAUGAGUCUGUUUUCGCUAUAUGUGCGGGUCGUGGGUGAAGCUCUGGAGGGA
[13]Ppr-1-14   GACGGAUGAUGACCUGUCCGAUAUGCCAAAACCCACCGACUCCUGCUGGUGGGUUUUGUGUUUGGGUGGUUUGGGCGGUUGAGAUUUGAUGCGUUUGUACGCGAAGGAGAUGACAUGAACGGUGGGAGCGAGAAAAAGUCGUUGUUG
Ppr-1-15       GACGGAUGAUGGUCCAUCUGAUCAUGGGAGCCUGCCUUUGUGAGACAGGCUCUUUUCAUUGCGGCUCAUUUCAGCAAGGAGACACCAUGCCAGAACAGAUAACCAUCAUCUGCCCCUGCUGCGCAUAUUCCAGGAUUGUUCCGCGAC
Pca-1-7        GGCUUGAGAGGCAACAUCUCUCAGGAUGAUUCGAUCAUUUGGGCCCGCCUUUUUUGGCGGGUUUUCAGUUUUUCAGGAUAGAUAAUAAGGAAUCGGAAAAGGAGACAACUGUGAGUUUGGGUCAACGGGUUCUAUUACUAUGCAUAG
Gur-1-29       GGUGAUGGAGACAGGCCUACCCUUUCGGGGAUAGGCUUUUUUUUAUCCUUCAGUGGCUGCGUCGUCGCCGAAACCGGCUCCGGCCAGGCGGGGUAAGGAAAAGGUGAAAGAUGAUAUGGUAGGGACCUGAUAUAAAGAUAAGUUAAG
Ppr-1-16       AUCAGCGCUUGUCACCACAGGGUGACGGUCGCCGCGCGUCCACUUUUUCGAAAGAAGAAGUGGAAUUUUUCGUUUCAGCUUGGAAACGGCUUUUAGUGCAGAAAAGGAGGACACACUAUGUCCGAAACCCAUAAACUCCCCGAGGAU
[13]Ppr-1-17   GGUACAGGGGGUGGUGCAUCCGUGCAUCCCCCAUUCAUGCCUGCCUUUUCAAAAGAAAAUGCGGGCUUUUUGUUUUUUGGGGCUUCCUUUGUCACGCAUCCGGCUGUUCGUGCGCCAUGGGGGCUCAGGUCAGCAGCCAUCAUAGAC
Pca-1-8        AAUCGAUGCCUUUCAGACACGACACCGCCUUGCGGCCGGUUUCAUGUCUCCAUCCGCCUUCGGUUUGCAAGGUGGAUUUUUUUAUUUCAAUCGACAUAUGCAUUCUGAGGAUUCGAGCCCCAAAGAUAUUGCUACACCAUCGCAUUU
Pca-1-9        GAGUGUGAUUGUCAGCUGUUUCUGUAUUUGUCUGACGAUUAACGCCCGCUUUCGGUUGAAGGCGGGCGUUUCUAUAUUGUGAAACGGAUCAUUUUUUUCGGUUUUUUUUAAAGCCAGGGAGGGCAAUGCGAACCAUCCAUAGUGUCA
Gme-1-19       GAGGCGACGGAGGCGUACCUGAAGGUACGUUGAGGAGCCCCGCACGAGCCAACGCCGGUAUACGUCAGAAGGCGAAUCCGGCCGAUGGAUGGCCGGGUUGCCUGGGGCGGUAAUGGAACUGACAGGACGUCGAAGCAGCACCAACA.
Gsp-1-7        AACAGGUGAAAAUCGCUUUUUUAAGCCGCUUACCGGGAAGCCGGUAAGCGGCUUUCUGUUAAUAUGCCGCGGGUGUAAUCCAUAUGCUGACAGUUUGACAGAAGGAGCAUCAAUGGCUGGUAGAGGCAGUGGAUUUGUUGAUCAGAC
[2,11]Gur-1-30 UGACCAUUUGUAUGGUUGCAUGCGUAUAACCCGGCCUUUUUAUUUGCCGGGUGAAGAAAGGAGCUGUUAACCGAUUAUCGACAUUUAACCCAACAAAAGGAGAAUGGCAUGAAAUCAAACAAAAUCUCACUAACAGUCUAUCUGAUC
Osp-1-1        AGAGAUUUUCUAUUUACCUUUCAAAGGGUGUCCAUCCGGAAGUUGUCUUUCUAUCCAUAGCGCUAAUUUAUGCCGCAUGUUGAGUUCGUUUUCCUGGUCCUUGGCUAACGCUUGAAUUAAUCCAUCUUUGACCAAUAAACGGUAUAC
Osp-1-2        AGAGUUACCCCAAGGAAUUUCGUUUUUGAGGGUCUCGCACGUUUAGAGUGAGGAGGUCUUUAUGAAAACGAUUAACGUCCGUCUUAACGCUAGGCACAUCCCUAUAGUGCUGAUGUUCUUUCUUGUGUCACCUGUAACAACGGCAAG
Tcr-1-1        AACGACACAUUUAUUUAAGUGGCAACACUUAACAACCUAUCGUUAUUAUGUUCUGUUUAGAUCCAUCCUGUGUGAUUGGAUUUCAUUUAGCCUUUUAAAAAAAUUAAUGACAUGGCGGUUAUGUAUAAAGCGGUUUAUCGUUUUUAU
Tcr-1-2        GAUUUAGAAAUGUGUUUCUUGAUAUUUCUUAUGGUUGAUUUGUUCCUAUCUUUCCGACUUGCCUCACAACAGACAUUCCUUUUCUUUCUAAUGAAAGUUUUUAAUUGCCAAACUUGUUGUGAAGCAGGACGGAAAGCCAUGGGUCU
[21]Tcr-1-3    GUACUUUCUACCACUCUUCAACACACGUUUCGCUUGAGUUCCCUCAAGGAGAUGAAGAUGUUGUUUGUAAAAAAAUAAAUCGUCAGGCUCGACUGACGAUCGCUUGUAUUAGCUUGAUGACAAUGGGCGCGCCCGAUGCACUGGCCG
Tcr-1-4        GAAGUGACAGAAUCAAAACCAUACAUUUUCUGCCCCGUUUAAGGACAAAUGGAGUUAAUAGUUAGGUCAAAAGAAAGGUAAUAUAGAUUGUUAUGAAGCUUAAAAGAAGAGAAAAUAUUUUUCGAACGCAAAAGAAAUCUCACUUUA
Osp-1-3        AAUAAAAGCAGUUGAAAAAGGAUGUAGACAAUGAAUAAAAAUUAUAACUUAAAGAUGGUUAUUUCUGGACUAGCGCUAAGCCUUCCCAUGACAACUCAAGCUGUUGUUUUAGACUAUGAAGCUGUACCAAAGGGCUACGUGUACGGA
Ptu-1-1        UUAUUGAGUUGUGCCUGAGUGCCUCGUUUUAUUACAAACGCGCCUUUAGCUUUUCUGCCAAAUUCUUAUUUUUAUACAUAUUAAAUGGUGCAUGGCAUGGUGUUUUCACCUUAUUAGCCAAAGGGGGCUGCUUUCAUGGUUAAGGUU
Pin-1-1        UGAAUAGACACUUUUGUUUUUUGAGCUGAAUUUUAUGUAUGGUCUUCUUUGGAAGGCCAUUUUUUUUGCUUUUUUUGACAAAAAACAUAAAAACUAAUAAAAGGAAGUAACACCGUUAAUAUGAAAAAAUACGCUAAAGCAUUUAUC
[23]Pin-1-2    CGUGUAGACGCUUUUGCCUUUUGAGCUGAUUUUUAUGUAUGGUCUUCCCUGGAAGGCCAUUUUUUUUGCUUUUUUUUGAUAAAAAAGAAAGCAAAACAGAAAAACUAAUAAUAGGAAGUAACGCGGUUAAUAUGAAAAAGUACGUUG
Pin-1-3        UGUAUAGACACUUUUGCUAUUGAGCUGAUUUUUAUGAAUGGCCUUCCUUGGAAGGCCAUUUUUUACCUUUUUUUGUGUACUGCUGUCUAAAUGCGCAUACAAGUACUUAAUUUAAAAAUGAAAUGUUUUUGAUCCAAAACAAGG
Ptu-1-2        AUACGUAUUUGAAAUUUCUUUGUAUUUGCCAAUUCCCAGAUUUUGUAGUUAAAAAGGUUGAGCAAAGAUGUCCGAGAUUAUUACCAGCAUUAGUUACGACUCAUCACUUGAUUUAGUGUUAGUAAAACUAACUGGCAGCGCCACUUC
Cps-1-1        UAGAUAUAGUGGGGGCUUAUCUUAAUUCGAUAGUUCACUAUCUGUUUCUUCCGGUUUCUUUGCGUGGCCUCUUUCCCUGUAUUUCUCAACCCAUCACUUGCUGCUCUUUACCUAUAAUUUAUUUAGCACUAUUUAAAAAUUGCUAAU
Ptu-1-3        UAACAGUAUCGGGGGCCCUAGGCUGCUUCGUAUGUACCUCCGAGUUCAAACAAUUUACUUUUCUAAAUUUCUGAACAAGGCGAAGCAGCCUGAUUUCCUUUUUAACCUUUUAGUUUUGUUUUUUCCUCUUUGAUGUAUUGCUCAAUU
Iba-1-1        AUAAUUAAUUUUGUUUAACUGCCAAACUCAUCGCGAGAUGAGGACGGAAAGCCACGGGUCUCAGAGUGUGUUGAGAUGUAUCUCAGAUAGCCGGGUUGCGGAACUUAGGAAUUUCGCAAAAUGAAAACUUCAGUAAAAACCCUCUUA
Cps-1-2        AACAGGUACAUCCCGUGGUUUUUUCUUUUAUUAACUCCAUUAUUAAGGUUUAUAAAAUGAAAAAAUUAUUAAUAGCUUCCACCAUCACAGUUUUAACAUUAGCGUCUGCUGUAACAACACCAAAGGCACAAGCUGCUGACAUUGCGA
Cps-1-3        CCAGCAUUAUUGUGCCAAUAAUAAUGUGCAAUGACAGGCACUAAGAAGUAAUAGGUUUUCUGUAGACUUCACACAUUGUGGAGUCACAAAAUGAAGUACGUUAGUAACAAUCAAAUCGAAAUCAGCUCACUAAUAAAACGCCAAAGC
Pin-1-4        GAUAGCACUUCUUAUCAUCCCUUCUAAUUUGCAAAUUAUAAAUAACUGCUUGAUCACUUUUGUGGAGACAAAAUGAAUCAACGAGUUAAUUUAAUCUAUAACCGCCUAAUUUUAAUUUUUUUAUUAGUUAUUAGUCCUGCUG.....
Pin-1-5        GAUAGCACUUCUUAUCAUCCCUUCUAAUUUGCAAAUUAUAAAUAACUGCUUGAUCACUUUUGUGGAGACAAAAUGAAUCAACGAGUUAAUUUAAUCUAUAACCGCCUAAUUUUAAUUUUUUUAUUAGUUAUUAGUCCUGCUGUACUA
Sde-1-1        GCGAGACACCACUCUUUCGCGACAGUCAUUUACCGGUGAUACCACCACUACUGGCGCGGUCGAUGGAAUUAUUCGAACAAAAAAUUGUGGUAAAUCAGCUUCAAAUAGGCAUGUACGUUUGCCGACUAGAUAGGUCGUGGCACGAAA
[3]Iba-1-2     UAGGAAUUUCGCAAAAUGAAAACUUCAGUAAAAACCCUCUUAGUUGCUAGUCUAUUCACCACCACCGCAGCAGUCGCUGAACAAAUGGAUUUAAGCAACACCGAUACCACGUCAUUUACUGUCAACGGUACCGUUGAGAAAGUCUGU
Sde-1-2        GGAAAAGCGACGCUGCUCCGUCUACUCUUCCCCACCAUUAACAUCAAUCACCUUUGGUGAUUAGCUAAUACAUGCUCGCAGGCCUAACAAAUUGACGACGAUUAAAAUUAAUGUGGCAGCAUUAAAAGAAGACCUGGCUUGGCUGUG
Ssp-2-1        AUUGAGCUGAUGCUUUCAAUUCUUUGAAAAAUAAACGGCUUUGCUUAAUGCUGUUUUGCCUAAAUACCACUUUUAUACCUAUGCCAAUCUCGUCGAGAAAGGCGCCCUACGGAGUACAAUAAAGUGGCACAUCUAACUGCUUUUACA
Ssp-3-1        AUUGAGCUGAUGCUUUCAAUUCUUUGAAAAAUAAACGGCUUUGCUUAAUGCUGUUUUGCCUAAAUACCACUUUUAUACCUAUGCCAAUCUCGUCGAGAAAGGCACCCUACGGAGUACAAUAAAGUGGCACAUCUAACUGCUUUUACA
Ssp-1-1        AUUGAUUUAAUGCUUUCAAUUCUUUGAAAAAUAAACGGCUUUGCUUAAUGCUGUUUUGCCUAAAUACCACUUUUAUACCUAUGCCAAUCUCGUCGAGAAAGGCACCCUACGGAGUACAUGAAAGUGGAACAUCAAACUGCUUUUUCA
Son-1-1        AUUCAGUGAAUGUAUUGAAUUCUUAUAAAAAUAAAUAGCAUUUGCUAAAUGCUGUUUUGCCUAAAUACCACUUUUAUACCAAUGCCAAUCUUGUAUCUAAAUAGGCACCCUACGGAGUACAUUAAAGUGGAGCAUUUAACUGCUUUU
Ssp-1-2        UUGUUUCUUCAUGCUGCCGUUGUACGCACAUGGUUUUAAGAGUGUUGUUUCCUGUCUACGCUUUGUUUGCCUUAUUGAUGCAUAACUUACUCAAUAGGGAAAUCUCAUGUUUAAGACUCAAAUCUCCACCUCGGCGAUGCUGGUGGC

               ....................................................................................................................................................................
               ....................................................................................................................................................................
               ●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●
```

	Sde-2-1	UUGUUUCUUCAUGCACUAACGUGAGUGCACGGUAUUAAGAGCGUUUUCGCCUGUUUAUGCUUUGUUUGCCUCAUUGAUGCAUCUAACUCAAUAGGGAUUUCAUAUGUUCAAGACUCAAAUAUCAUCAUCAGCUUGGCUGGUGGCAUC
	Ssp-2-3	UUGUUCUUCAUGUUGCCGUUGUACGCACAUGGUUUUAAGAGUGUUGUAUCCUGUCUAUGCUUUGUUUGCCUUAUUGAUGCAUAAUCUACUCAAUAGGGAAUUUUCAUGUUUAAGACUCAAAUCUCCACCACUGCGAUGCUGGUGGCU
	Ssp-3-2	UUGUUCUUCAUGCUGCCGUUGUACGCACAUGGUUUUAAGAGUGUUGUAUCCUGUCUAUGCUUUGUUUGCCUUAUUGAUGCAUAACUUACUCAAUAGGGAAUUUUCAUGUUUAAGACUCAAAUCUCCACCACUGCGAUGCUGGUGGCU
	Sba-1-1	UUUUUCUUCGUGCAAUUGUCAGUCGUGUUGCAUGGUCUUAAGAGAGUGUUGUUUCCUGUCUACGCUUUGUUUGCCUUAUUGAUUCAUAUAAAAAAAUUCAAUAGGGAUAUCUCAUGUUCAAGACUCAAAUUUCAACAUCAGCCUUGCU
	Sba-2-1	UUUUUCUUCGUGCAAUUGUCUGUCGUGUUGCAUGGUCUUAAGAGAGUGUUGUUUCCUGUCUACGCUUUGUUUGCCUUAUUGAUUCAUAUAAAAAAAUUCAAUAGGGAUAUCUCAUGUUCAAGACUCAAAUUUCAACAUCAGCCUUGCU
	Sba-1-2	AUUCGAUUUCUGUGAAAGAUAAACGACUGCUGUUUAAUGCUGUUGUUUGCCCAAAUACCACUUUUAUACCUAAGCCCAAGCCAAUACUGGCUAAGAUGGGUGCAAAUGGAGUGCAUUAAAGUGGCAAAUCCUACUGUUUUUUCCCCU
	Sba-2-2	AUUCGAUUUCUGUGAAAGAUAAACGACUGCUGUUUAAUGCUGUUGUUUGCCCAAAUACCACUUUUAUACCUAAGCCCAAGCCAAUACUGGCGAAGAUGGGUGCAAAUGGAGUGCAUUAAAGUGGCAAAUCCUACUGUUUUUUCCCCU
	Ssp-5-1	AGGUGAAAAAUCGUUCGUAUACGCUCAAAUCAUUAGCGUUUACAGAAUGUUUAACAAUAUUUUUUACCUGUCGGAGAAUAUAUGAAAUAUAAAGUACUCGCUUUAGUCUCUUUGUUUGCAGCUAUGUCGGCAAAUGCUUCCGUAAUC
	Ssp-2-5	AGGUGAAAAACCUUGGCGAUAUGCAAACCCAAUGCAUGCGCUAACGGGAUAUGGAUAUUUGUUUUUUACCUUUCGGAGAAUAUAUGAAAUUUAAAUCAACACUCGUAGCGUUAGCUACUUUAGUAUUUGCUGGCUCAGUUAAUGCCU
	Ssp-3-3	AGGUGAAAAACCUUGGCGAUAUGCAAACCCAAUGCAUGCGCUAACGGGAUAUGGAUAUUUGUUUUUUACCUUUCGGAGAAUAUAUGAAGAUAAAGCAUACCAUCGUGGCUUUAGCGAGCUUCGCAUUUAUGGGGUCAGUCAACGCUU
	Ssp-4-1	AGCGUUGAUAGAGCCUAGUACUUGUACUUGGAGCUUUCGAAGUGAUGUUGACUUUGAGAGGUCCAAGCUAUGAGUUCAGUUCAUUGCGCAGCUGCGCGUACCGAUUCUACUUUCCUGAAAACCUUUGGUUUGGGCCUAUUGGGCGCA
	Ppr-3-1	UUUAUUAACAGUUUUGAAUAAACCCAUAACUGUUAUAGUUUUCCUGCUUGGUUGCAGCCCUUUUCGUUAAUAACUCCUAACGAAUAAGGAUAUACAACAUGUUACAUCAAACACCUUCCCUGAUUUCCUCGGAACAACGUAAUUUCU
	Ppr-2-1	UUUAUCAACAGUUUUGAAUGGAUCAUAACUGUUUAUAGUUUUCCUGCUUGGUUGCAGCCCUUUUCGUUAAUAACUCCUAACGAAUAAGGAUAUACAACAUGUUACAUCAAACACCUUCACUGGUUUCCUCGGAACGACGCAACUUUU
	Vsp-3-1	AUAGGAUUUAAGUGUCAAUGCAUGCUGGAGUGUUCUCUUUUUUGUCGGCCAAAUACAGUAAGGACAAAUAAUGAAACAACUCCCAAAAAAAUCGCUCAUCGCACUAUCGCUACUUAGCCUAAGUGGAGCAAGCUUCGGUCACGGUUA
	Ppr-3-2	UUAAAGACGUCAAAUUGCAUCAAACAUGCCGAUUCUAUGGUUGUUUUUUCUAUAAAACAACGGAAUCAGUGAUAAUGAUUAUUACAAAAAAAACACUUCUGCCUGUUACCUUGGCGCUGUUUAGCUCAGGGGUAAUGGCACACGGU
	Ppr-2-2	UUCAAGACGUAAAAUUGCAUCAAGCAUGCCGAUUCUAUGGUUGUUUUUUCUAUAAAACAACGGGAUCAGUGAUAAUGAUUAUUACAAAAAACACACUUCUACCAGUUACCUUGGCGCUGUUUAGCUCAGGAUAUUGUAUUCAAUUG
24	Vch-2-1	CAAAAUGCAUACACUUUGUUGACUCAUCAUUGACACUAUGAAUGCAUGCUUUUGCUAAUUUUCUCGGACCUGACUUGGUGGCGUAUGUAACAAUACUCCGAGACUAACAGCAUGGAUAAACCAGUAUUAAAAGAUUCAAUGCGGCUC
	Vch-3-1	CAAAAUGCAUACACUUUGUUGACUCAUCAUUGACACUAUGAAUGCAUGCUUUUGCUAAUUUUCUCGGACCUGACUUGGUGGCGUAUGUAACAAUACUCCGAGACUAACAGCAUGGAUAAACCAGUAUUAAAAGAUUCAAUGCGGCUC
13	Vpa-1-1	CAAGUGUGCAACCAAUGUCUUUACCCUUUAAUGACAUUGCCUUUGCGCACGUUUGCUACUCUCUCGGACCUGACUUUUGGUGGCGUAAUAUUAAUUACACCGAGAUAACAAAACAUGGAUAAACCGAUACUCAAAGACUCUAUGAAG
	Vsp-1-1	CAAGUGUGCAAGCGAUGACUUUACCCUGUAAUGACAUUGCUUUUGCGCAUGUUUGCUACUCUCUCGGACCUGACUUUUGGUGGCGUAAUAUUAAUUACACCGAGAUAACAAAACAUGGAUAAACCGAUACUCAAAGACUCUAUGAAG
	Val-1-1	CAAGUGUGUAAGCAAUGACUUUACUCUGUAAUGACAUUGUUUUUACGCAUGUUUGCUACUCUCUCGGACCUGACUUUUGGUGGCGUAAUAUUAAUUACACCGAGAUAACAAAACAUGGAUAAACCGAUACUCAAAGACUCUAUGAAG
	Vvu-2-1	CAAAAAUGCAUAACAACAAGUUUACCCAUUAAUGACAUUGUAUUUGCAUUGGUUUGCUAAUCUCUCGGACCUGAUUUGGUGGCGUAAUAGUAAUUACACCGAGAUAACACAUCAUGGAUAAACCGAUACUCAAAGAUUCAAUGAAGC
	Vsp-3-2	CAAAAUGCAGUAACUACUAAUAAUUUAGUAAUUUACGGCUAAUUGACAAUUUAAUUGCAACAUUUUGCUAAUCUCUCGGACCUGCUUAUUUGGUGGCGUAGUUCAAAUACACCGAGAUAAAUAACAUGGAUAAACCGAUACUAAAGG
	Vsp-2-1	CAAAAUGCAGUAACUGCUAAUACUUUAGUAAUUUACAGCCUAUUGACAAUUUAAUUGCAACAUUUUGCUAAUCUCUCGGACCUGCUUAUUUGGUGGCGUAGUUCAAAUACACCGAGAUAAAUAACAUGGAUAAACCGAUACUAAAGG
	Vfi-1-1	CAAGUGGUCGAUUCUAUCCCUUGUUAAUUUUUACGUUAAAAUAACAAUACUUAGUUUUCAACUUGCUACUAUUUCCUCGGGCUCAGUUGUUUAGUGACAUAGACUUACUAACCGAGAAUUAUUGCAAUGGAUAAACCAGUAUUAAAA
	Van-1-1	AUUAUUAUUGAGCCGACAAUAUUAUGACACAAACUUAACAAUAUAUUGACAGUAUCAAUGUGUAGAUUAAUUAAUCUAUUUCCUCGCUCUUAAUGUUUAGGUGACACUCAUAUUUUUUCACCGAGUAGAAGCAUGGAUAAACCAAUA
	Psp-1-1	AUUAUUAUUGAACCGACAAUAUGAUGACGUAAACUUAACAAUAUAUUGACAGUAUCAAUGUAUAGAUUAAUUAAUCUAUUUCCUCGCUCUUAAUGUUUAGGUGACACUCAUAUUUUUUCACCGAGUAGAAGCAUGGAUAAACCAAUA
	Vvu-2-2	CACUAGAGAGUGUUCUCACCAAGCACGUCAGUGCGCUCUUUACGAGCCAUUCAUCGCCUUGGGAAUAACAUGCGUAUGCUGUGGUUGAACAUUGGCAAUAUAGCCAAAUAAAGUCAGCAAUCAAAAAGAGUAUGUUAUGAAAAAACA
	Vvu-1-2	CACUAGAGAGUGUUCUCACCAAGCACGUCAGUGCGCUCUUUACGAGCCAUUCAUCGCCUUGGGAAUAACAUGCGUAUGCUGUGGUUGAAGAUUGGCAAUAUAGCCAAAUAAAGUCAGCAAUCAAAAAGAGUAUGUUAUGAAAAAACA
	Val-1-2	GCAGUGUUACAGAAGUCCAAUUUCUGUAGCGUUGCUUUAUGCUCAGCAUGCGCCGGGGUGUGACACUAGUUUGUCAGCCAUAACAGAGGACAUGUGAUGAAAUCAUUUCCAACCAAAUCCGUACUCGCGUUAGCACUUGCAAGCAUU
	Vpa-1-2	CUAUCUGUAUCUAAUGAGUUCCAAUCAGAUGCGGACGAGUGUGCUCUGCAUGCGCCGAGUUGUGGCAUUCAUUUGUCAGCCAAACAAGAGGACGUGUAAUGAAAUCAUUCCCUAAUAAAUCGCUCGUUGCCCUUGCUAUCGCAAGCA
10	Vch-2-2	GGUUUAUACCGUUUUGUCUUAUGCGUUGCUGAAGGUGCUGAAGUUCAGAUUCACUCUGAACUGCGUCUGGCGCUUUGUCAGACAAACUAUGUCAGCUAACAAAGAAGAGUCUGUGAUGAAAAAACAACCUAAAAUGACCGCUAUUGC
	Vch-3-2	GGUUUAUGCCGUUUUGUCUUAUGCGUUGCUGAAGGUGCUGAAGUUCAGAUUCACUCUGAACUGCGUCUGGCGCUUUGCCAGACAAACUAUGUCAGCUAACAAAGAAGAGUCUGUGAUGAAAAAACAACCUAAAAUGACCGCUAUCGC
	Vch-5-2	GGUUUAUGCCGUUUUGUCUUAUGCGUUGCUGAAGGUGCUGAAGUUCAGAUUCACUCUGAACUGCGUCUGGCGCUUUGUCAGACAAACUAUGUCAGCUAACAAAGAAGAGUCUGUGAUGAAAAAACAACCUAAAAUGACCGCUAUCGC
	Vch-4-1	GGUUUAUACCGUUUUGUCUUAUGCGUUGCUGAAGGUGCUGAAGUUCAGAUUCCCUCUGAACUGCGUCUGGCGCUUUGCCAGACAAACUAUGUCAGCUAACAAAGAAGAGUCUAUGAUGAAAAAACAACCUAAAAUGACCGCUAUUGC
	Ppr-3-3	AUUAAAAAAAGUUUUGUCAUAAAUUUUACAUACUUUUGACACAAAUUAUCUGUAGUUUGAUAAUUUGGAUUCCUCGCUCUGAUGUUAGGUGACUCAUGGCAUAUCACCGAGAAGAAGCAUGGAUAAACCAGUACUAAAAGACGCAUU
	Ppr-2-3	AUUAACAAAAAGUUUUGUCAUAAAUUUUACAUAAUUUUGACACAGAUUAUCUGUAGUUUGAUAAUUUGAAUUCCUCGCUCUGAUGUUAGGUGACUCAUGGCAUAUCACCGAGAAGAAGCAUGGAUAAACCAGUACUAAAAGACGCAU
19	env-1	AUUAAGCUGAUGCUUUCAAUUCUUUGAAAAAUAAACGGCUUUGCUUAAUGCUGUUUUGCCUAAAUACCACUUUUAUACCUAUGCCAAUCUCAUCGAGAAAGGCACCCUACGGAGUACAUUAAAGUGGAACAUCAAACUGCUUUUUCA
	env-2	AUUGAUUUAAUGCUUUCAAUUCUUUGAAAAAUAAACGGCUUUGCUUAAUGCUAUUUUGCCUAAAUACCACUUUUAUACCUAUGCCAAUCUCGUCGAGAAAGGCACCCUACGGAGUACAUUAAAGUGGCACACCAAACUGCUUUUUCA
	env-3	AUUCAGCUAAUAUAUUGAAUUCUAUAAAAAAUAAACGGCUUUGCUUAAUGCUGUUUUGCCUAAAUACCACUUUUAUACCAAUGCCAAUCUUUUAUUGAAUGGCCCCUACGGAGUACAUUAAAGUGGAACAUCUGACUGCUUUUUCUC
	env-4	UGAGGGUGACGUUUGUGGCAGGUCCAUUGCCAGUUGUCUGCGCCCAUCACGCAGGUUGCGCUUGUUCUUUGCCUUGGUGACCAAUGCCAAGACAAAACAAGGAAGAUAACAAUAUGCAACAGACCCUAAGGCCCAGCAGGCUUGCCA
	env-5	UUGUUUCGUCCAUGGGCCAAGCCAUGGUGCGAUGUUGUCUGCGUUACCUGUCAGGUUGCGUUGCCUUUUUCGUUAUCUAUUCCAAGAACGAAAUAAGGAUAUUCAACUAUGUUAAGUCCAAAACCCUCCGUACUGGCGUUGUUGGUC
	env-6	UUGUUUCUUCAUGCCGCCGUUGUAGGCACAUGGUUUUAAGAGUGUUGUUUCCUGUCUACGCUUUGUUUGCCUUAUUGAUGCAUAACUUACUCAAUAGGGAAAUCUCAUGUUUAAGACUCAAAUCUCCACUUCGGCGAUGCUAGUGGC
7	env-7	UUGUUUCUUCAUGCUGCCGUUGUGCGCACAUGGUUUUAAGAGUGUUGUUUCCUGUCUACGCUUUGUUUGCCUUAUUGAUGCAUAACUUACUCAAUAGGGAAAUCUCAUGUUUAAGACUCAAAUCUCCACCUCGGCGAUGCUGGUGGC
	env-8	AGAAGAGCGAUGUCAUCAGUUGUAUCUGUGUCAUUCUCUUUGAUUGCGUUGCCCUUUUCGUACAACGACAUGUGACGAAAAGGACAUAAAAAUGAACACUUCCACCGCUUCCGUUCGCAACAGCGCCAUCCGUUUCGCUUCUUCUGC
	env-9	AAGCGGUCACCCAAAAGGUGAACCUGUAGGUUGAUCAAUACUGGUCAAGCCUCGGUUUUGCCUGUUCAGGUGGAUACACCACAGAUCCGUCCAGGAUUUGGCUGUAUGCCUGUCUGCUUUUUUUGCCUUUUAAUACUGUUUGUGGUU
8	env-10	CUGAUGAUUACGUAGGACGCAUCUACGCCAGCUCGCUGGUAUCAAAGCUCCUCCUUUCGUGUUCGCUGAUAAACAACGGCAAACAUACUGUUGUUGAAGUUUCAGCCUGUUUAUACAGACCAAUCAGCCUGAUUUUUAGAUCAGUGU
	env-11	GUGCUUCGGGGAACUCCCCCUGCCAACUUGACUGCCCACAUCUGUCGAACAGAGGAAUGUAUGGCAGCAAAAAUCCAACUCAAUCACGUCGCUGCGACCCUGGCCCUGCUGGUCAGUGGCAGCGUGUUCGCCCACGGCUACAUC...
		....
		....
		●●●

2.2 The SAH motif

Notes:

[1] BLAST indicates that the downstream gene is probably *metH*. Also, the first roughly 30 amino acids (90 nucleotides) are not conserved. Therefore, we have annotated a start codon after the predicted RNA that is highly consistent with other instances.

[2] No downstream is annotated in this shotgun fragment, but BLAST homology supports the annotated start codon, and that the gene is *acyH*.

abbrev. of hits	**taxonomy of species**
Fsp-1-1 to Fsp-1-1	Actinobacteria Actinobacteridae Actinomycetales Frankineae Frankiaceae *Frankia sp.* EAN1pec
Bma-1-1 to Bma-1-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* 10229
Bma-2-1 to Bma-2-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* 10399
Bma-3-1 to Bma-3-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* 2002721280
Bma-4-1 to Bma-4-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* ATCC 23344
Bma-5-1 to Bma-5-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* FMH
Bma-6-1 to Bma-6-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* GB8 horse 4
Bma-7-1 to Bma-7-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* JHU
Bma-8-1 to Bma-8-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* NCTC 10247
Bma-9-1 to Bma-9-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* SAVP1
Bps-1-1 to Bps-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1106a
Bps-2-1 to Bps-2-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1106b
Bps-3-1 to Bps-3-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1655
Bps-4-1 to Bps-4-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1710a
Bps-5-1 to Bps-5-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1710b
Bps-6-1 to Bps-6-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 406e
Bps-7-1 to Bps-7-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 668
Bps-8-1 to Bps-8-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* K96243
Bps-9-1 to Bps-9-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* Pasteur
Bps-10-1 to Bps-10-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* S13
Bth-1-1 to Bth-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia thailandensis* E264
Aeh-1-1 to Aeh-1-1	γ-proteobacteria Chromatiales Ectothiorhodospiraceae *Alkalilimnicola ehrlichei* MLHE-1
Mca-1-1 to Mca-1-1	γ-proteobacteria Methylococcales Methylococcaceae *Methylococcus capsulatus str.* Bath
Avi-1-1 to Avi-1-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Azotobacter vinelandii* AvOP
Pae-1-1 to Pae-1-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* 2192
Pae-2-1 to Pae-2-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PA7
Pae-3-1 to Pae-3-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PACS2
Pae-4-1 to Pae-4-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PAO1
Pae-5-1 to Pae-5-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* UCBPP-PA14
Pen-1-1 to Pen-1-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas entomophila* L48
Pfl-1-1 to Pfl-1-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf-5
Pfl-2-1 to Pfl-2-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* PfO-1
Ppu-1-1 to Ppu-1-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* F1
Ppu-2-1 to Ppu-2-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* KT2440
Psy-1-1 to Psy-1-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. phaseolicola* 1448A
Psy-2-1 to Psy-2-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. syringae* B728a
Psy-3-1 to Psy-3-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. tomato str.* DC3000
Xax-1-1 to Xax-1-1	γ-proteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas axonopodis pv. citri str.* 306

Xca-1-1 to Xca-1-1	γ-proteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas campestris pv. campestris str.* 8004
Xca-2-1 to Xca-2-1	γ-proteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas campestris pv. campestris str.* ATCC 33913
Xca-3-1 to Xca-3-1	γ-proteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas campestris pv. vesicatoria str.* 85-10
env-1	environmental sample

abbrev	**RefSeq accession**		**5′ at**	**3′ at**	**genes**
Pen-1-1	NC_008027.1	+	5348235	5348364	**RNA**→ ahcY (pfam05221)→ metF (cd00537)→
Ppu-2-1	NC_002947.3	+	5667861	5667990	**RNA**→ metF (cd00537)→
Ppu-1-1	NZ_AALM01000081.1	-	12364	12235	**RNA**→ ahcY (pfam05221)→ metF (cd00537)→
Pfl-2-1	NC_007492.1	+	5940635	5940768	**RNA**→ ahcY (pfam05221)→ metF (cd00537)→
Avi-1-1	NZ_AAAU03000001.1	+	1647262	1647384	**RNA**→ ahcY (pfam05221)→
[2]env-1	AACY01215322.1	+	598	726	**RNA**→
Pfl-1-1	NC_004129.6	+	6604857	6604997	**RNA**→ ahcY (pfam05221)→ metF (cd00537)→
Psy-3-1	NC_004578.1	+	5771299	5771444	**RNA**→ ahcY (pfam05221)→ metF (cd00537)→
Psy-1-1	NC_005773.3	-	515378	515232	**RNA**→ ahcY (pfam05221)→ metF (cd00537)→
Psy-2-1	NC_007005.1	-	499730	499584	**RNA**→ ahcY (pfam05221)→ metF (cd00537)→
Pae-4-1	NC_002516.2	-	484250	484110	**RNA**→ ahcY (pfam05221)→ hypo→ metF (cd00537)→
Pae-5-1	NZ_AABQ07000003.1	-	719407	719267	**RNA**→ ahcY (pfam05221)→ COG0317: Guanosine polyphosphate pyrophosphohydrolases/synthetases→ metF (cd00537)→
Pae-1-1	NZ_AAKW01000056.1	-	2066	1926	**RNA**→ ahcY (pfam05221)→ COG4120 (COG4120)→
Pae-2-1	NZ_AAQE01000015.1	+	59340	59480	**RNA**→ ahcY (pfam05221)→ hypo→ metF (cd00537)→
Pae-3-1	NZ_AAQW01000001.1	-	473183	473043	**RNA**→ ahcY (pfam05221)→ hypo→ metF (cd00537)→
[1]Fsp-1-1	NZ_AAII01000057.1	+	13972	14079	**RNA**→ MetH (COG0646)MetH (COG1410)→
Aeh-1-1	NZ_AALK01000011.1	+	76189	76296	**RNA**→ ahcY (pfam05221)→
Mca-1-1	NC_002977.6	-	144286	144123	**RNA**→ ahcY (pfam05221)→ metF (cd00537)→
Xax-1-1	NC_003919.1	-	955589	955419	**RNA**→ ahcY (pfam05221)→
Xca-3-1	NC_007508.1	-	977044	976874	**RNA**→ ahcY (pfam05221)→
Xca-1-1	NC_007086.1	+	4140239	4140408	**RNA**→ ahcY (pfam05221)→
Xca-2-1	NC_003902.1	-	900313	900144	**RNA**→ ahcY (pfam05221)→
Bth-1-1	NC_007651.1	-	3610424	3610238	**RNA**→ ahcY (pfam05221)→ COG1950 (COG1950)→ metF (cd00537)→
Bps-8-1	NC_006350.1	-	3907874	3907688	**RNA**→ ahcY (pfam05221)→ COG1950 (COG1950)→ metF (cd00537)→
Bps-5-1	NC_007434.1	-	62272	62086	**RNA**→ ahcY (pfam05221)→ COG1950 (COG1950)→ metF (cd00537)→
Bps-4-1	NZ_AAHS02000077.1	-	8493	8307	**RNA**→ ahcY (pfam05221)→ COG1950 (COG1950)→ metF (cd00537)→
Bps-7-1	NZ_AAHU01000007.1	-	30106	29920	**RNA**→ ahcY (pfam05221)→ COG1950 (COG1950)→ metF (cd00537)→
Bps-10-1	NZ_AAHW02000049.1	-	30124	29938	**RNA**→ ahcY (pfam05221)→ COG1950 (COG1950)→ metF (cd00537)→
Bma-4-1	NC_006348.1	-	2934899	2934713	**RNA**→ ahcY (pfam05221)→ COG1950 (COG1950)→ metF (cd00537)→
Bma-1-1	NZ_AAHM02000001.1	+	1743003	1743189	**RNA**→ ahcY (pfam05221)→ COG1950 (COG1950)→ metF (cd00537)→
Bma-2-1	NZ_AAHN02000009.1	-	203964	203778	**RNA**→ ahcY (pfam05221)→ COG1950 (COG1950)→ metF (cd00537)→
Bma-6-1	NZ_AAHO01000086.1	-	8024	7838	**RNA**→ ahcY (pfam05221)→ COG1950 (COG1950)→ metF (cd00537)→
Bma-8-1	NZ_AAHP01000058.1	-	14197	14011	**RNA**→ ahcY (pfam05221)→ COG1950 (COG1950)→ metF (cd00537)→
Bma-9-1	NZ_AAHQ02000003.1	-	219660	219474	**RNA**→ ahcY (pfam05221)→ COG1950 (COG1950)→ metF (cd00537)→
Bps-3-1	NZ_AAHR02000057.1	-	29952	29766	**RNA**→ ahcY (pfam05221)→ COG1950 (COG1950)→ metF (cd00537)→

Bps-9-1	NZ_AAHV02000048.1	-	30165	29979	**RNA**→ ahcY (pfam05221)→ COG1950 (COG1950)→ metF (cd00537)→
Bma-5-1	NZ_AAIQ02000088.1	-	8023	7837	**RNA**→ ahcY (pfam05221)→ COG1950 (COG1950)→ metF (cd00537)→
Bma-7-1	NZ_AAIR02000076.1	-	8010	7824	**RNA**→ ahcY (pfam05221)→ COG1950 (COG1950)→ metF (cd00537)→
Bps-1-1	NZ_AAMA01000084.1	-	8669	8483	**RNA**→ ahcY (pfam05221)→ COG1950 (COG1950)→ metF (cd00537)→
Bps-2-1	NZ_AAMB01000063.1	+	13940	14126	**RNA**→ ahcY (pfam05221)→ COG1950 (COG1950)→ metF (cd00537)→
Bps-6-1	NZ_AAMM02000096.1	+	13994	14180	**RNA**→ ahcY (pfam05221)→ COG1950 (COG1950)→ metF (cd00537)→
Bma-3-1	NZ_AANX02000034.1	+	13904	14090	**RNA**→ ahcY (pfam05221)→ COG1950 (COG1950)→ metF (cd00537)→

cd00537 Methylenetetrahydrofolate reductase (MTHFR).
COG0646 Methionine synthase I (cobalamin-dependent), methyltransferase domain [Amino acid transport and metabolism]
COG1410 Methionine synthase I, cobalamin-binding domain [Amino acid transport and metabolism]
COG1950 Predicted membrane protein [Function unknown]
COG4120 ABC-type uncharacterized transport system, permease component [General function prediction only]
pfam05221 S-adenosylhomocysteine hydrolase [Coenzyme metabolism]

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Bma-1-1, Bma-2-1, Bma-3-1, Bma-5-1, Bma-6-1, Bma-7-1, Bma-8-1, Bma-9-1, Bps-1-1, Bps-10-1, Bps-2-1, Bps-3-1, Bps-4-1, Bps-5-1, Bps-6-1, Bps-7-1, Bps-9-1, Pae-1-1, Pae-2-1, Pae-3-1, Pae-5-1, Ppu-1-1, Psy-2-1, Xca-2-1

alignment positions 1 · · · 152

```
 Pen-1-1   CCGCC.............A.CCCGGCCGAGGGGCGCUGC.AGCAGG........................UUCGU..........................CC.UGU.CAGGCUCGGUCGGGGCGUUGUCCGC............GCA....
 Ppu-2-1   CCGCC.............A.CCUGUCCGAGGGGCGCUGC.AGCAGGC........................UAG..........................GCC.UGU.CAGGCUCGGAUGGGGCGUU..GCUCGC.........UCCC....
 Pfl-2-1   CCGCC.............A.CCUGUCCGAGGGGCGCUGC.AGCAGGU........................UCA..........................ACC.UGU.CAGGCUCGGAUGGGGCGUU..GACCGGC........UCAG....
 Avi-1-1   CCGCC.............A.CCCUUCCGAGGGGCGCUGC.AGCAGGCC.....................UGCCAAC.......................GGCC.UGU.CAGGCUCGGAAGGGGCGUU.................UUCCG...
2env-1     CCGCC.............A.CCCGGUCGAGGGGCGCUGC.AGCAAACCCGC...................AUCAC.....................GCGGCUU.UGU.CAGGCUCGCCCGGGGCGUU.................AUCCA...
 Pfl-1-1   CCGCC.............A.CCUGUCCGAGGGGCGCUGC.AGCAGG........................CAUGA..........................CC.UGU.CAGGCUCGGAUGGGGCGUUGU.UG.GUACAGG....GCUC....
 Psy-3-1   CCGCC.............A.CCUGUCCGAGGGGCGCUGC.AGCAGG........................UUUAC..........................CC.UGU.CAGGCUCGGAUGGGGCGUUG.UUUGUCAGGUUG...AUCA....
 Psy-1-1   CCGCC.............A.CCUGUCCGAGGGGCGCUGC.AGCAGG........................UUUUC..........................CC.UGU.CAGGCUCGGAUGGGGCGUUG.UUUGCCAGGACG...AUCCU...
 Pae-4-1   CCGCC.............U.UCCGUUCGAGGGGCGCUGC.AGCAGGC.......................CUAG..........................GCC.UGU.CAGGCUCGAACGGAGCGCGCUCUCAGGCCGGC.....CUC....
1Fsp-1-1   UCGUC............GAUGCCGGUCGAGGGGCGCUGCAACGGAC.........................CGAG..........................GU.CCGCCACGCUCGAUCGG....................AACCUCAUCCG
 Aeh-1-1   GCGCC............G...CC..CCGAGGAGCGCUGCAACAGGGCC.......................GCC.........................GGCC.CUGCCAGGCUCGG..GG.AUGG..................CGCC....
 Mca-1-1   UUGUC................CGCGCCGAGGAGCGCUGCGACGGCCC.......................GUCGA.........................GGG.CCGCCAGGCUCGGCGGG.GAC...................AAUCG...
 Xax-1-1   GUGCCGAUAUAAUCGGGCACUCUCGUCGAGGGGCGCUGCGACCG.GUA...GGGCGGACUGAUCCGCUC...A...CUGCACC.ACA.GGUGCAGG...UACA.CGGCCAGGCUCGACCAG................CAGU...UACC....
 Xca-3-1   CUGCCGAUAUAAUCGGGCCCUCUCGUCGAGGGGCGCUGCGACCG.GUA...GAGCGGACUGAUCCGCUC...A...CUGCACC.ACA.GGUGCAGG...UACA.CGGCCAGGCUCGACCAG................CAGU...UACC....
 Xca-1-1   CUGCCGAUAUAAUC..AGACCCUCGUCGAGGGGCGCUGCGACCG.GUA...CGGCAGAUUCAUCUGCC....A...CCUGCACGAAAAGUGCGGGC...AACAACGGCCAGGCUCGACAGGGAU....................UUUUC...
 Bth-1-1   CCACC.............U.CGUUUCCGAGGAGCGUUGCGACGGGC.........................CCC...........................GC.CCGCCAGGCUCGGAAAUGGU......................CA....
 Bps-8-1   CCACC.............U.CGUUUCCGAGGAGCGUUGCGACGGGC.........................CCC...........................GC.CCGCCAGGCUCGGAAAUGGU......................CA....
 Bma-4-1   CCACC.............U.CGUUUCCGAGGAGCGUUGCGACGGGC.........................CCC...........................GC.CCGCCAGGCUCGGAAAUGGU......................CA....
           .<<<<...............<<<<<<<<<...<<.<<<...<<<<<<<<<<<<<<<<<....>>>>>>>......<<<<<<<.....>>>>>>>.>>>>>>>.>>>.>>>>>>>>>>>>>><<<<<<<<<<<<<<<<<<<...........
           .1210...............?2??2?000...00.?00...222?2??000?200200....002002?......0222220.....0222220.000??2?.222.00?00000?2??2?222????????22?2??22...........
           YYGCC-------------∘-∘CY••YCGAGGRGCGCUGC∘A••R∘•∘------------------------∘•∘--------------------------∘••-Y••∘CAGGCUCGR••GG∘R∘∘-------------------∘∘Y•----
```

```
           alignment positions 153···308
 Pen-1-1   ........GCGGACGCUCAACGCA....CAACGGCGCCC...AUUCGCACACU..ACG.AAU..........................................................................GGAGG.CUCUCA..AUGAGC
 Ppu-2-1   ......GCGAGCGC.UU.AACGCA....CAACGGCGCCC...AUUCGCACAUU..ACG.AAU..........................................................................GGAGG.CUCUC...AUGAGC
 Pfl-2-1   .....GCCGGACAC.U.AAACGCA....CAACGGCGCCC...AUUCGCAUACAUUACG.AAU..........................................................................GGAGG.CUCUC...AUGAGC
 Avi-1-1   ..................AACGCAU...CAACGGCGCCC...AUUCGCACAUU..ACG.AAU..........................................................................GGAGA.CUCUGC..AUGAGC
2env-1     ..................AACGCAU...CAACGGCGCCC...AUUCGCAGACA..ACG.AAU..........................................................................GGAGA.GUUACUGCAUGAGC
 Pfl-1-1   .CCUGUAUGGACAC.U.AAACGCA....CAACGGCGCCC...AUUCGCACACU..ACG.AAU..........................................................................GGAGG.CUCUUG..AUGAGC
 Psy-3-1   CAGCC.GGCAAGCC.UUAAACGCA....CAACGGCGCCC...AUUCGCAUUUUU.AUGGAAU..........................................................................GGAGA.CUCUCA..AUGAGU
 Psy-1-1   CGACC.GGCAAACC.UUAAACGCA....CAACGGCGCCC...AUUCGCAUUUUU.AUGGAAU..........................................................................GGAGA.CUCUCA..AUGAGU
 Pae-4-1   .GCCGGUCUG.....UCAA.CGCAC...CAACGGCGCCC...AUUCGCAGACA..ACG.AAU..........................................................................GGAGAGUUACGC..AUGAGC
1Fsp-1-1   ............................CAAGGGCGCCUCCCACAACCG.......................................................................................GGAGG.UGUGCGCGAUGGCA
 Aeh-1-1   ...................CCAUCGUUCCAACGGCGCUCAGCGAAAUC........................................................................................GGAG..UCUGCACCAUGAAU
 Mca-1-1   ....................GUU..UUCCAACGGCGCUCUGUUUAUCUCCACCGGCCGCGGCUGCCGACGGGUAAUCAGAGUCUCGCCCCGGUGGAGACCCGACAAA.............................GAGG..UAUUC...AUGAAC
 Xax-1-1   GUU.........................CAACGGCGCCCGGCAUU.GCAGACAUUCGUUUGCCUACC.....................................................................GGAG..CCAUUGC.AUGAAC
 Xca-3-1   GUU.........................CAACGGCGCCCGGCAUU.GCAGACAUUCGUUUGCCUACC.....................................................................GGAG..CCAUUGC.AUGAAC
 Xca-1-1   .....................GU.G...AAACGGCGCCCGGAAUU.GCAGACAUUCGCCUGCCUACC.....................................................................GGAG..CAACUG..AUGAAC
 Bth-1-1   .....................ACC....GAGCGGUGUGUCUGCGCAUCAUCGUCGCAUAGCGGCGAUCGUUUCCGCCGAUGCCGGCCGCUUACUUGAACGGCGCUCACGUCACAAUUUUCUAGAACUUUUUUAGAAAGGAGGGCGUG...AUGAAC
 Bps-8-1   .....................ACC....GAGCGGUGUGUCGGCGCAGCAUCGUCGCAUAGCGGCGAUCGUUUCCGCCGAGACCGGCCGCUUACUUGAACGGCGCUCACGUCACAAUUUUCUAGAACUUUUUUAGAAAGGAGGGCGUG...AUGAAC
 Bma-4-1   .....................ACC....GAGCGGUGUGUCGGCGCAGCAUCGUCGCAUAGCGGCGAUCGUUUUCGCCGAGACCGGCCGCUUACUUGAACGGCGCUCACGUCACAAUUUUCUAGAACUUUUUUAGAAAGGAGGGCGUG...AUGAAC
           >>>>>>>>>>>>>>....>>>>>.........>>>>........................................................................................................................
           22??2?22??????....??222.........0121........................................................................................................................
           --------------------○○Y○----CAACGGCGCYC○○○RY●○●CR○○○○--○○○-○○○--------------------------------------------------------------------------GGAG○-Y●●●Y○--AUGARY
                                                                                                                                                   SD++++        AUG
```

2.3 The COG4708 motif

Notes:

[1] The start codon of this putative homologs is annotated roughly 70 nucleotides 5′ to our annotation. We propose that this early start codon is false because there is clear sequence similarity throughout the motif region, so if that region is coding in some sequences it should be coding in all. (Also, the start codon is UUG, so our use of a putative UUG start codon is no worse.)

abbrev. of hits	taxonomy of species
Lla-1-1 to Lla-1-1	Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. lactis* Il1403
Sag-1-1 to Sag-1-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 18RS21
Sag-2-1 to Sag-2-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 2603V/R
Sag-3-1 to Sag-3-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 515
Sag-4-1 to Sag-4-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* A909
Sag-5-1 to Sag-5-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* CJB111
Sag-6-1 to Sag-6-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* COH1
Sag-7-1 to Sag-7-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* H36B
Sag-8-1 to Sag-8-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* NEM316
Smu-1-1 to Smu-1-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus mutans* UA159
Spn-1-1 to Spn-1-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* R6
Spn-2-1 to Spn-2-2	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* TIGR4
Spy-1-1 to Spy-1-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* M1 GAS
Spy-2-1 to Spy-2-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* M49 591
Spy-3-1 to Spy-3-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10270
Spy-4-1 to Spy-4-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10394
Spy-5-1 to Spy-5-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10750
Spy-6-1 to Spy-6-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS2096
Spy-7-1 to Spy-7-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS5005
Spy-8-1 to Spy-8-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS6180
Spy-9-1 to Spy-9-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS8232
Spy-10-1 to Spy-10-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS9429

abbrev	RefSeq accession		5′ at	3′ at	genes
Spy-4-1	NC_006086.1	+	587336	587498	**RNA**→ COG4708 (COG4708)→
Spy-7-1	NC_007297.1	+	571458	571620	**RNA**→ COG4708 (COG4708)→
Spy-8-1	NC_007296.1	+	569292	569454	**RNA**→ COG4708 (COG4708)→
Spy-10-1	NC_008021.1	+	609013	609175	**RNA**→ COG4708 (COG4708)→
Spy-3-1	NC_008022.1	+	602440	602602	**RNA**→ COG4708 (COG4708)→
Spy-6-1	NC_008023.1	+	610421	610583	**RNA**→ COG4708 (COG4708)→
Spy-5-1	NC_008024.1	+	629301	629463	**RNA**→ COG4708 (COG4708)→
Spy-1-1	NC_002737.1	+	610434	610596	**RNA**→ COG4708 (COG4708)→
Spy-9-1	NC_003485.1	+	649988	650150	**RNA**→ COG4708 (COG4708)→
Spy-2-1	NZ_AAFV01000127.1	+	2701	2863	**RNA**→ COG4708 (COG4708)→
Sag-4-1	NC_007432.1	+	925396	925551	**RNA**→ COG4708 (COG4708)→
Sag-2-1	NC_004116.1	+	847184	847339	**RNA**→ COG4708 (COG4708)→
Sag-8-1	NC_004368.1	+	891258	891413	**RNA**→ COG4708 (COG4708)→
Sag-1-1	NZ_AAJO01000019.1	-	14506	14351	**RNA**→ COG4708 (COG4708)→
Sag-3-1	NZ_AAJP01000038.1	-	19582	19427	**RNA**→ COG4708 (COG4708)→
Sag-5-1	NZ_AAJQ01000003.1	-	48527	48372	**RNA**→ COG4708 (COG4708)→
Sag-7-1	NZ_AAJS01000011.1	-	5688	5533	**RNA**→ COG4708 (COG4708)→
Sag-6-1	NZ_AAJR01000052.1	+	12937	13092	**RNA**→ COG4708 (COG4708)→
Spn-2-1	NC_003028.1	+	955901	956039	**RNA**→ COG4708 (COG4708)→ tRNA-Thr→
Spn-1-1	NC_003098.1	+	904161	904299	**RNA**→ COG4708 (COG4708)→ tRNA-Thr→
Spn-2-2	NZ_AAGY02000049.1	+	9540	9678	**RNA**→ COG4708 (COG4708)→
Smu-1-1	NC_004350.1	-	1474191	1474053	**RNA**→ COG4708 (COG4708)→
[1]Lla-1-1	NC_002662.1	+	802553	802700	**RNA**→ COG4708 (COG4708)→

```
YP_329594.1           ---------------------MNTFTTRDYAHMAIVTAIYIVLTITPPFNAIAYGAYQFRVSEMLNFLAFYHRKYLFAVTLGCMISN...
ZP_00786020.1         ---------------------MNTFTTRDYAHMAIVTAIYIVLTITPPFNAIAYGAYQFRVSEMLNFLAFYHRKYLFAVTLGCMISN...
YP_059907.1           ---------------------MTKLTVHDYVHIGLVAALYVVLTITPPLNAISYGMYQFRISEMMNFLAFYHRKYIIAVTLGCMIAN...
NP_721887.1           ---------------------MKQLTVRDLAHIAIVAALYVALTATPPLNAISYGGIQFRLSEMLNFLAFYNPKYIIAVTLGCMIAN...
NP_345488.1           ---------------------MKKLTIRDVADIAIVAAIYVVLTVTPPLNAISYGAYQFRISEMMNFMAFYNPKYIIGVTIGCMIAN...
NP_266951.1-extra5    MSLHPFAQAYYRRILSPKENLLKKSKTYDIVTIAIVAALYVILTMTPGLSAISYGPIQFRVSEMLNFTAFFNKKYIIAVTIGCMISN...
                      : :  :       ::     ::.  .  * . :.:*:*:*: ** ** :.**:**  ***:***:** **:: **::.**:****:*
```

Figure 3: Alignment of COG4708 proteins

COG4708 Predicted membrane protein [Function unknown]

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Sag-1-1, Sag-2-1, Sag-3-1, Sag-5-1, Sag-7-1, Sag-8-1, Spn-1-1, Spn-2-2, Spy-1-1, Spy-10-1, Spy-2-1, Spy-3-1, Spy-5-1, Spy-6-1, Spy-7-1, Spy-8-1, Spy-9-1

alignment positions 1 · · · 160

```
 Spy-4-1  UGUUUUAUGGUACAAUGUAACCGUUGC...UUU.....G.CAAC.CCUUGGUGCUUAGUUCCCUUUCACCAAGCAUAUUAGAAACGUAGAUAGCUUUUGA.GAGU.UAUCCAGACGUU.AAAGGAGAAAAA.AUGACAAAAUUGACUGUACACGACUAUG
 Sag-4-1  UGUGAUAUA...CUUAGUAA..GUUGU...UUUAC...A.CAAC.CCUUGAUGCUUAGUUCCCUUUCAUCAAGCAUAUUACAAACGUU.AGAAGUGUU.A..ACUAUCU..AAACGUU.AAAGGAGAAAUU.AUGAACACAUUUACAACUCGCGAUUACG
 Sag-6-1  UGUGAUAUA...CUUAGUAA..GUUGU...UUUAC...A.CAAC.CCUUGAUGCUUAGUUCCCUUUCAUCAAGCAUAUUACAAACGUU.AGAAGUGUU.A..ACUGUCU..AAACGUU.AAAGGAGAAAUU.AUGAACACAUUUACAACUCGCGAUUACG
 Spn-2-1  UGUGCUAUA...CU.AGUAAG.GUUGA...AUGAA...U.CAAC.CCUUGGUGCUUAGCUU.CUUUCACCAAGCAUAUUACACGCGG........AU.A...........ACCGCC.AAAGGAGAAAAG.AUGAAAAAAUUAACUAUUCGUGAUGUUG
 Smu-1-1  UAUGCUAUA...CUGGAUAA..GUUGU...CUG.....A.CAAC.CCUUGGUGCUUAUUUCCCUUUCACCAAGCAUAUUACAAACGGU.......AGAU............CCGUU.AAAGGAGAAACA.AUGAAACAACUAACUGUUCGUGAUUUAG
1Lla-1-1  UCUGCUAAA...AU.AAAAAC.GUUGCUCAAUAG.UGAGUUAAC.CCUUGAGCCUUCAUCCCU.UUGCUCAAGCAUAUUAUAGGCGAA.......UU.U..........UGUCGCC.AAGGAGAACCUUUUGAAAAAAUCAAAAACUUAUGAUAUUG
          ......................<<<<<<<<.....>>>>.>>>>..<<<<<<<.............>>>>>>>.........<<<<<..<<<<<<.......>>>.>>>....>>>>>..........................................
          ......................00012000.....0002.1000..0000222.............2220000.........?2002..020120.......021.020....2002?..........................................
          ...........................................................<<<<<<<.....................................................>>>>>.>>.................................
          ...........................................................?1?1?00.....................................................00?1?.1?.................................
          URUGYUAUA---CU○RRUAA○-GUUGY---●U●○----R-CAAC-CCUUGRUGCUUARUUCCCUUUCAYCAAGCAUAUUAYARRCG●○--------●U-●-----------○R●CGYY-AAAGGAGAAA●●-AUGAAAAAAUURAC●RYUCRYGAU●●YG
                                                                                                                                      SD+++       AUG
```

alignment positions 161 · · · 177

```
 Spy-4-1  UUCACAUUGGGCUAGUA
 Sag-4-1  CUCACAUGGCUAUUGUA
 Sag-6-1  CUCACAUGGCUAUUGUA
 Spn-2-1  CAGAUAUUGCAAUCGUC
 Smu-1-1  CCCAUAUCGCCAUUGUG
1Lla-1-1  UUACAAUCGCAAUUGUU
          .................
          .................
          .................
          .................
          YY●AYAUYGC●AUYGU●
```

2.3.1 Annotation of start codon in COG4708 genes

To demonstrate the plausibility of our start codon annotation, we extracted the downstream proteins of these sequences and aligned them with CLUSTALW [13]; see Figure 3 (page 39). We show the N-terminal portion (i.e., 5′ portion, in nucleotide terms) of the protein. This alignment demonstrates (1) that the start codon we annotated is more likely to be correct, given the unaligned 5′ region, and (2) that the aligned part of the sequences (corresponding with our annotated start codon) appears to be a conserved coding region, even though the N-terminal part is less well conserved, particularly at the nucleotide level. We note again that both the annotated start codon and our start codon in NP_266951 are UUG, so this does not make one prediction more likely to be correct than the other.

2.4 The *sucA* motif

The genes downstream of the RNA are all annotated (in some cases) as members of the *suc* operon. We therefore suspect that they are transcribed as a unit, and annotated them as such. Recall that our automated operon prediction scheme was simply to assume that any gap of more than 100 nucleotides would end an operon. Some genes downstream of sucA RNA have gaps above 100 nucleotides, so would not by default be shown together. *sucA*, *aceF*, *lpd* and COG1485 (putative ATPase) are conserved in the same order in all species with sucA RNA, although gap lengths vary.

There is a possible homolog in *Dechloromonas aromatica* in sequence NC_007298.1 from nucleotides 3080843 to 3080715. This sequence folds into a similar structure, with a much longer P1 loop and largely lacking P4 stems, but we were unconvinced by the homology, particularly at the sequence level. As well, it was assigned a marginal E-value even in a search confined to *sucA* UTRs within β-proteobacteria only.

Notes:

---

[1] Overlaps hypothetical gene, but the gene is not conserved in any bacteria.

[2] These sequences come from the Global Ocean Survey [12], and were downloaded from the CAMERA site at `http://camera.calit2.net/` related to CAMERA identifier CAM_PUB_Rusch07a.

[3] This sequence is derived from samples obtained from an estuary in Delaware Bay, New Jersey, USA.

[4] This sequence is derived from samples obtained from an estuary in Chesapeake Bay, Maryland, USA.

[5] This sequence is derived from samples obtained from a hypersaline lagoon at Punta Cormorant in Floreana Island, Ecuador.

[6] This sequence is derived from samples obtained from Lake Gatún in Panama.

[7] This genetic information downloaded from `http://camera.calit2.net` may be considered to be part of the genetic patrimony of the country from which the sample was obtained. Users of this information agree to: (1) acknowledge the country of origin in any publications where the genetic information is presented and (2) contact the CBD focal point identified on the CBD website (`http://www.biodiv.org/doc/info-centre.shtml`) if they intend to use the genetic information for commercial purposes.

abbrev. of hits	**taxonomy of species**
Bbr-1-1 to Bbr-1-1	β-proteobacteria Burkholderiales Alcaligenaceae *Bordetella bronchiseptica* RB50
Bpa-1-1 to Bpa-1-1	β-proteobacteria Burkholderiales Alcaligenaceae *Bordetella parapertussis* 12822
Bpe-1-1 to Bpe-1-1	β-proteobacteria Burkholderiales Alcaligenaceae *Bordetella pertussis* Tohama I
Bam-1-1 to Bam-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia ambifaria* AMMD
Bce-1-1 to Bce-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* PC184
Bdo-1-1 to Bdo-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia dolosa* AUO158
Bsp-1-1 to Bsp-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia sp.* 383
Bvi-1-1 to Bvi-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia vietnamiensis* G4
Bma-1-1 to Bma-1-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* 10229
Bma-2-1 to Bma-2-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* 10399
Bma-3-1 to Bma-3-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* 2002721280
Bma-4-1 to Bma-4-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* ATCC 23344
Bma-5-1 to Bma-5-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* FMH
Bma-6-1 to Bma-6-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* GB8 horse 4
Bma-7-1 to Bma-7-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* JHU
Bma-8-1 to Bma-8-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* NCTC 10247
Bma-9-1 to Bma-9-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* SAVP1
Bxe-1-1 to Bxe-1-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia xenovorans* LB400
Bps-1-1 to Bps-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1106a
Bps-2-1 to Bps-2-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1106b
Bps-3-1 to Bps-3-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1655
Bps-4-1 to Bps-4-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1710a
Bps-5-1 to Bps-5-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1710b
Bps-6-1 to Bps-6-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 406e
Bps-7-1 to Bps-7-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 668
Bps-8-1 to Bps-8-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* K96243
Bps-9-1 to Bps-9-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* Pasteur
Bps-10-1 to Bps-10-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* S13
Bth-1-1 to Bth-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia thailandensis* E264
Reu-1-1 to Reu-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Cupriavidus *Ralstonia eutropha* H16
Reu-2-1 to Reu-2-1	β-proteobacteria Burkholderiales Burkholderiaceae Cupriavidus *Ralstonia eutropha* JMP134
Rme-1-1 to Rme-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Cupriavidus *Ralstonia metallidurans* CH34
Rso-1-1 to Rso-1-1	β-proteobacteria Burkholderiales Burkholderiaceae *Ralstonia solanacearum* GMI1000
Rso-2-1 to Rso-2-1	β-proteobacteria Burkholderiales Burkholderiaceae *Ralstonia solanacearum* UW551
Aav-1-1 to Aav-1-1	β-proteobacteria Burkholderiales Comamonadaceae *Acidovorax avenae subsp. citrulli* AAC00-1
Asp-1-1 to Asp-1-1	β-proteobacteria Burkholderiales Comamonadaceae *Acidovorax sp.* JS42
Pna-1-1 to Pna-1-1	β-proteobacteria Burkholderiales Comamonadaceae *Polaromonas naphthalenivorans* CJ2
Psp-1-1 to Psp-1-1	β-proteobacteria Burkholderiales Comamonadaceae *Polaromonas sp.* JS666
Rfe-1-1 to Rfe-1-1	β-proteobacteria Burkholderiales Comamonadaceae *Rhodoferax ferrireducens* T118
Vei-1-1 to Vei-1-1	β-proteobacteria Burkholderiales Comamonadaceae *Verminephrobacter eiseniae* EF01-2
Rge-1-1 to Rge-1-1	β-proteobacteria Burkholderiales *Rubrivivax gelatinosus* PM1
env-1 to env-7	environmental samples

abbrev	RefSeq accession		5′ at	3′ at	genes
[2,3]env-1	JCVI_SCAF_1101668124968	-	170	84	**RNA**→
Vei-1-1	NZ_AASQ01000011.1	-	66218	66123	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Aav-1-1	NZ_AASX01000003.1	-	235516	235421	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
env-2	AACY01261444.1	-	619	533	**RNA**→ unknown→
Asp-1-1	NZ_AASD01000010.1	-	45337	45242	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
[2,5,7]env-3	JCVI_SCAF_1096627140354	-	957	870	**RNA**→
Rfe-1-1	NC_007908.1	-	2549152	2549046	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Pna-1-1	NZ_AANM01000006.1	-	135596	135492	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Psp-1-1	NC_007948.1	-	2758692	2758588	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
[2,4]env-4	JCVI_SCAF_1096627456667	+	900	988	**RNA**→
Rge-1-1	NZ_AAEM01000001.1	+	683107	683203	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ EmrE (COG2076)→ Lpd (COG1249)→ COG1485 (COG14
Reu-2-1	NC_007347.1	-	2250080	2249984	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Reu-1-1	NC_008313.1	-	2524712	2524616	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Rme-1-1	NC_007973.1	-	2222411	2222315	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Rso-1-1	NC_003295.1	+	1348529	1348625	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Rso-2-1	NZ_AAKL01000042.1	-	24703	24608	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bsp-1-1	NC_007510.1	+	1635427	1635523	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bce-1-1	NZ_AAKX01000009.1	-	11093	10997	**RNA**→ SucA (COG0567)→ 2-oxoacid_dh (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
[1]Bvi-1-1	NZ_AAEH02000086.1	-	11737	11641	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→
Bam-1-1	NZ_AAJL01000003.1	-	480235	480139	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bth-1-1	NC_007651.1	-	2917736	2917640	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bps-8-1	NC_006350.1	-	2274711	2274615	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bps-5-1	NC_007434.1	+	2098508	2098604	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bma-4-1	NC_006348.1	-	1100353	1100257	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bma-1-1	NZ_AAHM02000001.1	-	176237	176141	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bma-2-1	NZ_AAHN02000001.1	+	184426	184522	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bma-6-1	NZ_AAHO01000128.1	-	3390	3294	**RNA**→ SucA (COG0567)→
Bma-8-1	NZ_AAHP01000017.1	+	58415	58511	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bma-9-1	NZ_AAHQ02000001.1	+	1688315	1688411	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bps-3-1	NZ_AAHR02000071.1	-	20447	20351	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bps-4-1	NZ_AAHS02000055.1	+	9012	9108	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bps-7-1	NZ_AAHU01000012.1	-	81715	81619	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bps-9-1	NZ_AAHV02000071.1	+	10601	10697	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bps-10-1	NZ_AAHW02000073.1	+	9018	9114	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bma-5-1	NZ_AAIQ02000028.1	-	46755	46659	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bma-7-1	NZ_AAIR02000025.1	-	54481	54385	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bps-1-1	NZ_AAMA01000072.1	-	23180	23084	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bps-2-1	NZ_AAMB01000040.1	-	23582	23486	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bps-6-1	NZ_AAMM02000070.1	-	23147	23051	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bma-3-1	NZ_AANX02000014.1	-	50484	50388	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bdo-1-1	NZ_AAKY01000089.1	-	10360	10264	**RNA**→ SucA (COG0567)→ 2-oxoacid_dh (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bxe-1-1	NC_007951.1	+	1834734	1834831	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
env-5	AACY01084698.1	+	1793	1890	**RNA**→ unknown→
[2,6,7]env-6	JCVI_SCAF_1096627014744	+	1448	1534	**RNA**→
[2,6,7]env-7	JCVI_SCAF_1101668234623	-	978	892	**RNA**→
Bbr-1-1	NC_002927.3	-	3885643	3885548	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bpa-1-1	NC_002928.3	-	3469201	3469106	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→
Bpe-1-1	NC_002929.2	+	1184722	1184817	**RNA**→ SucA (COG0567)→ SucB/AceF (COG0508)→ Lpd (COG1249)→ COG1485 (COG1485)→

COG0508 Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase (E2) component, and related enzymes [Energy production and conversion]
COG0567 2-oxoglutarate dehydrogenase complex, dehydrogenase (E1) component, and related enzymes [Energy production and conversion]
COG1249 Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component, and related enzymes [Energy production and conversion]
COG1485 Predicted ATPase [General function prediction only]
COG2076 Membrane transporters of cations and cationic drugs [Inorganic ion transport and metabolism]

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Bam-1-1, Bma-1-1, Bma-2-1, Bma-3-1, Bma-4-1, Bma-5-1, Bma-6-1, Bma-7-1, Bma-8-1, Bma-9-1, Bpa-1-1, Bpe-1-1, Bps-1-1, Bps-10-1, Bps-2-1, Bps-3-1, Bps-4-1, Bps-5-1, Bps-6-1, Bps-7-1, Bps-8-1, Bps-9-1, env-5

alignment positions 1 · · · 117

```
2,3 env-1    ACAGUCGC.AUCCGCCAACCGGUCCAGCCGUGUCGCGGAAGGUAGUUC...........AACC..AACUAAU.GUUCCCUGAA..GGGG.AGCGAA.GGUCA.GCGAAAA.AUGAGU
    Vei-1-1  ACAGUCGC.AUCCGCCAACCGGUCCAGCCGUGUCGCGGAAGGUUUUCC...........AACC..AACUCAU.GCUUUCCGCA..GGAA.AGCGCA.GGUCA.GCGCAUC.AUGAGC
    Aav-1-1  ACAGUCGC.AUCCGCCAACCGGUUCAGCCGUGUCGCGGAAGGUUUUACC..........AACC..AGCUAAU.GCUUCCCGCA..GGGA.AGCGGA.GGUCA.GCGGAA..AUGAGC
    env-2    ACAGUCGC.AUCCGCCAACCGGUUCAGCCGUGUCGCGGAAGGUUUCUC...........AACC..AGCUAAU.GCUUUCCCAA..GGAA.AGCGGA.GGUCA.GCGGAAU.AUGAGC
    Asp-1-1  ACAGUCGC.AUCCGCCAACCGGUUCAGCCGUGACGCGGAAGGUUUGUU...........AACC..AGCUAAC.GCUUCCUUCA..GGGA.AGUGGA.GGUCA.GCGGAAU.AUGAGC
2,5,7 env-3  ACAGUCGC.AUCCGCCAAUCGGUCCAGCCGUGUCGCGGAAGGUUGUUU...........AACC.AGUUCAAA.GUUUCCUGCA..GGGA.AACGGA.GGUCA.GCGCA...AUGAGC
    Rfe-1-1  ACAGUCGC.AUCCGCCAAACGGUCAAGCCGUGACGCGGAAGGUUUUUAAAAACAGCUUUAACC..AGCUCAA.GUUUCCUGCA..GGGA.AACGGA.GGUCA.GCGCAAU.AUGAGU
    Pna-1-1  ACAGUCGC.AUCCGCUAAUCGGUUCAGCCGUGUUGCGGAAGGUUAAUUAA.CCAGCUUUAACC..AGCUAAU.GUUUCCCAAA..GGGA.AACGGA.GGUCA.GCGGAU..AUGAGU
    Psp-1-1  ACAGUCGC.AUCCGCUAAUCGGUCCAGCCGUGUUGCGGAAGGUUAAUCAA.CCAGCUUUAACC..AGCUAAU.GUUUCCCCAA..GGGA.AACGGA.GGUCA.GCGGAU..AUGAGU
2,4 env-4    NNNNNNNC.AUCCGCUAACCGGUUCAGCCGUGUCG.GGGAGGUUUUUU........UUCAACC..AACUAAU.GUUCCCUGCA..GGGG.AACGGA.GGUCA.GCGCAU..AUGAGU
    Rge-1-1  CGGGUCGCAAUCCGCCAACCGGUCCAGCCGUGUCGCGGAAGGUUUUUC...........AACC..AGCCAAC.GUCCUGGAUA..CCAG.GGUGAAAGGUGA.GCGAAG..AUGAUG
    Reu-2-1  UUGUUUGCGAUCCGCUAACCGGUCAAGCCGUGUCGCGGAAGGUUGAAU...........AACC..CGCUGAA..CUCCGGCAUACCCGG.AG.AAA.AGUGA.GCGUCCC.AUGAUG
    Reu-1-1  UUGUUUGCGAUCCGCUAACCGGUCAAGCCGUGUCGCGGAAGGUUGAAU...........AACC..CGCUGAA..CUCCGGCAUUCCCGG.AG.AAU.AGUGA.GCGUCCC.AUGAUG
    Rme-1-1  UUGUUUGCAAUCCGCUAACCGGUCGAGCCGUGUCGCGGAAGGUUGAUG...........AACC..CGCUUAA..CUCCGGCAUUCCCGG.AG.AAU.AGUGA.GCGCCCC.AUGAUG
    Rso-1-1  AAGUUUGCGAUCCGCUAACCGGUCAAGCCGUGUCGCGGAAGGUUGAUG...........AACC..CGCUGAA..CUCCGGCAGACCCGG.AG.AAA.GGUGA.GCGCCCC.AUGACU
    Rso-2-1  AAGUUUGCGAUCCGCUAACCGGUCAAGCCGUGUCGCGGAAGGUUGAUG...........AACC..CGCUGAA..CUCCGGCAGACCCGG.AG.AA..GGUGA.GCGCCCC.AUGACU
    Bsp-1-1  UGAUUCGCAAUCCGCUAAACGGUCAGGCCGUGUCGCGGAAGGUUGAGU...........AACC..CGCUAUU.UCUCGAGAAA..CUCGAAGAAA..GGUGA.GCGUAAA.AUGUCA
    Bce-1-1  UGAUUCGCAAUCCGCUAAACGGUCAGGCCGUGUCGCGGAAGGUUGAGU...........AACC..CGCUAUU.UCUCGAGAAG..CUCGAAGAAA..GGUGA.GCGUAAA.AUGUCA
1   Bvi-1-1  UGAUUCGCAAUCCGCUAAACGGUCAGGCCGUGUCGCGGAAGGUUGAGC...........AACC..CGCUAUU.UCUCGAGAAG..CUCGAAGAAA..GGUGA.GCGUAAA.AUGUCA
    Bth-1-1  AUUCGCGCAAUCCGCUAAACGGUCAGGCCGUGUCGCGGAAGGUUGAGU...........AACC..CGCUAUU.UCUCGAGAAG..CUCGAAGAAA..GGUGA.GCGUCAA.AUGUCA
    Bdo-1-1  AUUCGCGCAAUCCGCUAAACGGUCAGGCCGUGUCGCGGAAGGUUGAGU...........AACC..CGCUAUU.UCUCGAGAAG..CUCGAAGAAA..GGUGA.GCGUAAA.AUGUCA
    Bxe-1-1  AUUCGCGCAAUCCGCUAAACGGUCAGGCCGUGUCGCGGAAGGUUCUGU...........AACC..CGCUAUU.UCUCGAGAAA..CUCGAAGAAA..GGUGA.GCGUAAAAAUGAUG
2,6,7 env-6  UGCUUUGCGUUCUGCUAAUCGGUAAAGCCGUGUCGCAGAUGGUUUU.............AACCACAGUUU...UUUUCGGGAUACCCGAUGAAA...GGUGA.GCAUC...AUGAUG
2,6,7 env-7  UGCUUUGCGUUCUGCUAAUCGGUGAAGCCGUGUCGCAGAUGGUUUA.............AACCACAGUUU...UUUUCGGGAUACCCGAUGAAA...GGUGA.GCAUC...AUGAUG
    Bbr-1-1  GCCCCUGCUAUCCGCUAAACGGGAAGCCCGUGUCGCGGAAGG..UUUU.............CCUGCAUCGUAU.CGAGUGAAG..CACU.CG.UCAAGGUGAAGCGAUUU.AUGUCU
             .....<<<..<<<<<....<<<<...>>>>....>>>>>.<<<<...............>>>>..<<<<....<<<<<<<.....>>>>.>>>....>>>>..>>>...........
             .....200..102?0....0002...2000....0?201.000?...............?000..?112....2222222.....2222.222....211?..002...........
             ●●●●YYGC○AUCCGCYAA●CGGUY●RGCCGUGUCGCGGAAGGUU●●●Y-----------AACC--●RYUR●●-○YUY●●●R●R--●●●R-RR○R●○-GGU●A-GCG●●●○○-AUGA●●
                  P1   P2a      P2b                 P3                       P4a     P4b                 SD?+++            AUG
```

2.5 The 23S-methyl motif

Notes:

[1] This putative homolog is upstream of two genes on the opposite strand. The next gene, ≈ 2.5 Kb downstream, is an annotated rRNA methyltransferase gene, and is on the same strand. Using translated BLAST, we were unable to find any closer gene on the same strand. Note: this putative homolog was *not* found by looking upstream of rRNA methylase genes.

[2] This putative homolog is near the end of a sequence fragment, so the downstream gene is unknown.

abbrev. of hits	**taxonomy of species**
Efa-1-1 to Efa-1-1	Firmicutes Lactobacillales Enterococcaceae *Enterococcus faecalis* V583
Efa-2-1 to Efa-2-1	Firmicutes Lactobacillales Enterococcaceae *Enterococcus faecium* DO
Lca-1-1 to Lca-1-1	Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus casei* ATCC 334
Lpl-1-1 to Lpl-1-1	Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus plantarum* WCFS1
Lre-1-1 to Lre-1-1	Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus reuteri* 100-23
Lre-2-1 to Lre-2-1	Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus reuteri* JCM 1112
Lsa-1-1 to Lsa-1-1	Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus sakei subsp. sakei* 23K
Lsa-2-1 to Lsa-2-1	Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus salivarius subsp. salivarius* UCC118
Ppe-1-1 to Ppe-1-1	Firmicutes Lactobacillales Lactobacillaceae *Pediococcus pentosaceus* ATCC 25745
Lla-1-1 to Lla-1-1	Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. cremoris* SK11
Lla-2-1 to Lla-2-1	Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. lactis* Il1403
Sag-1-1 to Sag-1-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 18RS21
Sag-2-1 to Sag-2-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 2603V/R
Sag-3-1 to Sag-3-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 515
Sag-4-1 to Sag-4-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* A909
Sag-5-1 to Sag-5-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* CJB111
Sag-6-1 to Sag-6-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* COH1
Sag-7-1 to Sag-7-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* H36B
Sag-8-1 to Sag-8-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* NEM316
Smu-1-1 to Smu-1-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus mutans* UA159
Spn-1-1 to Spn-1-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* R6
Spn-2-1 to Spn-2-2	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* TIGR4
Spy-1-1 to Spy-1-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* M1 GAS
Spy-2-1 to Spy-2-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10270
Spy-3-1 to Spy-3-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10394
Spy-4-1 to Spy-4-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10750
Spy-5-1 to Spy-5-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS2096
Spy-6-1 to Spy-6-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS315
Spy-7-1 to Spy-7-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS5005
Spy-8-1 to Spy-8-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS6180
Spy-9-1 to Spy-9-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS8232
Spy-10-1 to Spy-10-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS9429
Spy-11-1 to Spy-11-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* SSI-1
Ssu-1-1 to Ssu-1-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus suis* 89/1591
Sth-1-1 to Sth-1-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* CNRZ1066
Sth-2-1 to Sth-2-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* LMD-9
Sth-3-1 to Sth-3-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* LMG 18311

abbrev	**RefSeq accession**		**5′ at**	**3′ at**	**genes**
Spy-3-1	NC_006086.1	+	348439	348550	**RNA**→ CspR (COG0219)→

Spy-7-1	NC_007297.1	+	318993	319104	**RNA**→ CspR (COG0219)→
Spy-1-1	NC_002737.1	+	318399	318510	**RNA**→ CspR (COG0219)→
Spy-8-1	NC_007296.1	+	319891	320002	**RNA**→ CspR (COG0219)→
Spy-10-1	NC_008021.1	+	322981	323092	**RNA**→ CspR (COG0219)→
Spy-5-1	NC_008023.1	+	323907	324018	**RNA**→ CspR (COG0219)→
Spy-4-1	NC_008024.1	+	316889	317000	**RNA**→ CspR (COG0219)→
Spy-9-1	NC_003485.1	+	349475	349586	**RNA**→ CspR (COG0219)→
Spy-2-1	NC_008022.1	+	320578	320689	**RNA**→ CspR (COG0219)→
Spy-11-1	NC_004606.1	-	1587863	1587752	**RNA**→ CspR (COG0219)→
Spy-6-1	NC_004070.1	+	308882	308993	**RNA**→ CspR (COG0219)→
Sag-5-1	NZ_AAJQ01000051.1	+	10437	10549	**RNA**→ CspR (COG0219)→
Sag-7-1	NZ_AAJS01000102.1	-	3919	3807	**RNA**→ CspR (COG0219)→
Sag-4-1	NC_007432.1	-	1597751	1597639	**RNA**→ CspR (COG0219)→
Sag-2-1	NC_004116.1	-	1593794	1593682	**RNA**→ CspR (COG0219)→
Sag-8-1	NC_004368.1	-	1698231	1698119	**RNA**→ CspR (COG0219)→
Sag-1-1	NZ_AAJO01000406.1	+	289	401	**RNA**→ CspR (COG0219)→
Sag-3-1	NZ_AAJP01000101.1	-	3231	3119	**RNA**→ CspR (COG0219)→
Sag-6-1	NZ_AAJR01000063.1	+	5861	5973	**RNA**→ CspR (COG0219)→
[2]Ssu-1-1	NZ_AAFA02000101.1	+	6262	6374	**RNA**→
Spn-2-1	NC_003028.1	+	466468	466571	**RNA**→ CspR (COG0219)→
Spn-1-1	NC_003098.1	+	432756	432859	**RNA**→ CspR (COG0219)→
Spn-2-2	NZ_AAGY02000109.1	+	2700	2803	**RNA**→ CspR (COG0219)→
Sth-3-1	NC_006448.1	+	267727	267838	**RNA**→ CspR (COG0219)→
Sth-1-1	NC_006449.1	+	267767	267878	**RNA**→ CspR (COG0219)→
Sth-2-1	NZ_AAGS01000063.1	-	8014	7903	**RNA**→ CspR (COG0219)→
Smu-1-1	NC_004350.1	-	1618054	1617939	**RNA**→ CspR (COG0219)→
Lla-2-1	NC_002662.1	-	1162967	1162862	**RNA**→ CspR (COG0219)→
Lla-1-1	NZ_AAGO01000033.1	-	788	683	**RNA**→ CspR (COG0219)→
Efa-2-1	NZ_AAAK03000109.1	+	25	138	**RNA**→ CspR (COG0219)→ GpmB (COG0406)→ RecD (COG0507)→
Efa-1-1	NC_004668.1	-	2578245	2578132	**RNA**→ CspR (COG0219)→
Lsa-1-1	NC_007576.1	+	496014	496123	**RNA**→ CspR (COG0219)→ hypo→
Lca-1-1	NZ_AAGR01000005.1	-	56003	55882	**RNA**→ CspR (COG0219)→
[1]Lpl-1-1	NC_004567.1	-	1998562	1998450	**RNA**→
Lre-2-1	NZ_AAOV01000032.1	-	14837	14718	**RNA**→ CspR (COG0219)→ FtsK (COG1674)→
Lre-1-1	NZ_AAPZ01000011.1	-	46363	46244	**RNA**→ CspR (COG0219)→ FtsK (COG1674)→
Lsa-2-1	NC_007929.1	-	1171104	1170996	**RNA**→ CspR (COG0219)→ FtsK (COG1674)→
Ppe-1-1	NZ_AAEV01000008.1	-	54736	54652	**RNA**→ CspR (COG0219)→ FtsK (COG1674)→

COG0219 Predicted rRNA methylase (SpoU class) [Translation, ribosomal structure and biogenesis]

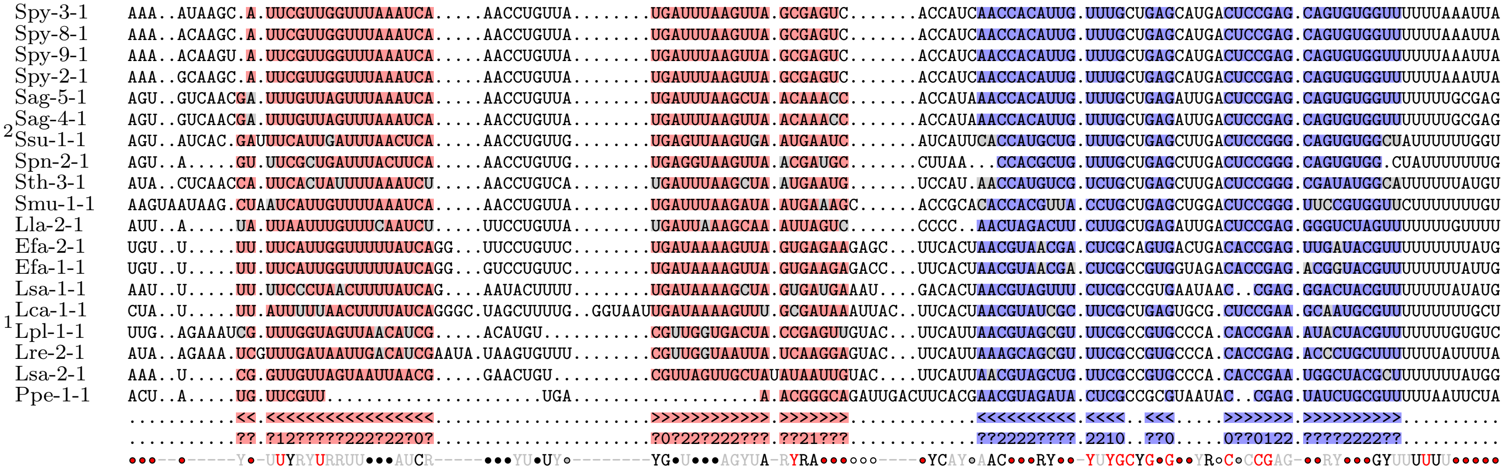
COG0406 Fructose-2,6-bisphosphatase [Carbohydrate transport and metabolism]

COG0507 ATP-dependent exoDNAse (exonuclease V), alpha subunit - helicase superfamily I member [DNA replication, recombination, and repair]

COG1674 DNA segregation ATPase FtsK/SpoIIIE and related proteins [Cell division and chromosome partitioning]

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Lla-1-1, Lre-1-1, Sag-1-1, Sag-2-1, Sag-3-1, Sag-6-1, Sag-7-1, Sag-8-1, Spn-1-1, Spn-2-2, Spy-1-1, Spy-10-1, Spy-11-1, Spy-4-1, Spy-5-1, Spy-6-1, Spy-7-1, Sth-1-1, Sth-2-1

alignment positions 1 · · · 139



```
  Spy-3-1   AAA..AUAAGC.A.UUCGUUGGUUUAAAUCA.....AACCUGUUA........UGAUUUAAGUUA.GCGAGUC.......ACCAUCAACCACAUUG.UUUGCUGAGCAUGACUCCGAG.CAGUGUGGUUUUUUAAAUUA
  Spy-8-1   AAA..ACAAGC.A.UUCGUUGGUUUAAAUCA.....AACCUGUUA........UGAUUUAAGUUA.GCGAGUC.......ACCAUCAACCACAUUG.UUUGCUGAGCAUGACUCCGAG.CAGUGUGGUUUUUUAAAUUA
  Spy-9-1   AAA..ACAAGU.A.UUCGUUGGUUUAAAUCA.....AACCUGUUA........UGAUUUAAGUUA.GCGAGUC.......ACCAUCAACCACAUUG.UUUGCUGAGCAUGACUCCGAG.CAGUGUGGUUUUUUAAAUUA
  Spy-2-1   AAA..GCAAGC.A.UUCGUUGGUUUAAAUCA.....AACCUGUUA........UGAUUUAAGUUA.GCGAGUC.......ACCAUCAACCACAUUG.UUUGCUGAGCAUGACUCCGAG.CAGUGUGGUUUUUUAAAUUA
  Sag-5-1   AGU..GUCAACGA.UUUGUUAGUUUAAAUCA.....AACCUGUUA........UGAUUUAAGCUA.ACAAACC.......ACCAUAAACCACAUUG.UUUGCUGAGAUUGACUCCGAG.CAGUGUGGUUUUUUUGCGAG
  Sag-4-1   AGU..GUCAACGA.UUUGUUAGUUUAAAUCA.....AACCUGUUA........UGAUUUAAGUUA.ACAAACC.......ACCAUAAACCACAUUG.UUUGCUGAGAUUGACUCCGAG.CAGUGUGGUUUUUUUGCGAG
2 Ssu-1-1   AGU..AUCAC.GAUUUCAUUGAUUUAACUCA.....AACCUGUUG........UGAGUUAAGUGA.AUGAAUC.......AUCAUUCACCAUGCUG.UUUGCUGAGCUUGACUCCGGG.CAGUGUGGCUAUUUUUUGGU
  Spn-2-1   AGU..A.....GU.UUCGCUGAUUUACUUCA.....AACCUGUUG........UGAGGUAAGUUA.ACGAUGC.......CUUAA...CCACGCUG.UUUGCUGAGCUUGACUCCGGG.CAGUGUGG.CUAUUUUUUUG
  Sth-3-1   AUA..CUCAACCA.UUCACUAUUUUAAAUCU.....AACCUGUCA........UGAUUUAAGCUA.AUGAAUG.......UCCAU.AACCAUGUCG.UCUGCUGAGCUUGACUCCGGG.CGAUAUGGCAUUUUUUAUGU
  Smu-1-1   AAGUAAUAAG.CUAAUCAUUGUUUUAAAUCA.....AACCUGUUA........UGAUUUAAGAUA.AUGAAAGC......ACCGCACACCACGUUA.CCUGCUGAGCUGGACUCCGGG.UUCCGUGGUUCUUUUUUUGU
  Lla-2-1   AUU..A.....UA.UUAAUUUGUUUCAAUCU.....UUCCUGUUA........UGAUUAAAGCAA.AUUAGUC.......CCCC..AACUAGACUU.CUUGCUGAGAUUGACUCCGAG.GGGUCUAGUUUUUUUGUUUU
  Efa-2-1   UGU..U.....UU.UUCAUUGGUUUUUAUCAGG...UUCCUGUUC........UGAUAAAAGUUA.GUGAGAAGAGC...UUCACUAACGUAACGA.CUCGCAGUGACUGACACCGAG.UUGAUACGUUUUUUUUUAUG
  Efa-1-1   UGU..U.....UU.UUCAUUGGUUUUUAUCAGG...GUCCUGUUC........UGAUAAAAGUUA.GUGAAGAGACC...UUCACUAACGUAACGA.CUCGCCGUGGUAGACACCGAG.ACGGUACGUUUUUUUUAUUG
  Lsa-1-1   AAU..U.....UU.UUCCCUAACUUUUAUCAG....AAUACUUUU........UGAUAAAAGCUA.GUGAUGAAAU....GACACUAACGUAGUUU.CUCGCCGUGAAUAAC..CGAG.GGACUACGUUUUUUUAUAUG
  Lca-1-1   CUA..U.....UU.AUUUUUAACUUUUAUCAGGGC.UAGCUUUUG..GGUAAUUGAUAAAAGUUU.GCGAUAAAUUAC..UUCACUAACGUAUCGC.UUCGCUGAGUGCG.CUCCGAA.GCAAUGCGUUUUUUUUUGCU
1 Lpl-1-1   UUG..AGAAAUCG.UUUGGUAGUUAACAUCG.....ACAUGU...........CGUUGGUGACUA.CCGAGUUGUAC...UUCAUUAACGUAGCGU.UUCGCCGUGCCCA.CACCGAA.AUACUACGUUUUUUGUGUC
  Lre-2-1   AUA..AGAAA.UCGUUUGAUAAUUGACAUCGAAUA.UAAGUGUUU........CGUUGGUAAUUA.UCAAGGAGUAC...UUCAUUAAAGCAGCGU.UUCGCCGUGCCCA.CACCGAG.ACCCUGCUUUUUUUAUUUUA
  Lsa-2-1   AAA..U.....CG.GUUGUUAGUAAUUAACG.....GAACUGU..........CGUUAGUUGCUAUAUAAUUGUAC....UUCAUUAACGUAGCUG.UUCGCCGUGCCCA.CACCGAA.UGGCUACGCUUUUUUUAUGG
  Ppe-1-1   ACU..A.....UG.UUCGUU......................UGA...................A.ACGGGCAGAUUGACUUCACGAACGUAGAUA.CUCGCCGCGUAAUAC..CGAG.UAUCUGCGUUUUUAAUUCUA
            ...........<<.<<<<<<<<<<<<<<<<<......................>>>>>>>>>>>>.>>>>>>>.............<<<<<<<<<<.<<<<..<<<.....>>>>>>>.>>>>>>>>>>..........
            ...........??.?12?????222?22?0?......................?0?22?222???.??21???.............??2222????.2210..??0.....0??0122.????2222??..........
            ●●●--●-----Y●-UUYRYURRUU●●●AUCR-----●●●YU●UY○--------YG●U●●●AGYUA-RYRA●●●○○○----●YCAY○AAC●●●RY●●-YUYGCYG●G●●YR○C○CCGAG-●●RY●●●GYUUUUUU●●●●●
```

2.6 The *hemB* motif

Note: many sequences overlap the 3′ end of “GTP-binding protein” by about 40 nucleotides. It is possible that the constraints that this overlap places on both the ORF and the RNA are not serious.

abbrev. of hits	**taxonomy of species**
Bam-1-1 to Bam-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia ambifaria* AMMD
Bce-1-1 to Bce-1-2	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* AU 1054
Bce-2-1 to Bce-2-2	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* HI2424
Bce-3-1 to Bce-3-2	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* PC184
Bdo-1-1 to Bdo-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia dolosa* AUO158
Bsp-1-1 to Bsp-1-2	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia sp.* 383
Bvi-1-1 to Bvi-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia vietnamiensis* G4
Bma-1-1 to Bma-1-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* 10229
Bma-2-1 to Bma-2-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* 10399
Bma-3-1 to Bma-3-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* 2002721280
Bma-4-1 to Bma-4-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* ATCC 23344
Bma-5-1 to Bma-5-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* FMH
Bma-6-1 to Bma-6-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* GB8 horse 4
Bma-7-1 to Bma-7-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* JHU
Bma-8-1 to Bma-8-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* NCTC 10247

Bma-9-1 to Bma-9-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* SAVP1
Bxe-1-1 to Bxe-1-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia xenovorans* LB400
Bps-1-1 to Bps-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1106a
Bps-2-1 to Bps-2-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1106b
Bps-3-1 to Bps-3-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1655
Bps-4-1 to Bps-4-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1710a
Bps-5-1 to Bps-5-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1710b
Bps-6-1 to Bps-6-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 406e
Bps-7-1 to Bps-7-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 668
Bps-8-1 to Bps-8-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* K96243
Bps-9-1 to Bps-9-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* Pasteur
Bps-10-1 to Bps-10-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* S13
Bth-1-1 to Bth-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia thailandensis* E264
Reu-1-1 to Reu-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Cupriavidus *Ralstonia eutropha* H16
Reu-2-1 to Reu-2-1	β-proteobacteria Burkholderiales Burkholderiaceae Cupriavidus *Ralstonia eutropha* JMP134
Rme-1-1 to Rme-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Cupriavidus *Ralstonia metallidurans* CH34
Rso-1-1 to Rso-1-1	β-proteobacteria Burkholderiales Burkholderiaceae *Ralstonia solanacearum* GMI1000
Rso-2-1 to Rso-2-1	β-proteobacteria Burkholderiales Burkholderiaceae *Ralstonia solanacearum* UW551
Asp-1-1 to Asp-1-1	β-proteobacteria Burkholderiales Comamonadaceae *Acidovorax sp.* JS42
Pna-1-1 to Pna-1-1	β-proteobacteria Burkholderiales Comamonadaceae *Polaromonas naphthalenivorans* CJ2
Psp-1-1 to Psp-1-1	β-proteobacteria Burkholderiales Comamonadaceae *Polaromonas sp.* JS666
Rfe-1-1 to Rfe-1-1	β-proteobacteria Burkholderiales Comamonadaceae *Rhodoferax ferrireducens* T118
Rge-1-1 to Rge-1-1	β-proteobacteria Burkholderiales *Rubrivivax gelatinosus* PM1
Tde-1-1 to Tde-1-1	β-proteobacteria Hydrogenophilales Hydrogenophilaceae *Thiobacillus denitrificans* ATCC 25259
Mfl-1-1 to Mfl-1-1	β-proteobacteria Methylophilales Methylophilaceae *Methylobacillus flagellatus* KT
Cvi-1-1 to Cvi-1-1	β-proteobacteria Neisseriales Neisseriaceae *Chromobacterium violaceum* ATCC 12472
Neu-1-1 to Neu-1-1	β-proteobacteria Nitrosomonadales Nitrosomonadaceae *Nitrosomonas europaea* ATCC 19718
Neu-2-1 to Neu-2-1	β-proteobacteria Nitrosomonadales Nitrosomonadaceae *Nitrosomonas eutropha* C71
Nmu-1-1 to Nmu-1-1	β-proteobacteria Nitrosomonadales Nitrosomonadaceae *Nitrosospira multiformis* ATCC 25196
Asp-2-1 to Asp-2-1	β-proteobacteria Rhodocyclales Rhodocyclaceae *Azoarcus sp.* EbN1
Dar-1-1 to Dar-1-1	β-proteobacteria Rhodocyclales Rhodocyclaceae *Dechloromonas aromatica* RCB

abbrev	RefSeq accession		$5'$ at	$3'$ at	genes
Bth-1-1	NC_007651.1	+	3483591	3483692	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)DsbDgamma (cd02953)
Bsp-1-1	NC_007510.1	-	340248	340152	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)
Bce-1-1	NC_008060.1	+	3003024	3003120	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)
Bce-2-1	NZ_AAHL01000010.1	+	118443	118539	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)
Bce-3-1	NZ_AAKX01000036.1	+	46066	46162	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)
Bam-1-1	NZ_AAJL01000014.1	-	12222	12126	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)
Bsp-1-2	NC_007510.1	-	340248	340149	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)
Bce-1-2	NC_008060.1	+	3003024	3003123	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)
Bce-2-2	NZ_AAHL01000010.1	+	118443	118542	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)
Bce-3-2	NZ_AAKX01000036.1	+	46066	46165	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)
Bdo-1-1	NZ_AAKY01000127.1	+	3544	3640	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)
Bma-4-1	NC_006348.1	+	2711794	2711889	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)DsbDgamma (cd02953)
Bma-1-1	NZ_AAHM02000001.1	-	1978857	1978762	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)DsbDgamma (cd02953)

Bma-2-1	NZ_AAHN02000036.1	+	27668	27763	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)DsbDgamma (cd02953)
Bma-6-1	NZ_AAHO01000046.1	+	26960	27055	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)DsbDgamma (cd02953)
Bma-8-1	NZ_AAHP01000073.1	-	15701	15606	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)DsbDgamma (cd02953)
Bma-9-1	NZ_AAHQ02000001.1	-	46540	46445	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)DsbDgamma (cd02953)
Bma-5-1	NZ_AAIQ02000040.1	+	26976	27071	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)DsbDgamma (cd02953)
Bma-7-1	NZ_AAIR02000039.1	+	26973	27068	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)DsbDgamma (cd02953)
Bma-3-1	NZ_AANX02000049.1	+	28921	29016	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)DsbDgamma (cd02953)
Bps-8-1	NC_006350.1	+	3790869	3790964	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)DsbDgamma (cd02953)
Bps-5-1	NC_007434.1	+	4084282	4084377	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)DsbDgamma (cd02953)
Bps-3-1	NZ_AAHR02000026.1	-	18475	18380	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)DsbDgamma (cd02953)
Bps-4-1	NZ_AAHS02000015.1	+	115543	115638	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)DsbDgamma (cd02953)
Bps-7-1	NZ_AAHU01000011.1	-	18488	18393	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)DsbDgamma (cd02953)
Bps-9-1	NZ_AAHV02000014.1	-	18480	18385	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)DsbDgamma (cd02953)
Bps-10-1	NZ_AAHW02000029.1	+	80251	80346	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)DsbDgamma (cd02953)
Bps-1-1	NZ_AAMA01000030.1	+	59482	59577	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)DsbDgamma (cd02953)
Bps-2-1	NZ_AAMB01000047.1	+	35379	35474	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)DsbDgamma (cd02953)
Bps-6-1	NZ_AAMM02000024.1	-	16212	16117	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)DsbDgamma (cd02953)
Bvi-1-1	NZ_AAEH02000055.1	-	14114	14018	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)
Bxe-1-1	NC_007951.1	+	4496259	4496363	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)DsbDgamma (cd02953)
Psp-1-1	NC_007948.1	-	269077	268981	←VanZ like protein ←COG1981 (COG1981) **RNA**→ ALAD (pfam00490)→ CorA (COG0598)→
Rfe-1-1	NC_007908.1	+	4244992	4245088	←VanZ like protein ←COG1981 (COG1981) **RNA**→ ALAD (pfam00490)→ CorA (COG0598)→
Rso-2-1	NZ_AAKL01000032.1	-	18786	18678	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)
Rso-1-1	NC_003295.1	+	3222003	3222105	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG4232 (COG4232)
Rme-1-1	NC_007973.1	+	3561927	3562029	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ CorA (COG0598)→
Reu-2-1	NC_007347.1	+	3460223	3460323	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ CorA (COG0598)→
Reu-1-1	NC_008313.1	+	3733625	3733720	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ CorA (COG0598)→
Mfl-1-1	NC_007947.1	-	372278	372176	Cpn10 (pfam00166)→ GroEL (cd03344)→ **RNA**→ UvrD (COG0210)→ ←hypo
Asp-2-1	NC_006513.1	-	626325	626226	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←COG1011 (COG1011)
Nmu-1-1	NC_007614.1	-	2372743	2372646	COG1521 (COG1521)→ COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←MrcA (COG5009)

Tde-1-1	NC_007404.1	+	201454	201549	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ hypo→
Cvi-1-1	NC_005085.1	+	4728513	4728607	CspR (COG0219)→ RlpA (COG0797)→ **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Pna-1-1	NZ_AANM01000002.1	+	129119	129218	←putative membrane protein ←COG1981 (COG1981) **RNA**→ ALAD (pfam00490)→ CorA (COG0598)→
Rge-1-1	NZ_AAEM01000003.1	+	151950	152052	←hypo ←COG1981 (COG1981) **RNA**→ ALAD (pfam00490)→ CorA (COG0598)→
Dar-1-1	NC_007298.1	+	690971	691067	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←AcrR (COG1309)
Neu-2-1	NZ_AAJE01000012.1	+	18796	18888	COG1521 (COG1521)→ COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←PstB (COG1117)
Neu-1-1	NC_004757.1	-	2684455	2684363	COG1521 (COG1521)→ COG0218 (COG0218)→ **RNA**→ ALAD (pfam00490)→ ←GckA (COG2379)
Asp-1-1	NZ_AASD01000001.1	-	356570	356483	←VanZ like protein ←COG1981 (COG1981) **RNA**→ ←LplA (COG0095) ALAD (pfam00490)→

cd02953 DsbD gamma family; DsbD gamma is the C-terminal periplasmic domain of the bacterial protein DsbD.
cd03344 GroEL_like type I chaperonin.
COG0095 Lipoate-protein ligase A [Coenzyme metabolism]
COG0210 Superfamily I DNA and RNA helicases [DNA replication, recombination, and repair]
COG0218 Predicted GTPase [General function prediction only]
COG0219 Predicted rRNA methylase (SpoU class) [Translation, ribosomal structure and biogenesis]
COG0598 Mg2+ and Co2+ transporters [Inorganic ion transport and metabolism]
COG0797 Lipoproteins [Cell envelope biogenesis, outer membrane]
COG1011 Predicted hydrolase (HAD superfamily) [General function prediction only]
COG1117 ABC-type phosphate transport system, ATPase component [Inorganic ion transport and metabolism]
COG1309 Transcriptional regulator [Transcription]
COG1521 Putative transcriptional regulator, homolog of Bvg accessory factor [Transcription]
COG1981 Predicted membrane protein [Function unknown]
COG2379 Putative glycerate kinase [Carbohydrate transport and metabolism]
COG2863 Cytochrome c553 [Energy production and conversion]
COG4232 Thiol:disulfide interchange protein [Posttranslational modification, protein turnover, chaperones / Energy production and conversion]
COG5009 Membrane carboxypeptidase/penicillin-binding protein [Cell envelope biogenesis, outer membrane]
pfam00166 Chaperonin 10 Kd subunit.
pfam00490 Delta-aminolevulinic acid dehydratase.

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Bce-1-1, Bce-1-2, Bce-2-1, Bce-2-2, Bce-3-1, Bce-3-2, Bma-1-1, Bma-2-1, Bma-3-1, Bma-5-1, Bma-6-1, Bma-7-1, Bma-8-1, Bma-9-1, Bps-1-1, Bps-10-1, Bps-2-1, Bps-3-1, Bps-4-1, Bps-5-1, Bps-6-1, Bps-7-1, Bps-9-1

alignment positions 1 · · · 144

```
Bth-1-1  UUCGG.GCUGCCCAUAAAAAA.ACCCGCCGU..UCA...........A.CGGCGGGU.UA.AA......CA.GCCU..AAUCGAA..A.....AACGAC...AGGCA.CCCGCU.CAGGGAGGAGAAGCGGG.GA..GCAUUGC
Bsp-1-1  UUCGG.CCGGCUCAUAAAAAA.ACCCGCCG...UUGAA...........CGGCGGGU.CA.AA......CA.GCCU..AAUCGAA..A.....AACGAC...AGGCA.CCCGCU.CAGGGAGGAGAAGCGGG.GA.......GC
Bam-1-1  UUCGG.CCGGCUCAUAAAAAA.ACCCGCCG...UCGAA...........CGGCGGGU.CA.AA......CA.GCCU..AAUCGAA..U.....AACGAC...AGGCA.CCCGCU.CAGGGAGGAGAAGCGGG.GA.......GC
Bsp-1-2  UUCGG.CCGGCUCAUAAAAAA.ACCCGCCG...UUGAA...........CGGCGGGU.CA.AA......CA.GCCU..AAUCGAA..A.....AACGAC...AGGCA.CCCGCU.CAGGGAGGAGAAGCGGG.GA....GCUGC
Bdo-1-1  UUCGU.CCGGCCCAUAAAAAA.ACCCGCCG...UUUAC...........CGGCGGGU.CA.AA......CA.GCCU..AAUCGAA..U.....AACGAC...AGGCA.CCCGCU.CAGGGAGGAGAAGCGGG.GA.......GC
Bma-4-1  UGCGG.CUGCCC.AUAAAAAA.ACCCGCCG...GUCAA...........CGGCGGGU.UA.AA......CA.GCCU..AAUCGAA..A.....AACGAC...AGGCA.CUCGCU.CAGGGAGGAGAAGCGGG.GA.......GC
Bps-8-1  UGCGG.CUGCCC.AUAAAAAA.ACCCGCCG...UUCAA...........CGGCGGGU.UA.AA......CA.GCCU..AAUCGAA..A.....AACGAC...AGGCA.CCCGCU.CAGGGAGGAGAAGCGGG.GA.......GC
Bvi-1-1  UUCGG.CCGGCUCAUAAAAAA.ACCCGCCGU..CGA...........A.CGGCGGGU.CA.AA......CA.GCCU..AAUCGAA..A.....AACGAC...AGGCA.CCCGCU.CAGGGAGGAGAAGCGGG.GA.......GC
Bxe-1-1  UUCGG.CCGCCC.AUAAAAAA.ACCCGCCG...CUUUAA..........CGGCGGGU.CA.AA......CA.GCCU..AAUCGAAA.A.....AACGAC...AGGCA.CCCGCU.CAGGGAGGAGAAGCGGG.GAGGUCAGCGC
Psp-1-1  ........GGC..A.AAAAAAUGCCCGGUUGGCACGAA.......GCC.AACCGGGCUGGUAA.........GCCUUU..UUGCU..GUC..ACACG.C.AAAGGC..CCCGCU.CAGGGAGGAAAAGCGGG..A......UGC
Rfe-1-1  AGCAU.CAGGC..AAAAAAAA.CCCCAGCU...CGAG............GGCUGGGGUGGUAA.........GCCU..AUUUGCU..GUC..ACACAUCA..AGGC..CCCGCU.CAGGGAGGAAAAGCGGG..A......GGC
Rso-2-1  ACCGA.ACAGGC.AAAAAAAA.CCCCGUCGC..CAGACGC.......G.CAACGGGG..AUAA.......C.GUCC..CAUCGAACGGUC...GUCGAU...GGGUA.CCCGCGUCAGGGAGGUGCCGCGGGGGA...CACCGC

         ......................<<<<<<<<<<<............>>>.>>>>>>>>...............<<<<<<..<<<............>>>..>>>>>>..<<<<<<............>>>>>>............
         ......................200022??220............022.??220002...............?12?22..??2............2??..22?21?..022002............200220............
         ○○C○○-○○○○○--A○AAAAAA-●CCCG●Y○---●RRR○○----------○R●CGGG●-○R-AA-------○-GCCY----●●●○○--●○○---○○●●●----AGGY○-CCCGCU-CAGGGAGG●RAAGCGGG○GA------○GC
```

```
Rso-1-1   ..CGAGCCGGC..AAAAAAAA.CCCCGUCGC..CAGACGC.......G.CAACGGGG.GG.AA.......U.GACC..CAUCGAGC.C.....GUCGAU...GGGG..CCCGCGUCAGGGAGGCGUCGCGGGGGA...CACCGC
Rme-1-1   UUCGAACCGGC..AAAAAAAAGCCCCGUUG...CGAA............CAACGGGGACG.AA......CUUUCCC....GCCGGU.U....CAGGGC....GGGUA.CCCGCU.CAGGGAGGAGUAGCGGGGGA..CAUCCGC
Reu-2-1   AGCGC.GCGGAC.AAAAAAAAGCCCCGUUG...CAAG............CAACGGGGACG.AA......CUUUCCC....GCCGGU.U....GA.GGC....GGGUA.CCCGCU.CAGGGAGGAGUAGCGGGGGA...CACCGC
Reu-1-1   AGCAC.GC.....AAAAAAAAGCCCCGCUG...CAAG............CAACGGGGACA.AA......CUUUCCC....GCCGCU.G......GGGC....GGGUA.CCCGCU.CAGGGAGGAGUAGCGGGGGA...CACCGC
Mfl-1-1   ........GGCA.A.AAAAAA.CCCCGAUU...AAAGGGGAGUAA....AAUCGGGG.UU.AAGCAUUUAA.GCCU....CAU...UGGG.....AUA....AGGCUUCCCGCU.CAGGGAGGA.AAGCGGGAGA..CGAGUGC
Asp-2-1   ..CGG...GGCA.AUAAAAAA.CCCCGGGCU..AAAGGGGAUAA...A.GCCCGGGGUCA.AA.......C.GCCU....UAA...CUGG..A..UUA....AGGCA.CCCGCU.CAGGGAGGUGAAGCGGGAGA.....UGGC
Nmu-1-1   ......CCAGAA.AUAAAAAACCCCCGGUU...AAAGGGGAAUAA....AACCGGGGGCA.AA......CU.ACCU....UAA....UAG..GGAUUA....AGGUA.CCCGCU.CAGGGAGGAAAAGCGGGAGA........C
Tde-1-1   AGCGG.GC.....AACAAAAAGCCCCGACC...AAGGGAGG........GAUCGGGG.CA.AAAA....AU.GCCU....UAA.GGGGA......UUA....AGGCAUCCCGCU.CAGGGAGGAGAAGCGGGAGA........C
Cvi-1-1   GGCCG.CGCCC..AAUAAAAAGCCCCGGCC...GAGAA...........AGCCGGGG.CA.AA.........GCCUG...GAUAGC.AAC....GAUU...CAGGC..CCCGCU.CAGGGAGCCA.AGCGGGAGG......UGC
Pna-1-1   UGCGG.CC.....AAAAAAAA.UCCCGGC.G..ACCUGCUGGC....CA.GCCGGGA.UA.AA.......U.GCCU..UUUUG..CUGUCACACGCAA....AGGC..CCCGCU.CAGGGAGGAAAAGCGGG..A......UGC
Rge-1-1   ACCGG.CCAGGC.AUAAAAAA.GCCCCGGC...AAA.............GCCGGGGCUUG.AA.......A.GCCUGG...CGCU..GUC..ACACGG..UCAGGCU.CCUGCU.CAGGGAGGUGAAGCAGG.GA.AUGAACGC
Dar-1-1   ........AG..AAAACAAAGGGCCCCCGGAU.AAGGGGA......GU.CCGGGGGC.CA.AA......AU.GCCC....UAA.CGUU..G....UCA....AGGCA.CCCGCU.CAGGGAGGUGAAGCGGGAGA...CAGCGC
Neu-2-1   AGCAA.UC....AAAAAAAAA.CCCCGGCUAA.AGGGGA.......UUAAGCCGGGG..AGAA.......UCGCCU....UG....CUUG......UA....AGGC..CUCGCCUCAGGGAGGGAAGGCGAG..A......GGC
Neu-1-1   AGCGA.GC....AAAAAAAAA.CCCCGGCUAA.AGGGGA.......UUGAGCCGGGG..AGAA.......UCGCCU....UG....CUUG......UA....AGGC..CUCGCCUCAGGGAGGGAAGGCGAG..A......GGU
Asp-1-1   ..CGA...GGC..AAAAAAAU.GCCCGG.....AAAU..............CCGGGCGGA.AAACC...CCGCCGC....GUC....GCA.....GGCA...GCGGGGCCCGCU.CAGGGAGGUGAAGCGGG.GG.......GC
          ......................<<<<<<<<<<<............>>>.>>>>>>>>...............<<<<<<..<<<............>>>..>>>>>>..<<<<<<............>>>>>>............
          ......................200022??220............022.??220002...............?12?22..??2............2??..22?21?..022002............200220............
          ◦◦C◦◦-◦◦◦◦◦--A◦AAAAAA-•CCCG•Y◦---•RRR◦◦----------◦R•CGGG•-◦R-AA-------◦-GCCY----•••◦◦--•◦◦---◦◦•••----AGGY◦-CCCGCU-CAGGGAGG•RAAGCGGG◦GA------◦GC
```

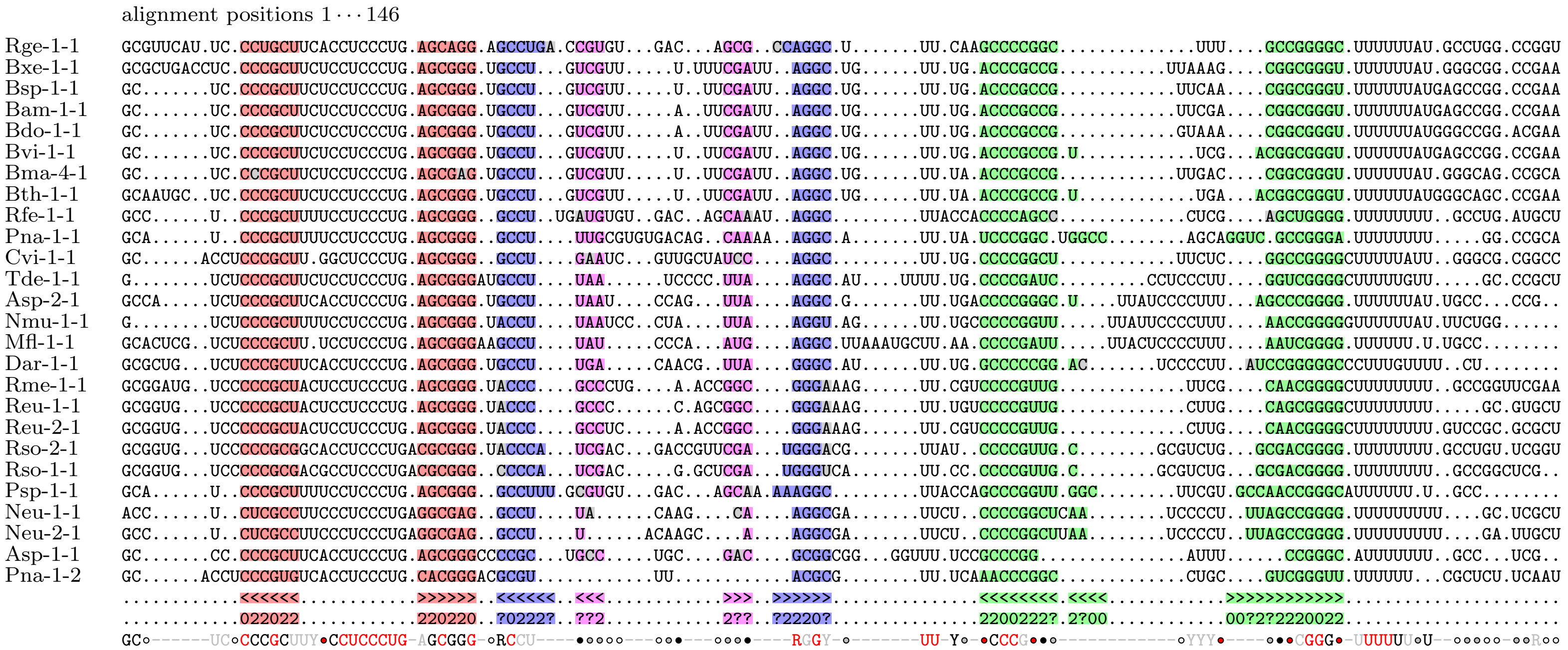
2.7 anti-hemB

abbrev. of hits	**taxonomy of species**
Bam-1-1 to Bam-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia ambifaria* AMMD
Bce-1-1 to Bce-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* AU 1054
Bce-2-1 to Bce-2-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* HI2424
Bce-3-1 to Bce-3-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* PC184
Bdo-1-1 to Bdo-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia dolosa* AUO158
Bsp-1-1 to Bsp-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia sp.* 383
Bvi-1-1 to Bvi-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia vietnamiensis* G4
Bma-1-1 to Bma-1-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* 10229
Bma-2-1 to Bma-2-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* 10399
Bma-3-1 to Bma-3-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* 2002721280
Bma-4-1 to Bma-4-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* ATCC 23344
Bma-5-1 to Bma-5-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* FMH
Bma-6-1 to Bma-6-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* GB8 horse 4
Bma-7-1 to Bma-7-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* JHU
Bma-8-1 to Bma-8-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* NCTC 10247
Bma-9-1 to Bma-9-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* SAVP1
Bxe-1-1 to Bxe-1-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia xenovorans* LB400
Bth-1-1 to Bth-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia thailandensis* E264
Reu-1-1 to Reu-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Cupriavidus *Ralstonia eutropha* H16
Reu-2-1 to Reu-2-1	β-proteobacteria Burkholderiales Burkholderiaceae Cupriavidus *Ralstonia eutropha* JMP134
Rme-1-1 to Rme-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Cupriavidus *Ralstonia metallidurans* CH34
Rso-1-1 to Rso-1-1	β-proteobacteria Burkholderiales Burkholderiaceae *Ralstonia solanacearum* GMI1000
Rso-2-1 to Rso-2-1	β-proteobacteria Burkholderiales Burkholderiaceae *Ralstonia solanacearum* UW551
Asp-1-1 to Asp-1-1	β-proteobacteria Burkholderiales Comamonadaceae *Acidovorax sp.* JS42
Pna-1-1 to Pna-1-2	β-proteobacteria Burkholderiales Comamonadaceae *Polaromonas naphthalenivorans* CJ2
Psp-1-1 to Psp-1-1	β-proteobacteria Burkholderiales Comamonadaceae *Polaromonas sp.* JS666
Rfe-1-1 to Rfe-1-1	β-proteobacteria Burkholderiales Comamonadaceae *Rhodoferax ferrireducens* T118
Rge-1-1 to Rge-1-1	β-proteobacteria Burkholderiales *Rubrivivax gelatinosus* PM1
Tde-1-1 to Tde-1-1	β-proteobacteria Hydrogenophilales Hydrogenophilaceae *Thiobacillus denitrificans* ATCC 25259
Mfl-1-1 to Mfl-1-1	β-proteobacteria Methylophilales Methylophilaceae *Methylobacillus flagellatus* KT
Cvi-1-1 to Cvi-1-1	β-proteobacteria Neisseriales Neisseriaceae *Chromobacterium violaceum* ATCC 12472
Neu-1-1 to Neu-1-1	β-proteobacteria Nitrosomonadales Nitrosomonadaceae *Nitrosomonas europaea* ATCC 19718
Neu-2-1 to Neu-2-1	β-proteobacteria Nitrosomonadales Nitrosomonadaceae *Nitrosomonas eutropha* C71
Nmu-1-1 to Nmu-1-1	β-proteobacteria Nitrosomonadales Nitrosomonadaceae *Nitrosospira multiformis* ATCC 25196
Asp-2-1 to Asp-2-1	β-proteobacteria Rhodocyclales Rhodocyclaceae *Azoarcus sp.* EbN1
Dar-1-1 to Dar-1-1	β-proteobacteria Rhodocyclales Rhodocyclaceae *Dechloromonas aromatica* RCB

abbrev	RefSeq accession		5′ at	3′ at	genes
Rge-1-1	NZ_AAEM01000003.1	-	152052	151950	←CorA (COG0598) ←ALAD (pfam00490) **RNA**→ COG1981 (COG1981)→ hypo→
Bxe-1-1	NC_007951.1	-	4496363	4496259	COG4232 (COG4232)DsbDgamma (cd02953)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Bsp-1-1	NC_007510.1	+	340152	340248	COG4232 (COG4232)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Bce-1-1	NC_008060.1	-	3003120	3003024	COG4232 (COG4232)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Bce-2-1	NZ_AAHL01000010.1	-	118539	118443	COG4232 (COG4232)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Bce-3-1	NZ_AAKX01000036.1	-	46162	46066	COG4232 (COG4232)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Bam-1-1	NZ_AAJL01000014.1	+	12126	12222	COG4232 (COG4232)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Bdo-1-1	NZ_AAKY01000127.1	-	3640	3544	COG4232 (COG4232)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Bvi-1-1	NZ_AAEH02000055.1	+	14018	14114	COG4232 (COG4232)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Bma-4-1	NC_006348.1	-	2711889	2711794	COG4232 (COG4232)DsbDgamma (cd02953)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Bma-1-1	NZ_AAHM02000001.1	+	1978762	1978857	COG4232 (COG4232)DsbDgamma (cd02953)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Bma-2-1	NZ_AAHN02000036.1	-	27763	27668	COG4232 (COG4232)DsbDgamma (cd02953)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Bma-6-1	NZ_AAHO01000046.1	-	27055	26960	COG4232 (COG4232)DsbDgamma (cd02953)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Bma-8-1	NZ_AAHP01000073.1	+	15606	15701	COG4232 (COG4232)DsbDgamma (cd02953)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Bma-9-1	NZ_AAHQ02000001.1	+	46445	46540	COG4232 (COG4232)DsbDgamma (cd02953)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Bma-5-1	NZ_AAIQ02000040.1	-	27071	26976	COG4232 (COG4232)DsbDgamma (cd02953)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Bma-7-1	NZ_AAIR02000039.1	-	27068	26973	COG4232 (COG4232)DsbDgamma (cd02953)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Bma-3-1	NZ_AANX02000049.1	-	29016	28921	COG4232 (COG4232)DsbDgamma (cd02953)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Bth-1-1	NC_007651.1	-	3483692	3483591	COG4232 (COG4232)DsbDgamma (cd02953)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Rfe-1-1	NC_007908.1	-	4245088	4244992	←CorA (COG0598) ←ALAD (pfam00490) **RNA**→ COG1981 (COG1981)→ VanZ like protein→
Pna-1-1	NZ_AANM01000002.1	-	129218	129119	←CorA (COG0598) ←ALAD (pfam00490) **RNA**→ COG1981 (COG1981)→ putative membrane protein→
Cvi-1-1	NC_005085.1	-	4728607	4728513	←COG2863 (COG2863) COG0218 (COG0218)→ **RNA**→ ←RlpA (COG0797) ←CspR (COG0219)
Tde-1-1	NC_007404.1	-	201549	201454	←hypo ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Asp-2-1	NC_006513.1	+	626226	626325	COG1011 (COG1011)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Nmu-1-1	NC_007614.1	+	2372646	2372743	MrcA (COG5009)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) ←COG1521 (COG1521)
Mfl-1-1	NC_007947.1	+	372176	372278	hypo→ ←UvrD (COG0210) **RNA**→ ←GroEL (cd03344) ←Cpn10 (pfam00166)
Dar-1-1	NC_007298.1	-	691067	690971	AcrR (COG1309)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Rme-1-1	NC_007973.1	-	3562029	3561927	←CorA (COG0598) ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Reu-1-1	NC_008313.1	-	3733720	3733625	←CorA (COG0598) ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Reu-2-1	NC_007347.1	-	3460323	3460223	←CorA (COG0598) ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Rso-2-1	NZ_AAKL01000032.1	+	18678	18786	COG4232 (COG4232)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Rso-1-1	NC_003295.1	-	3222105	3222003	COG4232 (COG4232)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) COG2863 (COG2863)→
Psp-1-1	NC_007948.1	+	268981	269077	←CorA (COG0598) ←ALAD (pfam00490) **RNA**→ COG1981 (COG1981)→ VanZ like protein→
Neu-1-1	NC_004757.1	+	2684363	2684455	GckA (COG2379)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) ←COG1521 (COG1521)
Neu-2-1	NZ_AAJE01000012.1	-	18888	18796	PstB (COG1117)→ ←ALAD (pfam00490) **RNA**→ ←COG0218 (COG0218) ←COG1521 (COG1521)
Asp-1-1	NZ_AASD01000001.1	+	356483	356570	←ALAD (pfam00490) LplA (COG0095)→ **RNA**→ COG1981 (COG1981)→ VanZ like protein→
Pna-1-2	NZ_AANM01000011.1	+	13971	14054	←hypo DUF932 (pfam06067)→ **RNA**→ RadC (pfam04002)→ ParB-like nuclease→

cd02953 DsbD gamma family; DsbD gamma is the C-terminal periplasmic domain of the bacterial protein DsbD.
cd03344 GroEL_like type I chaperonin.
COG0095 Lipoate-protein ligase A [Coenzyme metabolism]
COG0210 Superfamily I DNA and RNA helicases [DNA replication, recombination, and repair]
COG0218 Predicted GTPase [General function prediction only]
COG0219 Predicted rRNA methylase (SpoU class) [Translation, ribosomal structure and biogenesis]
COG0598 Mg2+ and Co2+ transporters [Inorganic ion transport and metabolism]
COG0797 Lipoproteins [Cell envelope biogenesis, outer membrane]
COG1011 Predicted hydrolase (HAD superfamily) [General function prediction only]
COG1117 ABC-type phosphate transport system, ATPase component [Inorganic ion transport and metabolism]
COG1309 Transcriptional regulator [Transcription]
COG1521 Putative transcriptional regulator, homolog of Bvg accessory factor [Transcription]
COG1981 Predicted membrane protein [Function unknown]
COG2379 Putative glycerate kinase [Carbohydrate transport and metabolism]
COG2863 Cytochrome c553 [Energy production and conversion]
COG4232 Thiol:disulfide interchange protein [Posttranslational modification, protein turnover, chaperones / Energy production and conversion]
COG5009 Membrane carboxypeptidase/penicillin-binding protein [Cell envelope biogenesis, outer membrane]
pfam00166 Chaperonin 10 Kd subunit.
pfam00490 Delta-aminolevulinic acid dehydratase.
pfam04002 RadC, DNA repair protein.
pfam06067 Domain of unknown function (DUF932).

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Bce-1-1, Bce-2-1, Bce-3-1, Bma-1-1, Bma-2-1, Bma-3-1, Bma-5-1, Bma-6-1, Bma-7-1, Bma-8-1, Bma-9-1



alignment positions 1 · · · 146

```
Rge-1-1   GCGUUCAU.UC.CCUGCUUCACCUCCCUG.AGCAGG.AGCCUGA.CCGUGU...GAC...AGCG..CCAGGC.U.......UU.CAAGCCCCGGC.............UUU....GCCGGGGC.UUUUUUAU.GCCUGG.CCGGU
Bxe-1-1   GCGCUGACCUC.CCCGCUUCUCCUCCCUG.AGCGGG.UGCCU...GUCGUU.....U.UUUCGAUU..AGGC.UG......UU.UG.ACCCGCCG..........UUAAAG....CGGCGGGU.UUUUUUAU.GGGCGG.CCGAA
Bsp-1-1   GC.......UC.CCCGCUUCUCCUCCCUG.AGCGGG.UGCCU...GUCGUU.....U..UUCGAUU..AGGC.UG......UU.UG.ACCCGCCG...........UUCAA....CGGCGGGU.UUUUUUAUGAGCCGG.CCGAA
Bam-1-1   GC.......UC.CCCGCUUCUCCUCCCUG.AGCGGG.UGCCU...GUCGUU.....A..UUCGAUU..AGGC.UG......UU.UG.ACCCGCCG...........UUCGA....CGGCGGGU.UUUUUUAUGAGCCGG.CCGAA
Bdo-1-1   GC.......UC.CCCGCUUCUCCUCCCUG.AGCGGG.UGCCU...GUCGUU.....A..UUCGAUU..AGGC.UG......UU.UG.ACCCGCCG...........GUAAA....CGGCGGGU.UUUUUUAUGGGCCGG.ACGAA
Bvi-1-1   GC.......UC.CCCGCUUCUCCUCCCUG.AGCGGG.UGCCU...GUCGUU.....U..UUCGAUU..AGGC.UG......UU.UG.ACCCGCCG.U...........UCG...ACGGCGGGU.UUUUUUAUGAGCCGG.CCGAA
Bma-4-1   GC.......UC.CCCGCUUCUCCUCCCUG.AGCGAG.UGCCU...GUCGUU.....U..UUCGAUU..AGGC.UG......UU.UA.ACCCGCCG...........UUGAC....CGGCGGGU.UUUUUUAU.GGGCAG.CCGCA
Bth-1-1   GCAAUGC..UC.CCCGCUUCUCCUCCCUG.AGCGGG.UGCCU...GUCGUU.....U..UUCGAUU..AGGC.UG......UU.UA.ACCCGCCG.U...........UGA...ACGGCGGGU.UUUUUUAUGGGCAGC.CCGAA
Rfe-1-1   GCC......U..CCCGCUUUUCCUCCCUG.AGCGGG..GCCU..UGAUGUGU..GAC..AGCAAAU..AGGC.........UUACCACCCCAGCC............CUCG....AGCUGGGG.UUUUUUUU..GCCUG.AUGCU
Pna-1-1   GCA......U..CCCGCUUUUCCUCCCUG.AGCGGG..GCCU....UUGCGUGUGACAG..CAAAA..AGGC.A.......UU.UA.UCCCGGC.UGGCC.......AGCAGGUC.GCCGGGA.UUUUUUUU.....GG.CCGCA
Cvi-1-1   GC......ACCUCCCGCUU.GGCUCCCUG.AGCGGG..GCCU....GAAUC...GUUGCUAUCC....AGGC.........UU.UG.CCCCGGCU...........UUCUC....GGCCGGGGCUUUUUAUU..GGGCG.CGGCC
Tde-1-1   G........UCUCCCGCUUCUCCUCCCUG.AGCGGGAUGCCU....UAA......UCCCC.UUA....AGGC.AU....UUUU.UG.CCCCGAUC........CCUCCCUU....GGUCGGGGCUUUUUGUU.....GC.CCGCU
Asp-2-1   GCCA.....UCUCCCGCUUCACCUCCCUG.AGCGGG.UGCCU....UAAU....CCAG...UUA....AGGC.G.......UU.UGACCCCGGGC.U...UUAUCCCCUUU...AGCCCGGGG.UUUUUUAU.UGCC...CCG..
Nmu-1-1   G........UCUCCCGCUUUUCCUCCCUG.AGCGGG.UACCU....UAAUCC..CUA....UUA....AGGU.AG......UU.UGCCCCCGGUU....UUAUUCCCCUUU....AACCGGGGGUUUUUUAU.UUCUGG......
Mfl-1-1   GCACUCG..UCUCCCGCUU.UCCUCCCUG.AGCGGGAAGCCU....UAU.....CCCA...AUG....AGGC.UUAAAUGCUU.AA.CCCCGAUU....UUACUCCCCUUU....AAUCGGGG.UUUUUU.U.UGCC........
Dar-1-1   GCGCUG...UCUCCCGCUUCACCUCCCUG.AGCGGG.UGCCU....UGA.....CAACG..UUA....GGGC.AU......UU.UG.GCCCCCGG.AC......UCCCCUU..AUCCGGGGGCCCUUUGUUUU..CU........
Rme-1-1   GCGGAUG..UCCCCCGCUACUCCUCCCUG.AGCGGG.UACCC....GCCCUG....A.ACCGGC....GGGAAAG......UU.CGUCCCCGUUG............UUCG....CAACGGGGCUUUUUUUU..GCCGGUUCGAA
Reu-1-1   GCGGUG...UCCCCCGCUACUCCUCCCUG.AGCGGG.UACCC....GCCC......C.AGCGGC....GGGAAAG......UU.UGUCCCCGUUG............CUUG....CAGCGGGGCUUUUUUUU.....GC.GUGCU
Reu-2-1   GCGGUG...UCCCCCGCUACUCCUCCCUG.AGCGGG.UACCC....GCCUC....A.ACCGGC....GGGAAAG......UU.CGUCCCCGUUG............CUUG....CAACGGGGCUUUUUUUU.GUCCGC.GCGCU
Rso-2-1   GCGGUG...UCCCCCGCGGCACCUCCCUGACGCGGG.UACCCA...UCGAC...GACCGUUCGA...UGGGACG.......UUAU..CCCCGUUG.C.......GCGUCUG...GCGACGGGG.UUUUUUUU.GCCUGU.UCGGU
Rso-1-1   GCGGUG...UCCCCCGCGACGCCUCCCUGACGCGGG..CCCCA...UCGAC.....G.GCUCGA...UGGGUCA.......UU.CC.CCCCGUUG.C.......GCGUCUG...GCGACGGGG.UUUUUUUU..GCCGGCUCG..
Psp-1-1   GCA......U..CCCGCUUUUCCUCCCUG.AGCGGG..GCCUUU.GCGUGU...GAC...AGCAA.AAAGGC.........UUACCAGCCCGGUU.GGC........UUCGU.GCCAACGGGCAUUUUUU.U..GCC........
Neu-1-1   ACC......U..CUCGCCUUCCCUCCCUGAGGCGAG..GCCU....UA......CAAG....CA....AGGCGA.......UUCU..CCCCGGCUCAA.......UCCCCU..UUAGCCGGGG.UUUUUUUUU....GC.UCGCU
Neu-2-1   GCC......U..CUCGCCUUCCCUCCCUGAGGCGAG..GCCU....U......ACAAGC....A....AGGCGA.......UUCU..CCCCGGCUUAA.......UCCCCU..UUAGCCGGGG.UUUUUUUUU....GA.UUGCU
Asp-1-1   GC.......CC.CCCGCUUCACCUCCCUG.AGCGGGCCCCGC...UGCC.....UGC....GAC....GCGGCGG...GGUUU.UCCGCCCGG..............AUUU......CCGGGC.AUUUUUUU..GCC...UCG..
Pna-1-2   GC......ACCUCCCGUGUCACCUCCCUG.CACGGGACGCGU............UU............ACGCG........UU.UCAAACCCGGC............CUGC....GUCGGGUU.UUUUUU...CGCUCU.UCAAU
          ............<<<<<<............>>>>>>..<<<<<<..<<<............>>>..>>>>>>...............<<<<<<<<.<<<<...........>>>>>>>>>>>>......................
          ............022022............220220..?0222?..??2............2??..?2220?...............2200222?.2?00...........00?2?2220022......................
          GC○------UC○CCCGCUUY●CCUCCCUG-AGCGGG-○RCCU----●○○○○---○○●---○○○●----RGGY-○-------UU-Y○-●CCCG●●○------------○YYY●----○●●CGGG●-UUUUUU○U--○○○○○-○○R○○
```

2.8 The MAEB motif

2.8.1 Additional analysis

Association with glycine/citrate cycle. In addition to *gcv* genes, MAEB is associated with acetyl-CoA carboxylase, aconitase A, thiolases and citrate synthase. MAEB is also associated with several coenzyme A (CoA) derivatives: acetyl-CoA (as before), acyl-CoA, enoyl-CoA and glutaryl-CoA. Aldehyde dehydrogenase, another MAEB-associated gene, has CoA-dependent and -independent forms. MAEB is associated with sarcosine oxidase, whose substrate is glycine.

Association with primary metabolism. We now list genes occurring more than once in association with MAEB (hypothetical genes are not included).

- Genes related to primary metabolism:
  1. Dihydrodipicolinate synthase (cd00408).(S)-2,3-dihydropyridine-2,6-dicarboxylate + 2 H2O
  2. Thiolases that cleave 3-ketoacyl-CoA into acyl-CoA and acetyl-CoA (cd00751)
  3. Glutaryl-CoA dehydrogenase (cd01151)
  4. Quinolinate phosphoribosyl transferase (cd01572)
  5. NCAIR synthase (COG0026)
  6. NCAIR mutase (COG0041)

7. Tryptophan synthase alpha chain (COG0159)
8. Acyl-CoA synthetase (AMP-forming)/AMP-acid ligases II (COG0318, COG0365)
9. Citrate synthase (COG0372)
10. gcvP, Glycine cleavage system protein P (COG0403)
11. gcvT, Glycine cleavage system protein P (COG0404)
12. uncharacterized NAD(FAD)-dependent dehydrogenases (COG0446)
13. Biotin carboxyl carrier (COG0511)
14. Predicted hydrolases or acetyltransferases (COG0596)
15. Predicted dehydrogenases (COG0673)
16. Acetyl-CoA carboxylase beta subunit (COG0777)
17. Glycine cleavage system protein P, C-terminal domain (COG1003)
18. NAD-dependent aldehyde dehydrogenase (COG1012)
19. Enoyl-CoA hydratase/carnitine racemace (COG1024)
20. Aconitase A (COG1048)
21. Lactate (or related) dehydrogenase (COG1052)
22. Dihydroneopterin aldolase (COG1539)
23. Predicted acyl-CoA transferase/carnitine dehydratase (COG1804)
24. Methylase involved in ubiquinone/menaquinone biosynthesis (COG2226)
25. ABC-type sugar transporter (COG1129)
26. Sugar phophate permease (COG2271)
27. Sugar (or other) transporter (pfam00083)
28. ABC-type sugar transporter, periplasmic component (COG1879)
29. Choline dehydrogenase (COG2303)
30. Predicted carboxypeptidase (COG2866)
31. Transcriptional activator of acetoin/glycerol metabolism (COG3284)
32. Ketosteroid isomerase (COG4319)
33. Sarcosine oxidase gamma subunit (COG4583)
34. fer2 (cd00207). 2Fe-2S iron-sulfur cluster binding domain.
35. alpha/beta hydrolase (COG4757)
36. Acetyl/propionyl-CoA carboxylase, alpha subunit (COG4770)
37. Short chain dehydrogenase (pfam00106)
38. Aldehyde dehydrogenase (pfam00171)
39. Quinolinate synthetase A (pfam02445)
40. Allophanate hydrolase subunit 1 (pfam02682)
41. ATP synthase, delta/epsilon chain (pfam02823)
42. Sarcosine oxidase, delta subunit (pfam04267)

- Genes not necessarily involved in primary metabolism:
    1. Flagellar basal body rod protein (COG4786)
    2. Putative translation factor (COG0009)
    3. SUA5 domain (pfam03481)
    4. Biopolymer transport protein (COG0848)

2.8.2 Gene context and alignment

Notes:

[1] Weak homology.

[2] The loop sequence is highly unusual, but other features are conserved.

[3] This homolog, which is in front of an unusual gene for this motif, has an atypical loop sequence, and may therefore be a false positive.

[4] Weak homology on 5′ end.

[5] The downstream gene is on the opposite strand to this mini-glycine instance.

[6] OverlapsNonconserved

[7] Overlaps 5′-end of hypothetical gene on opposite strand. The overlapping part of this gene does not appear to be conserved (according to BLAST).

[8] This putative homolog lies in between annotated tRNAs and rRNAs. It has to be regarded suspiciously; unrelated RNAs are a common source of false positives because conserved stems are much more likely to (approximately) appear in them than random, non-RNA sequence.

[9] Overlaps 5′-end of downstream gene; however, BLAST indicates that this part of the gene is not conserved.

abbrev. of hits	**taxonomy of species**
Bam-1-1 to Bam-1-15	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia ambifaria* AMMD
Bce-1-1 to Bce-1-17	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* AU 1054
Bce-2-1 to Bce-2-17	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* HI2424
Bce-3-1 to Bce-3-16	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* PC184
Bdo-1-1 to Bdo-1-22	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia dolosa* AUO158
Bsp-1-1 to Bsp-1-33	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia sp.* 383
Bvi-1-1 to Bvi-1-4	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia vietnamiensis* G4
Bma-1-1 to Bma-1-13	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* 10229
Bma-2-1 to Bma-2-13	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* 10399
Bma-3-1 to Bma-3-13	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* 2002721280
Bma-4-1 to Bma-4-13	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* ATCC 23344
Bma-5-1 to Bma-5-13	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* FMH
Bma-6-1 to Bma-6-13	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* GB8 horse 4
Bma-7-1 to Bma-7-13	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* JHU
Bma-8-1 to Bma-8-16	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* NCTC 10247
Bma-9-1 to Bma-9-13	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* SAVP1
Bxe-1-1 to Bxe-1-36	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia xenovorans* LB400
Bps-1-1 to Bps-1-35	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1106a
Bps-2-1 to Bps-2-31	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1106b
Bps-3-1 to Bps-3-26	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1655
Bps-4-1 to Bps-4-23	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1710a
Bps-5-1 to Bps-5-23	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1710b
Bps-6-1 to Bps-6-32	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 406e
Bps-7-1 to Bps-7-30	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 668
Bps-8-1 to Bps-8-25	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* K96243
Bps-9-1 to Bps-9-32	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* Pasteur
Bps-10-1 to Bps-10-29	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* S13
Bth-1-1 to Bth-1-21	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia thailandensis* E264
env-1 to env-75	environmental samples

abbrev	**RefSeq accession**	**5′ at**	**3′ at**	**genes**

Bam-1-1	NZ_AAJL01000001.1	+	272992	273034	**RNA**→ **RNA**→ DUF485 (pfam04341)→ DhlC (COG4147)→
Bam-1-3	NZ_AAJL01000001.1	+	273041	273083	(shown above)
Bsp-1-1	NC_007510.1	-	3085270	3085228	**RNA**→ PurE (COG0041)→ PurK (COG0026)→ SUA5 (COG0009)SUA5 (pfam03481)→
Bdo-1-1	NZ_AAKY01000041.1	+	58852	58894	**RNA**→ **RNA**→ AdhC (COG1062)→ COG0627 (COG0627)→
Bdo-1-12	NZ_AAKY01000041.1	+	58901	58943	(shown above)
env-1	AACY01085258.1	-	855	813	**RNA**→ unknown→
Bxe-1-4	NC_007953.1	+	35411	35453	**RNA**→ XdhA (COG4630)→ CoxS (COG2080)→ Ald_Xan_dh_C (pfam01315)Ald_Xan_dh_C2 (pfam02738)→
Bxe-1-5	NC_007951.1	-	3372361	3372319	**RNA**→ **RNA**→ BetA (COG2303)→
Bxe-1-1	NC_007951.1	-	3372312	3372270	(shown above)
env-2	AACY01029310.1	-	7922	7880	**RNA**→ **RNA**→ unknown→
env-3	AACY01029310.1	-	7872	7830	(shown above)
Bxe-1-6	NC_007952.1	-	232814	232772	**RNA**→ **RNA**→ **RNA**→ SseA (COG2897)→
Bxe-1-7	NC_007952.1	-	232765	232723	(shown above)
Bxe-1-8	NC_007952.1	-	232716	232674	(shown above)
Bps-7-8	NZ_AAHU01000005.1	+	92300	92342	**RNA**→ thiolase (cd00751)→ CaiD (COG1024)→
Bxe-1-9	NC_007951.1	+	1603440	1603482	**RNA**→ **RNA**→ ExbD (COG0848)→ hypo→
Bxe-1-10	NC_007951.1	+	1603538	1603580	(shown above)
Bps-10-4	NZ_AAHW02000016.1	+	20819	20862	**RNA**→ COG4319 (COG4319)→
Bps-3-7	NZ_AAHR02000001.1	+	42585	42628	**RNA**→ COG4319 (COG4319)→
Bsp-1-3	NC_007511.1	-	2435547	2435505	**RNA**→ TrpA (COG0159)→ AccD (COG0777)→
Bce-1-2	NC_008060.1	-	1793534	1793492	**RNA**→ COG2866 (COG2866)→
Bce-1-3	NC_008061.1	-	1639854	1639812	**RNA**→ TrpA (COG0159)→ AccD (COG0777)→
Bce-2-2	NZ_AAHL01000002.1	+	273545	273587	**RNA**→ COG2866 (COG2866)→
Bce-2-3	NZ_AAHL01000060.1	+	5914	5956	**RNA**→ TrpA (COG0159)→ AccD (COG0777)→
env-6	AACY01074305.1	-	3485	3443	**RNA**→ **RNA**→ unknown→ unknown→ unknown→
env-7	AACY01074305.1	-	3431	3389	(shown above)
Bps-5-5	NC_007434.1	-	182681	182639	**RNA**→ **RNA**→ **RNA**→ **RNA**→ ←hypo
Bps-5-6	NC_007434.1	-	182588	182546	(shown above)
^{7}Bps-5-20	NC_007434.1	-	182773	182731	(shown above)
Bps-5-22	NC_007434.1	-	182633	182593	(shown above)
Bth-1-5	NC_007651.1	-	3769469	3769427	**RNA**→ **RNA**→ **RNA**→ ATP-synt_DE_N (pfam02823)→
Bth-1-4	NC_007651.1	-	3769285	3769243	(shown above)
Bth-1-6	NC_007651.1	-	3769377	3769335	(shown above)
Bth-1-7	NC_007651.1	+	1478855	1478897	**RNA**→ COG4319 (COG4319)→
Bdo-1-2	NZ_AAKY01000115.1	+	22806	22848	**RNA**→ GlcD (COG0277)→ FAD_binding_4 (pfam01565)→ FAD-oxidase_C (pfam02913)→
env-8	AACY01054057.1	+	1681	1723	**RNA**→
env-10	AACY01078788.1	-	10646	10604	**RNA**→ **RNA**→ **RNA**→ unknown→ unknown→ unknown→ unknown→ unknown→
env-9	AACY01078788.1	-	10597	10555	(shown above)
env-11	AACY01078788.1	-	10548	10506	(shown above)
Bdo-1-5	NZ_AAKY01000054.1	+	26086	26128	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ COG2866 (COG2866)→
Bdo-1-3	NZ_AAKY01000054.1	+	26282	26324	(shown above)
Bdo-1-4	NZ_AAKY01000054.1	+	26331	26373	(shown above)
Bdo-1-7	NZ_AAKY01000054.1	+	26135	26177	(shown above)

Bdo-1-8	NZ_AAKY01000054.1	+	26184	26226	(shown above)
Bdo-1-9	NZ_AAKY01000054.1	+	26233	26275	(shown above)
Bam-1-4	NZ_AAJL01000013.1	-	101919	101877	**RNA**→ PurE (COG0041)→ PurK (COG0026)→ SUA5 (COG0009)SUA5 (pfam03481)→
env-14	AACY01174822.1	+	483	525	**RNA**→
env-15	AACY01198017.1	-	138	96	**RNA**→
Bdo-1-6	NZ_AAKY01000044.1	+	16111	16153	**RNA**→ NadA (pfam02445)→ QPRTase (cd01572)→
env-17	AACY01120349.1	+	192319	192361	**RNA**→ unknown→ unknown→
env-18	AACY01174822.1	+	532	574	**RNA**→
env-19	AACY01198017.1	-	89	47	**RNA**→
Bce-1-6	NC_008060.1	-	2091481	2091439	**RNA**→ **RNA**→ NadA (pfam02445)→ QPRTase (cd01572)→
Bce-1-8	NC_008060.1	-	2091383	2091341	(shown above)
Bce-2-6	NZ_AAHL01000062.1	+	16826	16868	**RNA**→ **RNA**→ NadA (pfam02445)→ QPRTase (cd01572)→
Bce-2-8	NZ_AAHL01000062.1	+	16924	16966	(shown above)
Bce-3-6	NZ_AAKX01000076.1	+	33252	33294	**RNA**→ NadA (pfam02445)→ QPRTase (cd01572)→
Bce-1-7	NC_008060.1	+	1995774	1995816	**RNA**→ **RNA**→ PutA (COG1012)→ adh_short (pfam00106)→
Bce-1-5	NC_008060.1	+	1995873	1995915	(shown above)
Bce-2-7	NZ_AAHL01000002.1	-	71292	71250	**RNA**→ **RNA**→ PutA (COG1012)→ adh_short (pfam00106)→
Bce-2-5	NZ_AAHL01000002.1	-	71193	71151	(shown above)
[6]env-20	AACY01057792.1	-	32713	32671	**RNA**→ **RNA**→ ←unknown
env-21	AACY01057792.1	-	32664	32622	(shown above)
env-22	AACY01057792.1	-	32615	32573	**RNA**→ unknown→
Bam-1-5	NZ_AAJL01000003.1	-	664343	664301	**RNA**→ **RNA**→ ACAD_fadE5 (cd01153)→
Bam-1-2	NZ_AAJL01000003.1	-	664294	664252	(shown above)
Bth-1-8	NC_007651.1	+	646278	646320	**RNA**→ thiolase (cd00751)→ CaiD (COG1024)→ hypo→
Bxe-1-12	NC_007953.1	-	228490	228448	**RNA**→ Sugar_tr (pfam00083)→ CaiB (COG1804)→ IclR (COG1414)→ hypo→
Bps-8-13	NC_006350.1	-	3368494	3368451	**RNA**→ **RNA**→ COG4319 (COG4319)→
Bps-8-2	NC_006350.1	-	3368395	3368353	(shown above)
Bps-5-9	NC_007434.1	-	3627099	3627056	**RNA**→ **RNA**→ COG4319 (COG4319)→
Bps-5-2	NC_007434.1	-	3627000	3626958	(shown above)
Bma-4-6	NC_006348.1	-	2411961	2411918	**RNA**→ **RNA**→ **RNA**→ COG4319 (COG4319)→
Bma-4-1	NC_006348.1	-	2411911	2411869	(shown above)
Bma-4-2	NC_006348.1	-	2411862	2411820	(shown above)
Bma-1-6	NZ_AAHM02000001.1	-	1111737	1111694	**RNA**→ **RNA**→ **RNA**→ COG4319 (COG4319)→
Bma-1-1	NZ_AAHM02000001.1	-	1111687	1111645	(shown above)
Bma-1-2	NZ_AAHM02000001.1	-	1111638	1111596	(shown above)
Bma-2-6	NZ_AAHN02000004.1	+	265112	265155	**RNA**→ **RNA**→ **RNA**→ COG4319 (COG4319)→
Bma-2-1	NZ_AAHN02000004.1	+	265162	265204	(shown above)
Bma-2-2	NZ_AAHN02000004.1	+	265211	265253	(shown above)
Bma-6-6	NZ_AAHO01000019.1	-	53214	53171	**RNA**→ **RNA**→ **RNA**→ COG4319 (COG4319)→
Bma-6-1	NZ_AAHO01000019.1	-	53164	53122	(shown above)
Bma-6-2	NZ_AAHO01000019.1	-	53115	53073	(shown above)
Bma-8-8	NZ_AAHP01000011.1	+	55145	55188	**RNA**→ **RNA**→ **RNA**→ COG4319 (COG4319)→
Bma-8-1	NZ_AAHP01000011.1	+	55195	55237	(shown above)

Bma-8-3	NZ_AAHP01000011.1	+	55244	55286	(shown above)
Bma-8-9	NZ_AAHP01000378.1	-	452	409	**RNA**→ **RNA**→ **RNA**→ COG4319: Ketosteroid isomerase homolog→
Bma-8-2	NZ_AAHP01000378.1	-	402	360	(shown above)
Bma-8-4	NZ_AAHP01000378.1	-	353	311	(shown above)
Bma-9-6	NZ_AAHQ02000001.1	-	2650854	2650811	**RNA**→ **RNA**→ **RNA**→ COG4319 (COG4319)→
Bma-9-1	NZ_AAHQ02000001.1	-	2650804	2650762	(shown above)
Bma-9-2	NZ_AAHQ02000001.1	-	2650755	2650713	(shown above)
Bps-4-9	NZ_AAHS02000002.1	-	272033	271990	**RNA**→ **RNA**→ COG4319 (COG4319)→
Bps-4-1	NZ_AAHS02000002.1	-	271934	271892	(shown above)
Bps-9-16	NZ_AAHV02000001.1	+	19484	19527	**RNA**→ **RNA**→ COG4319 (COG4319)→
Bps-9-1	NZ_AAHV02000001.1	+	19583	19625	(shown above)
Bma-5-6	NZ_AAIQ02000002.1	-	71804	71761	**RNA**→ **RNA**→ **RNA**→ COG4319 (COG4319)→
Bma-5-1	NZ_AAIQ02000002.1	-	71754	71712	(shown above)
Bma-5-2	NZ_AAIQ02000002.1	-	71705	71663	(shown above)
Bma-7-6	NZ_AAIR02000001.1	-	394278	394235	**RNA**→ **RNA**→ **RNA**→ COG4319 (COG4319)→
Bma-7-1	NZ_AAIR02000001.1	-	394228	394186	(shown above)
Bma-7-2	NZ_AAIR02000001.1	-	394179	394137	(shown above)
Bps-1-16	NZ_AAMA01000011.1	+	19427	19470	**RNA**→ **RNA**→ COG4319 (COG4319)→
Bps-1-4	NZ_AAMA01000011.1	+	19526	19568	(shown above)
Bps-2-13	NZ_AAMB01000001.1	-	373671	373628	**RNA**→ **RNA**→ COG4319 (COG4319)→
Bps-2-1	NZ_AAMB01000001.1	-	373572	373530	(shown above)
Bps-6-15	NZ_AAMM02000001.1	-	267283	267240	**RNA**→ **RNA**→ **RNA**→ COG4319 (COG4319)→
Bps-6-8	NZ_AAMM02000001.1	-	267233	267191	(shown above)
Bps-6-7	NZ_AAMM02000001.1	-	267135	267093	(shown above)
Bma-3-6	NZ_AANX02000012.1	+	80591	80634	**RNA**→ **RNA**→ **RNA**→ COG4319 (COG4319)→
Bma-3-1	NZ_AANX02000012.1	+	80641	80683	(shown above)
Bma-3-2	NZ_AANX02000012.1	+	80690	80732	(shown above)
Bsp-1-8	NC_007510.1	+	2852721	2852763	**RNA**→ **RNA**→ **RNA**→ UhpC (COG2271)→
Bsp-1-9	NC_007510.1	+	2852770	2852812	(shown above)
Bsp-1-10	NC_007510.1	+	2852819	2852861	(shown above)
Bdo-1-10	NZ_AAKY01000054.1	+	18382	18424	**RNA**→ **RNA**→ ACAD (cd00567)→ YdeM (cd03454)→ NodN (cd03450)→
Bdo-1-11	NZ_AAKY01000054.1	+	18431	18473	(shown above)
Bxe-1-13	NC_007951.1	+	2361599	2361641	**RNA**→ ←OmpR (COG0745)
env-28	AACY01077632.1	+	2716	2758	**RNA**→ unknown→
Bps-8-14	NC_006351.1	+	842808	842850	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ Acs (COG0365)→ Aldedh (pfam00171)→
Bps-8-9	NC_006351.1	+	842955	842997	(shown above)
Bps-8-10	NC_006351.1	+	843004	843046	(shown above)
Bps-8-11	NC_006351.1	+	843053	843095	(shown above)
Bps-8-16	NC_006351.1	+	842857	842899	(shown above)
env-29	AACY01119105.1	+	1026	1068	**RNA**→
Bxe-1-16	NC_007951.1	-	4652583	4652541	**RNA**→ **RNA**→ **RNA**→ SBP_bac_5 (pfam00496)→
Bxe-1-14	NC_007951.1	-	4652485	4652443	(shown above)
Bxe-1-15	NC_007951.1	-	4652387	4652345	(shown above)

env-31	AACY01170267.1	-	53	11	**RNA**→ **RNA**→ **RNA**→ **RNA**→ ←unknown
env-61	AACY01170267.1	-	151	109	(shown above)
env-65	AACY01170267.1	-	102	60	(shown above)
env-70	AACY01170267.1	-	200	158	(shown above)
env-32	AACY01174822.1	+	434	476	**RNA**→
env-33	AACY01198017.1	-	187	145	**RNA**→
env-34	AACY01074308.1	+	922	964	**RNA**→
Bth-1-9	NC_007650.1	+	1183543	1183585	**RNA**→ SoxD (pfam04267)→ GcvT (COG0404)HcaD (COG0446)→ COG4583 (COG4583)→ FolB (COG1539)→
Bdo-1-13	NZ_AAKY01000086.1	+	72	114	**RNA**→ CaiC (COG0318)→ hypo→
env-36	AACY01012789.1	+	514	556	**RNA**→ unknown→
env-37	AACY01202489.1	-	162	120	**RNA**→ unknown→
env-38	AACY01227980.1	+	534	576	**RNA**→ unknown→
env-39	AACY01005766.1	+	98374	98416	**RNA**→ **RNA**→ unknown→ unknown→
env-40	AACY01005766.1	+	98423	98465	(shown above)
Bsp-1-11	NC_007511.1	-	2023351	2023309	**RNA**→ COG3384 (COG3384)→ HpaF (COG3232)→ COG3971 (COG3971)→ HpcH (COG3836)→
env-41	AACY01120341.1	-	30946	30904	**RNA**→ **RNA**→ unknown→ unknown→ unknown→ unknown→
env-42	AACY01120341.1	-	30897	30855	(shown above)
Bps-8-15	NC_006351.1	-	1873032	1872990	**RNA**→ **RNA**→ **RNA**→ **RNA**→ SoxD (pfam04267)→ GcvT (COG0404)HcaD (COG0446)→ COG4583 (COG4583)→ FolB (COG1539)→
Bps-8-12	NC_006351.1	-	1872934	1872892	(shown above)
Bps-8-3	NC_006351.1	-	1872885	1872843	(shown above)
Bps-8-4	NC_006351.1	-	1872836	1872794	(shown above)
Bps-5-11	NC_007435.1	-	553692	553650	**RNA**→ **RNA**→ **RNA**→ **RNA**→ SoxD (pfam04267)→ GcvT (COG0404)HcaD (COG0446)→ COG4583 (COG4583)→ FolB (COG1539)→
Bps-5-8	NC_007435.1	-	553594	553552	(shown above)
Bps-5-3	NC_007435.1	-	553545	553503	(shown above)
Bps-5-12	NC_007435.1	-	553643	553601	(shown above)
Bps-4-11	NZ_AAHS02000004.1	-	27737	27695	**RNA**→ **RNA**→ **RNA**→ **RNA**→ SoxD (pfam04267)→ GcvT (COG0404)HcaD (COG0446)→ COG4583 (COG4583)→ FolB (COG1539)→
Bps-4-8	NZ_AAHS02000004.1	-	27639	27597	(shown above)
Bps-4-2	NZ_AAHS02000004.1	-	27590	27548	(shown above)
Bps-4-12	NZ_AAHS02000004.1	-	27688	27646	(shown above)
Bps-3-14	NZ_AAHR02000020.1	-	34017	33975	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ SoxD (pfam04267)→ GcvT (COG0404)HcaD (COG0446)→ COG4583 (COG4583)→ FolB (COG1539)→
Bps-3-13	NZ_AAHR02000020.1	-	33968	33926	(shown above)
Bps-3-4	NZ_AAHR02000020.1	-	33870	33828	(shown above)
Bps-3-5	NZ_AAHR02000020.1	-	33821	33779	(shown above)
Bps-3-22	NZ_AAHR02000020.1	-	33919	33877	(shown above)
Bps-9-19	NZ_AAHV02000010.1	-	29789	29747	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ SoxD (pfam04267)→ HcaD (COG0446)GcvT (COG0404)→ COG4583 (COG4583)→ FolB (COG1539)→
Bps-9-18	NZ_AAHV02000010.1	-	29740	29698	(shown above)
Bps-9-15	NZ_AAHV02000010.1	-	29691	29649	(shown above)
Bps-9-3	NZ_AAHV02000010.1	-	29642	29600	(shown above)
Bps-9-4	NZ_AAHV02000010.1	-	29593	29551	(shown above)
Bps-9-5	NZ_AAHV02000010.1	-	29544	29502	(shown above)

Bps-9-6	NZ_AAHV02000010.1	-	29495	29453	(shown above)
Bps-9-7	NZ_AAHV02000010.1	-	29446	29404	(shown above)
Bps-10-14	NZ_AAHW02000007.1	-	29495	29453	**RNA**→ **RNA**→ **RNA**→ **RNA**→ SoxD (pfam04267)→ HcaD (COG0446)GcvT (COG0404)→ COG4583 (COG4583)→ FolB (COG1539)→
Bps-10-13	NZ_AAHW02000007.1	-	29446	29404	(shown above)
Bps-10-10	NZ_AAHW02000007.1	-	29397	29355	(shown above)
Bps-10-1	NZ_AAHW02000007.1	-	29348	29306	(shown above)
Bps-1-19	NZ_AAMA01000007.1	+	147899	147941	**RNA**→ **RNA**→ **RNA**→ **RNA**→ SoxD (pfam04267)→ GcvT (COG0404)HcaD (COG0446)→ COG4583 (COG4583)→ FolB (COG1539)→
Bps-1-15	NZ_AAMA01000007.1	+	147997	148039	(shown above)
Bps-1-2	NZ_AAMA01000007.1	+	148046	148088	(shown above)
Bps-1-3	NZ_AAMA01000007.1	+	148095	148137	(shown above)
Bps-1-20	NZ_AAMA01000203.1	-	609	567	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ SoxD (pfam04267)→
Bps-1-18	NZ_AAMA01000203.1	-	560	518	(shown above)
Bps-1-14	NZ_AAMA01000203.1	-	511	469	(shown above)
Bps-1-7	NZ_AAMA01000203.1	-	462	420	(shown above)
Bps-1-5	NZ_AAMA01000203.1	-	413	371	(shown above)
Bps-1-6	NZ_AAMA01000203.1	-	364	322	(shown above)
Bps-2-16	NZ_AAMB01000012.1	-	34808	34766	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ SoxD (pfam04267)→ GcvT (COG0404)HcaD (COG0446)→ COG4583 (COG4583)→ FolB (COG1539)→
Bps-2-15	NZ_AAMB01000012.1	-	34759	34717	(shown above)
Bps-2-12	NZ_AAMB01000012.1	-	34710	34668	(shown above)
Bps-2-5	NZ_AAMB01000012.1	-	34661	34619	(shown above)
Bps-2-2	NZ_AAMB01000012.1	-	34612	34570	(shown above)
Bps-2-3	NZ_AAMB01000012.1	-	34563	34521	(shown above)
Bps-6-18	NZ_AAMM02000020.1	+	65168	65210	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ SoxD (pfam04267)→ GcvT (COG0404)HcaD (COG0446)→ COG4583 (COG4583)→ FolB (COG1539)→
Bps-6-17	NZ_AAMM02000020.1	+	65217	65259	(shown above)
Bps-6-14	NZ_AAMM02000020.1	+	65266	65308	(shown above)
Bps-6-2	NZ_AAMM02000020.1	+	65364	65406	(shown above)
Bps-6-3	NZ_AAMM02000020.1	+	65413	65455	(shown above)
Bps-6-4	NZ_AAMM02000020.1	+	65462	65504	(shown above)
Bps-9-20	NZ_AAHV02000019.1	+	108142	108184	**RNA**→ **RNA**→ **RNA**→ **RNA**→ Acs (COG0365)→
Bps-9-17	NZ_AAHV02000019.1	+	108191	108233	(shown above)
Bps-9-10	NZ_AAHV02000019.1	+	108240	108282	(shown above)
Bps-9-11	NZ_AAHV02000019.1	+	108289	108331	(shown above)
Bps-10-16	NZ_AAHW02000030.1	-	25522	25480	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ Acs (COG0365)→
Bps-10-11	NZ_AAHW02000030.1	-	25424	25382	(shown above)
Bps-10-7	NZ_AAHW02000030.1	-	25228	25186	(shown above)
Bps-10-12	NZ_AAHW02000030.1	-	25277	25235	(shown above)
Bps-10-15	NZ_AAHW02000030.1	-	25375	25333	(shown above)
Bps-10-22	NZ_AAHW02000030.1	-	25326	25284	(shown above)
Bps-1-22	NZ_AAMA01000020.1	-	25550	25508	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ Acs (COG0365)→
Bps-1-21	NZ_AAMA01000020.1	-	25452	25410	(shown above)

Bps-1-17	NZ_AAMA01000020.1	-	25403	25361	(shown above)
Bps-1-12	NZ_AAMA01000020.1	-	25354	25312	(shown above)
Bps-1-23	NZ_AAMA01000020.1	-	25501	25459	(shown above)
Bps-2-18	NZ_AAMB01000024.1	-	25325	25283	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ Acs (COG0365)→
Bps-2-17	NZ_AAMB01000024.1	-	25227	25185	(shown above)
Bps-2-14	NZ_AAMB01000024.1	-	25178	25136	(shown above)
Bps-2-10	NZ_AAMB01000024.1	-	25129	25087	(shown above)
Bps-2-19	NZ_AAMB01000024.1	-	25276	25234	(shown above)
Bps-5-13	NC_007435.1	+	2662751	2662793	**RNA**→ **RNA**→ **RNA**→ Aldedh (pfam00171)→ Acs (COG0365)→
Bps-5-10	NC_007435.1	+	2662800	2662842	(shown above)
Bps-5-7	NC_007435.1	+	2662849	2662891	(shown above)
Bps-4-13	NZ_AAHS02000001.1	+	338933	338975	**RNA**→ **RNA**→ **RNA**→ Acs (COG0365)→
Bps-4-10	NZ_AAHS02000001.1	+	338982	339024	(shown above)
Bps-4-7	NZ_AAHS02000001.1	+	339031	339073	(shown above)
Bps-7-19	NZ_AAHU01000026.1	+	73849	73891	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ Acs (COG0365)→
Bps-7-15	NZ_AAHU01000026.1	+	73898	73940	(shown above)
Bps-7-11	NZ_AAHU01000026.1	+	74143	74185	(shown above)
Bps-7-12	NZ_AAHU01000026.1	+	74192	74234	(shown above)
Bps-7-13	NZ_AAHU01000026.1	+	74241	74283	(shown above)
Bps-7-16	NZ_AAHU01000026.1	+	74094	74136	(shown above)
Bps-7-17	NZ_AAHU01000026.1	+	74045	74087	(shown above)
Bps-7-18	NZ_AAHU01000026.1	+	73947	73989	(shown above)
Bps-6-20	NZ_AAMM02000026.1	+	42501	42543	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ Acs (COG0365)→
Bps-6-19	NZ_AAMM02000026.1	+	42599	42641	(shown above)
Bps-6-16	NZ_AAMM02000026.1	+	42648	42690	(shown above)
Bps-6-11	NZ_AAMM02000026.1	+	42697	42739	(shown above)
Bps-6-21	NZ_AAMM02000026.1	+	42550	42592	(shown above)
Bps-3-16	NZ_AAHR02000013.1	+	105461	105503	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ Acs (COG0365)→
Bps-3-15	NZ_AAHR02000013.1	+	105559	105601	(shown above)
Bps-3-11	NZ_AAHR02000013.1	+	105608	105650	(shown above)
Bps-3-9	NZ_AAHR02000013.1	+	105657	105699	(shown above)
Bps-3-23	NZ_AAHR02000013.1	+	105510	105552	(shown above)
Bps-3-17	NZ_AAHR02000003.1	+	89234	89276	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ GcvP (COG0403)GcvP (COG1003)→
Bps-3-6	NZ_AAHR02000003.1	+	89528	89570	(shown above)
Bps-3-1	NZ_AAHR02000003.1	+	89577	89619	(shown above)
Bps-3-2	NZ_AAHR02000003.1	+	89626	89668	(shown above)
Bps-3-18	NZ_AAHR02000003.1	+	89332	89374	(shown above)
Bps-3-19	NZ_AAHR02000003.1	+	89381	89423	(shown above)
Bps-3-20	NZ_AAHR02000003.1	+	89283	89325	(shown above)
Bps-3-21	NZ_AAHR02000003.1	+	89430	89472	(shown above)
env-43	AACY01025908.1	-	5439	5397	**RNA**→ unknown→
Bce-1-9	NC_008062.1	+	981750	981792	**RNA**→ **RNA**→ ExbD (COG0848)→ hypo→
Bce-1-1	NC_008062.1	+	981848	981890	(shown above)

Bce-2-9	NZ_AAHL01000064.1	-	11010	10968	**RNA**→ **RNA**→ ExbD (COG0848)→ hypo→
Bce-2-1	NZ_AAHL01000064.1	-	10912	10870	(shown above)
env-44	AACY01057793.1	+	165527	165569	**RNA**→ unknown→
Bxe-1-17	NC_007953.1	+	1336341	1336383	**RNA**→ PL2_Passenger_AT (cd01344)Autotransporter (pfam03797)→
env-45	AACY01118859.1	-	327708	327666	**RNA**→ **RNA**→ unknown→ unknown→
env-23	AACY01118859.1	-	327659	327617	(shown above)
env-46	AACY01205166.1	+	357	399	**RNA**→ **RNA**→ unknown→
env-24	AACY01205166.1	+	406	448	(shown above)
env-47	AACY01218275.1	-	449	407	**RNA**→ **RNA**→ unknown→
env-25	AACY01218275.1	-	400	358	(shown above)
env-48	AACY01219723.1	+	288	330	**RNA**→ **RNA**→ unknown→
env-26	AACY01219723.1	+	337	379	(shown above)
env-49	AACY01247047.1	-	481	439	**RNA**→ **RNA**→ unknown→
env-27	AACY01247047.1	-	432	390	(shown above)
Bxe-1-18	NC_007953.1	+	273812	273854	**RNA**→ FixB (COG2025)→
Bam-1-7	NZ_AAJL01000025.1	+	5415	5457	**RNA**→ LldP (COG1620)→
Bsp-1-12	NC_007510.1	-	2952347	2952305	**RNA**→ NadA (pfam02445)→ QPRTase (cd01572)→
Bam-1-8	NZ_AAJL01000006.1	-	158281	158238	**RNA**→ **RNA**→ NadA (pfam02445)→ QPRTase (cd01572)→
Bam-1-6	NZ_AAJL01000006.1	-	158181	158139	(shown above)
Bdo-1-15	NZ_AAKY01000038.1	+	91564	91607	**RNA**→ thiolase (cd00751)→
Bce-1-11	NC_008062.1	+	467834	467877	**RNA**→ COG4757 (COG4757)→
Bce-2-11	NZ_AAHL01000021.1	+	100704	100747	**RNA**→ COG4757 (COG4757)→
Bce-3-8	NZ_AAKX01000082.1	+	51305	51348	**RNA**→ COG4757 (COG4757)→
Bce-3-9	NZ_AAKX01000001.1	+	109240	109282	**RNA**→ **RNA**→ **RNA**→ ExbD (COG0848)→ hypo→
Bce-3-1	NZ_AAKX01000001.1	+	109289	109331	(shown above)
Bce-3-10	NZ_AAKX01000001.1	+	109338	109380	(shown above)
Bps-7-26	NZ_AAHU01000003.1	+	19371	19414	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ COG4319 (COG4319)→
Bps-7-1	NZ_AAHU01000003.1	+	19421	19463	(shown above)
Bps-7-2	NZ_AAHU01000003.1	+	19470	19512	(shown above)
Bps-7-3	NZ_AAHU01000003.1	+	19519	19561	(shown above)
Bps-7-4	NZ_AAHU01000003.1	+	19568	19610	(shown above)
Bsp-1-15	NC_007510.1	-	2484166	2484124	**RNA**→ MhpC (COG0596)→
Bxe-1-19	NC_007952.1	-	3334149	3334107	**RNA**→ AcyP (COG1254)→
Bxe-1-20	NC_007951.1	-	4695181	4695139	**RNA**→ FlgC (COG1558)→ FlgD (COG1843)→ FlgE (COG1749)FlgG (COG4786)→ FlgF (COG4787)→ FlgG (COG4786)→
Bxe-1-21	NC_007951.1	+	3085495	3085537	**RNA**→ AceA (COG2224)→
env-53	AACY01015268.1	-	131492	131450	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ unknown→
env-51	AACY01015268.1	-	131394	131352	(shown above)
env-12	AACY01015268.1	-	131296	131254	(shown above)
env-52	AACY01015268.1	-	131345	131303	(shown above)
env-54	AACY01015268.1	-	131443	131401	(shown above)
Bsp-1-16	NC_007511.1	-	792372	792330	**RNA**→ **RNA**→ AcoR (COG3284)→ fer2 (cd00207)→ hypo→
Bsp-1-5	NC_007511.1	-	792323	792281	(shown above)
env-56	AACY01058207.1	-	414	372	**RNA**→

Bvi-1-2	NZ_AAEH02000034.1	-	65957	65914	**RNA**→ TrpA (COG0159)→ AccD (COG0777)→
Bth-1-12	NC_007650.1	-	2177506	2177464	**RNA**→ **RNA**→ **RNA**→ **RNA**→ Acs (COG0365)→ Aldedh (pfam00171)→ MmsB (COG2084)→ CaiD (COG1024)→ CaiD (COG1024)→
Bth-1-11	NC_007650.1	-	2177408	2177366	(shown above)
Bth-1-2	NC_007650.1	-	2177359	2177317	(shown above)
Bth-1-3	NC_007650.1	-	2177310	2177268	(shown above)
env-57	AACY01077631.1	-	84681	84639	**RNA**→ **RNA**→ unknown→ unknown→ unknown→ unknown→
env-55	AACY01077631.1	-	84632	84590	(shown above)
Bsp-1-17	NC_007511.1	+	2165800	2165842	**RNA**→ ←DadA (COG0665)
env-58	AACY01084398.1	+	134115	134157	**RNA**→ **RNA**→ unknown→
env-59	AACY01084398.1	+	134164	134206	(shown above)
env-62	AACY01174822.1	+	336	378	**RNA**→
env-63	AACY01198017.1	-	285	243	**RNA**→
env-66	AACY01174822.1	+	385	427	**RNA**→
env-67	AACY01198017.1	-	236	194	**RNA**→
Bce-1-12	NC_008060.1	+	1998352	1998394	**RNA**→ UhpC (COG2271)→
Bce-2-12	NZ_AAHL01000002.1	-	68714	68672	**RNA**→ UhpC (COG2271)→
Bxe-1-22	NC_007951.1	+	2205005	2205047	**RNA**→ **RNA**→ LysR (COG0583)→
Bxe-1-11	NC_007951.1	+	2205103	2205145	(shown above)
Bsp-1-18	NC_007511.1	+	154791	154833	**RNA**→ GltA (COG0372)→ AcnA (COG1048)→ DUF453 (pfam04303)→
Bxe-1-24	NC_007952.1	-	2349464	2349422	**RNA**→ **RNA**→ Sugar_tr (pfam00083)→
Bxe-1-23	NC_007952.1	-	2349364	2349322	(shown above)
Bce-3-11	NZ_AAKX01000077.1	-	19064	19022	**RNA**→ UhpC (COG2271)→
Bps-8-22	NC_006350.1	-	4034545	4034503	**RNA**→ **RNA**→ **RNA**→ ATP-synt_DE_N (pfam02823)→
Bps-8-8	NC_006350.1	-	4034452	4034410	(shown above)
Bps-8-24	NC_006350.1	-	4034497	4034457	(shown above)
Bma-4-9	NC_006348.1	+	3049507	3049549	**RNA**→ **RNA**→ **RNA**→ **RNA**→ ATP-synt_DE_N (pfam02823)→
Bma-4-5	NC_006348.1	+	3049786	3049828	(shown above)
Bma-4-11	NC_006348.1	+	3049741	3049781	(shown above)
[4]Bma-4-12	NC_006348.1	+	3049552	3049595	(shown above)
Bma-1-9	NZ_AAHM02000001.1	-	1628401	1628359	**RNA**→ **RNA**→ **RNA**→ **RNA**→ ATP-synt_DE_N (pfam02823)→
Bma-1-5	NZ_AAHM02000001.1	-	1628122	1628080	(shown above)
Bma-1-11	NZ_AAHM02000001.1	-	1628167	1628127	(shown above)
Bma-1-12	NZ_AAHM02000001.1	-	1628356	1628313	(shown above)
Bma-2-9	NZ_AAHN02000006.1	-	100548	100506	**RNA**→ **RNA**→ **RNA**→ **RNA**→ ATP-synt_DE_N (pfam02823)→
Bma-2-5	NZ_AAHN02000006.1	-	100269	100227	(shown above)
Bma-2-11	NZ_AAHN02000006.1	-	100314	100274	(shown above)
Bma-2-12	NZ_AAHN02000006.1	-	100503	100460	(shown above)
Bma-6-9	NZ_AAHO01000007.1	+	62144	62186	**RNA**→ **RNA**→ **RNA**→ **RNA**→ ATP-synt_DE_N (pfam02823)→
Bma-6-5	NZ_AAHO01000007.1	+	62423	62465	(shown above)
Bma-6-11	NZ_AAHO01000007.1	+	62378	62418	(shown above)
Bma-6-12	NZ_AAHO01000007.1	+	62189	62232	(shown above)
Bma-8-12	NZ_AAHP01000038.1	+	54905	54947	**RNA**→ **RNA**→ **RNA**→ **RNA**→ ATP-synt_DE_N (pfam02823)→
Bma-8-7	NZ_AAHP01000038.1	+	55184	55226	(shown above)

Bma-8-14	NZ_AAHP01000038.1	+	55139	55179	(shown above)
Bma-8-15	NZ_AAHP01000038.1	+	54950	54993	(shown above)
Bma-9-9	NZ_AAHQ02000003.1	-	154865	154823	**RNA**→ **RNA**→ **RNA**→ **RNA**→ ATP-synt_DE_N (pfam02823)→
Bma-9-5	NZ_AAHQ02000003.1	-	154586	154544	(shown above)
Bma-9-11	NZ_AAHQ02000003.1	-	154631	154591	(shown above)
Bma-9-12	NZ_AAHQ02000003.1	-	154820	154777	(shown above)
Bps-3-24	NZ_AAHR02000003.1	-	137592	137550	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ ATP-synt_DE_N (pfam02823)→
Bps-3-12	NZ_AAHR02000003.1	-	137500	137458	(shown above)
Bps-3-10	NZ_AAHR02000003.1	-	137407	137365	(shown above)
Bps-3-8	NZ_AAHR02000003.1	-	137314	137272	(shown above)
Bps-3-26	NZ_AAHR02000003.1	-	137359	137319	(shown above)
Bps-7-27	NZ_AAHU01000007.1	-	161243	161201	**RNA**→ **RNA**→ **RNA**→ **RNA**→ ATP-synt_DE_N (pfam02823)→
Bps-7-14	NZ_AAHU01000007.1	-	161151	161109	(shown above)
Bps-7-10	NZ_AAHU01000007.1	-	161058	161016	(shown above)
Bps-7-28	NZ_AAHU01000007.1	-	161103	161063	(shown above)
Bps-9-27	NZ_AAHV02000003.1	-	148230	148188	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ ATP-synt_DE_N (pfam02823)→
Bps-9-12	NZ_AAHV02000003.1	-	148137	148095	(shown above)
Bps-9-9	NZ_AAHV02000003.1	-	147951	147909	(shown above)
Bps-9-14	NZ_AAHV02000003.1	-	147858	147816	(shown above)
Bps-9-13	NZ_AAHV02000003.1	-	148044	148002	(shown above)
Bps-9-29	NZ_AAHV02000003.1	-	147996	147956	(shown above)
Bps-9-31	NZ_AAHV02000003.1	-	148089	148049	(shown above)
Bps-9-30	NZ_AAHV02000003.1	-	148182	148142	(shown above)
Bps-10-24	NZ_AAHW02000024.1	+	2266	2308	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ ATP-synt_DE_N (pfam02823)→
Bps-10-8	NZ_AAHW02000024.1	+	2359	2401	(shown above)
Bps-10-6	NZ_AAHW02000024.1	+	2545	2587	(shown above)
Bps-10-9	NZ_AAHW02000024.1	+	2452	2494	(shown above)
Bps-10-26	NZ_AAHW02000024.1	+	2500	2540	(shown above)
Bps-10-28	NZ_AAHW02000024.1	+	2407	2447	(shown above)
Bps-10-27	NZ_AAHW02000024.1	+	2314	2354	(shown above)
Bma-5-9	NZ_AAIQ02000004.1	+	55075	55117	**RNA**→ **RNA**→ **RNA**→ **RNA**→ ATP-synt_DE_N (pfam02823)→
Bma-5-5	NZ_AAIQ02000004.1	+	55354	55396	(shown above)
Bma-5-11	NZ_AAIQ02000004.1	+	55309	55349	(shown above)
Bma-5-12	NZ_AAIQ02000004.1	+	55120	55163	(shown above)
Bma-7-9	NZ_AAIR02000013.1	+	2460	2502	**RNA**→ **RNA**→ **RNA**→ **RNA**→ ATP-synt_DE_N (pfam02823)→
Bma-7-5	NZ_AAIR02000013.1	+	2739	2781	(shown above)
Bma-7-11	NZ_AAIR02000013.1	+	2694	2734	(shown above)
Bma-7-12	NZ_AAIR02000013.1	+	2505	2548	(shown above)
Bps-1-32	NZ_AAMA01000004.1	-	108193	108151	**RNA**→ **RNA**→ **RNA**→ **RNA**→ ATP-synt_DE_N (pfam02823)→
Bps-1-13	NZ_AAMA01000004.1	-	108101	108059	(shown above)
Bps-1-11	NZ_AAMA01000004.1	-	108008	107966	(shown above)
Bps-1-34	NZ_AAMA01000004.1	-	108053	108013	(shown above)
Bps-2-28	NZ_AAMB01000197.1	+	95373	95415	**RNA**→ **RNA**→ **RNA**→ **RNA**→ ATP-synt_DE_N (pfam02823)→

Bps-2-11	NZ_AAMB01000197.1	+	95465	95507	(shown above)
Bps-2-9	NZ_AAMB01000197.1	+	95558	95600	(shown above)
Bps-2-30	NZ_AAMB01000197.1	+	95513	95553	(shown above)
Bps-6-27	NZ_AAMM02000015.1	+	49612	49654	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ ATP-synt_DE_N (pfam02823)→
Bps-6-12	NZ_AAMM02000015.1	+	49705	49747	(shown above)
Bps-6-10	NZ_AAMM02000015.1	+	49891	49933	(shown above)
Bps-6-13	NZ_AAMM02000015.1	+	49798	49840	(shown above)
Bps-6-29	NZ_AAMM02000015.1	+	49846	49886	(shown above)
Bps-6-31	NZ_AAMM02000015.1	+	49753	49793	(shown above)
Bps-6-30	NZ_AAMM02000015.1	+	49660	49700	(shown above)
Bma-3-9	NZ_AANX02000011.1	+	99386	99428	**RNA**→ **RNA**→ **RNA**→ **RNA**→ ATP-synt_DE_N (pfam02823)→
Bma-3-5	NZ_AANX02000011.1	+	99665	99707	(shown above)
Bma-3-11	NZ_AANX02000011.1	+	99620	99660	(shown above)
Bma-3-12	NZ_AANX02000011.1	+	99431	99474	(shown above)
Bps-4-20	NZ_AAHS02000111.1	-	6443	6401	**RNA**→ **RNA**→ **RNA**→ **RNA**→ ATP-synt_DE_N (pfam02823)→
Bps-4-5	NZ_AAHS02000111.1	-	6351	6309	(shown above)
Bps-4-6	NZ_AAHS02000111.1	-	6258	6216	(shown above)
Bps-4-22	NZ_AAHS02000111.1	-	6303	6263	(shown above)
Bce-3-12	NZ_AAKX01000081.1	+	57358	57400	**RNA**→ **RNA**→ COG2866 (COG2866)→
Bce-3-2	NZ_AAKX01000081.1	+	57407	57449	(shown above)
Bsp-1-19	NC_007510.1	-	819745	819703	**RNA**→ **RNA**→ GCD (cd01151)→
Bsp-1-4	NC_007510.1	-	819647	819605	(shown above)
Bvi-1-3	NZ_AAEH02000010.1	+	42925	42967	**RNA**→ Sugar_tr (pfam00083)→ DHDPS-like (cd00408)→ MviM (COG0673)→ RbsB (COG1879)→ MglA (COG1129)→ AraH (COG1172)→ SMP-30 (pfam03758)→
Bce-1-13	NC_008062.1	-	81313	81271	**RNA**→ COG4770 (COG4770)AHS1 (pfam02682)AccB (COG0511)→
Bce-2-13	NZ_AAHL01000006.1	-	89378	89336	**RNA**→ COG4770 (COG4770)AHS1 (pfam02682)AccB (COG0511)→ Amidase (pfam01425)→ TauC (COG0600)→ ABC_NrtD_SsuB_transporters (cd03293)→
Bsp-1-20	NC_007510.1	+	3594427	3594469	**RNA**→ hypo→
Bxe-1-25	NC_007951.1	-	3173059	3173016	**RNA**→ Eno (COG0148)→
Bxe-1-26	NC_007951.1	+	3250414	3250456	**RNA**→ DUF495 (pfam04362)→ hypo→ Sugar_tr (pfam00083)→
Bsp-1-22	NC_007510.1	+	3000900	3000942	**RNA**→ **RNA**→ adh_short (pfam00106)→
Bsp-1-21	NC_007510.1	+	3000950	3000992	(shown above)
env-69	AACY01057792.1	-	35437	35395	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ unknown→
env-60	AACY01057792.1	-	35388	35346	(shown above)
env-30	AACY01057792.1	-	35290	35248	(shown above)
env-13	AACY01057792.1	-	35241	35199	(shown above)
env-16	AACY01057792.1	-	35192	35150	(shown above)
env-64	AACY01057792.1	-	35339	35297	(shown above)
env-71	AACY01174822.1	+	287	329	**RNA**→
env-72	AACY01198017.1	-	334	292	**RNA**→
Bsp-1-23	NC_007510.1	-	2060448	2060406	**RNA**→ PurR (COG1609)→ COG4377 (COG4377)→ hypo→
Bdo-1-16	NZ_AAKY01000189.1	-	11327	11285	**RNA**→ GltA (COG0372)→ AcnA (COG1048)→ DUF453 (pfam04303)→ CaiB (COG1804)→
Bxe-1-27	NC_007951.1	-	4181410	4181367	**RNA**→ DHBP_synthase (pfam00926)→

Bth-1-13	NC_007650.1	+	577773	577815	**RNA**→ Ring_hydroxylating_alpha (cd00680)Rieske_RO_Alpha_N (cd03469)→ ring_hydroxylating_dioxygenases_beta (cd00667)→ Rieske_RO_ferredoxin (cd03528)→ Pyr_redox (pfam00070)→
Bsp-1-24	NC_007510.1	-	843927	843885	**RNA**→ ←XdhB (COG4631)
[3]Bth-1-14	NC_007651.1	+	894209	894250	**RNA**→ NadA (pfam02445)→ QPRTase (cd01572)→
Bth-1-15	NC_007651.1	+	3261390	3261432	**RNA**→ ThiC (pfam01964)→
Bxe-1-28	NC_007951.1	+	509233	509275	**RNA**→ Putative NAD(P) transhydrogenase subunit alpha→ PNTB (pfam02233)→
[9]Bsp-1-25	NC_007511.1	+	809158	809200	**RNA**→ Acs (COG0365)→
Bxe-1-29	NC_007951.1	+	3709602	3709644	**RNA**→ ←RpmF (COG0333)
Bce-3-13	NZ_AAKX01000077.1	-	21741	21699	**RNA**→ **RNA**→ **RNA**→ **RNA**→ PutA (COG1012)→ adh_short (pfam00106)→
Bce-3-7	NZ_AAKX01000077.1	-	21643	21601	(shown above)
Bce-3-4	NZ_AAKX01000077.1	-	21593	21551	(shown above)
Bce-3-5	NZ_AAKX01000077.1	-	21543	21501	(shown above)
Bxe-1-30	NC_007951.1	-	4807806	4807764	**RNA**→ AtpG (COG0224)→
env-73	AACY01057790.1	+	17114	17156	**RNA**→ **RNA**→ **RNA**→ unknown→ unknown→
env-5	AACY01057790.1	+	17212	17254	(shown above)
env-4	AACY01057790.1	+	17261	17303	(shown above)
Bvi-1-4	NZ_AAEH02000031.1	-	44042	44000	**RNA**→ **RNA**→ NadA (pfam02445)→ QPRTase (cd01572)→
Bvi-1-1	NZ_AAEH02000031.1	-	43944	43902	(shown above)
env-74	AACY01007479.1	+	104478	104520	**RNA**→ unknown→
Bam-1-12	NZ_AAJL01000007.1	-	93026	92984	**RNA**→ **RNA**→ ATP-synt_DE_N (pfam02823)→
Bam-1-9	NZ_AAJL01000007.1	-	92934	92892	(shown above)
Bsp-1-26	NC_007510.1	-	2572433	2572390	**RNA**→ **RNA**→ COG4757 (COG4757)→
Bsp-1-30	NC_007510.1	-	2572339	2572300	(shown above)
Bsp-1-27	NC_007510.1	+	2849959	2850001	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ PutA (COG1012)→ adh_short (pfam00106)→
Bsp-1-13	NC_007510.1	+	2850057	2850099	(shown above)
Bsp-1-6	NC_007510.1	+	2850204	2850246	(shown above)
Bsp-1-7	NC_007510.1	+	2850253	2850295	(shown above)
Bsp-1-14	NC_007510.1	+	2850155	2850197	(shown above)
Bsp-1-28	NC_007511.1	-	3528699	3528657	**RNA**→ Sugar_tr (pfam00083)→
Bce-1-14	NC_008062.1	+	549993	550035	**RNA**→ **RNA**→ MhpC (COG0596)→
Bce-1-10	NC_008062.1	+	550091	550133	(shown above)
Bce-2-14	NZ_AAHL01000106.1	-	1038	996	**RNA**→ **RNA**→ hypo→ hypo→
Bce-2-10	NZ_AAHL01000106.1	-	940	898	(shown above)
Bce-3-14	NZ_AAKX01000083.1	+	46203	46245	**RNA**→ MhpC (COG0596)→
Bxe-1-31	NC_007951.1	-	239435	239393	**RNA**→ **RNA**→ **RNA**→ SCAD/SBCAD (cd01158)→ COG3012 (COG3012)→
Bxe-1-2	NC_007951.1	-	239386	239344	(shown above)
Bxe-1-3	NC_007951.1	-	239337	239295	(shown above)
Bdo-1-19	NZ_AAKY01000054.1	+	16871	16913	**RNA**→ **RNA**→ ACAD_fadE6_17_26 (cd01152)→
Bdo-1-14	NZ_AAKY01000054.1	+	16970	17013	(shown above)
Bce-1-15	NC_008060.1	-	3284104	3284064	**RNA**→ BetA (COG2303)→
Bce-2-15	NZ_AAHL01000003.1	-	212830	212790	**RNA**→ BetA (COG2303)→
Bps-8-23	NC_006350.1	+	738466	738508	**RNA**→ **RNA**→ thiolase (cd00751)→ CaiD (COG1024)→
Bps-8-5	NC_006350.1	+	738570	738612	(shown above)
Bps-5-21	NC_007434.1	+	896083	896125	**RNA**→ **RNA**→ thiolase (cd00751)→

Bps-5-4	NC_007434.1	+	896187	896229	(shown above)
Bps-3-25	NZ_AAHR02000045.1	-	54726	54684	**RNA**→ **RNA**→ thiolase (cd00751)→ CaiD (COG1024)→
Bps-3-3	NZ_AAHR02000045.1	-	54674	54632	(shown above)
Bps-4-21	NZ_AAHS02000019.1	+	56711	56753	**RNA**→ **RNA**→ thiolase (cd00751)→ CaiD (COG1024)→
Bps-4-4	NZ_AAHS02000019.1	+	56815	56857	(shown above)
Bps-9-28	NZ_AAHV02000004.1	-	23429	23387	**RNA**→ **RNA**→ thiolase (cd00751)→ CaiD (COG1024)→
Bps-9-8	NZ_AAHV02000004.1	-	23325	23283	(shown above)
Bps-10-25	NZ_AAHW02000035.1	-	55229	55187	**RNA**→ **RNA**→ **RNA**→ thiolase (cd00751)→ CaiD (COG1024)→
Bps-10-23	NZ_AAHW02000035.1	-	55177	55135	(shown above)
Bps-10-3	NZ_AAHW02000035.1	-	55125	55083	(shown above)
Bps-1-33	NZ_AAMA01000003.1	+	192246	192288	**RNA**→ **RNA**→ thiolase (cd00751)→ CaiD (COG1024)→
Bps-1-8	NZ_AAMA01000003.1	+	192350	192392	(shown above)
Bps-2-29	NZ_AAMB01000198.1	+	124014	124056	**RNA**→ **RNA**→ thiolase (cd00751)→ CaiD (COG1024)→
Bps-2-6	NZ_AAMB01000198.1	+	124118	124160	(shown above)
Bps-6-28	NZ_AAMM02000059.1	+	14165	14207	**RNA**→ **RNA**→ **RNA**→ thiolase (cd00751)→ CaiD (COG1024)→
Bps-6-5	NZ_AAMM02000059.1	+	14321	14363	(shown above)
Bps-6-6	NZ_AAMM02000059.1	+	14373	14415	(shown above)
Bma-4-10	NC_006348.1	+	213794	213836	**RNA**→ **RNA**→ thiolase (cd00751)→ CaiD (COG1024)→
Bma-4-3	NC_006348.1	+	213898	213940	(shown above)
Bma-1-10	NZ_AAHM02000001.1	+	2365557	2365599	**RNA**→ **RNA**→ thiolase (cd00751)→ CaiD (COG1024)→
Bma-1-3	NZ_AAHM02000001.1	+	2365661	2365703	(shown above)
Bma-2-10	NZ_AAHN02000011.1	-	54795	54753	**RNA**→ **RNA**→ thiolase (cd00751)→ CaiD (COG1024)→
Bma-2-3	NZ_AAHN02000011.1	-	54691	54649	(shown above)
Bma-6-10	NZ_AAHO01000031.1	+	2570	2612	**RNA**→ **RNA**→ thiolase (cd00751)→ CaiD (COG1024)→
Bma-6-3	NZ_AAHO01000031.1	+	2674	2716	(shown above)
Bma-8-13	NZ_AAHP01000016.1	+	33498	33540	**RNA**→ **RNA**→ thiolase (cd00751)→ CaiD (COG1024)→
Bma-8-5	NZ_AAHP01000016.1	+	33602	33644	(shown above)
Bma-9-10	NZ_AAHQ02000001.1	+	433035	433077	**RNA**→ **RNA**→ thiolase (cd00751)→ CaiD (COG1024)→
Bma-9-3	NZ_AAHQ02000001.1	+	433139	433181	(shown above)
Bma-5-10	NZ_AAIQ02000007.1	+	87028	87070	**RNA**→ **RNA**→ thiolase (cd00751)→ CaiD (COG1024)→
Bma-5-3	NZ_AAIQ02000007.1	+	87132	87174	(shown above)
Bma-7-10	NZ_AAIR02000005.1	+	86605	86647	**RNA**→ **RNA**→ thiolase (cd00751)→ CaiD (COG1024)→
Bma-7-3	NZ_AAIR02000005.1	+	86709	86751	(shown above)
Bma-3-10	NZ_AANX02000057.1	-	23403	23361	**RNA**→ **RNA**→ thiolase (cd00751)→ CaiD (COG1024)→
Bma-3-3	NZ_AANX02000057.1	-	23299	23257	(shown above)
Bxe-1-32	NC_007951.1	-	3373147	3373105	**RNA**→ hypo→
[1]Bxe-1-33	NC_007952.1	+	1362200	1362242	**RNA**→ REC (cd00156)BaeS (COG0642)→
Bce-1-16	NC_008060.1	-	365461	365419	**RNA**→ **RNA**→ GCD (cd01151)→
Bce-1-4	NC_008060.1	-	365359	365317	(shown above)
Bce-2-16	NZ_AAHL01000005.1	+	93999	94041	**RNA**→ **RNA**→ GCD (cd01151)→
Bce-2-4	NZ_AAHL01000005.1	+	94101	94143	(shown above)
Bce-3-15	NZ_AAKX01000025.1	+	4237	4279	**RNA**→ **RNA**→ GCD (cd01151)→
Bce-3-3	NZ_AAKX01000025.1	+	4335	4377	(shown above)

^{8}Bth-1-16	NC_007650.1	-	2493736	2493694	**RNA**→ 23S ribosomal RNA→
Bth-1-17	NC_007651.1	+	1239634	1239676	**RNA**→ 23S ribosomal RNA→
Bth-1-18	NC_007651.1	-	3528168	3528126	**RNA**→ 23S ribosomal RNA→
Bth-1-19	NC_007651.1	-	3331232	3331190	**RNA**→ 23S ribosomal RNA→
Bsp-1-29	NC_007510.1	-	2063173	2063131	**RNA**→ **RNA**→ UhpC (COG2271)→ LdhA (COG1052)→
Bsp-1-2	NC_007510.1	-	2063075	2063033	(shown above)
^{5}Bxe-1-34	NC_007951.1	+	2727968	2728011	**RNA**→ ←Lrp (COG1522)
Bth-1-20	NC_007650.1	-	768907	768865	**RNA**→ MmgE_PrpD (pfam03972)→ AcnA (COG1048)→
Bam-1-13	NZ_AAJL01000009.1	-	288481	288439	**RNA**→ **RNA**→ AcoR (COG3284)→ fer2 (cd00207)→ DUF704 (pfam05146)→
Bam-1-10	NZ_AAJL01000009.1	-	288383	288341	(shown above)
Bce-1-17	NC_008060.1	-	2010050	2010008	**RNA**→ UbiE (COG2226)→
Bce-2-17	NZ_AAHL01000002.1	+	57016	57058	**RNA**→ UbiE (COG2226)→
Bce-3-16	NZ_AAKX01000077.1	+	7370	7412	**RNA**→ UbiE (COG2226)→
Bps-7-29	NZ_AAHU01000004.1	-	32067	32026	**RNA**→ **RNA**→ SoxD (pfam04267)→ HcaD (COG0446)GcvT (COG0404)→ COG4583 (COG4583)→ FolB (COG1539)→
Bps-7-7	NZ_AAHU01000004.1	-	31970	31928	(shown above)
Bth-1-21	NC_007651.1	-	3713376	3713333	**RNA**→ **RNA**→ **RNA**→ GcvP (COG0403)GcvP (COG1003)→
Bth-1-10	NC_007651.1	-	3713277	3713235	(shown above)
Bth-1-1	NC_007651.1	-	3713179	3713137	(shown above)
Bma-4-13	NC_006348.1	-	3099338	3099295	**RNA**→ **RNA**→ **RNA**→ **RNA**→ GcvP (COG0403)GcvP (COG1003)→
Bma-4-8	NC_006348.1	-	3099288	3099246	(shown above)
Bma-4-7	NC_006348.1	-	3099239	3099197	(shown above)
Bma-4-4	NC_006348.1	-	3099190	3099148	(shown above)
Bma-1-13	NZ_AAHM02000001.1	+	1576998	1577041	**RNA**→ **RNA**→ **RNA**→ **RNA**→ GcvP (COG0403)GcvP (COG1003)→
Bma-1-8	NZ_AAHM02000001.1	+	1577048	1577090	(shown above)
Bma-1-7	NZ_AAHM02000001.1	+	1577097	1577139	(shown above)
Bma-1-4	NZ_AAHM02000001.1	+	1577146	1577188	(shown above)
Bma-2-13	NZ_AAHN02000006.1	+	50744	50787	**RNA**→ **RNA**→ **RNA**→ **RNA**→ GcvP (COG0403)GcvP (COG1003)→
Bma-2-8	NZ_AAHN02000006.1	+	50794	50836	(shown above)
Bma-2-7	NZ_AAHN02000006.1	+	50843	50885	(shown above)
Bma-2-4	NZ_AAHN02000006.1	+	50892	50934	(shown above)
Bma-6-13	NZ_AAHO01000007.1	-	111975	111932	**RNA**→ **RNA**→ **RNA**→ **RNA**→ GcvP (COG0403)GcvP (COG1003)→
Bma-6-8	NZ_AAHO01000007.1	-	111925	111883	(shown above)
Bma-6-7	NZ_AAHO01000007.1	-	111876	111834	(shown above)
Bma-6-4	NZ_AAHO01000007.1	-	111827	111785	(shown above)
Bma-8-16	NZ_AAHP01000033.1	+	50690	50733	**RNA**→ **RNA**→ **RNA**→ **RNA**→ GcvP (COG0403)GcvP (COG1003)→
Bma-8-11	NZ_AAHP01000033.1	+	50740	50782	(shown above)
Bma-8-10	NZ_AAHP01000033.1	+	50789	50831	(shown above)
Bma-8-6	NZ_AAHP01000033.1	+	50838	50880	(shown above)
Bma-9-13	NZ_AAHQ02000003.1	+	105093	105136	**RNA**→ **RNA**→ **RNA**→ **RNA**→ GcvP (COG0403)GcvP (COG1003)→
Bma-9-8	NZ_AAHQ02000003.1	+	105143	105185	(shown above)
Bma-9-7	NZ_AAHQ02000003.1	+	105192	105234	(shown above)
Bma-9-4	NZ_AAHQ02000003.1	+	105241	105283	(shown above)
Bma-5-13	NZ_AAIQ02000004.1	-	104906	104863	**RNA**→ **RNA**→ **RNA**→ **RNA**→ GcvP (COG0403)GcvP (COG1003)→

Bma-5-8	NZ_AAIQ02000004.1	-	104856	104814	(shown above)
Bma-5-7	NZ_AAIQ02000004.1	-	104807	104765	(shown above)
Bma-5-4	NZ_AAIQ02000004.1	-	104758	104716	(shown above)
Bma-7-13	NZ_AAIR02000013.1	-	52291	52248	**RNA**→ **RNA**→ **RNA**→ **RNA**→ GcvP (COG0403)GcvP (COG1003)→
Bma-7-8	NZ_AAIR02000013.1	-	52241	52199	(shown above)
Bma-7-7	NZ_AAIR02000013.1	-	52192	52150	(shown above)
Bma-7-4	NZ_AAIR02000013.1	-	52143	52101	(shown above)
Bma-3-13	NZ_AANX02000007.1	-	17302	17259	**RNA**→ **RNA**→ **RNA**→ **RNA**→ GcvP (COG0403)GcvP (COG1003)→
Bma-3-8	NZ_AANX02000007.1	-	17252	17210	(shown above)
Bma-3-7	NZ_AANX02000007.1	-	17203	17161	(shown above)
Bma-3-4	NZ_AANX02000007.1	-	17154	17112	(shown above)
Bps-8-25	NC_006350.1	+	3986071	3986114	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ GcvP (COG0403)GcvP (COG1003)→
Bps-8-18	NC_006350.1	+	3986121	3986163	(shown above)
Bps-8-17	NC_006350.1	+	3986170	3986212	(shown above)
Bps-8-6	NC_006350.1	+	3986366	3986408	(shown above)
Bps-8-1	NC_006350.1	+	3986513	3986555	(shown above)
Bps-8-7	NC_006350.1	+	3986464	3986506	(shown above)
Bps-8-21	NC_006350.1	+	3986415	3986457	(shown above)
Bps-8-19	NC_006350.1	+	3986268	3986310	(shown above)
Bps-8-20	NC_006350.1	+	3986317	3986359	(shown above)
Bps-5-23	NC_007434.1	+	134501	134544	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ GcvP (COG0403)GcvP (COG1003)→
Bps-5-15	NC_007434.1	+	134551	134593	(shown above)
Bps-5-14	NC_007434.1	+	134649	134691	(shown above)
Bps-5-1	NC_007434.1	+	134894	134936	(shown above)
Bps-5-17	NC_007434.1	+	134698	134740	(shown above)
Bps-5-18	NC_007434.1	+	134747	134789	(shown above)
Bps-5-19	NC_007434.1	+	134796	134838	(shown above)
Bps-5-16	NC_007434.1	+	134600	134642	(shown above)
Bps-4-23	NZ_AAHS02000026.1	+	58188	58231	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ GcvP (COG0403)GcvP (COG1003)→
Bps-4-15	NZ_AAHS02000026.1	+	58238	58280	(shown above)
Bps-4-14	NZ_AAHS02000026.1	+	58336	58378	(shown above)
Bps-4-3	NZ_AAHS02000026.1	+	58581	58623	(shown above)
Bps-4-17	NZ_AAHS02000026.1	+	58385	58427	(shown above)
Bps-4-18	NZ_AAHS02000026.1	+	58434	58476	(shown above)
Bps-4-19	NZ_AAHS02000026.1	+	58483	58525	(shown above)
Bps-4-16	NZ_AAHS02000026.1	+	58287	58329	(shown above)
Bps-7-30	NZ_AAHU01000007.1	+	112875	112918	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ GcvP (COG0403)GcvP (COG1003)→
Bps-7-21	NZ_AAHU01000007.1	+	112925	112967	(shown above)
Bps-7-20	NZ_AAHU01000007.1	+	113023	113065	(shown above)
Bps-7-9	NZ_AAHU01000007.1	+	113219	113261	(shown above)
Bps-7-5	NZ_AAHU01000007.1	+	113268	113310	(shown above)
Bps-7-6	NZ_AAHU01000007.1	+	113317	113359	(shown above)

Bps-7-23	NZ_AAHU01000007.1	+	113072	113114	(shown above)
Bps-7-24	NZ_AAHU01000007.1	+	113121	113163	(shown above)
Bps-7-25	NZ_AAHU01000007.1	+	113170	113212	(shown above)
Bps-7-22	NZ_AAHU01000007.1	+	112974	113016	(shown above)
Bps-9-32	NZ_AAHV02000003.1	+	99586	99629	**RNA→ RNA→ RNA→ RNA→ RNA→ RNA→ RNA→ RNA→** GcvP (COG0403)GcvP (COG1003)→
Bps-9-22	NZ_AAHV02000003.1	+	99636	99678	(shown above)
Bps-9-21	NZ_AAHV02000003.1	+	99734	99776	(shown above)
Bps-9-2	NZ_AAHV02000003.1	+	99979	100021	(shown above)
Bps-9-24	NZ_AAHV02000003.1	+	99783	99825	(shown above)
Bps-9-25	NZ_AAHV02000003.1	+	99832	99874	(shown above)
Bps-9-26	NZ_AAHV02000003.1	+	99881	99923	(shown above)
Bps-9-23	NZ_AAHV02000003.1	+	99685	99727	(shown above)
Bps-10-29	NZ_AAHW02000024.1	-	51066	51023	**RNA→ RNA→ RNA→ RNA→ RNA→ RNA→ RNA→ RNA→** GcvP (COG0403)GcvP (COG1003)→
Bps-10-18	NZ_AAHW02000024.1	-	51016	50974	(shown above)
Bps-10-17	NZ_AAHW02000024.1	-	50967	50925	(shown above)
Bps-10-5	NZ_AAHW02000024.1	-	50722	50680	(shown above)
Bps-10-2	NZ_AAHW02000024.1	-	50673	50631	(shown above)
Bps-10-19	NZ_AAHW02000024.1	-	50918	50876	(shown above)
Bps-10-20	NZ_AAHW02000024.1	-	50820	50778	(shown above)
Bps-10-21	NZ_AAHW02000024.1	-	50771	50729	(shown above)
Bps-1-35	NZ_AAMA01000004.1	+	59606	59649	**RNA→ RNA→ RNA→ RNA→ RNA→ RNA→ RNA→ RNA→ RNA→ RNA→ RNA→ RNA→** GcvP (COG0403)GcvP (COG1003)→
Bps-1-26	NZ_AAMA01000004.1	+	59656	59698	(shown above)
Bps-1-25	NZ_AAMA01000004.1	+	59705	59747	(shown above)
Bps-1-24	NZ_AAMA01000004.1	+	59754	59796	(shown above)
Bps-1-9	NZ_AAMA01000004.1	+	59999	60041	(shown above)
Bps-1-1	NZ_AAMA01000004.1	+	60195	60237	(shown above)
Bps-1-10	NZ_AAMA01000004.1	+	60146	60188	(shown above)
Bps-1-30	NZ_AAMA01000004.1	+	60048	60090	(shown above)
Bps-1-31	NZ_AAMA01000004.1	+	60097	60139	(shown above)
Bps-1-27	NZ_AAMA01000004.1	+	59803	59845	(shown above)
Bps-1-28	NZ_AAMA01000004.1	+	59901	59943	(shown above)
Bps-1-29	NZ_AAMA01000004.1	+	59950	59992	(shown above)
Bps-2-31	NZ_AAMB01000197.1	-	143960	143917	**RNA→ RNA→ RNA→ RNA→ RNA→ RNA→ RNA→ RNA→ RNA→ RNA→ RNA→ RNA→** GcvP (COG0403)GcvP (COG1003)→
Bps-2-22	NZ_AAMB01000197.1	-	143910	143868	(shown above)
Bps-2-21	NZ_AAMB01000197.1	-	143861	143819	(shown above)
Bps-2-20	NZ_AAMB01000197.1	-	143812	143770	(shown above)
Bps-2-7	NZ_AAMB01000197.1	-	143567	143525	(shown above)
Bps-2-4	NZ_AAMB01000197.1	-	143371	143329	(shown above)
Bps-2-8	NZ_AAMB01000197.1	-	143420	143378	(shown above)
Bps-2-26	NZ_AAMB01000197.1	-	143518	143476	(shown above)
Bps-2-27	NZ_AAMB01000197.1	-	143469	143427	(shown above)
Bps-2-23	NZ_AAMB01000197.1	-	143763	143721	(shown above)

Bps-2-24	NZ_AAMB01000197.1	-	143665	143623	(shown above)
Bps-2-25	NZ_AAMB01000197.1	-	143616	143574	(shown above)
Bps-6-32	NZ_AAMM02000015.1	-	99538	99495	**RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ **RNA**→ GcvP (COG0403)GcvP (COG1003)→
Bps-6-23	NZ_AAMM02000015.1	-	99488	99446	(shown above)
Bps-6-22	NZ_AAMM02000015.1	-	99439	99397	(shown above)
Bps-6-9	NZ_AAMM02000015.1	-	99194	99152	(shown above)
Bps-6-1	NZ_AAMM02000015.1	-	99145	99103	(shown above)
Bps-6-24	NZ_AAMM02000015.1	-	99390	99348	(shown above)
Bps-6-25	NZ_AAMM02000015.1	-	99292	99250	(shown above)
Bps-6-26	NZ_AAMM02000015.1	-	99243	99201	(shown above)
Bsp-1-31	NC_007510.1	-	2043614	2043572	**RNA**→ ExbD (COG0848)→ hypo→
Bam-1-14	NZ_AAJL01000005.1	+	388286	388329	**RNA**→ GCD (cd01151)→
Bsp-1-32	NC_007510.1	-	1666823	1666781	**RNA**→ CaiC (COG0318)→ hypo→
Bam-1-15	NZ_AAJL01000010.1	-	293775	293733	**RNA**→ **RNA**→ NarK (COG2223)→ DHDPS-like (cd00408)→ MviM (COG0673)→ RbsB (COG1879)→ MglA (COG1129)→
Bam-1-11	NZ_AAJL01000010.1	-	293678	293636	(shown above)
Bdo-1-20	NZ_AAKY01000015.1	-	11421	11379	**RNA**→ ATP-synt_DE_N (pfam02823)→
Bdo-1-21	NZ_AAKY01000068.1	+	11702	11742	**RNA**→ UhpC (COG2271)→ LdhA (COG1052)→
Bsp-1-33	NC_007510.1	-	171796	171754	**RNA**→ GcvP (COG0403)GcvP (COG1003)→
Bdo-1-22	NZ_AAKY01000192.1	-	7943	7902	**RNA**→ **RNA**→ **RNA**→ UhpC (COG2271)→ DHDPS-like (cd00408)→ MviM (COG0673)→ RbsB (COG1879)→ MglA (COG1129)→ MglA (COG1129)→
Bdo-1-17	NZ_AAKY01000192.1	-	7846	7804	(shown above)
Bdo-1-18	NZ_AAKY01000192.1	-	7748	7706	(shown above)
env-75	AACY01118859.1	-	347368	347326	**RNA**→ **RNA**→ **RNA**→ **RNA**→ unknown→ unknown→
env-50	AACY01118859.1	-	347319	347277	(shown above)
env-35	AACY01118859.1	-	347222	347180	(shown above)
env-68	AACY01118859.1	-	347270	347229	(shown above)
[2]Bxe-1-35	NC_007951.1	-	1872412	1872362	**RNA**→ CaiC (COG0318)→
Bxe-1-36	NC_007951.1	+	1587813	1587854	**RNA**→ AidB (cd01154)→ CaiB (COG1804)→ Sugar_tr (pfam00083)→

cd00156 Signal receiver domain; originally thought to be unique to bacteria (CheY, OmpR, NtrC, and PhoB), now recently identified in eukaroytes ETR1 Arabidopsis thaliana; this domain receives the signal from the sensor partner in a two-component systems; contains a phosphoacceptor site that is phosphorylated by histidine kinase homologs; usually found N-terminal to a DNA binding effector domain; forms homodimers

cd00207 2Fe-2S iron-sulfur cluster binding domain.

cd00408 Dihydrodipicolinate synthase family.

cd00567 Acyl-CoA dehydrogenase (ACAD).

cd00667 Ring hydroxylating dioxygenase beta subunit. [2Fe-2S]

cd00680 Ring hydroxylating alpha subunit (catalytic domain).

cd00751 Thiolase are ubiquitous enzymes that catalyze the reversible thiolytic cleavage of 3-ketoacyl-CoA into acyl-CoA and acetyl-CoA, a 2-step reaction involving a covalent intermediate formed with a catalytic cysteine.

cd01151 Glutaryl-CoA dehydrogenase (GCD).

cd01152 Putative acyl-CoA dehydrogenase (ACAD).

cd01153 Putative acyl-CoA dehydrogenase (ACAD).

cd01154 AidB.

cd01158 Short chain acyl-CoA dehydrogenase (SCAD).

cd01344 Pertactin-like passenger domains (virulence factors), C-terminal, subgroup 2, of autotransporter proteins of the type V secretion system of Gram-negative bacteria.

cd01572 Quinolinate phosphoribosyl transferase (QAPRTase or QPRTase), also called nicotinate-nucleotide pyrophosphorylase, is involved in the de novo synthesis of NAD in both prokaryotes and eukaryotes.

cd03293 NrtD and SsuB are the ATP-binding subunits of the bacterial ABC-type nitrate and sulfonate transport systems, respectively.

cd03450 NodN (nodulation factor N) contains a single hot dog fold similar to those of the peroxisomal Hydratase-Dehydrogenase-Epimerase (HDE) protein, and the fatty acid synthase beta subunit.

cd03454 YdeM is a Bacillus subtilis protein that belongs to a family of prokaryotic proteins of unkown function.

cd03469 Rieske non-heme iron oxygenase (RO) family, N-terminal Rieske domain of the oxygenase alpha subunit; The RO family comprise a large class of aromatic ring-hydroxylating dioxygenases found predominantly in microorganisms. [2Fe-2S] cluster and a C-terminal catalytic domain with a mononuclear Fe(II) binding site. The Rieske [2Fe-2S]

cd03528 Rieske non-heme iron oxygenase (RO) family, Rieske ferredoxin component; composed of the Rieske ferredoxin component of some three-component RO systems including biphenyl dioxygenase (BPDO) and carbazole 1,9a-dioxygenase (CARDO). [2Fe-2S]

COG0009 Putative translation factor (SUA5) [Translation, ribosomal structure and biogenesis]

COG0026 Phosphoribosylaminoimidazole carboxylase (NCAIR synthetase) [Nucleotide transport and metabolism]

COG0041 Phosphoribosylcarboxyaminoimidazole (NCAIR) mutase [Nucleotide transport and metabolism]

COG0148 Enolase [Carbohydrate transport and metabolism]

COG0159 Tryptophan synthase alpha chain [Amino acid transport and metabolism]

COG0224 F0F1-type ATP synthase, gamma subunit [Energy production and conversion]

COG0277 FAD/FMN-containing dehydrogenases [Energy production and conversion]

COG0318 Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II [Lipid metabolism / Secondary metabolites biosynthesis, transport, and catabolism]

COG0333 Ribosomal protein L32 [Translation, ribosomal structure and biogenesis]

COG0365 Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases [Lipid metabolism]

COG0372 Citrate synthase [Energy production and conversion]

COG0403 Glycine cleavage system protein P (pyridoxal-binding), N-terminal domain [Amino acid transport and metabolism]

COG0404 Glycine cleavage system T protein (aminomethyltransferase) [Amino acid transport and metabolism]

COG0446 Uncharacterized NAD(FAD)-dependent dehydrogenases [General function prediction only]

COG0511 Biotin carboxyl carrier protein [Lipid metabolism]

COG0583 Transcriptional regulator [Transcription]

COG0596 Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily) [General function prediction only]

COG0600 ABC-type nitrate/sulfonate/bicarbonate transport system, permease component [Inorganic ion transport and metabolism]

COG0627 Predicted esterase [General function prediction only]

COG0642 Signal transduction histidine kinase [Signal transduction mechanisms]

COG0665 Glycine/D-amino acid oxidases (deaminating) [Amino acid transport and metabolism]

COG0673 Predicted dehydrogenases and related proteins [General function prediction only]

COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain [Signal transduction mechanisms / Transcription]

COG0777 Acetyl-CoA carboxylase beta subunit [Lipid metabolism]

COG0848 Biopolymer transport protein [Intracellular trafficking and secretion]

COG1003 Glycine cleavage system protein P (pyridoxal-binding), C-terminal domain [Amino acid transport and metabolism]

COG1012 NAD-dependent aldehyde dehydrogenases [Energy production and conversion]

COG1024 Enoyl-CoA hydratase/carnithine racemase [Lipid metabolism]

COG1048 Aconitase A [Energy production and conversion]

COG1052 Lactate dehydrogenase and related dehydrogenases [Energy production and conversion / Coenzyme metabolism / General function prediction only]

COG1062 Zn-dependent alcohol dehydrogenases, class III [Energy production and conversion]

COG1129 ABC-type sugar transport system, ATPase component [Carbohydrate transport and metabolism]

COG1172 Ribose/xylose/arabinose/galactoside ABC-type transport systems, permease components [Carbohydrate transport and metabolism]

COG1254 Acylphosphatases [Energy production and conversion]

COG1414 Transcriptional regulator [Transcription]

COG1522 Transcriptional regulators [Transcription]

COG1539 Dihydroneopterin aldolase [Coenzyme metabolism]

COG1558 Flagellar basal body rod protein [Cell motility and secretion]

COG1609 Transcriptional regulators [Transcription]

COG1620 L-lactate permease [Energy production and conversion]

COG1749 Flagellar hook protein FlgE [Cell motility and secretion]

COG1804 Predicted acyl-CoA transferases/carnitine dehydratase [Energy production and conversion]

COG1843 Flagellar hook capping protein [Cell motility and secretion]

COG1879 ABC-type sugar transport system, periplasmic component [Carbohydrate transport and metabolism]

COG2025 Electron transfer flavoprotein, alpha subunit [Energy production and conversion]

COG2080 Aerobic-type carbon monoxide dehydrogenase, small subunit CoxS/CutS homologs [Energy production and conversion]

COG2084 3-hydroxyisobutyrate dehydrogenase and related beta-hydroxyacid dehydrogenases [Lipid metabolism]

COG2223 Nitrate/nitrite transporter [Inorganic ion transport and metabolism]

COG2224 Isocitrate lyase [Energy production and conversion]

COG2226 Methylase involved in ubiquinone/menaquinone biosynthesis [Coenzyme metabolism]

COG2271 Sugar phosphate permease [Carbohydrate transport and metabolism]

COG2303 Choline dehydrogenase and related flavoproteins [Amino acid transport and metabolism]

COG2866 Predicted carboxypeptidase [Amino acid transport and metabolism]

COG2897 Rhodanese-related sulfurtransferase [Inorganic ion transport and metabolism]

COG3012 Uncharacterized protein conserved in bacteria [Function unknown]

COG3232 5-carboxymethyl-2-hydroxymuconate isomerase [Amino acid transport and metabolism]

COG3284 Transcriptional activator of acetoin/glycerol metabolism [Secondary metabolites biosynthesis, transport, and catabolism / Transcription]

COG3384 Uncharacterized conserved protein [Function unknown]

COG3836 2,4-dihydroxyhept-2-ene-1,7-dioic acid aldolase [Carbohydrate transport and metabolism]

COG3971 2-keto-4-pentenoate hydratase [Secondary metabolites biosynthesis, transport, and catabolism]

COG4147 Predicted symporter [General function prediction only]

COG4319 Ketosteroid isomerase homolog [Function unknown]

COG4377 Predicted membrane protein [Function unknown]

COG4583 Sarcosine oxidase gamma subunit [Amino acid transport and metabolism]

COG4630 Xanthine dehydrogenase, iron-sulfur cluster and FAD-binding subunit A [Nucleotide transport and metabolism]

COG4631 Xanthine dehydrogenase, molybdopterin-binding subunit B [Nucleotide transport and metabolism]

COG4757 Predicted alpha/beta hydrolase [General function prediction only]

COG4770 Acetyl/propionyl-CoA carboxylase, alpha subunit [Lipid metabolism]

COG4786 Flagellar basal body rod protein [Cell motility and secretion]

COG4787 Flagellar basal body rod protein [Cell motility and secretion]

pfam00070 Pyridine nucleotide-disulphide oxidoreductase.

pfam00083 Sugar (and other) transporter.

pfam00106 short chain dehydrogenase.

pfam00171 Aldehyde dehydrogenase family.

pfam00496 Bacterial extracellular solute-binding proteins, family 5.

pfam00926 3,4-dihydroxy-2-butanone 4-phosphate synthase.

pfam01315 Aldehyde oxidase and xanthine dehydrogenase, a/b hammerhead domain.

pfam01425 Amidase.

pfam01565 FAD binding domain.

pfam01964 ThiC family.

pfam02233 NAD(P) transhydrogenase beta subunit.

pfam02445 Quinolinate synthetase A protein.

pfam02682 Allophanate hydrolase subunit 1.

pfam02738 Aldehyde oxidase and xanthine dehydrogenase, molybdopterin binding domain.

pfam02823 ATP synthase, Delta/Epsilon chain, beta-sandwich domain.

pfam02913 FAD linked oxidases, C-terminal domain.

pfam03481 SUA5 domain.

pfam03758 Senescence marker protein-30 (SMP-30).

pfam03797 Autotransporter beta-domain.

pfam03972 MmgE/PrpD family.

pfam04267 Sarcosine oxidase, delta subunit family.

pfam04303 Protein of unknown function (DUF453).

pfam04341 Protein of unknown function, DUF485.

pfam04362 Protein of unknown function (DUF495).

pfam05146 Aha1 domain.

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Bam-1-2, Bam-1-3, Bce-1-1, Bce-1-2, Bce-1-3, Bce-1-6, Bce-2-1, Bce-2-10, Bce-2-11, Bce-2-12, Bce-2-13, Bce-2-14, Bce-2-15, Bce-2-16, Bce-2-17, Bce-2-2, Bce-2-3, Bce-2-4, Bce-2-5, Bce-2-6, Bce-2-7, Bce-2-8, Bce-2-9, Bce-3-1, Bce-3-14, Bce-3-15, Bce-3-16, Bce-3-2, Bce-3-4, Bce-3-5, Bce-3-6, Bce-3-8, Bdo-1-1, Bdo-1-11, Bdo-1-18, Bdo-1-6, Bdo-1-8, Bdo-1-9, Bma-1-1, Bma-1-10, Bma-1-11, Bma-1-12, Bma-1-13, Bma-1-2, Bma-1-3, Bma-1-4, Bma-1-5, Bma-1-6, Bma-1-7, Bma-1-8, Bma-1-9, Bma-2-1, Bma-2-10, Bma-2-11, Bma-2-12, Bma-2-13, Bma-2-2, Bma-2-3, Bma-2-4, Bma-2-5, Bma-2-6, Bma-2-7, Bma-2-8, Bma-2-9, Bma-3-1, Bma-3-10, Bma-3-11, Bma-3-12, Bma-3-13, Bma-3-2, Bma-3-3, Bma-3-4, Bma-3-5, Bma-3-6, Bma-3-7, Bma-3-8, Bma-3-9, Bma-4-11, Bma-4-2, Bma-4-3, Bma-4-5, Bma-4-6, Bma-4-8, Bma-4-9, Bma-5-1, Bma-5-10, Bma-5-11, Bma-5-12, Bma-5-13, Bma-5-2, Bma-5-3, Bma-5-4, Bma-5-5, Bma-5-6, Bma-5-7, Bma-5-8, Bma-5-9, Bma-6-1, Bma-6-10, Bma-6-11, Bma-6-12, Bma-6-13, Bma-6-2, Bma-6-3, Bma-6-4, Bma-6-5, Bma-6-6, Bma-6-7, Bma-6-8, Bma-6-9, Bma-7-1, Bma-7-10, Bma-7-11, Bma-7-12, Bma-7-13, Bma-7-2, Bma-7-3, Bma-7-4, Bma-7-5, Bma-7-6, Bma-7-7, Bma-7-8, Bma-7-9, Bma-8-1, Bma-8-10, Bma-8-11, Bma-8-12, Bma-8-13, Bma-8-14, Bma-8-15, Bma-8-16, Bma-8-2, Bma-8-3, Bma-8-4, Bma-8-5, Bma-8-6, Bma-8-7, Bma-8-8, Bma-8-9, Bma-9-1, Bma-9-10, Bma-9-11, Bma-9-12, Bma-9-13, Bma-9-2, Bma-9-3, Bma-9-4, Bma-9-5, Bma-9-6, Bma-9-7, Bma-9-8, Bma-9-9, Bps-1-1, Bps-1-10, Bps-1-11, Bps-1-12, Bps-1-13, Bps-1-14, Bps-1-16, Bps-1-17, Bps-1-18, Bps-1-19, Bps-1-2, Bps-1-20, Bps-1-21, Bps-1-23, Bps-1-24, Bps-1-25, Bps-1-26, Bps-1-27, Bps-1-28, Bps-1-29, Bps-1-3, Bps-1-30, Bps-1-31, Bps-1-32, Bps-1-33, Bps-1-34, Bps-1-35, Bps-1-4, Bps-1-5, Bps-1-6, Bps-1-7, Bps-1-8, Bps-1-9, Bps-10-1, Bps-10-10, Bps-10-11, Bps-10-12, Bps-10-14, Bps-10-16, Bps-10-17, Bps-10-18, Bps-10-19, Bps-10-2, Bps-10-20, Bps-10-21, Bps-10-22, Bps-10-24, Bps-10-25, Bps-10-26, Bps-10-27, Bps-10-28, Bps-10-29, Bps-10-3, Bps-10-5, Bps-10-6, Bps-10-7, Bps-10-8, Bps-10-9, Bps-2-1, Bps-2-10, Bps-2-11, Bps-2-12, Bps-2-13, Bps-2-14, Bps-2-15, Bps-2-16, Bps-2-17, Bps-2-18, Bps-2-19, Bps-2-2, Bps-2-20, Bps-2-21, Bps-2-22, Bps-2-23, Bps-2-24, Bps-2-25, Bps-2-26, Bps-2-27, Bps-2-28, Bps-2-29, Bps-2-3, Bps-2-30, Bps-2-31, Bps-2-4, Bps-2-5, Bps-2-6, Bps-2-7, Bps-2-8, Bps-2-9, Bps-3-1, Bps-3-11, Bps-3-13, Bps-3-19, Bps-3-2, Bps-3-20, Bps-3-21, Bps-3-24, Bps-3-25, Bps-3-26, Bps-3-3, Bps-3-5, Bps-3-6, Bps-3-8, Bps-3-9, Bps-4-1, Bps-4-10, Bps-4-11, Bps-4-12, Bps-4-13, Bps-4-14, Bps-4-15, Bps-4-16, Bps-4-17, Bps-4-18, Bps-4-19, Bps-4-2, Bps-4-20, Bps-4-21, Bps-4-22, Bps-4-23, Bps-4-3, Bps-4-4, Bps-4-5, Bps-4-6, Bps-4-7, Bps-4-8, Bps-4-9, Bps-5-1, Bps-5-11, Bps-5-12, Bps-5-17, Bps-5-18, Bps-5-19, Bps-5-2, Bps-5-21, Bps-5-23, Bps-5-3, Bps-5-4, Bps-5-5, Bps-5-6, Bps-5-7, Bps-5-8, Bps-5-9, Bps-6-1, Bps-6-10, Bps-6-11, Bps-6-12, Bps-6-13, Bps-6-14, Bps-6-15, Bps-6-16, Bps-6-17, Bps-6-19, Bps-6-2, Bps-6-21, Bps-6-22, Bps-6-23, Bps-6-24, Bps-6-25, Bps-6-26, Bps-6-27, Bps-6-28, Bps-6-29, Bps-6-3, Bps-6-30, Bps-6-31, Bps-6-32, Bps-6-4, Bps-6-5, Bps-6-6, Bps-6-9, Bps-7-1, Bps-7-10, Bps-7-11, Bps-7-12, Bps-7-13, Bps-7-15, Bps-7-16, Bps-7-17, Bps-7-18, Bps-7-19, Bps-7-2, Bps-7-20, Bps-7-21, Bps-7-22, Bps-7-23, Bps-7-24, Bps-7-25, Bps-7-27, Bps-7-28, Bps-7-3, Bps-7-30, Bps-7-4, Bps-7-5, Bps-7-6, Bps-7-8, Bps-7-9, Bps-8-1, Bps-8-10, Bps-8-11, Bps-8-2, Bps-8-3, Bps-8-4, Bps-9-1, Bps-9-10, Bps-9-11, Bps-9-13, Bps-9-16, Bps-9-17, Bps-9-18, Bps-9-2, Bps-9-21, Bps-9-22, Bps-9-23, Bps-9-24, Bps-9-25, Bps-9-26, Bps-9-27, Bps-9-28, Bps-9-29, Bps-9-3, Bps-9-30, Bps-9-31, Bps-9-32, Bps-9-4, Bps-9-5, Bps-9-6, Bps-9-7, Bps-9-8, Bps-9-9, Bsp-1-10, Bsp-1-14, Bsp-1-2, Bsp-1-7, Bth-1-1, Bth-1-17, Bth-1-18, Bth-1-19, Bth-1-3, Bth-1-4, Bth-1-6, Bxe-1-2, Bxe-1-3, Bxe-1-5, Bxe-1-7, Bxe-1-8, env-1, env-11, env-13, env-14, env-15, env-16, env-17, env-18, env-19, env-2, env-21, env-22, env-24, env-25, env-26, env-27, env-3, env-31, env-32, env-33, env-35, env-37, env-38, env-40, env-42, env-46, env-47, env-48, env-49, env-55, env-59, env-6, env-61, env-62, env-63, env-65, env-66, env-67, env-7, env-70, env-71, env-72

alignment positions 1 · · · 56

```
Bam-1-1     GC.AUGG.GACCC.GGGU.AAGUGCUGAA........GCACUAACUCUGG.UCGAC
Bsp-1-1     GC.AUGG.GACCA.GAGU.AAGUGCUGAA........GCACUAACUCUGG.UCGAC
Bma-4-1     GC.AUGG.GACCA.GAGU.AAGCGCUGAA........GCGCCAACUCUGG.UCGAC
Bxe-1-1     GC.AUGG.GACCA.GAGU.AAGCGCUGAA........GCGCUAACUCUGG.UCGAC
Bxe-1-4     GC.AUGG.GACCA.GAGU.AAGCGCUAAA........GCGCUAACUCUGG.UCGAC
Bps-7-7     GC.AUGG.GACCA.GAGU.AAGUGCUAAA........GCGCUAACUCUGG.UCGAC
Bxe-1-6     GC.AUGG.GACCA.GAGU.AAGCGCUAAA........GCGCCAACUCUGG.UCGAC
Bps-8-5     AC.AUGG.GACCA.GAGU.AAGCGCUAAA........GCGCUAACUCUGG.UCGAC
Bxe-1-9     GC.AUGG.GACCA.GAGU.AAGCGCUAAA........GCACUAACUCUGG.UCGAC
Bxe-1-10    GC.AUGG.GACCA.AAGU.AAGCGCUAAA........GCGCUAACUCUGG.UCGAC
Bps-3-4     GC.AUGG.GACCA.GAGU.AAGCGCUCAA........GCGCUAACUCUGG.UCGAC
Bma-4-4     GC.AUGG.GACCA.GCGU.AAGCGCUAAA........GCGCUAACUCUGG.UCGAC
env-4       GC.AUGG.GACCA.GCGU.AAGCGCUAAA........GCGCUAACGCUGG.UCGAC
Bps-8-6     GC.AUGG.GACCG.GAGU.AAGCGCUGAA........GCGCUAACUCUGG.UCGAC
env-5       GC.AUGG.GACCA.GCAU.AAGCGCUAAA........GCGCUAACGCUGG.UCGAC
Bps-6-7     GC.ACGG.GACCA.GAGU.AAGCGCUGAA........GCGCUAACUCUGG.UCGAC

            ........<<<<<.<<<<..<<<<<............>>>>>.>>>>>>>.>>...
            ........???2?.????..?221?............?122?.?????2?.??...
            GC-AUGG-GRYC•-RRGY-RRGCGCYRAA--------GCGCYRACYY•GR-YCGAC
```

```
  Bps-6-8     GC.ACGG.GACCA.GAGU.AAGCGCUAAA........GCGCCAACUCUGG.UCGAC
  Bps-10-4    GCCAUGG.GACCA.GAGU.AAGCGCUAAA........GCGCUAACUCUGG.UCGAC
  Bps-3-7     GCCAUGG.GAUCA.GAGU.AAGCGCUAAA........GCGCUAACUCUGG.UCGAC
  Bps-8-7     GC.AUGG.GACCG.GAGU.AAGCGCUAAA........GCGCUAACUCUGG.UCGAC
  Bsp-1-3     GC.AUGG.GACCG.GAGU.AAGCGCUAAA........GCGCUAACUCCGG.UCGAC
  Bps-8-8     GC.AUGG.GAUCG.GAGU.AAGCGCUGAA........GCGCUGACUCCGA.UCGAC
  Bps-8-9     GC.AUGG.GAUCG.GAGU.AAGCGCUGAA........GCGCUAACUCCGA.UCGAC
  Bth-1-2     GC.AUGG.GAUCG.GAGU.AAGCGCUAAA........GCGCUAACUCCGA.UCGAC
  Bps-3-10    GC.AUGG.GAUCG.GAGU.AAGCGCUGAA........GCGUUGACUCCGA.UCGAC
  Bps-9-12    GC.AUGG.GAUCG.GAGU.AAGUGCUGAA........GCGUUGACUCCGA.UCGAC
  Bps-9-14    GC.AUGG.GAUCG.GAGU.AAGCGCUGAA........GCGCUGACUCCGA.UUGAC
  Bth-1-5     GC.AUGG.GAUCG.GAGU.AAGCGCUAAG........GCGCUAACUCCGA.UCGAC
  Bps-7-14    GC.AUGG.GAUCG.GAGU.GAGCGCUGAA........GCGCUGACUCCGA.UCGAC
  Bth-1-7     GC.AUGG.GACCG.CAGU.AAGCGCUAAA........GCGCCAACUGCGG.UCGAC
  Bxe-1-11    GC.AUGG.GACCG.CAGU.AAGCGCUAAA........GCGCUAACUGCGG.UCGAC
  Bdo-1-2     GC.AUGG.GACCG.CAGU.AAGCGCUGAA........GCGCUAACUGCGG.UCGAC
  env-8       GC.AUGG.GAUCG.GAGU.AAGCACUAAA........GCGCCAACUCCGA.UCGAC
  env-9       GC.AUGG.GAUCA.GAAU.AAGCGCUAAA........GCGCUAACUCUGA.UCGAC
  env-10      GC.AUGG.GAUCA.GAGU.AAGCGCUAAA........GCGCUAACUCUGA.UCGAC
  Bdo-1-3     GC.AUGG.GACCG.GAGU.AAGCGCUGAA........GCACUAACUCCGG.UCGAC
  Bdo-1-4     GC.AUGG.GACCG.GAGU.AAGUGCUGAA........GCACUAACUCCGG.UCGAC
  Bdo-1-5     GU.AUGG.GACCG.GAGU.AAGUGCUGAA........GCACUAACUCCGG.UCGAC
  Bce-3-3     GC.AUGG.GGCCC.GAGU.AAGCGCUGAA........GCGCCAACUCGGG.CCGAC
  Bce-1-4     GC.AUGG.GACCC.GAGU.AAGCGCUGAA........GCGCCAACUCGGG.CCGAC
  env-12      GC.AUGG.GACCC.GAGU.AAGCGCUGAA........GCGCUAACUCGGG.CCGAC
  Bam-1-4     GC.AUGG.GACCC.GAGU.AAGCGCUGAA........GCGCUAACUCGGG.UCGAC
  Bsp-1-4     GC.AUGG.GACCC.GAGU.AAGCGCUAAA........GCGCUAACUCGGG.UCGAC
  Bsp-1-5     GC.AUGG.GACAG.CAGU.AAGCGCUGAA........GCGCUAACUGCUG.UCGAC
  Bce-1-5     GC.AUGG.GACCC.AAGU.AAGCGCUAAA........GCGCUAACUUGGG.UCGAC
  Bsp-1-6     GC.AUGG.GACCC.AAGU.AAGCGCUGAA........GCGCUAACUUGGG.UCGAC
  Bce-3-7     GC.AUGG.GACCC.AAGU.AACCGCUAAA........GCGCUAACUUGGG.UCGAC
  Bce-1-7     GC.AUGG.GACCC.AAGU.AACCACUAAA........GCGCUAACUUGGG.UCGAC
6 env-20      GC.AUGG.GACCC.AAGU.AAGCGCUGAA........GCGCUAACUUUGA.UCGAC
  Bvi-1-1     GC.AUGG.GACCC.ACGU.AAGCGCUAAA........GCGCUAACGUGGG.UCGAC
  env-23      GC.AUGG.GACCC.ACGU.AAGCGCUAAA........GCGCUAACGUGGA.UCGAC
  Bdo-1-7     GC.AUGG.GACCG.AAGU.AAGCGCUGAA........GCGCUAACUCCGG.UCGAC
  Bps-9-15    GC.AUGG.GACCA.GAGU.AAGUGCUCAA........GCGCCAACUCGGG.UCGAC
  Bps-8-12    GC.AUGG.GACCA.GAGU.AAGUGCUCAA........GCGCCAACUCUGG.UCGAC
  Bam-1-5     GC.AUGG.AACCA.GAGC.AAGCGCUAAA........GCGCCAACUCUGG.UCGAC
  Bth-1-8     GC.AUGG.GACCA.GAGU.CAGCGCUGAG........GCGCCAACUCUGG.UCGAC
  Bps-1-15    GC.AUGG.GACCA.GAGU.AAGYGCUMAA........GCGCCAACUCUGG.UCGAC
  Bxe-1-12    GC.AUGG.GACCG.CGGU.AAGCGCCGAA........GCGCUAACUGCGG.UCGAC
  Bps-8-13    GCCAUGG.GAUCA.GAGU.AAGCGCUGAA........GCGCUAUCUCUGG.UCGAC
  Bam-1-6     GC.AUGG.GGCCA.GAGU.AAGCGCUAAA........GCGCCAACUCUGG.CCGAC
  Bsp-1-8     GC.AUGG.GAUCC.AAGU.AACCGCUAAA........GCGCUAACUUUGA.UCGAC
  Bsp-1-9     GC.AUGG.GACCC.AAGU.AACCGCUAAA........GCGCUAACUUUGA.UCGAC
  Bdo-1-10    GC.AUGG.GGCCC.GAGU.AAGCGCUAAA........GCGCUAACUCGGG.CCGAC
  Bxe-1-13    GC.AUCG.GACCA.GAGU.AAGCGCUGAA........GCGCCAACUCUGA.UCGAC
  env-28      GC.AUGG.GACGA.GAGU.AAGCGCUAAA........GCGCUAACUCUCG.UCGAC
  Bps-8-14    GC.AUGG.GAUCG.GGGU.AAGCGCUGAA........GCGCUAACUCUGA.UCGAC
  env-29      GC.AUGG.GACCG.GAAU.AAGCGCUAAA........GCGCCAACUCCGG.UCGAC

              ........<<<<<.<<<<..<<<<<............>>>>>.>>>>>>>.>>...
              ........???2?.????..?221?............?122?.?????2?.??...
              GC-AUGG-GRYC•-RRGY-RRGCGCYRAA--------GCGCYRACYY•GR-YCGAC
```

```
Bps-5-10    GC.AUGG.GAUCG.GGGC.AAGCGCUGAA........GCGCUAACUCCGA.UCGAC
Bxe-1-14    GC.AUGG.GAUCG.GAGU.AACCGCUAAA........GCGGUAACUCCGG.UCGAC
Bxe-1-15    GC.AUGG.GAUCG.GAGU.AACCGCUAAA........GCGGUAACUCCGA.UCGAC
Bxe-1-16    GC.AUGG.GAUCG.GAGU.AACGGCUAAA........GCGGUAACUCCGA.UCGAC
env-30      GC.AUGG.GACCC.GAGU.AAGCGCUGAA........GCGCCAACCCGGG.UCGAC
env-34      GC.AUGG.GACCG.GAGU.AAGUGCUGAA........GCGCCAACUCUGG.UCGAC
Bps-3-12    GC.AUGG.GAUCG.GAGU.CAGCGCUGMA........GCGCUGACUCCGA.UCGAC
Bdo-1-12    GC.AUGG.GGCCG.GAGU.AAGCGCUAAA........GCGCUAACUUCGG.CCGAC
Bth-1-9     GC.AUGG.GGCCG.GAGU.AAGCGCUAAA........GCGCUAACUCCGG.CCGAC
Bdo-1-13    AC.AUGG.GGCCG.GAGU.AAGCGCUGAA........GCGCUAACUCCGG.CCGAC
env-36      GC.AUGG.GGCCG.GACU.AAGCGCUAAA........GCGCUAAGUCCGG.CCGAC
env-39      GC.AUGG.GGCCG.GACU.AAGCGCUGAA........GCGCCAAGUCCGG.CCGAC
Bsp-1-11    GC.AUGG.GGCUG.GAGU.AAGCGCUAAA........GCGCUAACUCCAG.CCGAC
env-41      GC.AUGG.GGUUG.GAGU.AAGCGCUAAA........GCGCUAACUCCAA.CCGAC
Bce-1-8     GC.AUGG.GGCCG.GAGU.AAGUGCUAAA........GCACUCACUCCGG.CCGAC
Bps-10-13   GC.AUGA.GGCCG.GAGU.AAGUGCUCAA........GCGCCAACUCGGG.UCGAG
Bps-8-15    GC.AUGG.GGCCG.GAGU.AAGUGCUCAA........GCGCCAACUCGGG.UCGAG
Bps-3-14    GC.GUGG.GGCCG.GAGU.AAGUGCUCAA........GCGCCAACUCGGG.UCGAG
Bps-9-19    GC.GUGG.GGCCG.GAGU.AAGCGCUCAA........GCGCCAACUCGGG.UCGAG
Bps-6-18    GC.GUGG.GGCCG.GAGU.AAGCGCUCAA........GCGCCAACUCGGG.GCGAG
Bps-3-15    GC.AUGG.GAUCG.GGGC.AAGCGCCGAA........GCGCCGACUCCGA.UCGAC
Bps-8-16    GC.AUGG.GAUCG.GGGC.AAGCGCCGAA........GCGCCGACUUCGA.UCGAC
Bps-10-15   GC.AUGG.GAUCG.GGGC.AAGCGCUGAA........GCGCCGACUUCGA.UCGAC
Bps-9-20    GC.AGGG.GAUCG.GGGC.AAGCGCCGAA........GCGCCGACUCCGA.UCGAC
Bps-1-22    GC.AUGG.GAUCG.GGGC.AAGCGCCGAA........GCGCCGACUUCGA.UCGGC
Bps-5-13    GC.AUGG.GAUCG.GGGU.AAGCGCUGAA........GCGCCGACUUCGA.UCGAC
Bps-6-20    GC.AUGG.GAUCG.GGGC.AAGCGCCGAA........GCGCUAACUUCGA.UCGGC
Bps-3-16    GC.AGGG.GAUCG.GGGC.AAGCGCCGAA........GCGCUAACUUCGA.UCGAC
Bps-5-14    GC.AUGG.GAUCG.GGGU.ACGCGCUGAA........GCGCCAACUUCGG.UCGAC
Bma-4-7     GC.AUGG.GAUCG.GGGU.ACGCGCUGAA........GCGCCAACUUCGG.UCGAU
Bps-8-17    AC.AUGG.GAUCG.GGGU.ACGCGCUGAA........GCGCCAACUUCGG.UCGAC
Bps-3-17    AC.AUGG.GAUCG.GGGU.ACGCGCUGAA........GCGCCAACUUCGG.UCGAU
Bps-3-18    AC.AUGG.GAUCG.GGGU.ACGCGCUGAA........GCGUCAACUUCGG.UCGAU
Bps-5-15    GC.AUGG.GAUCG.GGGU.ACGCGCUGAA........GCGCCAGCUCCGG.UCGAC
Bps-5-16    GC.AUGG.GAUCG.GGGU.ACGCGCUGAA........GCGCCAGCUCCGG.UCGAU
Bps-8-18    GC.AUGG.GACCG.GGGU.ACGCGCUGAA........GCGCCAGCUCCGG.UCGAU
Bps-8-19    GC.AUGG.GACCG.GAGU.AAGCGCUGAA........GCGCCAGCUCCGG.UCGAC
Bps-8-20    GC.AUGG.GACCG.GGGU.AAGCGCUGAA........GCGCCAGCUCCGG.UCGAC
Bps-8-21    GC.AUGG.GACCG.GGGU.AAGCGCUAAA........GCGCCAGCUCUGG.UCGAC
env-43      GC.AUAG.GACCA.GAGU.AAGCGCCAAA........GCGCUAACUCUGA.UCGAC
Bce-1-9     GC.AUGG.GACCA.GAAC.AACCGCUAAA........GCGCCAACUCUGG.UCGAC
env-44      AC.AUGG.GACCG.GCGU.AAGCGCUAAA........GCGCUAACGCCGG.UCGAC
Bxe-1-17    GC.AUGG.GACCU.GAGC.AAGCGCUAAA........GCGCCAGCUCCGG.UCGAC
Bps-3-22    GC.AUGG.GGCCG.GAGU.AAGCGCUCAA........GCGCCAACUCUGG.UCGAC
env-45      GC.AUGG.GACCC.ACGU.AAGCGCUGAA........GGGCUGACGUGGA.UCGAC
env-50      GC.AUGG.AGCCG.GAGU.AAGCGCUGAA........GCGCCGACUCUGG.UCGAC
Bxe-1-18    GC.AUGG.GACUG.GAAU.AAGCGCUAAA........GCGCUAACUCCAG.UCGAC
Bam-1-7     GC.AUGG.GACAG.GAGU.AAGCGCUAAA........GCGCUAACUCCUA.UCGAC
Bce-1-10    GC.AUGG.GGCGG.GAGU.AAGUGCUGAA........GCACUCACUCCCG.CCGAC
Bsp-1-12    GC.AUGG.GGUGG.GAGU.AAGUGCUAAA........GCACUCACUCCCG.CCGAC
Bam-1-8     CC.AUGG.GGCCAUGAGU.AAGCGCUAAA........GCGCCAACUCUGG.UCGAC

            ........<<<<<.<<<<..<<<<<............>>>>>.>>>>>>>.>>...
            ........????2?.????..?221?............?122?.?????2?.??...
            GC-AUGG-GRYC•-RRGY-RRGCGCYRAA--------GCGCYRACYY•GR-YCGAC
```

```
 Bdo-1-14     GC.AUGG.GACCG.GAGUAAGGCGCUAAA........GCGCCAACUCCGG.UCGAC
 Bdo-1-15     GC.AUGG.GAUCG.GAGUAAGGCGCUAAA........GCGCCAACUCCGA.UCGAC
 Bth-1-10     GC.AUCG.GACCA.GAGC.AAGCGCUAAA........GCGCUAACUCCGG.UCGAC
 Bce-1-11     GC.AUGG.GACCC.CAGU.AAGCGCUAAA........GCGCUAACUGGGGAUCGAC
 Bce-3-9      GC.AUGG.GAUCU.GAGU.AAGCACUAAA........GCGCUAACUCAGA.UCGAC
 Bce-3-10     GC.AUGG.GAUCU.GAGC.AAGCGCUAAA........GCGCUAACUCAGA.UCGAC
 Bsp-1-13     GC.AUGG.GGCCC.AAGC.AAGCGCUAAA........GCGCUAACUCGGG.UCGAC
 Bps-7-26     GCCAUGG.GAUCA.GAGU.AAGCGCUGAA........GCGCCAACUCUGG.UUGAC
 env-51       GC.AUGG.GACCC.GAGC.AAGCGCUGAA........GUGCUGACUCGGG.UCGAC
 env-52       GC.AUGG.GACCC.GAGC.AAGCGCUGAA........GUGCUGACUCGGG.CCGAC
 Bsp-1-15     GC.AUGG.GACGA.GAGU.AAGUGCUGAA........GCACUAACUCUCG.CCGAC
 Bxe-1-19     GC.AUGG.GACCG.GAGU.CAGCGCUGAA........GCGCUAACUCCGG.CCGAC
 Bxe-1-20     GC.ACGG.GACCG.GAGU.AGGCGCUGAA........GCGCUAACUCCGG.CCGAC
 Bxe-1-21     GC.AUGG.GACCG.GAGU.AAGCGUCGAA........GCGCCAACUCUGG.UCGAC
 Bps-3-23     GC.AUGG.GAUCG.GGGC.AAGCGUCGAA........GCGCUAACUCCGC.UCGAC
 env-53       GC.GUGG.GAUCU.GAGC.AAGCGCUGAA........GGGCUGACUCGGG.UUGAC
 env-54       GC.GUGG.GCUCU.GAGC.AAGCGCUGAA........GUGCUGACUCGGG.UCGAC
 Bsp-1-16     GC.AUGG.GGUCC.CAGU.AAGCGCUAAA........GCGCUAACUGGGA.CCGAC
 env-56       GC.AUGG.GACCA.GAGC.AAGUGCUGAA........GCACUAACUCUGG.UCACA
 Bvi-1-2      GC.AUGG.GAUCG.GAGUUGAGCGCUGAA........GCGCCCGCUCCGG.UCGAC
 Bam-1-9      GC.AUGG.GACGG.GAGU.AAGCGCUAAA........GCGCCAACUCCCG.UCGGC
 Bth-1-11     GC.AUGG.GAUCG.GGGU.AGGCGCUGAA........GCGCCAACGCCGA.UCGAC
 Bth-1-12     GC.AUGG.GAUCG.GGAA.AGGCGCUGAA........GCGCCAACGCCGA.UCGAC
 Bam-1-10     GC.AUGG.GAUCA.GAGC.AAGCGUUAAA........GCGCCAACUCUGG.CCGAC
 env-57       GC.AUGG.GACAU.CAGU.AAGCGCUAAA........GCGCUAACUGGGA.CCGAC
 Bsp-1-17     GC.AUGG.GACCG.GAGU.CAGCGCUGAA........GCGCCCACUGCGG.UCGAC
 env-58       GC.AUGG.GACGA.CAGU.AAGUGCUAAA........GCACUAACUGUCG.UCGAC
 env-60       GC.AUGG.GGCUC.GGGU.GGGCGCUGAA........GCGCCAACUCGGG.CCGAC
 env-64       GC.AUGG.GGCUC.GGGU.GAGCGCUGAA........GCGCCAACUCGGG.UCGAC
 Bce-1-12     GC.AUGG.GACUC.AGGU.AACCGCUAAA........GCGCUAACCUGGA.UCGAC
 env-68       GC.AUGG.GAUCC.AGGU.AAGCGCUAAA........GCG.UAACUUGGA.UCGAC
 Bxe-1-22     GC.AUGG.GGGCG.GAGU.GAGCGCUGAA........GCGCUAACUGCGG.UCGAC
 Bsp-1-18     GC.AUGG.GGCCG.CACC.AGGCGCUGAA........GCGCCCAGUCCGG.CCGAC
 Bxe-1-23     GC.AUGG.GGUCG.GAGU.AAGCGCUAAA........GCACCAACUCCGA.CCGAC
 Bxe-1-24     GC.AUGG.AGUCG.GAGU.AGGCGCUAAA........GCACCAACUCCGA.CCGAC
 Bps-10-23    AC.AUGG.GACCA.GAGC.AGGUGCUGAA........GCGCUAACUCUGG.UCGAC
 Bce-3-11     GC.AUGG.GAUCC.AGGU.AACCGCUAAA........GCGCCAACCCGGA.UCGAC
7Bps-8-22     GC.AUGG.GAUCG.GAGU.GAGCGCUGAA........GCGUUCGCUUCGG.UCGAC
 Bps-5-20     GC.AUGG.GAUCG.GGGU.GAGCGCUGAA........GCGUUCGCUUCGG.UCGAC
 Bce-3-12     GC.AUGA.GACUG.GUGU.AAGCGCCAAA........GCGCUAACUCCGG.UCGAC
 Bsp-1-19     GC.AUGG.GACCC.GAGU.AAGCGCUGAA........GUGCCGACUCGGA.CCGGC
 Bam-1-11     GC.AUGG.GGCCC.CCGU.AAGCGCUGAA........GCGCCAACGUGGG.CCGAC
 Bvi-1-3      GC.AUGG.GGUCC.ACGU.AGGCGCUGAA........GCGCCAACGUGGA.CCGAC
 Bce-1-13     GC.AUGG.GGCCG.GAGU.ACGCGCUGAA........GCGC.UACUCCGGGCCGAC
 Bsp-1-20     GC.AUGG.GGCCG.GAGC.ACGCGCGCAA........GCGCCCACUCCGG.UCGAC
 Bxe-1-25     GC.AUGC.GACCG.GAGU.AACCGCUAAA........GCGCUGACUCCGGAUCGCC
 Bxe-1-26     UU.ACGG.GACCG.GAGG.AAGUGCUGAA........GCACUCACUCCGC.UCGAC
 Bsp-1-21     GC.AUGG.GACCG.CCCU.AAGCGCUGAA........GCGCCAAGGGCGG.UCGAC
 Bsp-1-22     GC.AUGG.GACCU.CCCU.AAGCGCUAAA........GCGCCAAGGGCGG.UCGAC
 env-69       GC.AUGG.GGCCU.GGGG.GAGCGCUGAA........GCGCCAAUUCGGG.GCGAC
 Bsp-1-23     GC.AUGG.GGUCA.GCGU.AAGCGCUGAA........GCGCUAACGCUGA.CCGAC

              ........<<<<<.<<<<..<<<<<............>>>>>.>>>>>>>.>>...
              ........???2?.????..?221?............?122?.??????2?.??...
              GC-AUGG-GRYC•-RRGY-RRGCGCYRAA--------GCGCYRACYY•GR-YCGAC
```

```
 Bdo-1-16    GC.AUGG.GACUG.GAGC.AAGCGCUGAA........GCGGUGACUCCGG.UCGAC
 Bxe-1-27    GC.AUGG.GACCG.GAGUAAUGCGCCAAA........GCGCGAACUCCGG.CCAAC
 Bth-1-13    GC.AUGG.GAUCG.GAGU.CAACGCCGAA........GCGCCAGCUCCGA.UCGAC
 Bsp-1-24    GC.AUGG.GACAG.GAGC.AAGCGCUGAA........GCGCCAACUCCCG.UAGAC
 Bdo-1-17    GC.AUGG.GGUAG.CCGU.AAGCGCUAAA........GCGCUAACGCCUA.CCGAC
3Bth-1-14    GC.AUGG.AACCA.GAGC.AUGCGCCAA.........GCGCUAACUCCGG.UCGAC
 Bth-1-15    GC.AUGG.GACCC.GGGC.GAGCGCCAAA........GCGCCAACUCGGG.UCGAC
 Bxe-1-28    GC.AUGG.GACCG.AGGU.AAGCGCCGAC........GCGCUUACUCUGG.CCGAC
9Bsp-1-25    GC.AUGG.GAUCC.ACGU.AAGCGACAAG........GCGCUCACGCGGG.UCGAC
 Bxe-1-29    GC.AUGG.GAUCG.CAGC.AAGCGUCAGA........GCGCCAACUGCGA.UCGAC
 Bce-3-13    GC.AUGG.CGCCC.GGGC.AAGCGCUGAA........GCGCUCACUUGGG.UCGAC
 Bxe-1-30    GC.AUGG.GACCC.GAGU.UACCGCGAAA........GCGGUAACCCGGG.CCGAC
 env-73      GG.AUGG.GAUCG.GAGU.GGGCGCUAAC........GCGCCAACGCUGA.UCGAC
 Bvi-1-4     GU.GUGG.AGUUC.ACGU.AGGCGCUGAA........GCGCCAAGGUGGG.CCGAC
 env-74      GC.AUGG.GAUCG.GCGC.AGGCGCGCAA........GUGCCCGCUCCGA.UCGAC
 Bam-1-12    GC.AUGG.GAUCG.GAGU.CAGCGCGCAA........GUGUCCGCUCCGA.UCGAC
 Bsp-1-26    GG.AGGGGGGCC.GAGU.AAGCGCUGAA........GCGCUAACUCGGA.CCGAC
 Bsp-1-27    AC.AUGG.GGCCC.GGGU.GAGCGCUGAA........GUGCUGACUCGGG.UCGAU
 Bsp-1-28    GC.AUGG.GAUCC.AGGC.AAGCACCAAA........CCGCCAACCUGGA.UCGAC
 Bce-1-14    GC.AUGG.GACUG.GCGU.GAGAGCCAAA........GCGCCAACGCCGG.UCGAC
 Bxe-1-31    GG.AUGG.GAGCG.GAGU.UGGCGGUAAA........GCGCUGACUCUGG.UCGAC
 Bdo-1-19    GC.AUGG.GAUCG.GAGC.GGCCGCUGAA........GCGGCAACUCCGA.UCGAU
 Bce-1-15    GC.AU...GACCC.GAGU.CAGUGCCAAA........GCACUCACUCGGG.UCGAC
 Bps-8-23    GC.AUGG.GAUUG.GGGU.CGGCGCUGAA........AUGCCGACUCCGG.UCGAC
 Bma-4-10    GC.AUGG.GAUUG.GGGG.CGGCGCUGAA........AUGCCGACCCCGG.UCGAC
 Bps-5-22    ....CGG.GGCCA.GAGUAAAGCGCUGAA........GCGCUGACUCUGG.CCAAC
 Bps-8-24    ....CGG.GGCCA.GAGUAAAGCGCUGAA........GCGCUGACUCUGG.CCGAC
 Bxe-1-32    GC.AUGG.GACCG.GAGC.CGGCGCUGAG........GCGCUGACCCCGG.UCUAC
1Bxe-1-33    GC.ACGG.GACCA.GAGC.AGGCGCCGGA........GCGCCAACUCCGA.UCGAC
 Bce-1-16    GC.ACGG.GCUCG.GGGC.CGGCGCCGAA........GCGCCCGCUCCGA.CCGAC
4Bma-4-12    CC.GCGG.GACCG.GAGUAAAGCGCUAAA........GCGCUGACUCUGG.CUAAC
8Bth-1-16    GU.AUCG.GAUCA.UCGU.AAGCGCUCAA........GCGCUAACGAUGA.UCGAG
 Bsp-1-29    GC.AUGG.GGUUC.AGGU.GAGUGCUAAA........ACACUGACCUGGA.UCGAC
5Bxe-1-34    GC.CUGG.GACCG.GAGUGACGCGCUCAA........GCGCGAACUCUGG.CCGGU
 Bsp-1-30    ....UGG.GGGUG.GAGU.AAGCGCUGAA........GCGCUAACUCCAC.CCAUC
 Bth-1-20    GC.AUGG.GACCG.GAAU.CGGCGCCGAG........GCGCCCGAUCCGG.UCCAC
 Bam-1-13    GC.AUAG.GGCCC.AAGU.GCGCGCUGAA........GUGCCAAGCUGGG.ACGAC
 Bce-1-17    GC.ACGG.GGCCG.GAUC.GCGCGCUGAA........GCGCCGGGAUCGG.CCGAC
 Bps-7-29    GC.AUGG.GGCCG.GGGG.A.GCGUUGAA........GCGUGAACUCGGG.ACGAG
 Bth-1-21    GU.AUCG.GACUA.UGGGCAGGCGCUGAG........GUACCGACUUCGG.UCGAC
 Bma-4-13    GC.AUCG.GACGA.UGGGCAAGCGCUGAA........GCGCCAAUUUCGG.UCGAC
 Bps-8-25    GC.AUCG.GACGA.UGGGCAAGCGCUGAA........GCGCCAAUUUCGG.UCGAU
 Bsp-1-31    CG.GUGG.GAUCG.GCGU.AAGUGCUCAG........GCACUAACGCAGG.UCGAC
 Bam-1-14    GC.AUGG.AAUCA.GGCA.AAGCGCUGGA.......UCCGCCGACUCUGA.CCGAC
 Bsp-1-32    GC.ACGG.GGCAC.GGGC.CGGCGCAGAA........GCGCCCGCCCUCG.CCGAG
 Bam-1-15    GC.AUGG.GGCCU.CCGC.AAGCGCCAAG........ACGCAGACGGAGG.CCGAC
 Bdo-1-20    GG.CAUG.GAUCG.GAGU.CGGCGCCGAC........GCGCUCGCUCCGG.UCGAC
 Bdo-1-21    GG.ACGG.GACCG.GAG...CGCGCGAAC........GCGCGACCGCCGG.UCGAC
 Bsp-1-33    GC.AUGC.GGCCC.GAGC.CGGUGUCGAA........GCACCCCCUCGGG.CCGAC
 Bdo-1-22    GC.AUGGGCGGCU...GU.AAGCGCUAAG........GCGCUAACGUGGA.CGGAU
 env-75      GU.ACGG.GGCCG.GGAC.GGGCGCUGAA........GCGUUUGUUCCGG.UCGAU

             ........<<<<<.<<<<..<<<<<............>>>>>.>>>>>>>.>>...
             ........???2?.????..?221?............?122?.?????2?.??...
             GC-AUGG-GRYC•-RRGY-RRGCGCYRAA--------GCGCYRACYY•GR-YCGAC
```

```
²Bxe-1-35    GC.AUGG.GACCG.GGAC.GAGCGCCACAGCACCACAGCGCCGGCUCCGG.UCGAC
 Bxe-1-36    GC.AGGG.ACCAG.AGUA.AUGCGCCAAA........GCGCGA.ACUCUG.GCGAU
             ........<<<<<.<<<<..<<<<<............>>>>>.>>>>>>>.>>...
             ........???2?.????..?221?............?122?.?????2?.??...
             GC-AUGG-GRYC•-RRGY-RRGCGCYRAA--------GCGCYRACYY•GR-YCGAC
```

2.9 The mini-*ykkC* motif

2.9.1 Gene families associated with the *ykkC* motif.

The following conserved domains are found downstream of both the *ykkC* motif [1] and our new mini-*ykkC* motif:

- COG0511 Biotin carboxyl carrier protein [Lipid metabolism]
- COG0715 ABC-type nitrate/sulfonate/bicarbonate transport systems, periplasmic components [Inorganic ion transport and metabolism]
- COG2076 Membrane transporters of cations and cationic drugs [Inorganic ion transport and metabolism]
- COG3665 Uncharacterized conserved protein [Function unknown]
- COG4770 Acetyl/propionyl-CoA carboxylase, alpha subunit [Lipid metabolism]
- pfam00893 Small Multidrug Resistance protein.
- pfam01425 Amidase.
- pfam02626 Allophanate hydrolase subunit 2.
- pfam02682 Allophanate hydrolase subunit 1.
- Note: various cation transporters appear in *ykkC* and mini-*ykkC*, although their CDD assignments are different.

2.9.2 False positive rates and mini-*ykkC*

Because the mini-*ykkC* motif is small, false positives will inevitably be a problem in searches of large genome databases. A concern is that our mini-*ykkC* motif might be merely a subset of these false positives. The following results of automated refinement of the motif, however, argue against this concern. (Alignments discussed here appear in the following subsections.) CMfinder's initial alignment consisted of two unique sequences, both upstream of *emrE* (COG2076). After the RaveNnA homology search, CMfinder added one sequence, upstream of a gene containing allophanate hydrolase domains (pfam02626 and pfam02682). It was at this point that we saw the motif and decided to investigate it.

We ran another automated iteration of RaveNnA/CMfinder, and this resulted in four new sequences. Two were upstream of *emrE*, one was upstream of an allophanate hydrolase gene, and the remainder was upstream of a Ca^{2+}/H^+ antiporter (COG0387).

These data imply that mini-*ykkC* is a genuine motif for three reasons. Primarily, the seven sequences show considerable covariation in most of CMfinder's proposed base pairs, while conserving the ACGR loop motif. Second, the four gene families discussed above are common in our manual alignment of mini-*ykkC*. Finally, all but COG0387 are also found downstream of the *ykkC* motif, and COG0387 is a putative cation transporter, like many other *ykkC*-associated genes. Thus, while our manual alignment of mini-*ykkC* could contain an unknown number of false positives, the claim is secure that mini-*ykkC* is an RNA-like motif whose gene context resembles that of the *ykkC* motif. (Note: our manual alignment found four additional gene families typical of *ykkC*; see above.)

2.9.3 our manual mini-*ykkC* alignment, and gene context information

Notes:

[1] Annotated start codon overlaps predicted RNA, but BLAST finds no homology with overlapping ORF. Correct start codon is unknown.

[2] Annotated start codon lacked plausible Shine-Dalgarno, so a later start was selected. However, the earlier start did not overlap the RNA, so either start is consistent with our predictions.

[3] Downstream gene points wrong way, but sequence homology suggests that the RNA may be functional.

[4] RNA overlaps annotated gene. However, highly similar sequences annotate the homologous gene's start as being after the RNA.

[5] According to BLAST, downstream gene is homology both to COG2076 (Membrane transporters of cations and cationic drugs) and *sugE*.

[6] According to BLAST, downstream gene is homology both to allophanate hydrolase subunit 2 and urea amidolyase

[7] According to BLAST, downstream gene is homology to *sugE*.

[8] Predicted mini-ykkC overlaps $5'$ part of downstream gene, but BLAST finds no conservation of the overlapping coding region. We therefore selected a downstream start codon that is consistent with other mini-ykkC instances.

[9] Deviates from consensus, but upstream of canonical mini-ykkC gene.

[10] UCGA in loops violates consensus, but it is encouraging that both are the same. No detectable gene downstream, including by translated BLAST search; sequence fragment is truncated roughly 300 nucleotides after RNA.

[11] Overlaps $3'$ end of hypothetical gene.

[12] Overlaps $3'$ end of putative ATPase.

[13] *sugE* gene overlaps RNA. However, BLAST reports no homology for the first 15 codons, which places the conserved part within the RNA. The first AUG within the RNA that is in-frame with the predicted ORF is beyond the structured region, and seems a plausible start codon.

[14] Overlaps $3'$ end of putative ABC transporter.

[15] Somewhat dubious. UCGA in loops violates consensus, although it is encouraging that both are the same. Stems have more A-U pairs than normal. This predicted mini-ykkC instance is the only predicted mini-ykkC within the taxon Bacteroidetes/Chlorobi. BLAST finds no significant homology for the downstream hypothetical gene.

abbrev. of hits	**taxonomy of species**
Cef-1-1 to Cef-1-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Corynebacteriaceae *Corynebacterium efficiens* YS-314
Cgl-1-1 to Cgl-1-2	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Corynebacteriaceae *Corynebacterium glutamicum* ATCC 13032
Bli-1-1 to Bli-1-1	Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Brevibacteriaceae *Brevibacterium linens* BL2
Pgi-1-1 to Pgi-1-1	Bacteroidetes Bacteroidetes (class) Bacteroidales Porphyromonadaceae *Porphyromonas gingivalis* W83
Gvi-1-1 to Gvi-1-1	Cyanobacteria Gloeobacteria Gloeobacterales *Gloeobacter violaceus* PCC 7421
Ccr-1-1 to Ccr-1-1	α-proteobacteria Caulobacterales Caulobacteraceae *Caulobacter crescentus* CB15
Csp-1-1 to Csp-1-1	α-proteobacteria Caulobacterales Caulobacteraceae *Caulobacter sp.* K31
Asp-1-1 to Asp-1-1	α-proteobacteria Rhizobiales Aurantimonadaceae *Aurantimonas sp.* SI85-9A1
Bja-1-1 to Bja-1-1	α-proteobacteria Rhizobiales Bradyrhizobiaceae *Bradyrhizobium japonicum* USDA 110
Nsp-1-1 to Nsp-1-1	α-proteobacteria Rhizobiales Bradyrhizobiaceae *Nitrobacter sp.* Nb-311A
Nwi-1-1 to Nwi-1-1	α-proteobacteria Rhizobiales Bradyrhizobiaceae *Nitrobacter winogradskyi* Nb-255

Rpa-1-1 to Rpa-1-2	α-proteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* BisA53
Rpa-2-1 to Rpa-2-4	α-proteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* BisB18
Rpa-3-1 to Rpa-3-2	α-proteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* BisB5
Rpa-4-1 to Rpa-4-2	α-proteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* CGA009
Mlo-1-1 to Mlo-1-1	α-proteobacteria Rhizobiales Phyllobacteriaceae *Mesorhizobium loti* MAFF303099
Msp-1-1 to Msp-1-2	α-proteobacteria Rhizobiales Phyllobacteriaceae *Mesorhizobium sp.* BNC1
Atu-1-1 to Atu-1-2	α-proteobacteria Rhizobiales Rhizobiaceae Rhizobium/Agrobacterium group *Agrobacterium tumefaciens str.* C58
Ret-1-1 to Ret-1-2	α-proteobacteria Rhizobiales Rhizobiaceae Rhizobium/Agrobacterium group *Rhizobium etli* CFN 42
Sme-1-1 to Sme-1-1	α-proteobacteria Rhizobiales Rhizobiaceae Sinorhizobium/Ensifer group *Sinorhizobium meliloti* 1021
Xau-1-1 to Xau-1-2	α-proteobacteria Rhizobiales Xanthobacteraceae *Xanthobacter autotrophicus* Py2
Oal-1-1 to Oal-1-1	α-proteobacteria Rhodobacterales Hyphomonadaceae *Oceanicaulis alexandrii* HTCC2633
Jsp-1-1 to Jsp-1-1	α-proteobacteria Rhodobacterales Rhodobacteraceae *Jannaschia sp.* CCS1
Oba-1-1 to Oba-1-2	α-proteobacteria Rhodobacterales Rhodobacteraceae *Oceanicola batsensis* HTCC2597
Ogr-1-1 to Ogr-1-1	α-proteobacteria Rhodobacterales Rhodobacteraceae *Oceanicola granulosus* HTCC2516
Rnu-1-1 to Rnu-1-1	α-proteobacteria Rhodobacterales Rhodobacteraceae *Roseovarius nubinhibens* ISM
Rsp-1-1 to Rsp-1-1	α-proteobacteria Rhodobacterales Rhodobacteraceae *Roseovarius sp.* 217
Ssp-1-1 to Ssp-1-1	α-proteobacteria Rhodobacterales Rhodobacteraceae *Sulfitobacter sp.* EE-36
Ssp-2-1 to Ssp-2-1	α-proteobacteria Rhodobacterales Rhodobacteraceae *Sulfitobacter sp.* NAS-14.1
Rba-1-1 to Rba-1-1	α-proteobacteria *Rhodobacterales bacterium* HTCC2654
Acr-1-1 to Acr-1-1	α-proteobacteria Rhodospirillales Acetobacteraceae *Acidiphilium cryptum* JF-5
Mma-1-1 to Mma-1-1	α-proteobacteria Rhodospirillales Rhodospirillaceae *Magnetospirillum magneticum* AMB-1
Rru-1-1 to Rru-1-1	α-proteobacteria Rhodospirillales Rhodospirillaceae *Rhodospirillum rubrum* ATCC 11170
Bbr-1-1 to Bbr-1-1	β-proteobacteria Burkholderiales Alcaligenaceae *Bordetella bronchiseptica* RB50
Bpa-1-1 to Bpa-1-1	β-proteobacteria Burkholderiales Alcaligenaceae *Bordetella parapertussis* 12822
Bam-1-1 to Bam-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia ambifaria* AMMD
Bce-1-1 to Bce-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* AU 1054
Bce-2-1 to Bce-2-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* HI2424
Bce-3-1 to Bce-3-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* PC184
Bdo-1-1 to Bdo-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia dolosa* AUO158
Bsp-1-1 to Bsp-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia sp.* 383
Bvi-1-1 to Bvi-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia vietnamiensis* G4
Bma-1-1 to Bma-1-2	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* 10229
Bma-2-1 to Bma-2-2	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* 10399
Bma-3-1 to Bma-3-2	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* 2002721280
Bma-4-1 to Bma-4-2	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* ATCC 23344
Bma-5-1 to Bma-5-2	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* FMH
Bma-6-1 to Bma-6-2	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* GB8 horse 4
Bma-7-1 to Bma-7-2	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* JHU
Bma-8-1 to Bma-8-2	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* NCTC 10247
Bma-9-1 to Bma-9-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* SAVP1
Bxe-1-1 to Bxe-1-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia xenovorans* LB400
Bps-1-1 to Bps-1-2	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1106a
Bps-2-1 to Bps-2-2	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1106b
Bps-3-1 to Bps-3-2	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1655
Bps-4-1 to Bps-4-2	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1710a
Bps-5-1 to Bps-5-2	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 1710b
Bps-6-1 to Bps-6-2	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 406e
Bps-7-1 to Bps-7-2	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* 668
Bps-8-1 to Bps-8-2	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* K96243
Bps-9-1 to Bps-9-2	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* Pasteur
Bps-10-1 to Bps-10-2	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* S13
Bth-1-1 to Bth-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia thailandensis* E264
Reu-1-1 to Reu-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Cupriavidus *Ralstonia eutropha* JMP134
Pna-1-1 to Pna-1-1	β-proteobacteria Burkholderiales Comamonadaceae *Polaromonas naphthalenivorans* CJ2

Neu-1-1 to Neu-1-1	β-proteobacteria Nitrosomonadales Nitrosomonadaceae *Nitrosomonas europaea* ATCC 19718
Dar-1-1 to Dar-1-1	β-proteobacteria Rhodocyclales Rhodocyclaceae *Dechloromonas aromatica* RCB
Gsu-1-1 to Gsu-1-1	δ-proteobacteria Desulfuromonadales Geobacteraceae *Geobacter sulfurreducens* PCA
Wsu-1-1 to Wsu-1-1	ε-proteobacteria Campylobacterales Helicobacteraceae *Wolinella succinogenes* DSM 1740
Sde-1-1 to Sde-1-1	γ-proteobacteria Alteromonadales Alteromonadaceae *Saccharophagus degradans* 2-40
Pat-1-1 to Pat-1-1	γ-proteobacteria Alteromonadales Pseudoalteromonadaceae *Pseudoalteromonas atlantica* T6c
Pin-1-1 to Pin-1-1	γ-proteobacteria Alteromonadales Psychromonadaceae *Psychromonas ingrahamii* 37
Sba-1-1 to Sba-1-1	γ-proteobacteria Alteromonadales Shewanellaceae *Shewanella baltica* OS155
Sba-2-1 to Sba-2-1	γ-proteobacteria Alteromonadales Shewanellaceae *Shewanella baltica* OS195
Sfr-1-1 to Sfr-1-1	γ-proteobacteria Alteromonadales Shewanellaceae *Shewanella frigidimarina* NCIMB 400
Son-1-1 to Son-1-1	γ-proteobacteria Alteromonadales Shewanellaceae *Shewanella oneidensis* MR-1
Spu-1-1 to Spu-1-1	γ-proteobacteria Alteromonadales Shewanellaceae *Shewanella putrefaciens* CN-32
Ssp-3-1 to Ssp-3-1	γ-proteobacteria Alteromonadales Shewanellaceae *Shewanella sp.* ANA-3
Ssp-4-1 to Ssp-4-2	γ-proteobacteria Alteromonadales Shewanellaceae *Shewanella sp.* MR-4
Ssp-5-1 to Ssp-5-1	γ-proteobacteria Alteromonadales Shewanellaceae *Shewanella sp.* MR-7
Ssp-6-1 to Ssp-6-1	γ-proteobacteria Alteromonadales Shewanellaceae *Shewanella sp.* W3-18-1
Aeh-1-1 to Aeh-1-1	γ-proteobacteria Chromatiales Ectothiorhodospiraceae *Alkalilimnicola ehrlichei* MLHE-1
Eco-1-1 to Eco-1-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* 101-1
Eco-2-1 to Eco-2-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* 536
Eco-3-1 to Eco-3-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* 53638
Eco-4-1 to Eco-4-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* B171
Eco-5-1 to Eco-5-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* B7A
Eco-6-1 to Eco-6-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* CFT073
Eco-7-1 to Eco-7-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* E110019
Eco-8-1 to Eco-8-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* E22
Eco-9-1 to Eco-9-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* E24377A
Eco-10-1 to Eco-10-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* F11
Eco-11-1 to Eco-11-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* HS
Eco-12-1 to Eco-12-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* K12
Eco-13-1 to Eco-13-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 EDL933
Eco-14-1 to Eco-14-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* O157:H7 str. Sakai
Eco-15-1 to Eco-15-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* UTI89
Eco-16-1 to Eco-16-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Escherichia coli* W3110
Eca-1-1 to Eca-1-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae Pectobacterium *Erwinia carotovora subsp. atroseptica* SCRI1043
Plu-1-1 to Plu-1-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Photorhabdus luminescens subsp. laumondii* TTO1
Sen-1-1 to Sen-1-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Salmonella enterica subsp. enterica serovar* Paratyphi A str. ATCC
Sen-3-1 to Sen-3-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Salmonella enterica subsp. enterica serovar* Typhi str. CT18
Sen-2-1 to Sen-2-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Salmonella enterica subsp. enterica serovar* Typhi Ty2
Sty-1-1 to Sty-1-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Salmonella typhimurium* LT2
Sbo-1-1 to Sbo-1-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Shigella boydii* BS512
Sbo-2-1 to Sbo-2-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Shigella boydii* Sb227
Sdy-1-1 to Sdy-1-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Shigella dysenteriae* Sd197
Sfl-1-1 to Sfl-1-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Shigella flexneri* 2a str. 2457T
Sfl-2-1 to Sfl-2-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Shigella flexneri* 2a str. 301
Sfl-3-1 to Sfl-3-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Shigella flexneri* 5 str. 8401
Sgl-1-1 to Sgl-1-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Sodalis glossinidius str. ’morsitans’*
Ybe-1-1 to Ybe-1-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Yersinia bercovieri* ATCC 43970
Yfr-1-1 to Yfr-1-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Yersinia frederiksenii* ATCC 33641
Yin-1-1 to Yin-1-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Yersinia intermedia* ATCC 29909
Ype-1-1 to Ype-1-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* Angola
Ype-2-1 to Ype-2-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* Antiqua
Ype-6-1 to Ype-6-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Microtus str. 91001
Ype-7-1 to Ype-7-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis biovar* Orientalis str. IP275
Ype-3-1 to Ype-3-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* CO92

Ype-4-1 to Ype-4-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* KIM
Ype-5-1 to Ype-5-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pestis* Nepal516
Yps-1-1 to Yps-1-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pseudotuberculosis* IP 31758
Yps-2-1 to Yps-2-1	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Yersinia pseudotuberculosis* IP 32953
Abo-1-1 to Abo-1-1	γ-proteobacteria Oceanospirillales Alcanivoracaceae *Alcanivorax borkumensis* SK2
Msp-2-1 to Msp-2-1	γ-proteobacteria Oceanospirillales *Marinomonas sp.* MED121
Osp-1-1 to Osp-1-1	γ-proteobacteria Oceanospirillales *Oceanobacter sp.* RED65
Asp-2-1 to Asp-2-1	γ-proteobacteria Pseudomonadales Moraxellaceae *Acinetobacter sp.* ADP1
Pae-1-1 to Pae-1-3	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* 2192
Pae-2-1 to Pae-2-2	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* C3719
Pae-3-1 to Pae-3-2	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PA7
Pae-4-1 to Pae-4-3	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PACS2
Pae-5-1 to Pae-5-3	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PAO1
Pae-6-1 to Pae-6-3	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* UCBPP-PA14
Pen-1-1 to Pen-1-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas entomophila* L48
Pfl-1-1 to Pfl-1-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf-5
Pfl-2-1 to Pfl-2-2	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* PfO-1
Ppu-1-1 to Ppu-1-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* F1
Ppu-2-1 to Ppu-2-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* KT2440
Psy-1-1 to Psy-1-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. phaseolicola* 1448A
Psy-2-1 to Psy-2-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. syringae* B728a
Psy-3-1 to Psy-3-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. tomato str.* DC3000
Val-1-1 to Val-1-1	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio alginolyticus* 12G01
Vch-1-1 to Vch-1-1	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio cholerae* MO10
Vch-2-1 to Vch-2-1	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio cholerae* O1 biovar eltor str. N16961
Vch-3-1 to Vch-3-1	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio cholerae* O395
Vch-4-1 to Vch-4-1	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio cholerae* RC385
Vch-5-1 to Vch-5-1	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio cholerae* V51
Vch-6-1 to Vch-6-1	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio cholerae* V52
Vpa-1-1 to Vpa-1-1	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio parahaemolyticus* RIMD 2210633
Vsp-1-1 to Vsp-1-1	γ-proteobacteria Vibrionales Vibrionaceae *Vibrio sp.* Ex25
Xca-1-1 to Xca-1-1	γ-proteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas campestris pv. campestris str.* 8004
Xca-2-1 to Xca-2-1	γ-proteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas campestris pv. campestris str.* ATCC 33913
Xor-1-1 to Xor-1-1	γ-proteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas oryzae pv. oryzae* KACC10331
Xor-2-1 to Xor-2-1	γ-proteobacteria Xanthomonadales Xanthomonadaceae *Xanthomonas oryzae pv. oryzae* MAFF 311018
Sen-4-1 to Sen-4-2	γ-proteobacteria Enterobacteriales Enterobacteriaceae *Salmonella enterica subsp. enterica serovar* Choleraesuis str
env-1 to env-10	environmental samples

abbrev	**RefSeq accession**		**5′ at**	**3′ at**	**genes**
Osp-1-1	NZ_AAQH01000002.1	-	112825	112760	**RNA**→ COG4770 (COG4770)AHS2 (pfam02626)AHS1 (pfam02682)AccB (COG0511)→ Amidase (pfam01425)→
Pen-1-1	NC_008027.1	-	1469847	1469761	**RNA**→ EmrE (COG2076)→
Ppu-1-1	NZ_AALM01000051.1	+	20063	20150	**RNA**→ EmrE (COG2076)→
Ppu-2-1	NC_002947.3	-	1897151	1897065	**RNA**→ EmrE (COG2076)→
Pae-5-1	NC_002516.2	+	3653053	3653143	**RNA**→ EmrE (COG2076)→
Pae-6-1	NZ_AABQ07000001.1	+	953543	953633	**RNA**→ EmrE (COG2076)→
Pae-2-1	NZ_AAKV01000043.1	+	2119	2209	**RNA**→ EmrE (COG2076)→
Pae-1-1	NZ_AAKW01000028.1	+	300	390	**RNA**→ EmrE (COG2076)→
Pae-4-1	NZ_AAQW01000001.1	+	4351419	4351509	**RNA**→ EmrE (COG2076)→
Pae-3-1	NZ_AAQE01000062.1	-	30463	30376	**RNA**→ EmrE (COG2076)→

Oba-1-1	NZ_AAMO01000001.1	+	123399	123479	**RNA**→ BTP (pfam05232)→ 5-FTHF_cyc-lig (pfam01812)→
Oba-1-2	NZ_AAMO01000002.1	+	61290	61370	**RNA**→ BTP (pfam05232)→ 5-FTHF_cyc-lig (pfam01812)→
Sba-2-1	NZ_AATK01000011.1	+	87254	87327	**RNA**→ EmrE (COG2076)→
Sba-1-1	NZ_AAIO01000002.1	+	35270	35343	**RNA**→ EmrE (COG2076)→
Ssp-5-1	NC_008322.1	+	1984998	1985070	**RNA**→ EmrE (COG2076)→
Spu-1-1	NZ_AALB01000014.1	-	56724	56651	**RNA**→ EmrE (COG2076)→
Ssp-6-1	NZ_AALN01000010.1	+	35498	35571	**RNA**→ EmrE (COG2076)→
Ssp-4-1	NC_008321.1	+	1903956	1904032	**RNA**→ EmrE (COG2076)→
Ssp-3-1	NZ_AALH01000010.1	+	34502	34578	**RNA**→ EmrE (COG2076)→
Ssp-4-2	NZ_AALX01000008.1	+	108108	108184	**RNA**→ EmrE (COG2076)→
env-1	AACY01003452.1	+	246	322	**RNA**→ unknown→
env-2	AACY01010645.1	-	30763	30687	**RNA**→ unknown→
env-3	AACY01051730.1	-	47405	47329	**RNA**→ unknown→
Son-1-1	NC_004347.1	+	2001423	2001499	**RNA**→ EmrE (COG2076)→
Yps-2-1	NC_006155.1	+	487916	487993	**RNA**→ Multi_Drug_Res (pfam00893)→
Ype-2-1	NC_008150.1	-	4425181	4425104	**RNA**→ Multi_Drug_Res (pfam00893)→
Ype-5-1	NC_008149.1	-	3779292	3779215	**RNA**→ Multi_Drug_Res (pfam00893)→
Ype-4-1	NC_004088.1	+	692489	692566	**RNA**→ Multi_Drug_Res (pfam00893)→
Ype-3-1	NC_003143.1	+	366341	366418	**RNA**→ Multi_Drug_Res (pfam00893)→
Ype-6-1	NC_005810.1	+	547367	547444	**RNA**→ Multi_Drug_Res (pfam00893)→
Ype-1-1	NZ_AAKS01000043.1	+	17455	17532	**RNA**→ Multi_Drug_Res (pfam00893)→
Yps-1-1	NZ_AAKT01000001.1	+	3458104	3458181	**RNA**→ Multi_Drug_Res (pfam00893)→
Ype-7-1	NZ_AAOS01000048.1	-	15639	15562	**RNA**→ Multi_Drug_Res (pfam00893)→
Ybe-1-1	NZ_AALC01000031.1	+	22429	22506	**RNA**→ Multi_Drug_Res (pfam00893)→
Yin-1-1	NZ_AALF01000043.1	+	25328	25408	**RNA**→ Multi_Drug_Res (pfam00893)→
[2]Yfr-1-1	NZ_AALE01000017.1	+	76722	76802	**RNA**→ Multi_Drug_Res (pfam00893)→
Bps-8-1	NC_006350.1	+	1359248	1359329	**RNA**→ EmrE (COG2076)→
Bma-4-1	NC_006348.1	-	1960227	1960146	**RNA**→ EmrE (COG2076)→
Bma-1-1	NZ_AAHM02000001.1	-	797739	797658	**RNA**→ EmrE (COG2076)→
Bma-2-1	NZ_AAHN02000015.1	+	51107	51188	**RNA**→ EmrE (COG2076)→
Bma-6-1	NZ_AAHO01000018.1	-	71928	71847	**RNA**→ EmrE (COG2076)→
Bma-8-1	NZ_AAHP01000018.1	-	73502	73421	**RNA**→ EmrE (COG2076)→
Bma-9-1	NZ_AAHQ02000001.1	-	2089171	2089090	**RNA**→ EmrE (COG2076)→
Bps-3-1	NZ_AAHR02000067.1	-	3929	3848	**RNA**→ EmrE (COG2076)→
Bps-7-1	NZ_AAHU01000020.1	+	51134	51215	**RNA**→ EmrE (COG2076)→
Bps-9-1	NZ_AAHV02000030.1	-	72204	72123	**RNA**→ EmrE (COG2076)→
Bma-5-1	NZ_AAIQ02000015.1	-	74739	74658	**RNA**→ EmrE (COG2076)→
Bma-7-1	NZ_AAIR02000052.1	-	20120	20039	**RNA**→ EmrE (COG2076)→
Bma-3-1	NZ_AANX02000037.1	-	6000	5919	**RNA**→ EmrE (COG2076)→
Bps-5-1	NC_007434.1	+	1472149	1472230	**RNA**→ EmrE (COG2076)→
Bps-4-1	NZ_AAHS02000030.1	+	14417	14498	**RNA**→ EmrE (COG2076)→
Bps-10-1	NZ_AAHW02000015.1	+	51289	51370	**RNA**→ EmrE (COG2076)→
Bps-1-1	NZ_AAMA01000091.1	+	14429	14510	**RNA**→ EmrE (COG2076)→

Bps-2-1	NZ_AAMB01000018.1	-	71919	71838	**RNA**→ EmrE (COG2076)→
Bps-6-1	NZ_AAMM02000027.1	+	1635	1716	**RNA**→ EmrE (COG2076)→
Bth-1-1	NC_007651.1	+	1166403	1166484	**RNA**→ EmrE (COG2076)→
Bce-1-1	NC_008060.1	-	1851717	1851639	**RNA**→ EmrE (COG2076)→
Bce-2-1	NZ_AAHL01000002.1	+	215360	215438	**RNA**→ EmrE (COG2076)→
Bce-3-1	NZ_AAKX01000080.1	+	42964	43042	**RNA**→ EmrE (COG2076)→
[4]Bvi-1-1	NZ_AAEH02000013.1	+	160299	160377	**RNA**→ EmrE (COG2076)→
[7]env-4	AACY01057793.1	+	107368	107445	**RNA**→ unknown→
[4]Bsp-1-1	NC_007510.1	-	2708132	2708055	**RNA**→ EmrE (COG2076)→
Bam-1-1	NZ_AAJL01000016.1	-	32654	32577	**RNA**→ EmrE (COG2076)→
Bdo-1-1	NZ_AAKY01000050.1	+	24768	24845	**RNA**→ EmrE (COG2076)→
[5]env-5	AACY01023341.1	+	1360	1436	**RNA**→ unknown→
Bxe-1-1	NC_007951.1	+	1349400	1349474	**RNA**→ EmrE (COG2076)→
Vch-2-1	NC_002505.1	+	1484752	1484829	**RNA**→ EmrE (COG2076)→
Vch-1-1	NZ_AAKF01000013.1	-	39410	39333	**RNA**→ EmrE (COG2076)→
Vch-3-1	NZ_AAKG01000001.1	+	1043220	1043297	**RNA**→ EmrE (COG2076)→
Vch-5-1	NZ_AAKI01000006.1	+	34875	34952	**RNA**→ EmrE (COG2076)→
Vch-6-1	NZ_AAKJ01000010.1	+	11362	11439	**RNA**→ EmrE (COG2076)→
Vch-4-1	NZ_AAKH01000007.1	-	25723	25646	**RNA**→ EmrE (COG2076)→
Val-1-1	NZ_AAPS01000011.1	-	72329	72250	**RNA**→ EmrE (COG2076)→
Vpa-1-1	NC_004605.1	+	787489	787565	**RNA**→ EmrE (COG2076)→
Vsp-1-1	NZ_AAKK01000007.1	-	80612	80537	**RNA**→ EmrE (COG2076)→
Pat-1-1	NC_008228.1	-	1686458	1686380	**RNA**→ EmrE (COG2076)→
Sen-1-1	NC_006511.1	+	4331168	4331244	**RNA**→ EmrE (COG2076)→
Sen-4-1	NC_006905.1	+	4516129	4516205	**RNA**→ EmrE (COG2076)→
Sty-1-1	NC_003197.1	+	4581429	4581505	**RNA**→ EmrE (COG2076)→
Sen-2-1	NC_004631.1	+	4540824	4540900	**RNA**→ EmrE (COG2076)→
Sen-3-1	NC_003198.1	+	4557922	4557998	**RNA**→ EmrE (COG2076)→
Eco-16-1	AC_000091.1	+	4381485	4381560	**RNA**→ EmrE (COG2076)→
Eco-12-1	NC_000913.2	+	4374828	4374903	**RNA**→ EmrE (COG2076)→
Eco-5-1	NZ_AAJT01000005.1	-	59704	59629	**RNA**→ EmrE (COG2076)→
Eco-8-1	NZ_AAJV01000034.1	+	41002	41077	**RNA**→ EmrE (COG2076)→
Eco-7-1	NZ_AAJW01000004.1	+	193907	193982	**RNA**→ EmrE (COG2076)→
Eco-4-1	NZ_AAJX01000024.1	+	53773	53848	**RNA**→ EmrE (COG2076)→
Eco-11-1	NZ_AAJY01000001.1	+	4402101	4402176	**RNA**→ EmrE (COG2076)→
Eco-9-1	NZ_AAJZ01000001.1	+	2492556	2492631	**RNA**→ EmrE (COG2076)→
Eco-1-1	NZ_AAMK01000006.1	+	194583	194658	**RNA**→ EmrE (COG2076)→
[4]Sdy-1-1	NC_007606.1	+	4110745	4110820	**RNA**→ EmrE (COG2076)→
Sbo-2-1	NC_007613.1	-	4357286	4357211	**RNA**→ EmrE (COG2076)→
Eco-14-1	NC_002695.1	+	5231086	5231161	**RNA**→ EmrE (COG2076)→
Eco-13-1	NC_002655.2	+	5261071	5261146	**RNA**→ EmrE (COG2076)→
Sbo-1-1	NZ_AAKA01000002.1	-	207042	206967	**RNA**→ EmrE (COG2076)→
Eco-3-1	NZ_AAKB01000006.1	+	199084	199159	**RNA**→ EmrE (COG2076)→

[4]Eco-2-1	NC_008253.1	+	4606978	4607053	**RNA**→ EmrE (COG2076)→
Eco-15-1	NC_007946.1	+	4648604	4648679	**RNA**→ EmrE (COG2076)→
Eco-6-1	NC_004431.1	+	4985526	4985601	**RNA**→ EmrE (COG2076)→
Eco-10-1	NZ_AAJU01000026.1	+	56626	56701	**RNA**→ EmrE (COG2076)→
[4]Sfl-3-1	NC_008258.1	+	4449786	4449861	**RNA**→ EmrE (COG2076)→
Sfl-2-1	NC_004337.1	+	4481192	4481267	**RNA**→ EmrE (COG2076)→
Sfl-1-1	NC_004741.1	+	4473385	4473460	**RNA**→ EmrE (COG2076)→
[4]Sen-4-2	NC_006856.1	-	72205	72131	**RNA**→ EmrE (COG2076)→
Gvi-1-1	NC_005125.1	-	3187387	3187310	**RNA**→ EmrE (COG2076)→
Msp-1-1	NC_008254.1	+	1077816	1077892	**RNA**→ EmrE (COG2076)→
Msp-1-2	NZ_AAED02000011.1	+	121362	121438	**RNA**→ EmrE (COG2076)→
[5]env-6	AACY01084250.1	+	426	503	**RNA**→ unknown→
[2]Nwi-1-1	NC_007406.1	+	2219773	2219857	**RNA**→ EmrE (COG2076)→
[8]Bja-1-1	NC_004463.1	-	4473831	4473748	**RNA**→ EmrE (COG2076)→
Nsp-1-1	NZ_AAMY01000019.1	-	9610	9522	**RNA**→ EmrE (COG2076)→
Gsu-1-1	NC_002939.4	-	755450	755372	**RNA**→ EmrE (COG2076)→ UhpC (COG2271)→
Plu-1-1	NC_005126.1	-	4818299	4818218	**RNA**→ EmrE (COG2076)→
Rru-1-1	NC_007643.1	+	336932	337015	**RNA**→ Small multidrug resistance protein→
Rpa-4-1	NC_005296.1	-	1558599	1558516	**RNA**→ COG3665 (COG3665)→ COG3665 (COG3665)→ AccB (COG0511)COG4770 (COG4770)AHS1 (pfam02682)→ Amidase (pfam01425)→
Rpa-1-1	NZ_AALA01000001.1	-	169823	169739	**RNA**→ COG3665 (COG3665)→ COG3665 (COG3665)→ COG4770 (COG4770)AHS1 (pfam02682)AccB (COG0511)→ Amidase (pfam01425)→
Rpa-3-1	NC_007958.1	+	2440537	2440621	**RNA**→ COG3665 (COG3665)→ COG3665 (COG3665)→ COG4770 (COG4770)AHS1 (pfam02682)AccB (COG0511)→
Rpa-2-1	NC_007925.1	-	5064915	5064828	**RNA**→ COG3665 (COG3665)→ COG3665 (COG3665)→ AccB (COG0511)COG4770 (COG4770)AHS1 (pfam02682)→
Rpa-2-2	NZ_AALR01000001.1	-	61324	61237	**RNA**→ COG3665 (COG3665)→ COG3665 (COG3665)→ AccB (COG0511)COG4770 (COG4770)AHS1 (pfam02682)→
Pae-5-2	NC_002516.2	+	2052871	2052946	**RNA**→ EmrE (COG2076)→ NuoA (COG0838)→
Pae-6-2	NZ_AABQ07000002.1	+	1373619	1373694	**RNA**→ EmrE (COG2076)→ NuoA (COG0838)→
Pae-2-2	NZ_AAKV01000014.1	+	117148	117223	**RNA**→ EmrE (COG2076)→ NuoA (COG0838)→
Pae-1-2	NZ_AAKW01000013.1	+	5552	5627	**RNA**→ EmrE (COG2076)→ NuoA (COG0838)→
Pae-4-2	NZ_AAQW01000001.1	+	2613925	2614000	**RNA**→ EmrE (COG2076)→ NuoA (COG0838)→
Pae-3-2	NZ_AAQE01000010.1	+	122283	122358	**RNA**→ EmrE (COG2076)→ NuoA (COG0838)→
Asp-1-1	NZ_AAPJ01000003.1	-	422926	422845	**RNA**→ BTP (pfam05232)→
Rsp-1-1	NZ_AAMV01000001.1	+	957202	957275	**RNA**→ BTP (pfam05232)→ hypo→
Rba-1-1	NZ_AAMT01000031.1	-	4129	4057	**RNA**→ BTP (pfam05232)→
Rpa-2-3	NC_007925.1	+	3222811	3222888	**RNA**→ EmrE (COG2076)→
Rpa-2-4	NZ_AALR01000002.1	-	59822	59745	**RNA**→ EmrE (COG2076)→
Eca-1-1	NC_004547.2	+	2171193	2171271	**RNA**→ AccB (COG0511)COG4770 (COG4770)AHS2 (pfam02626)AHS1 (pfam02682)→
[6]env-7	AACY01195930.1	+	484	563	**RNA**→ unknown→
[6]env-8	AACY01274299.1	-	294	212	**RNA**→ unknown→
Pfl-2-1	NC_007492.1	-	1515939	1515859	**RNA**→ AccB (COG0511)COG4770 (COG4770)AHS2 (pfam02626)AHS1 (pfam02682)→ Amidase (pfam01425)→
Pna-1-1	NZ_AANM01000030.1	+	36824	36895	**RNA**→ COG4770 (COG4770)AHS2 (pfam02626)→ AccB (COG0511)AHS2 (pfam02626)AHS1 (pfam02682)→ Amidase (pfam01425)→

Reu-1-1	NC_007348.1	-	1606553	1606471	**RNA**→ AccB (COG0511)COG4770 (COG4770)AHS2 (pfam02626)AHS1 (pfam02682)→ Amidase (pfam01425)→
Sgl-1-1	NC_007712.1	+	2850584	2850658	**RNA**→ EmrE (COG2076)→
[13]Mma-1-1	NC_007626.1	+	1410471	1410551	**RNA**→ EmrE (COG2076)→
Sde-1-1	NC_007912.1	+	1459868	1459944	**RNA**→ AccB (COG0511)COG4770 (COG4770)AHS1 (pfam02682)→ Amidase (pfam01425)→
Msp-2-1	NZ_AANE01000002.1	+	193996	194070	**RNA**→ AccB (COG0511)COG4770 (COG4770)AHS2 (pfam02626)AHS1 (pfam02682)→ Amidase (pfam01425)→ hypo→
Ssp-1-1	NZ_AALV01000005.1	-	80416	80337	**RNA**→ BTP (pfam05232)→
Ssp-2-1	NZ_AALZ01000005.1	-	234115	234036	**RNA**→ BTP (pfam05232)→
Oal-1-1	NZ_AAMQ01000006.1	-	81077	81001	**RNA**→ EmrE (COG2076)→
Jsp-1-1	NC_007802.1	-	1252878	1252804	**RNA**→ EmrE (COG2076)→ cysteine_hydrolases (cd00431)→
Dar-1-1	NC_007298.1	+	88470	88546	**RNA**→ AccB (COG0511)COG4770 (COG4770)AHS2 (pfam02626)AHS1 (pfam02682)→ Amidase (pfam01425)→
Atu-1-1	NC_003304.1	+	1627014	1627094	**RNA**→ BTP (pfam05232)→
Atu-1-2	NC_003062.1	+	1626901	1626981	**RNA**→ BTP (pfam05232)→
Pfl-1-1	NC_004129.6	+	5115326	5115409	**RNA**→ EmrE (COG2076)→
Bbr-1-1	NC_002927.3	+	862218	862288	**RNA**→ ChaA (COG0387)→
Bpa-1-1	NC_002928.3	+	777303	777373	**RNA**→ COG5361 (COG5361)→
Ccr-1-1	NC_002696.2	-	3687061	3686980	**RNA**→ EmrE (COG2076)→
[12]Xau-1-1	NZ_AAPC01000001.1	+	920657	920745	**RNA**→ COG3665 (COG3665)→ COG3665 (COG3665)→ AccB (COG0511)COG4770 (COG4770)AHS1 (pfam02682)→ probable urea amidolyase→
Pfl-2-2	NC_007492.1	-	1858455	1858372	**RNA**→ EmrE (COG2076)→
Xau-1-2	NZ_AAPC01000001.1	+	37444	37526	**RNA**→ EmrE (COG2076)→ KdsA (COG2877)→
[14]Ret-1-1	NC_007765.1	-	143278	143199	**RNA**→ COG3665 (COG3665)→ COG3665 (COG3665)→ AccB (COG0511)COG4770 (COG4770)AHS1 (pfam02682)→ Amidase (pfam01425)→
[2]Cef-1-1	NC_004369.1	+	1194192	1194276	**RNA**→ ChaA (COG0387)→ COG2120 (COG2120)→
Bli-1-1	NZ_AAGP01000042.1	-	24816	24739	**RNA**→ ChaA (COG0387)→
[9]Asp-2-1	NC_005966.1	+	680725	680803	**RNA**→ EmrE (COG2076)→ Ada (COG0350)→
Neu-1-1	NC_004757.1	-	2389363	2389281	**RNA**→ EmrE (COG2076)→
Ogr-1-1	NZ_AAOT01000014.1	+	63921	64008	**RNA**→ BTP (pfam05232)→
Ret-1-2	NC_007761.1	+	2332738	2332822	**RNA**→ EmrE (COG2076)→
Sme-1-1	NC_003047.1	+	1498546	1498630	**RNA**→ EmrE (COG2076)→
Rnu-1-1	NZ_AALY01000001.1	-	1481671	1481584	**RNA**→ BTP (pfam05232)→
Abo-1-1	NC_008260.1	+	2165340	2165437	**RNA**→ COG4770 (COG4770)AHS2 (pfam02626)AHS1 (pfam02682)AccB (COG0511)→
Wsu-1-1	NC_005090.1	+	1071422	1071507	**RNA**→ AccB (COG0511)COG4770 (COG4770)AHS2 (pfam02626)AHS1 (pfam02682)→ RarD (COG2962)→ Amidase (pfam01425)→ ALLOPHANATE HYDROLASE→ HTH_ARSR (smart00418)→ ZntA (COG2217)→
[11]Cgl-1-1	NC_003450.3	+	1129574	1129660	**RNA**→ ChaA (COG0387)→ COG2120 (COG2120)→
Cgl-1-2	NC_006958.1	+	1131042	1131128	**RNA**→ ChaA (COG0387)→ COG2120 (COG2120)→
Pae-5-3	NC_002516.2	-	1934936	1934852	**RNA**→ TauA (COG0715)→ AmiR (COG3707)→
Pae-4-3	NZ_AAQW01000001.1	-	2495257	2495173	**RNA**→ TauA (COG0715)→ AmiR (COG3707)→
Pae-6-3	NZ_AABQ07000002.1	-	1253426	1253342	**RNA**→ TauA (COG0715)→ AmiR (COG3707)→
Pae-1-3	NZ_AAKW01000008.1	-	176158	176074	**RNA**→ TauA (COG0715)→ AmiR (COG3707)→
Psy-3-1	NC_004578.1	+	4775587	4775690	**RNA**→ COG4770 (COG4770)AHS2 (pfam02626)AHS1 (pfam02682)AccB (COG0511)→ Amidase (pfam01425)→
Psy-2-1	NC_007005.1	+	4723995	4724098	**RNA**→ COG4770 (COG4770)AHS2 (pfam02626)AHS1 (pfam02682)AccB (COG0511)→ Amidase (pfam01425)→
Psy-1-1	NC_005773.3	+	4535817	4535920	**RNA**→ AccB (COG0511)COG4770 (COG4770)AHS2 (pfam02626)AHS1 (pfam02682)→ Amidase (pfam01425)→ Transposase_mut (pfam00872)→
[3]env-9	AACY01014706.1	+	332	411	**RNA**→

^{1}Bps-8-2	NC_006351.1	+	858880	858963	**RNA**→ OmpR (COG0745)→
Bps-5-2	NC_007435.1	+	2678630	2678713	**RNA**→ OmpR (COG0745)→
Bma-4-2	NC_006349.2	+	842068	842151	**RNA**→ OmpR (COG0745)→
Bma-1-2	NZ_AAHM02000002.1	+	25517	25600	**RNA**→ OmpR (COG0745)→
Bma-2-2	NZ_AAHN02000040.1	-	9516	9433	**RNA**→ OmpR (COG0745)→
Bma-6-2	NZ_AAHO01000096.1	+	9249	9332	**RNA**→ OmpR (COG0745)→
Bma-8-2	NZ_AAHP01000048.1	-	9408	9325	**RNA**→ OmpR (COG0745)→
Bps-3-2	NZ_AAHR02000013.1	+	121483	121566	**RNA**→ OmpR (COG0745)→
Bps-4-2	NZ_AAHS02000001.1	+	354812	354895	**RNA**→ OmpR (COG0745)→
Bps-7-2	NZ_AAHU01000026.1	+	90048	90131	**RNA**→ OmpR (COG0745)→
Bps-9-2	NZ_AAHV02000104.1	-	9383	9300	**RNA**→ OmpR (COG0745)→
Bps-10-2	NZ_AAHW02000030.1	-	9353	9270	**RNA**→ OmpR (COG0745)→
Bma-5-2	NZ_AAIQ02000066.1	+	16564	16647	**RNA**→ OmpR (COG0745)→
Bma-7-2	NZ_AAIR02000049.1	+	28076	28159	**RNA**→ OmpR (COG0745)→
Bps-1-2	NZ_AAMA01000020.1	-	9570	9487	**RNA**→ OmpR (COG0745)→
Bps-2-2	NZ_AAMB01000024.1	-	9345	9262	**RNA**→ OmpR (COG0745)→
Bps-6-2	NZ_AAMM02000026.1	+	58505	58588	**RNA**→ OmpR (COG0745)→
Bma-3-2	NZ_AANX02000032.1	-	9370	9287	**RNA**→ OmpR (COG0745)→
^{10}env-10	AACY01772680.1	+	233	313	**RNA**→
Xca-1-1	NC_007086.1	+	577384	577505	**RNA**→ EmrE (COG2076)→
Xca-2-1	NC_003902.1	+	575294	575415	**RNA**→ EmrE (COG2076)→
^{1}Xor-1-1	NC_006834.1	-	4433364	4433279	**RNA**→ EmrE (COG2076)→
Xor-2-1	NC_007705.1	-	4430838	4430753	**RNA**→ EmrE (COG2076)→
Rpa-4-2	NC_005296.1	-	2586491	2586351	**RNA**→ EmrE (COG2076)→ Plasmid_killer (pfam05015)→ VapI (COG3093)→
Rpa-3-2	NC_007958.1	-	2608528	2608410	**RNA**→ EmrE (COG2076)→
Rpa-1-2	NZ_AALA01000013.1	+	14849	14990	**RNA**→ EmrE (COG2076)→
^{1}Sfr-1-1	NZ_AAIV01000010.1	+	122744	122845	**RNA**→ hypo→
Mlo-1-1	NC_002678.2	+	3864262	3864387	**RNA**→ EmrE (COG2076)→
Csp-1-1	NZ_AATH01000001.1	+	336429	336571	**RNA**→ EmrE (COG2076)→
Aeh-1-1	NZ_AALK01000006.1	-	55971	55807	**RNA**→ Allophanate hydrolase→
^{12}Acr-1-1	NZ_AAOO01000022.1	+	8563	8750	**RNA**→ COG3665 (COG3665)→ COG3665 (COG3665)→ COG4770 (COG4770)AHS1 (pfam02682)AccB (COG0511)→ Amidase (pfam01425)→
Pin-1-1	NZ_AAQS01000011.1	+	97253	97514	**RNA**→ Na_H_antiport_1 (pfam06965)→
^{15}Pgi-1-1	NC_002950.2	+	1541841	1542127	**RNA**→ hypo→ STE14 (COG2020)→ UbiE (COG2226)→

cd00431 Cysteine hydrolases; This family contains amidohydrolases, like CSHase (N-carbamoylsarcosine amidohydrolase), involved in creatine metabolism and nicotinamidase, converting nicotinamide to nicotinic acid and ammonia in the pyridine nucleotide cycle.

COG0350 Methylated DNA-protein cysteine methyltransferase [DNA replication, recombination, and repair]

COG0387 Ca2+/H+ antiporter [Inorganic ion transport and metabolism]

COG0511 Biotin carboxyl carrier protein [Lipid metabolism]

COG0715 ABC-type nitrate/sulfonate/bicarbonate transport systems, periplasmic components [Inorganic ion transport and metabolism]

COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain [Signal transduction mechanisms / Transcription]

COG0838 NADH:ubiquinone oxidoreductase subunit 3 (chain A) [Energy production and conversion]

COG2020 Putative protein-S-isoprenylcysteine methyltransferase [Posttranslational modification, protein turnover, chaperones]

COG2076 Membrane transporters of cations and cationic drugs [Inorganic ion transport and metabolism]

COG2120 Uncharacterized proteins, LmbE homologs [Function unknown]

COG2217 Cation transport ATPase [Inorganic ion transport and metabolism]

COG2226 Methylase involved in ubiquinone/menaquinone biosynthesis [Coenzyme metabolism]

COG2271 Sugar phosphate permease [Carbohydrate transport and metabolism]

COG2877 3-deoxy-D-manno-octulosonic acid (KDO) 8-phosphate synthase [Cell envelope biogenesis, outer membrane]

COG2962 Predicted permeases [General function prediction only]

COG3093 Plasmid maintenance system antidote protein [General function prediction only]

COG3665 Uncharacterized conserved protein [Function unknown]

COG3707 Response regulator with putative antiterminator output domain [Signal transduction mechanisms]

COG4770 Acetyl/propionyl-CoA carboxylase, alpha subunit [Lipid metabolism]

COG5361 Uncharacterized conserved protein [Function unknown]

pfam00872 Transposase, Mutator family.

pfam00893 Small Multidrug Resistance protein.

pfam01425 Amidase.

pfam01812 5-formyltetrahydrofolate cyclo-ligase family.

pfam02626 Allophanate hydrolase subunit 2.

pfam02682 Allophanate hydrolase subunit 1.

pfam05015 Plasmid maintenance system killer protein.

pfam05232 Bacterial Transmembrane Pair family.

pfam06965 Na+/H+ antiporter 1.

smart00418 helix_turn_helix, Arsenical Resistance Operon Repressor;

alignment positions 1 · · · 131

```
 Osp-1-1   UAACCU..........CG.AUGG.ACGGCCAU.CG.............UUGAAAGUACACA...........................CCGGG.ACGGCCCGG.CUC........................
 Pen-1-1   GCGCAC......U...GU.GAGG.ACGACCUC.AC....G........ACGCAGUGCGUGAUCGCU.....................ACGGGG.ACGGCCCCAUCAUUGAAACGAU...............
 Ppu-1-1   GCACAC......U...GU.GAGG.ACGACCUC.AC....G........ACGCACCGCGUGACCAUUCC.....................GGGG.ACGGCCCC..AUCAUUCAAACGAU.............
 Ppu-2-1   CCCACG..........GU.GAGG.ACGACCUC.AC.............GACGCACCGCGUGACCACUCC....................GGGG.ACGGCCCC..AUCAUUCAAACGAU.............
 Pae-5-1   CACGCA..........GU.GAGG.ACGACCUC.AC.............CACCCGCCUCCGGCCGGUGAUCGAA...............UGGGG.ACGGCCCCG.CCACCCGUUGCCCU.............
 Pae-3-1   GCCGCU..........GU.GGGG.ACGACCUC.AC.............CACCCGCCGCCGGCGGGUGACCGAAC...............GGGG.ACGGCCCC..ACCAUCCGCCCU...............
 Oba-1-1   GGGCAA..........GC.GAGG.ACGACCUC.GC.............CCGGCCAGACGGUCAGCACG.....................CGGG.ACGGCCCG..ACCUACAAUC.................
 Oba-1-2   GGGCAA..........GC.GAGG.ACGACCUC.GC.............CCGGCCAGACGGUCAGCUCG.....................CGGG.ACGGCCCG..AGCUACAAUC.................
 Sba-2-1   ACCGGC..........UU.AGGG.ACGACCCU.AA.............ACAUGCCUUAAUUUGG.......................UAGGGG.ACGACCCCUAAUU........................
 Sba-1-1   ACCGGC..........UU.AGGG.ACGACCCU.AA.............ACAUGCCUUAAUUUGG.......................UAGGGG.ACGACCCCUAAUG........................
 Ssp-5-1   AACGGC..........UU.AGGG.ACGACCUU.AA.............AUACGCCGAUAUCAAUU.......................UGGGG.ACGACCCCG.UAUG.......................
 Spu-1-1   AAUGGC..........UU.AGGG.ACGACCCU.AA.............AUACGCCUAAUCUAACAU.......................GGGG.ACGACCCC..UUUAUG.....................
 Ssp-4-1   UCACAA...CGGC...UU.AGGG.ACGACCCU.AAAUACG.C.CG...AUAUCAAUU...............................UGGGG.ACGACCCCG.UAU........................
 Son-1-1   UCAUAA...CGGC...UU.AGGG.ACGACCCU.AAAUACG.C.CG...AUAUCAAUU...............................UGGGG.ACGACCCCG.UAU........................
 Yps-2-1   CUCUGC.....UG...GC.UGGG.ACGACCCA.GU....U.A......CGCAUCUGUUUUU...........................CGGGG.ACGGCCCCG.CCUGAGCG...................
 Ybe-1-1   UUCGGC.....UG...UU.CGGG.ACGACCUG.AG....C.A......CGCAUCUGUUUUC...........................CGGGG.ACGACCCCG.CCUAAGC....................
 Yin-1-1   CCUUUC..UGCUG...UU.CAGG.ACGACCUG.AG....C.ACGCA..UCUGUUUUU...............................CGGGG.ACGACCCCG.CCUGAGC....................
2Yfr-1-1   UCUAUC...UGCU...GCUUGGG.ACGACCCA.GC....ACG.CA...UCUGUUUUCU..............................CGGGG.ACGACCCCG.CAUGAGC....................
 Bps-8-1   UUCGCU.....GC...CC.GGGG.ACGACCCC.GG....G.C......CGUCGAUGUUCAAC...........................GGGG.ACGACCCC..AAUUGAUUCAU................
 Bps-5-1   UUCGCU.....GC...CC.GGGG.ACGACCCC.GG....G.C......CGUCGAUGUUCAAC...........................GGGG.ACGACCCC..AAUUGAUUCAU................
 Bth-1-1   UCCGCU.....GC...CC.GGGG.ACGACCCC.GG....G.C......CGUCGAUGUCCAAC...........................GGGG.ACGACCCC..AUUUGAUUCAC................
 Bce-1-1   GGCGCC.....GC...CC.GGGG.ACGACCCC.GG....G.C......CAUUCGCUUCAAC............................GGGG.ACGGCCCC..AUCCGAUCU..................
4Bvi-1-1   GGCGCC.....GC...CC.GGGG.ACGGCCCC.GG....G.C......CGUUCGCUUCAAC............................GGGG.ACGACCCC..AUCCGAUCC..................
7env-4     GCGCCG......C...CC.GGGG.ACGACCCC.GG....G........UCAUUCGUUUCAAC...........................GGGG.ACGGCCCC..AUCCGAUAC..................
4Bsp-1-1   GCGCCG......C...CC.GGGG.ACGACCCC.GG....G........UCAUUCGCUUCAAC...........................GGGG.ACGGCCCC..AUCCGAUAC..................
 Bam-1-1   GCGCCG......C...CC.GGGG.ACGACCCC.GG....G........UCAUUCGAUUCAAC...........................GGGG.ACGGCCCC..AUCCGAGAC..................
 Bdo-1-1   GCGCCG......C...CC.GCGG.ACGACCGC.GG....G........ACAUUCGCUUCAAC...........................GGGG.ACGGCCCC..AUCCGAUAC..................

           ......<<<<<<<...<<.<<<<.....>>>>.>>....>.>.>>>>>.......................................<<<<<<.....>>>>>>...........................
           ......0222222...?2.2222.....2222.2?....2.2.22220.......................................222222.....222222...........................
           ●●●●●●-----○○---●●-●RGG-ACGACCY●-●●----○-○------●●●●●●●●●●●●●○--------------------------○●GGG-ACGRCCC●○-●●●●●●●○○○-----------------
```

```
 5 env-5      ACAGUU......C...GU.AAGG.ACGACCUU.GC....G........CAUCUGCUUCAAAC...........................GGGG.ACGGCCCC..AUCUGAUCAU.................
   Bxe-1-1    ACGGUU......C...GU.AAGG.ACGACCUU.GC....G........UUUUCACGAAC..............................GGGG.ACGGCCCC..AUCUGAUCAU.................
   Vch-2-1    ACAUCA....GGA..AGU.GAGG.ACGACCUC.ACCGC.U.C.C....AUCUUUUCUUC..............................GGGG.ACGACCCC..AUAUAU.....................
   Vch-4-1    ACAUUA....GGA..AGU.GAGG.ACGACCUC.ACCGC.U.C.C....AUCUUUUCUUC..............................GGGG.ACGACCCC..AUAUAU.....................
   Val-1-1    UCAAUC....AGA..AGU.GAGG.ACGACCUC.ACCGC.U.C.U....AUCUUUUCUC...............................GGGG.ACGACCCC..AUCUCGUUG..................
   Vpa-1-1    ACAGAA..........GU.GAGG.ACGACCUC.AC.............CGCUCUAUCUUUUCUC.........................GGGG.ACGACCCC..AAUCUCGUUG.................
   Vsp-1-1    AAAGAA..........GU.GAGG.ACGACCUC.AC.............CGCUCUAUCUUUUCUC.........................GGGG.ACGACCCC..AUCUCGUUG..................
   Pat-1-1    AACGAA..........AU.GAGG.ACGACCUC.AU.............CCCUCGCGCUUUUCU.........................CUGGG.ACGACCCAG.CUCAAAUG...................
   Sen-1-1    UAUACG......U...UA.CAGG.ACGACCUG.UA....A........ACGCUAUUCUCACC...........................GGGG.ACGGCCCC..AUUAAUUAUCU................
   Eco-16-1   AUCAGC.....UU...UG.CAGG.ACGACCUG.CA....A.A......CGCCUCUUUUCACC...........................GGGG.ACGGCCCC..AAUUCUCC...................
 4 Sdy-1-1    AUCAGC.....UU...UG.CAGG.ACGACCUG.CA....A.A......CGCGUCUUUUCACC...........................GGGG.ACGGCCCC..AAUUCUCC...................
 4 Eco-2-1    AUCAGC.....UU...UG.CAGG.ACGACCUG.CA....A.A......CGCGUCUUUUCACC...........................GGGG.ACGGCCCC..AAUUUUCC...................
 4 Sfl-3-1    AUCAGC.....UU...UG.CAGG.ACGACCUG.CA....A.A......CGCCUCUUUUCACC...........................GGGG.ACGGCCCC..AAAUUUCC...................
 4 Sen-4-2    AUUAGC.....UU...UG.CAGG.ACGACCUG.CA....A.A......CGCUUC.UUUCACC...........................GGGG.ACGGCCCC..AUUGUUCU...................
   Gvi-1-1    CAAAAC......G...GU.GGGG.ACGACCCC.AC....C........CUUCGCAAAGUUUC.........................GCGGGG.ACGACCCCGCACCCAC.....................
   Msp-1-1    GGCGGU..........GC.AGGG.ACGGCCUU.GC.............UUUCUUCUAUCCGG.........................AUGAGG.ACGGCCUCAUCGCUAUUGU..................
 5 env-6      GCUGCC.....CC...UC.GGGG.ACGACCCC.GA....G.G......CCUCCUGAUAC............................GCAGGG.ACGACCCUGCUUCCUGC....................
 2 Nwi-1-1    GGGUCG......C...GG.AAGG.ACGACCUU.UC....G........UAUCCCACACAACCACGC.....................GCGGGG.ACGGCCCUGCCUCAUGUCUUG................
 8 Bja-1-1    GCCGUU.....UG...GG.AGGG.ACGGCCCU.CC....U.A......UCCACCAGCAUAUCAGCC.....................GCGGGG.ACGACCCUGCCUUGAAU....................
   Nsp-1-1    CAAGGC....UGC...GG.AAGG.ACGGCCUU.CC....G.C.A....UUCUCACACAACCGCCC......................GCGGGG.ACGGCCCUGCCUUCACGAUCUG...............
   Gsu-1-1    GACCGA.....CG...GC.GGGG.ACGACUCC.GC....C.G......CCCGAACAUCCAC............................GGGG.ACGACCCC..ACGAAUCGGAA................
   Plu-1-1    AAUGGU.....UA...UC.CGGG.ACGACCUG.GA....U.A......UACUGCCGUUUUC...........................UGGGG.ACGACUCCA.CAACACUAUU.................
   Rru-1-1    UCGUCA......G...GC.GGGG.ACGACCCC.GC....U........GGAGGUGUUUUUCAG........................GCGGGG.ACGACCCCGCCAUGGAUCCAGAG..............
   Rpa-4-1    CCGCUC.....CG...GC.CGGG.ACGACCCG.GC....C.G......CACAGGGAGCAUGA.........................GCCGGG.ACGACCCGGCGUUUGUUCACCGU..............
   Rpa-1-1    CAGCUC.....CG...GC.CGGG.ACGACCCG.GC....C.G......ACCAGGGAGCAUGA.........................GCCGGG.ACGACCCGGCGUUUGUCUGAAGU..............
   Rpa-3-1    CCGCUC.....CG...GC.CGGG.ACGACCCG.GC....C.G......AACAGGGAGCAUGA.........................GCCGGG.ACGACCCGGCGUUUGUCUGAUGU..............
   Rpa-2-1    UCGCUC.....CG...GC.CGGG.ACGACCCG.GC....C.G......AACAAGGAGCAUGA.........................GCCGGG.ACGGCCCGGCGCAUUGUCUUUCAGA............
   Pae-5-2    UCAGCC....GGA...AG.CGGG.ACGACCCG.UU....U.U.C....CCUCUUUCAUUGC..........................GCGGGG.ACGACCCUGCAGA........................
   Pae-3-2    UCAGCC....GGA...AG.CGGG.ACGACCCG.UU....UUC.C....UUCUUUCAUUGC...........................GCGGGG.ACGACCCUGCAG.........................
   Asp-1-1    UGCGCC......G...UG.AAGG.ACGACCUU.CC....U........UUCCUCCCAUCAUAUC.......................GCCGGG.ACGACCCGGCACAGGACA...................
   Rsp-1-1    UCCAGC.............GAGG.ACGACCUC................CACAUCUGGACAACAGC.......................CCGGG.ACGACCCGG.CACCUAGUUU.................
   Rba-1-1    GUCAUC..............AGG.ACGACCU.................UCCCAUCAUGGAUCAACA......................CCGGG.ACGACCCGG.AAAGCACGC..................
   Rpa-2-3    GGCCAA.............GAGG.ACGACCUC................CCAAUCACCCAAAUGAC......................GCAGGG.ACGACCCUGCCCCGCUCUCGU................
 6 Eca-1-1    CGUUAC.....CG...CG.AUGG.ACGACCAU.CA....U.G......CGGAUCGCAUUUU..........................GGCGGG.ACGACCUGCCCUUGC......................
 6 env-7      CCCCAA..........CC.ACGG.ACGGCCGU.GG.............CGCAGAUGAAUAAA.........................GGCGGG.ACGGCCCGCCAGCCCUUC...................
   env-8      CUUUUC.....UG...CC.ACGG.ACGACCGU.GG....C.A......CAUGACGAAUGAC..........................GGCGGG.ACGGCCCGCCAUGACCUUC..................
   Pfl-2-1    UUUUGC.....CG...UC.GGGG.ACGACCCC.GG....C.G......CUCAUCGAAAAAC..........................UGCGGG.ACGGCCCGCAUCCCCUUC...................
   Pna-1-1    ACCAGC......C...CA.CGGG.ACGGCCCG.UG....G........CGCCGCACCACCC..........................GGCGGG.ACGGCCCGCCAA.........................
   Reu-1-1    GGCGCC.....GC...CG.CGGG.ACGACCCG.UG....G.C......CGAUACGCAACC...........................GGCAGG.ACGACCUGCCUGACAUCGA..................
   Sgl-1-1    UAGUAC...........U.CAAG.ACGACCUG.G..............CCGCCAGCCUUAUUCUG........................CGGG.ACGACCCG..ACCAAUAC...................
13 Mma-1-1    AGCGGU....AGG...GC.CGGG.ACGACCCG.CC....C.C.U....UUUCCCCUCACCU...........................CCGGG.ACGGCCCGG.CCUUGAC....................
   Sde-1-1    GCGCUG....CUG...UU.UAGG.ACGACCUA.AG....C.A.G....GCUCUCAAU...............................GCGGG.ACGACCCGC.CCCAGUUC...................
   Msp-2-1    GCAGUU...UUUA...AU.CAGG.ACGACCUG.AU....U.A.AA...GGAUUUAUUCU..............................CGGG.ACGACCCG..CAAGUUU....................
   Ssp-1-1    CUAUGG....CAG..UGU.GAGG.ACGACCUC.CC....C.U.G....AAUGGAAAACU............................GUCGGG.ACGACCCGGCCCUUCAAU...................
   Oal-1-1    GCCUGA.............GAGG.ACGACCUC................CCCGCACUCGGGAUCAGCA....................UUCGGG.ACGGCCCGAAACCUGA.....................
   Jsp-1-1    UGGCCC..........GC.CGGG.ACGACCCG.GC.............CAUGUACGGGUAGUUGU........................CAGG.ACGACCUG..CGAUAUGU...................
   Dar-1-1    GCAACA....GCU...UC.CGGG.ACGACCCG.GA....G.G.C....AAAAGAAUCC.............................GGCGGG.ACGACCCGUCAAACGA.....................
   Atu-1-1    GGUCAA.....GC..G.C.GAGG.ACGACCUC.G.....G.C......CAAAUCGGCGACAGCAUG.......................CGGG.ACGGCCCG..GUGAUAUACC.................
   Pfl-1-1    UGCGCA.....UC...GC.GAGG.ACGACCUC.GC....G.A......CUCAUUCGAGCGAUCAAUUCC....................GGGG.ACGACCCC..AUCAACCGAU.................
   Bbr-1-1    CAAUCC....CGC...GC.CGGG.ACGACCCG.GC....G.C.G....UGUACUGGCC..............................UGAGG.ACGACCUCG.C..........................

              ......<<<<<<<...<<.<<<<.....>>>>.>>....>.>.>>>>>.......................................<<<<<<.....>>>>>>...........................
              ......0222222...?2.2222.....2222.2?....2.2.22220.......................................222222.....222222...........................
              ●●●●●●-----○○---◍●-●RGG-ACGACCY●-●◍----○-○------●●●●●●●●●●◍◍◍○--------------------------○●GGG-ACGRCCC●○-●●●●●◍◍○○○-----------------
```

```
   Ccr-1-1   CUAGCG..AAGGC...GC.GGGG.ACGACCCC.GC....G.C.CUU..UUGCUUUUC...............................CGGGG.ACGACCCCG.CCUUUCGA...................
12 Xau-1-1   ACCCCA......G...GC.CGGG.ACGACCCG.GC....C........AUUCCCUCGGGACCCG.......................GCCGGG.ACGACCCGGCGCUUCAGACGGCUUG............
   Pfl-2-2   GCGCCA......C...GC.GGGG.ACGACCCU.GC....G........GCUCAAUCGAGUUUUCAAUC....................UGGGG.ACGGCCCCG.UCAAUCGAU..................
   Xau-1-2   CCCGAU....GGG...AC.GCGG.ACGGCCGC.GUG...C.C.C....UUGCGUAUCGAGCG.........................GCGGGG.ACGACCUCGCCACCC......................
14 Ret-1-1   AGCCCC......G...AC.CGGG.ACGGCCCG.GU....C........GCUUCCCUGGGGCACA.......................GCCGGG.ACGACCUGGCAUCAUCA....................
 2 Cef-1-1   GGUCGU....GAU...CG.CGGG.ACGACCCG.CG..C.A.U.C....GGGUACCGGA..............................UGGGG.ACGGCCCCG.CCUGAGCUUUCACA.............
   Bli-1-1   AAGAAA.....CG..C.U.CGGG.ACGACCCG.A.A...C.G......CAAUCAGUCACGA...........................UGAGG.ACGGCCUCG.AAGAAC.....................
 9 Asp-2-1   ACUUUU...........G.GAGGAACGACCUC.CC.............CCAAUGUGGGAAAAU........................AUCAGG.CCGACCUGAUCUUGAUAAA..................
   Neu-1-1   UUUCCU.....GA..U.U.GAGG.ACGACCUC.A.....U.C......UUCGGGACAAUCUACUAU......................CCGGG.ACGACCUGG.UAUCUUUCAAU................
   Ogr-1-1   UGUUGC...GCCGCAAGC.GGGG.ACGGCCCC.GCC...C.G.GC...CGAUGGGUCUGCACA..........................CGGG.ACGGCCCG..GGACAACUCG.................
   Ret-1-2   GCACUA..........UC.GCAG.ACGACUGC.GG.............CGUCCGGAGGAUGCAACUGUC..................GCGGGG.ACGGCCUCGCUCUUGAACA..................
   Sme-1-1   CACAUC.............GCAG.ACGGCUGC................ACGGAUCGAACAGAUCCGAACUUCC..............GCGGGG.ACGGCCUCGCUCCAUAUGAA.................
   Rnu-1-1   UCCCGC....CAU...GC.GAGG.ACGACCUCCGCCCC.A.U.G....CAACCCGGGGUUUUUAGA.....................UUCGGG.ACGGCCCGAGUGAC.......................
   Abo-1-1   CAAUUU.....CC...CU.GGGG.ACGACCCC.GG..C.G.G......CCAAGGCUAUAAA..........................GACGGG.ACGGCCCGUCAUCUGUCCCAGAGACAGCCCGACAA..
   Wsu-1-1   AAAUGG.....CC...CC.UAGG.ACGACCUA.GA....G.G......GUGACUUCACCGAUGA...........................UGGG.ACGACCCA..AAUCCAAACCCCAAA............
11 Cgl-1-1   AUGAGG......C...UU.CGGG.ACGACCCG.AA....G........AAAUCUAUGACAGC..........................CUGGGAACGGCCCAG.AGUUCUUAAGAAAGUUUGACU......
   Pae-5-3   UGGAAA.....AG..CGC.GCCG.ACGACGGC.GC....C.U......GCGCACUGGAAAA............................CGGG.ACGGCCCG..CGCCGCUCUCCCGCACUC.........
   Pae-6-3   UGGAAA.....AGC..GC.GCCG.ACGACGGC.GC....C.U......GCGCACUGGAAAA............................CGGG.ACGGCCCG..CGCCGCCCUCCCGCACUC.........
   Psy-3-1   UCCUCA......G...CA.ACGG.ACGACCGU.UG....C........GUAGAUAGAUCAAU.........................AGCGGG.ACGGCCCGCUUCCCACCACGAAACCGAGCGCGGCUGG
   Psy-2-1   CCCUCA......G...CA.ACGG.ACGACCGU.UG....C........GUAGAUUGAUCAAU.........................AGCGGG.ACGGCCCGCUUCCCAUCACGAAACCGAGUGCGACUGG
   Psy-1-1   CCCUCA......G...CA.ACGG.ACGACCGU.UG....C........GUAGAUAGAUCAAUA.........................GCGGG.ACGGCCCGC.UUCCCAUCACGAAACCGAGCGCGUCUG
 3 env-9     UGUGGC.....CU...GU.UGGG.ACGACCCA.AU....A.G......CGCAUCUAUUUAGC...........................GGGG.ACGACCCC..AACUUCAAACACUGGAAUACCUUAUGA
 1 Bps-8-2   CGACACCGUGUUGA..AC.GGCG.ACGACGCC.GU....C.G.ACGCGAUUCUGA..................................CGGG.ACGCCCCG..AUCUCGUGAUCCUCGACGUGAACCUGC
10 env-10    CGGAUC.....GA...UC.CGGA.UCGAUCCG.GA....U.C......AAUUCUGAUCGA...........................UCCGGA.UCGAUCCGGAUGCCAACACUGCGUCCUCGUGCCGGUC
   Xca-1-1   UGACAC..........GG.CGGG.ACGGCCCG.CC.............ACGCGAUGACCUCGCGUUCCAUA..................GCCG.ACGGCGGC..CCGGUGUGCAGACACCGGGUCGCUUUU
 1 Xor-1-1   CCACAA..........GG.CGGG.ACGGCCCG.CC.............ACGCGAUGACUUCGCGUUCCACA..................GCCG.ACGGCGGC..CCAGUGUGCAAACACUGGGUCGCUUUU
   Rpa-4-2   GGGAUC.............GAGG.ACGACCUC................AAAAUCACCUUUUCG........................GCCGGG.ACGACCCGGCCUCCUCGAAGACCGCGCAUUGCGCCGG
   Rpa-3-2   GUGCUC.............GAGG.ACGACCUC................CCAAUCACCCUUUCG........................GCCGGG.ACGACCCGGCUUCCCGAGACUGCGCAUGUUCCGCGCC
   Rpa-1-2   ACGCUC.............GAGG.ACGACCUC................AAAAUCACCCUUUCG........................GCCGGG.ACGACCCGGCCUCCUGUCGACGACGCAAGCGAAGGCG
 1 Sfr-1-1   AACCCA..........UU.GAGG.ACGACCUC.AG.............CCUGACAUGCACCACUUUUCAUCGACAGGUAUUCACAUC..GGGG.ACGGCCCC..AUCUGCCAUCUACACAGAAGAGCCGCU
   Mlo-1-1   UUCCAU....GGC...GC.GGGG.ACGACCCC.GC....G.C.C....UCCCCAAGAAUUCG.........................GUCGGG.ACGACCCGCCAUAAUGAUCGGCAUCUGAGGAUCGGAU
   Csp-1-1   CUGAUC.AAAGGC...GC.GGGG.ACGACCCC.GC....G.C.CUUU.UCCUUUUUCUC.............................CGGGG.ACGGCCCCG.CUCAUUUGCCGAGCGCAUUUUCUUCAC
   Aeh-1-1   AUCCCG.....GG...CC.GGGG.ACGACCCU.GG....C.C......GCCAGAACGAAAACGG.........................CGGG.AGGACCUG..GACGUGGCCCUGCAGGCCGCCGAGGGC
12 Acr-1-1   GCAACG....GGC...CG.GGGG.ACGACCCCACG....G.C.C....GGGACGAUCGCCC..........................GCGAGG.ACGGCCUCGCUAGAGCAUCAUCCGAUCAGAUGGAAUC
   Pin-1-1   CGGGAA......G....U.AAGG.ACGACCUC.AG....C........GACCUAUCCAAUUCA..........................GGGG.ACAGCCUC..AUAUAUAAGGUUAACUAUAUGGUUUGG
15 Pgi-1-1   GACCGU...........U.UAGG.UCGACCUA.A..............CUGAACGA................................UUAGG.UCGACCUAA.AACAAACAACAGGUCGUCCCGACACAU
             ......<<<<<<<...<<.<<<<.....>>>>.>>....>.>.>>>>>.......................................<<<<<<.....>>>>>>...........................
             ......0222222...?2.2222.....2222.2?....2.2.22220.......................................222222.....222222...........................
             ●●●●●●-----○○---●●-●RGG-ACGACCY●-●●----○-○------●●●●●●●●●●●●●○--------------------------○●GGG-ACGRCCC●○-●●●●●●●○○○-----------------
```

alignment positions 132 · · · 166

```
   Osp-1-1   ..(  0 nucs)AGGUG...AAC......AUGUUU
   Pen-1-1   ..(  0 nucs)GGAGA...CUCACC...AUGUCC
   Ppu-1-1   ..(  0 nucs)GGAG....ACACACUC.AUGUCC
   Ppu-2-1   ..(  0 nucs)GGAGA...CACACUC..AUGUCC
   Pae-5-1   ..(  0 nucs)GGAGA...UCGCC....AUGUCC
   Pae-3-1   ..(  0 nucs)GGAGA...UCGCC....AUGUCC

             ..          .......................
             ..          .......................
             --          RRRR○---●●●●○○---AUG●●Y
                         SD++++++         AUG
```

```
  Oba-1-1   ..(  0 nucs)GGAG....CCAACCA..AUGCGC
  Oba-1-2   ..(  0 nucs)GGAG....CCAACCC..AUGCGC
  Sba-2-1   ..(  0 nucs)AGGAG...UUCACU...AUGAGU
  Sba-1-1   ..(  0 nucs)AGGAG...UUCACU...AUGAGU
  Ssp-5-1   ..(  0 nucs)GGAG....UUCUCC...AUGAGC
  Spu-1-1   ..(  0 nucs)GGAG....UUCACU...AUGAGU
  Ssp-4-1   ..(  0 nucs)GGGAG...UUCUCU...AUGAGC
  Son-1-1   ..(  0 nucs)GGGAG...UUCUCU...AUGAGC
  Yps-2-1   ..(  0 nucs)AGGUGG..UUACC....AUGGCU
  Ybe-1-1   ..(  0 nucs)GAGGUGG.UUACC....AUGGCU
  Yin-1-1   ..(  0 nucs)GAGGUGG.UUACC....AUGGCU
2 Yfr-1-1   ..(  0 nucs)GAGGUGG.UUACC....AUGGCU
  Bps-8-1   ..(  0 nucs)GGAGAG..UCUCCGC..AUGCCU
  Bps-5-1   ..(  0 nucs)GGAGAG..UCUCCCC..AUGCCU
  Bth-1-1   ..(  0 nucs)GGAG....AGUCUCCGCAUGCCU
  Bce-1-1   ..(  0 nucs)GGAG....UACGUCGAGAUGGCG
4 Bvi-1-1   ..(  0 nucs)GGAG....UACGUCGAGAUGGCG
7 env-4     ..(  0 nucs)GGAG....UACGUCGAGAUGGCG
4 Bsp-1-1   ..(  0 nucs)GGAG....UACGUCGAGAUGGCG
  Bam-1-1   ..(  0 nucs)GGAG....UAGGUCGAGAUGGCG
  Bdo-1-1   ..(  0 nucs)GGAG....UACGUCGCGAUGGCG
5 env-5     ..(  0 nucs)GGAG....CUGUCUC..AUGUCC
  Bxe-1-1   ..(  0 nucs)GGAG....CUUGUCUC.AUGUCC
  Vch-2-1   ..(  0 nucs)GAGG....UGUAUAU..AUGGCG
  Vch-4-1   ..(  0 nucs)GAGG....UGUAUAU..AUGGCG
  Val-1-1   ..(  0 nucs)GGAA....AUGUGAC..AUGGCU
  Vpa-1-1   ..(  0 nucs)GGAA....AUAUGAC..AUGGCU
  Vsp-1-1   ..(  0 nucs)GGAA....AUAUGAC..AUGGCU
  Pat-1-1   ..(  0 nucs)GGUAGG..UAUAAAA..AUGAAC
  Sen-1-1   ..(  0 nucs)GGAG....CCUGCU...AUGUCC
  Eco-16-1  ..(  0 nucs)GGAG....CCUGAU...AUGUCC
4 Sdy-1-1   ..(  0 nucs)GGAG....CCUGAU...AUGUCC
4 Eco-2-1   ..(  0 nucs)GGAG....CCUGAU...AUGUCC
4 Sfl-3-1   ..(  0 nucs)GGAG....CCUGAU...AUGUCC
4 Sen-4-2   ..(  0 nucs)GGAG....CCUGAU...AUGUCC
  Gvi-1-1   ..(  0 nucs)AGGAGG..UGCGCC...AUGUCU
  Msp-1-1   ..(  0 nucs)GGAG....AUCAGC...AUGGCC
5 env-6     ..(  0 nucs)GGAA....UCUGCCUU.AUGUCC
2 Nwi-1-1   ..(  0 nucs)GGAGG...ACCGC....AUGGCC
8 Bja-1-1   ..(  0 nucs)GGAGG...GUCGUC...AUGGCC
  Nsp-1-1   ..(  0 nucs)GGAGG...GCCAU....AUGGCC
  Gsu-1-1   ..(  0 nucs)AGGACG..ACAUC....AUGACG
  Plu-1-1   ..(  0 nucs)AUGGGG..CGUUGAU..AUGUCC
  Rru-1-1   ..(  0 nucs)GGAGG...CAUCC....AUGGCU
  Rpa-4-1   ..(  0 nucs)GAGGA...CAGC.....AUGAUC
  Rpa-1-1   ..(  0 nucs)GAGGA...CAACC....AUGAUC
  Rpa-3-1   ..(  0 nucs)GAGGA...CAUCC....AUGAUC
  Rpa-2-1   ..(  0 nucs)AAGGG...CAAUCC...AUGAUC
  Pae-5-2   ..(  0 nucs)GAGG....CUGAAA...AUGGCG
  Pae-3-2   ..(  0 nucs)AGAGG...CUGAAA...AUGGCG

            ..          .......................
            ..          .......................
            --          RRRR○---●●●●●○---AUG●●Y
                        SD++++++         AUG
```

```
  Asp-1-1    ..(  0 nucs)AGCAGG..GUCGCC...AUGCGC
  Rsp-1-1    ..(  0 nucs)GAGG....UCGCC....AUGCGC
  Rba-1-1    ..(  0 nucs)GGGG....UUUCCC...AUGCGC
  Rpa-2-3    ..(  0 nucs)GGAGG...UCGUC....AUGGCC
  Eca-1-1    ..(  0 nucs)AUGAG...UCUAGCGU.AUGUUC
 6env-7      ..(  0 nucs)GAGGG...UUACUGAACAUGUUC
 6env-8      ..(  0 nucs)GAGGG...UCACCCUG.AUGUUU
  Pfl-2-1    ..(  0 nucs)AGGG....GUUUUGCC.AUGUUC
  Pna-1-1    ..(  0 nucs)AGGGA...UUGCCC...AUGUUC
  Reu-1-1    ..(  0 nucs)GAGGGA..ACAUCCCC.AUGUCC
  Sgl-1-1    ..(  0 nucs)GGAA....AACGUCCU.AUGCAG
13Mma-1-1    ..(  0 nucs)GGAG....CAUGGUCACAUGCCG
  Sde-1-1    ..(  0 nucs)GGUA....UUUAACUU.AUGUUC
  Msp-2-1    ..(  0 nucs)GGAG....CCGUU....AUGUUU
  Ssp-1-1    ..(  0 nucs)GAGG....UAUAAC...AUGCGC
  Oal-1-1    ..(  0 nucs)AAGG....UUUCGACC.AUGGCC
  Jsp-1-1    ..(  0 nucs)GGAGG...UCUAA....AUGGCU
  Dar-1-1    ..(  0 nucs)GAGACG..ACAAC....AUGUUC
  Atu-1-1    ..(  0 nucs)GGAG....AAUGAA...AUGCGC
  Pfl-1-1    ..(  0 nucs)GAGG....UAGGC....AUGUCC
  Bbr-1-1    ..(  0 nucs)GAGA....UCCCAUCC.GUGUCG
  Ccr-1-1    ..(  0 nucs)AGUGGG..GCUCGUC..GUGGCC
12Xau-1-1    ..(  0 nucs)AGGAG...ACACGAC..AUGACC
  Pfl-2-2    ..(  0 nucs)GGAGG...UCUUUC...AUGUCC
  Xau-1-2    ..(  0 nucs)GAGG....UACCCGCCAAUGGCG
14Ret-1-1    ..(  0 nucs)GGAGA...GACGCU...AUGAUG
 2Cef-1-1    ..(  0 nucs)AGGAUG..UUCACC...AUGCCG
  Bli-1-1    ..(  0 nucs)AGGA....AUUCCCGUCAUGACU
 9Asp-2-1    ..(  0 nucs)AGGA....AAGCAAU..AUGGCU
  Neu-1-1    ..(  0 nucs)GGAGA...UGAG.....AUGGCU
  Ogr-1-1    ..(  0 nucs)GGAG....CAUCACC..UUGCGC
  Ret-1-2    ..(  0 nucs)AGGAG...AGGCGU...AUGGCC
  Sme-1-1    ..(  0 nucs)GGAG....UUGGAA...AUGGCC
  Rnu-1-1    ..(  0 nucs)AGGAG...AUCUGCC..AUGCGC
  Abo-1-1    ..(  0 nucs)GAGAA...AAAGUC...AUGUUC
  Wsu-1-1    ..(  0 nucs)GAGGA...ACCACU...AUGUUU
11Cgl-1-1    ..(  0 nucs)AGAG....AAC......AUGCCG
  Pae-5-3    ..(  0 nucs)GAGGUGG.UG.......AUGACA
  Pae-6-3    ..(  0 nucs)GAGG....UGGUG....AUGACA
  Psy-3-1    GU(  4 nucs)GUGG....UUAUGCC..AUGUUC
  Psy-2-1    GU(  4 nucs)GGGG....UUAUGCC..AUGUUC
  Psy-1-1    GG(  5 nucs)GGGG....UUAUGCC..AUGUUC
 3env-9      A.(  0 nucs).......................
 1Bps-8-2    C.(  0 nucs).......................
10env-10     ..(  0 nucs).......................
  Xca-1-1    UU( 18 nucs)AGGAG...UGAUGUG..AUGCCC
 1Xor-1-1    UU(  0 nucs).......................
  Rpa-4-2    CG( 46 nucs)AGGAGG..CCGCU....AUGGCU
  Rpa-3-2    GC( 24 nucs)AGGAGG..CCCUC....AUGGCU
  Rpa-1-2    GC( 47 nucs)AAGGG...GGCCCG...AUGGCC

             ..          .......................
             ..          .......................
             --          RRRR○---●●●●●○---AUG●●Y
                         SD++++++         AUG
```

```
 1Sfr-1-1    AA(  0 nucs).......................
  Mlo-1-1    CC( 22 nucs)GGAG....CGUUCCG..AUGUCC
  Csp-1-1    GC( 38 nucs)GGGAG...CAGGUC...AUGGCC
  Aeh-1-1    AU( 66 nucs)GGCGGG..UGCU.....GUGCGA
12Acr-1-1    AU( 86 nucs)GAGGG...CAUG.....AUGGAC
  Pin-1-1    UU(165 nucs)AGG.....UUCGAUAUAAUGAAC
15Pgi-1-1    CA(199 nucs)AAGAA...AAAGGU...AUGGAU
             ..          .......................
             ..          .......................
             --          RRRR○---●●●●●○---AUG●●Y
                         SD++++++         AUG
```

2.9.4 Automated, iteration 0: CMfinder's original prediction for mini-*ykkC*

(Duplicate sequences are not shown in this case, only unique sequences.)

abbrev. of hits	**taxonomy of species**
Bps-1-1 to Bps-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* K96243
Neu-1-1 to Neu-1-1	β-proteobacteria Nitrosomonadales Nitrosomonadaceae *Nitrosomonas europaea* ATCC 19718

abbrev	**RefSeq accession**		**5′ at**	**3′ at**	**genes**
Neu-1-1	NC_004757.1	-	2389361	2389308	**RNA**→ EmrE (COG2076)→
Bps-1-1	NC_006350.1	+	1359249	1359299	**RNA**→ EmrE (COG2076)→

COG2076 Membrane transporters of cations and cationic drugs [Inorganic ion transport and metabolism]

alignment positions 1 · · · 59

```
Neu-1-1   UCC..UGAUUGAGGACGACCUCAUCUUCG...GGACAAUCUACUAUCCGGGACGACCUG
Bps-1-1   UCGCUGCC.CGGGGACGACCCCG...GGCCGUCGAUGUUCAAC....GGGGACGACCCC
          <<<..<<<.<<<<<....>>>>>...>>>...>>>............<<<<....>>>>
          002..222.20200....00202...222...200............2100....0012
          UC●○○●●●○YGRGGACGACCYCR○○○●●●○○○●GAYR●UC●AC○○○○●GGGACGACCY●
```

2.9.5 Automated, iteration 1: Pipeline's prediction for mini-*ykkC*, after one homology search

(Duplicate sequences are not shown in this case, only unique sequences.)

abbrev. of hits	**taxonomy of species**
Bps-1-1 to Bps-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia pseudomallei* K96243
Reu-1-1 to Reu-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Cupriavidus *Ralstonia eutropha* JMP134
Neu-1-1 to Neu-1-1	β-proteobacteria Nitrosomonadales Nitrosomonadaceae *Nitrosomonas europaea* ATCC 19718

abbrev	**RefSeq accession**		**5′ at**	**3′ at**	**genes**
Bps-1-1	NC_006350.1	+	1359249	1359299	**RNA**→ EmrE (COG2076)→
Reu-1-1	NC_007348.1	-	1606552	1606502	**RNA**→ AccB (COG0511)COG4770 (COG4770)AHS2 (pfam02626)AHS1 (pfam02682)→ Amidase (pfam01425)→
Neu-1-1	NC_004757.1	-	2389361	2389308	**RNA**→ EmrE (COG2076)→

COG0511 Biotin carboxyl carrier protein [Lipid metabolism]
COG2076 Membrane transporters of cations and cationic drugs [Inorganic ion transport and metabolism]
COG4770 Acetyl/propionyl-CoA carboxylase, alpha subunit [Lipid metabolism]
pfam01425 Amidase.
pfam02626 Allophanate hydrolase subunit 2.
pfam02682 Allophanate hydrolase subunit 1.

alignment positions $1\cdots59$

```
Bps-1-1  UCGCUGCC.CGGGGACGACCCCG...GGCCGUCGAUGUUCAAC....GGGGACGACCCC
Reu-1-1  GCGCCGCC.GCGGGACGACCCGU...GGC...CGAUACGCAACCGG.CAGGACGACCUG
Neu-1-1  UCC..UGAUUGAGGACGACCUCAUCUUCG...GGACAAUCUACUAUCCGGGACGACCUG
         <<<..<<<.<<<<<....>>>>>...>>>...>>>...........<<<<....>>>>
         ?02..222.22200....00222...222...20?...........2200....0022
         ●C●○○●●●–●●RGGACGACCY●●–––●●●–––●GAYR●●C●AC○○○–●RGGACGACCY●
```

2.9.6 Automated, iteration 2: mini-*ykkC* prediction after 2 iterations of homology search/CMfinder

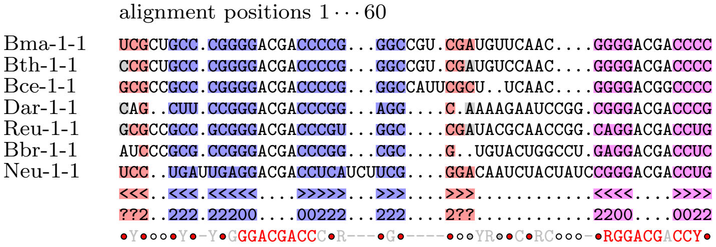
(Duplicate sequences are not shown in this case, only unique sequences.)

The database searched for homologs in this extra automated iteration was the IGRs of β-proteobacteria, i.e., the same database as the previous iteration for this motif, which was originall found in β-proteobacteria.

abbrev. of hits	**taxonomy of species**
Bbr-1-1 to Bbr-1-1	β-proteobacteria Burkholderiales Alcaligenaceae *Bordetella bronchiseptica* RB50
Bce-1-1 to Bce-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia Burkholderia cepacia complex *Burkholderia cenocepacia* AU 1054
Bma-1-1 to Bma-1-1	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia mallei* ATCC 23344
Bth-1-1 to Bth-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Burkholderia pseudomallei group *Burkholderia thailandensis* E264
Reu-1-1 to Reu-1-1	β-proteobacteria Burkholderiales Burkholderiaceae Cupriavidus *Ralstonia eutropha* JMP134
Neu-1-1 to Neu-1-1	β-proteobacteria Nitrosomonadales Nitrosomonadaceae *Nitrosomonas europaea* ATCC 19718
Dar-1-1 to Dar-1-1	β-proteobacteria Rhodocyclales Rhodocyclaceae *Dechloromonas aromatica* RCB

abbrev	**RefSeq accession**		**5′ at**	**3′ at**	**genes**
Bma-1-1	NC_006348.1	-	1960226	1960176	**RNA**→ EmrE (COG2076)→
Bth-1-1	NC_007651.1	+	1166404	1166454	**RNA**→ EmrE (COG2076)→
Bce-1-1	NC_008060.1	-	1851716	1851667	**RNA**→ EmrE (COG2076)→
Dar-1-1	NC_007298.1	+	88474	88521	**RNA**→ AccB (COG0511)COG4770 (COG4770)AHS2 (pfam02626)AHS1 (pfam02682)→ Amidase (pfam01425)→
Reu-1-1	NC_007348.1	-	1606552	1606502	**RNA**→ AccB (COG0511)COG4770 (COG4770)AHS2 (pfam02626)AHS1 (pfam02682)→ Amidase (pfam01425)→
Bbr-1-1	NC_002927.3	+	862220	862268	**RNA**→ ChaA (COG0387)→
Neu-1-1	NC_004757.1	-	2389361	2389308	**RNA**→ EmrE (COG2076)→

COG0387 Ca2+/H+ antiporter [Inorganic ion transport and metabolism]
COG0511 Biotin carboxyl carrier protein [Lipid metabolism]
COG2076 Membrane transporters of cations and cationic drugs [Inorganic ion transport and metabolism]
COG4770 Acetyl/propionyl-CoA carboxylase, alpha subunit [Lipid metabolism]
pfam01425 Amidase.
pfam02626 Allophanate hydrolase subunit 2.
pfam02682 Allophanate hydrolase subunit 1.



alignment positions 1 · · · 60

```
Bma-1-1   UCGCUGCC.CGGGGACGACCCCG...GGCCGU.CGAUGUUCAAC....GGGGACGACCCC
Bth-1-1   CCGCUGCC.CGGGGACGACCCCG...GGCCGU.CGAUGUCCAAC....GGGGACGACCCC
Bce-1-1   GCGCCGCC.CGGGGACGACCCCG...GGCCAUUCGCU..UCAAC....GGGGACGGCCCC
Dar-1-1   CAG..CUU.CCGGGACGACCCGG...AGG....C.AAAAGAAUCCGG.CGGGACGACCCG
Reu-1-1   GCGCCGCC.GCGGGACGACCCGU...GGC....CGAUACGCAACCGG.CAGGACGACCUG
Bbr-1-1   AUCCCGCG.CCGGGACGACCCGG...CGC....G..UGUACUGGCCU.GAGGACGACCUC
Neu-1-1   UCC..UGAUUGAGGACGACCUCAUCUUCG....GGACAAUCUACUAUCCGGGACGACCUG
          <<<..<<<.<<<<<....>>>>>...>>>....>>>............<<<<....>>>>
          ??2..222.22200....00222...222....2??............2200....0022
          ●Y●○○●Y●-Y●GGGACGACCC●R---●G●----●○○YR●●C●RC○○○-●RGGACGACCY●
```

2.10 The *purD* motif

2.10.1 In-line probing

Using in-line probing, we tested whether purD RNA can bind each of several metabolites. We tested two RNA molecules, of different lengths, both from *Campylobacter jejuni subsp. jejuni* NCTC 11168, RefSeq accession NC_002163.1

The “small” RNA is

GUGAUCCAACACAUUUUUAUUAACUGCGAUUGUGUGAUUUACCGUGUUCUGUGGCAUCGU
UUGAGCUUUGAAAAAAGCGAGAAGUUGCAGCCUUUAAAAAUUACCUAGCGGUUUUCUUUG
ACUUUUUG

The “long” RNA is extended on the 3′ end. Its sequence is

GUGAUCCAACACAUUUUUAUUAACUGCGAUUGUGUGAUUUACCGUGUUCUGUGGCAUCGU
UUGAGCUUUGAAAAAAGCGAGAAGUUGCAGCCUUUAAAAAUUACCUAGCGGUUUUCUUUG
ACUUUUUGGGGUCAAUCUUUGUUAAACGGCUGUUUGGGUUUAAAAGG

The concentrations of compounds used were dictated by the fact that many purine derivatives are not highly soluble.

The following compounds were tested with both the short and long RNAs: 100 µM guanine, adenine, GAR and AICAR ribotide. We also tested AIR. Since this stock was old and AIR is unstable in solution, we tested at the ideal concentrations of 3.8 mM and 100 µM, but we expect the true concentration were lower, due to degradation.

The following compounds were tested with the long RNA only: 10 µM 2′deoxyguanosine 3′P, deoxyguanosine, 3′deoxyguanosine, adenosine, guanosine, ATP, GTP, GDP, G2′MP, GMP, dATP, dGTP, dGDP, dAMP, dGMP, 2′deoxyadenosine, 2′deoxyguanosine.

We found no evidence of structural modulation of the RNA with any of these compounds.

2.10.2 Genome context and alignment

abbrev. of hits	taxonomy of species
Cco-1-1 to Cco-1-1	ε-proteobacteria Campylobacterales Campylobacteraceae *Campylobacter coli* RM2228
Cco-2-1 to Cco-2-1	ε-proteobacteria Campylobacterales Campylobacteraceae *Campylobacter concisus* 13826
Ccu-1-1 to Ccu-1-1	ε-proteobacteria Campylobacterales Campylobacteraceae *Campylobacter curvus* 525.92
Cfe-1-1 to Cfe-1-1	ε-proteobacteria Campylobacterales Campylobacteraceae *Campylobacter fetus subsp. fetus* 82-40
Cje-1-1 to Cje-1-1	ε-proteobacteria Campylobacterales Campylobacteraceae *Campylobacter jejuni* RM1221
Cje-2-1 to Cje-2-1	ε-proteobacteria Campylobacterales Campylobacteraceae *Campylobacter jejuni subsp. doylei* 269.97
Cje-3-1 to Cje-3-1	ε-proteobacteria Campylobacterales Campylobacteraceae *Campylobacter jejuni subsp. jejuni* 260.94
Cje-4-1 to Cje-4-2	ε-proteobacteria Campylobacterales Campylobacteraceae *Campylobacter jejuni subsp. jejuni* 81-176
Cje-5-1 to Cje-5-1	ε-proteobacteria Campylobacterales Campylobacteraceae *Campylobacter jejuni subsp. jejuni* 84-25
Cje-6-1 to Cje-6-1	ε-proteobacteria Campylobacterales Campylobacteraceae *Campylobacter jejuni subsp. jejuni* CF93-6
Cje-7-1 to Cje-7-1	ε-proteobacteria Campylobacterales Campylobacteraceae *Campylobacter jejuni subsp. jejuni* HB93-13
Cje-8-1 to Cje-8-1	ε-proteobacteria Campylobacterales Campylobacteraceae *Campylobacter jejuni subsp. jejuni* NCTC 11168
Cla-1-1 to Cla-1-1	ε-proteobacteria Campylobacterales Campylobacteraceae *Campylobacter lari* RM2100
Cup-1-1 to Cup-1-1	ε-proteobacteria Campylobacterales Campylobacteraceae *Campylobacter upsaliensis* RM3195
Hac-1-1 to Hac-1-1	ε-proteobacteria Campylobacterales Helicobacteraceae *Helicobacter acinonychis str.* Sheeba
Hhe-1-1 to Hhe-1-1	ε-proteobacteria Campylobacterales Helicobacteraceae *Helicobacter hepaticus* ATCC 51449
Hpy-1-1 to Hpy-1-1	ε-proteobacteria Campylobacterales Helicobacteraceae *Helicobacter pylori* 26695
Hpy-2-1 to Hpy-2-1	ε-proteobacteria Campylobacterales Helicobacteraceae *Helicobacter pylori* HPAG1
Tde-1-1 to Tde-1-1	ε-proteobacteria Campylobacterales Helicobacteraceae *Thiomicrospira denitrificans* ATCC 33889
Wsu-1-1 to Wsu-1-1	ε-proteobacteria Campylobacterales Helicobacteraceae *Wolinella succinogenes* DSM 1740

abbrev	RefSeq accession		5′ at	3′ at	genes
Hpy-2-1	NC_008086.1	-	1226386	1226207	**RNA**→ PurD (COG0151)→ RDD (pfam06271)→ Imp (COG1452)→ COG1926 (COG1926)→ Pnp (COG1185)→
Hpy-1-1	NC_000915.1	-	1295565	1295386	**RNA**→ PurD (COG0151)→ hypo→ Imp (COG1452)→ hypo→ COG1926 (COG1926)→ Pnp (COG1185)→
Hac-1-1	NC_008229.1	-	1390725	1390548	**RNA**→ PurD (COG0151)→ hypo→ Imp (COG1452)→ COG1926 (COG1926)→ Pnp (COG1185)→
Hhe-1-1	NC_004917.1	-	579160	578967	**RNA**→ PurD (COG0151)→ RDD (pfam06271)→ Imp (COG1452)→ COG1926 (COG1926)→ Pnp (COG1185)→
Wsu-1-1	NC_005090.1	+	844000	844211	**RNA**→ PurD (COG0151)→ RDD (pfam06271)→ Imp (COG1452)→ Ttg2C (COG1463)→ COG1926 (COG1926)→ Pnp (COG1185)→
Ccu-1-1	NZ_AARA01000020.1	+	17545	17732	**RNA**→ PurD (COG0151)→ RDD (pfam06271)→ Imp (COG1452)→ COG1926 (COG1926)→ Pnp (COG1185)→
Cco-2-1	NZ_AAQZ01000001.1	-	131159	130972	**RNA**→ PurD (COG0151)→ RDD (pfam06271)→ Imp (COG1452)→ COG1926 (COG1926)→ Pnp (COG1185)→
Cfe-1-1	NZ_AANR02000001.1	-	1253272	1253081	**RNA**→ PurD (COG0151)→ RDD (pfam06271)→ Imp (COG1452)→ COG1926 (COG1926)→ Pnp (COG1185)→
Cje-2-1	NZ_AARB01000012.1	-	8787	8605	**RNA**→ PurD (COG0151)→ RDD (pfam06271)→ Imp (COG1452)→ Pnp (COG1185)→
Cje-8-1	NC_002163.1	+	1179610	1179792	**RNA**→ PurD (COG0151)→ RDD (pfam06271)→ Imp (COG1452)→ Pnp (COG1185)→
Cje-6-1	NZ_AANJ01000010.1	+	29408	29590	**RNA**→ PurD (COG0151)→ RDD (pfam06271)→ Imp (COG1452)→ Pnp (COG1185)→
Cje-7-1	NZ_AANQ01000001.1	-	423569	423387	**RNA**→ PurD (COG0151)→ RDD (pfam06271)→ Imp (COG1452)→ Pnp (COG1185)→
Cje-5-1	NZ_AANT02000001.1	-	475699	475517	**RNA**→ PurD (COG0151)→ RDD (pfam06271)→ Imp (COG1452)→ Pnp (COG1185)→
Cje-1-1	NC_003912.7	+	1300389	1300571	**RNA**→ PurD (COG0151)→ RDD (pfam06271)→ Imp (COG1452)→ Pnp (COG1185)→
Cje-3-1	NZ_AANK01000001.1	-	129510	129328	**RNA**→ PurD (COG0151)→ RDD (pfam06271)→ Imp (COG1452)→ Pnp (COG1185)→
Cje-4-1	NZ_AANY01000002.1	+	30153	30335	**RNA**→ PurD (COG0151)→ RDD (pfam06271)→ Imp (COG1452)→ Pnp (COG1185)→
Cje-4-2	NZ_AASL01000001.1	+	1180898	1181080	**RNA**→
Cco-1-1	NZ_AAFL01000002.1	-	145403	145219	**RNA**→ PurD (COG0151)→ RDD (pfam06271)→ Imp (COG1452)→ Pnp (COG1185)→
Cup-1-1	NZ_AAFJ01000002.1	+	15693	15872	**RNA**→ PurD (COG0151)→ RDD (pfam06271)→ Imp (COG1452)→ Pnp (COG1185)→
Cla-1-1	NZ_AAFK01000006.1	+	40242	40434	**RNA**→ PurD (COG0151)→ RDD (pfam06271)→ Imp (COG1452)→ COG1926 (COG1926)→
Tde-1-1	NC_007575.1	+	424828	425074	**RNA**→ PurD (COG0151)→ RDD (pfam06271)→ Imp (COG1452)→ COG1926 (COG1926)→ Pnp (COG1185)→

COG0151 Phosphoribosylamine-glycine ligase [Nucleotide transport and metabolism]
COG1185 Polyribonucleotide nucleotidyltransferase (polynucleotide phosphorylase) [Translation, ribosomal structure and biogenesis]
COG1452 Organic solvent tolerance protein OstA [Cell envelope biogenesis, outer membrane]
COG1463 ABC-type transport system involved in resistance to organic solvents, periplasmic component [Secondary metabolites biosynthesis, transport, and catabolism]
COG1926 Predicted phosphoribosyltransferases [General function prediction only]
pfam06271 RDD family.

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Cje-3-1, Cje-4-2, Cje-5-1, Cje-6-1, Cje-7-1

```
         alignment positions 1···157

Hpy-2-1  CAAGUUCGUUGAAAAAAUAAGUUAG..GUUAAGGU..AGU.GGUUUU..UGUGUGAU.GGCA.UUU.............U.....UU.........AAAUCUGCC.CCUAAG..AAAAUC.C.UCCUAGCCUUAUUAGUGUAGGUU.UGU.UG....
Hpy-1-1  CAAGUUCGUUGAAAAAAUAAGUUAG..GCUAAGGU..AGU.GGUUUU..UGUGUGAU.GGCA.UUU.............U.....UU.........AAAUCUGCC.CCUAAG..AAAAUC.C.UCCUAGCCUUAUUAGUGUAGGUU.UGU.UG....
Hac-1-1  CAAGUUCGUUGAAAAAAUAAGUUAG..GCUAAGGU..AGU.GGUUUU..UGUGUGAU.GGCA.UUU.............U.....UU.........AAAUCUGCC.CCUAAU..AAGAUU.C.UCCUAGCCUUAUUGGUGUAGGUU.AU..UG....
Hhe-1-1  UAGUCUGUUACUAAAAGAUUGUUUGGGUUAAUCGU..AGU.AAUCU..UUGUGUGU.CGGUGUUUGCGCUUG.AGCUA.CUUUG.AGA.UAGUGAGCAUU.CGCCGCCUAAUG..AGAUU.C.UCUUAACCCUAUUU..GUGGCUU.UU..UU....
Wsu-1-1  UGGGGUCCGACAAUCUAAGUAUAG.UCAGGCAGGU..AGAUGAUCU...UGUGUGUUCGGUGUUUCCGUUUGUAGCCU.C......UA.AGGUGAGGAUU.UGCCGCCU.AAU..AGGUCGC.UCUUGUCUGAGUUG..UUGGCUU.UUUUUG....
Ccu-1-1  ACGGGGUCCAACACUUUUAUGUUAGCGAAGAC.GU..AC.AGGUGC...UAUGCG..AUGUGAUUUCGUUUG.AGUGC.G.....AAA.GUGCGAGAGAU.UGCAACCUUGAG..GCGCUUG.UCUAUUUGCAAGA...UUGGCUU.UG..UUUAC.
Cco-2-1  ACGGGGUCCAACACUUUUUUGUUAG.CAAAAAGGU..AU..GGGUGCC..UUGCGA..UGUGGCUUCGUUUG.AGCCU.U.....AAACA.GCGAGAAGU.UGCAACCUUUU.GGUAUCC.A.UCUAUUUGCAAGA...UUGGCUU.UG..UUUAU.
Cfe-1-1  UGAGGGUCCAACACAUUAGUAUGGGCGGAAAA.GU..AC..GCAAA..UGGUGUGA.U.GCGAUCUCGUUUGUAGCCA.U.....UUU.UAGCGAGAGAU.UGCGACCUAAAG..UUUGA.G.UCUAUCUGCAAGG...UUGGCUUUUUA.UU....
Cje-2-1  AAGUGAUCCAACACAUUUUUAUUAACUGCGAUUGU..GU..GAUUU...ACCGUGUUCUGUGGCAUCGUUUG.AGCUUUG.....AAAAAAGCGAGAAGU.UGCAGCCUUUAA..AAAUU.A.CCUAGCGUUUUUCU..UUGACUU.UU..UG....
Cje-8-1  AAGUGAUCCAACACAUUUUUAUUAACUGCGAUUGU..GU..GAUUU...ACCGUGUUCUGUGGCAUCGUUUG.AGCUUUG.....AAAAAAGCGAGAAGU.UGCAGCCUUUAA..AAAUU.A.CCUAGCGGUUUUCU..UUGACUU.UU..UG....
Cje-1-1  AAGUGAUCCAACACAUUUUUAUUAACUGCGAUUGU..GU..GAUUU...ACCGUGUUCUGUGGCAUCGUUUG.AGCUUUG.....AAAAAAGUGAGAAGU.UGCAGCCUUUAA..AAAUU.A.CCUAGCGGUUUUCU..UUGACUU.UU..UG....
Cje-4-1  AAGUGAUCCAACACAUUUUUAUUAACUGCGAUUGU..GU..GAUUU...ACCGUGUUCUGUGGCAUUGUUUG.AGCUUUG.....AAAAAAGCGAGAAGU.UGCAGCCUUUAA..AAAUU.A.CCUAGCGGUUUUCU..UUGACUU.UU..UG....
Cco-1-1  AAGUGAUCCAACAUAUUUUUAUUAACUGUGAUUGU..GC..GAUUU...ACCGUGUUCUGUGGCGUCGUUUG.AGUUUCG.....AAAGAAGCGAGAAGU.UACAGCCUUUAA..AAAUC.G.CCUAACGGUUUUCU..UUGACUU.UU..UG....
Cup-1-1  AAGUGAUCCAACACAUUAUUACUAGUUGCGAGUGU..GC..GAUUU...ACCGUGU.CUGUGGCGUCGUUUG.AGUUUCG.....AAAGAAGCGAGAAGU.UGCAGCCUUUAA..AAAUU.G.CCUAGCG.ACUUUU..UUGACUU.UU..UG....
Cla-1-1  GAGGAUCCAACACAUUAAUUAUAAGCUGUGA.UGU..GU..GAUUU...ACCGUGUUCUGUGACUUCGUUUG.AGUUUUG.....AAAAAAGCGAGAAGC.UGCAGCCUUUAA..AAAUC.A.CCUAAUGGCAUACU..UUGACUU.UU..UGA...
Tde-1-1  AUGAGGGUUCUACAUUUUCAAUUAG.CGAACUAGUA.AU..GCUCU...UGUGUGU.CGGUGCUAUCGUUUG.AGAUA.UGUUUACUC.UGUCGAGAUAA.UGCCGCCGAUAA..AGAGC.ACUCUCUUCGCCCAAAU.ACGGCUU.UU..AGAACC
         .....................................<<.<<<<<<<<.........<<<<<.<<<<......<<<<<..........>>>>>..>>>>..>>>>>.......>>>>>>>>>.>...............<<<<<<<<..........
         .....................................22.2?222220.........??022.??21......12?2?..........?2?21..12??..220??.......022222?22.2...............??222110..........
         ●●G●●●YYY●●●A●●●●●●●RU●RG○○R●RA●○GU--A●--GRUYY---●●UGUG●-○○GUG○Y●UCGYUUG-AG○Y○-●-----○●○-○○○YGAGARRU-UGC●RCCU○●●○--RRRYY-●-UCU●●●Y●Y●●●●○--●UGGCUU-UU--U●----

         alignment positions 158···280

Hpy-2-1  .AACUUAUGCGCA.UGUAAAAGUAGCGGUCAUCUGGGUUUAAAAACAACCACAUUGU...............................................AGGAAAGAC....AUGAAA
Hpy-1-1  .AACUUAUGCGCA.UGUAAAAGUAGCGGUCAUCUGGGUUUAAAAACAACCACAUUGC...............................................AGGAAAGAC....AUGAAA
Hac-1-1  .AACUUAUGCACA.UGUAAAAGUAGCGGUCAUCUGGGUUUAAAA.CAACCACAUUUU...............................................AGGGAAGAC....AUGAAA
Hhe-1-1  .AGGCUAC...................GCAAAUCGGGGUAACGCUCAUUUGGGGAUUUGUAAGAUA......................................AGG..A..UUAUAAUGCAA
Wsu-1-1  .GGGCCAC...................GCAACUUGGUAACGCCCGCUCGGGCAUCUCCUUGACCAUCCAUACGACCAAAAUUCAUC...................GG..AG.UUUUCAUGAAU
Ccu-1-1  .GGGCCACUCUU..UUAUGCG...ACGCUCACUUGGGUUUUUAAAUUUAC.......................................................GG..AGA.CAAUAUGAAC
Cco-2-1  .GGGCCACUCUU..CUAUGCG...ACGCCCACUUGGGUUUUUAAAUUUAC.......................................................GG..AGA.UAAUAUGAAC
Cfe-1-1  AGGGCCGAUUCUACUGCGUAA....CGGCCAUUUGGGUUUUUAUUUAAACUCU...................................................AGG..AAA..AAUAUGAAU
Cje-2-1  .GGGUCAAUCUU..GGUUAAA....CGGCUGUUUGGGUUUAAA.............................................................AGG..AGAUAUAAAUGAAA
Cje-8-1  .GGGUCAAUCUU..UGUUAAA....CGGCUGUUUGGGUUUAAA.............................................................AGG..AGAUAUAAAUGAAA
Cje-1-1  .GGGUCAAUCUU..UGUUAAA....CGGCUGUUUGGGUUUAAA.............................................................AGG..AGAUAUAAAUGAAA
Cje-4-1  .GGGUCAAUCUU..UGUUAAA....CGGCUGUUUGGGUUUAAA.............................................................AGG..AGAUAUAAAUGAAA
Cco-1-1  .GGGUCAAUCUU..UAUUAAG....CGGCUGUUUGGGUUUUAU.ACA.........................................................AGG..AGA.AUAAAUGAAA
Cup-1-1  .GGGUCAAUCUU..UGUAAAA....CGGCUACUUGGGUUUUUA.............................................................AGG..AGA.AAUAAUGAAG
Cla-1-1  .GGGUCAG....ACACUUAUACUC.CGGCCAUUUGGGUUUUUGUAAUAUUUUUA..................................................AGG..AAA..AUAAUGAAA
Tde-1-1  .AAGCCAA.....CUGCGAGA....CGGCUACUCGGGCUAUUAUCAAAGCGUAUUUACUGAAAAUAUUUUGAUGUUUUGAGUAGAUAUUUUUUAACUUAUUUUU.GG..AG.CUUUAAUGAAG
         >>>>>>>>...................................................................................................................
         011222??...................................................................................................................
         -RRGYYA●○○○○--○○○○○○R----CGGCYAYUUGGGYUY●YR○○○○○○○○○○○○○○-----------------------------------------------○GG--AR○○○○○○AUGAA●
                                                                                                                 SD++++++     AUG
```

2.11 The 6C motif

6C is typically moderately far (100-300 nucleotides) from the start codon, so the start codon is not shown with the motif.

Notes:

[1] Unusual gene context (downstream gene points wrong way), though RNA fits consensus reasonably well. May be false positive.

[2] Appears to overlap 5′ coding part of type II secretion gene ZP00619578.1, but BLAST shows that this part of the gene (actually the first 1 Kb) is not conserved.

[3] Appears to overlap 5′ coding part of hypothetical gene NP736907.1, but BLAST shows this part of gene is not conserved.

abbrev. of hits	**taxonomy of species**
Cdi-1-1 to Cdi-1-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Corynebacteriaceae *Corynebacterium diphtheriae* NCTC 13129
Cef-1-1 to Cef-1-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Corynebacteriaceae *Corynebacterium efficiens* YS-314
Cgl-1-1 to Cgl-1-2	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Corynebacteriaceae *Corynebacterium glutamicum* ATCC 13032
Mfl-1-1 to Mfl-1-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium flavescens* PYR-GCK
Msp-1-1 to Msp-1-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium sp.* JLS
Msp-2-1 to Msp-2-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium sp.* KMS
Msp-3-1 to Msp-3-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium sp.* MCS
Mva-1-1 to Mva-1-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium vanbaalenii* PYR-1
Mav-1-1 to Mav-1-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium avium subsp. paratuberculosis* K-10
Mbo-1-1 to Mbo-1-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium bovis* AF2122/97
Mtu-1-1 to Mtu-1-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium tuberculosis* C
Mtu-2-1 to Mtu-2-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium tuberculosis* CDC1551
Mtu-3-1 to Mtu-3-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium tuberculosis* F11
Mtu-4-1 to Mtu-4-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium tuberculosis* H37Rv
Nfa-1-1 to Nfa-1-2	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Nocardiaceae *Nocardia farcinica* IFM 10152
Rsp-1-1 to Rsp-1-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Nocardiaceae *Rhodococcus sp.* RHA1
Ace-1-1 to Ace-1-1	Actinobacteria Actinobacteridae Actinomycetales Frankineae Acidothermaceae *Acidothermus cellulolyticus* 11B
Fal-1-1 to Fal-1-1	Actinobacteria Actinobacteridae Actinomycetales Frankineae Frankiaceae *Frankia alni* ACN14a
Kra-1-1 to Kra-1-1	Actinobacteria Actinobacteridae Actinomycetales Frankineae Kineosporiaceae *Kineococcus radiotolerans* SRS30216
Jsp-1-1 to Jsp-1-1	Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Intrasporangiaceae *Janibacter sp.* HTCC2649
Str-1-1 to Str-1-1	Actinobacteria Actinobacteridae Actinomycetales Micromonosporineae Micromonosporaceae *Salinispora tropica* CNB-440
Sav-1-1 to Sav-1-1	Actinobacteria Actinobacteridae Actinomycetales Streptomycineae Streptomycetaceae *Streptomyces avermitilis* MA-4680
Sco-1-1 to Sco-1-1	Actinobacteria Actinobacteridae Actinomycetales Streptomycineae Streptomycetaceae *Streptomyces coelicolor* A3(2)
Tfu-1-1 to Tfu-1-1	Actinobacteria Actinobacteridae Actinomycetales Streptosporangineae Nocardiopsaceae *Thermobifida fusca* YX
mac-1-1 to mac-1-1	Actinobacteria marine actinobacterium PHSC20C1

abbrev	RefSeq accession		5′ at	3′ at	genes
[1]Nfa-1-1	NC_006361.1	-	54340	54212	**RNA**→
Str-1-1	NZ_AATJ01000003.1	-	329719	329592	**RNA**→ hypo→ CpaF (COG4962)→ hypo→ type II secretion system protein→ hypo (cons)→ hypo→ hypo→
Jsp-1-1	NZ_AAMN01000001.1	-	855459	855340	**RNA**→ hypo→ CpaF (COG4962)→ hypo→ hypo→
Tfu-1-1	NC_007333.1	-	153005	152904	**RNA**→ putative septum site determining protein→ CpaF (COG4962)→ TadB (COG4965)→ putative integral membrane protein→
Ace-1-1	NZ_AAOL01000002.1	-	345966	345854	**RNA**→ putative septum site determining protein→ CpaF (COG4962)→ TadB (COG4965)→ putative integral membrane protein→
Rsp-1-1	NC_008268.1	+	4571438	4571555	**RNA**→ hypo→ CpaF (COG4962)→ hypo→ hypo→ hypo→ hypo→ hypo→
Nfa-1-2	NC_006361.1	+	368533	368649	**RNA**→ hypo→ CpaF (COG4962)→ hypo→ hypo→ hypo→
Mtu-2-1	NC_002755.2	-	4091740	4091621	**RNA**→ hypo→
Mtu-4-1	NC_000962.2	-	4099482	4099363	**RNA**→ hypo→
Mtu-3-1	NZ_AAIX01000046.1	-	4469	4350	**RNA**→ hypo→ CpaF (COG4962)→ TadB (COG4965)→ COG2064: Flp pilus assembly protein TadC→ hypo→ hypo→
Mtu-1-1	NZ_AAKR01000113.1	-	5086	4967	**RNA**→ hypo→ CpaF (COG4962)→ TadB (COG4965)→ COG2064: Flp pilus assembly protein TadC→ hypo→ hypo→ hypo→
Mbo-1-1	NC_002945.3	-	4036829	4036709	**RNA**→ hypo→
Mav-1-1	NC_002944.2	+	439643	439762	**RNA**→ hypo→
Mfl-1-1	NZ_AAPA01000025.1	-	2695	2578	**RNA**→ hypo (cons)→ CpaF (COG4962)→
Mva-1-1	NZ_AAPF01000001.1	+	1361895	1362011	**RNA**→ COG0455 (COG0455)→ CpaF (COG4962)→
Msp-3-1	NC_008146.1	-	5091214	5091098	**RNA**→ hypo→ CpaF (COG4962)→ TadB (COG4965)→ type II secretion system protein→ hypo→
Msp-2-1	NZ_AAQD01000022.1	+	49134	49250	**RNA**→ hypo (cons)→ CpaF (COG4962)→ TadB (COG4965)→ type II secretion system protein→ hypo (cons)→
Msp-1-1	NZ_AAQC01000003.1	-	2777	2660	**RNA**→ hypo (cons)→ CpaF (COG4962)→
Sco-1-1	NC_003888.3	-	3933621	3933505	**RNA**→ septum site determining protein→ CpaF (COG4962)→ integral membrane protein→ integral membrane protein→ small membrane protein→ hypo→ membrane spanning protein→
Sav-1-1	NC_003155.3	+	5623199	5623315	**RNA**→ septum site determining protein→ CpaF (COG4962)→
[2]Kra-1-1	NZ_AAEF02000159.1	+	2080	2202	**RNA**→ CpaF (COG4962)→
Fal-1-1	NC_008278.1	+	7144284	7144409	**RNA**→ ParA (cd02042)→ CpaF (COG4962)→ hypothetical protein; putative signal peptide→ hypo→ hypo→ hypothetical protein; putative TONB Box domain→
mac-1-1	NZ_AAOB01000010.1	+	29188	29304	**RNA**→ CpaF (COG4962)→ putative integral membrane protein→
Cdi-1-1	NC_002935.2	+	278513	278629	**RNA**→ Putative morphological differentiation-related protein→ CpaF (COG4962)→ TadB (COG4965)→ Putative secreted protein→ hypo→ Putative secreted protein→ Putative secreted protein→
Cgl-1-1	NC_003450.3	+	314679	314804	**RNA**→ hypo→ CpaF (COG4962)→ TadB (COG4965)→ TadC (COG2064)→ hypo→ hypo→ hypo→
Cgl-1-2	NC_006958.1	+	314679	314804	**RNA**→ PUTATIVE SEPTUM SITE DETERMINING PROTEIN→ CpaF (COG4962)→ TadB (COG4965)→ TadC (COG2064)→ hypo→ conserved secreted protein→ conserved hypothetical secreted protein→
[3]Cef-1-1	NC_004369.1	+	322055	322178	**RNA**→ hypo→ CpaF (COG4962)→ TadB (COG4965)→ hypo→

cd02042 ParA and ParB of Caulobacter crescentus belong to a conserved family of bacterial proteins implicated in chromosome segregation.

COG0455 ATPases involved in chromosome partitioning [Cell division and chromosome partitioning]

COG2064 Flp pilus assembly protein TadC [Cell motility and secretion / Intracellular trafficking and secretion]

COG4962 Flp pilus assembly protein, ATPase CpaF [Intracellular trafficking and secretion]

COG4965 Flp pilus assembly protein TadB [Intracellular trafficking and secretion]

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Cgl-1-2, Msp-2-1, Mtu-1-1, Mtu-3-1, Mtu-4-1

alignment positions 1 · · · 160

```
[1]Nfa-1-1  GCA...CAAUC.....GCGUCUUAUCGCGUAG.....GAUCGUUCACG.UCGGCCCCU........CCCCCCC...GGGC.CGACCU..UAC....GCGGGCCUCGGC.UAACUCCCCCCCUUC..GCCGAGGCCCGCUCCAUGUCCAGGGUCGGUCCGG
Str-1-1     CCU...CGGCGGUUGCACCCCCCCCC......GUGACCGCUGAGUGGU.UCGGCUCAC........CCCCCCCG..GAGC.CGGACC..CC.....GACGACCCCCGUCU...UCCCCCC.....GACGGGGGUCGUCCCUUCUUCUCCUGUUGUCGGCC
Jsp-1-1     CCC...CCUCG.....GGGUUGCAG............CGACGGGUGGU.CCGGCCUCU........CCCCCCCG..GGGCUCGGCCC..GCCC...GACGGCCCCGGUCA...UCCCCCC.....GACCGGGGCCGUCGCCUUUUCUUUUUCAGGUGCCG
Tfu-1-1     GGA.............UA........................CGGG.A.CCGACUCUU........CCCCCCG...GAGU.CGGACCG.UCG....GACGGCCCCCGCCUU..UCCCCCCU....GGCGGGGGCCGUCCCCUUGGACGCCGGAGCUUUCC
Ace-1-1     GGA...AGCGU.....UACC.................GUGUUGGGGGU.GCGGCUCUU........CCCCCCA...GAGC.CGUACCGGGC.....AGUGGCCCCCGCU....ACCCCCCC....GGCGGGGGCCACUCCUUUUGUGCGCGUUUUCGCCG
Rsp-1-1     AUG...CCCGGA....UUUGA...............UCCGGGCGAUGU.UCAGCCCGA........CCCCCCG...GGGC.UGAACA..CU.....GACGACCCUCGCCUCC.UCCCCCCCU...GGCGAGGGUCGUCCUCUGUCUGCGGCGGGUCUGCC
Nfa-1-2     GCA...CCUGGA....CAUGU...............UCCAGGCGUG.U.UCAGCCCGA........CCCCCCG...GGGC.UGAACC..CC.....GACGGCCCCAGCCUCC.UCCCCCCCC...GGCUGGGGCCGUCCUCUAUUUCCGGCCGGUGAUAU
Mtu-2-1     GCAUUACUCGGC....CAUA................GCCGAGGGUGUA.UCAGUCCGA........CCCCCCG...GGGC.UGAUGC..AC.....GACGACCCUCGCCUCC.UCCCCCCCU...GGCGGGGGUCGUCCCUUUUGGCAGCCACGCAAACG
Mbo-1-1     GCAUUACUCGGC....CAUA................GCCGAGGGUGUA.UCAGUCCGA........CCCCCCG...GGGC.UGAUGC..AC.....GACGACCCUCGCCUCC.UCCCCCCCCU..GGCGGGGGUCGUCCCUUUUGGCAGCCACGCAAACG
Mav-1-1     GCAUUACUCGGC....CUGA................GCCGAGGGUGUA.UCAGCCCGA........CCCCCCG...GGGC.UGAUGC..AC.....GACGACCCUCGCCUCC.UCCCCCCCU...GGCGGGGGUCGUCCCUUUUCUCGGGUCUUUCCGGG
Mfl-1-1     UUA...CUCGGC....UUAU................GCCGAGUGUGCA.UCAGCCCGA........CCCCCCG...GGGC.UGAUAC..AC.....GACGACCUCCGCCUCC.UCCCCCCCCU..GGCGGGGGUCGUCCCUUUUGUGGGGUGUUUUCGUC
Mva-1-1     UUA...CUCGGC....UUAU................GCCGGGUGUGCA.UCAGCCCGA........CCCCCCG...GGGC.UGAUAC..AC.....GACGACCUCCGCCUCC.UCCCCCCCU...GGCGGGGGUCGUCCCUUUUCUGGGGUGUUUCGCCC
Msp-3-1     UUA...CCCGG.....GUUAUG...............CCGGGUGUGCA.UCAGCCCGA........CCCCCCG...GGGC.UGAUAC..AC.....GACGACCUCCGCCUCC.UCCCCCCCU...GGCGGGGGUCGUCCUUUUUCGGGGUGUUGCAUUC
Msp-1-1     UUA...CCCGG.....GUUAUG...............CCGGGUGUGCA.UCAGCCCGA........CCCCCCG...GGGC.UGAUAC..AC.....GACGACCUCCGCCUCC.UCCCCCCCCU..GGCGGGGGUCGUCCUUUUUCGGGGUGUUGCAUUC
Sco-1-1     UCG...UUCGC.....GACCGU...............GCUAA.CCCGGUCCGACUCCU........CCCCCCCC..GAGU.CGGCCGU.GG....AGACGACCCCCGC.UC..UCCCCCCC....GGCGGGGGUCGUCGCAUGUCCGGGUGGGUUUUCCG
Sav-1-1     UCG...UUCGC.....GACCGU...............GCUAA.CCCGGUCCGACUCCU........CCCCCCCC..GAGU.CGGCCGU.GG....GGACGACCCCCGC.UC..UCCCCCCC....GGCGGGGGUCGUCGCAUGUCCGGAUGGGUUUUCCU
[2]Kra-1-1  GGA...UGCAGC....AGUACGCCCUGUGG......GCUACACGCGGC.CCGGCUCAU........CCCCCCC...GAGC.CGGACCG.UC.....GACGGCCCCCGC.....UCCCCCCCA....GCGGGGGCCGUCUCCACUCCUCCUCCGGUCGCAC
Fal-1-1     UGG...UGGCC.....CUGA.................GGCCACGAGAC.CCGGCCCGAGACGGGCAGCCCCCAUACGGGC.CAUGUC..CC.....GAGGCGACCCCCGCAC.UCCCCCCUG...UGGGGGUCGCCUCCCCCAGACCGUGCCUCUCCGGU
mac-1-1     ACG...CCGAA.....UGCAAG...............UUCGGCACGCG.GCAAUUUUGAUU.....CCCCCCAAU.CAAU.UGCCGCG.GU.....GGCGGUGCCUAU.....UCCCCCCA.....AUGGGCGCCGCCCCUCUCUCGUUAAAGCGUCGCU
Cdi-1-1     GCA...AGGCC.....CCGAUAUACAGUGU.......GGCCUG.....CCCCGGCCCG........CCCCCCUGU.GGCC.GGG...........UGGCGGUUCGCGC.AC..ACCCCCCC.GA..GCGCGAACCGCCCUUAAUUGUUUUAUGGGGUGUU
Cgl-1-1     GCA...AGGCC.....CCGAUAUACAGUGU.......GGCCUG.....CCCCGGCACAAUCC....CCCCCCGU..UGCC.GGG.......UUACUGACGGCCCGCGC.AC..ACCCCCCCCGA.GGCGCGGGCCGUCCCUUAUUUUCUAAACUAGUGCC
[3]Cef-1-1  GCA..AGGCCCCG...AUAUACUGUG.........C.GGCCUG.....CCCCGGCCACAC......CCCCCCGA..GGUC.GGG.......UUACCGACGGCCCGCGC.AC..ACCCCCCCCGA.GGCGCGGGCCGUCCCCAUUUCCACCCCUCGUCCAU
            ......<<<<<<<<<<................>>>>>>>>>>.<<<<<.<<<<<<<....................>>>>.>>>>>>..>>.....<<<<<<<<<<<<<................>>>>>>>>>>>>>......................
            ......2??22?2000................0002?22??2.2?2??.??2222?....................?222.2????2..?2.....222022222222?................?222222220222......................
            ●●R---●●●●●-----●●●●○○---------------●●●●●●●●●○●-YCRGYYC●●--------CCCCCC●○--GRGC-YGR●●●--●●-----GRCGRCCCYYGC○●○--UCCCCCCY○---○GCGGGGGYCGYCYCYY●Y●Y●●●●●●●●●Y●●●●
```

alignment positions 161 · · · 164

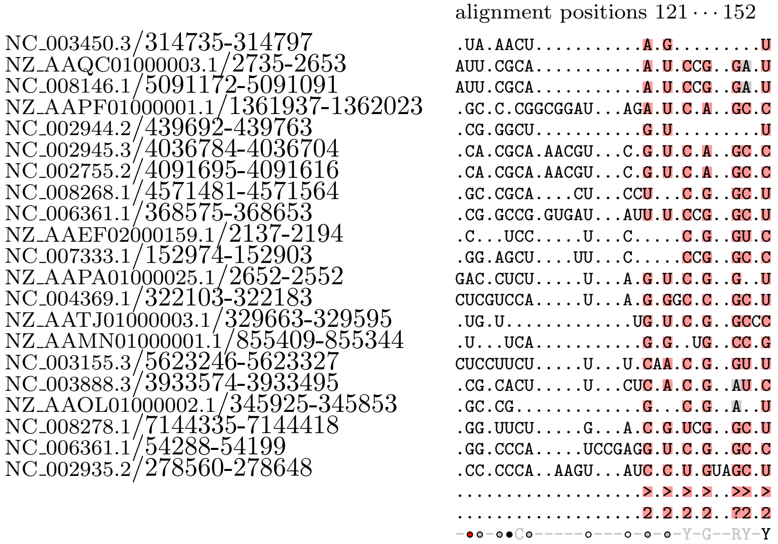
```
[1]Nfa-1-1  CCCA
Str-1-1     CAGG
Jsp-1-1     UUUU
Tfu-1-1     CGGC
Ace-1-1     GCGA
Rsp-1-1     GCAC
Nfa-1-2     UUCC
Mtu-2-1     UCGU
Mbo-1-1     UCGU
Mav-1-1     CUGU
Mfl-1-1     ACGC
Mva-1-1     GGCG
Msp-3-1     GCAA
Msp-1-1     GCAA
Sco-1-1     CACU
Sav-1-1     CCUU
[2]Kra-1-1  ACCG
Fal-1-1     UCUG
mac-1-1     GUGU
Cdi-1-1     UUCC
Cgl-1-1     CUUA
[3]Cef-1-1  AGGG
            ....
            ....
            ●Y●●
```

2.11.1 Possible alternate structure surrounding P2 stem

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: , , , ,

alignment positions 1 · · · 120

```
NC_003450.3/314735-314797          G.......U.U..A.C..U.....GA..CGGCCCGCGC.ACAC.CCCCCCC.GAG.GCGCGGGCCG..UC.CCUU.......A...........U..UUU...................C
NZ_AAQC01000003.1/2735-2653        G.G.CUG.A.U..A.C..A...C.GA..CGACCUCCGCCUC.CUCCCCCCCCU..GGCGGGGGUCG..UC.CCUU.......U...UUCG.GGGU.GUUG...................C
NC_008146.1/5091172-5091091        G.G.CUG.A.U..A.C..A...C.GA..CGACCUCCGCCUC.CUCCCCCCC.U..GGCGGGGGUCG..UC.CCUU.......U...UUCG.GGGU.GUUG...................C
NZ_AAPF01000001.1/1361937-1362023  G.G.CUG.A.U..A.C..A...C.GA..CGACCUCCGCCUC.CUCCCCCCC.U..GGCGGGGGUCG..UC.CCUU.......U...UCUG.GGGU.GUUU...................C
NC_002944.2/439692-439763          G.......A.U..G.C..A...C.GA..CGACCCUCGCCUC.CUCCCCCCC.U..GGCGGGGGUCG..UC.CCUU.......U...UCUC.GGGU.CUUU...................C
NC_002945.3/4036784-4036704        G.G.CUG.A.U..G.C..A...C.GA..CGACCCUCGCCUC.CUCCCCCCCCU..GGCGGGGGUCG..UC.CCUU.......U...UGGC.AG..........................C
NC_002755.2/4091695-4091616        G.G.CUG.A.U..G.C..A...C.GA..CGACCCUCGCCUC.CUCCCCCCC.U..GGCGGGGGUCG..UC.CCUU.......U...UGGC.AG..........................C
NC_008268.1/4571481-4571564        G.G.CUG...A..A.C..A...CUGA..CGACCCUCGCCUC.CUCCCCCCC.U..GGCGAGGGUCG..UC.C.UC.......UGUCUGCG.GCGG..GUC...................U
NC_006361.1/368575-368653          G.G.CUG.A.A..C.C..C...C.GA..CGGCCCCAGCCUC.CUCCCCCCC.C..GGCUGGGGCCG..UC.C.UC.......U...........A..UUU...................C
NZ_AAEF02000159.1/2137-2194        G.A.CCG.......UC........GA..CGGCCCCCGCU.....CCCCCCC....AGCGGGGGCCG..UCUCCAC.......U....................................C
NC_007333.1/152974-152903          G.GACCG....UCG..........GA..CGGCCCCCGCCU...UUCCCCCC.U..GGCGGGGGCCG..UC.CCCU.......U...GGAC.GC..........................C
NZ_AAPA01000025.1/2652-2552        G...CUG.A.U..A.C..A...C.GA..CGACCUCCGCCUC.CUCCCCCCCCU..GGCGGGGGUCG..UC.CCUU.......U...UGUG.GGGU.GUUUUCGUCACGCGGAUAGUUGAC
NC_004369.1/322103-322183          G.GUCGGGU.U..A.C..C.....GA..CGGCCCGCGC.ACAC.CCCCCCC.GAG.GCGCGGGCCG..UCCCCAU.......U...UCCA.CC..........................C
NZ_AATJ01000003.1/329663-329595    GAG.CCGGA.C....C..C...C.GA..CGACCCCCGUCU...UCCCCCC.....GACGGGGGUCG..UC.CCUU......CU...........U..CUC...................C
NZ_AAMN01000001.1/855409-855344    C.G.GC..C.C..G.C..C...C.GA..CGGCCCCGGUCA...UCCCCCC.....GACCGGGGCCG..UCGCCUU.......U...........U..CUU...................U
NC_003155.3/5623246-5623327        G.G.CCG.UGG..G..........GA..CGACCCCCGCU.C..UCCCCCCC....GGCGGGGGUCG..UC.GCAU......GU.CCGGAU.GGGU..UUU...................C
NC_003888.3/3933574-3933495        G.G.CCG.UGG..A..........GA..CGACCCCCGCU.C..UCCCCCCC....GGCGGGGGUCG..UC.GCAU......GU.CCGGGU.GGGU..UUU...................C
NZ_AAOL01000002.1/345925-345853    G...CCG...U..A.C..CGGGC.AG..UGGCCCCCGCUAC...CCCCCC.....GGCGGGGGCCA..CU.CCUU.......U...UGUGCGCGU..UUU...................C
NC_008278.1/7144335-7144418        G.G.CCA.U.G..U.C..C...C.GAGGCGACCCCCGCA.C..UCCCCCC.....UGUGGGGGUCGCCUC.CCCC.......A...GACC.GUGCCUCUC...................C
NC_006361.1/54288-54199            G.G.CCG.A.C....CUUA...C.GC..GGGCCUCGGC.UAACUCCCCCCC.UUC.GCCGAGGCCC..GCUCCAU.......G...UCCA.GGGUCGGUC...................C
NC_002935.2/278560-278648          G.G.CCG.G.G..U..........GG..CGGUUCGCGC.AC.A.CCCCCCC.GA..GCGCGAACCG..CC.C.UUAAUUGUUU...UAUG.GGGU.GUUU...................U
                                   <.<.<<<.<.<.............<<..<<<<<<<<<<<................>>>>>>>>>>>..>>..................................................
                                   2.2.?22.2.2.............22..20222222122................22122222202..22..................................................
                                   G-G-CYG-●-●--●-C--○---○-GA--CGRCCCYCGC○●●-○○CCCCCC○-○--○GCGGGGGYCG--UC-CC●U-------U---○○○○-○○○●--●UY-------------------C
```



alignment positions 121 · · · 152

```
NC_003450.3/314735-314797          .UA.AACU...........A.G.........U
NZ_AAQC01000003.1/2735-2653        AUU.CGCA...........A.U.CCG..GA.U
NC_008146.1/5091172-5091091        AUU.CGCA...........A.U.CCG..GA.U
NZ_AAPF01000001.1/1361937-1362023  .GC.C.CGGCGGAU...AGA.U.C.A..GC.C
NC_002944.2/439692-439763          .CG.GGCU...........G.U.........U
NC_002945.3/4036784-4036704        .CA.CGCA.AACGU...C.G.U.C.A..GC.C
NC_002755.2/4091695-4091616        .CA.CGCA.AACGU...C.G.U.C.A..GC.C
NC_008268.1/4571481-4571564        .GC.CGCA....CU...CCU...C.G..GC.U
NC_006361.1/368575-368653          .CG.GCCG.GUGAU...AUU.U.CCG..GC.U
NZ_AAEF02000159.1/2137-2194        .C...UCC.....U...C.....C.G..GU.C
NC_007333.1/152974-152903          .GG.AGCU....UU...C.....CCG..GC.C
NZ_AAPA01000025.1/2652-2552        GAC.CUCU.....U...A.G.U.C.G..G..U
NC_004369.1/322103-322183          CUCGUCCA.....U...A.G.GGC.C..GC.U
NZ_AATJ01000003.1/329663-329595    .UG.U.............UG.U.C.G..GCCC
NZ_AAMN01000001.1/855409-855344    .U...UCA...........G.G..UG..CC.G
NC_003155.3/5623246-5623327        CUCCUUCU.....U...U.CAA.C.G..GU.U
NC_003888.3/3933574-3933495        .CG.CACU.....U...CUC.A.C.G..AU.C
NZ_AAOL01000002.1/345925-345853    .GC.CG.............G...C.G..A..U
NC_008278.1/7144335-7144418        .GG.UUCU.....G...A.C.G.UCG..GC.U
NC_006361.1/54288-54199            .GG.CCCA.....UCCGAGG.U.C.G..GC.C
NC_002935.2/278560-278648          .CC.CCCA..AAGU...AUC.C.U.GUAGC.U
                                   ...................>.>.>.>..>>.>
                                   ...................2.2.2.2..?2.2
                                   -●○-○●C○------○---○-○-○-Y-G--RY-Y
```

2.12 alpha-transposase motif

abbrev. of hits	**taxonomy of species**
Asp-1-1 to Asp-1-1	α-proteobacteria Rhizobiales Aurantimonadaceae *Aurantimonas sp.* SI85-9A1
Bja-1-1 to Bja-1-15	α-proteobacteria Rhizobiales Bradyrhizobiaceae *Bradyrhizobium japonicum* USDA 110
Bsp-1-1 to Bsp-1-1	α-proteobacteria Rhizobiales Bradyrhizobiaceae *Bradyrhizobium sp.* BTAi1
Rpa-1-1 to Rpa-1-8	α-proteobacteria Rhizobiales Bradyrhizobiaceae *Rhodopseudomonas palustris* BisB18
Mlo-1-1 to Mlo-1-3	α-proteobacteria Rhizobiales Phyllobacteriaceae *Mesorhizobium loti* MAFF303099
Atu-1-1 to Atu-1-2	α-proteobacteria Rhizobiales Rhizobiaceae Rhizobium/Agrobacterium group *Agrobacterium tumefaciens str.* C58
Ret-1-1 to Ret-1-1	α-proteobacteria Rhizobiales Rhizobiaceae Rhizobium/Agrobacterium group *Rhizobium etli*
Rsp-1-1 to Rsp-1-3	α-proteobacteria Rhizobiales Rhizobiaceae Rhizobium/Agrobacterium group *Rhizobium sp.* NGR234
Sme-1-1 to Sme-1-1	α-proteobacteria Rhizobiales Rhizobiaceae Sinorhizobium/Ensifer group *Sinorhizobium medicae* WSM419
Sme-2-1 to Sme-2-1	α-proteobacteria Rhizobiales Rhizobiaceae Sinorhizobium/Ensifer group *Sinorhizobium meliloti* 1021
Pde-1-1 to Pde-1-1	α-proteobacteria Rhodobacterales Rhodobacteraceae *Paracoccus denitrificans* PD1222
Rsp-2-1 to Rsp-2-11	α-proteobacteria Rhodobacterales Rhodobacteraceae *Rhodobacter sphaeroides* ATCC 17025
Acr-1-1 to Acr-1-3	α-proteobacteria Rhodospirillales Acetobacteraceae *Acidiphilium cryptum* JF-5
Rru-1-1 to Rru-1-4	α-proteobacteria Rhodospirillales Rhodospirillaceae *Rhodospirillum rubrum* ATCC 11170
Ssp-1-1 to Ssp-1-2	α-proteobacteria Sphingomonadales Sphingomonadaceae *Sphingomonas sp.* KA1
Ssp-2-1 to Ssp-2-25	α-proteobacteria Sphingomonadales Sphingomonadaceae *Sphingomonas sp.* SKA58
Sal-1-1 to Sal-1-2	α-proteobacteria Sphingomonadales Sphingomonadaceae *Sphingopyxis alaskensis* RB2256
Bxe-1-1 to Bxe-1-6	β-proteobacteria Burkholderiales Burkholderiaceae *Burkholderia xenovorans* LB400
Asp-2-1 to Asp-2-11	β-proteobacteria Rhodocyclales Rhodocyclaceae *Azoarcus sp.* EbN1
env-1	environmental sample

abbrev	RefSeq accession		5′ at	3′ at	genes
Mlo-1-1	NC_002678.2	+	4876744	4876821	COG3293 (COG3293)→ **RNA**→ rve (pfam00665)→
Sme-2-1	NC_003047.1	-	3397596	3397504	←rve (pfam00665) **RNA**→ rve (pfam00665)→
Rsp-1-1	NC_000914.1	-	383476	383381	←DdlA (COG1181) **RNA**→ rve (pfam00665)→
Rsp-1-2	NC_000914.1	-	282996	282901	←Y4nF **RNA**→ rve (pfam00665)→
Rsp-1-3	NC_000914.1	-	189705	189610	←Transposase_8 (pfam01527) **RNA**→ rve (pfam00665)→
Bja-1-1	NC_004463.1	+	1870915	1871020	Peptidase_C48 (pfam02902)→ **RNA**→ rve (pfam00665)→
Bja-1-2	NC_004463.1	-	5545647	5545542	EFh (cd00051)→ **RNA**→ rve (pfam00665)→
Bja-1-3	NC_004463.1	-	3893645	3893540	Bug (pfam03401)→ **RNA**→ rve (pfam00665)→
Bja-1-4	NC_004463.1	-	2157859	2157754	←hypo **RNA**→ rve (pfam00665)→
Bja-1-5	NC_004463.1	-	2056179	2056074	←hypo **RNA**→ rve (pfam00665)→
Mlo-1-2	NC_002678.2	-	5043561	5043457	COG3039 (COG3039)→ **RNA**→ rve (pfam00665)→
Pde-1-1	NZ_AAIT01000005.1	-	235363	235276	hypo→ **RNA**→ rve (pfam00665)→
Asp-2-1	NC_006513.1	+	437018	437094	←hypo **RNA**→ ←IstB (pfam01695)
Asp-2-2	NC_006513.1	+	3309135	3309211	←hypo **RNA**→ ←IstB (pfam01695)
Asp-2-3	NC_006513.1	-	4018059	4017983	←CypX (COG2124) **RNA**→ ←IstB (pfam01695)
Asp-2-4	NC_006513.1	-	3765238	3765162	←IstB (pfam01695) **RNA**→ ←IstB (pfam01695)
Asp-2-5	NC_006513.1	-	1952520	1952444	←hypo **RNA**→ ←IstB (pfam01695)
Asp-2-6	NC_006513.1	-	1189836	1189760	COG4453 (COG4453)→ **RNA**→ ←IstB (pfam01695)
Asp-2-7	NC_006513.1	-	225780	225704	hypo→ **RNA**→ ←IstB (pfam01695)
Asp-2-8	NC_006824.1	+	200486	200562	Transposase_mut (pfam00872)→ **RNA**→ ←IstB (pfam01695)
Asp-2-9	NC_006824.1	-	158387	158311	putative replication protein→ **RNA**→ ←IstB (pfam01695)
Asp-2-10	NC_006824.1	-	138080	138004	←Peptidase_S24 (pfam00717) **RNA**→ ←IstB (pfam01695)
Asp-2-11	NC_006824.1	-	87887	87811	←HNHc (smart00507) **RNA**→ ←IstB (pfam01695)
Atu-1-1	NC_003063.1	-	28127	28054	COG1708 (COG1708)HEPN (smart00748)→ **RNA**→ ←DnaC (COG1484)
Atu-1-2	NC_003305.1	+	2046630	2046703	COG1708 (COG1708)HEPN (smart00748)→ **RNA**→ ←DnaC (COG1484)
Acr-1-1	NZ_AAOO01000004.1	-	76651	76567	←hypo **RNA**→ rve (pfam00665)→
Acr-1-2	NZ_AAOO01000001.1	+	286345	286429	←Transposase_2 (pfam01385)Transposase_35 (pfam07282) **RNA**→ rve (pfam00665)→
Acr-1-3	NZ_AAOO01000001.1	-	355999	355915	←COG4227 (COG4227) **RNA**→ rve (pfam00665)→
Rru-1-1	NC_007641.1	+	10421	10476	←Glycos_transf_1 (pfam00534)RgpF (pfam05045) **RNA**→ putative fragment of transposase protein→
Rru-1-2	NC_007641.1	-	15435	15380	←INT_FimBE_C (cd01197) **RNA**→ putative fragment of transposase protein→
Ssp-2-1	NZ_AAQG01000001.1	+	5195	5250	tRNA-Met→ **RNA**→ Transposase_mut (pfam00872)→
Ssp-2-2	NZ_AAQG01000002.1	+	175125	175180	HTH_LUXR (smart00421)→ **RNA**→ Transposase_mut (pfam00872)→
Ssp-2-3	NZ_AAQG01000003.1	-	41183	41128	rve (pfam00665)→ **RNA**→ Transposase_mut (pfam00872)→
Ssp-2-4	NZ_AAQG01000004.1	-	26158	26103	←hypo **RNA**→ Transposase_mut (pfam00872)→
Ssp-2-5	NZ_AAQG01000004.1	-	15038	14983	rve (pfam00665)→ **RNA**→ Transposase_mut (pfam00872)→
Ssp-2-6	NZ_AAQG01000013.1	+	24577	24632	←hypo **RNA**→ Transposase_mut (pfam00872)→
Ssp-2-7	NZ_AAQG01000013.1	+	51046	51101	XynE_like (cd01830)→ **RNA**→ Transposase_mut (pfam00872)→
Ssp-2-8	NZ_AAQG01000014.1	+	63965	64020	←ligand_gated_channel (cd01347) **RNA**→ Transposase_mut (pfam00872)→
Ssp-2-9	NZ_AAQG01000018.1	-	11185	11130	ISPsy20, transposase IstB, transcriptional regulator, LysR family - fusion→ **RNA**→ Transposase_mut (pfam00872)→
Ssp-2-10	NZ_AAQG01000024.1	-	47296	47241	←Transposase_8 (pfam01527) **RNA**→ Transposase_mut (pfam00872)→
Rru-1-3	NC_007643.1	-	3348856	3348799	Protein of unknown function DUF6, transmembrane→ **RNA**→ transposase→
Ssp-2-11	NZ_AAQG01000001.1	+	5195	5285	tRNA-Met→ **RNA**→ Transposase_mut (pfam00872)→

Ssp-2-12	NZ_AAQG01000002.1	+	175125	175215	HTH_LUXR (smart00421)→ **RNA**→ Transposase_mut (pfam00872)→
Ssp-2-13	NZ_AAQG01000003.1	-	41183	41093	rve (pfam00665)→ **RNA**→ Transposase_mut (pfam00872)→
Ssp-2-14	NZ_AAQG01000004.1	-	26158	26068	←hypo **RNA**→ Transposase_mut (pfam00872)→
Ssp-2-15	NZ_AAQG01000004.1	-	15038	14948	rve (pfam00665)→ **RNA**→ Transposase_mut (pfam00872)→
Ssp-2-16	NZ_AAQG01000013.1	+	24577	24667	←hypo **RNA**→ Transposase_mut (pfam00872)→
Ssp-2-17	NZ_AAQG01000013.1	+	51046	51136	XynE_like (cd01830)→ **RNA**→ Transposase_mut (pfam00872)→
Ssp-2-18	NZ_AAQG01000014.1	+	63965	64055	←ligand_gated_channel (cd01347) **RNA**→ Transposase_mut (pfam00872)→
Ssp-2-19	NZ_AAQG01000018.1	-	11185	11095	ISPsy20, transposase IstB, transcriptional regulator, LysR family - fusion→ **RNA**→ Transposase_mut (pfam00872)→
Ssp-2-20	NZ_AAQG01000024.1	-	47296	47206	←Transposase_8 (pfam01527) **RNA**→ Transposase_mut (pfam00872)→
Sal-1-1	NC_008048.1	+	440766	440859	hypo→ **RNA**→ ←IstB (pfam01695)
Sal-1-2	NC_008048.1	+	856443	856536	←hypo **RNA**→ ←IstB (pfam01695)
Asp-1-1	NZ_AAPJ01000001.1	-	626620	626533	Transposase_25 (pfam03050)→ **RNA**→ transposase→
Rsp-2-1	NZ_AAME01000001.1	+	167776	167872	←HipA (COG3550) **RNA**→ rve (pfam00665)→
Bja-1-6	NC_004463.1	+	1764755	1764848	hypo→ **RNA**→ ←rve (pfam00665)
Bja-1-7	NC_004463.1	+	2293017	2293110	hypo→ **RNA**→ ←rve (pfam00665)
Bja-1-8	NC_004463.1	+	8330281	8330374	PorG (COG1014)COG4231 (COG4231)→ **RNA**→ ←rve (pfam00665)
Bja-1-9	NC_004463.1	-	9051687	9051594	←hypo **RNA**→ ←rve (pfam00665)
Bja-1-10	NC_004463.1	-	9028428	9028335	hypo→ **RNA**→ ←rve (pfam00665)
Bja-1-11	NC_004463.1	-	2081222	2081129	Metallophos (pfam00149)→ **RNA**→ ←rve (pfam00665)
Bja-1-12	NC_004463.1	-	1898433	1898340	←HypF (COG0068)AcyP (COG1254)HupF_HypC (pfam01455) **RNA**→ ←rve (pfam00665)
Bja-1-13	NC_004463.1	-	15937	15844	←hypo **RNA**→ ←rve (pfam00665)
Bja-1-14	NC_004463.1	-	9016772	9016679	←rve (pfam00665) **RNA**→ ←putative transposase
Bja-1-15	NC_004463.1	+	1906118	1906211	COG3415 (COG3415)rve (pfam00665)→ **RNA**→ ←rve (pfam00665)
Mlo-1-3	NC_002678.2	-	5052265	5052159	COG4977 (COG4977)→ **RNA**→ rve (pfam00665)→
Ssp-2-21	NZ_AAQG01000001.1	-	251931	251822	Ku protein→ **RNA**→ rve (pfam00665)→
Bsp-1-1	NZ_AALJ01000035.1	+	4382	4484	←IstB (pfam01695) **RNA**→ ←IstB (pfam01695)
Ssp-2-22	NZ_AAQG01000001.1	-	420998	420915	←hypo **RNA**→ ←rve (pfam00665)
Ssp-2-23	NZ_AAQG01000014.1	+	25932	26015	COG2944 (COG2944)→ **RNA**→ ←rve (pfam00665)
Ssp-2-24	NZ_AAQG01000015.1	-	955	872	Glyoxalase (pfam00903)→ **RNA**→ ←DnaC (COG1484)
Ssp-1-1	NC_008308.1	+	248587	248694	hypo→ **RNA**→ ←IstB (pfam01695)
Ssp-2-25	NZ_AAQG01000015.1	-	955	850	Glyoxalase (pfam00903)→ **RNA**→ ←DnaC (COG1484)
Rsp-2-2	NZ_AAME01000006.1	-	78528	78426	←rve (pfam00665) **RNA**→ rve (pfam00665)→
Rsp-2-3	NZ_AAME01000046.1	+	29195	29297	←tISRso9, RSp0584; ISRSO9-transposase protein **RNA**→ rve (pfam00665)→
Rsp-2-4	NZ_AAME01000056.1	+	17471	17573	IstB (pfam01695)→ **RNA**→
Rsp-2-5	NZ_AAME01000059.1	+	15569	15671	rve (pfam00665)→ **RNA**→ putative transposase→
Rsp-2-6	NZ_AAME01000065.1	+	9439	9541	hypo→ **RNA**→ putative transposase→
Rpa-1-1	NC_007925.1	+	654127	654230	hypo→ **RNA**→ rve (pfam00665)→
Rpa-1-2	NC_007925.1	+	5001678	5001781	←COG3513 (COG3513) **RNA**→ rve (pfam00665)→
Rpa-1-3	NC_007925.1	-	3393798	3393695	INT_P4 (cd00801)→ **RNA**→ rve (pfam00665)→
Rpa-1-4	NC_007925.1	-	1102506	1102403	hypo→ **RNA**→ rve (pfam00665)→
Rpa-1-5	NZ_AALR01000013.1	+	167077	167180	←COG3513 (COG3513) **RNA**→
Rpa-1-6	NZ_AALR01000014.1	+	133089	133192	similar to Phosphoenolpyruvate carboxykinase (ATP)→ **RNA**→ rve (pfam00665)→
Rpa-1-7	NZ_AALR01000021.1	+	101592	101695	hypo→ **RNA**→

Rpa-1-8	NZ_AALR01000042.1	+	13218	13321	INT_P4 (cd00801)→ **RNA**→ putative transposase→
Rru-1-4	NC_007643.1	+	2540975	2541078	←COG4231 (COG4231)PorG (COG1014) **RNA**→ putative transposase→
env-1	AACY01005623.1	+	200	301	**RNA**→ ←unknown
Ret-1-1	NC_004041.2	-	53080	52973	hypo→ **RNA**→ rve (pfam00665)→
Ssp-1-2	NC_008308.1	-	254516	254419	hypo→ **RNA**→ ←DnaC (COG1484)
Bxe-1-1	NC_007952.1	-	3240751	3240655	AtoS (COG2202)→ **RNA**→ rve (pfam00665)→
Bxe-1-2	NC_007951.1	+	794210	794306	hypo→ **RNA**→ rve (pfam00665)→
Bxe-1-3	NC_007951.1	+	2105927	2106023	UvrD-helicase (pfam00580)→ **RNA**→ rve (pfam00665)→
Bxe-1-4	NC_007951.1	+	2219944	2220040	rve (pfam00665)→ **RNA**→ rve (pfam00665)→
Bxe-1-5	NC_007951.1	-	1953011	1952915	Transposase_9 (pfam01548)→ **RNA**→ rve (pfam00665)→
Bxe-1-6	NC_007951.1	-	1434502	1434406	Transposase_25 (pfam03050)→ **RNA**→ rve (pfam00665)→
Sme-1-1	NZ_AATG01000011.1	+	110837	110931	←COG1936 (COG1936) **RNA**→ ←IstB
Rsp-2-7	NZ_AAME01000006.1	+	75965	76099	tRNA_anti (pfam01336)→ **RNA**→ ←IstB (pfam01695)
Rsp-2-8	NZ_AAME01000009.1	+	96960	97094	ISGsu6, transposase OrfB→ **RNA**→ ←similar to DNA replication protein
Rsp-2-9	NZ_AAME01000018.1	+	87461	87595	tISRso9, RSp0584; ISRSO9-transposase protein→ **RNA**→ ←IstB (pfam01695)
Rsp-2-10	NZ_AAME01000042.1	+	34094	34228	←COG4584 (COG4584) **RNA**→ ←IstB (pfam01695)
Rsp-2-11	NZ_AAME01000057.1	+	16250	16384	←Adk (COG0563) **RNA**→ ←IstB (pfam01695)

cd00051 EF-hand, calcium binding motif; A diverse superfamily of calcium sensors and calcium signal modulators; most examples in this alignment model have 2 active canonical EF hands.
cd00801 Bacteriophage P4 integrase.
cd01197 FimB and FimE and related proteins, DNA breaking-rejoining enzymes, integrase/recombinases, catalytic domain.
cd01347 TonB dependent/Ligand-Gated channels are created by a monomeric 22 strand (22,24) anti-parallel beta-barrel.
cd01830 SGNH_hydrolase subfamily, similar to the putative arylesterase/acylhydrolase from the rumen anaerobe Prevotella bryantii XynE.
COG0068 Hydrogenase maturation factor [Posttranslational modification, protein turnover, chaperones]
COG0563 Adenylate kinase and related kinases [Nucleotide transport and metabolism]
COG1014 Pyruvate:ferredoxin oxidoreductase and related 2-oxoacid:ferredoxin oxidoreductases, gamma subunit [Energy production and conversion]
COG1181 D-alanine-D-alanine ligase and related ATP-grasp enzymes [Cell envelope biogenesis, outer membrane]
COG1254 Acylphosphatases [Energy production and conversion]
COG1484 DNA replication protein [DNA replication, recombination, and repair]
COG1708 Predicted nucleotidyltransferases [General function prediction only]
COG1936 Predicted nucleotide kinase (related to CMP and AMP kinases) [Nucleotide transport and metabolism]
COG2124 Cytochrome P450 [Secondary metabolites biosynthesis, transport, and catabolism]
COG2202 FOG: PAS/PAC domain [Signal transduction mechanisms]
COG2944 Predicted transcriptional regulator [Transcription]
COG3039 Transposase and inactivated derivatives, IS5 family [DNA replication, recombination, and repair]
COG3293 Transposase and inactivated derivatives [DNA replication, recombination, and repair]
COG3415 Transposase and inactivated derivatives [DNA replication, recombination, and repair]
COG3513 Uncharacterized protein conserved in bacteria [Function unknown]
COG3550 Uncharacterized protein related to capsule biosynthesis enzymes [General function prediction only]
COG4227 Antirestriction protein [DNA replication, recombination, and repair]
COG4231 Indolepyruvate ferredoxin oxidoreductase, alpha and beta subunits [Energy production and conversion]
COG4453 Uncharacterized protein conserved in bacteria [Function unknown]
COG4584 Transposase and inactivated derivatives [DNA replication, recombination, and repair]
COG4977 Transcriptional regulator containing an amidase domain and an AraC-type DNA-binding HTH domain [Transcription]
pfam00149 Calcineurin-like phosphoesterase.
pfam00534 Glycosyl transferases group 1.
pfam00580 UvrD/REP helicase.
pfam00665 Integrase core domain.

pfam00717	Peptidase family S24.
pfam00872	Transposase, Mutator family.
pfam00903	Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily.
pfam01336	OB-fold nucleic acid binding domain.
pfam01385	Probable transposase.
pfam01455	HupF/HypC family.
pfam01527	Transposase.
pfam01548	Transposase.
pfam01695	IstB-like ATP binding protein.
pfam02902	Ulp1 protease family, C-terminal catalytic domain.
pfam03050	Transposase IS66 family.
pfam03401	Bordetella uptake gene (bug) product.
pfam05045	Rhamnan synthesis protein F.
pfam07282	Putative transposase DNA-binding domain.
smart00421	helix_turn_helix, Lux Regulon; lux regulon (activates the bioluminescence operon
smart00507	HNH nucleases;
smart00748	Higher Eukarytoes and Prokaryotes Nucleotide-binding domain;

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Asp-2-10, Asp-2-11, Asp-2-2, Asp-2-3, Asp-2-4, Asp-2-5, Asp-2-6, Asp-2-7, Asp-2-8, Asp-2-9, Atu-1-2, Bja-1-10, Bja-1-11, Bja-1-12, Bja-1-13, Bja-1-2, Bja-1-3, Bja-1-4, Bja-1-5, Bja-1-7, Bja-1-8, Bja-1-9, Bxe-1-2, Bxe-1-3, Bxe-1-4, Bxe-1-5, Bxe-1-6, Rpa-1-2, Rpa-1-3, Rpa-1-4, Rpa-1-5, Rpa-1-6, Rpa-1-7, Rpa-1-8, Rru-1-2, Rsp-1-2, Rsp-1-3, Rsp-2-10, Rsp-2-11, Rsp-2-3, Rsp-2-4, Rsp-2-5, Rsp-2-6, Rsp-2-8, Rsp-2-9, Sal-1-2, Ssp-2-10, Ssp-2-12, Ssp-2-13, Ssp-2-14, Ssp-2-15, Ssp-2-16, Ssp-2-17, Ssp-2-18, Ssp-2-19, Ssp-2-2, Ssp-2-20, Ssp-2-23, Ssp-2-24, Ssp-2-3, Ssp-2-4, Ssp-2-5, Ssp-2-6, Ssp-2-7, Ssp-2-8, Ssp-2-9

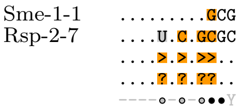
alignment positions 1 · · · 158

```
Mlo-1-1   UCGGCGGGGGUGGUGUGACGCAGU.GAGG.GCGUA............................................GC.CCGA.ACGGAGUUACACCACCCCCGCGG..CG.UC...................ACUGCC................
Sme-2-1   UCGGCGGCGGGGGUGUGACGAAGG.GAGG.GCGUA............................................GC.CCGA.CCGGAGUUACACUCCCGCCGCG...GC.GC.......AA....UUUUUAGGCGUC.UCA.UGAUC......
Rsp-1-1   UCGGCGGUGGGGGUGUGACGAAGG.GAGG.GCGUA............................................GC.CCGA.CCGGAGUUACACCCCCGCCGC.....G.GC..GCAAUU.......UUUAGGCGUC.UCA.U....GAUCGC
Bja-1-1   GCUACGCCGGGCUCGUGACGCAGG.GAGG.GCGUA............................................GC.CCGA.CCGGAGUUACGAGCCCGGCGU....CG.GC..GCGAUCCA...CAGCGGACCGCG.CCGACUGGUGAUCGC
Mlo-1-2   AAUGCCGGCGGAUUGGAACGCAGG.GAGG.GCGUA............................................GC.CCGA.CCGGAGUUUCAAUCCGCCGGC.....G.GC..GCGUUCCU...UUUCGGGAGCGU.UCGUCUGGUGAUCAC
Pde-1-1   GGAUCGGCGGGAGCGCGGCGGAGG.GAGG.GCGUA............................................GC.CCGA.CCGGAGGCGCGCUCCCGCCGA....UC................UGCCCGAUUUCC.CGG.GGUCU......
Asp-2-1   AUGUGCGACGGAGCGCCCCGAAGG.GAGGGCCGAA............................................GGCCCGA.CCGCAGGGGCGCUCCGUCGCG....CG.GC....................CCUCA................
Atu-1-1   UCGGCGCACGCAUGAGACCUCCGG.GAGU.GCGUA............................................GC.CCGA.GCGGGGUUCUCAUGCGUGCGCG...UC.GA......................UUU................
Acr-1-1   AAGACGGCUGGGGAGGCACCGGCG.AAGG.GNGNA............................................GC.CCGG.AGCCGGUGCCUCCCCAGCCGGG...CG.GC...............GUCCUUGCCA.GGG.C..........
Acr-1-2   AAUGCGGCUGGGGAGGCACCGGCG.GAGG.GCGUA............................................GC.CCGG.AGCCGGUGCCUCCCCAGCCGGG...CG.GC...............GUCCUUGCCA.GGG.C..........
Acr-1-3   AAUGCGGCUGGAGAGGCACCGGCG.GAGG.GCGUA............................................GC.CCGG.AGCCGGUGCCUCCCCAGCCGGG...CG.GC...............GUCCUUGCCA.GGG.C..........
Rru-1-1   UGUGCCUGGGGAUGAUGGCGCAGG...GC.GCGGA............................................GC.GC...CCGGAGGCAUCAUCCCCAGGC..................................................
Ssp-2-1   CACCGUUGGGGCGUGAAGCGGAGG.GAGC...GUA...............................................GCGA.CCGGAGCUUCAUGCCCCAACG..................................................
Rru-1-3   GCGCUGCCAGGAUAGCCCCGUAGG.G.CG.GCGUA............................................GU.CGG..CCGAAGGGGCUAUCCUGGCAG..................................................
Ssp-2-11  CACCGUUGGGGCGUGAAGCGGAGG.GA...GCGUA............................................GC...GA.CCGGAGCUUCAUGCCCCAACGG...UG.GC..AUGGCU.........UCUGUCUA.GGC......GGCCAU
Sal-1-1   UUUUCGCGCUGCUGGCGGCGUAGG.GAGG.GCGUA............................................GC.CCGA.CCGGAGGCGUCAGCAGCGCGA....AG.GC..G..........GUUUUCUGGCUG.GGUCAGCUG.....C
Asp-1-1   GCUUCGCGAGCUUGGUUGCGAAGG.GAGG.GCGUA............................................GC.CCGA.CCGGAGGAACCGAGUCCGCGA....UG.GC.............GCGAACUUUCGG.GUU.CGU........
Rsp-2-1   UUGGCCGCUGGGUUCAGCCGGAGG.ACGG.GCGUA............................................GC.CCGUCCCGGAGGCUGAACCCAGCGGC....UG.GC..CCU........CGUUUCGGGGGA.GAA.GGGU....AGG
Bja-1-6   UUUACGAUUUUGUCGUUGCGUCGG.GAGG.GCGUA............................................GC.CCGA.CCAGAGCGACGACAAAAUCGU....CG.GC.......GAC...GGUCAUGCGGCC.AUC.ACCAU......
Bja-1-14  UUUACGGUUUUGUCGUUGCGUCGG.GAGG.GCGUA............................................GC.CCGA.CCAGAGCGACGACAAAACCGU....CG.GC.......GAC...GGUCAUGCGGCC.AUC.ACCAU......
Bja-1-15  UUUACGAUUUUGUCGUUGCGUCGG.GAGG.GCGUA............................................GC.CCGA.CCAGAGCGACGACAAAAUCG............UCGGCG.....ACGGUCAUGC...GGC.CAU..UGCGGA
Mlo-1-3   .CGCCCGUCGGGCCCGGUCGUAGG.GAGGGGCGGA............................................GCCCCGA.CCGGAGAGCGGGUCUGACGGG....CGGGC..GCGACACA...AGCCCCUGCGA..GGG.GCUCUGAUCGG
Ssp-2-21  UCGGUUGUCUGGGCUGAUCGUAGG.GAGGGGCGUA............................................GCCCCGA.CCGGAGAGCGGACCAGACAAC....CG.GUG.GCGAUCUUUUUCCCCUUCUUUCG.GGG.GGGCUGAUCGG
Bsp-1-1   UUUCCGCCGGUGACGGGCCGGAGG.GA.G.GCGGA............................................GC.C.GA.CCGGAGGGCCGUCACCGGCGGU...CG.GCGUUUUGAU.....GAGACCGCUCUU.GCG.CUUGGCUCGGA
Ssp-2-22  AUCGCCGCCAUAUCCGGACGGAGG.UGGA.GCGAA............................................GC.GGAA.CCGGAGGCCGGAUAUGGCGGUGU..CG.CC..GUUUCC...........................GGUGCC
Ssp-1-1   UCGGCCGCCAUUGCCGGUCGGAGG.UGGA.GCGGA............................................GC.GGAA.CCGGAGACCGGCAAUGGCGGUGU..CG.CC..GCUUCC.....GGCGCCGGCGGGCGGUAGCCGAGGCUGC

          ...<<<<<<<<<<<<<<<<...<<...<<.<<...<<<<<..<<<<<<.<<<<....>>.>>.>>>>>>>>>>>.....>>.>>...>>...>>>>>>>>>>>>>>>>....<<.<<..<<<<<<.....<<<<<<.......>>>.>>>..>>>>>>
          ...??22222222222??2...2?...??.21...00000..000000.0000....00.00.00000000000.....12.??...?2...2??22222222222??....??.??..??????.....2????2.......2??.??2..??????
          ●●●●Y●●●●●●R●●R●●●CG●RGG-GRGG-GCGUA--------------------------------------------GC-CCGA-●CGRAG●●●Y●●Y●●●●●●R●----●●-●●--○○○○○○-----○○●●●●●●●●●○-●●●-●○○○○○○○○○○
```

```
Ssp-2-25  AUCGCCGCCAUAUCCGGACGGAGG.UGGA.GCGAA............................................GC.GGAA.CCGGAGGCCGGAUAUGGCGGUGU..CG.CC..GUUUCC.....GGUGCCGGCGGCCGGUAGCCCAGGCUGC
Rsp-2-2   GGCGAGCGGGGCGCGGAGUGAAGG.GAGG.GCGUA............................................GC.CCGA.CCGAAGUCCGCGCCCCGCUC.....G.CU..GCGGCCGA...ACUUCCGGUCU..GGA.GGUCGGGUGG
Rpa-1-1   UUGGCCGGUGUGGCCGGACGGCGG.GAGG.GCGUA............................................GC.CCGA.CCAGAGUCCGGCCACACCGGC....UG.CG..CCGUUUUG...UGGGGUCAUUUG.ACG.UCACU.GGCGG
Rru-1-4   GCGCUGCCAGGAUGGCCCCUUAGG.G.CG.GCGUA............................................GC.CGA..CCGGAGGGGCUGUCCUGGCGG....GG.GC....GGCCAU...CGGCGCCAAAAA.GAU.CCGCGGACGAA
env-1     CGGACGCAGGCAUGAGGCCCGCGG.GAGG.GCUUA............................................GC.CCGA.GCGGGGGGCUCAUGCCUGCGU....UG.CG..AUUUUCUGA..AGGGUUUCAGCC.GGC.CUUUCGGG...
Ret-1-1   ACUCGCGCCGGGGUGUGACGGAGU.GAGG.GCGUA............................................GC.CCGA.UCGUAGUCAUACCCCGGCGUGGCGCAG.AU..UUUCGG.....CGUCUCAUGAUC.GCGGUCGGUCCGGUA
Ssp-1-2   GUUACAUGCCGGGGGGAGUGGCGU.GAGG.GCGUA............................................GC.CCGA.AGGCCACUCCCCCCGGCAUUG....GC.UU..GAUUUUC....AGGGUCUGGUUUUGGC.CUUUC......
Bxe-1-1   CUGGCGUUGCGGCAGCGGUGGA.....GG.GCGUA............................................GC.CC.....GGAACCGCUGCCGCAACGCCG..CG.CC..GAGGAGUC...CGCCCCGCUGCC.GGG.GUGC.CUUGUC
Sme-1-1   UCGCGCGCGCAUGAGACCCCCGGG.GAGG.GCGUA............................................GC.CCGA.GCG..GGGUCUCAUGCGCGCG....C......GACGCUUU...CUGGAGAGGGG..UUU.CAGCCAGCCUU
Rsp-2-7   GAGAGCGCGACAAUCAGGAUGAGA.G.UC.GUGGUUGUCGCGGCGGGAUCAGGAUUGCCGUGUUCCCGUCGGCACUCAAGC.GGUU.UCGCAUCCUGAUUGUCGCGGA....GC.GU..CAAUCA.........GAGGUGU..UCU......UGAUUG
          ...<<<<<<<<<<<<<<<<...<<...<<.<<...<<<<<..<<<<<<.<<<<....>>.>>.>>>>>>>>>>>.....>>.>>...>>...>>>>>>>>>>>>>>>>....<<.<<..<<<<<<.....<<<<<<.......>>>.>>>..>>>>>>
          ...??22222222222??2...2?...??.21...00000..000000.0000....00.00.00000000000.....12.??...?2...2??22222222222??....??.??..??????.....2????2.......2??.??2..??????
          ●●●●Y●●●●●●R●●R●●●CG●RGG-GRGG-GCGUA--------------------------------------------GC-CCGA-●CGRAG●●●Y●●Y●●●●●●R●----○○-○○--○○○○○○-----○○○○○○○○○●○○-○○○-○○○○○○○○○○○
```

alignment positions 159 · · · 170

```
Mlo-1-1   ....G.A.CGUC
Sme-2-1   ....G.C.GGUC
Rsp-1-1   ...GG.U.CGGC
Bja-1-1   ...GG.C.CGGC
Mlo-1-2   ...GG.U.CGAC
Pde-1-1   ....G.U.GAUC
Asp-2-1   ....G.C.CGGC
Atu-1-1   ....U.U.GAAA
Acr-1-1   ....G.C.CGCC
Acr-1-2   ....G.C.CGCC
Acr-1-3   ....G.C.CGCC
Rru-1-1   ............
Ssp-2-1   ............
Rru-1-3   ............
Ssp-2-11  ...CG.C.CAAU
Sal-1-1   ...GG.U.UUUU
Asp-1-1   ....G.C.CGCC
Rsp-2-1   ...UG.C.UGGC
Bja-1-6   ...AG.C.UUGC
Bja-1-14  ...AG.C.UUGC
Bja-1-15  ...GA.U.AAAC
Mlo-1-3   ...GG.CGCGUC
Ssp-2-21  ...GG.C.UGU.
Bsp-1-1   ....G.C.CGGU
Ssp-2-22  ....G.G.CGGC
Ssp-1-1   ...UGUG.CGGC
Ssp-2-25  ...UAUG...GU
Rsp-2-2   ...UC.G.CGGU
Rpa-1-1   ...UC.G.UUUC
Rru-1-4   AUUGG.C.CUCU
env-1     ....C.G.CGGC
Ret-1-1   ...CA.U.UUGC
Ssp-1-2   ....G.G.GCCC
Bxe-1-1   ....G.G.UGGU

          ....>.>.>>..
          ....?.?.??..
          ----○-○-○●●Y
```



```
Sme-1-1   .........GCG
Rsp-2-7   ....U.C.GCGC
          ....>.>.>>..
          ....?.?.??..
          ----○-○-○●●Y
```

2.13 excisionase motif

2.13.1 Possible misannotation of excisionase genes

We note that the excisionase genes associated with this motif are often annotated merely as unknown DNA-binding proteins, and we can find only one gene (a putative prophage regulatory gene) nearby to Xis that relates to phages. Therefore the excisionases could be misclassified, and may encode some other type of DNA-binding protein.

2.13.2 Genome context and alignment

Notes:

[1] Downstream gene is annotated as overlapping RNA. The overlapping region is only found in one other excisionase, despite several excisionases with homology detectable by BLAST. (BLAST therefore implies that this gene is also an excisionase.) Moreover, the indicate start codon is consistent with the sequence motifs in the other sequences.

[2] Downstream gene is annotated as overlapping RNA. However, BLAST suggests that the overlap region is not conserved, and therefore may be an overprediction. The start codon we selected has a consistent placement with those of the other sequences.

[3] The homology for this hit is unconvincing, but plausible. Gene context is highly consistent, but some conserved features are broken. This hit was found in a search of UTRs of Actinobacterial excisionases found by PSIBLAST, and the CM estimated an E-value of above 1 even in this small database.

[4] Annotated start is the GUG nine nucleotides upstream of the putative start codon we have selected. Our candidate start codon is more consistent with those of other similar sequences.

abbrev. of hits	**taxonomy of species**
Cdi-1-1 to Cdi-1-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Corynebacteriaceae *Corynebacterium diphtheriae* NCTC 13129
Cef-1-1 to Cef-1-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Corynebacteriaceae *Corynebacterium efficiens* YS-314
Cgl-1-1 to Cgl-1-2	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Corynebacteriaceae *Corynebacterium glutamicum* ATCC 13032
Cje-1-1 to Cje-1-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Corynebacteriaceae *Corynebacterium jeikeium* K411
Mfl-1-1 to Mfl-1-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium flavescens* PYR-GCK
Mle-1-1 to Mle-1-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium leprae* TN
Msp-1-1 to Msp-1-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium sp.* JLS
Msp-2-1 to Msp-2-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium sp.* KMS
Msp-3-1 to Msp-3-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium sp.* MCS
Mva-1-1 to Mva-1-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium vanbaalenii* PYR-1
Mav-1-1 to Mav-1-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium avium subsp. paratuberculosis* K-10
Mbo-1-1 to Mbo-1-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium bovis* AF2122/97
Mtu-1-1 to Mtu-1-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium tuberculosis* C
Mtu-2-1 to Mtu-2-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium tuberculosis* CDC1551
Mtu-3-1 to Mtu-3-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium tuberculosis* F11
Mtu-4-1 to Mtu-4-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Mycobacteriaceae *Mycobacterium tuberculosis* H37Rv
Nfa-1-1 to Nfa-1-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Nocardiaceae *Nocardia farcinica* IFM 10152
Rsp-1-1 to Rsp-1-1	Actinobacteria Actinobacteridae Actinomycetales Corynebacterineae Nocardiaceae *Rhodococcus sp.* RHA1
Ace-1-1 to Ace-1-2	Actinobacteria Actinobacteridae Actinomycetales Frankineae Acidothermaceae *Acidothermus cellulolyticus* 11B
Kra-1-1 to Kra-1-1	Actinobacteria Actinobacteridae Actinomycetales Frankineae Kineosporiaceae *Kineococcus radiotolerans* SRS30216
Jsp-1-1 to Jsp-1-1	Actinobacteria Actinobacteridae Actinomycetales Micrococcineae Intrasporangiaceae *Janibacter sp.* HTCC2649
Str-1-1 to Str-1-1	Actinobacteria Actinobacteridae Actinomycetales Micromonosporineae Micromonosporaceae *Salinispora tropica* CNB-440
Nsp-1-1 to Nsp-1-1	Actinobacteria Actinobacteridae Actinomycetales Propionibacterineae Nocardioidaceae *Nocardioides sp.* JS614
Sav-1-1 to Sav-1-1	Actinobacteria Actinobacteridae Actinomycetales Streptomycineae Streptomycetaceae *Streptomyces avermitilis* MA-4680
Sco-1-1 to Sco-1-1	Actinobacteria Actinobacteridae Actinomycetales Streptomycineae Streptomycetaceae *Streptomyces coelicolor* A3(2)

abbrev	**RefSeq accession**		**5′ at**	**3′ at**	**genes**
Cgl-1-1	NC_003450.3	+	435866	435954	**RNA**→ hypo→
Cgl-1-2	NC_006958.1	+	435865	435953	**RNA**→ hypo→ extremely conserved POSSIBLE DNA-BINDING PROTEIN→
[4]Cef-1-1	NC_004369.1	+	467815	467903	**RNA**→ hypo→
Cdi-1-1	NC_002935.2	+	361856	361943	**RNA**→ Putative DNA-binding (excisionase) protein→ hypo→ DsbA_Com1_like (cd03023)→
Nfa-1-1	NC_006361.1	-	5475483	5475397	**RNA**→ hypo→
Rsp-1-1	NC_008268.1	-	2173234	2173148	**RNA**→ possible excisionase→
Mav-1-1	NC_002944.2	+	4449736	4449823	**RNA**→ hypo→
Mtu-2-1	NC_002755.2	+	592395	592484	**RNA**→ excisionase, putative→
Mtu-4-1	NC_000962.2	+	591027	591116	**RNA**→ hypo→
Mbo-1-1	NC_002945.3	+	592178	592267	**RNA**→ hypo→
Mtu-3-1	NZ_AAIX01000005.1	+	131569	131658	**RNA**→ COG1910: Periplasmic molybdate-binding protein/domain→
Mtu-1-1	NZ_AAKR01000010.1	+	721	810	**RNA**→ COG1910: Periplasmic molybdate-binding protein/domain→
Mle-1-1	NC_002677.1	-	2905434	2905344	**RNA**→ possible DNA-binding protein→
Mva-1-1	NZ_AAPF01000001.1	+	930429	930517	**RNA**→ Excisionase/Xis, DNA-binding→
Mfl-1-1	NZ_AAPA01000002.1	-	278262	278176	**RNA**→ Excisionase/Xis, DNA-binding→
Msp-3-1	NC_008146.1	+	729011	729098	**RNA**→ Excisionase/Xis, DNA-binding protein→
Msp-1-1	NZ_AAQC01000005.1	-	449098	449011	**RNA**→ Excisionase/Xis, DNA-binding→
Msp-2-1	NZ_AAQD01000006.1	+	17383	17470	**RNA**→ Excisionase/Xis, DNA-binding→
Cje-1-1	NC_007164.1	-	2221287	2221196	**RNA**→ hypo→
Sco-1-1	NC_003888.3	-	3680342	3680263	**RNA**→ hypo→
[2]Sav-1-1	NC_003155.3	+	5769564	5769643	**RNA**→ hypo→
Nsp-1-1	NZ_AAJB01000035.1	+	44622	44705	**RNA**→ Phage_AlpA (pfam05930)→
Str-1-1	NZ_AATJ01000007.1	+	1327	1402	**RNA**→ Uncharacterized membrane-associated protein-like→ GlcD (COG0277)→
[1]Jsp-1-1	NZ_AAMN01000001.1	-	762516	762442	**RNA**→ hypo→
Ace-1-1	NZ_AAOL01000004.1	+	111543	111630	**RNA**→ Excisionase/Xis, DNA-binding→
[3]Kra-1-1	NZ_AAEF02000084.1	-	4439	4349	**RNA**→ Excisionase/Xis, DNA-binding→
[3]Ace-1-2	NZ_AAOL01000002.1	-	432049	431961	**RNA**→ HTH_MlrA (cd01104)→

cd01104 Helix-turn-helix transcription regulator MlrA (merR-like regulator A).

cd03023 DsbA family, Com1-like subfamily; composed of proteins similar to Com1, a 27-kDa outer membrane-associated immunoreactive protein originally found in both acute and chronic disease strains of the pathogenic bacteria Coxiella burnetti.

COG0277 FAD/FMN-containing dehydrogenases [Energy production and conversion]

pfam05930 Prophage CP4-57 regulatory protein (AlpA).

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Cgl-1-2, Mbo-1-1, Msp-1-1, Msp-2-1, Mtu-1-1, Mtu-3-1, Mtu-4-1

alignment positions 1 · · · 104

```
  Cgl-1-1  CUAAGCU.GAACCA.GGCAA.AAGCGUGA...UCGUA.CUGCGGG.AGGGGAAGCCCGCAG.CG..CGCGAAU.UGCUGAAGGGU.AAUGAUAUAU..AUGGCU
4 Cef-1-1  CUAAGCU.GAACCA.GGCGU.AUGCGUGA...UCGUA.CUGUGGG.AGGGGAAGCCCGCAG.CG..CGCGUAA.GGCUGAAGGGU.AGUGAUGAAU..AUGGCU
  Cdi-1-1  CUAAGCU.GAUUUG.AGCAC.GUGCGUGA...UCGUU.CUGCGGG.AGGGGAAGCCUGCAG.CG..CGCG.ACUUGUUGAAGGGUU.AGGUAA.AU..AUGGCU
  Nfa-1-1  CUAAGCU.UCAAGG.AGCAC.GUGCGUGU...CUGUU.CCGCCGG.UGGGGAAGCCGGCGG.C..GCG.G.AC.GUGCCGGAGGUCGAUGGUGCA...AUGAUG
  Rsp-1-1  CUAAGCUCGAAGU..AGCACGUG.CGUGU..CUGUU..CCGCCGG.AGGGGAAGCCGGCGG..C.GCGAGACG.UGCCGG.AGGUAAGUGGUU.G...AUGACG
  Mav-1-1  CUAUUCU..CCUUG.UGUAA.GCGCGUGU..CGGUG..CCAGCGG.AGGGGAAGCCGCUGGCAU.GCGCGUGC.CUGACAC.GAUU.GGGUUGCG...AUGACG
  Mtu-2-1  CUAUUCU.CCUCUU.UGUAA.GCGCGUGU..GGGUG..CCAGCGG.AGGGGAAGCCGCUGGGACUGCGCGUGC.CUGACAC.GAUU.GGGUUGCG...AUGACG
  Mle-1-1  CUAUCCUCUUUUAU.UGUAU.GCACGUGU..GGGUG..CCAGCGG.AGGGGAAGCCGCUGGCAUUGCGCGUGC.CUGACAC.GAUU.GGGUUGCG...AUGACG
  Mva-1-1  CUAUUCU..CCCAGUGGUAA.GCACGGGU..UGGUG..CCAGCGG.UGGGGAAGCCGCUGGAACUGCCCGCGC...CUGAUGGAUU.GGGUUGCG...AUGACG
  Mfl-1-1  CUAUUCU..CCCAGUGGUAA.GCACGGGU..GGGU...CCAGCGG.UGGGGAAGCCGCUGGAACUGCCCGCGC...CUGA.GGAUU.GGGUUGCG...AUGACG
  Msp-3-1  CUAUUCU..CCUCG.UGUAU.GCACGGGU..CGGUG..CCAGCGG.UGGGGAAGCCGCUGGAACUGCCCGCGC...CUGAUGGAUU.GGGUUGCG...AUGACG
  Cje-1-1  UUAUGAUGGUUCGCGAGCGA.UCACGUGU.GACGUC..CUGCAGG.AGGGGAAGCCUGCAG..C.GCAUGAGA..GCGCAAAGAAA.GUGAUGAUCU.AUGGCU
  Sco-1-1  GUGAGCA........CG....CAGCGACG..AAGAG..UCACCGGAAGGGGAAGCCGGUGGC...CGUCG.AG.UGCGGA.AGGUUCAGGU.GUGUC.AUGGCU
2 Sav-1-1  GUGA....GCACGC.......A.ACGACG..AAGAG..UCACCGGAAGGGGAAGCCGGUGGC...CGUCGAGU..GCGGA.AGGUUCAGGU.GUGUC.AUGGCU
  Nsp-1-1  CUAUUCU..CACCG.UGCG..C.GACGCG..UGCAC..UCACCGG.CGGGGAAGUCGGUGG....CGCGUCAG..ACAGA.AAGUAUCGGU.GCACG.AUGGCU
  Str-1-1  UGACCG...........AA.GCCCCCAG....CCG..CCCCUGA.AGGGGAAGUCAGGGGCGG.CUGGGGGCUACUCCG..GAU....GUCCAGGG.AUGCGG
1 Jsp-1-1  UCUUG...........CAC.GCCCGUAG....UGA..ACGCCGA.CGGGGAAGUCGGCG...A..UGCGAGCAUGCAGA.AGAUCGAGGUGGCG...AUGUCU
  Ace-1-1  ..CAGCU.ACCCUC.AGUAACACACGUGCUGAUCGU..CCGCCGG.AGGGGAAGCCGGCGUCCGCGCGUGCCG......A.GGGUCU.GGU.GCAAACAUGAGC
3 Kra-1-1  CUACAGU.GGUCUG.GAAAC.GCGCCCCC..ACCGGGCCGUCCGC.GGGGGACGGCG..CGCCG.CGGCGCGCGAACUGA.AGGGUG.CUGUGGG..AUGCCU
3 Ace-1-2  CCAUCCC.GUAGUG.GGGG..GCCGACGC...UCCG..UCGGCUCAAGGAGGAGGAGUCGGCAGCGCGUCGGC.UCUGAUUGGAGG..AAGUUCC...GUGGUA
           .....................<<<<<<<<....<<<..<<<<<<<.........>>>>>>>>>>.>>>>>>>>...............................
           .....................???2?22?....???..?2??222.........222??2????.?22?2???...............................
           YYA●●○Y-○○○○○○-○R○R○-R○●CRYR●--○○●●●--YYRYYGR-●GGGGAAGYCR●YRG○○○-○YR●G○R●-○○Y●●R-RGRU○○○●G●○R○R○--AUGRY●
                                                                                     SD?+++              AUG
```

2.14 The ATPC motif

The part of the alignment past the start codon has the codons' wobble position marked with '`w`'. The pattern of mutations suggests a coding sequence, which supports our annotations of start codons. Note that the consensus line is somewhat misleading because we have not produced an overall global alignment. However, by comparing mutations among highly similar sequences (which are aligned well without adding gaps), the pattern of mutations is clear.

The gene context of environmental sequences is not shown because they are usually annotated only sparsely.

No Shine-Dalgarno sequence is annotated because the Shine-Dalgarno sequence appears to be optional in Cyanobacteria [10].

When looking at the alignment, it is worth noting that most of the sequences fit into two clusters of highly similar sequence.

Notes:

[1] No start codon was annotated in this region. However, this sequence is highly similar to another sequence with an annotated start codon; we therefore used the same start codon.

[2] This sequence fragment is cut off before the AUG start codon, by analogy with other highly similar sequences.

[3] The two indicated RNAs in strains of *Prochlorococcus marinus* have annotated start codons that are likely to be $5'$ of the real start. First, one is 28 codons longer than the other, and second the extra coding region in both annotations is not conserved in any other bacteria. Finally, sequence conservation after (but not before) the common start codon we have selected is very high, and most mutations are in the wobble position. We therefore think our indicated start codon is the likely true start codon.

[4] This RNA overlaps an annotated gene (not shown here) running in the reverse direction. However, this putative gene is only 36 amino acids and BLAST indicates that it is not conserved in any bacteria.

[5] These sequences come from the Global Ocean Survey [12], and were downloaded from the CAMERA site at `http://camera.calit2.net/` related to CAMERA identifier CAM_PUB_Rusch07a. We estimated the start codon based on homology with other sequences.

[6] This genetic information downloaded from `http://camera.calit2.net` may be considered to be part of the genetic patrimony of the country from which the sample was obtained. Users of this information agree to: (1) acknowledge the country of origin in any publications where the genetic information is presented and (2) contact the CBD focal point identified on the CBD website (`http://www.biodiv.org/doc/info-centre.shtml`) if they intend to use the genetic information for commercial purposes.

[7] This sequence is derived from Open Ocean samples obtained from Sargasso Sea, Hydrostation S, Bermuda (UK)

[8] This sequence is derived from Coastal samples obtained from Northeast of Colón, Panama

[9] This sequence is derived from Coastal samples obtained from Gulf of Panama, Panama

[10] This sequence is derived from Open Ocean samples obtained from 30 miles from Cocos Island, Costa Rica

[11] This sequence is derived from Open Ocean samples obtained from Yucatan Channel, Mexico

[12] This sequence is derived from Open Ocean samples obtained from Sargasso Sea, Station 13, Bermuda (UK)

[13] This sequence is derived from Open Ocean samples obtained from 134 miles NE of Galápagos, Ecuador

[14] This sequence is derived from Coastal samples obtained from North James Bay, Santigo Island, Ecuador

[15] This sequence is derived from Open Ocean samples obtained from 250 miles from Panama City, Panama

[16] This sequence is derived from Fringing Reef samples obtained from Dirty Rock, Cocos Island, Costa Rica

[17] This sequence is derived from Coral Reef Atoll samples obtained from Rangirora Atoll, Fr. Polynesia

[18] This sequence is derived from Open Ocean samples obtained from Sargasso Sea, Station 11, Bermuda (UK)

[19] This sequence is derived from Hypersaline samples obtained from Punta Cormorant, Hypersaline Lagoon, Floreana Island, Ecuador

[20] This sequence is derived from Coastal Sea samples obtained from Gulf of Mexico, USA

[21] This sequence is derived from Coastal samples obtained from Off Key West, FL, USA

[22] This sequence is derived from Coastal samples obtained from Devil's Crown, Floreana Island, Ecuador

[23] This sequence is derived from Coastal samples obtained from Coastal Floreana, Ecuador

[24] This sequence is derived from Coastal samples obtained from North Seamore Island, Ecuador

[25] This sequence is derived from Open Ocean samples obtained from Rosario Bank, Honduras

[26] This sequence is derived from Open Ocean samples obtained from 201 miles from F. Polynesia, International

abbrev. of hits	**taxonomy of species**
Ssp-1-1 to Ssp-1-1	Cyanobacteria Chroococcales *Synechococcus sp.* CC9311
Ssp-2-1 to Ssp-2-1	Cyanobacteria Chroococcales *Synechococcus sp.* CC9605
Ssp-3-1 to Ssp-3-1	Cyanobacteria Chroococcales *Synechococcus sp.* CC9902
Ssp-4-1 to Ssp-4-1	Cyanobacteria Chroococcales *Synechococcus sp.* RS9917
Ssp-5-1 to Ssp-5-1	Cyanobacteria Chroococcales *Synechococcus sp.* WH 5701
Ssp-6-1 to Ssp-6-1	Cyanobacteria Chroococcales *Synechococcus sp.* WH 7805
Ssp-7-1 to Ssp-7-1	Cyanobacteria Chroococcales *Synechococcus sp.* WH 8102
Pma-1-1 to Pma-1-1	Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9211
Pma-2-1 to Pma-2-1	Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9312
Pma-3-1 to Pma-3-1	Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9313
Pma-4-1 to Pma-4-1	Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* NATL2A
Pma-5-1 to Pma-5-1	Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus subsp. marinus str.* CCMP1375
Pma-6-1 to Pma-6-1	Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus subsp. pastoris str.* CCMP1986
env-1 to env-87	environmental samples

abbrev	**RefSeq accession**		**5′ at**	**3′ at**	**genes**
Ssp-4-1	NZ_AANP01000006.1	+	228716	228901	ATP synthase subunit A→ ATP-synt_A (pfam00119)→ **RNA**→ AtpE (COG0636)→ ATP-synt_B (pfam00430)→ AtpF (COG0711)→ OSCP (pfam00213)→ AtpA (COG0056)→ ATP-synt (pfam00231)→
Ssp-6-1	NZ_AAOK01000001.1	+	272018	272205	possible ATP synthase protein 1→ ATP-synt_A (pfam00119)→ **RNA**→ AtpE (COG0636)→ ATP-synt_B (pfam00430)→ ATP-synt_B (pfam00430)→ OSCP (pfam00213)→ AtpA (COG0056)→ ATP-synt (pfam00231)→
Ssp-3-1	NC_007513.1	+	500648	500850	possible ATP synthase protein 1→ ATP-synt_A (pfam00119)→ **RNA**→ AtpE (COG0636)→ ATP-synt_B (pfam00430)→ ATP-synt_B (pfam00430)→ OSCP (pfam00213)→ AtpA (COG0056)→ ATP-synt (pfam00231)→ hypo→ fer2 (cd00207)→ hypo→ hypo→ COG3012 (COG3012)→
Ssp-7-1	NC_005070.1	+	491082	491284	possible ATP synthase protein 1→ ATP-synt_A (pfam00119)→ **RNA**→ AtpE (COG0636)→ ATP-synt_B (pfam00430)→ ATP-synt_B (pfam00430)→ OSCP (pfam00213)→ AtpA (COG0056)→ ATP-synt (pfam00231)→ hypo→ fer2 (cd00207)→ hypo→ hypo→ COG3012 (COG3012)→
Ssp-2-1	NC_007516.1	-	2006437	2006235	possible ATP synthase protein 1→ ATP-synt_A (pfam00119)→ **RNA**→ AtpE (COG0636)→ ATP-synt_B (pfam00430)→ ATP synthase subunit B→ OSCP (pfam00213)→ AtpA (COG0056)→ ATP-synt (pfam00231)→ hypo→ fer2 (cd00207)→ hypo→ hypo→ COG3012 (COG3012)→
Ssp-1-1	NC_008319.1	-	2030819	2030612	possible ATP synthase protein 1→ ATP-synt_A (pfam00119)→ **RNA**→ AtpE (COG0636)→ ATP-synt_B (pfam00430)→ ATP-synt_B (pfam00430)→ OSCP (pfam00213)→ AtpA (COG0056)→ ATP-synt (pfam00231)→ hypo→ fer2 (cd00207)→ hypo→ Uncharacterized membrane protein→ COG3012 (COG3012)→

[3]Pma-3-1	NC_005071.1	-	1556654	1556456	possible H+-transporting ATP synthase→ ATP-synt_A (pfam00119)→ **RNA**→ AtpE (COG0636)→
Pma-5-1	NC_005042.1	-	1464170	1463966	Uncharacterized membrane protein→ ATP-synt_A (pfam00119)→ **RNA**→ AtpE (COG0636)→
Pma-4-1	NC_007335.1	-	927802	927602	uncharacterized membrane protein→ ATP-synt_A (pfam00119)→ **RNA**→ AtpE (COG0636)→
[3,4]Pma-1-1	NZ_AALP01000001.1	+	1642039	1642240	ATP-synt_A (pfam00119)→ ←ATP synthase subunit C **RNA**→ AtpE (COG0636)→
Pma-2-1	NC_007577.1	-	1442726	1442539	ATP synthase subunit 1-like→ ATP-synt_A (pfam00119)→ **RNA**→ AtpE (COG0636)→ ATP-synt_B (pfam00430)→ ATP-synt_B (pfam00430)→ OSCP (pfam00213)→ AtpA (COG0056)→ ATP-synt (pfam00231)→ fer2 (cd00207)→ hypo→ hypo→
Pma-6-1	NC_005072.1	-	1390580	1390392	possible ATP synthase subunit 1→ ATP-synt_A (pfam00119)→ **RNA**→ AtpE (COG0636)→ ATP-synt_B (pfam00430)→ ATP-synt_B (pfam00430)→ OSCP (pfam00213)→ AtpA (COG0056)→ ATP-synt (pfam00231)→ fer2 (cd00207)→ hypo→ hypo→
Ssp-5-1	NZ_AANO01000002.1	-	213710	213495	possible H+-transporting ATP synthase→ ATP-synt_A (pfam00119)→ **RNA**→ AtpE (COG0636)→

cd00207	2Fe-2S iron-sulfur cluster binding domain.
COG0056	F0F1-type ATP synthase, alpha subunit [Energy production and conversion]
COG0636	F0F1-type ATP synthase, subunit c/Archaeal/vacuolar-type H+-ATPase, subunit K [Energy production and conversion]
COG0711	F0F1-type ATP synthase, subunit b [Energy production and conversion]
COG3012	Uncharacterized protein conserved in bacteria [Function unknown]
pfam00119	ATP synthase A chain.
pfam00213	ATP synthase delta (OSCP) subunit.
pfam00231	ATP synthase.
pfam00430	ATP synthase B/B' CF(0).

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: env-31

```
                  alignment positions 1···160
[5,6,19]env-1     UUCGCCGGUCCCCGGCGA..UCUG.CUAAACACAUC.GCGCGC....AGGUCCGGGUGG....AACCGGGCCCACUCCUCUCAG.......GGGAGUGU.CCCGG.GCGCGCUCUGCCCUU...UC........CGCG.CUCAUCCUGCGCAACCCCAAC
Ssp-4-1           UUCGCCGGUCCCAGGCGA..UCUG.CUAAACACUUC.GCGCGC....AGGUCCGGGUGG....AACCGGGCCCACUCCUUUCAG.......GGGAGUGU.CCCGG.GCGCGCUCUGCCCUU...UC........CGCG.CUCAUCCUGCGCAACCCCAAC
Ssp-6-1           UUCGCAGGUCCCCAGUGA..UCUG.CUAAACACUUC.GCGCGC....AGGUCCGGGUGG....AACCGGGCCCGCCCUCCAUCGU......GAGGGUGU.CCCGG.GCGCGUCCUGCCCUU...UC........CGCG.CUCAUCCUGCGCACCUCCAAC
[5,6,7]env-2      UUCGCUCUAACCCGGCGA..UCUG.CUAAACACUUC.GCGCGC....AGGUCCGGCUGCU...GACCGGGCCCUUCCCUGAUUCC......AGGGUUGU.CCCAG.GCGCCCAUGGGGUCCCA.AC........CCCA.CCCAUCGCCCACUCAGCGCUC
[5,6,22]env-3     CUUGCCGUUAA..GGCGA..UCUG.CUAAACACUUCUGCGCGU....AGGUCCGACGCUCAUGACUCGGGCCCGCUCUCUUGCAACA....GAGAGUGU.CCCU..GCGCA.GGGGGAACC...AA........UCCC.CACUUCAUUUGCUCAGCGCUA
[5,6,22]env-4     CUCGCCGUUAA..GGCGA..UCUG.CUAAACACUUCUGCGCGU....AGGUCCGACGCUCAUGACUCGGGCCCGCUCUCUUGCAACA....GAGAGUGU.CCCU..GCGCA.GGGGGAACC...AA........UCCC.CACUUCAUUUGCUCAGCGCUA
[5,6,23]env-5     CUCGCCGUUAA..GGCGA..UCUG.CUAAACACUUCUGCGCGU....AGGUCCGACGCUCAUGACUCGGGCCCGCUCUCUUGCAAUA....GAGAGUGU.CCCU..GCGCA.GGGGGAACC...AA........UCCC.CACUUCAUUUGCUCAGCGCUA
[5,6,24]env-6     UUCGCUGAAUUUCGGCGA..UCUG.CUAAACACAUCCGCGCGC....AGGUCCGAAUUGC...AACCGGGCCCUCUCCUUCCU.........GGAGUGU.CCCU..GCGCG.GGGGGAACC...GA........UCCC.CAACCUCAUUCAGUUCAGCGC
Ssp-3-1           UUCGCUUUUACCUGGCGA..UCUG.CUAAACACAUCCGCGCGU....AGGUCCGGACUGC...AACCGGGCCCUCUCUUCAC.........AAGAGUGU.CCCU..GCGCG.GGGGGAACC...GA........UCCC.CAACCCAAUUCAGUUCAGCGC
[5,6,7]env-7      UUCGCCAAUUGAUGGCGA..UCUG.CUAAACACAUCCGCGCGC....AGGUCCGGCGAUC...ACCCGGGCCCGCUCCUUCA.........AGGAGUGU.CCCU..GCGCG.GGGGGAACC...GA........UCCC.CAAGUCCAUUCCGUUCAGCGC
Ssp-7-1           UUCGCCAUUUGAGGGCGA..UCUG.CUAAACACGUCCGCGCGC....AGGUCCGGAUAGC...ACCCGGGCCCGCUCCUUUC.........AGGAGUGU.CCCU..GCGCG.GGGGGAACC...GA........UCCC.CAAGUCCAUUCCGUUCAGCGC
[1]env-8          UUCGCCUUUUGAGGGCGA..UCUG.CUAAACACAUCCGCGCGC....AGGUCCGGAUAGC...ACCCGGGCCCGCUCCUUUC.........AGGAGUGU.CCCU..GCGCG.GGGGGAACC...GA........UCCC.CAAGUCCAUUCCGUUCAGCGC
[1]env-9          UUUGCCUUUUGAGGGCGA..UCUG.CUAAACACAUCCGCGCGC....AGGUCCGGAAAAC...ACCCGGGCCCGCUCCUUUC.........AGGAGUGU.CCCU..GCGCG.GGGGGAACC...GA........UCCC.CAAGUCCAUUCCGUUCAGCGC
[5,6,7]env-10     UUCGCCUUUUGAGGGCGA..UCUG.CUAAACACAUCCGCGCGC....AGGUCCGGACGAC...AACCGGGCCCGCUCCUUCC.........GGGAGUGU.CCCU..GCGCG.GGGGGAACC...GA........UCCC.CAAGCCCAUUCAGUUCAGCGC
[5,6,24]env-11    AUCGCCUGAGAACGGCGA..UCUG.CUAAACACAUCCGCGCGC....AGGUCCGGCAUUGC..ACCCGGGCCCAUUCCUUC..........GGGAAUGU.CCCU..GCGCG.GGGGGAACC...GA........UCCC.CAAGUCCAUUCCGUACAGCGC
[5,6,16]env-12    AUCGCCUGAGAACGGCGA..UCUG.CUAAACACAUCCGCGCGC....AGGUCCGGAAUUGC..ACCCGGGCCCGUUCCUUC..........GGGAAUGU.CCCU..GCGCG.GGGGGAACC...GA........UCCC.CAAGUCCAUUCCGUACAGCGC
Ssp-2-1           AUCGCCUGAGAACGGCGA..UCUG.CUAAACACAUCCGCGCGC....AGGUCCGGCAUUGC..ACCCGGGCCCGUUCCUUC..........GGGAAUGU.CCCU..GCGCG.GGGGGAACC...GA........UCCC.CAAGUCCAUUCCGUACAGCGC
[5,6,7]env-13     UUCGCCUUUUCGAGGCGA..UCUG.CUAAACACAUCCGCGCGC....AGGUCCGGCGCUC...GAUCGGGCCCACUCCAGAA..........GGAGUGU.CCCA..GCGCG.GGGGGAACC...GA........UCCC.CAUCCCAUUCAGUUCAGCGCA
Ssp-1-1           CGGCCUUUUCGA.AGGCUCGUCUG.CUAGAAAAUUUUGCGCGC....AGGUCCGACCAUCAACAUUCGGGCCCACUUUCCUCAGU......GAAGGUGU.CCCU..GCGCA.GGGUGAUCU...AA........CUCC.CAC.UUCAUAACUCUCAGCGC
[3]Pma-3-1        CCAGGCCUAACCCCUCGG..UCUGGC.AAAC.UCCAUGCGCGC....AGGUCCGGCUU.....GUUCGGGCCCAUCUCACAAUCUC.....CGAGAUGU.CCCUG.GCGCUGGUAUCAACCUCUCG........GUCCCCAUACGCGCUGUUUUAGCGCC
Pma-5-1           UCAGAGAUAAAUUAUUUU..UCUGGCAAAAC.U.....CUUGCGAAUAGCUUCGAUUUUC....UUCGAACGCAUCUUAAUUAACUA....UGAGAUGUACCCCUCGCAGGUUUAAACUUC.CG..........GUUU.CUCUCAAAUACGCGCUUAGAA

                  .<<<<<.......>>>>>....................<<<<<.....<<<<<<<..........>>>>>>><<<<<<<............>>>>>>>........>>>>>.<<<<..................>>>>......................
                  .??2??.......??2??....................?22??.....???2?02..........20?2?????22?2?............?2?22??........??22?.??2?..................?2??......................
                  YYYR●●●●●●●●○○●YR●--UCUG○C○AAAY○Y○Y○○○CGCGY----AGGUCYGR●●●○----●●YCRGRCCCRYYYY●●●●○○○------●RRRRUGU-CC●●--GCGYR○●●R●●A●YY-○-○○--------●YY●-C●Y●●●Y●●●●●●YYY●●●●●
```

```
         Pma-4-1 AUUGAAUUAACUUUUUAA....UGGC.AAAUUU.....CCUGCUUUUAGGCCUGAAAUUC...UUUCAGUAUCAUCUCAAGAACUA.....UGAGAUGCACCCCUCGCAGGUAUAAACUCC.CG..........GUCA.CAUCUUUAUAAGCGUUUAAAG
     3,4 Pma-1-1 GCCAGAACUUUUUCUUGU..UCUGUCAAAGUC....ACUGCAAGC..AGGUCUGAAAUC....UUUCUGACCUGUCCUAAUGAACUU....UGAGAUUUACCC...UCGCAGGUAUAAACUCC.CG........GUUC.UGUUCUUUAUGCGCUUGAUAG
  5,6,22 env-14  AGAGCAUCAUUA.GGUUC..UUUAGGUACUUUUGAACCUGCGC....AGGUCCGAUGAG....UUUCGGGCCCAUUUCAAAAAAAUUU..UGAGAUGUCCCCC..ACGCAGGUAUAAACU.............CUCC.GUCAAUUAAUCAGCGUUUAAA
    5,21 env-15  .UUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAACA......AAGAAUGU.CCAA..GCGUGUAUGAC...........................................
    5,21 env-16  .UUACAAAGAGUCUGUAA..UCUGGCAAAAUGUCUAAUCGCGU....AGGUCUGAAUA.....UAUCAGACCCAUUCUUAAAAUA......AAGGAUGU.CCAA..GCGUGUA...............................................
  5,6,14 env-17  UUUACAAACUGCAAAUAA..UCUGCCAAAAUACCUAAUCGCGU....AGGUCUGAAAA.....AUUCAGACCCAUUCUUAAAUAAAAA...AAGAAUGU.CCAA..GCGUGUUUGACAAAGAU.UU........AUCA.CAAGCUCAAA...........
  5,6,15 env-18  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAAAA......GAGAAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CUCAAGUAUUAAGCUCUUUAU
  5,6,25 env-19  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAAUA......AAGAAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGCAUUAAGCUAUUUAU
  5,6,25 env-20  UUUACAAAUAGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAUA.....UAUCAGACCCAUUCUUAAAAGA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
       1 env-21  UUUACAAAGAGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAUA.....UAUCAGACCCAUUCUUAAAAGA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
  5,6,11 env-22  UUUACAAAGAGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAUA.....UAUCAGACCCAUUCUUAAAAGA......AAGGAUGU.CCAA..GCGUGAAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
    5,21 env-23  UUUACAAAGAGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAUA.....UAUCAGACCCAUUCUUAAAAGA......AAGGAUGU.CCAA..GCGUGUAUGACAAAAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
  5,6,16 env-24  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....AAUCAGACCCAUUCUUAAAACA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
    5,20 env-25  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAACA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
    5,20 env-26  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAACU......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
  5,6,18 env-27  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UCUCAGACCCAUUCUUAAAACA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
   5,6,8 env-28  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAACA......AAGGAUGU.CCAA..GCGUGUAUGACAAAAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
  5,6,11 env-29  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAUCA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
       1 env-30  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAAGA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
       1 env-32  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAACA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CAAAAGUAUUAAGCUCUUUAU
  5,6,16 env-33  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAAUA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
    5,20 env-34  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAACA......AAGAAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CAAAAGUAUUAAGCUCUUUAU
   5,6,7 env-35  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAUAGA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGCAUUAAGCUCUUUAU
  5,6,13 env-36  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAACA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
  5,6,13 env-37  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAAGC......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUUUUUAU
       1 env-38  UUUACAAAGAGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAUAU....AAUCAGACCCAUUCUUAAAAGA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
       1 env-39  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAUAU....AAUCAGACCCAUUCUUAAAAGA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUUUUUAU
       1 env-40  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAUAU....AAUCAGACCCAUUCUUAAAAGA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAC.CACAAGUAUUAAGCUUUUUAU
    5,21 env-41  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAAAAA.....AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
  5,6,10 env-42  UUUACAAAGUGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....CAUCAGACCCAUUCUUAAAACA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
  5,6,11 env-43  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAACCA......AAGAAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
  5,6,10 env-44  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAA........AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
  5,6,25 env-45  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAACA........AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
  5,6,11 env-46  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAC.A......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
       1 env-47  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAAGA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGCAUUAAGCUUUUUAU
  5,6,12 env-48  UUUACAAAGAGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAACA.....UAUCAGACCCAUUCUUAAAAGA......AAGGAUGU.CCAA..GCGUGGAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
       1 env-49  UUUACAAAGAGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAACA.....UAUCAGACCCAUUCUUAAAAGA......AAGGAUGU.CCAA..GCGUGGAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
  5,6,17 env-50  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAACUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAACA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUUUUUAU
  5,6,13 env-51  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAACCA.....AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
    5,21 env-52  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAA.A......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
   5,6,7 env-53  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAACUAAUCGCGU....AGGUCUGAGAA.....UAUCAGACCCAUUCUUAAAAGA......AAGGAUGU.CCAA..GCGUUUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
  5,6,15 env-54  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGUAAU......AUCAGACCCAUUCUUAAAA........AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
    5,21 env-55  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAUAUA...UAUCAGACCCAUUCUUAAAAUA......AAGGAUGU.CCAA..GCGUGAAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
  5,6,13 env-56  UUUACAAAGAGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAG.....UAUCAGACCCAUUCUUAAAAAA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUUUUUAU
  5,6,10 env-57  UUUACAAAGAGUCUGUAA..UCUGGCAAAAUAACUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUCUUUAAAAGA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
       1 env-58  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAUAAA.....AAGGAUGU.CCAA..GCGUGUAUGACACAAAA.GAU.......UUAU.CACAAGUAUUAAGCUCUUUAU
         Pma-2-1 UUUACAAAGGGUCUGUAA..UCUGGCAAGAUAUCUAAUUGCGU....AGGUCUGAAAAA....UAUCAGACCCAUUCUUAAAAGA......AAGGAUGU.CCAA..GCGUUUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
  5,6,13 env-59  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCAAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUACAAAGA.....AAGGAUGU.CCAA..GCGUGAAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
  5,6,18 env-60  UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAUAA......AAGGAUGU.CCAA..GCGUGUAUGACAAAAAAGAU........UUAU.CACAAGUAUUAAGCUCUUUAU
    5,20 env-61  UUUACAAAGUGUCUGUGA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAUUA......AAGGAUGU.CCAA..GCGUGUAUGACAAAAAAGAU........UUAU.CACAAGUAUUAAGCUCUUUAU

                 .<<<<<.......>>>>>....................<<<<<.....<<<<<<<..........>>>>>>><<<<<<<............>>>>>>>........>>>>>.<<<<..................>>>>......................
                 .??2??.......??2??....................?22??.....???2?02..........20?2???????22?2?............?2?22??........??22?.???2..................?2??......................
                 YYYR●●●●●●●●○●●YR●--UCUG○C○AAAY○Y○Y○○○CGCGY----AGGUCYGR●●●○----●●YCRGRCCCRYYYY●●●●○○○------●RRRRUGU-CC●●--GCGYR○●●R●●A●YY-○-○○--------●YY●-C●Y●●●Y●●●●●●YYY●●●●●
```

```
5,6,12    env-62    UUUACAAAGGUUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAUAAA.....AAGGAUGU.CCAA..GCGUGUAUGACACAAAAGAU........UUAU.CACAAGUAUUAAGCUCUUUAU
1         env-63    UUUACAAAAGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAUAA......AAGAAUGU.CCAA..GCGUGUAUGACAAAAAAAGAU.......UUAU.CACAAGUAUUAAGCUCUUUAU
5,6,12    env-64    NNNACAAAGGGUCUGUAA..UCUGGCAAAAUAUUUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAAGA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CAAAAGUAUUAAGCUCUUUAU
5,21      env-65    UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAG.....UAUCAGACCCAUUCUUAAUUUA......AAGAAUGU.CCAA..GCGUGUAUGACAAAAAAAGA........UUUAUCACAAGUAUUAAGCUCUUUAU
5,6,13    env-66    UUUACAAAAGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAUUA......AAGGAUGU.CCAA..GCGUGUAUGACAAAAAAAGA........UUUAUCACAAGUAUUAAGCUCUUUAU
5,6,12,24 env-67    UUUACAAAGUGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAU.....UAUCAGACCCAUUCUUAAAUAA......AAGAAUGU.CCAA..GCGUGUAUGACAAAAAAAGA........UUUAUCACAAGUAUUAAGCUCUUUAU
5,6,8     env-68    UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAUUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAGAGA......AAGAAUGU.CCAA..GCGUGAACGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
5,20      env-69    .UUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAACA......AAGAAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CAAAAGUAUUAAGCUCUUUAU
5,6,8     env-70    UUUACAAAUUGUCUGUAA..UCUGGCACAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUUAAUUA......AAGAAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUUUUUAU
5,6,9     env-71    UUUACAAAAGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....AAUCAGACCCAUUCUUAAAUUC......AAGGAUGU.CCAA..GCGUGAAUGACAAAAAAGAU........UUAU.CAAAAGUAUUUAGCUCUUUAU
5,6,16    env-72    UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....AAUCAGACCCAUUCUUAUAACA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
5,6,13    env-73    UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAACA......AAGGAUGU.CCAA..GCGUGUUUGACAACAAAG..........GUUAUCACAAGUAUUAAGCUUUUUAU
5,6,18    env-74    NNNNNNAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAACA.......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGCAUUAAGCUCUUUAU
5,6,25    env-75    UUUACAAAGGGUCUGUAA..UCUGGCACAAUAUCUAAUCGCGU....AGGUCUGACAA.....UAUCAGACCCAUUCUUAAAUUAA......AGGAUGU.CCAA..GCGUGUAUGACAAAAAA.GA........UUUUUUACAAGAAUUAAGCUCUUUAU
1         env-76    UUUACAAAGGGCCUGUAA..UCUGGCAAAAUAUUUAAUCGCGU....AGGUCUGAAGU.....AAUCAGACCCAUUCUUAAAAAA......AAGAAUGU.CCAA..GCGUGUUAGACAA.AAA.GA........UUUAUCACAAGCAUUAAGCUCUUUAU
          Pma-6-1   UUUACAAAGGGCCUGUAA..UCUGGCAAAAUAUUUAAUCGCGU....AGGUCUGAAAA.....AUUCAGACCCAUUCUUAAAAAAAA....AAGAAUGU.CCAA..GCGUGUUAGACAA.AAA.GA........UUUAUCACAAGCAUUAAGCUCUUUAU
5,6,12    env-77    UUUACAAAAGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAUAU....AAUCAGACCCAUUCUUAAAAGA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAAUAAGCUUUUUAU
1         env-78    UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCAAAUCACGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAACA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CAAAAGUAUUAAGCUCUUUAU
5,6,23    env-79    UUUACAAAGAGCAUAUAA..UCUGCCAAAAUACCUAAUCGCGU....AGGUCUGUUAA.....GUUCAGACCCAUUCUUUAAAAAACAA..AAGGAUGU.CCAA..GCGUGUUCGACAAAA...GA........UUUAUCUCAAGCUUUAAGCUUUUUAU
5,6,15    env-80    UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAUACA......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
5,26      env-81    UUUACAAAAGGCAUAUAA..UCUGCCAAAAUACCUAAUCGCGU....AGGUCUGAAAA.....AUUCAGACCCAUUCUUAAUUAGAUUAAU.AGAAUGU.CCAA..GCGCGUUUGACAAAU...AU........UUAU.CACAAGCUAAAAGCUUUU.AU
5,26      env-82    UUUACAAAAGGCAAAUAA..UCUGCCAAAAUACCUAAUCGCGU....AGGUCUGAAAA......AUCAGGCCCAUCCUAAAUAAACUAG..UAGGAUGU.CCAA..GCGCGUUUGACAAAU...AU........UUAU.CACAAGCUAAAAGCUUUU.AU
5,6,24    env-83    UUUACAAAAGGCAUAUAA..UCUGCCAAAAUACCUAAUCGCGU....AGGUCUGAAAA......AUCAGGCCCAUUCUAAAUUAACUAG..UAGGAUGU.CCAA..GCGCGUUUGACAAAUAU.UA........AUCA.C..AAGCUAAAAGCUUUU.AU
5,6,24    env-84    UUUACAAAACGCAUAUAA..UCUGCCAAAAUACCUAAUCGCGU....AGGUCUGAAAA......AUCAGGCCCAUUCUAAAUACAUUAAA.UAGAAUGU.CCAA..GCGCGUUUGACAAAU...AU........UUAU.CACAAGCUAAAAGCUUUUU..
5,6,18    env-85    UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCUAAUCGCGU....AGGUCUGAAUA.....UAUCAGACCCAUUCUUAAAACA......AAGGAUGU.CCAA..GCGUG.UAGACAACAAAG..........AUUAUCACAAGUAUUAAGCUCUUUAU
2         env-86    UUUACAAAGAGUCUGUAA..UCUGGCACAAUAUCUAAUCGCGU....AGGUCUGAAUA.....UAUCAGACCCAUUCUUAUAAAGA.....AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CACAAGUAUUAAGCUCUUUAU
5,6,12    env-87    UUUACAAAGGGUCUGUAA..UCUGGCAAAAUAUCAAAUCGCGU....AGGUCUGAAAA.....UAUCAGACCCAUUCUUAAAACG......AAGGAUGU.CCAA..GCGUGUAUGACAACAAA.GA........UUAU.CAAAAGUAUUAAGCUCUUUAU
          Ssp-5-1   CCCGGGCCGGUCGGCCGU..UCUG.CUAAAUCUCC.UGCGCGC....AGGUCCGGUCUG....AACCGGGCCCGUCUCCCCAGC.......GGAGAUGU.CCCAG.GCGCCAGGUUCAACCCGCGCUCUUCCGGCGCCCCACAUCCCCAACGUUCCACCA
                    .<<<<<.......>>>>>....................<<<<<.....<<<<<<<..........>>>>>>><<<<<<<............>>>>>>>........>>>>>.<<<<..................>>>>......................
                    .??2??.......??2??....................?22??.....???2?02..........20?2?????22?2?............?2?22??........??22?.??2?..................?2??......................
                    YYYR●●●●●●●●○●●YR●--UCUG○C○AAAY○Y○Y○○○CGCGY----AGGUCYGR●●●○----●●YCRGRCCCRYYYY●●●●○○○------●RRRRUGU-CC●●--GCGYR○●●R●●A●YY-○-○○--------●YY●-C●Y●●●Y●●●●●●YYY●●●●●
```

alignment positions 161 · · · 238

```
5,6,19    env-1     GCUUUCCACCAAUGAGUGAUCUGACCUCCGCCGCCUCCGUCCUGGCCGCCGCCCUCGCGGUGGGUCUCGCCGCCAUCG
          Ssp-4-1   G.CUUUCCACCA..................AUGAGUGAUCUGACCUCCGCCGCCUCCGUCCUGGCCGCCGCUCUCGCG
          Ssp-6-1   CGCUUUCCACCA..................AUGAGUGAUCUGACCUCCGCCGCUUCCGUUCUGGCCGCCGCUCUCGCG
5,6,7     env-2     CCGCAGCGCCCCCCAUAACUC.CUCUCAACAUGGAUUCCAUUACCUCCGCUGCUUCCGUUGUGGCCGCUGGUCUGGCU
5,6,22    env-3     CUGCGCAUCCCCCCCAACG...CUCUCAACAUGGAUUCCAUUACUUCCGCAGCUUCCGUCGUUGCUGCUGGCCUGGCU
5,6,22    env-4     CUGCGCAUCCCCCCCCCCAACGCUCUCAACAUGGAUUCCAUUACUUCCGCAGCUUCCGUCGUUGCUGCUGGCCUGGCU
5,6,23    env-5     CUGCGCAUCCCCCCCCCAACGCU.CUCAACAUGGAUUCCAUCACUUCCGCAGCUUCCGUCGUUGCUGCUGGCCUGGCU
5,6,24    env-6     UCCCGAGCGCUUCCCCUUAUCCCACCCAACAUGGAUUCCAUCACCUCCGCUGCUUCCGUUGUGGCUGCCGGUCUGGCU
          Ssp-3-1   UCCGUAGCGCUUCCCCUUAUCCCACCCAACAUGGAUUCCAUCACCUCCGCCGCUUCUGUUGUGGCUGCUGGCCUGGCC
5,6,7     env-7     AUCCCCGCGCUUCCCCUUACCCCACCCAACAUGGAUUCCAUCACCUCCGCCGCUUCCGUUGUGGCUGCUGGCCUGGCA
          Ssp-7-1   AUCCCCGCGCUUCCCCUUACCCCACCCAACAUGGAUUCCAUCACCUCCGCCGCUUCCGUUGUGGCUGCUGGCCUGGCA
1         env-8     AUCCCCGCGCUUCCCCUUACCCCACCCAACAUGGAUUCCAUCACCUCCGCCGCUUCCGUUGUGGCUGCUGGCCUGGCA
1         env-9     AUCCCCGCGCUUCCCCUUAUCCCACCCAACAUGGAUUCCAUCACCUCCGCCGCUUCCGUUGUGGCUGCUGGCCUGGCA
5,6,7     env-10    GUCUCAGCGCUUCCCCUAACCCCACUCAACAUGGAUUCCAUCACCUCCGCCGCUUCCGUUGUGGCUGCUGGCCUGGCA

                    ..............................................................................
                    ..............................................................................
                    Y●●●●●●Y●○○○○○○○○○○○○-○------AUGGAUUC●AUYACY●CCGCYGC●UC●GUYGU●GCYGCYGGYYURGC●
                                                  AUG  w  w  w  w  w  w  w  w  w  w  w  w  w  w  w
```

```
5,6,24 env-11    UUCCCAGCGCUUCCCCUUACACCACCCAACAUGGAUUCCAUUACCUCCGCCGCUUCCGUUGUGGCUGCUGGCCUGGCA
5,6,16 env-12    UUCCCAGCGCUUCCCCUUACACCACCCAACAUGGAUUCCAUCACCUCCGCCGCUUCCGUUGUGGCUGCUGGCCUGGCA
       Ssp-2-1   UCCCCUGCGCUUCCCCUUACACCACCCAACAUGGAUUCCAUCACCUCCGCCGCUUCCGUUGUGGCUGCUGGCCUGGCA
 5,6,7 env-13    CAACAGCGCUUCCCCUAACCCCA.CCCAACAUGGAUUCCAUCACCUCCGCUGCUUCCGUCGUGGCCGCUGGUCUGGCU
       Ssp-1-1   CUUCCGGCGCGCCCCCAAACGC.ACUCAACAUGGAUUCCAUCACUUCUGCAGCCUCCGUCGUUGCUGCUGGUCUGGCU
     3 Pma-3-1   CCACACCUUUAGUUCAACC...........AUGGAUUCCAUUACCACCGCCGCGUCUGUUGUUGCCGCUGGUCUAGCU
       Pma-5-1   GGCGCUUCAUUUCCUUAGCUCAAUU.....AUGGAUUCCAUUACCACCGCCGCUUCUGUUGUUGCCGCUGGUUUAGCU
       Pma-4-1   CGCUACACAUCUUUAGCAGAAUU.......AUGGAUUCCAUUACCACCGCCGCAUCAGUUGUUGCAGCUGGCUUAGCU
   3,4 Pma-1-1   CGCUUCUCUUCCUUACCUCAAUU.......AUGGAUUCCAUUACCACCGCCGCUUCUGUUGUUGCCGCUGGUUUAGCU
5,6,22 env-14    CGCAUUUUAUUUAAAACU............AUGGAUUCAAUUACCACUGCUGCUUCUGUUGUUGCUGCUGGCUUAGCU
  5,21 env-15    ..............................................................................
  5,21 env-16    ..............................................................................
5,6,14 env-17    ..............................................................................
5,6,15 env-18    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGGUUUAGC
5,6,25 env-19    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,25 env-20    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
     1 env-21    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,11 env-22    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
  5,21 env-23    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,16 env-24    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
  5,20 env-25    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
  5,20 env-26    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,18 env-27    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
 5,6,8 env-28    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,11 env-29    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
     1 env-30    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
     1 env-32    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,16 env-33    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUGGCA
  5,20 env-34    UUUAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
 5,6,7 env-35    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,13 env-36    UUCAAAAUU.....................AUGGAUUCGAUCACUUCCGCUGCAUCUGUUGUAGCUGCUGGUUUAGCA
5,6,13 env-37    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
     1 env-38    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
     1 env-39    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
     1 env-40    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
  5,21 env-41    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,10 env-42    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,11 env-43    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,10 env-44    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,25 env-45    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGCUUAGCA
5,6,11 env-46    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
     1 env-47    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,12 env-48    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
     1 env-49    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,17 env-50    UUUAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,13 env-51    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUGGCA
  5,21 env-52    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAG..
 5,6,7 env-53    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,15 env-54    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
  5,21 env-55    UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA

                 ..............................................................................
                 ..............................................................................
                 Y●●●●●●Y●○○○○○○○○○○○○-○-------AUGGAUUC●AUYACY●CCGCYGC●UC●GUYGU●GCYGCYGGYYURGC●
                                               AUG  w  w  w  w  w  w  w  w  w  w  w  w  w  w  w
```

```
5,6,13 env-56      UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,10 env-57      UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
     1 env-58      UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
       Pma-2-1     UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,13 env-59      UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,18 env-60      UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
  5,20 env-61      UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,12 env-62      UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
     1 env-63      UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,12 env-64      UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
  5,21 env-65      UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,13 env-66      UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,12,24 env-67   UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
 5,6,8 env-68      UUCAAA.UU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
  5,20 env-69      UUUAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGU......
 5,6,8 env-70      UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
 5,6,9 env-71      UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,16 env-72      UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUG........
5,6,13 env-73      UUUAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUUUAGCA
5,6,18 env-74      UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCAGUUGUAGCUGCUGGUCUAGCA
5,6,25 env-75      UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCUGUUGUAGCUGCUGGUUUGGCA
     1 env-76      UUCAAAAUU.....................AUGGAUUCCAUUACUUCCGCUGCAUCAGUUGUAGCUGCCGGUUUAGCA
       Pma-6-1     UUCAAAAUU.....................AUGGAUUCCAUUACUUCCGCUGCAUCUGUUGUAGCUGCUGGCUUAGCA
5,6,12 env-77      UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUGCAUCA.....................
     1 env-78      UUCAAAAUU.....................AUGGAUUCGAUUACUUCCGCUG..........................
5,6,23 env-79      UUCAAAAUU.....................AUGGAUUCUAUUACUUCCGCUGCAUCAGUUGUAGCUGCAGGUUUAGCA
5,6,15 env-80      UUCAAAAUU.....................AUGGAUUCGAUUACUUCGCUGCAUCAGUGUAGCUGCUGGUUUAGCANN
  5,26 env-81      UUCAAAAUU.....................AUGGAUUCCAUUACUUCCGCUGCAUCUGUUGUAGCUGCUGGUCUAGCA
  5,26 env-82      UUCAAAAUU.....................AUGGAUUCCAUUACUUCCGCUGCAUCUGUUGUAGCUGCUGGUCUAGCA
5,6,24 env-83      UUCAAAAUU.....................AUGGAUUCCAUUACUUCCGCUGCAUCUGUUGUAGCUGCUGGCCUAGCA
5,6,24 env-84      UUCAAAAUU.....................AUGGAUUCAAUUACUUCCGCUGCAUCUGUUGUGGCUGCUGGUUUAGCA
5,6,18 env-85      UUCAAAAUU.....................................................................
     2 env-86      UUCAAAA.......................................................................
5,6,12 env-87      UUCAAAAUU.....................AUGGAUUCGAUNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
       Ssp-5-1     UGGAUUCCAUCACUUCUGCAGCGUCUGUU.GUGGCUGCCGGUCUGGCCGUCGGCCUCGGCGCCAUCGGCCCCGGCAUC
                   ..............................................................................
                   ..............................................................................
                   Y●●●●●●Y●○○○○○○○○○○○○-○-------AUGGAUUC●AUYACY●CCGCYGC●UC●GUYGU●GCYGCYGGYYURGC●
                                                 AUG  w  w  w  w  w  w  w  w  w  w  w  w  w  w  w
```

2.15 The cyano-30S motif

No Shine-Dalgarno sequence is annotated because the Shine-Dalgarno sequence appears to be optional in Cyanobacteria [10].

Notes:

[1] The start codon is annotated in GenBank 183 nucleotides downstream of our annotation. Our annotation is based on homology with the sequence listed below, the fact that the ORF with the extra 183 nucleotides does not contain a stop codon, and the conservation of the start codon position relative to the putative RNA motif in the other sequences.

abbrev. of hits	**taxonomy of species**
Ssp-1-1 to Ssp-1-1	Cyanobacteria Chroococcales *Synechococcus sp.* CC9311
Ssp-2-1 to Ssp-2-1	Cyanobacteria Chroococcales *Synechococcus sp.* CC9605
Ssp-3-1 to Ssp-3-1	Cyanobacteria Chroococcales *Synechococcus sp.* CC9902
Ssp-4-1 to Ssp-4-1	Cyanobacteria Chroococcales *Synechococcus sp.* RS9917
Ssp-5-1 to Ssp-5-1	Cyanobacteria Chroococcales *Synechococcus sp.* WH 5701
Ssp-6-1 to Ssp-6-1	Cyanobacteria Chroococcales *Synechococcus sp.* WH 7805
Ssp-7-1 to Ssp-7-1	Cyanobacteria Chroococcales *Synechococcus sp.* WH 8102
Pma-1-1 to Pma-1-1	Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9211
Pma-2-1 to Pma-2-1	Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9312
Pma-3-1 to Pma-3-1	Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* MIT 9313
Pma-4-1 to Pma-4-1	Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus str.* NATL2A
Pma-5-1 to Pma-5-1	Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus subsp. marinus str.* CCMP1375
Pma-6-1 to Pma-6-1	Cyanobacteria Prochlorales Prochlorococcaceae *Prochlorococcus marinus subsp. pastoris str.* CCMP1986
env-1 to env-13	environmental samples

abbrev	RefSeq accession		5′ at	3′ at	genes
[1]Ssp-1-1	NC_008319.1	-	514260	514162	**RNA**→ RpsA (COG0539)→ Gph (COG0546)→ MetK (COG0192)→ XylB (COG1070)→
Ssp-6-1	NZ_AAOK01000002.1	+	36286	36384	**RNA**→ RpsA (COG0539)→ Gph (COG0546)→ MetK (COG0192)→ XylB (COG1070)→
Ssp-4-1	NZ_AANP01000005.1	+	18103	18201	**RNA**→ RpsA (COG0539)→ Gph (COG0546)→ MetK (COG0192)→ XylB (COG1070)→
Pma-3-1	NC_005071.1	+	1772187	1772284	**RNA**→ RpsA (COG0539)→ Hydrolase (pfam00702)→ MetK (COG0192)→ XylB (COG1070)→
Pma-1-1	NZ_AALP01000001.1	+	1029265	1029364	**RNA**→ RpsA (COG0539)→ Gph (COG0546)→ MetK (COG0192)→ XylB (COG1070)→
Pma-5-1	NC_005042.1	-	339581	339484	**RNA**→ RpsA (COG0539)→ Gph (COG0546)→ MetK (COG0192)→ XylB (COG1070)→
Pma-4-1	NC_007335.1	-	1650069	1649969	**RNA**→ RpsA (COG0539)→ Gph (COG0546)→ MetK (COG0192)→ XylB (COG1070)→
Ssp-5-1	NZ_AANO01000001.1	-	404983	404887	**RNA**→ RpsA (COG0539)→ Gph (COG0546)→ MetK (COG0192)→ XylB (COG1070)→
Ssp-7-1	NC_005070.1	+	1891495	1891596	**RNA**→ RpsA (COG0539)→ Gph (COG0546)→ MetK (COG0192)→ XylB (COG1070)→
env-1	AACY01076090.1	+	325	426	**RNA**→ unknown→ unknown→ unknown→ unknown→
Ssp-2-1	NC_007516.1	-	458247	458146	**RNA**→ RpsA (COG0539)→ Hydrolase (pfam00702)→ MetK (COG0192)→ XylB (COG1070)→
Ssp-3-1	NC_007513.1	+	1794279	1794380	**RNA**→ RpsA (COG0539)→ Gph (COG0546)→ MetK (COG0192)→ XylB (COG1070)→
env-2	AACY01100904.1	+	89	183	**RNA**→ unknown→
Pma-2-1	NC_007577.1	-	304844	304750	**RNA**→ RpsA (COG0539)→
env-3	AACY01012870.1	+	315	409	**RNA**→ unknown→
env-4	AACY01114579.1	-	151	57	**RNA**→
env-5	AACY01287182.1	-	295	201	**RNA**→ unknown→
env-6	AACY01124570.1	+	40	135	**RNA**→ unknown→
env-7	AACY01809377.1	+	553	648	**RNA**→ unknown→
env-8	AACY01246174.1	-	117	22	**RNA**→
env-9	AACY01052204.1	-	3538	3443	**RNA**→ unknown→
env-10	AACY01667173.1	-	725	630	**RNA**→ unknown→
env-11	AACY01737885.1	-	223	128	**RNA**→
env-12	AACY01171845.1	+	167	261	**RNA**→ unknown→
Pma-6-1	NC_005072.1	-	300337	300241	**RNA**→ RpsA (COG0539)→
env-13	AACY01035039.1	-	993	897	**RNA**→ unknown→

COG0192 S-adenosylmethionine synthetase [Coenzyme metabolism]
COG0539 Ribosomal protein S1 [Translation, ribosomal structure and biogenesis]
COG0546 Predicted phosphatases [General function prediction only]
COG1070 Sugar (pentulose and hexulose) kinases [Carbohydrate transport and metabolism]
pfam00702 haloacid dehalogenase-like hydrolase.

alignment positions 1 · · · 107

```
[1]Ssp-1-1  UUCUGU.AAAAUU.AAAAAUUACUGACCCUUUGUCGGCGGGCGC.UG.GGUCAAUUUUUAUCGCACUGUCC...UCGGGCAGACCGCCAUCCAUCC.AUGACUGUUA
Ssp-6-1     CUCUGU.AAAGUGAGGGAUUGCCUGGAC.CAGGUCGGCGGGCCU..G.CUCCAGUUUCCUUCGCUCUGCUCA..ACGGGCAGACCGCCACCCCUCC.AUGACUGUCA
Ssp-4-1     AUCUGACAGAAUGACGGAUUCGAUCGUU.GUGGUCGGCGGGCCA..C.CACGUCAUUCCUUCGCACUGUCU...CCGGGCAGACCGCCAUCGCUCC.AUGUCUGUUU
Pma-3-1     AUCUGU.AGAGUGAUCAAUUAAUUCCGA.AAGGUCGGCGGGCCA..C.UCGGAUUUCCUUGCGCACUGCCA...UAGGGCAGACCGCCAUCCAACA.AUGUCUGCGA
Pma-1-1     AUAUGU.AGAGUG.UUGUAUUAGUUCCUCUAGGUAGGCAGGCCU..A.AGGGAAAUACUUUCGCAUUGCCAUA.ACAGGCAUACUGCUAAUUAAUC.AUGGUUGAAA
Pma-5-1     AGGUGU.AGAGUGAGUUAUCUGUUCCUU.UAGGUUGGCAGGCCU..A.AAGGGCAAAUGUACGCAUUGCCU...CAAGGCAUACUGCCCGCAAACC.AUGUCUGAAA
Pma-4-1     AUCAAU.AAAGUAGAAUUUAAUCUCAAU.UUGGUCGGCGGGCCA..U.UUUGUAUUCCUAUCGCACUGUUCUAAAAGGCAGACCGCCAUUCAAGC.AUGUCUGAAA
Ssp-5-1     AUUUGU.AUGAUC.AUGGAUCGCAUGCG.UCUGUCGGCGGGCAGGCA.CGCGUAU.CCCUUCCUUCUGCCC...UGA.GCAGACCGCCAGCCUUCC.AUGACCGUGA
Ssp-7-1     AUCUGU.AGGAUU.CAUAGUUCUCCUUGAACCGUCGGCGGGCGG.UU.CAAGCGAACCACUCGUACUACCCGCAUGAGGUAGACCGCCAUACCCCC.AUGUCCGCGA
env-1       AUCUGU.AGGAUU.CAUAAUUCUCCUUGAACCGUCGGCGGGCGG.UU.CAUGCGAACCACUCGUACUACCCGCAUGAGGUAGACCGCCAUACCCCC.AUGUCCGCGA
Ssp-2-1     GUCUGU.AAAGUC.AGUAGUUCUCCAUGAAACGUCGGCGGGCGU..UCCAUGCGAUUUUUUCGAACUACCCACCUGAGGUAGACCGCCAGCCCCCC.AUGUCUGCGA
Ssp-3-1     AUCUGU.AGAGUC.ACACAUUCUCCGUUCAACGUCGGCGGGCGU..UGGACGUGAAUCCUUCGUACUACCCACCUGAGGUAGACCGCCAACCCCCC.AUGGCCGCGA
env-2       AUACAU.AUCAAA.AUG.UUUUCUGCUA.UUAGUUUGCAGGCUA..A.UGGCGUU.CCUUUCUAACUACC....GAAGGUAGUCUGCGAUGUAACUAAUGAACGAAA
Pma-2-1     AUACAU.AACACA.AAU.AUUUUUGCUA.UUGGUUUGCAGGCUA..A.UAGCACU.CCUUUCUAACUACC....UAAGGUAGUCUGCGAUGCAACAAAUGAACGAAA
env-3       AUACAU.AUCACA.AAC.GUUUUUGCUA.UUGGUUUGCAGGCUA..G.UAGCAUU.CCUUUCUAACUACC....UUAGGUAGUCUGCGAUACAACAAAUGAACGAAA
env-4       AUACAU.AUCACA.AAU.GUUUUUGCUA.UUGGUUUGCAGGCUA..G.UAGCAUU.CCUUUCUAACUACC....UUAGGUAGUCUGCGAUACGACAAAUGAACGAAA
env-5       AUACAU.AACACA.ACU.GUUUUUGCUA.UUGGUUUGCAGGCUA..G.UAGCAUU.CCUUUCUAACUACC....UUAGGUAGUCUGCGUUACAACAAAUGAACGAAA
env-6       AUACAU.AUCACA.AAUGUUUUUUGCUA.UUGGUUUGCAGGCUA..G.UAGCAUU.CCUUUCUAACUACC....UUAGGUAGUCUGCGACACAACAAAUGAACGAAA
env-7       AUACAU.AUCACA.AAUGUAUUUUGCUA.UUGGUUUGCAGGCUA..G.UAGCAUU.CCUUUCUAACUACC....UUAGGUAGUCUGCGACGCAACAAAUGAACGAAA
env-8       AUACAU.AUCACA.AAUGUUUUUUGCUA.UUAGUUUGCAGGCUA..A.UAGCAUU.CCUUUCUAACUACC....UUAGGUAGUCUGCGAUACAACAAAUGAACGAAA
env-9       AUACAU.AACACA.AAUGUAUAUUGCUA.UUGGUUUGCAGGCUA..G.UAGCAUC.CCUUUCUAACUACC....UUAGGUAGUCUGCGAUACAACAAAUGAACGAAA
env-10      AUACAG.AAUACA.AAUGUUUUUUGCUA.CUAGUUUGCAGGCUA..G.UAACAUU.CCUUUCUAACUACC....UUAGGUAGUCUGCGAUACAACAAAUGAACGAAA
env-11      AUACAU.AGCACA.GAUAUUUUUUGCUA.UUAGUUUGCAGGCUG..A.UAGCAUU.CCUUUCUAACUACC....AUAGGUAGUCUGCGAUGCUAAAAAUGAACGAAA
env-12      AUAAAU.AACACA.AAU.GUUUUUGCUA.CUAGUUUGCAGGCUG..G.UAGCAUU.CCUUUCUAACUACC....UUAGGUAGUCUGCGAUGCACAAAAUGAACGAAA
Pma-6-1     GAAUAU.AACGUA.CAAAUAAUUUGCUCUUUAGUUUGCAGGCUA..A.AAGCAUU.UCUUUCUAACCAUC....UUAGGUGGUCUGCGUUUUAACAAAUGAUCGAAA
env-13      GAAUAU.AACUUA.CAAAUAAUUUGCUCUUUAGUUUGCAGGCUA..A.AAGCGUU.CCUUUCUAACCAUC....UUAGGUGGUCUGCGUUUUAACAAAUGAUCAAAA
            ........................<<<<.<<<<<......>>>>..>.>>>>.............<<<<<.......>>>>>.........................
            ........................2?2?.??200......002?..?.?2?2.............?221?.......?122?.........................
            ...................................<<<<<...........................................>>>>>...................
            ...................................20020...........................................02002...................
            RU•YRU-AR•RU•-•••••U••YU••Y•-•••GUY•GCRGGC••--•-•R•••••○YCYUUC••ACURYC○---••RGGYAG•CYGC•A••Y••C•-AUG•YYG•RA
                                                                                                       AUG
```

2.16 The lacto-1 *(Lactobacillus)* motif

Notes:

[1] These putative homolog overlaps S(MK) box RNA (a SAM-binding riboswitch) [3]. They are all between the main stem of this riboswitch class and the Shine-Dalgarno sequence. However, there are some instances of the S(MK) riboswitch that have about 50 nucleotides between the main stem and the SD, but for which we have found no homolog for the current motif.

[2] This homolog is truncated on its 5′ end because it at the end of a sequence fragment.

[3] This homolog overlaps a short ORF, but the ORF is not conserved in any bacteria

abbrev. of hits **taxonomy of species**

Lla-1-1 to Lla-1-4	Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. cremoris* SK11
Lla-2-1 to Lla-2-4	Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. lactis* Il1403
Smu-1-1 to Smu-1-6	Firmicutes Lactobacillales Streptococcaceae *Streptococcus mutans* UA159
Spn-1-1 to Spn-1-5	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* R6
Spn-2-1 to Spn-2-12	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* TIGR4
Spy-1-1 to Spy-1-5	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* M1 GAS
Spy-2-1 to Spy-2-3	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* M49 591
Spy-3-1 to Spy-3-5	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10270
Spy-4-1 to Spy-4-3	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10394
Spy-5-1 to Spy-5-6	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10750
Spy-6-1 to Spy-6-3	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS2096
Spy-7-1 to Spy-7-5	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS315
Spy-8-1 to Spy-8-5	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS5005
Spy-9-1 to Spy-9-4	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS6180
Spy-10-1 to Spy-10-5	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS8232
Spy-11-1 to Spy-11-2	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS9429
Spy-12-1 to Spy-12-5	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* SSI-1
Ssu-1-1 to Ssu-1-9	Firmicutes Lactobacillales Streptococcaceae *Streptococcus suis* 89/1591
Sth-1-1 to Sth-1-2	Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* CNRZ1066
Sth-2-1 to Sth-2-2	Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* LMD-9
Sth-3-1 to Sth-3-2	Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* LMG 18311

abbrev	**RefSeq accession**		**5′ at**	**3′ at**	**genes**
Lla-2-1	NC_002662.1	+	739381	739491	←MenF (COG1169) **RNA**→ RimL (COG1670)→
Lla-1-1	NZ_AAGO01000139.1	-	3688	3578	←Gnd (COG1023) **RNA**→ RpiR (COG1737)→
Spy-2-1	NZ_AAFV01000029.1	-	9122	9044	**RNA**→ NrdF (COG0208)→
Spy-10-1	NC_003485.1	+	1161915	1161993	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-12-1	NC_004606.1	-	804250	804172	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-7-1	NC_004070.1	+	1092633	1092711	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-3-1	NC_008022.1	+	1151406	1151484	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-4-1	NC_006086.1	+	1096428	1096506	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-6-1	NC_008023.1	+	1152407	1152485	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-4-2	NC_006086.1	+	1096423	1096506	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-11-1	NC_008021.1	+	1128501	1128573	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-8-1	NC_007297.1	+	1100560	1100655	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-1-1	NC_002737.1	+	1144210	1144305	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
[3]Spy-8-2	NC_007297.1	+	1100431	1100538	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-9-1	NC_007296.1	+	1123736	1123843	RNR_1 (cd01679)→ **RNA**→ ←hypo
Spy-3-2	NC_008022.1	+	1151277	1151384	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-5-1	NC_008024.1	+	1173974	1174081	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-5-2	NC_008024.1	+	1174103	1174210	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-1-2	NC_002737.1	+	1144081	1144188	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-10-2	NC_003485.1	+	1161786	1161893	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-12-2	NC_004606.1	-	804379	804272	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-7-2	NC_004070.1	+	1092504	1092611	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-8-3	NC_007297.1	+	1100426	1100538	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→

Spy-9-2	NC_007296.1	+	1123731	1123843	RNR_1 (cd01679)→ **RNA**→ ←hypo
Spy-3-3	NC_008022.1	+	1151272	1151384	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-5-3	NC_008024.1	+	1173969	1174081	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-1-3	NC_002737.1	+	1144076	1144188	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-10-3	NC_003485.1	+	1161781	1161893	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-12-3	NC_004606.1	-	804384	804272	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-7-3	NC_004070.1	+	1092499	1092611	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-6-2	NC_008023.1	+	1152284	1152385	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
[2]Spy-8-4	NC_007297.1	+	1100459	1100538	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-9-3	NC_007296.1	+	1123764	1123843	RNR_1 (cd01679)→ **RNA**→ ←hypo
Spy-3-4	NC_008022.1	+	1151305	1151384	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-5-4	NC_008024.1	+	1174002	1174081	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-5-5	NC_008024.1	+	1174131	1174210	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-1-4	NC_002737.1	+	1144109	1144188	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-10-4	NC_003485.1	+	1161814	1161893	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-12-4	NC_004606.1	-	804351	804272	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-7-4	NC_004070.1	+	1092532	1092611	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
Spy-2-2	NZ_AAFV01000029.1	-	9223	9144	**RNA**→ NrdF (COG0208)→
Smu-1-1	NC_004350.1	-	627986	627874	RNR_1 (cd01679)→ **RNA**→ NrdF (COG0208)→
[1]Smu-1-2	NC_004350.1	-	1500823	1500739	hypo→ **RNA**→ MetK (COG0192)→
Spn-2-1	NC_003028.1	+	497109	497217	GrpE (cd00446)→ **RNA**→ HSP70 (pfam00012)→
Spn-2-2	NZ_AAGY02000057.1	+	7485	7593	GrpE (cd00446)→ **RNA**→ HSP70 (pfam00012)→
Spn-1-1	NC_003098.1	+	461692	461800	GrpE (cd00446)→ **RNA**→ HSP70 (pfam00012)→
Spn-2-3	NC_003028.1	+	413993	414106	PrfC (COG4108)→ **RNA**→ RpmB (COG0227)→
Spn-1-2	NC_003098.1	+	393914	394027	PrfC (COG4108)→ **RNA**→ RfaJ (COG1442)→
Spn-2-4	NZ_AAGY02000036.1	-	9090	8977	PrfC (COG4108)→ **RNA**→ hypo→
Spn-2-5	NC_003028.1	+	1989945	1990059	ArgS (COG0018)→ **RNA**→ ←COG3293 (COG3293)
Spn-2-6	NZ_AAGY02000106.1	-	254	140	ArgS (COG0018)→ **RNA**→
Spn-1-3	NC_003098.1	+	1870878	1870992	ArgS (COG0018)→ **RNA**→ ←COG3293 (COG3293)
Spn-2-7	NC_003028.1	-	557548	557438	hypo→ **RNA**→ ←MTHFR (cd00537)
Spn-1-4	NC_003098.1	-	515396	515286	←Pnp (COG1185) **RNA**→ ←MTHFR (cd00537)
Spn-2-8	NZ_AAGY02000055.1	+	7747	7857	←Pnp (COG1185) **RNA**→ ←MTHFR (cd00537)
Lla-2-2	NC_002662.1	-	2274174	2274060	SSB (pfam00436)→ **RNA**→ Ribosomal_S18 (pfam01084)→
Lla-1-2	NZ_AAGO01000139.1	+	898	1012	SSB (pfam00436)→ **RNA**→ Ribosomal_S18 (pfam01084)→
Sth-3-1	NC_006448.1	-	1391788	1391675	PrfC (COG4108)→ **RNA**→ ←DUF894 (pfam05977)
Sth-1-1	NC_006449.1	-	1399434	1399321	PrfC (COG4108)→ **RNA**→ ←DUF894 (pfam05977)
Sth-2-1	NZ_AAGS01000029.1	+	9621	9734	PrfC (COG4108)→ **RNA**→ ←DUF894 (pfam05977)
Sth-3-2	NC_006448.1	-	1389119	1389008	←HTH_XRE (cd00093) **RNA**→ ←COG3859 (COG3859)
Sth-1-2	NC_006449.1	-	1396765	1396654	←HTH_XRE (cd00093) **RNA**→ ←COG3859 (COG3859)
Sth-2-2	NZ_AAGS01000029.1	+	12290	12401	←HTH_XRE (cd00093) **RNA**→ ←COG3859 (COG3859)
Lla-1-3	NZ_AAGO01000069.1	+	2935	3046	NrdA (COG0209)→ **RNA**→ NrdF (COG0208)→
Lla-2-3	NC_002662.1	-	1002628	1002515	NrdA (COG0209)→ **RNA**→ NrdF (COG0208)→
Smu-1-3	NC_004350.1	+	1752431	1752549	CorA (COG0598)→ **RNA**→ DUF1129 (pfam06570)→

Lla-1-4	NZ_AAGO01000030.1	+	7079	7190	Biotin_lipoyl (pfam00364)AceF (COG0508)→ **RNA**→ Lpd (COG1249)→
Lla-2-4	NC_002662.1	-	60576	60465	Biotin_lipoyl (pfam00364)AceF (COG0508)→ **RNA**→ Lpd (COG1249)→
Ssu-1-1	NZ_AAFA02000013.1	-	20262	20151	Obg (COG0536)→ **RNA**→ hypo (cons)→
Smu-1-4	NC_004350.1	+	747449	747565	Obg (COG0536)→ **RNA**→ hypo→
Ssu-1-2	NZ_AAFA02000130.1	-	3896	3810	COG5562 (COG5562)→ **RNA**→ GlyRS_alpha_core (cd00733)→
Ssu-1-3	NZ_AAFA02000013.1	-	20387	20276	Obg (COG0536)→ **RNA**→ hypo (cons)→
Ssu-1-4	NZ_AAFA02000075.1	+	88	198	**RNA**→ ←COG4043 (COG4043)
Ssu-1-5	NZ_AAFA02000084.1	-	4034	3925	CarA (COG0505)→ **RNA**→ MGS_CPS_II (cd01424)CarB (COG0458)CPSase_L_D3 (pfam02787)→
Smu-1-5	NC_004350.1	-	751253	751138	GlnQ (COG1126)→ **RNA**→ ←hypo
Spn-2-9	NC_003028.1	+	50741	50853	PurF (COG0034)→ **RNA**→ PurM (COG0150)→
Spn-2-10	NZ_AAGY02000011.1	+	9950	10062	PurF (COG0034)→ **RNA**→ PurM (COG0150)→
Smu-1-6	NC_004350.1	+	87538	87653	HSP70 (pfam00012)→ **RNA**→ DnaJ (COG0484)→
Ssu-1-6	NZ_AAFA02000019.1	-	7733	7618	PheS (COG0016)→ **RNA**→ Acetyltransf_1 (pfam00583)→
Ssu-1-7	NZ_AAFA02000019.1	-	7385	7291	PheS (COG0016)→ **RNA**→ Acetyltransf_1 (pfam00583)→
Spn-2-11	NC_003028.1	+	1989817	1989935	ArgS (COG0018)→ **RNA**→ ←COG3293 (COG3293)
Spn-1-5	NC_003098.1	+	1870750	1870868	ArgS (COG0018)→ **RNA**→ ←COG3293 (COG3293)
Spn-2-12	NZ_AAGY02000106.1	-	382	264	ArgS (COG0018)→ **RNA**→
Ssu-1-8	NZ_AAFA02000072.1	-	1767	1658	Peptidase_U32 (pfam01136)→ **RNA**→ Peptidase_U32 (pfam01136)→
[1]Spy-4-3	NC_006086.1	-	1079086	1078976	COG4886 (COG4886)→ **RNA**→ MetK (COG0192)→
Spy-8-5	NC_007297.1	-	1083089	1082979	COG4886 (COG4886)→ **RNA**→ MetK (COG0192)→
Spy-9-4	NC_007296.1	-	1104930	1104820	COG4886 (COG4886)→ **RNA**→ MetK (COG0192)→
Spy-11-2	NC_008021.1	-	1109698	1109588	COG4886 (COG4886)→ **RNA**→ MetK (COG0192)→
Spy-3-5	NC_008022.1	-	1133935	1133825	COG4886 (COG4886)→ **RNA**→ MetK (COG0192)→
Spy-6-3	NC_008023.1	-	1133481	1133371	COG4886 (COG4886)→ **RNA**→ MetK (COG0192)→
Spy-5-6	NC_008024.1	-	1155128	1155018	COG4886 (COG4886)→ **RNA**→ MetK (COG0192)→
Spy-1-5	NC_002737.1	-	1126738	1126628	COG4886 (COG4886)→ **RNA**→ MetK (COG0192)→
Spy-10-5	NC_003485.1	-	1144443	1144333	COG4886 (COG4886)→ **RNA**→ MetK (COG0192)→
Spy-12-5	NC_004606.1	+	821711	821821	COG4886 (COG4886)→ **RNA**→ MetK (COG0192)→
Spy-7-5	NC_004070.1	-	1075172	1075062	COG4886 (COG4886)→ **RNA**→ MetK (COG0192)→
Spy-2-3	NZ_AAFV01000074.1	-	1507	1397	COG4886 (COG4886)→ **RNA**→ MetK (COG0192)→
Ssu-1-9	NZ_AAFA02000019.1	-	7614	7514	PheS (COG0016)→ **RNA**→ Acetyltransf_1 (pfam00583)→

cd00093 Helix-turn-helix XRE-family like proteins.

cd00446 GrpE is the adenine nucleotide exchange factor of DnaK (Hsp70)-type ATPases.

cd00537 Methylenetetrahydrofolate reductase (MTHFR).

cd00733 Class II Glycyl-tRNA synthetase (GlyRS) alpha subunit core catalytic domain.

cd01424 Methylglyoxal synthase-like domain from type II glutamine-dependent carbamoyl phosphate synthetase (CSP).

cd01679 RNR, class I.

COG0016 Phenylalanyl-tRNA synthetase alpha subunit [Translation, ribosomal structure and biogenesis]

COG0018 Arginyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]

COG0034 Glutamine phosphoribosylpyrophosphate amidotransferase [Nucleotide transport and metabolism]

COG0150 Phosphoribosylaminoimidazole (AIR) synthetase [Nucleotide transport and metabolism]

COG0192 S-adenosylmethionine synthetase [Coenzyme metabolism]

COG0208 Ribonucleotide reductase, beta subunit [Nucleotide transport and metabolism]

COG0209 Ribonucleotide reductase, alpha subunit [Nucleotide transport and

	metabolism]
COG0227	Ribosomal protein L28 [Translation, ribosomal structure and biogenesis]
COG0458	Carbamoylphosphate synthase large subunit (split gene in MJ) [Amino acid transport and metabolism / Nucleotide transport and metabolism]
COG0484	DnaJ-class molecular chaperone with C-terminal Zn finger domain [Posttranslational modification, protein turnover, chaperones]
COG0505	Carbamoylphosphate synthase small subunit [Amino acid transport and metabolism / Nucleotide transport and metabolism]
COG0508	Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase (E2) component, and related enzymes [Energy production and conversion]
COG0536	Predicted GTPase [General function prediction only]
COG0598	Mg2+ and Co2+ transporters [Inorganic ion transport and metabolism]
COG1023	Predicted 6-phosphogluconate dehydrogenase [Carbohydrate transport and metabolism]
COG1126	ABC-type polar amino acid transport system, ATPase component [Amino acid transport and metabolism]
COG1169	Isochorismate synthase [Coenzyme metabolism / Secondary metabolites biosynthesis, transport, and catabolism]
COG1185	Polyribonucleotide nucleotidyltransferase (polynucleotide phosphorylase) [Translation, ribosomal structure and biogenesis]
COG1249	Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component, and related enzymes [Energy production and conversion]
COG1442	Lipopolysaccharide biosynthesis proteins, LPS:glycosyltransferases [Cell envelope biogenesis, outer membrane]
COG1670	Acetyltransferases, including N-acetylases of ribosomal proteins [Translation, ribosomal structure and biogenesis]
COG1737	Transcriptional regulators [Transcription]
COG3293	Transposase and inactivated derivatives [DNA replication, recombination, and repair]
COG3859	Predicted membrane protein [Function unknown]
COG4043	Uncharacterized conserved protein [Function unknown]
COG4108	Peptide chain release factor RF-3 [Translation, ribosomal structure and biogenesis]
COG4886	Leucine-rich repeat (LRR) protein [Function unknown]
COG5562	Phage envelope protein [General function prediction only]
pfam00012	Hsp70 protein.
pfam00364	Biotin-requiring enzyme.
pfam00436	Single-strand binding protein family.
pfam00583	Acetyltransferase (GNAT) family.
pfam01084	Ribosomal protein S18.
pfam01136	Peptidase family U32.
pfam02787	Carbamoyl-phosphate synthetase large chain, oligomerisation domain.
pfam05977	Bacterial protein of unknown function (DUF894).
pfam06570	Protein of unknown function (DUF1129).

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Spn-1-2, Spn-1-4, Spn-1-5, Spn-2-10, Spn-2-12, Spn-2-2, Spn-2-4, Spn-2-6, Spn-2-8, Spy-1-1, Spy-1-2, Spy-1-3, Spy-1-4, Spy-1-5, Spy-10-2, Spy-10-3, Spy-10-4, Spy-10-5, Spy-11-2, Spy-12-2, Spy-12-3, Spy-12-4, Spy-12-5, Spy-2-2, Spy-3-2, Spy-3-3, Spy-3-4, Spy-3-5, Spy-5-1, Spy-5-2, Spy-5-3, Spy-5-4, Spy-5-5, Spy-5-6, Spy-6-1, Spy-6-3, Spy-7-1, Spy-7-2, Spy-7-3, Spy-7-4, Spy-7-5, Spy-8-5, Spy-9-1, Spy-9-2, Spy-9-3, Spy-9-4, Sth-1-1, Sth-1-2, Sth-2-1, Sth-2-2

alignment positions 1 · · · 131

```
Lla-2-1    CUGCUUAUAA.CAC.GC.U.AAA.GC.GUGA.ACGCGCAG..ACAGAAGCCACU.AU.GGUGCCUUUCCUAGAUACCUUAUUA.C.UCACUCAGC.GGAAA.AUUAUCGUUCCCGCUUCGUU.....CGGU
Lla-1-1    CUGCUUAUAA.CAC.GC.U.AAA.GC.GUGA.ACGCGCAG..ACGGAAGUCGCU.AU.AGCGAGCUUCCUAGAUACCUUAUUA.C.UCACUCAGC.GGAAA.AUUAUCGUUUUCGCAUCGUU.....CGGU
Spy-2-1    .....UAAU..GAU.GC.U.AAA.GC.AUCUAUUAUUUCU..UGGGCAAUCGCU.UUAGGCGAUCUGCCUAAACUACUUACUAAC.UC....................A...........UU.....CGUU
Spy-10-1   .....UAAU..GAU.GC.U.AAA.AC.AUCUAUUACUUCU..UGGGCAAUCGCU.UUAGGCGUUCCGCCUAAACCACUCACUAAC.UC....................A...........UU.....CGUU
Spy-12-1   .....UAAU..GAU.GC.U.AAA.GC.AUCUAUUACUUCU..UGGGCAAUCGCU.UUAGGCGUUCCGCCUAAACCACUCACUAAC.UC....................A...........UU.....CGUU
Spy-3-1    .....UAAU..GAU.GC.U.AAA.GC.AUCUAUUACUUAU..UGGGCAAUCGCU.UUAGGCGUUCCGCCUAAACCACUCACUAAC.UC....................A...........UU.....CGUU
Spy-4-1    .....UAAU..GAU.GC.U.AAA.GC.AUCUAUUACUUCU..ACGGCAGUCGCU.UUAGGCGAUCUGCCUAAACCACUCACUAACUCA................................UU.....CGUU
Spy-4-2    AGAAGUAAU..GAU.GC.U.AAA.GC.AUCUAUUACUUCU..ACGGCAGUCGCU.UUAGGCGAUCUGCCUAAACCACUCACUAAC.UC....................A...........UU.....CGUU
Spy-11-1   AGAAGUAAU..GAU.GC.U.AAA.............UUCU..ACGGCAGUCGCU.UUAGGCGAUCUGCCUAAACCACUCACUAAC.UC....................A...........UU.....CGUU
Spy-8-1    .....UAAU..GAU.GC.U.AAA.AC.AUCUAUUACUUCU..UGGGCAAUCGCU.UUAGGCGAUCUGCCUAAACCACUCACUAAC.UCAUUCGUU.A.........CCC....UUUUGAAUC.....UAUU

           <<<<<<<<<..<<<.<<.......>>.>>>.>>>>>>>>>....<<<<<<<<<......>>>>.>>>>>....................<<<<<<.<<<<<........>>>>>>>>>>>...........
           ?????????..2??.??.......??.??2.?????????....?2222????......????.2222?....................?????2.22?22........22?222?????...........
           ○○○○○●○○○--R○●-RC-U-●AA-RY-●○YU○●Y●●●●●●--ACG●●ARYY●CU-●U-RG●R●●YU●●CUAR●Y●YCUYACUARY-U●○●●●●●●-●●AR○-AR●AUCG○YU●●●●●●○●●●-----CGU●
```

```
[3]Spy-8-2   .....UAAU..GAU.GC.U.AAA.GC.AUCUAUUACUUCU..ACGGCAGUCGCU.UUAGGCGAUCUGCCUAAACCACUCACUAAC.UACGUAAGG.AAAGU.AUUAUCGACUUUCCUUACUC.....CGUG
   Spy-8-3   AGAAGUAAU..GAU.GC.U.AAA.GC.AUCUAUUACUUCU..ACGGCAGUCGCU.UUAGGCGAUCUGCCUAAACCACUCACUAAC.UACGUAAGG.AAAGU.AUUAUCGACUUUCCUUACUC.....CGUG
   Spy-6-2   AGAAGUAAU..GAU.GC.U.AAA.............UUCU..ACGGCAGUCGCU.UUAGGCGAUCUGCCUAAACCACUCACUAAC.UAUGUAAGG.AAAGU.AUUAUCGACUUUCCUUACUC.....CGUG
[2]Spy-8-4   ..........................................ACGGCAGUCGCU.UUAGGCGAUCUGCCUAAACCACUCACUAAC.UACGUAAGG.AAAGU.AUUAUCGACUUUCCUUACUC.....CGUG
   Smu-1-1   GUAAGGAUU..GGA.GC.U.AAA.GC.UCCUAAUCCUUAC..ACGGCAAUCGCU.AAUGGCGAAUUGCCUAGUUCACUUACUAAC.UUCGAAAAU.GAAGU.AUCAUCGACUUCAUUUUCUA.....CGUG
[1]Smu-1-2   UAAAUGACU..GCA.UC.U.AAA.GU.UGCUAGUCAUUUG..AUGGUAGCCGCU.AUUGGCGGUCUGCCUCAGCGCCUUACUAGU.UU...................AA...........CA.....CUUG
   Spn-2-1   AUUCUGACG..GUA.GC.U.AAA.GC.AACUCGUCAGAAA..ACGGCAAUCGCU.AU.GGCGU.UUGCCUAGCUUCCUUACUAAC.UC.GUCGUC.GAAAU.AAAAUCGAUUUCGACU.CCU.....CGUG
   Spn-1-1   AUUCUGACG..GUA.GC.U.AAA.GC.AACUCGUCAGAAA..ACGGCAAUCGCU.AU.GACGU.UUGCCUAGCUUCCUUACUAAC.UC.GUCGUC.GAAAU.AAAAUCGAUUUCGACU.CCU.....CGUG
   Spn-2-3   CCAACAAUU..GGA.AC.U.AAA.GU.UCCAAAUUGUUGG..ACGCUAGUCGCUAUUUGGCGAACUAGCUAACUGCCUCACUAAC.UACGUUUGG.CAAAU.GAAAUCGAUUUGUAAAACUC.....CGUG
   Spn-2-5   UUA....UU..GGA.AC.U.AAA.GU.UCCUAAUUGUUAG..ACGCUAGCCGCUUAUAUGCGAACUAGCUAACUGCUUCACUAGU.CAUGACCCC.UAAAU.AUAAUCGAUUUAGAGUCCAUGACGUCGUA
   Spn-1-3   UUA....UU..GGA.AC.U.AAA.GU.UCCUAAUUGUUAG..ACGCUAGCCGCUUAUAUGCGAACUAGCUAACUGCUUCACUAGU.CAUGACCCC.UAAAU.AUAAUCGAUUUAGAGGCCAUGACGUCGUA
   Spn-2-7   GAAAGGUUU..AAU.AC.U.AAA.GU.AUCUAAAGCUUUC..ACGCUAAUCGCUAUCGGGCGA.UUAGCUAAAUGCUUUACUAAC.UC.UCUCGU.CAAAU.AACAUCGAUUUGAC....UC..ACUCGUG
   Lla-2-2   GAAAUAAUU..GAC.GC.U.AAA.GC.GCCUAAUUAUUUC..ACGGAAACCGCU.AU.GGCGGAUUUCCUAACCCUCUUACUAGU.UA.UUAUGA.AAAGC.AGAAUCGGCUUUUCAUAAUA.ACGUCGUA
   Lla-1-2   GGAAUAAUU..GAC.GC.U.AAA.GC.GCCUAAUUAUUUC..ACGGAAACCGCU.AU.GGCGGAUUUCCUAGUCCUCUUAUUAGU.UA.UUAUGA.AAAGC.AGAAUCGGCUUUUCAUAAUA.ACGUCGUA
   Sth-3-1   CAAAAGAUG..ACG.AUUU.UAA.AU.CGUUCAUCUUUUG..ACGCUAGCCCCU.AUUAGGUGUCUAGCUAGCCGUCUCACUAAC.UGCGUCAGA.CGAAU.AAUAUCGAUUCGUCUGACUU.....CGUG
   Sth-3-2   ..AUGUAUA..GUUUAU.G.UAA.AU.CGUUCAUCUUUUG..ACACUGGCGGCU.AUCUUGUGGCUAGCUAGCCGUCUCACUAAC.UACGUCAGA.CGAAU.AAUAUCGAUUCGUCUGACUU.....CGUG
   Lla-1-3   AUAAACAAU..GAU.GC.U.AAA.GC.ACCUAUUGUUUAU..GCGAAAGCCGCU.AG.AGCGAACUUUC.UAUUUUCUUACUUGU.UUCAUUUU..UCCGA.AAAAUCGUCGGUCAAAGUGC...GACGUA
   Lla-2-3   AGAAGCAUU..GGU.GC.U.AAA.GC.AUCUAUUGCUUCU..ACGGAAGCCGCU.AA.GACGGACUUUC.UAUUUUCUUACUUGU.UU.UCAUUC.AUUAA.AGUAUCAUCAAUGAAUGAAA.AUGGGUA
   Smu-1-3   AGAAUGUCG..GAU.AC.U.UAA.GC.AUCUCGCCAUUCU..ACGGCAUCCGCUAUUAGGCGGUAUGCCUAACCCACUUACUUGC.GCAAGGAAG.AGAAU.AGUAUCGAUUCUCUCCCUUGUGCUUCGUA
   Lla-1-4   GUAAUGAAU..GGU.GC.U.GAA.UC.CCCUAUUCAUUAC..ACGCGAAUUCCU.AA.CGGACAUUCGCUAAUCGUCUUAUUAGA.AA.ACUUAU.GAAAA.AAUAUCGUUUUCAUAAGUUU....CCGUC
   Lla-2-4   GUAAUAAAU..GGU.GC.U.AAA.GC.AUCUAUUUAUUAC..AGGCGUGUCUCC.UA..GAGGAUACUCGCUAAUUGUCUUAUUAGUAAAGGAAC.GAAAA.AAUAUCAUUUUCGUUUCUUU.....CUGU
   Ssu-1-1   AACACACUUA.AAU.CC.U.AAA.GG.ACUU.GAUGUGAU..ACGGUAGCCGCU.AU.GGCGGACUACCUACCAACCUUACUAAU.UCCGCACAG.GCAGU.GGUAUCGACUGCCUGUACUC.....CAUA
   Smu-1-4   ..GCAUAUCUAAGU.AC.U.AAA.GU.ACUUGAUAUGCAA..ACGGCAGCCACU.AU.GAUGAUCUGCCUGUCAACCUUACUAAU.UAUGUCAAG.UCAGU.AAUAUCGACUGACUUGACUGCAUGUCGUA
   Ssu-1-2   .............U.UG.C.....CA.A..............ACGCUCAUCGCU.AU.GGCGACUGAGCUAGACGUUGUACUAGU.UCGCCUUUC.CUAGC.GGUAUCGGCUAGGAAGGCUCA....CGUC
   Ssu-1-3   AGUUUGACU..GCA.AG.U.UAA.CU.UGCUAGUCAAACU..ACGCCAAAUCCU.AU.GGGACCUUGGCUAGCUACCUUACUAAC.UUCACCCAA.AUGGU.AGUAUCGACCAUUUGGGCUC.....CGUG
   Ssu-1-4   UGCCAGAUG..GUA.AG.U.AAA.CU.UACUCAUCUGGCA..ACGACAAAUCCU.AUAAGA..UUUGUCUAGGGCACGCACUAGC.UC.GAGUUU.UGAAA.AAUAUCGUUUCAAAACUCUC....ACGUC
   Ssu-1-5   AGAUCAAUG..GUU.GC....GA.GC.AACUCAUUUCCCU..ACGCCUGCCUCU.AU.GGAGGUCAGGCUAUCCCUUGCACUAGC.UC.GAAAAU.GGAUU.AAUGUCGAAUCCAUUUUCUC....ACGUC
   Smu-1-5   AGUUUAUCU..GUG.AC.U.AAA.GU.CACUAGCUAAACU..AUGCAAAGUCCU.AU.AAGAC.UUUGCUAAACACCUUACUAGU.AAAAGCUCU.CACUU.AAUAUCGAACUGAGAGCUUUUACGUCGGA
   Spn-2-9   CUACUGACU..AAAGGC.U.UAA.GC.AUUUAGUCAGUAG..ACGCUUUGUCCU.AU.AGGAUCAAAGCUAGAGCCCUGACUA.G.UAUUUUUAG.AUAAA.AAGAUGGUUUAUCUAAAAAU....ACGUC
   Smu-1-6   AGAAUGACUA.AUU.AC.U.AAA.GU.GAUU.GUCUUUCUCUACAGCAAUUGUUAUUUGGCGAAUGUCCCAAACGUCUUACUAGC.CU.GAUUUU.CGAAA.AGUAUCAUUUCGAAAAUCUC....ACGUC
   Ssu-1-6   GCUGUUUCU..GAU.AC.UGAAA.GU.AUCUAGAAACAGG..ACGGUAGCCGCU.AU.GGCGGUCUACCUACUCCCCUCACUAGG.UUCAAGGUU.GAAAAAAAUGUCGUUUUCAACCUUAG...AACGUC
   Ssu-1-7   .............U.UCAGUACGAGA.A..............ACGGUAGCCGCU.AU.GGCGGUCUACCUACUCCCCUCACUAGG.UUCAAGGUU.GAAAAAAAUGUUGUUUUCAACCUUGAAC...CGUC
   Spn-2-11  AAGUUACUU..GAUUGC...AAA.GCAAUCUAAGUAACUA..ACGCCAAAUCCU.AUUCGGACUUUGGCUAGGCGCUUCACUAGU.UUUAGGACA.UAAAU.AUGAUCGAUUUAUGUCCUAAAACGUCGUA
   Ssu-1-8   GCUCUGUCA..CUU.AC...AAU.GU.AAG..GUCAGAGC..ACGAAAACGCCU.UU.GAGCGAUUUUCUAAACUCUGCACUAGG.UUUCCACAUGGAAA..AGUAUCG.UUUCAUGUGGAAAC...CGUC
[1]Spy-4-3   CUAUUGUCA..GAC.AU.U.CUU..U.GUCUGGUUAAUGG..ACGCUAGCCACU.A..GGUGGACUAGCUAGUUGCCUUACUAGUAAAACCGUAA.CUGAU.UAGCUAAGUUGGUUUGGCC.....ACUU
   Spy-2-3   CUAUUGUCA..GAC.AU.U.CUU..U.GUCUGGUUAAUGG..ACGCUAGCCACU.A..GGUGGACUAGCUAGUUGCCUUACUAGUAAAACCGUAA.CUGAU.UAGCUAAGUUGGUCUUGGCC.....ACUU
   Ssu-1-9   AUCUCG............C.ACU........AUUUAAGGU..GUGGGAAUCGCUGCUCUGCGACUUUCCUAGCUACCUUACUAGG.CU.UUCAAG.CGAAA.AAUAUCGUUUCGCUUGAAAUAC...CGUC
             <<<<<<<<<..<<<.<<.......>>.>>>.>>>>>>>>>....<<<<<<<<<......>>>>.>>>>>....................<<<<<<.<<<<<........>>>>>>>>>>>..........
             ?????????..2??.??.......??.??2.?????????....?2222????......????.2222?....................?????2.22?22........22?222?????..........
             ○○○○○●○○○--R○●-RC-U-●AA-RY-●○YU○●Y●●●●●●--ACG●●ARYY●CU-●U-RG●R●●YU●●CUAR●Y●YCUYACUARY-U●○●●●●●●-●●AR○-AR●AUCG○YU●●●●●●○●●●-----CGU●
```

2.17 The lacto-2 *(Lactobacillus)* motif

Notes:

[1] This putative homologs is in a region of annotated tRNAs and rRNAs, either between annotated RNAs or overlapping an extreme end. Therefore, the homolog may represent an unrelated RNA structure; however it does match the consensus well.

abbrev. of hits **taxonomy of species**

Lla-1-1 to Lla-1-1	Firmicutes Lactobacillales Streptococcaceae *Lactococcus lactis subsp. lactis* Il1403
Sag-1-1 to Sag-1-5	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 18RS21
Sag-2-1 to Sag-2-3	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 2603V/R
Sag-3-1 to Sag-3-2	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 515
Sag-4-1 to Sag-4-2	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* A909
Sag-5-1 to Sag-5-2	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* CJB111
Sag-6-1 to Sag-6-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* COH1
Sag-7-1 to Sag-7-2	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* H36B
Sag-8-1 to Sag-8-2	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* NEM316
Smu-1-1 to Smu-1-33	Firmicutes Lactobacillales Streptococcaceae *Streptococcus mutans* UA159
Spn-1-1 to Spn-1-20	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* R6
Spn-2-1 to Spn-2-43	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* TIGR4
Spy-1-1 to Spy-1-17	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* M1 GAS
Spy-2-1 to Spy-2-10	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* M49 591
Spy-3-1 to Spy-3-16	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10270
Spy-4-1 to Spy-4-18	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10394
Spy-5-1 to Spy-5-19	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10750
Spy-6-1 to Spy-6-18	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS2096
Spy-7-1 to Spy-7-18	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS315
Spy-8-1 to Spy-8-17	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS5005
Spy-9-1 to Spy-9-17	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS6180
Spy-10-1 to Spy-10-18	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS8232
Spy-11-1 to Spy-11-18	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS9429
Spy-12-1 to Spy-12-17	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* SSI-1
Ssu-1-1 to Ssu-1-15	Firmicutes Lactobacillales Streptococcaceae *Streptococcus suis* 89/1591
Sth-1-1 to Sth-1-7	Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* CNRZ1066
Sth-2-1 to Sth-2-7	Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* LMD-9
Sth-3-1 to Sth-3-9	Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* LMG 18311

abbrev	**RefSeq accession**		**5′ at**	**3′ at**	**genes**
Sag-4-1	NC_007432.1	+	687878	687950	hypo→ **RNA**→ COG5280 (COG5280)COG5412 (COG5412)→
Sag-5-1	NZ_AAJQ01000012.1	+	7502	7574	hypo→ **RNA**→ COG5280 (COG5280)COG5412 (COG5412)→
Spn-2-1	NC_003028.1	+	557821	557894	←hypo **RNA**→ Pnp (COG1185)→
Spn-1-1	NC_003098.1	+	515669	515742	MTHFR (cd00537)→ **RNA**→ Pnp (COG1185)→
Spn-2-2	NZ_AAGY02000055.1	-	7474	7401	MTHFR (cd00537)→ **RNA**→ Pnp (COG1185)→
Spn-2-3	NC_003028.1	-	497087	497014	←HSP70 (pfam00012) **RNA**→ ←GrpE (cd00446)
Spn-1-2	NC_003098.1	-	461670	461597	←HSP70 (pfam00012) **RNA**→ ←GrpE (cd00446)
Spn-2-4	NZ_AAGY02000057.1	-	7463	7390	←HSP70 (pfam00012) **RNA**→ ←GrpE (cd00446)
[1]Spy-8-1	NC_007297.1	+	18760	18834	tRNA-Ala→ **RNA**→ large subunit 23S ribosomal RNA→
Spy-8-2	NC_007297.1	+	24761	24835	tRNA-Ala→ **RNA**→ large subunit 23S ribosomal RNA→
Spy-8-3	NC_007297.1	+	81567	81641	tRNA-Ala→ **RNA**→ large subunit 23S ribosomal RNA→
Spy-8-4	NC_007297.1	+	266686	266760	tRNA-Ala→ **RNA**→ large subunit 23S ribosomal RNA→
Spy-8-5	NC_007297.1	-	1578487	1578413	tRNA-Ala→ **RNA**→ large subunit 23S ribosomal RNA→
Spy-8-6	NC_007297.1	-	1289424	1289350	tRNA-Ala→ **RNA**→ large subunit 23S ribosomal RNA→
Spy-9-1	NC_007296.1	+	18763	18837	tRNA-Ala→ **RNA**→ 16S small subunit ribosomal RNA→
Spy-9-2	NC_007296.1	+	24764	24838	tRNA-Ala→ **RNA**→ 16S small subunit ribosomal RNA→
Spy-9-3	NC_007296.1	+	81053	81127	tRNA-Ala→ **RNA**→ 16S small subunit ribosomal RNA→

Spy-9-4	NC_007296.1	+	270235	270309	tRNA-Ala→ **RNA**→ 16S small subunit ribosomal RNA→
Spy-9-5	NC_007296.1	-	1608329	1608255	tRNA-Ala→ **RNA**→ 23S large subunit ribosomal RNA→
Spy-3-1	NC_008022.1	+	18760	18834	16S small subunit ribosomal RNA→ **RNA**→ 16S small subunit ribosomal RNA→
Spy-3-2	NC_008022.1	+	24761	24835	16S small subunit ribosomal RNA→ **RNA**→ 16S small subunit ribosomal RNA→
Spy-3-3	NC_008022.1	+	81747	81821	16S small subunit ribosomal RNA→ **RNA**→ 16S small subunit ribosomal RNA→
Spy-3-4	NC_008022.1	+	270926	271000	16S small subunit ribosomal RNA→ **RNA**→ 16S small subunit ribosomal RNA→
Spy-3-5	NC_008022.1	-	1383326	1383252	tRNA-Ala→ **RNA**→ 23S large subunit ribosomal RNA→
Spy-6-1	NC_008023.1	+	18765	18839	16S small subunit ribosomal RNA→ **RNA**→ 16S small subunit ribosomal RNA→
Spy-6-2	NC_008023.1	+	24766	24840	16S small subunit ribosomal RNA→ **RNA**→ 16S small subunit ribosomal RNA→
Spy-6-3	NC_008023.1	+	274273	274347	16S small subunit ribosomal RNA→ **RNA**→ 16S small subunit ribosomal RNA→
Spy-6-4	NC_008023.1	-	1595517	1595443	tRNA-Ala→ **RNA**→ 23S large subunit ribosomal RNA→
Spy-6-5	NC_008023.1	-	1307842	1307768	tRNA-Ala→ **RNA**→ 23S large subunit ribosomal RNA→
Spy-5-1	NC_008024.1	+	18761	18835	16S small subunit ribosomal RNA→ **RNA**→ 16S small subunit ribosomal RNA→
Spy-5-2	NC_008024.1	+	24762	24836	16S small subunit ribosomal RNA→ **RNA**→ 16S small subunit ribosomal RNA→
Spy-5-3	NC_008024.1	+	267021	267095	16S small subunit ribosomal RNA→ **RNA**→ 16S small subunit ribosomal RNA→
Spy-5-4	NC_008024.1	-	1613064	1612990	tRNA-Ala→ **RNA**→ 23S large subunit ribosomal RNA→
Spy-5-5	NC_008024.1	-	1366498	1366424	tRNA-Ala→ **RNA**→ 23S large subunit ribosomal RNA→
Spy-1-1	NC_002737.1	+	18791	18865	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
Spy-1-2	NC_002737.1	+	24792	24866	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
Spy-1-3	NC_002737.1	+	81010	81084	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
Spy-1-4	NC_002737.1	+	266076	266150	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
Spy-1-5	NC_002737.1	-	1581401	1581327	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
Spy-1-6	NC_002737.1	-	1334465	1334391	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
[1]Spy-9-6	NC_007296.1	-	1359051	1358977	tRNA-Ala→ **RNA**→ 23S large subunit ribosomal RNA→
Spy-11-1	NC_008021.1	+	18763	18837	tRNA-Ala→ **RNA**→ 16S small subunit ribosomal RNA→
Spy-11-2	NC_008021.1	+	24764	24838	tRNA-Ala→ **RNA**→ 16S small subunit ribosomal RNA→
Spy-11-3	NC_008021.1	+	81375	81449	tRNA-Ala→ **RNA**→ 16S small subunit ribosomal RNA→
Spy-11-4	NC_008021.1	+	273350	273424	tRNA-Ala→ **RNA**→ 16S small subunit ribosomal RNA→
Spy-11-5	NC_008021.1	-	1571632	1571558	tRNA-Ala→ **RNA**→ 23S large subunit ribosomal RNA→
Spy-11-6	NC_008021.1	-	1284000	1283926	tRNA-Ala→ **RNA**→ 23S large subunit ribosomal RNA→
Spy-6-6	NC_008023.1	+	81351	81425	16S small subunit ribosomal RNA→ **RNA**→ 16S small subunit ribosomal RNA→
Spy-2-1	NZ_AAFV01000024.1	-	7102	7028	**RNA**→ COG1782: Predicted metal-dependent RNase, consists of a metallo-beta-lactamase domain and an RNA-binding KH domai
[1]Spy-4-1	NC_006086.1	+	19849	19923	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
Spy-4-2	NC_006086.1	+	25850	25924	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
Spy-4-3	NC_006086.1	+	122425	122499	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
Spy-4-4	NC_006086.1	+	298786	298860	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
Spy-4-5	NC_006086.1	-	1629285	1629211	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
Spy-4-6	NC_006086.1	-	1361522	1361448	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
Spy-3-6	NC_008022.1	-	1633817	1633743	tRNA-Ala→ **RNA**→ 23S large subunit ribosomal RNA→
Spy-12-1	NC_004606.1	+	18790	18864	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
Spy-12-2	NC_004606.1	+	75769	75843	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
Spy-12-3	NC_004606.1	+	251824	251898	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
Spy-12-4	NC_004606.1	+	538960	539034	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→

Spy-12-5	NC_004606.1	-	1638833	1638759	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
Spy-7-1	NC_004070.1	+	18790	18864	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
Spy-7-2	NC_004070.1	+	24792	24866	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
Spy-7-3	NC_004070.1	+	81770	81844	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
Spy-7-4	NC_004070.1	+	257912	257986	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
Spy-7-5	NC_004070.1	-	1645047	1644973	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
Spy-7-6	NC_004070.1	-	1357926	1357852	tRNA-Ala→ **RNA**→ 23S ribosomal RNA→
[1]Spy-5-6	NC_008024.1	+	83938	84012	16S small subunit ribosomal RNA→ **RNA**→ 16S small subunit ribosomal RNA→
[1]Ssu-1-1	NZ_AAFA02000006.1	-	24636	24563	tRNA-Ala→ **RNA**→ tRNA-Asn→
Spy-10-1	NC_003485.1	+	18758	18834	16S ribosomal RNA→ **RNA**→ 23S ribosomal RNA→
Spy-10-2	NC_003485.1	+	24759	24835	16S ribosomal RNA→ **RNA**→ 23S ribosomal RNA→
Spy-10-3	NC_003485.1	+	81661	81737	16S ribosomal RNA→ **RNA**→ 23S ribosomal RNA→
Spy-10-4	NC_003485.1	+	259718	259794	16S ribosomal RNA→ **RNA**→ 23S ribosomal RNA→
Spy-10-5	NC_003485.1	-	1637097	1637021	16S ribosomal RNA→ **RNA**→ 23S ribosomal RNA→
Spy-10-6	NC_003485.1	-	1347473	1347397	16S ribosomal RNA→ **RNA**→ 23S ribosomal RNA→
Smu-1-1	NC_004350.1	-	1551272	1551200	PyrH (COG0528)→ **RNA**→ RRF (cd00520)→
Smu-1-2	NC_004350.1	+	431965	432037	TrxB (COG0492)→ **RNA**→ NAPRTase_A (cd01570)→
Smu-1-3	NC_004350.1	-	1229449	1229376	COG1481 (COG1481)→ **RNA**→ Peptidase_U34 (pfam03577)→
Ssu-1-2	NZ_AAFA02000140.1	-	1943	1871	←EF_TS (pfam00889) **RNA**→ ←Ribosomal_S2 (pfam00318)
[1]Smu-1-4	NC_004350.1	+	1655824	1655898	←5S ribosomal RNA **RNA**→ ←23S ribosomal RNA
Smu-1-5	NC_004350.1	+	1684126	1684200	←5S ribosomal RNA **RNA**→ ←23S ribosomal RNA
Smu-1-6	NC_004350.1	+	1873403	1873477	←5S ribosomal RNA **RNA**→ ←23S ribosomal RNA
Smu-1-7	NC_004350.1	-	190683	190609	←5S ribosomal RNA **RNA**→ ←23S ribosomal RNA
Smu-1-8	NC_004350.1	-	21811	21737	←5S ribosomal RNA **RNA**→ ←23S ribosomal RNA
Spy-4-7	NC_006086.1	+	719170	719245	Pnp (COG0005)→ **RNA**→ DeoD (COG0813)→
Spy-8-7	NC_007297.1	+	701823	701898	Pnp (COG0005)→ **RNA**→ DeoD (COG0813)→
Spy-11-7	NC_008021.1	+	740414	740489	Pnp (COG0005)→ **RNA**→ DeoD (COG0813)→
Spy-6-7	NC_008023.1	+	741823	741898	Pnp (COG0005)→ **RNA**→ DeoD (COG0813)→
Spy-5-7	NC_008024.1	+	764771	764846	Pnp (COG0005)→ **RNA**→ DeoD (COG0813)→
Spy-1-7	NC_002737.1	+	740803	740878	Pnp (COG0005)→ **RNA**→ DeoD (COG0813)→
Spy-10-7	NC_003485.1	+	781774	781849	Pnp (COG0005)→ **RNA**→ DeoD (COG0813)→
Spy-12-6	NC_004606.1	-	1224640	1224565	Pnp (COG0005)→ **RNA**→ DeoD (COG0813)→
Spy-7-7	NC_004070.1	+	672094	672169	Pnp (COG0005)→ **RNA**→ DeoD (COG0813)→
Spy-2-2	NZ_AAFV01000128.1	+	2543	2616	Pnp (COG0005)→ **RNA**→ DeoD (COG0813)→
Ssu-1-3	NZ_AAFA02000067.1	+	1638	1712	Era (COG1159)→ **RNA**→ Nei (COG0266)→
Spy-2-3	NZ_AAFV01000004.1	-	6737	6664	hypo→ **RNA**→ Adenylsuccinate_lyase_1 (cd01360)→
Spy-4-8	NC_006086.1	+	92612	92685	Hypothetical cytosolic protein→ **RNA**→ Adenylsuccinate_lyase_1 (cd01360)→
Spy-11-8	NC_008021.1	+	51552	51625	←hypo **RNA**→ Adenylsuccinate_lyase_1 (cd01360)→
Spy-6-8	NC_008023.1	+	51527	51600	←hypo **RNA**→ Adenylsuccinate_lyase_1 (cd01360)→
Spy-12-7	NC_004606.1	+	45954	46027	hypo→ **RNA**→ Adenylsuccinate_lyase_1 (cd01360)→
Spy-7-8	NC_004070.1	+	51955	52028	hypo→ **RNA**→ Adenylsuccinate_lyase_1 (cd01360)→
Spy-5-8	NC_008024.1	+	53030	53103	hypo→ **RNA**→ Adenylsuccinate_lyase_1 (cd01360)→
Ssu-1-4	NZ_AAFA02000024.1	-	3629	3556	COG4894 (COG4894)→ **RNA**→ tRNA-synt_2c (pfam01411)DHHA1 (pfam02272)→

Spy-10-8	NC_003485.1	+	51836	51911	hypo→ **RNA**→ Adenylsuccinate_lyase_1 (cd01360)→
Spn-2-5	NC_003028.1	-	200720	200648	←RplF (COG0097) **RNA**→ ←Ribosomal_S8 (pfam00410)
Spn-1-3	NC_003098.1	-	203230	203158	←RplF (COG0097) **RNA**→ ←Ribosomal_S8 (pfam00410)
Spn-2-6	NZ_AAGY02000002.1	+	22267	22339	←RplF (COG0097) **RNA**→ ←Ribosomal_S8 (pfam00410)
Spn-1-4	NC_003098.1	+	736253	736326	Conserved hypothetical protein, truncation→ **RNA**→ DeoD (COG0813)→
Spn-2-7	NC_003028.1	+	783399	783472	hemolysin-related protein→ **RNA**→ DeoD (COG0813)→
Spn-2-8	NZ_AAGY02000193.1	+	151	224	**RNA**→ DeoD (COG0813)→
Smu-1-9	NC_004350.1	-	1887437	1887362	Ribosomal_S8 (pfam00410)→ **RNA**→ RplF (COG0097)→
Spy-10-9	NC_003485.1	-	230617	230543	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Spy-8-8	NC_007297.1	-	237610	237536	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Spy-9-7	NC_007296.1	-	241133	241059	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Spy-3-7	NC_008022.1	-	241825	241751	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Spy-5-9	NC_008024.1	-	237917	237843	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Spy-1-8	NC_002737.1	-	236999	236925	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Spy-4-9	NC_006086.1	-	269684	269610	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Spy-11-9	NC_008021.1	-	244246	244172	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Spy-6-9	NC_008023.1	-	245168	245094	←Protein Translation Elongation Factor G (EF-G) **RNA**→ ←RpsG (COG0049)
Spy-12-8	NC_004606.1	-	222719	222645	←FusA (COG0480) **RNA**→ hypo→
Spy-7-9	NC_004070.1	-	228809	228735	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Spn-2-9	NC_003028.1	-	200039	199965	←Ribosomal_S8 (pfam00410) **RNA**→ hypo→
Spn-1-5	NC_003098.1	-	202549	202475	←Ribosomal_S8 (pfam00410) **RNA**→ ←Ribosomal_S14 (pfam00253)
Spn-2-10	NZ_AAGY02000002.1	+	22948	23022	←Ribosomal_S8 (pfam00410) **RNA**→ ←Ribosomal_S14 (pfam00253)
Spn-2-11	NC_003028.1	+	1528165	1528239	←Cad (pfam03596) **RNA**→ ←Ribosomal_S15p_S13e (cd00353)
Spn-2-12	NZ_AAGY02000066.1	+	1354	1428	←Cad (pfam03596) **RNA**→ ←Ribosomal_S15p_S13e (cd00353)
Spn-2-13	NC_003028.1	-	1701359	1701286	NUDIX (pfam00293)→ **RNA**→ PrmA (pfam06325)→
Spn-1-6	NC_003098.1	-	1585026	1584953	NUDIX (pfam00293)→ **RNA**→ PrmA (pfam06325)→
Spn-2-14	NZ_AAGY02000087.1	-	5269	5196	NUDIX (pfam00293)→ **RNA**→ PrmA (pfam06325)→
Spy-8-9	NC_007297.1	+	64664	64738	RpsJ (COG0051)→ **RNA**→ Ribosomal_L3 (pfam00297)→
Spy-1-9	NC_002737.1	+	64075	64149	RpsJ (COG0051)→ **RNA**→ Ribosomal_L3 (pfam00297)→
Spy-4-10	NC_006086.1	+	105495	105569	RpsJ (COG0051)→ **RNA**→ Ribosomal_L3 (pfam00297)→
Spy-9-8	NC_007296.1	+	64118	64192	RpsJ (COG0051)→ **RNA**→ Ribosomal_L3 (pfam00297)→
Spy-11-10	NC_008021.1	+	64440	64514	RpsJ (COG0051)→ **RNA**→ Ribosomal_L3 (pfam00297)→
Spy-3-8	NC_008022.1	+	64812	64886	RpsJ (COG0051)→ **RNA**→ Ribosomal_L3 (pfam00297)→
Spy-6-10	NC_008023.1	+	64415	64489	RpsJ (COG0051)→ **RNA**→ Ribosomal_L3 (pfam00297)→
Spy-5-10	NC_008024.1	+	65915	65989	RpsJ (COG0051)→ **RNA**→ LSU ribosomal protein L3P→
Spy-10-10	NC_003485.1	+	64724	64798	RpsJ (COG0051)→ **RNA**→ Ribosomal_L3 (pfam00297)→
Spy-12-9	NC_004606.1	+	58832	58906	RpsJ (COG0051)→ **RNA**→ hypo→
Spy-7-10	NC_004070.1	+	64833	64907	RpsJ (COG0051)→ **RNA**→ Ribosomal_L3 (pfam00297)→
Spy-10-11	NC_003485.1	-	230502	230428	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Spy-8-10	NC_007297.1	-	237496	237422	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Spy-9-9	NC_007296.1	-	241019	240945	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Spy-1-10	NC_002737.1	-	236885	236811	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Spy-4-11	NC_006086.1	-	269569	269495	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)

Spy-12-10	NC_004606.1	-	222604	222530	hypo→ **RNA**→ ←RpsG (COG0049)
Spy-7-11	NC_004070.1	-	228694	228620	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Spy-5-11	NC_008024.1	-	237803	237729	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Spy-11-11	NC_008021.1	-	244132	244058	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Spy-6-11	NC_008023.1	-	245054	244980	←Protein Translation Elongation Factor G (EF-G) **RNA**→ ←RpsG (COG0049)
Spy-5-12	NC_008024.1	-	237686	237612	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Spy-12-11	NC_004606.1	-	222487	222413	hypo→ **RNA**→ ←RpsG (COG0049)
Spy-7-12	NC_004070.1	-	228577	228503	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Spn-1-7	NC_003098.1	-	1977823	1977749	←hypo **RNA**→ Transposase, uncharacterized, truncation→
Spn-2-15	NC_003028.1	+	228934	229008	Acetyltransf_1 (pfam00583)→ **RNA**→ ←hypo
Spn-2-16	NC_003028.1	-	2100224	2100150	←hypo **RNA**→ Transposase_11 (pfam01609)→
Spn-2-17	NZ_AAGY02000102.1	-	1423	1349	Acetyltransf_1 (pfam00583)→ **RNA**→ hypo→
Spy-3-9	NC_008022.1	-	71976	71902	←RplF (COG0097) **RNA**→ ←Ribosomal_S8 (pfam00410)
Spy-10-12	NC_003485.1	-	71888	71814	←RplF (COG0097) **RNA**→ ←Ribosomal_S8 (pfam00410)
Spy-2-4	NZ_AAFV01000130.1	-	529	455	←RplF (COG0097) **RNA**→ ←Ribosomal_S8 (pfam00410)
Spy-4-12	NC_006086.1	-	112658	112584	←RplF (COG0097) **RNA**→ ←Ribosomal_S8 (pfam00410)
Spy-8-11	NC_007297.1	-	71796	71722	←RplF (COG0097) **RNA**→ ←Ribosomal_S8 (pfam00410)
Spy-9-10	NC_007296.1	-	71283	71209	←RplF (COG0097) **RNA**→ ←Ribosomal_S8 (pfam00410)
Spy-11-12	NC_008021.1	-	71605	71531	←RplF (COG0097) **RNA**→ ←Ribosomal_S8 (pfam00410)
Spy-6-12	NC_008023.1	-	71581	71507	←RplF (COG0097) **RNA**→ ←Ribosomal_S8 (pfam00410)
Spy-5-13	NC_008024.1	-	73077	73003	←RplF (COG0097) **RNA**→ ←Ribosomal_S8 (pfam00410)
Spy-1-11	NC_002737.1	-	71239	71165	←RplF (COG0097) **RNA**→ ←Ribosomal_S8 (pfam00410)
Spy-12-12	NC_004606.1	-	65996	65922	←RplF (COG0097) **RNA**→ ←Ribosomal_S8 (pfam00410)
Spy-7-13	NC_004070.1	-	71997	71923	←RplF (COG0097) **RNA**→ ←Ribosomal_S8 (pfam00410)
Smu-1-10	NC_004350.1	+	302316	302391	DapD (COG2171)→ **RNA**→ AbgB (COG1473)→
Smu-1-11	NC_004350.1	+	523903	523978	IleS (COG0060)→ **RNA**→ ←hypo
Smu-1-12	NC_004350.1	+	636158	636233	Icd (COG0538)→ **RNA**→ DUF1113 (pfam06541)→
Smu-1-13	NC_004350.1	-	1229334	1229259	COG1481 (COG1481)→ **RNA**→ Peptidase_U34 (pfam03577)→
Sth-3-1	NC_006448.1	-	1092657	1092583	RfbC (COG1898)→ **RNA**→ RfbB (COG1088)→
Sth-1-1	NC_006449.1	-	1090530	1090456	RfbC (COG1898)→ **RNA**→ RfbB (COG1088)→
Sth-2-1	NZ_AAGS01000001.1	-	37997	37923	RfbC (COG1898)→ **RNA**→ RfbB (COG1088)→
Ssu-1-5	NZ_AAFA02000083.1	+	6117	6191	←Ribosomal_S8 (pfam00410) **RNA**→ ←RpsN (COG0199)
Spn-2-18	NC_003028.1	-	1466566	1466492	hypo→ **RNA**→ DapB (COG0289)→
Spn-1-8	NC_003098.1	-	1402012	1401938	DegV (COG1307)→ **RNA**→ DapB (COG0289)→
Spn-2-19	NZ_AAGY02000009.1	-	12408	12334	DegV (COG1307)→ **RNA**→ DapB (COG0289)→
Sag-2-1	NC_004116.1	-	1296376	1296302	←COG5340 (COG5340) **RNA**→ TolA (pfam06519)→
Sag-1-1	NZ_AAJO01000011.1	-	12628	12554	←COG5340 (COG5340) **RNA**→ TolA (pfam06519)→
Sag-7-1	NZ_AAJS01000027.1	-	9916	9842	←COG5340 (COG5340) **RNA**→ TolA (pfam06519)→
Ssu-1-6	NZ_AAFA02000008.1	-	15037	14964	hypo→ **RNA**→ hypo→
Ssu-1-7	NZ_AAFA02000007.1	+	6002	6075	←COG5340 (COG5340) **RNA**→ agglutinin receptor→
Ssu-1-8	NZ_AAFA02000007.1	+	14354	14427	TolA (COG3064)→ **RNA**→ TolA (COG3064)→
Spn-2-20	NC_003028.1	-	1211408	1211335	CarA (COG0505)→ **RNA**→ MGS_CPS_II (cd01424)CarB (COG0458)CPSase_L_D3 (pfam02787)→
Spn-1-9	NC_003098.1	-	1158023	1157950	CarA (COG0505)→ **RNA**→ MGS_CPS_II (cd01424)CarB (COG0458)CPSase_L_D3 (pfam02787)→

Spn-2-21	NZ_AAGY02000008.1	-	12311	12238	CarA (COG0505)→ **RNA**→ CarB (COG0458)CPSase_L_D3 (pfam02787)MGS_CPS_II (cd01424)→
Spn-2-22	NC_003028.1	+	499967	500040	hypo→ **RNA**→ DnaJ (COG0484)→
Spn-1-10	NC_003098.1	+	464542	464615	HSP70 (pfam00012)→ **RNA**→ DnaJ (COG0484)→
Spn-2-23	NZ_AAGY02000057.1	+	10343	10416	HSP70 (pfam00012)→ **RNA**→ DnaJ (COG0484)→
Ssu-1-9	NZ_AAFA02000040.1	+	4756	4830	HSP70 (pfam00012)→ **RNA**→ DnaJ (COG0484)→
Sth-3-2	NC_006448.1	+	49506	49580	hypo→ **RNA**→ Adenylsuccinate_lyase_1 (cd01360)→
Sth-1-2	NC_006449.1	+	49533	49607	hypo→ **RNA**→ Adenylsuccinate_lyase_1 (cd01360)→
Sth-2-2	NZ_AAGS01000009.1	-	16336	16262	hypo→ **RNA**→ Adenylsuccinate_lyase_1 (cd01360)→
Smu-1-14	NC_004350.1	+	59145	59219	hypo→ **RNA**→ Adenylsuccinate_lyase_1 (cd01360)→
Ssu-1-10	NZ_AAFA02000004.1	+	1214	1288	Tdk (COG1435)→ **RNA**→ PrfA (COG0216)→
Smu-1-15	NC_004350.1	-	1551161	1551087	PyrH (COG0528)→ **RNA**→ RRF (cd00520)→
Smu-1-16	NC_004350.1	-	1551050	1550976	PyrH (COG0528)→ **RNA**→ RRF (cd00520)→
Smu-1-17	NC_004350.1	+	431854	431927	TrxB (COG0492)→ **RNA**→ NAPRTase_A (cd01570)→
Smu-1-18	NC_004350.1	-	1162751	1162677	PyrE (COG0461)→ **RNA**→ COG1881 (COG1881)→
Smu-1-19	NC_004350.1	-	1837479	1837400	PTS_IIB_man (cd00001)→ **RNA**→ EII-Sor (pfam03609)→
Spy-8-12	NC_007297.1	-	1358358	1358286	GlyRS_alpha_core (cd00733)→ **RNA**→ GlyS (COG0751)→
Spy-9-11	NC_007296.1	-	1428014	1427942	GlyRS_alpha_core (cd00733)→ **RNA**→ GlyS (COG0751)→
Spy-11-13	NC_008021.1	-	1352906	1352834	GlyRS_alpha_core (cd00733)→ **RNA**→ GlyS (COG0751)→
Spy-3-10	NC_008022.1	-	1451931	1451859	GlyRS_alpha_core (cd00733)→ **RNA**→ GlyS (COG0751)→
Spy-6-13	NC_008023.1	-	1376800	1376728	GlyRS_alpha_core (cd00733)→ **RNA**→ GlyS (COG0751)→
Spy-1-12	NC_002737.1	-	1403396	1403324	GlyRS_alpha_core (cd00733)→ **RNA**→ GlyS (COG0751)→
Spy-2-5	NZ_AAFV01000113.1	-	2717	2645	GlyRS_alpha_core (cd00733)→ **RNA**→ GlyS (COG0751)→
Spy-4-13	NC_006086.1	-	1430473	1430401	GlyRS_alpha_core (cd00733)→ **RNA**→ GlyS (COG0751)→
Spy-5-14	NC_008024.1	-	1435465	1435393	GlyRS_alpha_core (cd00733)→ **RNA**→ GlyS (COG0751)→
Spy-10-13	NC_003485.1	-	1416427	1416355	GlyRS_alpha_core (cd00733)→ **RNA**→ GlyS (COG0751)→
Spy-12-13	NC_004606.1	+	429984	430056	GlyRS_alpha_core (cd00733)→ **RNA**→ GlyS (COG0751)→
Spy-7-14	NC_004070.1	-	1466886	1466814	GlyRS_alpha_core (cd00733)→ **RNA**→ GlyS (COG0751)→
Spy-9-12	NC_007296.1	+	700206	700277	Pnp (COG0005)→ **RNA**→ DeoD (COG0813)→
Spy-3-11	NC_008022.1	+	734299	734370	Pnp (COG0005)→ **RNA**→ DeoD (COG0813)→
Spy-2-6	NZ_AAFV01000128.1	+	2692	2763	Pnp (COG0005)→ **RNA**→ DeoD (COG0813)→
Smu-1-20	NC_004350.1	-	1196168	1196097	GATase1_IGP_Synthase (cd01748)→ **RNA**→ HisA (COG0106)→
Ssu-1-11	NZ_AAFA02000081.1	+	2952	3025	COG0750 (COG0750)→ **RNA**→ ProS (COG0442)→
Spn-2-24	NC_003028.1	+	193403	193478	RpsJ (COG0051)→ **RNA**→ Ribosomal_L3 (pfam00297)→
Spn-1-11	NC_003098.1	+	195913	195988	RpsJ (COG0051)→ **RNA**→ Ribosomal_L3 (pfam00297)→
Spn-2-25	NZ_AAGY02000002.1	-	29584	29509	RpsJ (COG0051)→ **RNA**→ Ribosomal_L3 (pfam00297)→
Smu-1-21	NC_004350.1	+	281417	281490	HTH_ARAC (smart00342)→ **RNA**→ hypo→
Spy-9-13	NC_007296.1	-	1481736	1481661	HSP70 (pfam00012)→ **RNA**→ DnaJ (COG0484)→
Spy-5-15	NC_008024.1	-	1489199	1489124	HSP70 (pfam00012)→ **RNA**→ DnaJ (COG0484)→
Spy-10-14	NC_003485.1	-	1511962	1511887	HSP70 (pfam00012)→ **RNA**→ DnaJ (COG0484)→
Spy-8-13	NC_007297.1	-	1452936	1452861	HSP70 (pfam00012)→ **RNA**→ DnaJ (COG0484)→
Spy-1-13	NC_002737.1	-	1457153	1457078	HSP70 (pfam00012)→ **RNA**→ DnaJ (COG0484)→
Spy-11-14	NC_008021.1	-	1447373	1447298	HSP70 (pfam00012)→ **RNA**→ DnaJ (COG0484)→
Spy-3-12	NC_008022.1	-	1505706	1505631	HSP70 (pfam00012)→ **RNA**→ DnaJ (COG0484)→

Spy-6-14	NC_008023.1	-	1471263	1471188	HSP70 (pfam00012)→ **RNA**→ DnaJ (COG0484)→
Spy-12-14	NC_004606.1	+	374862	374937	HSP70 (pfam00012)→ **RNA**→ DnaJ (COG0484)→
Spy-7-15	NC_004070.1	-	1522008	1521933	HSP70 (pfam00012)→ **RNA**→ DnaJ (COG0484)→
Spy-4-14	NC_006086.1	-	1484261	1484186	HSP70 (pfam00012)→ **RNA**→ DnaJ (COG0484)→
Spy-2-7	NZ_AAFV01000158.1	-	117	42	HSP70 (pfam00012)→ **RNA**→
Smu-1-22	NC_004350.1	-	1162638	1162563	PyrE (COG0461)→ **RNA**→ COG1881 (COG1881)→
Smu-1-23	NC_004350.1	-	1837594	1837519	PTS_IIB_man (cd00001)→ **RNA**→ EII-Sor (pfam03609)→
Ssu-1-12	NZ_AAFA02000016.1	-	13724	13649	MrcB (COG0744)→ **RNA**→ RpoB (COG0085)→
Spn-1-12	NC_003098.1	+	515429	515503	MTHFR (cd00537)→ **RNA**→ Pnp (COG1185)→
Spn-2-26	NC_003028.1	+	557581	557655	←hypo **RNA**→ ←hypo
Spn-2-27	NZ_AAGY02000055.1	-	7714	7640	MTHFR (cd00537)→ **RNA**→ Pnp (COG1185)→
Spy-4-15	NC_006086.1	+	84957	85029	Formyl_trans_N (pfam00551)→ **RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Spy-8-14	NC_007297.1	+	44009	44081	Formyl_trans_N (pfam00551)→ **RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Spy-9-14	NC_007296.1	+	43718	43790	Formyl_trans_N (pfam00551)→ **RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Spy-11-15	NC_008021.1	+	43745	43817	Formyl_trans_N (pfam00551)→ **RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Spy-3-13	NC_008022.1	+	44113	44185	Formyl_trans_N (pfam00551)→ **RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Spy-6-15	NC_008023.1	+	43720	43792	Formyl_trans_N (pfam00551)→ **RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Spy-5-16	NC_008024.1	+	44005	44077	Formyl_trans_N (pfam00551)→ **RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Spy-1-14	NC_002737.1	+	43421	43493	Formyl_trans_N (pfam00551)→ **RNA**→ ←COG3942 (COG3942)
Spy-10-15	NC_003485.1	+	44028	44100	Formyl_trans_N (pfam00551)→ **RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Spy-12-15	NC_004606.1	+	38144	38216	Formyl_trans_N (pfam00551)→ **RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Spy-7-16	NC_004070.1	+	44145	44217	Formyl_trans_N (pfam00551)→ **RNA**→ MGS (pfam02142)AICARFT_IMPCHas (pfam01808)→
Spy-2-8	NZ_AAFV01000004.1	-	14570	14498	Formyl_trans_N (pfam00551)→ **RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Spy-4-16	NC_006086.1	+	84956	85030	Formyl_trans_N (pfam00551)→ **RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Spy-8-15	NC_007297.1	+	44008	44082	Formyl_trans_N (pfam00551)→ **RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Spy-9-15	NC_007296.1	+	43717	43791	Formyl_trans_N (pfam00551)→ **RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Spy-11-16	NC_008021.1	+	43744	43818	Formyl_trans_N (pfam00551)→ **RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Spy-3-14	NC_008022.1	+	44112	44186	Formyl_trans_N (pfam00551)→ **RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Spy-6-16	NC_008023.1	+	43719	43793	Formyl_trans_N (pfam00551)→ **RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Spy-5-17	NC_008024.1	+	44004	44078	Formyl_trans_N (pfam00551)→ **RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Spy-1-15	NC_002737.1	+	43420	43494	Formyl_trans_N (pfam00551)→ **RNA**→ ←COG3942 (COG3942)
Spy-10-16	NC_003485.1	+	44027	44101	Formyl_trans_N (pfam00551)→ **RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Spy-12-16	NC_004606.1	+	38143	38217	Formyl_trans_N (pfam00551)→ **RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Spy-7-17	NC_004070.1	+	44144	44218	Formyl_trans_N (pfam00551)→ **RNA**→ MGS (pfam02142)AICARFT_IMPCHas (pfam01808)→
Spy-2-9	NZ_AAFV01000004.1	-	14571	14497	Formyl_trans_N (pfam00551)→ **RNA**→ AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→
Spy-4-17	NC_006086.1	+	84209	84281	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Spy-8-16	NC_007297.1	+	43261	43333	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Spy-9-16	NC_007296.1	+	42970	43042	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Spy-11-17	NC_008021.1	+	42997	43069	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Spy-3-15	NC_008022.1	+	43365	43437	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Spy-6-17	NC_008023.1	+	42972	43044	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Spy-5-18	NC_008024.1	+	43257	43329	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Spy-1-16	NC_002737.1	+	42673	42745	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→

Spy-10-17	NC_003485.1	+	43280	43352	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Spy-12-17	NC_004606.1	+	37396	37468	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Spy-7-18	NC_004070.1	+	43397	43469	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Spy-2-10	NZ_AAFV01000004.1	-	15318	15246	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Sag-6-1	NZ_AAJR01000025.1	-	19764	19691	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Sag-4-2	NC_007432.1	+	43094	43168	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Sag-7-2	NZ_AAJS01000005.1	+	26855	26929	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Sag-2-2	NC_004116.1	+	43284	43358	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Sag-8-1	NC_004368.1	+	43179	43253	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Sag-1-2	NZ_AAJO01000137.1	-	3851	3777	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Sag-3-1	NZ_AAJP01000052.1	-	2653	2579	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Sag-5-2	NZ_AAJQ01000011.1	+	17395	17469	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Spy-4-18	NC_006086.1	+	84208	84282	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Spy-8-17	NC_007297.1	+	43260	43334	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Spy-9-17	NC_007296.1	+	42969	43043	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Spy-11-18	NC_008021.1	+	42996	43070	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Spy-3-16	NC_008022.1	+	43364	43438	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Spy-6-18	NC_008023.1	+	42971	43045	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Spy-5-19	NC_008024.1	+	43256	43330	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Spy-1-17	NC_002737.1	+	42672	42746	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Spy-10-18	NC_003485.1	+	43279	43353	PurM (COG0150)→ **RNA**→ Formyl_trans_N (pfam00551)→
Smu-1-24	NC_004350.1	-	1146715	1146639	hypo→ **RNA**→ GyrA (COG0188)→
Smu-1-25	NC_004350.1	-	1388698	1388624	hypo→ **RNA**→ RfbB (COG1088)→
Sth-3-3	NC_006448.1	+	46470	46544	PurD (COG0151)→ **RNA**→ PurE (COG0041)→
Sth-1-3	NC_006449.1	+	46497	46571	PurD (COG0151)→ **RNA**→ PurE (COG0041)→
Sth-2-3	NZ_AAGS01000009.1	-	19372	19298	PurD (COG0151)→ **RNA**→ PurE (COG0041)→
Sag-2-3	NC_004116.1	+	1960019	1960092	←hypo **RNA**→ COG1247 (COG1247)→
Sag-8-2	NC_004368.1	+	2034313	2034386	←hypo **RNA**→ COG1247 (COG1247)→
Sag-1-3	NZ_AAJO01000111.1	+	187	260	**RNA**→ COG1247 (COG1247)→
Sag-1-4	NZ_AAJO01000170.1	-	323	250	←hypo (cons) **RNA**→
Sag-1-5	NZ_AAJO01000315.1	+	351	424	←hypo (cons) **RNA**→ COG1247 (COG1247)→
Sag-3-2	NZ_AAJP01000002.1	+	34862	34935	←hypo (cons) **RNA**→ COG1247 (COG1247)→
Ssu-1-13	NZ_AAFA02000084.1	-	3841	3765	CarA (COG0505)→ **RNA**→ MGS_CPS_II (cd01424)CarB (COG0458)CPSase_L_D3 (pfam02787)→
Ssu-1-14	NZ_AAFA02000084.1	-	5430	5354	PyrB (COG0540)→ **RNA**→ CarA (COG0505)→
Sth-3-4	NC_006448.1	+	494669	494743	PyrR (COG2065)→ **RNA**→ UraA (COG2233)→
Sth-1-4	NC_006449.1	+	489857	489931	PyrR (COG2065)→ **RNA**→ UraA (COG2233)→
Sth-2-4	NZ_AAGS01000019.1	+	23952	24027	PyrR (COG2065)→ **RNA**→ UraA (COG2233)→
Sth-3-5	NC_006448.1	-	1434137	1434063	DUF143 (pfam02410)→ **RNA**→ UbiE (COG2226)→
Sth-2-5	NZ_AAGS01000030.1	+	485	559	DUF143 (pfam02410)→ **RNA**→ UbiE (COG2226)→
Sth-1-5	NC_006449.1	-	1441782	1441708	DUF143 (pfam02410)→ **RNA**→ UbiE (COG2226)→
Smu-1-26	NC_004350.1	-	1892977	1892902	30S ribosomal protein S10→ **RNA**→ Ribosomal_L3 (pfam00297)→
Sth-3-6	NC_006448.1	+	559043	559119	TOP2c (smart00433)→ **RNA**→ GyrA (COG0188)→
Sth-1-6	NC_006449.1	+	554227	554303	TOP2c (smart00433)→ **RNA**→ GyrA (COG0188)→

Smu-1-27	NC_004350.1	+	85544	85617	GrpE (cd00446)→ **RNA**→ HSP70 (pfam00012)→
Spn-2-28	NC_003028.1	+	1527833	1527908	←Cad (pfam03596) **RNA**→ ←Ribosomal_S15p_S13e (cd00353)
Spn-1-13	NC_003098.1	+	1449240	1449315	←Cad (pfam03596) **RNA**→ ←Ribosomal_S15p_S13e (cd00353)
Spn-2-29	NZ_AAGY02000066.1	+	1022	1097	←Cad (pfam03596) **RNA**→ ←Ribosomal_S15p_S13e (cd00353)
Spn-2-30	NC_003028.1	+	557695	557770	←hypo **RNA**→ Pnp (COG1185)→
Spn-1-14	NC_003098.1	+	515543	515618	MTHFR (cd00537)→ **RNA**→ Pnp (COG1185)→
Spn-2-31	NZ_AAGY02000055.1	-	7600	7525	MTHFR (cd00537)→ **RNA**→ Pnp (COG1185)→
Spn-2-32	NC_003028.1	-	248220	248147	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Spn-2-33	NZ_AAGY02000014.1	+	9253	9326	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Spn-1-15	NC_003098.1	-	248677	248604	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Smu-1-28	NC_004350.1	+	807440	807512	CarA (COG0505)→ **RNA**→ MGS_CPS_II (cd01424)CarB (COG0458)CPSase_L_D3 (pfam02787)→
Ssu-1-15	NZ_AAFA02000011.1	+	28171	28244	hypo (cons)→ **RNA**→ ThrS (COG0441)TGS (pfam02824)→
Smu-1-29	NC_004350.1	-	869469	869388	←GdhA (COG0334) **RNA**→ hypo→
Smu-1-30	NC_004350.1	+	745037	745110	←hypo **RNA**→ ←COG5496 (COG5496)
Spn-2-34	NC_003028.1	-	248105	248030	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Spn-1-16	NC_003098.1	-	248562	248487	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Spn-2-35	NZ_AAGY02000014.1	+	9368	9443	←FusA (COG0480) **RNA**→ ←RpsG (COG0049)
Sth-3-7	NC_006448.1	+	498504	498579	CarA (COG0505)→ **RNA**→ CarB (COG0458)CPSase_L_D3 (pfam02787)MGS_CPS_II (cd01424)→
Sth-2-6	NZ_AAGS01000019.1	+	27788	27863	CarA (COG0505)→ **RNA**→ CarB (COG0458)CPSase_L_D3 (pfam02787)MGS_CPS_II (cd01424)→
Spn-2-36	NC_003028.1	+	800607	800679	hypo→ **RNA**→ GyrA (COG0188)→
Spn-1-17	NC_003098.1	+	752159	752231	TOP2c (smart00433)→ **RNA**→ GyrA (COG0188)→
Spn-2-37	NZ_AAGY02000058.1	+	3061	3133	TOP2c (smart00433)→ **RNA**→ GyrA (COG0188)→
Smu-1-31	NC_004350.1	+	87911	87986	HSP70 (pfam00012)→ **RNA**→ DnaJ (COG0484)→
Sth-3-8	NC_006448.1	+	125360	125435	DnaK (COG0443)→ **RNA**→ DnaJ (COG0484)→
Sth-1-7	NC_006449.1	+	125387	125462	DnaK (COG0443)→ **RNA**→ DnaJ (COG0484)→
Sth-2-7	NZ_AAGS01000005.1	+	51195	51270	DnaK (COG0443)→ **RNA**→ DnaJ (COG0484)→
Spn-2-38	NC_003028.1	-	1099198	1099122	AcoB (COG0022)→ **RNA**→ AceF (COG0508)→
Spn-1-18	NC_003098.1	-	1045178	1045102	AcoB (COG0022)→ **RNA**→ AceF (COG0508)→
Spn-2-39	NZ_AAGY02000015.1	+	9437	9513	AcoB (COG0022)→ **RNA**→ AceF (COG0508)→
Smu-1-32	NC_004350.1	+	841975	842048	GalT (COG4468)→ **RNA**→ GalE (COG1087)→
Sth-3-9	NC_006448.1	-	938799	938726	AceF (COG0508)→ **RNA**→ Biotin_lipoyl (pfam00364)Lpd (COG1249)→
Smu-1-33	NC_004350.1	+	609025	609098	COG4894 (COG4894)→ **RNA**→ tRNA-synt_2c (pfam01411)DHHA1 (pfam02272)→
Spn-2-40	NC_003028.1	+	200081	200156	←hypo **RNA**→ Ribosomal_S8 (pfam00410)→
Spn-2-41	NZ_AAGY02000002.1	-	22906	22831	Ribosomal_S14 (pfam00253)→ **RNA**→ Ribosomal_S8 (pfam00410)→
Spn-1-19	NC_003098.1	+	202591	202666	Ribosomal_S14 (pfam00253)→ **RNA**→ Ribosomal_S8 (pfam00410)→
Lla-1-1	NC_002662.1	-	735311	735237	MhpC (COG0596)→ **RNA**→ MenB (COG0447)→
Spn-2-42	NC_003028.1	-	1466677	1466605	hypo→ **RNA**→ hypo→
Spn-1-20	NC_003098.1	-	1402123	1402051	DegV (COG1307)→ **RNA**→ DapB (COG0289)→
Spn-2-43	NZ_AAGY02000009.1	-	12519	12447	DegV (COG1307)→ **RNA**→ DapB (COG0289)→

cd00001 PTS_IIB, PTS system, Mannose/sorbose specific IIB subunit.

cd00353 Ribosomal protein S15 (prokaryotic)_S13 (eukaryotic) binds the central domain of 16S rRNA and is required for assembly of the small ribosomal subunit and for intersubunit association, thus representing a key element in the assembly of the whole ribosome.

cd00446 GrpE is the adenine nucleotide exchange factor of DnaK (Hsp70)-type ATPases.

cd00520 Ribosome recycling factor (RRF).

cd00537 Methylenetetrahydrofolate reductase (MTHFR).

cd00733 Class II Glycyl-tRNA synthetase (GlyRS) alpha subunit core catalytic domain.

cd01360 Adenylsuccinate lyase_1: Adenylsuccinate lyase (ASL)_subgroup 1.

cd01424 Methylglyoxal synthase-like domain from type II glutamine-dependent carbamoyl phosphate synthetase (CSP).

cd01570 Nicotinate phosphoribosyltransferase (NAPRTase), subgroup A.

cd01748 Type 1 glutamine amidotransferase (GATase1) domain found in imidazole glycerol phosphate synthase (IGPS). [(5'-phosphoribulosyl)-formimino]

COG0005 Purine nucleoside phosphorylase [Nucleotide transport and metabolism]

COG0022 Pyruvate/2-oxoglutarate dehydrogenase complex, dehydrogenase (E1) component, eukaryotic type, beta subunit [Energy production and conversion]

COG0041 Phosphoribosylcarboxyaminoimidazole (NCAIR) mutase [Nucleotide transport and metabolism]

COG0049 Ribosomal protein S7 [Translation, ribosomal structure and biogenesis]

COG0051 Ribosomal protein S10 [Translation, ribosomal structure and biogenesis]

COG0060 Isoleucyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]

COG0085 DNA-directed RNA polymerase, beta subunit/140 kD subunit [Transcription]

COG0097 Ribosomal protein L6P/L9E [Translation, ribosomal structure and biogenesis]

COG0106 Phosphoribosylformimino-5-aminoimidazole carboxamide ribonucleotide (ProFAR) isomerase [Amino acid transport and metabolism]

COG0150 Phosphoribosylaminoimidazole (AIR) synthetase [Nucleotide transport and metabolism]

COG0151 Phosphoribosylamine-glycine ligase [Nucleotide transport and metabolism]

COG0188 Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), A subunit [DNA replication, recombination, and repair]

COG0199 Ribosomal protein S14 [Translation, ribosomal structure and biogenesis]

COG0216 Protein chain release factor A [Translation, ribosomal structure and biogenesis]

COG0266 Formamidopyrimidine-DNA glycosylase [DNA replication, recombination, and repair]

COG0289 Dihydrodipicolinate reductase [Amino acid transport and metabolism]

COG0334 Glutamate dehydrogenase/leucine dehydrogenase [Amino acid transport and metabolism]

COG0441 Threonyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]

COG0442 Prolyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]

COG0443 Molecular chaperone [Posttranslational modification, protein turnover, chaperones]

COG0447 Dihydroxynaphthoic acid synthase [Coenzyme metabolism]

COG0458 Carbamoylphosphate synthase large subunit (split gene in MJ) [Amino acid transport and metabolism / Nucleotide transport and metabolism]

COG0461 Orotate phosphoribosyltransferase [Nucleotide transport and metabolism]

COG0480 Translation elongation factors (GTPases) [Translation, ribosomal structure and biogenesis]

COG0484 DnaJ-class molecular chaperone with C-terminal Zn finger domain [Posttranslational modification, protein turnover, chaperones]

COG0492 Thioredoxin reductase [Posttranslational modification, protein turnover, chaperones]

COG0505 Carbamoylphosphate synthase small subunit [Amino acid transport and metabolism / Nucleotide transport and metabolism]

COG0508 Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase (E2) component, and related enzymes [Energy production and conversion]

COG0528 Uridylate kinase [Nucleotide transport and metabolism]

COG0538 Isocitrate dehydrogenases [Energy production and conversion]

COG0540 Aspartate carbamoyltransferase, catalytic chain [Nucleotide transport and metabolism]

COG0596 Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily) [General function prediction only]

COG0744 Membrane carboxypeptidase (penicillin-binding protein) [Cell envelope biogenesis, outer membrane]

COG0750 Predicted membrane-associated Zn-dependent proteases 1 [Cell envelope biogenesis, outer membrane]

COG0751 Glycyl-tRNA synthetase, beta subunit [Translation, ribosomal structure and biogenesis]

COG0813 Purine-nucleoside phosphorylase [Nucleotide transport and metabolism]

COG1087 UDP-glucose 4-epimerase [Cell envelope biogenesis, outer membrane]

COG1088 dTDP-D-glucose 4,6-dehydratase [Cell envelope biogenesis, outer membrane]

COG1159 GTPase [General function prediction only]

COG1185 Polyribonucleotide nucleotidyltransferase (polynucleotide phosphorylase) [Translation, ribosomal structure and biogenesis]

COG1247 Sortase and related acyltransferases [Cell envelope biogenesis, outer mem-

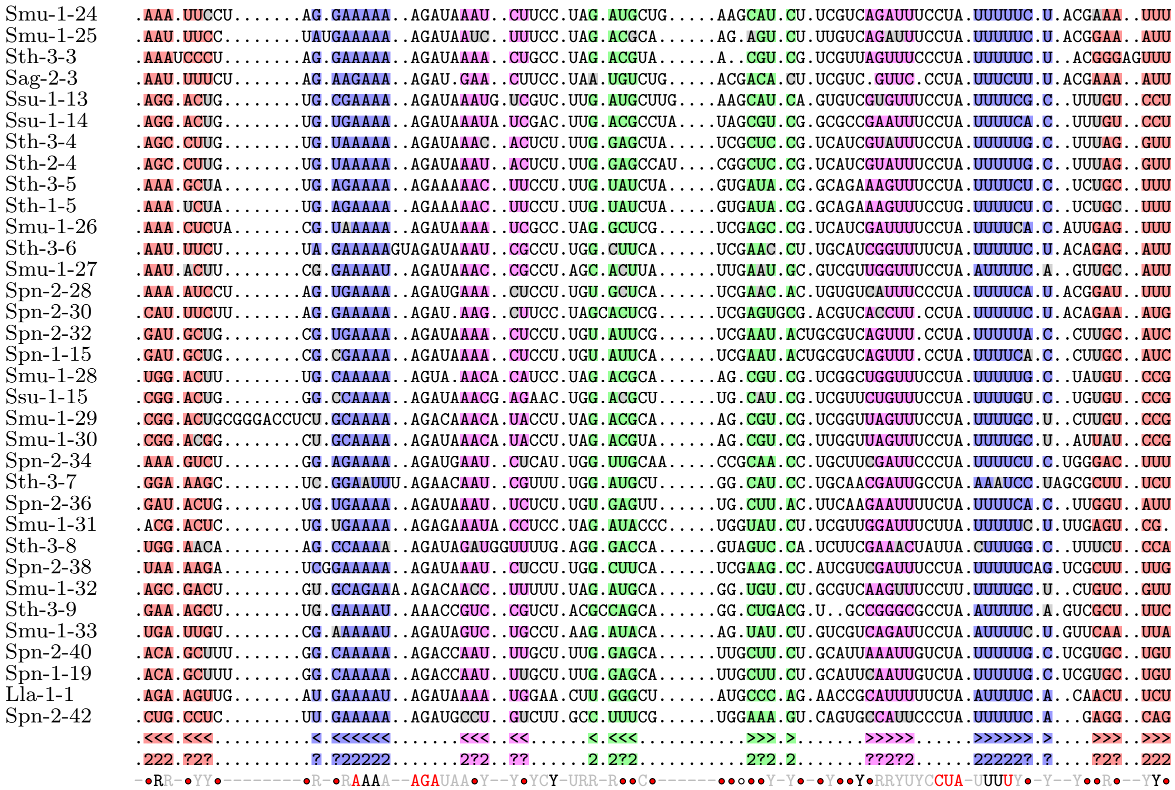
brane]
COG1249 Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component, and related enzymes [Energy production and conversion]
COG1307 Uncharacterized protein conserved in bacteria [Function unknown]
COG1435 Thymidine kinase [Nucleotide transport and metabolism]
COG1473 Metal-dependent amidase/aminoacylase/carboxypeptidase [General function prediction only]
COG1481 Uncharacterized protein conserved in bacteria [Function unknown]
COG1881 Phospholipid-binding protein [General function prediction only]
COG1898 dTDP-4-dehydrorhamnose 3,5-epimerase and related enzymes [Cell envelope biogenesis, outer membrane]
COG2065 Pyrimidine operon attenuation protein/uracil phosphoribosyltransferase [Nucleotide transport and metabolism]
COG2171 Tetrahydrodipicolinate N-succinyltransferase [Amino acid transport and metabolism]
COG2226 Methylase involved in ubiquinone/menaquinone biosynthesis [Coenzyme metabolism]
COG2233 Xanthine/uracil permeases [Nucleotide transport and metabolism]
COG3064 Membrane protein involved in colicin uptake [Cell envelope biogenesis, outer membrane]
COG3942 Surface antigen [General function prediction only]
COG4468 Galactose-1-phosphate uridyltransferase [Carbohydrate transport and metabolism]
COG4894 Uncharacterized conserved protein [Function unknown]
COG5280 Phage-related minor tail protein [Function unknown]
COG5340 Predicted transcriptional regulator [Transcription]
COG5412 Phage-related protein [Function unknown]
COG5496 Predicted thioesterase [General function prediction only]
pfam00012 Hsp70 protein.
pfam00253 Ribosomal protein S14p/S29e.
pfam00293 NUDIX domain.
pfam00297 Ribosomal protein L3.
pfam00318 Ribosomal protein S2.
pfam00364 Biotin-requiring enzyme.
pfam00410 Ribosomal protein S8.
pfam00551 Formyl transferase.
pfam00583 Acetyltransferase (GNAT) family.
pfam00889 Elongation factor TS.
pfam01411 tRNA synthetases class II (A).
pfam01609 Transposase DDE domain.
pfam01808 AICARFT/IMPCHase bienzyme.
pfam02142 MGS-like domain.
pfam02272 DHHA1 domain.
pfam02410 Domain of unknown function DUF143.
pfam02787 Carbamoyl-phosphate synthetase large chain, oligomerisation domain.
pfam02824 TGS domain.
pfam03577 Peptidase family U34.
pfam03596 Cadmium resistance transporter.
pfam03609 PTS system sorbose-specific iic component.
pfam06325 Ribosomal protein L11 methyltransferase (PrmA).
pfam06519 TolA protein.
pfam06541 Protein of unknown function (DUF1113).
smart00342 helix_turn_helix, arabinose operon control protein;
smart00433 TopoisomeraseII; Eukaryotic DNA topoisomerase II, GyrB, ParE

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Sag-1-1, Sag-1-2, Sag-1-3, Sag-1-4, Sag-1-5, Sag-3-1, Sag-3-2, Sag-5-1, Sag-5-2, Sag-7-1, Sag-7-2, Sag-8-1, Sag-8-2, Smu-1-12, Smu-1-5, Smu-1-6, Smu-1-7, Smu-1-8, Spn-1-1, Spn-1-10, Spn-1-11, Spn-1-13, Spn-1-14, Spn-1-16, Spn-1-17, Spn-1-18, Spn-1-2, Spn-1-20, Spn-1-3, Spn-1-5, Spn-1-6, Spn-1-8, Spn-1-9, Spn-2-10, Spn-2-12, Spn-2-14, Spn-2-16, Spn-2-17, Spn-2-19, Spn-2-2, Spn-2-21, Spn-2-23, Spn-2-25, Spn-2-27, Spn-2-29, Spn-2-31, Spn-2-33, Spn-2-35, Spn-2-37, Spn-2-39, Spn-2-4, Spn-2-41, Spn-2-43, Spn-2-6, Spn-2-8, Spy-1-1, Spy-1-10, Spy-1-11, Spy-1-12, Spy-1-13, Spy-1-14, Spy-1-15, Spy-1-16, Spy-1-17, Spy-1-2, Spy-1-3, Spy-1-4, Spy-1-5, Spy-1-6, Spy-1-7, Spy-1-8, Spy-1-9, Spy-10-10, Spy-10-13, Spy-10-14, Spy-10-15, Spy-10-16, Spy-10-17, Spy-10-18, Spy-10-2, Spy-10-3, Spy-10-4, Spy-10-5, Spy-10-6, Spy-10-7, Spy-11-1, Spy-11-10, Spy-11-12, Spy-11-13, Spy-11-15, Spy-11-16, Spy-11-17, Spy-11-18, Spy-11-2, Spy-11-3, Spy-11-4, Spy-11-5, Spy-11-6, Spy-11-7, Spy-11-8, Spy-11-9, Spy-12-1, Spy-12-10, Spy-12-12, Spy-12-14, Spy-12-15, Spy-12-16, Spy-12-2, Spy-12-3, Spy-12-4, Spy-12-5, Spy-12-6, Spy-12-7, Spy-12-8, Spy-12-9, Spy-2-1, Spy-2-10, Spy-2-4, Spy-2-5, Spy-2-6, Spy-2-8, Spy-2-9, Spy-3-1, Spy-3-10, Spy-3-11, Spy-3-12, Spy-3-13, Spy-3-14, Spy-3-15, Spy-3-16, Spy-3-2, Spy-3-3, Spy-3-4, Spy-3-5, Spy-3-7, Spy-3-8, Spy-4-2, Spy-4-3, Spy-4-4, Spy-4-5, Spy-4-6, Spy-5-1, Spy-5-10, Spy-5-12, Spy-5-13, Spy-5-14, Spy-5-15, Spy-5-16, Spy-5-17, Spy-5-18, Spy-5-19, Spy-5-2, Spy-5-3, Spy-5-4, Spy-5-5, Spy-5-7, Spy-5-9, Spy-6-1, Spy-6-10, Spy-6-11, Spy-6-12, Spy-6-13, Spy-6-14, Spy-6-15, Spy-6-16, Spy-6-17, Spy-6-18, Spy-6-2, Spy-6-3, Spy-6-4, Spy-6-5, Spy-6-6, Spy-6-7, Spy-6-8, Spy-6-9, Spy-7-1, Spy-7-10, Spy-7-11, Spy-7-12, Spy-7-13, Spy-7-14, Spy-7-15, Spy-7-16, Spy-7-17, Spy-7-18, Spy-7-2, Spy-7-3, Spy-7-4, Spy-7-5, Spy-7-6, Spy-7-7, Spy-7-8, Spy-7-9, Spy-8-11, Spy-8-14, Spy-8-15, Spy-8-16, Spy-8-17, Spy-8-2, Spy-8-3, Spy-8-4, Spy-8-5, Spy-8-6, Spy-8-7, Spy-9-1, Spy-9-10, Spy-9-11, Spy-9-14, Spy-9-15, Spy-9-16, Spy-9-17, Spy-9-2, Spy-9-3, Spy-9-4, Spy-9-5, Spy-9-7, Spy-9-8, Spy-9-9, Ssu-1-8, Sth-1-2, Sth-1-3, Sth-1-4, Sth-1-6, Sth-1-7, Sth-2-1, Sth-2-3, Sth-2-5, Sth-2-6, Sth-2-7

alignment positions 1 · · · 105

```
  Sag-4-1     .AAA.GCUG........UG.UAAAAA..AGAGAGAU..GGACU.UUA.GUGCA......AG.CAC.CA.UUCUUCCAUCUCCUA.UUUUUA.C..UUUGC..UUU
  Spn-2-1     .AAA.GCUG........UG.UAAAAA..AGAUAAAC..UGUCU.UGU.CUUCA......UCGAAG.AC.UUCGUCAGUUUCCUA.UUUUUA.C..UUUGC..UUU
  Spn-2-3     .AGC.UCUG........AG.UAAAAA..AGAUAGAU..UUUCU.UGU.GUUCA......UCGAAC.AC.UUCGUCAAUCUCCUA.UUUUUA.C..UGCGA..GCU
1 Spy-8-1     .ACU.UCCU........AG.GAAAAU..GGACAAUC..UUCCCUUGU.GUGCA......AGGCCC.AC.AUGGUCAGAUUCCUA.AUUUUC.U..ACAGA..AGU
1 Spy-9-6     .ACU.UCCU........AG.GAAAAU..AGACAAUC..UUCCCUUGU.GUGCA......AGGCCC.AC.AUGGUCAGAUUCCUA.AUUUUC.U..ACAGA..AGU
1 Spy-4-1     .ACU.UCCU........AG.GAAAAU..AGACAAUCU.UCGCU.UGU.CUGCA......AGGCAC.AC.AUGGUCAGAUUCCUA.AUUUUC.U..ACAGA..AGU
  Spy-3-6     .ACU.UCCU........AG.GAAAAU..AGACAAUCU.UCGCU.UGU.GUGCA......AGGCAC.AC.AUGGUCAGAUUCCUA.AUUUUC.U..ACAGA..AGU
1 Spy-5-6     .ACU.UCCU........AG.GAAAAU..AGACAAUCU.UCGCA.UGU.GUGCA......AGGCAC.AC.AUGGUCAGAUUCCUA.AUUUUC.U..ACAGA..AGU
1 Ssu-1-1     .ACU.UCUU........AG.GAAAAU..AGAUAAUC..UUCCU.UGU.GUGCA......AGACAC.AC.AAUGUCAGAUUCCUA.AUUUUC.U..ACAGA..AGU
  Spy-10-1    .AAC.UUCCU.......AG.GAAAAU..AGACAAUCU.UCGUU.UGU.GUGCA......AGGCAC.AC.AUGGUCAGAUUCCUA.AUUUUC.U.ACAGAA..GUU
  Smu-1-1     .AAA.GCGU........CG.AGAAAA..AGAUAGAC..UGCCU.UGU.GCGCC......AG.CGC.AC.AGCGUUAGUCUCCUA.UUUUCU.U..CUCGC..UUU
  Smu-1-2     .AAA.ACGU........CG.AGAAAA..AGAUAGCC..UGCCU.UAU.GUGCA......AG.CGC.AC.GGUGUCAGCCUCCUA.UUUUCU.U..CUCGU..UUU
  Smu-1-3     .AAA.GCGU........CG.UGAAAA..AGAUAAAC..UUCCU.UGU.GUGCA......AGUCAC.AC.GGCGUCAGUUUCCUA.UUUUCA.U..CUCGC..UUU
  Ssu-1-2     .AAA.UCUU........AG.UGAAAA..AGAUAAAC..UUCCU.UGU.AUGCA......AG.CAU.AC.UGCGUCAGUUUCCUA.UUUUCA.U..ACAGA..UUU
1 Smu-1-4     .AAA.UCCU........UG.GAAAAA..AGAGAAAU..GGUCU.GGU.GUUCA......UCGAAC.AC.UGCACCCAUUUCCUA.UUUUUC.A.GGCGGA..UUU
  Spy-4-7     .GGA.AAGC........UCGGAAACU..AGAUAAAC..UAACU.UUU.GAGCU......CA.CAC.AU.GGCGUUAGUUCCCUA.AAUUUCAG.UCGCUU..UCU
  Spy-2-2     .GAA.AGCU........CG.GAAACU..AGAUAAAC..UAACU.UUU.GUGCU......CA.CAC.AU.GGCGUUAGUUCCCUA.AAUUUC.A.GUCGCU..UUC
  Ssu-1-3     .AAA.AGCU........CG.GAAAAU..AGAUAGAG..CUCCU.UGU.GUGCA......AGUCAC.AC.AGCGUCGCUCUCCUA.AUUUUC.A.GUCGCU..UUU
  Spy-2-3     .GAA.AGCU........CG.GAAUUU..AGAUAAAC..CGCCU.AGG.ACGCA......AG.CGU.CC.GUUGUUGGUUUCCUA.AAAUUC.A.GUCGCU..UUC
  Spy-4-8     .GAA.AGCU........CG.GAAUUU..AGAUAAAC..CGCCU.AGG.ACGCA......AG.CGU.CC.AUUGUUGGUUUCCUA.AAAUUC.A.GUCGCU..UUC
  Spy-5-8     .GAA.AGCU........CG.GAAUUU..AGAUAAAC..CGCCU.AGG.ACGCA......AG.CGU.CC.AUUGUCGGUUUUCUA.AAAUUC.A.GUCGCU..UUC
  Ssu-1-4     .GAA.GGCU........CG.GAAAAA..AGAUAAAC..UGCCU.AGG.AAAUCA.....AGAUUU.C...CUGUCAGUUUCCUA.UUUUUC.A.GUCGCC..UUC
  Spy-10-8    .GGA.AAGC........UC.GGAAUUU.AGAUAAAC..CGCCU.AGG.ACGCA......AG.CGU.CC.AUUGUUGGUUUUCUAAAAUUCA.G.UCGCUU..UCU
  Spn-2-5     .AGA.ACUG........UG.CAAAAU..AGAUAAGU..UGGCU.UGU.UUGUA......AA.CAA.AC.UUCUCCAACUUCCUA.AUUUUG.C..UUUGU..UCU
  Spn-1-4     .AAA.GCAG........UG.GGAAAA..AGAUAGUU..GGUCU.AGC.GAGCA......UCGCUC.AC.UGCGCCCAACUCCUA.UUUUCC.C..UUCGC..UUU
  Spn-2-7     .AAA.GUAG........UG.GGAAAA..AGAU.AGU..UGAUC.UAGCGAGCA......UCGCUCACU.GCGCCCAACU.CCUA.UUUUCC.C..UUCGC..UUU
  Smu-1-9     .AAA.CCUG........UG.UGAAAA..AGAUAAAU..UUUCC.UAG.AAACCU.....GUGUUU.CU.UCGUCAAGUUAUCUA.UUUUCA.C.UUUGGC..UUU
  Spy-10-9    .AAC.GCUG........UG.UGAAAA..AGAUAAAC..UUUCC.UAG.AGACAA.....AGGUUU.CU.UCAUCAAAUUUCCUA.UUUUCA.C..UUUGC..GUU
  Spy-8-8     .AAC.GCUG........UG.UGAAAA..AGAUAAAC..UUUCC.UAG.AGACAA.....AGGUUU.CU.UCGUCAAAUUUCCUA.UUUUCA.C..UUUGC..GUU
  Spy-4-9     .AAC.GCUG........UG.UGAAAA..AGAUAAAC..UUUCC.UAG.AAACAA.....AGGUUU.CU.UCGUCAAAUUUCCUA.UUUUCA.C..UUUGC..GUU
  Spn-2-9     .AAC.UCUG........UG.UGAAAA..AGAUAAAC..UUUCC.UAG.AAACUU.....UAGUUU.CU.UCGUCAAGUUUCCUA.UUUUUA.C..UUGGA..GUU
  Spn-2-11    .AGC.ACUU........GG.CAAAAA..AGAUAAAA..UCUCC.UAG.AAACUG.....AAGUUU.CU.UCGUCAAUUUUCCUA.UUUUUG.C..UUUGU..GCU
  Spn-2-13    .AAC.ACUG........UG.CCAAAA..AGAUAAACU.UCUCU.UAG.ACACA......AG.CGU.CU.UCAGAGAAUUUCCUA.UUUUGG.U..UUUGU..GUU
  Spy-8-9     .AAC.UCUA........CA.UGAAAA..AGAUAAAU..CUUCC.UCG.AAACAG.....AAGCUU.UU.GUGUUAGAUUUUCUA.UUUUUA.U..UUUGA..GUU
  Spy-4-10    .AAC.UCUA........CA.UGAAAA..AGAUAAAU..CUUCC.UCG.AAACAG.....AAGUUU.UU.GUGUUAGAUUUUCUA.UUUUUA.U..UUUGA..GUU
  Spy-10-11   .AAA.UCAC........UG.ACAAAA..AGAUAAUA..GUUCC.UAG.AAACUG.....AAGUUU.CU.GCGUCAAUUUCCCUA.UUUUGU.C..UAGCA..UUU
  Spy-8-10    .AAG.CCAC........UG.ACAAAA..AGAUAAUA..GUUCC.UAG.AAACUG.....AAGUUU.CU.GCGUCAAUUUCCCUA.UUUUGU.C..UAGCA..UUU
  Spy-4-11    .AAA.CCAC........UG.ACAAAA..AGAUAAUA..GUUCC.UAG.AAACUG.....AAGUUU.CU.GCGUCACUAUUCCUA.UUUUGU.C..UUGGA..CUU
  Spy-5-11    .AAG.CCAC........UG.ACAAAA..AGAUAAUA..GUUCC.UAG.AAACUG.....AAGUUU.CU.GCGUCACUAUUCCUA.UUUUGU.C..UUGGA..CUU
  Spy-11-11   .AAA.UGUC........UG.CCACAA..AGAUAAAA..UUACC.UAG.AAACUG.....AAGUUU.CU.GCGUCAAUUUCCCUA.UUUUGG.C..UAGCA..UUU
  Spy-12-11   .AAA.UGUC........UG.CCAAAG..AGAUAAAA..UUGCC.UAG.AAACUG.....AAGUUU.CU.GCGUCAAUUUCCCUA.UUUUGG.C..UAGCA..UUU
  Spn-1-7     AAGG.UUUG........UG.UGAAAA...GAUAAAU..CUUUC.UAG.AGUCUA.....AAGACU.CU.GCGUCAGAUUUCCUA.UUUUCA.C..UGUAA..CCU
  Spn-2-15    .AGG.UUUG........UG.UGAAAA..AGAUAAAU..CUUUC.UAG.AGUCUA.....AAGACU.CU.GCGUCAGAUUUCCUA.UUUUCA.C..UGUAA..CCU

              .<<<.<<<..........<.<<<<<<.......<<<..<<......<.<<<...........>>>.>.......>>>>>......>>>>>>.>....>>>..>>>
              .222.?2?..........?.?22222.......2?2..??......2.2?2...........2?2.2.......??2?2......22222?.?....?2?..222
              -•RR-•YY•--------•R-•RAAAA--AGAUAA•Y--Y•YCY-URR-R••C•------••○••Y-Y•-•Y••Y•RRYUYCCUA-UUUUY•-Y--Y••R•--YY•
```

```
Spy-3-9     .UGG.UCUG........UG.UGAAAA..AGAUAAAU..CUUCC.UUG.AGUAUA.....AACACG.CU.ACGUCAGAUUUCCUA.UUUUCA.C..UUUGU..CCA
Spy-10-12   .UGG.UCUG........UG.UGAAAA..AGAUAAAU..CCUCC.UUGAGUGUA......AA.CACGCU.ACGUCAGAUUUCCUA.UUUUCA.C..UUUGU..CCA
Spy-4-12    .UGG.UCUG........UG.UGAAAA..AGAUAAAU..CUUCC.UUGAGUGUA......AA.CACGCU.ACGUCAGAUUUCCUA.UUUUCA.C..UUUGU..CCA
Smu-1-10    .CGG.AUUG........UG.GCAAAA..AGAUGAGU..CCUCC.UAG.AAUCUUA....AUGAUU.CC.GCGUCGUACUCCCUA.UUUUAC.C..UUUAU..CCG
Smu-1-11    .CGG.AUUG........UG.GCAAAA..AGAUGAGU..CCUCC.UAG.AAUCUUA....AUGAUU.CC.GCGUCGUACUCCCUA.UUUUGC.C..UUUAU..CCG
Smu-1-13    .CGG.AUUG........UG.GCAAAA..AGAUGAGU..CCUCC.UAG.AAUCUUA....AUGAUU.CC.GCGUCGUACUCCCUA.UUUUGC.C..UUUAU..UCG
Sth-3-1     .AGG.AUCG........UG.UCAAAA..AGAUAAGU..UCUCC.UAG.AAUCGA.....AAGAUU.CA.GCAUUAAACUUCCUA.UUUUGA.C..UUUAU..CCG
Sth-1-1     .AGG.AUCG........UG.UCAAAA..AGAUAAGU..UCUCC.UAG.AAUCGA.....AAGAUU.CA.GCGUUAAAUUUCCUA.UUUUGA.C..UUUAU..CCG
Ssu-1-5     .CGG.CCUG........UG.UGAAAA..AGAUAACU..CUAUC.UAGAACGCA......AG.CGUUCU.UCGCUAGAGUUCCUA.UUUUCA.C..UUUGU..CCG
Spn-2-18    .AAC.CCUG........UG.UGAAAA..AGAUAGUU..CUUCC.AAG.AAGUAG.....ACACUU.CU.UGAUUGGAAUUCCUA.UUUUCA.C..UUUGU..GUU
Sag-2-1     .CGA.GCUG........UA.CAAAAAA.AGAGAGUU..UCUCA.UAG.AACGG......AAAGUU.CU.UCGUCAAAACUCCUA.UUUUUG.U..CGUGC..UCG
Ssu-1-6     .CGA.GCUG........UA.CAAAAA..AGAGAGUU..UCUCC.UAG.AACGG......AAAGUU.CU.UCGUCAAAACUCCUA.UUUUUG.U..CAUGC..UCG
Ssu-1-7     .CGA.GCUG........UA.CAAAAA..AGAGAGUU..UCUCC.UAG.AACGG......AAAGUU.CU.UCGUCAAAACUCCUA.UUUUUG.U..CGUGC..UCG
Spn-2-20    .CGG.ACGG........UG.CUAAAA..AGAUAGUU..AUUCC.UAG.AACUG......ACAGUU.CU.UCGUCAUAACUCCUA.UUUUAG.C..UUUGU..CCG
Spn-2-22    .UGG.ACUG........UG.CCAAAA..AGAUAGUU..UUUUC.UAG.GACGUA.....AGCGUC.CG.UCGUCAAAACUCCUA..GAUGG.C..UGUGU..CCG
Ssu-1-9     .CGG.ACUG........UG.CCAAAA..AGAUAGUU..UUUCC.UAG.GACGGU.....GAGGUC.CU.GCGUCAAAACUCCUA.UUUUGG.C..UGUGU..CCG
Sth-3-2     .CGG.ACUU........GG.UCGAAA..AGAUAGUU..UUUCC.UAG.AAGCUG.....ACGCUU.CU.UCGUCUAAACUCCUA.UUUUGA.C..UGUGU..CCG
Sth-2-2     .CGG.ACUU........GG.UCGAAA..AGAUAGUU..UUUCC.UAG.AAGCUG.....AUGCUU.CU.UCGUCUAAACUCCUA.UUUUGA.C..UGUGU..CCG
Smu-1-14    .CGG.ACUA........GG.UCAAAA..AGAUAGUU..UUUCC.UAG.AAGCAG.....UCGCUU.CU.UCGUCGAAACUCCUA.UUUUGA.C..UGUGU..CCG
Ssu-1-10    .CGA.ACUG........UG.CAAAAA..AGAUAGUU..UUUCC.UAGGACGCU......UU.CGUCCU.UCGUCAAAACUCCUA.UUUUUG.C..UUUGU..UCG
Smu-1-15    .CGA.CACU........GU.GCAAAA..AGAUAGCU..UCUCC.UAG.AGUUCU.....GAGACU.CU.UGGUUGGACCUCCUA.UUUUGC.U..UUGUG..UCG
Smu-1-16    .CGA.CACU........GU.GCAAAA..AGAUAGCU..UCUCC.UAG.AGUUCU.....GAGACU.CU.UCGUUGAACCUCCUA.UUUUGC.U..UUGUG..UCG
Smu-1-17    .CGA.UCUG........UG.UCAAAA..AGACAGCU..UCUCC.UAG.AGUCCU.....GAAACU.CU.UC.ACGAACCUCCUA.UUUUGC.C..UUUGC..UCG
Smu-1-18    .CGA.GCUG........UA.UUAAAG..AGAUAGUU..UCUCC.UAG.CGUCUU.....GUGACU.CU.UCGUCAAAACUUCUA.UUUUGC.C..UUUGC..UCG
Smu-1-19    .CGA.GCUA........UG.UCAAAA..AGGUAGUU..UCUUC.UAG.AGUCCAGUCCUGUGACU.CU.UCGUCAAAACUCCUA.UUUUGC.C..UUUGC..UCG
Spy-8-12    .CGG.ACUA........UG.UCAAAA..AGACCAAGA.CUUCC.UAG.GUGCA......AG.CAC.CU.GCGUC.GCUUGCUA.UUUUGA.U..GGUGU..CCG
Spy-4-13    .CGG.ACUA........UG.UCAAAA..AGACCAAGA.CUUCC.UAG.GUGCA......AG.CAC.CU.GCGUC.GCUUGCUA.UUUUGA.U..GGUGU..CCA
Spy-12-13   .CGG.ACUA........UG.UCAAAA..AGACAAAGA.CUUCC.UAG.GUGCA......AG.CAC.CU.GCGUU.GCUUGCUA.UUUUGA.U..GGGGU..CCA
Spy-9-12    .AGA.ACUG........UG.UAAAAA..AGAU.AAA..CUGUC.UAG.CAUCU......GGGAUG.CG.UCGUCAGUUU.CCUA.UUUUUA.C..UUUGU..UCU
Smu-1-20    .AAA.GCAU........CG.AGAAAA..AGAU.AAA..CUUCC.UAG.UGACA......AAGUCA.CG.UCGUCAGUUU.CCUA.UUUUCU.A..CUUGC..UUU
Ssu-1-11    .AAA.ACUU........UA.CGAAAA..AGAUAAAC..CUGCC.UAG.UGCCA......AAAGCA.CU.GCGUUAGGUUUCCUA.UUUUCG.U..UUCGU..UUU
Spn-2-24    .AAA.CUCGG.......UG.UGAAAA..AGAU.AAA..CUUCC.UAG.UGUCUGC....UAGACA.CU.GCGUCAGUUU.CCUA.UUUUCA.C.GUUGAG..UUU
Smu-1-21    .AAA.UCCU........AG.UAAAAA..AGAU.AAA..CUUCC.UUG.UGUCUUG....GCGACA.CG.GCGUCAGUUU.CCUA.UUUUUA.U..ACGGA..UUU
Spy-9-13    .CGA.CUCU........GC.GAAAAA..AGAGAAAUA.CCUCC.UAG.AUACGA.....GAGUAU.CU.UUGUCGGAUUUCCUA.UUUUUC.A..UUGAG..UCG
Spy-8-13    .CGA.CUCU........GC.GAAAAA..AGAGAAAUA.CCUCC.UAG.AUACGA.....GAGUAU.CU.UUGUCGGAUUUCCUA.UUUUUC.A..UUAAG..UCG
Spy-11-14   .CGA.CCCU........GC.GAAAAA..AGAGAAAUA.CCUCC.UAG.AUACGA.....GAGUAU.CU.UUGUCGGAUUUCCUA.UUUUUC.A..UUGAG..UCG
Spy-4-14    .CGA.UUCU........GC.GAAAAA..AGAGAAAUA.CCUCC.UAG.AUACGA.....GAGUAU.CU.UUGUCGGAUUUCCUA.UUUUUC.A..UUGAG..UCG
Spy-2-7     .CGA.CUCU........GC.GAAAAA..AGAGAAAUA.CCUCC.UAG.AUACAA.....GAGUAU.CU.UCGUCGGAUUUCCUA.UUUUUC.A..UUGAG..UCG
Smu-1-22    .CGA.CACU........GU.GCAAAA..AGAUAAAGA.CUGCC.UAG.AUGCUG.....GUGCAU.CU.UCGUCAGCUUUCCUA.UUUUGC.G..GUGUG..UCG
Smu-1-23    .CGA.CACU........GU.GCAAAA..AGAUAAAGA.CUGCC.UAA.AUGCUG.....UCGCAC.CU.UCAUCAGCUUUCCUA.UUUUGC.G..GUGUG..UCG
Ssu-1-12    .CAA.CUCU........GU.GAAAAA..AGAUAAACA.GCUCC.UAG.GAGGCU.....UGCCUC.CG.UCGUCGUGUUUCCUA.UUUUUC.U..UUGAG..UUG
Spn-1-12    .CAA.CUCU........GU.GGAAAA..AGAUAAAU..CUCCC.UAG.GAGCAG.....UCGCUC.CG.UCGUUCGAUUUCCUA.UUUUCC.U..UUGAG..UUG
Spn-2-26    .CAA.CUCU........GU.GGAAAA..AGAUAAAU..CUCCC.UAG.GAGUAG.....UCGCUC.CG.UCGUUCGAUUUCCUA.UCUUCC.U..UUGAG..UUG
Spy-4-15    .AAG.GCA.........CA.GAAAAA..AGAUAAACA.UCUCC.UAG.GCGCA......AG.CGU.CU.UCGUCGAGUUUCCUA.UUUUUC.C..UUGGC..CUU
Spy-4-16    .GAA.GGCA........CA.GAAAAA..AGAUAAACA.UCUCC.UAG.GCGCA......AG.CGU.CU.UCGUCGAGUUUCCUA.UUUUUC.C.UUGGCC..UUC
Spy-4-17    .AAU.UUCU........AG.GAAAAA..AGAUAAAA..UGACU.UAG.ACGCA......AG.CGU.CG.UUGUUCGUUUUCCUA.UUUUUC.U..ACGAA..AUU
Spy-12-17   .AAU.UUCU........AG.GAAAAA..AGAUAAAA..CGACU.UAG.ACGCA......AG.CGU.CG.UUGUUCGUUUUUCUA.UUUUUC.U..ACGAA..AUU
Sag-6-1     .AAA.UUUCU.......GG.GAAAAA..AGAUAAAA..CGACU.UAG.ACGCA......AG.CGU.CG.UUGUUCGUUUUCCUA.UUUUUC.U..AGAAA..UUU
Sag-4-2     .AAA.UUUCU.......AG.GAAAAA..AGAUAAAA..CGACU.UAG.ACGCA......AG.UGU.CG.UUGUUCGUUUUCCUA.UUUUUC.U.ACGAAA..UUU
Sag-2-2     .AAA.UUUCU.......AG.GAAAAA..AGAUAAAA..CGACU.UAG.ACGCA......AG.CGU.CG.UUGUUCGUUUUCCUA.UUUUUC.U.ACGAAA..UUU
Spy-4-18    .AAA.UUUCU.......AG.GAAAAA..AGAUAAAA..UGACU.UAG.ACGCA......AG.CGU.CG.UUGUUCGUUUUCCUA.UUUUUC.U.ACGAAA..UUU

            .<<<.<<<..........<.<<<<<<.......<<<..<<......<.<<<...........>>>.>.......>>>>>.....>>>>>>.>....>>>..>>>
            .222.?2?..........?.?22222.......2?2..??......2.2?2...........2?2.2.......??2?2.....22222?.?....?2?..222
            -•RR-•YY•--------•R-•RAAAA--AGAUAA•Y--Y•YCY-URR-R••C•------••○••Y-Y•-•Y••Y•RRYUYCCUA-UUUUY•-Y--Y••R•--YY•
```



```
Smu-1-24  .AAA.UUCCU.......AG.GAAAAA..AGAUAAAU..CUUCC.UAG.AUGCUG.....AAGCAU.CU.UCGUCAGAUUUCCUA.UUUUUC.U.ACGAAA..UUU
Smu-1-25  .AAU.UUCC........UAUGAAAAA..AGAUAAUC..UUUCC.UAG.ACGCA......AG.AGU.CU.UUGUCAGAUUUCCUA.UUUUUC.U.ACGGAA..AUU
Sth-3-3   .AAAUCCCU........AG.GAAAAA..AGAUAAAA..CUGCC.UAG.ACGUA......A..CGU.CG.UCGUUAGUUUCCCUA.UUUUUC.U..ACGGGAGUUU
Sag-2-3   .AAU.UUUCU.......AG.AAGAAA..AGAU.GAA..CUUCC.UAA.UGUCUG.....ACGACA.CU.UCGUC.GUUC.CCUA.UUUCUU.U.ACGAAA..AUU
Ssu-1-13  .AGG.ACUG........UG.CGAAAA..AGAUAAAUG.UCGUC.UUG.AUGCUUG....AAGCAU.CA.GUGUCGUGUUUCCUA.UUUUCG.C..UUUGU..CCU
Ssu-1-14  .AGG.ACUG........UG.UGAAAA..AGAUAAAUA.UCGAC.UUG.ACGCCUA....UAGCGU.CG.GCGCCGAAUUUCCUA.UUUUCA.C..UUUGU..CCU
Sth-3-4   .AGC.CUUG........UG.UAAAAA..AGAUAAAC..ACUCU.UUG.GAGCUA.....UCGCUC.CG.UCAUCGUAUUUCCUA.UUUUUG.C..UUUAG..GUU
Sth-2-4   .AGC.CUUG........UG.UAAAAA..AGAUAAAU..ACUCU.UUG.GAGCCAU....CGGCUC.CG.UCAUCGUAUUUCCUA.UUUUUG.C..UUUAG..GUU
Sth-3-5   .AAA.GCUA........UG.AGAAAA..AGAAAAAC..UUCCU.UUG.UAUCUA.....GUGAUA.CG.GCAGAAAGUUUCCUA.UUUUCU.C..UCUGC..UUU
Sth-1-5   .AAA.UCUA........UG.AGAAAA..AGAAAAAC..UUCCU.UUG.UAUCUA.....GUGAUA.CG.GCAGAAAGUUUCCUG.UUUUCU.C..UCUGC..UUU
Smu-1-26  .AAA.CUCUA.......CG.UAAAAA..AGAUAAAA..UCGCC.UAG.GCUCG......UCGAGC.CG.UCAUCGAUUUUCCUA.UUUUCA.C.AUUGAG..UUU
Sth-3-6   .AAU.UUCU........UA.GAAAAAGUAGAUAAAU..CGCCU.UGG.CUUCA......UCGAAC.CU.UGCAUCGGUUUUCUA.UUUUUC.U.ACAGAG..AUU
Smu-1-27  .AAU.ACUU........CG.GAAAAU..AGAUAAAC..CGCCU.AGC.ACUUA......UUGAAU.GC.GUCGUUGGUUUCCUA.AUUUUC.A..GUUGC..AUU
Spn-2-28  .AAA.AUCCU.......AG.UGAAAA..AGAUGAAA..CUCCU.UGU.GCUCA......UCGAAC.AC.UGUGUCAUUUCCCUA.UUUUCA.U.ACGGAU..UUU
Spn-2-30  .CAU.UUCUU.......AG.GAAAAA..AGAU.AAG..CUUCC.UAGCACUCG......UCGAGUGCG.ACGUCACCUU.CCUA.UUUUUC.U.ACAGAA..AUG
Spn-2-32  .GAU.GCUG........CG.UGAAAA..AGAUAAAA..CUCCU.UGU.AUUCG......UCGAAU.ACUGCGUCAGUUU.CCUA.UUUUUA.C..CUUGC..AUC
Spn-1-15  .GAU.GCUG........CG.CGAAAA..AGAUAAAA..CUCCU.UGU.AUUCA......UCGAAU.ACUGCGUCAGUUU.CCUA.UUUUCA.C..CUUGC..AUC
Smu-1-28  .UGG.ACUU........UG.CAAAAA..AGUA.AACA.CAUCC.UAG.ACGCA......AG.CGU.CG.UCGGCUGGUUUCCUA.UUUUUG.C..UAUGU..CCG
Ssu-1-15  .CGG.ACUG........GG.CCAAAA..AGAUAAACG.AGAAC.UGG.ACGCU......UG.CAU.CG.UCGUUCUGUUUCCUA.UUUUGU.C..UGUGU..CCG
Smu-1-29  .CGG.ACUGCGGGACCUCU.GCAAAA..AGACAAACA.UACCU.UAG.ACGCA......AG.CGU.CG.UCGGUUAGUUUCCUA.UUUUGC.U..CUUGU..CCG
Smu-1-30  .CGG.ACGG........CU.GCAAAA..AGAUAAACA.UACCU.UAG.ACGUA......AG.CGU.CG.UUGGUUAGUUUCCUA.UUUUGC.U..AUUAU..CCG
Spn-2-34  .AAA.GUCU........GG.AGAAAA..AGAUGAAU..CUCAU.UGG.UUGCAA.....CCGCAA.CC.UGCUUCGAUUCCCUA.UUUUCU.C.UGGGAC..UUU
Sth-3-7   .GGA.AAGC........UC.GGAAUUU.AGAACAAU..CGUUU.UGG.AUGCU......GG.CAU.CC.UGCAACGAUUGCCUA.AAAUCC.UAGCGCUU..UCU
Spn-2-36  .GAU.ACUG........UG.UGAAAA..AGAUAAAU..UCUCU.UGU.GAGUU......UG.CUU.AC.UUCAAGAAUUUUCUA.UUUUCA.C..UUGGU..AUU
Smu-1-31  .ACG.ACUC........UG.UGAAAA..AGAGAAAUA.CCUCC.UAG.AUACCC.....UGGUAU.CU.UCGUUGGAUUUCUUA.UUUUUC.U.UUGAGU..CG.
Sth-3-8   .UGG.AACA........AG.CCAAAA..AGAUAGAUGGUUUUG.AGG.GACCA......GUAGUC.CA.UCUUCGAAACUAUUA.CUUUGG.C..UUUCU..CCA
Spn-2-38  .UAA.AAGA........UCGGAAAAA..AGAUAAAU..CUCCU.UGG.CUUCA......UCGAAG.CC.AUCGUCGAUUUCCUA.UUUUUCAG.UCGCUU..UUG
Smu-1-32  .AGC.GACU........GU.GCAGAAA.AGACAACC..UUUUU.UAG.AUGCA......GG.UGU.CU.GCGUCAAGUUUCCUU.UUUUGC.U..CUGUC..GUU
Sth-3-9   .GAA.AGCU........UG.GAAAAU..AAACCGUC..CGUCU.ACGCCAGCA......GG.CUGACG.U..GCCGGGCGCCUA.AUUUUC.A.GUCGCU..UUC
Smu-1-33  .UGA.UUGU........CG.AAAAAU..AGAUAGUC..UGCCU.AAG.AUACA......AG.UAU.CU.GUCGUCAGAUUCCUA.AUUUUC.U.GUUCAA..UUA
Spn-2-40  .ACA.GCUUU.......GG.CAAAAA..AGACCAAU..UUGCU.UUG.GAGCA......UUGCUU.CU.GCAUUAAAUUGUCUA.UUUUUG.C.UCGUGC..UGU
Spn-1-19  .ACA.GCUUU.......GG.CAAAAA..AGACCAAU..UUGCU.UUG.GAGCA......UUGCUU.CU.GCAUUCAAUUGUCUA.UUUUUG.C.UCGUGC..UGU
Lla-1-1   .AGA.AGUUG.......AU.GAAAAU..AGAUAAAA..UGGAA.CUU.GGGCU......AUGCCC.AG.AACCGCAUUUUUCUA.AUUUUC.A..CAACU..UCU
Spn-2-42  .CUG.CCUC........UU.GAAAAA..AGAUGCCU..GUCUU.GCC.UUUCG......UGGAAA.GU.CAGUGCCAUUCCCUA.UUUUUC.A...GAGG..CAG
          .<<<.<<<..........<.<<<<<<.......<<<..<<......<.<<<...........>>>.>.......>>>>>......>>>>>>.>....>>>..>>>
          .222.?2?..........?.?22222.......2?2..??......2.2?2...........2?2.2.......??2?2......22222?.?....?2?..222
          -•RR-•YY•--------•R-•RAAAA--AGAUAA•Y--Y•YCY-URR-R••C•------••○••Y-Y•-•Y••Y•RRYUYCCUA-UUUUY•-Y--Y••R•--YY•
```

2.18 The TD-1 *T. denticola* motif

Many of these hits overlap hypothetical short ORFs. BLAST is able to find putative homologs in *T. denticola* with marginal E-values (10^{-2} to 10^{-5}). The TD-1 RNA alignment has many insertions between conserved regions whose sizes are not multiples of 3, so we do not think that TD-1 RNA is fully a coding region.

Notes:

[1] Overlaps the reverse complement of a putative homolog in the TD-2 motif

abbrev. of hits	**taxonomy of species**
Tde-1-1 to Tde-1-29	Bacteria Spirochaetes Spirochaetales Spirochaetaceae *Treponema denticola* ATCC 35405

abbrev	**RefSeq accession**		**5′ at**	**3′ at**	**genes**
[1]Tde-1-1	NC_002967.9	-	72871	72748	**RNA**→ PfkA (COG0205)→
Tde-1-2	NC_002967.9	-	2425623	2425496	**RNA**→ 60KD_IMP (pfam02096)→
Tde-1-3	NC_002967.9	-	1987259	1987128	**RNA**→ DNA polymerase III domain protein→ NtrY (COG5000)→ AtoC (COG2204)→ FtsJ (COG0293)→
Tde-1-4	NC_002967.9	-	2522242	2522121	**RNA**→ hypo→ MA (smart00283)→
Tde-1-5	NC_002967.9	-	2422749	2422628	**RNA**→ ←hypo
[1]Tde-1-6	NC_002967.9	-	2121240	2121118	**RNA**→ hypo→
[1]Tde-1-7	NC_002967.9	-	678502	678381	**RNA**→ UdpNAET (cd01555)→
Tde-1-8	NC_002967.9	+	1560665	1560785	**RNA**→ SalY (COG0577)→ LolE (COG4591)→ ABC_MJ0796_Lo1CDE_FtsE (cd03255)→ COG4939 (COG4939)→ LolE (COG4591)→
Tde-1-9	NC_002967.9	-	1749446	1749325	**RNA**→ hypo→ hypo→
Tde-1-10	NC_002967.9	+	1922427	1922548	**RNA**→ ←hypo
Tde-1-11	NC_002967.9	-	2824639	2824515	**RNA**→ Flg_hook (pfam02120)→ FlgD (COG1843)→ FlgE (COG1749)→ FlbD (pfam06289)→ MotA (COG1291)→ MotB (COG1360)→ FliL (pfam03748)→ FliM (COG1868)→ CheC (COG1776)FliN (COG1886)→ hypo→ FliP (COG1338)→
Tde-1-12	NC_002967.9	-	1361831	1361708	**RNA**→ COG4254 (COG4254)→
Tde-1-13	NC_002967.9	+	779849	779975	**RNA**→ TRX_family (cd02947)→
[1]Tde-1-14	NC_002967.9	+	568504	568630	**RNA**→ hypo→
Tde-1-15	NC_002967.9	-	1631980	1631857	**RNA**→ GlgA (COG0297)→ hypo→ hypo→ COG1293 (COG1293)→
Tde-1-16	NC_002967.9	-	2507946	2507818	**RNA**→ hypo→ COG1512 (COG1512)→
[1]Tde-1-17	NC_002967.9	-	1933322	1933202	**RNA**→ COG1979 (COG1979)→ RuvB (COG2255)→
Tde-1-18	NC_002967.9	-	2040778	2040650	**RNA**→ hypo→ lipoprotein, putative→
[1]Tde-1-19	NC_002967.9	-	2423681	2423553	**RNA**→ Jag (COG1847)→
Tde-1-20	NC_002967.9	-	630085	629955	**RNA**→ AccD (COG0777)→ AccA (COG0825)→
Tde-1-21	NC_002967.9	+	293260	293387	**RNA**→ MarR (pfam01047)→ WrbA (COG0655)COG1146 (COG1146)→ FldA (COG0716)→
Tde-1-22	NC_002967.9	-	2548186	2548056	**RNA**→ hypo→
Tde-1-23	NC_002967.9	-	2444069	2443941	**RNA**→ STE14 (COG2020)→ COG3467 (COG3467)→ hypo→ hypo→ hypo→ hypo→ acetyltransferase, GNAT family→
Tde-1-24	NC_002967.9	-	551054	550941	**RNA**→ hypo→ HipA (COG3550)→
[1]Tde-1-25	NC_002967.9	+	1992272	1992465	**RNA**→ hypo→
Tde-1-26	NC_002967.9	+	2264813	2265000	**RNA**→ hypo→ FcbC (COG0824)→ AdcA (cd01017)→ DUF1361 (pfam07099)→
Tde-1-27	NC_002967.9	+	729883	730072	**RNA**→ ThiD (COG0351)→
Tde-1-28	NC_002967.9	+	1163353	1163547	**RNA**→ hypo→ Chorismate_synt (pfam01264)→
Tde-1-29	NC_002967.9	-	1229988	1229789	**RNA**→ COG1505 (COG1505)→

cd01017 Metal binding protein AcdA.

cd01555 UDP-N-acetylglucosamine enolpyruvyl transferase catalyzes enolpyruvyl transfer as part of the first step in the biosynthesis of peptidoglycan, a component of the bacterial cell wall.

cd02947 TRX family; composed of two groups: Group I, which includes proteins that exclusively encode a TRX domain; and Group II, which are composed of fusion proteins of TRX and additional domains.

cd03255 This family is comprised of MJ0796 ATP-binding cassette, macrolide-specific ABC-type efflux carrier (MacAB), and proteins involved in cell division (FtsE) and release of liporoteins from the cytoplasmic membrane (LolCDE).

COG0205 6-phosphofructokinase [Carbohydrate transport and metabolism]

COG0293 23S rRNA methylase [Translation, ribosomal structure and biogenesis]

COG0297 Glycogen synthase [Carbohydrate transport and metabolism]

COG0351 Hydroxymethylpyrimidine/phosphomethylpyrimidine kinase [Coenzyme metabolism]

COG0577 ABC-type antimicrobial peptide transport system, permease component [Defense mechanisms]

COG0655 Multimeric flavodoxin WrbA [General function prediction only]

COG0716 Flavodoxins [Energy production and conversion]

COG0777 Acetyl-CoA carboxylase beta subunit [Lipid metabolism]

COG0824 Predicted thioesterase [General function prediction only]

COG0825 Acetyl-CoA carboxylase alpha subunit [Lipid metabolism]

COG1146 Ferredoxin [Energy production and conversion]

COG1291 Flagellar motor component [Cell motility and secretion]

COG1293 Predicted RNA-binding protein homologous to eukaryotic snRNP [Transcription]

COG1338 Flagellar biosynthesis pathway, component FliP [Cell motility and secretion / Intracellular trafficking and secretion]

COG1360 Flagellar motor protein [Cell motility and secretion]

COG1505 Serine proteases of the peptidase family S9A [Amino acid transport and metabolism]

COG1512 Beta-propeller domains of methanol dehydrogenase type [General function prediction only]

COG1749 Flagellar hook protein FlgE [Cell motility and secretion]

COG1776 Chemotaxis protein CheC, inhibitor of MCP methylation [Cell motility and secretion / Signal transduction mechanisms]

COG1843 Flagellar hook capping protein [Cell motility and secretion]

COG1847 Predicted RNA-binding protein [General function prediction only]

COG1868 Flagellar motor switch protein [Cell motility and secretion]

COG1886 Flagellar motor switch/type III secretory pathway protein [Cell motility and secretion / Intracellular trafficking and secretion]

COG1979 Uncharacterized oxidoreductases, Fe-dependent alcohol dehydrogenase family [Energy production and conversion]

COG2020 Putative protein-S-isoprenylcysteine methyltransferase [Posttranslational modification, protein turnover, chaperones]

COG2204 Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains [Signal transduction mechanisms]

COG2255 Holliday junction resolvasome, helicase subunit [DNA replication, recombination, and repair]

COG3467 Predicted flavin-nucleotide-binding protein [General function prediction only]

COG3550 Uncharacterized protein related to capsule biosynthesis enzymes [General function prediction only]

COG4254 Uncharacterized protein conserved in bacteria [Function unknown]

COG4591 ABC-type transport system, involved in lipoprotein release, permease component [Cell envelope biogenesis, outer membrane]

COG4939 Major membrane immunogen, membrane-anchored lipoprotein [Function unknown]

COG5000 Signal transduction histidine kinase involved in nitrogen fixation and metabolism regulation [Signal transduction mechanisms]

pfam01047 MarR family.

pfam01264 Chorismate synthase.

pfam02096 60Kd inner membrane protein.

pfam02120 Flagellar hook-length control protein.

pfam03748 Flagellar basal body-associated protein FliL.

pfam06289 Flagellar protein (FlbD).

pfam07099 Protein of unknown function (DUF1361).

smart00283 Methyl-accepting chemotaxis-like domains (chemotaxis sensory transducer).

alignment positions 1 · · · 133

```
1 Tde-1-1    ACUGUUUAUCUU.C.GAGUUUUG......UG.C..UUU..G.CA..CAAAACAUCGCAUU........AUUGUAUGCAGUU.UG..UGAA..........CA.AACUGCUUGAAAAACUUU.UUUCGCAAAU
  Tde-1-2    ACUGUUUAUCUU.C.GAGUUUUG......UG.C..AAA..G.CA..CAAAACAUCGCAUU........AUUGUAUGUAGUUUUGC.AAAA.........GCAAAACUACUUGAAAAACUUU.UUUCGCAAAU
  Tde-1-3    ACUGUUUAUCUG.CCGAGUUUUG......UA.C..AUACGG.UA..CAAAACAUCGCAUU........AUUGUAUGUAGUUUUGCAUAAA.........GCAAAACUACUUGAAAAACUUU.UUUCGGAAAU
  Tde-1-4    ACUGUUUAUCUG.CCGAGUUU.G......C..C..UUUC.G..G..C.AAACGUCGCAUU........AUUGUAUGCAGUU.UG..UAAA..........CA.AACUGCGUGAAAAACUUU.UUUCGCAAAU
  Tde-1-5    ACUGUUUAUCUG.CCGAGUUU.G......C..C..UUUC.G..G..C.AAACGUCGCAUU........AUUGUAUGCAGUU.UG..UAAA..........CA.AACUGCGUGAAAAACUUU.UUUCGCAAAU
1 Tde-1-6    ACUGUUUAUCUG.CCGAGUUU.G......C..C..UUUC.G..G..C.AAACAUCGCAUU........AUUGUAUGCGGUUUG..UAAA..........CA.AACCGCUUGAAAAACUUU.UUUCAGGAUU
1 Tde-1-7    ACUGUUUAUCUG.CCAAGUUU.G......C..C..UUUC.G..G..C.AAACGUCGCAUU........AUUGUAUGCAGUU.UG..UAAA..........CA.AACUGCGUGAAAAACUUU.UUUCAGGAUU
  Tde-1-8    .CUGUUUAGCUGAC.GAGUUU.G......C..C..UUUC.G..G..C.AAACAUCGCAUU........AUUGUAUGCGGUU.UG..UAAA..........CA.AACCGCUUGAAAAACUUU.UUUCGCAAAU
  Tde-1-9    UCUGUUUAUCUG.CCGAGUUUGC......C..U..UUC.....G..GCAAACGUCGCAUU........AUUGUAUGCAGUU.UG..UAAA..........CA.AACUGCGUGAAAAACUUU.UUUCGCAAAU
  Tde-1-10   ACUGUUUAUCUG.CCGAGUUUGC......C..U..UUC.....G..GCAAACGUCGCAUU........AUUGUAUGCAGUU.UG..UAAA..........CA.AACUGCGUGAAAAACUUU.UUUCGCAAAU
  Tde-1-11   .CUGUUUAUCUG.CCGAGUUUUGUA....C..C..AAA..G..GUACAAAACGUCGCAUU........AUUGUAUGCAGUU.UG..UAAA..........CA.AACUGCGUGAAAAACUUU.UUUUGGAAAU
  Tde-1-12   UCUGUUUAUCUU.C.GAGUUUUG......UG.C..AAA..G.CA..CAAAACAUCAUGUG........GUUGUAUGCGGUU.UG..UUUA..........CA.AACCGCUUGAAAAACUUU.UUUCGGAAAU
  Tde-1-13   ACUGUUUAUCUAUC.GAGUUUUGUA....C..C..AAA..G..GUACAAAACAUCGCAUU........AUUAUAUGCGGUU.UG..UAAA..........CA.AACCGCUUGAAAAACUUU.UUUCGGAAAC
1 Tde-1-14   ACUGUUUAUCUGAC.GAGUUUUG......UG.C..UUU..G.CA..CAAAACAUCGCAUU........AUUAUAUGU.GUG.UGCUUUUCG........GCA.CAC.ACUUGAAAAACUUU.UUUCGCAAAU
  Tde-1-15   ACUGUUUAUCUG.C.GAGUUUUG......UG.C..UUU..A.CA..CAAAACAUCGUUU.........AUUGUAUGUAGUUUUGC.UUU..........GCAAAACUACUUGAAAAACUUU.U.UCGCAAAU
  Tde-1-16   ACUGUUCAUCUGAC.GAGUUUUGUA....C..C..AAUA.G..GUACAAAACGUCGCAUU........AUUAUAUGU.GUG.UGC.UUUCG........GCA.CAC.ACUUGAAAAACUUU.UUUUGCAAAU
1 Tde-1-17   ACUGUUUAUCUU.C.GAGUUU.G......C..CGUAAAC.G..G..C.AAACAUCGUUU.........AUCGUAUGCGGUU.UG..UAAA..........CA.AACCGCUUGAAAAACUUU.UUUCGGAAAC
  Tde-1-18   UCUAUUUAUCUG.CCGAGUUUGC......C..U..UUC.....G..GCAAACGUCGCAUU........AUUAUAUGUGCUU.UU..UCACUUCGUUGC..AA.AAGCAC.GGAAAAAACUUU.UUUCGGAAGC
1 Tde-1-19   GCUGUUUAUCUU.C.GAGUUUGCC........U..UUC..A....GGCAAACAUCGUAUU........AUUAUAUGUGCUU.UU..UCACUUC.GUUGC..AAAAGCACG.GAAAAAACUUU.UUUCAUGAUU
  Tde-1-20   ACUGUUUAUCUGAC.GAGUUUUG......UG.C..UUU..G.CA..CAAAACAUCGCAUU........AUUGAUUGCCGUU.UC..UCACAGGUGUUGC.GA.AACGGCGU..AAAAAGUCA.AAUCGAAAAU
  Tde-1-21   CUAAAAACGCUG..CGAGUUUGC.........U..UAC........GCAAACGUCGCUUU........AUUGAUUGCACUU.UU..UCAUUUCAUUGC..AA.AAGUGCGGCAAUUCGCCAAUCCUCGCCUUU
  Tde-1-22   CUAAAAACACUG.C.GAGUUUGC......C..U..UGC..G..G..GCAAACUUCGCUUU........AUUGUAUGCCGUU.UC..UCAC.UUUGUUGC.GA.AACGGCGGAAAAUAAGCAAUGCUCAGAAAC
  Tde-1-23   AAAAAAACGCUG.UCGAGUUUGC......C..U..UUC.....G..GCAAACAUCGCUUU........AUUGUAUGCCGUU.UC..UCACUGACGUUAC.GA.AACGGC..GAAUUGAACAG.CUUCCAAAAU
  Tde-1-24   CUCGUUAACCGU.C.GAGUUUGC.........U..GAC..C.....GCAAACAUCGCAUA........UUACAAUGUGCGC.GU..U.............AC.GCGCACCA.AUGCAACAGU.UUUCGAGAAU
1 Tde-1-25   ACUGUUUAUCUG.CCGAGUUUGC......C..U..UUC.....G..GCAAACGUCGCAUU........AUUAUAUAUGCUU.UU..UCACUUU.GUUGC.AA.AAGCACGGAAAAAACUUUU..UUCGCAAAU
  Tde-1-26   ACUGUUUAUCUG.CCGAGUUUGC......C..U..UUCA....G..GCAAACAUCGCAUU........AUUGUAUGCAGUU.UG..UAAA..........CA.AACUGCGUGAAAAACUUU.UUUCGCAAGC
  Tde-1-27   ACUGUUUAUCUAUC.GAGUUUUG......UG.C..GUU..G.CA..CAAAACAUCGUAUU........AUUGUAUGCGGUU.UG..UAAA..........CA.AACCGCUUGAAAAACUUU.UUUUGCAAAU
  Tde-1-28   ACUGUUUAUCUG.CCGAGUUUUG..C...ACUU..AUG...AGUA.CAAAACUUCGCAUU........AUUAUAUGUGUGU.GC..UUUUCG........GC.ACACACUUGAAAAACUUU.UUUCACAAAU
  Tde-1-29   ...UUUUUUUUA.C.UAUUUUUUUUAAACU..A..CUU..G..AAAAAAAACAUAGUUUCUUCUAUAUAUUAUAUGCGGUU.UG..UAAA..........CA.AACCGCUUGAAAAACUUU.UUUCAAGAUU
             ...............<<<<<<<<<<....<<<.........>>>>>>>>>>>.>>....................<<<<<<<<<<..............>>>>>>>>>>....<<<<<<<<<.<<..<<<<<<
             ...............200000222?....220.........022?2220000.02....................022222?220..............022?222220....22?22?222.2?..2?22?2
             .....................................................................................................................................
             .....................................................................................................................................
             ●Y●RUU●●●CUR-C-GAGUUU○●------○--Y--●●Y--○--○--●○AAACRUCGC●UU--------AUURUAUGY●●UU-U●--U●●○----------●A-AA●●RC○○RAAR●AACY●●-YUUCR●RARU
```

alignment positions 134 · · · 228

```
[1]Tde-1-1   UGAU.................................................................AUUUGCUGCAAAAAGUUUUUAUGGGA
   Tde-1-2   UGAU.................................................................AUUUGCUGCAAAAAGUUUUUAUUGGA
   Tde-1-3   UGAU.................................................................AUUUCCUGCAAAAAGUUUUUAUAGGA
   Tde-1-4   UGAU.................................................................AUUUGCUGCAAAAAGUUUUUAUGGGA
   Tde-1-5   UGCU.................................................................AUUUGCUACAAAAAGUUUUUAUGGGA
[1]Tde-1-6   UGAU.................................................................AAUCCUUGCAAAAAGUUUUUAUGGGA
[1]Tde-1-7   UGAU.................................................................AAUCCUUGCAAAAAGUUUUUAUAGGA
   Tde-1-8   UAAU.................................................................AUUUGCUGCAAAAAGUUUUUAUAGGA
   Tde-1-9   UGCU.................................................................AUUUGCUGCAAAAAGUUUUUAUUGGA
   Tde-1-10  UGAU.................................................................AUUUGUUGCAAAAAUUUUUUAUAGGA
   Tde-1-11  UGAU.................................................................AUUUCCUGCAAAAAGUUUUUAUG.GA
   Tde-1-12  UGAU.................................................................AUUUCCUGCAAAAAGUUUUUAUAGGA
   Tde-1-13  UGAG.................................................................GUUUCCUGCAAAAAGUCUUUAUAGGA
[1]Tde-1-14  UGAU.................................................................AUUUGCUGCAAAAAGUUUUUAUAGGA
   Tde-1-15  UAAU.................................................................AUUUGCUGC.AAAAGUUUUUAUAGGG
   Tde-1-16  UAAU.................................................................AUUUGCUGCAAAAAGUUUUUAUUGGG
[1]Tde-1-17  UAAU.................................................................GUUUCCUGCAAAAAGUUUUUAGA.GA
   Tde-1-18  UGAG.................................................................GCUUCCUGCAAAAAGUUUUUAUGGGA
[1]Tde-1-19  UGAU.................................................................AAUCCUUGCAAAAAGUUUUUAUGGGA
   Tde-1-20  UGAU.................................................................A.UUUUUUCAUUGACUUUUUAUAGGA
   Tde-1-21  UGAU.................................................................AAAGGCUGCGGUUGUCGAAUUUUAAG
   Tde-1-22  UGAU.................................................................GUUUCUUCCGCUUGCUUAUUUUAAUA
   Tde-1-23  UGAU.................................................................AUUUUGUUCAGCUGUUCAAUUUUAAG
   Tde-1-24  UGAU.................................................................AUUCUCU..CAACUGUUCCAUUUUAA
[1]Tde-1-25  UUAUUUGCUAUCCGCUUUCGCGGAUACGAUGAAAAAUUUCCUUACAGAACAAAGCCGAAAGGCUCUGAUAUUUGCUGCAAAAAGUUUUUAUAGGG
   Tde-1-26  UGAUUCGCUAUCCGCUUUAGCGGAUACGAUGAAAAAUUUCCUUACAGAACAGAGCCGAAAGGCUCUGAUAUUUGCUGCAAAAAGUUUUUAUAGGA
   Tde-1-27  UGAUUCGCUAUCCGCUUUAGCGGAUACGAUAAAAAAUUUCCUUGCAGAACAGAACCGCAAGGGUCUGAUAUUUGCUACAAAAAGUUUUUAUAGGA
   Tde-1-28  UGAUUCGCUAUCCGCUUUAGCGGAUACGAUGAAAAAUUUCCUUACAGAACAGAGCCGAAAGGCUCUGAUAUUUGCUGCAAAAAGUUUUUAUAGGA
   Tde-1-29  UGAUUCGCUAUCCGCUUUAGCGGAUACGAUGAAAAAGUUCCUUGCAGAACAUAGCCGCAAGGCUCUGAUAAUCCUUGCAAAAAGUUUUUAUAGGA
             ....<<<.<<<<<<<....>>>>>>>>>>.........<<<<<<<...........>>>>>>>......>>>>>>...>>>>>>>>>>>......
             ....010.0000000....0000000010.........?0000?0...........0?0000?......2?22?2...?2222?22?22......
             .................................................<<.<<<<....>>>>.>>............................
             .................................................00.0?00....00?0.00............................
             UGAU-----------------------------------------------------------------A•UY•YU•CRR••R•UUYYU•U•RRR
```

2.19 The TD-2 *T. denticola* motif

Notes:

---

[1] Homology is moderately good on 3′-most hairpin, but unclear elsewhere.

[2] 30 nucleotides of the RNA’s 5′ end overlap a small hypothetical gene, but there is no evidence that this protein is conserved in any bacteria.

[3] Overlaps a hypothetical gene (not shown), which may have distant homology in *Treponema denticola*.

[4] Overlaps a hypothetical gene that does not appear to be conserved in any bacteria.

[5] Overlaps a small hypothetical gene. The conserved part of this gene does not overlap.

[6] Overlaps the reverse complement of a putative homolog in the TD-1 motif

abbrev. of hits	taxonomy of species
Tde-1-1 to Tde-1-36	Bacteria Spirochaetes Spirochaetales Spirochaetaceae *Treponema denticola* ATCC 35405

abbrev	RefSeq accession		5′ at	3′ at	genes
[6]Tde-1-1	NC_002967.9	+	72799	72905	hypo→ ←PfkA (COG0205) **RNA**→ ←DUF192 (pfam02643) ←hypo
[6]Tde-1-2	NC_002967.9	+	678433	678536	←Cache (pfam02743)MA (smart00283)HAMP (smart00304) ←UdpNAET (cd01555) **RNA**→ ←E_set_proteins_like (cd02861) ←PTH (cd00462)
[3,6]Tde-1-3	NC_002967.9	+	2121169	2121274	←hypo ←hypo **RNA**→ ←hypo AmpC (COG1680)→
[3,6]Tde-1-4	NC_002967.9	-	1992344	1992238	hypo→ hypo→ **RNA**→ hypo→ hypo→
[6]Tde-1-5	NC_002967.9	+	2423609	2423715	←hypo ←Jag (COG1847) **RNA**→ ←60KD_IMP (pfam02096) ←hypo
[6]Tde-1-6	NC_002967.9	-	568579	568470	hypo→ ←hypo **RNA**→ PnbA (COG2272)→ SMI1 (COG4282)→
[6]Tde-1-7	NC_002967.9	+	1933247	1933356	←RuvB (COG2255) ←COG1979 (COG1979) **RNA**→ ←GlpK (COG0554) ←SpsF (COG1861)UbiG (COG2227)spsG (COG3980)
Tde-1-8	NC_002967.9	-	1000082	999985	←DUF323 (pfam03781) ←hypo **RNA**→ ←hypo ←hypo
Tde-1-9	NC_002967.9	-	599658	599561	←hypo ←hypo **RNA**→ ←hypo ←hypo
Tde-1-10	NC_002967.9	-	1953077	1952980	←hypo ←hypo **RNA**→ hypo→ hypo→
Tde-1-11	NC_002967.9	+	843599	843700	←MdlB (COG1132) ←MdlB (COG1132) **RNA**→ ←AcrR (COG1309) hypo→
Tde-1-12	NC_002967.9	-	1911156	1911055	←hypo ←hypo **RNA**→ Branch_AA_trans (pfam05525)→ DUF583 (pfam04519)→
Tde-1-13	NC_002967.9	+	1129831	1129931	COG3330 (COG3330)→ COG1543 (COG1543)→ **RNA**→ Ala_racemase (cd00430)→ LYTB (pfam02401)→
Tde-1-14	NC_002967.9	+	1075188	1075287	Ribosomal_S12 (cd03368)→ RpsG (COG0049)→ **RNA**→ FusA (COG0480)→ hypo→
Tde-1-15	NC_002967.9	-	2164499	2164393	←hypo ←hypo **RNA**→ CbiQ (pfam02361)→ ABC_cobalt_transport_domain1 (cd03225)→
[2]Tde-1-16	NC_002967.9	+	2573259	2573364	←COG1284 (COG1284) hypo→ **RNA**→ Cache (pfam02743)MA (smart00283)HAMP (smart00304)→ ←COG1451 (COG1451)
[4]Tde-1-17	NC_002967.9	+	2276539	2276644	←COG4865 (COG4865) ←hypo **RNA**→ ←DUF323 (pfam03781) ←hypo
Tde-1-18	NC_002967.9	-	1086059	1085954	hypo→ hypo→ **RNA**→ ←DUF323 (pfam03781) NorM (COG0534)→
[4]Tde-1-19	NC_002967.9	+	2125442	2125546	←hypo hypo→ **RNA**→ ←DUF323 (pfam03781) ←hypo
Tde-1-20	NC_002967.9	+	2374805	2374908	Peptidase_M29 (pfam02073)→ ←hypo **RNA**→ LeuS (COG0495)→ ←COG3976 (COG3976)
Tde-1-21	NC_002967.9	+	854281	854383	hypo→ DUF323 (pfam03781)→ **RNA**→ ←RnfA (COG4657) ←RnfE (COG4660)
[5]Tde-1-22	NC_002967.9	+	1941707	1941805	←hypo ←hypo **RNA**→ ←hypo SlpA (COG1047)→
Tde-1-23	NC_002967.9	+	1084708	1084807	MA (smart00283)→ ←NorM (COG0534) **RNA**→ DUF323 (pfam03781)→ ←hypo
Tde-1-24	NC_002967.9	-	2302899	2302794	HRDC (pfam00570)→ RVT (pfam00078)→ **RNA**→ ←hypo MdlB (COG1132)→
Tde-1-25	NC_002967.9	+	994718	994821	←EFP (pfam01132) Prc (COG0793)→ **RNA**→ hypo→ ArgS (COG0018)→
Tde-1-26	NC_002967.9	-	2164591	2164489	←hypo ←hypo **RNA**→ CbiQ (pfam02361)→ ABC_cobalt_transport_domain1 (cd03225)→
[4]Tde-1-27	NC_002967.9	-	1952896	1952794	←hypo ←hypo **RNA**→ hypo→ hypo→
Tde-1-28	NC_002967.9	-	1952985	1952892	←hypo ←hypo **RNA**→ hypo→ hypo→
Tde-1-29	NC_002967.9	-	1875793	1875692	hypo→ ←hypo **RNA**→ hypo→ hypo→
Tde-1-30	NC_002967.9	+	1660531	1660631	←VapI (COG3093) Mug (COG3663)→ **RNA**→ Upp (COG0035)→ TPR (cd00189)→
Tde-1-31	NC_002967.9	+	2540940	2541042	GidB (COG0357)→ ←COG2129 (COG2129) **RNA**→ ←DUF323 (pfam03781) hypo→
[3]Tde-1-32	NC_002967.9	-	968502	968400	hypo→ FusA (COG0480)→ **RNA**→ ←DUF323 (pfam03781) ←hypo
Tde-1-33	NC_002967.9	-	2512716	2512618	COG1573 (COG1573)→ PriA (COG1198)→ **RNA**→ PrfA (COG0216)→ HemK (COG2890)→
Tde-1-34	NC_002967.9	+	2770519	2770618	hypo→ ←SpoU (COG0566) **RNA**→ ←DUF323 (pfam03781) hypo→
Tde-1-35	NC_002967.9	-	2516687	2516588	←hypo DUF323 (pfam03781)→ **RNA**→ hypo→ COG1573 (COG1573)→
[1]Tde-1-36	NC_002967.9	+	825321	825421	←COG1959 (COG1959) ←COG2910 (COG2910) **RNA**→ hypo→ ←UPF0027 (pfam01139)

cd00189 Tetratricopeptide repeat domain; typically contains 34 amino acids [WLF]-X(2)-[LIM]-[GAS]-X(2)-[YLF]-X(8)-[ASE]-X(3)-[FYL]-X(2)-[ASL]-X(4)-[PKE]
cd00430 Alanine racemase.
cd00462 Peptidyl-tRNA hydrolase (PTH) is a monomeric protein that cleaves the ester bond linking the nascent peptide and tRNA when peptidyl-tRNA is released prematurely from the ribosome.
cd01555 UDP-N-acetylglucosamine enolpyruvyl transferase catalyzes enolpyruvyl transfer as part of the first step in the biosynthesis of peptidoglycan, a component of the bacterial cell wall.
cd02861 E or "early" set-like proteins.
cd03225 Domain I of the ATPase component of a cobalt transport family found in both bacteria and archaea.
cd03368 S12-like family, 30S ribosomal protein S12 subfamily; S12 is located at the interface of the large and small ribosomal subunits of prokaryotes, chloroplasts and mitochondria, where it plays an important role in both tRNA and ribosomal subunit interactions.
COG0018 Arginyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
COG0035 Uracil phosphoribosyltransferase [Nucleotide transport and metabolism]
COG0049 Ribosomal protein S7 [Translation, ribosomal structure and biogenesis]
COG0205 6-phosphofructokinase [Carbohydrate transport and metabolism]
COG0216 Protein chain release factor A [Translation, ribosomal structure and biogenesis]
COG0357 Predicted S-adenosylmethionine-dependent methyltransferase involved in bacterial cell division [Cell envelope biogenesis, outer membrane]
COG0480 Translation elongation factors (GTPases) [Translation, ribosomal structure and biogenesis]
COG0495 Leucyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]
COG0534 Na+-driven multidrug efflux pump [Defense mechanisms]
COG0554 Glycerol kinase [Energy production and conversion]
COG0566 rRNA methylases [Translation, ribosomal structure and biogenesis]
COG0793 Periplasmic protease [Cell envelope biogenesis, outer membrane]
COG1047 FKBP-type peptidyl-prolyl cis-trans isomerases 2 [Posttranslational modification, protein turnover, chaperones]
COG1132 ABC-type multidrug transport system, ATPase and permease components [Defense mechanisms]
COG1198 Primosomal protein N' (replication factor Y) - superfamily II helicase [DNA replication, recombination, and repair]
COG1284 Uncharacterized conserved protein [Function unknown]
COG1309 Transcriptional regulator [Transcription]
COG1451 Predicted metal-dependent hydrolase [General function prediction only]
COG1543 Uncharacterized conserved protein [Function unknown]
COG1573 Uracil-DNA glycosylase [DNA replication, recombination, and repair]
COG1680 Beta-lactamase class C and other penicillin binding proteins [Defense mechanisms]
COG1847 Predicted RNA-binding protein [General function prediction only]
COG1861 Spore coat polysaccharide biosynthesis protein F, CMP-KDO synthetase homolog [Cell envelope biogenesis, outer membrane]
COG1959 Predicted transcriptional regulator [Transcription]
COG1979 Uncharacterized oxidoreductases, Fe-dependent alcohol dehydrogenase family [Energy production and conversion]
COG2129 Predicted phosphoesterases, related to the Icc protein [General function prediction only]
COG2227 2-polyprenyl-3-methyl-5-hydroxy-6-metoxy-1,4-benzoquinol methylase [Coenzyme metabolism]
COG2255 Holliday junction resolvasome, helicase subunit [DNA replication, recombination, and repair]
COG2272 Carboxylesterase type B [Lipid metabolism]
COG2890 Methylase of polypeptide chain release factors [Translation, ribosomal structure and biogenesis]
COG2910 Putative NADH-flavin reductase [General function prediction only]
COG3093 Plasmid maintenance system antidote protein [General function prediction only]
COG3330 Uncharacterized protein conserved in bacteria [Function unknown]
COG3663 G:T/U mismatch-specific DNA glycosylase [DNA replication, recombination, and repair]
COG3976 Uncharacterized protein conserved in bacteria [Function unknown]
COG3980 Spore coat polysaccharide biosynthesis protein, predicted glycosyltransferase [Cell envelope biogenesis, outer membrane]
COG4282 Protein involved in beta-1,3-glucan synthesis [Carbohydrate transport and metabolism]
COG4657 Predicted NADH:ubiquinone oxidoreductase, subunit RnfA [Energy production and conversion]
COG4660 Predicted NADH:ubiquinone oxidoreductase, subunit RnfE [Energy production and conversion]
COG4865 Glutamate mutase epsilon subunit [Amino acid transport and metabolism]
pfam00078 Reverse transcriptase (RNA-dependent DNA polymerase).
pfam00570 HRDC domain.
pfam01132 Elongation factor P (EF-P).
pfam01139 Uncharacterized protein family UPF0027.
pfam02073 Thermophilic metalloprotease (M29).
pfam02096 60Kd inner membrane protein.

pfam02361 Cobalt transport protein.
pfam02401 LytB protein.
pfam02643 Uncharacterized ACR, COG1430.
pfam02743 Cache domain.
pfam03781 Domain of unknown function (DUF323).
pfam04519 Protein of unknown function, DUF583.
pfam05525 Branched-chain amino acid transport protein.
smart00283 Methyl-accepting chemotaxis-like domains (chemotaxis sensory transducer).
smart00304 HAMP (Histidine kinases, Adenylyl cyclases, Methyl binding proteins, Phosphatases) domain;

alignment positions 1 · · · 134

```
6   Tde-1-1    AAGCAGUUUGU.UCA.CAAACUGCAUACA...A..UAAUGCGAUGUUU.UGUG....CAAAG......CACAAA.AC....UCGAA.GAUAAACAGUGAGG.CUGAAUUUCUG.CCGAACUGUUUA.UCUUACC
6   Tde-1-2    .CGCAGUUUG..UUUACAAACUGCAUACA...A..UAAUGCGACGUUU.GCC.....GAAA........GGCAA.AC....UUGGCAGAUAAACAGUGAGG.CUGAAUUCCUU.CCGAACUGUUUA.UCAUAUC
3,6 Tde-1-3    AAGCGGUUUGU.UUACAAAACCGCAUACA...A..UAAUGCGAUGUUU.GCC.....GAAA........GGCAA.AC....UCGGCAGAUAAACAGUGAGG.CUGAAUUCCUG.CCGAACUGUUUA.UCAAACC
3,6 Tde-1-4    UUUUGCAACAAAGUGAAAAAGCAUAUAUA...A..UAAUGCGACGUUU.GCC.....GAAA........GGCAA.AC....UCGGCAGAUAAACAGUGAGG.CUGAAUUCCUG.CCGAACUGUUUA.UCAUACC
6   Tde-1-5    UUUUGCAACGAAGUGAAAAAGCACAUAUA...A..UAAUACGAUGUUU.GCCU....GAA........AGGCAA.AC....UCGAA.GAUAAACAGCGAGGC.AGAAUUCCUG.CCGAACUGUUUA.UCGUAUC
6   Tde-1-6    .AAGUGUGUGCCGAAAAGCACACACAUAU...AA.UAAUGCGAUGUUU.UGUG....CAAAG......CACAAA.AC....UCGUCAGAUAAACAGUGGGG.CUGAAUUUCUG.CCGAACUGUUUA.UCAUAUC
6   Tde-1-7    UUUUCAAGCGGUUUGUUUACAAACCGCAU...ACGAUAAACGAUGUUU.GCCG....UUUA.......CGGCAA.AC....UCGAA.GAUAAACAGUGAGGC.UGAAUUCCUG.CCGAGCUGUUUA.UCGUACC
    Tde-1-8    CAGUACCGCCA.GGAUGGCGGUGGCUCCA...AGCAAAAUCGA.GUUC.GC......ACU..........GCGA.AC....UCGCA.GCUAAAGAAUGAAC.ACGGAC.....GUUCAUUCUUUA.GCGUGCC
    Tde-1-9    CAGUACCGCCA.GGAUGGCGGUGGCUCCA...AGCAGAAUCGA.GUUC.GC......ACU..........GCGA.AC....UCGCA.GCUAAAGAAUGAAC.ACGGAC.....GUUCAUUCUUUA.GCGUGCC
    Tde-1-10   CUUCAGCGGCACGGAUGCCGCUGGUUCCA...CU.CCUUACGA.GUUU.GC......AAC..........GCAA.AC....UCGAA.GCUGAAAAAUGUAC.ACGGAU.....GUCCGAUCAUAA.GCCGGAC
    Tde-1-11   CAGAACCGCCACGGAUGGCGGUAGUGCCA...AGCAGUAGCAA.GUUU.UUC.....GUUA........GAAAA.AC....UCGUC.GUUGAAAAGUGUAC.ACGGAG.....GUACGCUUUUCA.ACACACC
    Tde-1-12   CCCAACCGCCAUGGACGGCGGUGGUUUCA...AACUUC.GAGA.GUUU.GCG.....CUUGC.......CGCAA.AC....UCUAA.GAUUAAAAAUGUGC.AUGGAC.....GCACAUUUUUAA.UCCACCC
    Tde-1-13   CCCAACCGCCAUGGAAGGCGGUGGUUCCA...UACUUC.GAGA.GUUU.UGC.....UUUA........GCAAA.AC....UCUAA.GAUGAAAAAUGUAC.ACGGAU.....GUACAUUUUUCA.UCGGUCU
    Tde-1-14   AUAAAUGGACACGGAUGUCCAUGGUUCCA...A..ACUUGAGA.GUUU.UGC.....GUAA........GCAAA.AC....UCUAA.GAUUAAAAAUGUAC.AAGGAU.....GUACAUUUUUAA.UCGACCC
    Tde-1-15   AUAAACGGACACGGAUGUCCGUGGUUCCA...AACUUCCAAGA.GUUU.UAGAC...GCAA......GUCUAAA.AC....UCUAA.GAUUAAAAAUGUACUCAGGAU.....GUACAUUUUUAA.UCGCCCC
2   Tde-1-16   AUUAACCGGCACGGAUGCCGGUGGUUCCA...AACUUCCAAGA.GUUU.UAGAC...GCAA......GUCUAAA.AC....UCUAA.GCUGAAAAAUGUAC.AAGGAU.....GUACAUUUUUCA.GCGGAUG
4   Tde-1-17   CUUUACCGCCAUGGACGGCGGUGGUUCCA...AACUUCCAAGA.GUUU.UAGAC...GCAA......GUCUAAA.AC....UCUAA.GAUGAAAAAUGUAC.AAGGAU.....GUAUAUUUUUAA.GCGGCUC
    Tde-1-18   CAACACCGCCAUGGACGGCGGUGGUUCCA...AACUUCCAAGA.GUUU.UAGAC...ACAA......GUCUAAA.AC....UCUAA.GAUGAAAAAUGUAC.ACGGAU.....GUACAUUUUUCA.UCGACUC
4   Tde-1-19   CUUUACCGCCAUGGACGGCAGUGGUUCCA...AACUUCCAGGA.GUUU.UAGAC...GCAA......GUCUAAA.AC....UCUAA.GAUGAAAAAUGUAC..AGGAC.....GUACAUUCUUCA.ACCGGCA
    Tde-1-20   CCCAACCGGCACGGAUGCCGGUGUUUCCA...G..CCUUGAGA.GUUU.UAGAC...GCAA......GUCUAAA.AC....UCUAA.GAUGAAAAAUGUAC.AUGGAA.....GUACAUUUUUCA.UCGGCAC
    Tde-1-21   CACAACCGGCACGGAUGCCGGUGGUUUAA...A..CUU.ACGA.GUUU.UUUGC...GAA.......GCAAAAAUAC....UCGAA.GCUGAAAAAUGUAC.AAGGAC.....GUACAUUUUUCA.GCGGUUA
5   Tde-1-22   GCAUACCGCUAUGGAUGGCGGUGGUUCCU...GC..UUCGAGA.GUUU.UGC.....ACA.........GCAAA.AC....UCUUA.GAUUAAAAAUGUAC.AAGGAG.....GUACAUUUUUAA.UCAAUCG
    Tde-1-23   UAAAACCGCCACGGAUGGCGGUGGUUAAG...A..CUCUUAGA.GUUU.UGC.....UUAC........GCAAA.AC....UCUAA.GAUCAAAAAUGUGC.GUCCAU.....GCACAUUUUUGA.UCGCUCG
    Tde-1-24   CAAAGCGGACA.GGAUGUCCGCCUGUAUA...AGCUGU.ACGA.GUUU.UUUGC...GUAA......GCAAAAACAC....UCGAA.GAUGAAAAAUGUACUUUGGAU.....GUACAUUUUUCA.UCGGUCA
    Tde-1-25   GGGAGCGAACA.GGAUGUUCGCUGAUUUA...A..UCUUACGAUGUUU.UAGAC...GUAA......GUCUAAA.AC....UCGAAAGCUAAAAAAUGUAC..AGGAU.....GUACAUUUUUUA.GCUCCCC
    Tde-1-26   CCCAACCGUCACGGAUGGCGGCAGUUCCA...C..GCUCGCAA.GUUU.UAGCC...UGC.......GGCUAAA.AC....UUGAA.GAUUAAAAAUGUAC.AAGAAU.....GUACAUUCAUAA.ACGGACA
4   Tde-1-27   UAAGCCGGACA.GGAUGUCCGGUGGUUCC...AU.UCUCACGAUGUUU.UGCC....GUAA.......GGCAAA.AC....UCGAA.GCUGAAAAAUGUAC.ACGGAU.....GUACAUUUUUCA.GCGAGUA
    Tde-1-28   ....CCGGACA.GGAUGUCCGGUGGUUCC...AU.UCUCGCGAUGUUU.UGCC....GUAA.......GGCAAA.AC....UCGAA.GCUGAAAAAUGUAC.ACGGAU.....GUCCGAUCAUAA.GC.....
    Tde-1-29   UGCGCCCGCCACGGAUGGCGGGUGGUUUC...AGCUUA.AAGA.GUUU.GUG.....CUUGU.......CACAA.AC....UCGAA.GAUCAAAAAUAUAU.AAGGAU.....GUAUAUUUUUGA.UCGAGUA
    Tde-1-30   AAUAGCGGACAUGGAUGUCCGUGGUUCCA...A..CUUCAAGA.GUUU.GCG.....CUUGC.......CACAA.AC....UCUAA.GCUGAAAAUCGGAC.ACGGAU.....GUCCGAUUUUCA.GCGAACU
    Tde-1-31   UUUUACCGCCAUGGACGGCGGUGGUUCCA...U..UCUUGCGAUGUUU.UGCC....GUGA.......GGCAAA.AC....UCGAA.ACUCAAAAUCGGAC.ACAGAU.....GUCCGAUUUUGA.GUACCUC
3   Tde-1-32   CUUUACCGCCACGGACGGCGGUGGUUACA...U..UCUUGCGAUGUUU.UGCC....GAAA.......GGCAAA.AC....UCGAA.ACUCAAAAUCGGAC.AAGGAC.....GUCCGAUUUUGA.GUACCUC
    Tde-1-33   ......GUUUUUUGGCUGCCGGUGGUUUC...AC.GCUAGCGACGUUU.UGCC....GUGA.......GGCAAA.AC....UUGAA.GCUUAAAAAAUGUA.CCUCUCUG....UACAUUUUUUAA.GCGGUA.
    Tde-1-34   CCUGUCACGGACGGCAAAAAAACGUCAU...AAGUCUUGCGA.GUUU.GGC.....GAA.........GACAA.AC....UCGG..UAUGAAAAUUGCAC.ACGGAG.....GUGCAAUUCCCA.UACCGAC
    Tde-1-35   AAUCGACCUCC.UGUCGAUGUUUGAGCCU...UACUUCCGCGAUGUUU.UGCC....GUCA.......GGCAAA.AC....UCGAA.GUUCAAUUCUGUC..ACA..........GGCAGAAUUGAUACAAACU
1   Tde-1-36   UUUGAGAUUUAUAUAUAGUUCUUUUUUAUUUACAAAUUAUGA.AUUA.C.......ACGG..........GU..AU...GUCGAA.GUUGAACAUUGGAC.ACUGAU.....CUCCGAUUUUCA.ACGGUAG
               ....<<<<<<......>>>>>>..................<<<.<<<<.<<<<<<<<.......>>>>>>>>>>.>>....>>>...<<<<<<<<<<<<<<.............>>>>>>>>>>>>.>>.....
               ....??????......??????..................220.20?2.2?220???.......???022?22?.02....022...2?0????2?2??2?.............?2??2?2????0.?2.....
               ●●●●R●●●●YA○RGRYR●Y●RY●●●UYY●---A○-●●●●R●GA-GUUU-●●●○----●●●●-------○●●●AA-AC----UC●RA-G●U●AA●ARUG●R●-●●GRAY------●Y●YR●UYUU●A-●CR●●Y●
```

2.20 gamma-150

We presume that gamma-150, if a genuine structured RNA, is a non-coding RNA, because the gene context is not typical of known *cis*-regulatory RNAs.

Notes:

[1] Genome annotation has overlapping gene, but this gene is hypothetical and does not have homologs in any other bacteria, according to BLAST.

[2] This gene is ≈ 1.3 Kb from its downstream gene, according to the annotation. However, there appears to be an unannotated Isovaleryl-CoA dehydrogenase gene in this reason. (A translated BLAST query of this region returns a match to this gene, and includes the start codon, and no apparent premature stop.)

abbrev. of hits	**taxonomy of species**
Avi-1-1 to Avi-1-2	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Azotobacter vinelandii* AvOP
Pae-1-1 to Pae-1-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* 2192
Pae-2-1 to Pae-2-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* C3719
Pae-3-1 to Pae-3-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PA7
Pae-4-1 to Pae-4-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PACS2
Pae-5-1 to Pae-5-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* PAO1
Pae-6-1 to Pae-6-1	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas aeruginosa* UCBPP-PA14
Pen-1-1 to Pen-1-2	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas entomophila* L48
Pfl-1-1 to Pfl-1-2	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* Pf-5
Pfl-2-1 to Pfl-2-2	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas fluorescens* PfO-1
Ppu-1-1 to Ppu-1-2	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* F1
Ppu-2-1 to Ppu-2-2	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas putida* KT2440
Psy-1-1 to Psy-1-3	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. phaseolicola* 1448A
Psy-2-1 to Psy-2-3	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. syringae* B728a
Psy-3-1 to Psy-3-3	γ-proteobacteria Pseudomonadales Pseudomonadaceae *Pseudomonas syringae pv. tomato str.* DC3000

abbrev	RefSeq accession		5′ at	3′ at	genes
Psy-2-1	NC_007005.1	-	940989	940819	PutP (COG0591)BaeS (COG0642)→ AtoC (COG2204)→ **RNA**→ PcnB (COG0617)→ FolK (COG0801)→
Psy-1-1	NC_005773.3	-	1024846	1024676	PutP (COG0591)BaeS (COG0642)→ AtoC (COG2204)→ **RNA**→ PcnB (COG0617)→ FolK (COG0801)→
Psy-3-1	NC_004578.1	-	1046817	1046647	PutP (COG0591)BaeS (COG0642)PAS (pfam00989)→ AtoC (COG2204)→ **RNA**→ PcnB (COG0617)→ FolK (COG0801)→
Pfl-2-1	NC_007492.1	+	5425361	5425532	PutP (COG0591)BaeS (COG0642)→ AtoC (COG2204)→ **RNA**→ PcnB (COG0617)→ FolK (COG0801)→
Pfl-1-1	NC_004129.6	+	6041094	6041265	PutP (COG0591)BaeS (COG0642)PAS (pfam00989)→ AtoC (COG2204)→ **RNA**→ PcnB (COG0617)→ FolK (COG0801)→
Ppu-2-1	NC_002947.3	+	5338409	5338585	PutP (COG0591)COG4191 (COG4191)→ AtoC (COG2204)→ **RNA**→ PcnB (COG0617)→ FolK (COG0801)→
Ppu-1-1	NZ_AALM01000108.1	-	6023	5847	**RNA**→ PcnB (COG0617)→ FolK (COG0801)→
Pen-1-1	NC_008027.1	+	5029985	5030162	PutP (COG0591)COG4191 (COG4191)→ AtoC (COG2204)→ **RNA**→ PcnB (COG0617)→ FolK (COG0801)→
Avi-1-1	NZ_AAAU03000017.1	+	11677	11852	PutP (COG0591)NtrY (COG5000)→ AtoC (COG2204)→ **RNA**→ PcnB (COG0617)→ FolK (COG0801)→
Pae-5-1	NC_002516.2	+	5308721	5308901	PutP (COG0591)COG4191 (COG4191)→ AtoC (COG2204)→ **RNA**→ PcnB (COG0617)→ FolK (COG0801)→
Pae-2-1	NZ_AAKV01000104.1	-	66034	65854	HATPase_c (smart00387)→ AtoC (COG2204)→ **RNA**→ PcnB (COG0617)→ FolK (COG0801)→
Pae-4-1	NZ_AAQW01000001.1	+	5519005	5519185	PutP (COG0591)COG4191 (COG4191)→ AtoC (COG2204)→ **RNA**→ PcnB (COG0617)→ FolK (COG0801)→
Pae-6-1	NZ_AABQ07000004.1	+	39642	39822	PutP (COG0591)COG4191 (COG4191)→ AtoC (COG2204)→ **RNA**→ PcnB (COG0617)→ FolK (COG0801)→
Pae-1-1	NZ_AAKW01000066.1	-	195081	194901	AtoC (COG2204)→ hypo→ **RNA**→ PcnB (COG0617)→ FolK (COG0801)→
Pae-3-1	NZ_AAQE01000003.1	+	24552	24732	PutP (COG0591)COG4191 (COG4191)→ AtoC (COG2204)→ **RNA**→ PcnB (COG0617)→ FolK (COG0801)→
Psy-1-2	NC_005773.3	-	1738748	1738566	←DUF485 (pfam04341) ←ProX (COG2113) **RNA**→ hypo→ VacJ (pfam04333)→
Psy-2-2	NC_007005.1	+	4476246	4476428	←DUF485 (pfam04341) ←ProX (COG2113) **RNA**→ hypo→ VacJ (pfam04333)→
Psy-3-2	NC_004578.1	-	1778604	1778422	←DUF485 (pfam04341) ←ProX (COG2113) **RNA**→ hypo→ VacJ (pfam04333)→
Pfl-2-2	NC_007492.1	+	4136134	4136304	HTH_CueR (cd01108)→ LeuA (COG0119)→ **RNA**→ ←CaiC (COG0318) IVD (cd01156)→
Pfl-1-2	NC_004129.6	+	4546343	4546513	HTH_MERR (smart00422)→ LeuA (COG0119)→ **RNA**→ ←CaiC (COG0318) IVD (cd01156)→
Psy-2-3	NC_007005.1	-	2864869	2864701	HTH_CueR (cd01108)→ LeuA (COG0119)→ **RNA**→ IVD (cd01156)→ COG4799 (COG4799)→
[2]Psy-1-3	NC_005773.3	-	3034238	3034071	HTH_CueR (cd01108)→ LeuA (COG0119)→ **RNA**→ COG4799 (COG4799)→ CaiD (COG1024)→
Psy-3-3	NC_004578.1	-	3045151	3044971	←DnaC (COG1484) ←COG4584 (COG4584) **RNA**→ IVD (cd01156)→ COG4799 (COG4799)→
Ppu-2-2	NC_002947.3	+	4013374	4013536	HTH_CueR (cd01108)→ LeuA (COG0119)→ **RNA**→ ←SapB (COG1285) hypo→
Ppu-1-2	NZ_AALM01000002.1	-	287203	287041	HTH_CueR (cd01108)→ LeuA (COG0119)→ **RNA**→ ←SapB (COG1285) hypo (cons)→
[1]Avi-1-2	NZ_AAAU03000009.1	+	123848	124019	hypo (cons)→ ←hypo **RNA**→ ←hypo ←DUF480 (pfam04337)
Pen-1-2	NC_008027.1	+	3150359	3150522	HTH_CueR (cd01108)→ LeuA (COG0119)→ **RNA**→ ←SapB (COG1285) NapH (COG0348)→

cd01108 Helix-turn-helix transcription regulator CueR.

cd01156 Isovaleryl-CoA dehydrogenase (IVD) is an is an acyl-CoA dehydrogenase, which catalyzes the third step in leucine catabolism, the conversion of isovaleryl-CoA (3-methylbutyryl-CoA) into 3-methylcrotonyl-CoA.

COG0119 Isopropylmalate/homocitrate/citramalate synthases [Amino acid transport and metabolism]

COG0318 Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases II [Lipid metabolism / Secondary metabolites biosynthesis, transport, and catabolism]

COG0348 Polyferredoxin [Energy production and conversion]

COG0591 Na+/proline symporter [Amino acid transport and metabolism / General function prediction only]

COG0617 tRNA nucleotidyltransferase/poly(A) polymerase [Translation, ribosomal structure and biogenesis]

COG0642 Signal transduction histidine kinase [Signal transduction mechanisms]

COG0801 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase [Coenzyme metabolism]

COG1024 Enoyl-CoA hydratase/carnithine racemase [Lipid metabolism]

COG1285 Uncharacterized membrane protein [Function unknown]

COG1484 DNA replication protein [DNA replication, recombination, and repair]

COG2113 ABC-type proline/glycine betaine transport systems, periplasmic components [Amino acid transport and metabolism]

COG2204 Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains [Signal transduction mechanisms]

COG4191 Signal transduction histidine kinase regulating C4-dicarboxylate transport system [Signal transduction mechanisms]

COG4584 Transposase and inactivated derivatives [DNA replication, recombination, and repair]

COG4799 Acetyl-CoA carboxylase, carboxyltransferase component (subunits alpha and beta) [Lipid metabolism]

COG5000 Signal transduction histidine kinase involved in nitrogen fixation and metabolism regulation [Signal transduction mechanisms]

pfam00989 PAS domain.

pfam04333 VacJ like lipoprotein.

pfam04337 Protein of unknown function, DUF480.

pfam04341 Protein of unknown function, DUF485.

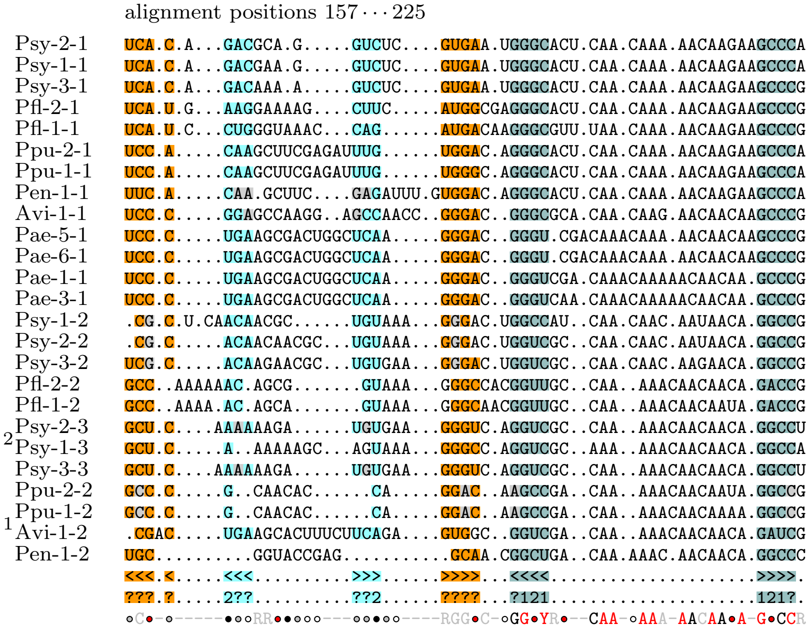
smart00387 Histidine kinase-like ATPases; Histidine kinase-, DNA gyrase B-, phytochrome-like ATPases.

smart00422 helix_turn_helix, mercury resistance;

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Pae-2-1, Pae-4-1

alignment positions 1 · · · 156

```
  Psy-2-1  CAACCAGGC.CGAGAACAACAAAAACUACC.CUUA...GGUAGCGCCUGAACU.GG.UUG.GAUC.GAAU...GAUCAU................UGCA....ACGCAGCGAC.CAAAGCAAUCCGUUUGCUCUUGA.CUCCCG.ACUGGGAGGUU
  Psy-1-1  CAACCAGGC.CGAGAACAACAAAAACUACC.CUUA...GGUAGAGCCUGAACU.GG.UUG.GAUC.GAAU...GAUCAU................UGCA....ACGCAGCGAC.CAAAGCAAUCCGUUUGCUCUUGA.CUCCCG.AUUGGGAGGUU
  Psy-3-1  CAACCAGGC.CGAGAACAAUAAAAACUACC.CUUA...GGUAGCGCCUGAACU.GG.UUG.GAUC.GAAU...GAUCAU................UGCA....ACGCAGCGAC.CAAAGCAAUCCGUUUGCUCUUGA.CUCCCG.ACUGGGAGGGU
  Pfl-2-1  CGACCAGGC.CGAGAACAACAAAAACUGUC.UUAA...GACAGAGCCUGUACU.GG.UUG.GAUC.GCAA...GAUCAC................UGCA....ACACAGCGAC.CAAAGCAAUCCGUUUGCUCUUGG.CUCCCG.AUUGGGAG.GG
  Pfl-1-1  CAACCAGGC.CGAGAACAACAAAAACUGCC.CUAA...GGCAGAGCCUGAACU.GG.UUG.GAUC.GAAA...GGUCAU................UGCAAC.U...CAGCGAU.CAAAGCAAUCCGUUUGCUCUUGG.CUCCCG.AUUGGGAG.GG
  Ppu-2-1  CAACCAGGC.AGAGAACAACAAAAACUGC.ACUAAAA..GCAGCGCCUGAACU.GG.UUG.GAUC.GACA...GAUCAACGU.............GACA...U...CAGCGGC.CAAAGCAAUCCGUUUGCUCUUUA.CCCCCGGUUUGGGGGUCG
  Ppu-1-1  CAACCAGGC.AGAGAACAACAAAAACUGC.ACUAAAA..GCAGCGCCUGAACU.GG.UUG.GAUC.GACA...GAUCAACGU.............GUCA...U...CAGCGGC.CAAAGCAAUCCGUUUGCUCUUUA.CCCCCGGUUUGGGGGUCG
  Pen-1-1  CGAUCAGGC.AGAGAACAACAAAAACUGC.ACUCAAAA.GCAGCACCUGAACU.GA.UUG.GAUC.GAAA...GAUCAGCA..............UCAUC..U...CGGCGAC.CAAAGCAAUCCGUUUGCUCUUCA.CCCCAGGUUUUGGGGUCG
  Avi-1-1  CAGCCAGAC.CGAGAACAACAAAA.CUGC.CACGAA...GCAGCACCCGAACU.GG.UAG.GUCC.GCA....GGACCCUA..............GGCAAUAU...CUGCAAC.CAAAGAAAUCCGUUUGCUAUUGG.CUCCCC.AUUGGGGGCCA
  Pae-5-1  CAGCCAGUCGGAAGAAGAAUAAAA.CUGCC.UUGA...GGCAGCGCACAGACU.GG.UUG.GAUC.GCUCGACGAUCAUGGC.............AGCA...U...CAGCGAC.CAAAGCAAUCCGUUUGCUAUUGAACUCCCAGCCUGGGAGAUA
  Pae-6-1  CAGCCAGUCGGAAGAAGAAUAAAA.CUGCC.UCGA...GGCAGCGCACAGACU.GG.UUG.GAUC.GCUCGACGAUCAUGGC.............AGCA...U...CAGCGAC.CAAAGCAAUCCGUUUGCUAUUGAACUCCCAGCCUGGGAGAUA
  Pae-1-1  CAGCCAGUCGAAAGAAGAAUAAAA.CUGCC.UUGA...GGCAGCGCACAGACU.GG.UUGGAUCG.CUCGA..CGAUCAUGGC............AGCA...U...CAGCGAC.CAAAGCAAUCCGUUUGCUAUUGAACUCCCAGCCUGGGAGAUA
  Pae-3-1  UAGCCAGUCGGAAGAAGAAUAAAA.CUGCC.UCGA...GGCAGCGCACAGACU.GG.UUGGAUCG.CUUGG..CGAUCAUGGC............AGCA...U...CAGCGAC.CAAAGCAAUCCGUUUGCUAUUGAACUCCCAGCCUGGGAGAUA
  Psy-1-2  CGACCGGAC.AGAGAACAAUAAAA.CUACC.UUCA...GGUAGCUUCCGAACC.GG.UGG.GAUC.ACGA...UGUUGUGCAGACAGGUGAUUAAAGCGAG.U...CAGCG.CUCAAAAAAAUACGUUUGCUCUUG..ACCCCGGAUGGGGGU...
  Psy-2-2  CGACCGGAC.CGAGAACAAUAAAA.CUACC.GUCA...GGUAGCUCCAGAACC.GG.UGG.GAUC.ACGC...UGUUGUGCAGGCAGGUGAUUAAAGCGAA.U...CAGCG.CUCAAAAAAAUACGUUUGCUCUUG..ACCCCGGAUAGGGGU...
  Psy-3-2  CGACCGGAC.AGAGAACAAUAAAA.CUACC.GUCA...GGUAGCUCCAGAACC.GG.UGG.GAUC.AGGC...UGUUGUGCAGACAGAUGAUUAAAGCGAA.U...CAGCG.CUCAAAAAAAUACGUUUGCUCUUGA.CCCUG.AUU..GGGGG..
  Pfl-2-2  CAACCGGGC.AGAGAACAAUAAAA.CUACC.UUCA...GGUAGCUCCCGAAUC.GG.UUG.GAUC.GCUA...GACGAGAA..............AGCAGA.U...CAGCG.CUCAAAAAAAUACGUUUGCUCUUG..ACCCCGGAUGGGGGU..C
  Pfl-1-2  CAACCGGUC.AGAGAACAAUAAAA.CUACC.CUGA...GGUAGCUCCCGAACU.GG.UUG.GAUC.GCCC...UGGCGACACAU...........UGAA...U...CAGCG.CUCAAAAAAAUACGUUUGCUCUUG..AUCCCGGAUGGGGAU..C
  Psy-2-3  CAGCCAGAC.AGAGAACAACAAAA.CUGCC.CCGA...GGCAGUUGCCGAACU.GG.UUG.GAUC.ACUA..CGAUGAAAGU.............GUCA...U...CAGCG.CUCAAAAAAAUACGUUUGCUCUUG..AUCCCGUCUGCGGAU..C
 2Psy-1-3  CAGCCAGAC.AGAGAAUAACAAAA.CUGCC.CUGA...GGCAGCUCCCGAACU.GG.UUG.GAUC.ACAU...GAUGAA................AGCUUGAU...CAGCG.CUCAAAAAAAUACGUUUGCUCUUG..GUCCCGCAUGCGGAC..C
  Psy-3-3  CAGCCAGAC.AGAGAACAAUAAAA.CUGCC.CUGA...GGCAGCUCCGGAACU.GG.UUGAAUCG.UUG....CGAUAGAAAGCACGAUGAAAGUGACA...U...CAGCG.CUCAAAAAAAUACGUUUGCUCUUG..GUCCAGCCUCUGGAC..C
  Ppu-2-2  CGGUCUGGC.ACAGAACAAUAAAA.CUACC.UCAA...GGUAGGGGCAGC.CA.GGUUCG.GAUCAGCAA..UGAUGAA................AGCAGG.U...CAGCG.CUCAAAAAAAUACGUUUGCUCUUG..ACCCCGCCUGGGGGU..C
  Ppu-1-2  CGGUCUGGC.ACAGAACAAUAAAA.CUACC.UCAA...GGUAGCGGCAGC.CA.GGUUCG.GAUCAGCAA..UGAUGAA................AGCAGG.U...CAGCG.CUCAAAAAAAUACGUUUGCUCUUG..ACCCCGCAUGGGGGU..C
 1Avi-1-2  CAGCCGGAC.AGAGAACAAUAAAA.CUACC.UCGA...GGUAGCGCCCGGACU.GG.UUG.GAUC.ACUUGGUGAUCAU................CGCGAUAU...CGGCGAU.CAAAGAAAUACGUUUGCUCUUG..GCCCCGGCU.GGGGC...
  Pen-1-2  CAACCUGAC.ACAGAACAAUAAAA.CUACC.UGCAA..GGUAGCGGCAGU.CACGG.UUG.GAUC.ACGUUG.GAUGAACGC.............AGG....U...CAGCG.CUCAAAAAAAUACGUUUGCUCUUG..GCCCCGGGUAGGGGCCC.
           <<<<<<<..................<<<<<........>>>>>........>>.>>.>>>.<<<<........>>>>.............................................................<<<<<.....>>>>>...
           1?12021..................00220........02200........12.02.1?1.??2?........?2??.............................................................2201?.....?1022...
           CRRCCRGRC-•GAGAACAAYAAAA-CURCC-YY•A---GGYAGC•CC•GRACY-GG-UUG-GAUC-R••∘---•RY•R•∘∘--------------•GCR∘--U---CAGCG∘C-CAAAR•AAU•CGUUUGCUCUUG--•YCCCG∘•U∘GGGR•-∘∘
```



```
         alignment positions 157···225
 Psy-2-1 UCA.C.A...GACGCA.G.....GUCUC....GUGAA.UGGGCACU.CAA.CAAA.AACAAGAAGCCCA
 Psy-1-1 UCA.C.A...GACGAA.G.....GUCUC....GUGAA.UGGGCACU.CAA.CAAA.AACAAGAAGCCCA
 Psy-3-1 UCA.C.A...GACAAA.A.....GUCUC....GUGAA.UGGGCACU.CAA.CAAA.AACAAGAAGCCCG
 Pfl-2-1 UCA.U.G...AAGGAAAAG....CUUC.....AUGGCGAGGGCACU.CAA.CAAA.AACAAGAAGCCCG
 Pfl-1-1 UCA.U.C...CUGGGUAAAC...CAG......AUGACAAGGGCGUU.UAA.CAAA.AACAAGAAGCCCG
 Ppu-2-1 UCC.A.....CAAGCUUCGAGAUUUG......UGGAC.AGGGCACU.CAA.CAAA.AACAAGAAGCCCA
 Ppu-1-1 UCC.A.....CAAGCUUCGAGAUUUG......UGGGC.AGGGCACU.CAA.CAAA.AACAAGAAGCCCA
 Pen-1-1 UUC.A.....CAA.GCUUC....GAGAUUU.GUGGAC.AGGGCACU.CAA.CAAA.AACAAGAAGCCCA
 Avi-1-1 UCC.C.....GGAGCCAAGG..AGCCAACC..GGGAC..GGGCGCA.CAA.CAAG.AACAACAAGCCCG
 Pae-5-1 UCC.C.....UGAAGCGACUGGCUCAA.....GGGAC..GGGU.CGACAAACAAA.AACAACAAGCCCG
 Pae-6-1 UCC.C.....UGAAGCGACUGGCUCAA.....GGGAC..GGGU.CGACAAACAAA.AACAACAAGCCCG
 Pae-1-1 UCC.C.....UGAAGCGACUGGCUCAA.....GGGAC..GGGUCGA.CAAACAAAAACAACAA.GCCCG
 Pae-3-1 UCC.C.....UGAAGCGACUGGCUCAA.....GGGAC..GGGUCAA.CAAACAAAAACAACAA.GCCCG
 Psy-1-2 .CG.C.U.CAACAACGC......UGUAAA...GGGAC.UGGCCAU..CAA.CAAC.AAUAACA.GGCCG
 Psy-2-2 .CG.C.....ACAACAACGC...UGUAAA...GGGAC.UGGUCGC..CAA.CAAC.AAUAACA.GGCCG
 Psy-3-2 UCG.C.....ACAAGAACGC...UGUGAA...GGGAC.UGGUCGC..CAA.CAAC.AAGAACA.GGCCG
 Pfl-2-2 GCC..AAAAAAC.AGCG.......GUAAA...GGGCCACGGUUGC..CAA..AAACAACAACA.GACCG
 Pfl-1-2 GCC..AAAA.AC.AGCA.......GUAAA...GGGCAACGGUUGC..CAA..AAACAACAAUA.GACCG
 Psy-2-3 GCU.C....AAAAAAGA......UGUGAA...GGGUC.AGGUCGC..CAA..AAACAACAACA.GGCCU
2Psy-1-3 GCU.C.....A..AAAAAGC...AGUAAA...GGGCC.AGGUCGC..AAA..AAACAACAACA.GGCCA
 Psy-3-3 GCU.C....AAAAAAGA......UGUGAA...GGGUC.AGGUCGC..CAA..AAACAACAACA.GGCCU
 Ppu-2-2 GCC.C.....G..CAACAC......CA.....GGAC..AAGCCGA..CAA..AAACAACAAUA.GGCCG
 Ppu-1-2 GCC.C.....G..CAACAC......CA.....GGAC..AAGCCGA..CAA..AAACAACAAAA.GGCCG
1Avi-1-2 .CGAC.....UGAAGCACUUUCUUCAGA....GUGGC..GGUCGA..CAA..AAACAACAACA.GAUCG
 Pen-1-2 UGC.........GGUACCGAG...........GCAA.CGGCUGA..CAA.AAAC.AACAACA.GGCCC
         <<<.<.....<<<..........>>>.....>>>>...<<<<....................>>>>.
         ???.?.....2??..........??2.....????...?121....................121?.
         ○C●-○-----●○○RR●●○○○---○○●○○----RGG●C-○GG●YR●--CAA-○AAA-AACAA●A-G●CCR
```

2.21 coccus-1

This possible ncRNA is present in many Firmicutes in the genera *Staphylococcus*, *Streptococcus* and *Lactococcus*. The motif consists of two hairpins. Although there is covariation, there are also mutations that break the structure, and many putative homologs have strong homology to one hairpin, but not the other.

abbrev. of hits	**taxonomy of species**
Sau-1-1 to Sau-1-2	Firmicutes Bacillales *Staphylococcus aureus* RF122
Sau-2-1 to Sau-2-6	Firmicutes Bacillales *Staphylococcus aureus subsp. aureus* COL
Sau-3-1 to Sau-3-8	Firmicutes Bacillales *Staphylococcus aureus subsp. aureus* JH1
Sau-4-1 to Sau-4-8	Firmicutes Bacillales *Staphylococcus aureus subsp. aureus* JH9
Sau-5-1 to Sau-5-10	Firmicutes Bacillales *Staphylococcus aureus subsp. aureus* MRSA252
Sau-6-1 to Sau-6-8	Firmicutes Bacillales *Staphylococcus aureus subsp. aureus* MSSA476
Sau-8-1 to Sau-8-8	Firmicutes Bacillales *Staphylococcus aureus subsp. aureus* Mu50
Sau-7-1 to Sau-7-8	Firmicutes Bacillales *Staphylococcus aureus subsp. aureus* MW2
Sau-9-1 to Sau-9-8	Firmicutes Bacillales *Staphylococcus aureus subsp. aureus* N315
Sau-10-1 to Sau-10-6	Firmicutes Bacillales *Staphylococcus aureus subsp. aureus* NCTC 8325
Sau-11-1 to Sau-11-6	Firmicutes Bacillales *Staphylococcus aureus subsp. aureus* USA300
Sep-1-1 to Sep-1-1	Firmicutes Bacillales *Staphylococcus epidermidis* ATCC 12228

Sep-2-1 to Sep-2-2	Firmicutes Bacillales *Staphylococcus epidermidis* RP62A
Sha-1-1 to Sha-1-39	Firmicutes Bacillales *Staphylococcus haemolyticus* JCSC1435
Ssa-1-1 to Ssa-1-7	Firmicutes Bacillales *Staphylococcus saprophyticus subsp. saprophyticus* ATCC 15305
Lre-1-1 to Lre-1-17	Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus reuteri* 100-23
Lre-2-1 to Lre-2-22	Firmicutes Lactobacillales Lactobacillaceae *Lactobacillus reuteri* JCM 1112
Sag-1-1 to Sag-1-2	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 18RS21
Sag-2-1 to Sag-2-2	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 2603V/R
Sag-3-1 to Sag-3-2	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* 515
Sag-4-1 to Sag-4-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* A909
Sag-5-1 to Sag-5-2	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* CJB111
Sag-6-1 to Sag-6-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* COH1
Sag-7-1 to Sag-7-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* H36B
Sag-8-1 to Sag-8-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus agalactiae* NEM316
Smu-1-1 to Smu-1-2	Firmicutes Lactobacillales Streptococcaceae *Streptococcus mutans* UA159
Spn-1-1 to Spn-1-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* R6
Spn-2-1 to Spn-2-2	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pneumoniae* TIGR4
Spy-1-1 to Spy-1-4	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* M1 GAS
Spy-2-1 to Spy-2-3	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* M49 591
Spy-3-1 to Spy-3-3	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10270
Spy-4-1 to Spy-4-3	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10394
Spy-5-1 to Spy-5-3	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS10750
Spy-6-1 to Spy-6-3	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS2096
Spy-7-1 to Spy-7-3	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS315
Spy-8-1 to Spy-8-4	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS5005
Spy-9-1 to Spy-9-3	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS6180
Spy-10-1 to Spy-10-3	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS8232
Spy-11-1 to Spy-11-3	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* MGAS9429
Spy-12-1 to Spy-12-3	Firmicutes Lactobacillales Streptococcaceae *Streptococcus pyogenes* SSI-1
Ssu-1-1 to Ssu-1-22	Firmicutes Lactobacillales Streptococcaceae *Streptococcus suis* 89/1591
Sth-1-1 to Sth-1-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* CNRZ1066
Sth-2-1 to Sth-2-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* LMD-9
Sth-3-1 to Sth-3-1	Firmicutes Lactobacillales Streptococcaceae *Streptococcus thermophilus* LMG 18311

abbrev	**RefSeq accession**		**5′ at**	**3′ at**	**genes**
Ssu-1-1	NZ_AAFA02000017.1	+	3520	3641	COG3464 (COG3464)→ **RNA**→ ←PrfC (COG4108)
Spy-8-1	NC_007297.1	+	48492	48578	PurD (COG0151)→ **RNA**→ PurE (COG0041)→
Spy-1-1	NC_002737.1	+	47904	47990	**RNA**→ FldA (COG0716)→
Spy-4-1	NC_006086.1	+	89441	89527	PurD (COG0151)→ **RNA**→ PurE (COG0041)→
Spy-9-1	NC_007296.1	+	48201	48287	**RNA**→ FldA (COG0716)→
Spy-12-1	NC_004606.1	+	42630	42716	**RNA**→
Spy-7-1	NC_004070.1	+	48631	48717	**RNA**→ FldA (COG0716)→
Spy-2-1	NZ_AAFV01000004.1	-	10086	10000	**RNA**→
Spy-11-1	NC_008021.1	+	48228	48314	**RNA**→
Spy-3-1	NC_008022.1	+	48597	48683	**RNA**→ FldA (COG0716)→
Spy-6-1	NC_008023.1	+	48203	48289	**RNA**→
Spy-5-1	NC_008024.1	+	48488	48574	PurD (COG0151)→ **RNA**→ PurE (COG0041)→
Spy-10-1	NC_003485.1	+	48513	48599	PurD (COG0151)→ **RNA**→ PurE (COG0041)→
Sag-8-1	NC_004368.1	+	59292	59378	PurD (COG0151)→ **RNA**→ PurE (COG0041)→

Sag-3-1	NZ_AAJP01000029.1	-	7801	7715	**RNA**→
Sag-6-1	NZ_AAJR01000025.1	-	3651	3565	PurD (COG0151)→ **RNA**→ PurE (COG0041)→
Sag-4-1	NC_007432.1	+	59209	59295	PurD (COG0151)→ **RNA**→ PurE (COG0041)→
Sag-7-1	NZ_AAJS01000005.1	+	42970	43056	**RNA**→
Sag-2-1	NC_004116.1	+	59396	59482	PurD (COG0151)→ **RNA**→ PurE (COG0041)→
Sag-1-1	NZ_AAJO01000037.1	+	12850	12936	**RNA**→
Sag-5-1	NZ_AAJQ01000011.1	+	33507	33593	**RNA**→
Spy-8-2	NC_007297.1	+	756948	757054	←hypo **RNA**→ ←hypo
Spy-1-2	NC_002737.1	+	838364	838470	AceF (COG0508)→ **RNA**→ Biotin_lipoyl (pfam00364)Lpd (COG1249)→
Spy-9-2	NC_007296.1	+	754881	754987	putative extracellular matrix binding protein→ **RNA**→
Spy-11-2	NC_008021.1	+	836161	836267	**RNA**→
Spy-3-2	NC_008022.1	+	828063	828169	←hypo **RNA**→ ←hypo
Spy-6-2	NC_008023.1	+	796738	796844	**RNA**→
Spy-12-2	NC_004606.1	-	1169720	1169614	**RNA**→
Spy-7-2	NC_004070.1	+	727036	727142	COG2374 (COG2374)→ **RNA**→
Spy-2-2	NZ_AAFV01000056.1	-	5123	5017	**RNA**→
Spy-5-2	NC_008024.1	+	859336	859442	←hypo **RNA**→ ←hypo
Spy-10-2	NC_003485.1	+	836933	837039	AceF (COG0508)→ **RNA**→ Lpd (COG1249)Biotin_lipoyl (pfam00364)→
Spy-4-2	NC_006086.1	+	783144	783250	←hypo **RNA**→ ←hypo
Spy-8-3	NC_007297.1	+	756950	757054	←hypo **RNA**→ ←hypo
Spy-1-3	NC_002737.1	+	838366	838470	AceF (COG0508)→ **RNA**→ Biotin_lipoyl (pfam00364)Lpd (COG1249)→
Ssu-1-2	NZ_AAFA02000083.1	-	5326	5222	Ribosomal_S8 (pfam00410)→ **RNA**→ RplF (COG0097)→
Ssu-1-3	NZ_AAFA02000016.1	-	13879	13775	MrcB (COG0744)→ **RNA**→ RpoB (COG0085)→
Smu-1-1	NC_004350.1	+	53721	53824	CPSase_L_D2 (pfam02786)→ **RNA**→ PurE (COG0041)→
Ssu-1-4	NZ_AAFA02000093.1	-	1031	928	←AcoB (COG0022) **RNA**→ ←TPP_E1_PDC_ADC_BCADC (cd02000)
Ssu-1-5	NZ_AAFA02000066.1	-	2732	2628	←COG4129 (COG4129) **RNA**→ ←Phosphoglycerate_kinase (cd00318)
Spy-8-4	NC_007297.1	-	548987	548871	←COG4120 (COG4120) **RNA**→ ←PutA (COG1012)EutG (COG1454)
Spy-5-3	NC_008024.1	-	608890	608774	←COG4120 (COG4120) **RNA**→ ←yrhL_like (cd01840)COG1835 (COG1835)
Spy-1-4	NC_002737.1	-	587963	587847	COG3547 (COG3547)→ **RNA**→ ←enolase (cd03313)
Spy-12-3	NC_004606.1	+	1367615	1367731	**RNA**→
Spy-7-3	NC_004070.1	-	529126	529010	←streptolysin S associated protein **RNA**→ ←enolase (cd03313)
Spy-3-3	NC_008022.1	-	590126	590010	←Streptolysin S precursor **RNA**→ ←enolase (cd03313)
Spy-11-3	NC_008021.1	-	588877	588761	**RNA**→
Spy-6-3	NC_008023.1	-	590278	590162	**RNA**→
Spy-2-3	NZ_AAFV01000009.1	+	16935	17051	Transposase_20 (pfam02371)→ **RNA**→ ←Eno (COG0148)
Spy-9-3	NC_007296.1	-	546590	546474	Transposase_20 (pfam02371)→ **RNA**→ ←enolase (cd03313)
Spy-4-3	NC_006086.1	-	575023	574907	←Streptolysin S precursor **RNA**→ ←Eno (COG0148)
Spy-10-3	NC_003485.1	-	637677	637561	←streptolysin S associated protein **RNA**→ ←enolase (cd03313)
Sau-6-1	NC_002953.3	+	1244039	1244142	←hypo **RNA**→ ←FtsK (COG1674)
Sau-7-1	NC_003923.1	+	1215366	1215469	adh_short (pfam00106)→ **RNA**→ AcpP (COG0236)→
Sau-3-1	NZ_AAPK01000033.1	-	8812	8709	**RNA**→
Sau-4-1	NZ_AAPL01000016.1	-	7809	7706	adh_short (pfam00106)→ **RNA**→ AcpP (COG0236)→
Sau-11-1	NC_007793.1	+	1231269	1231372	**RNA**→

Sau-10-1	NC_007795.1	+	1151323	1151426	AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→ **RNA**→
Sau-2-1	NC_002951.2	+	1255089	1255192	AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→ **RNA**→
Sau-8-1	NC_002758.2	+	1291833	1291936	**RNA**→
Sau-9-1	NC_002745.2	+	1215504	1215607	adh_short (pfam00106)→ **RNA**→ AcpP (COG0236)→
Sau-5-1	NC_002952.2	+	1257472	1257575	adh_short (pfam00106)→ **RNA**→ AcpP (COG0236)→
Sau-11-2	NC_007793.1	-	2866749	2866627	**RNA**→
Sau-10-2	NC_007795.1	-	2815313	2815191	**RNA**→ ←AICARFT_IMPCHas (pfam01808)MGS (pfam02142)
Sau-2-2	NC_002951.2	-	2803401	2803279	**RNA**→ ←AICARFT_IMPCHas (pfam01808)MGS (pfam02142)
Sau-8-2	NC_002758.2	-	2872466	2872344	**RNA**→
Sau-9-2	NC_002745.2	-	2808753	2808631	←hypo **RNA**→ hypo→
Sau-3-2	NZ_AAPK01000001.1	-	284096	283974	**RNA**→ DapF (COG0253)→
Sau-4-2	NZ_AAPL01000001.1	-	265140	265018	**RNA**→ hypo (cons)→
Sau-1-1	NC_007622.1	-	2265660	2265538	**RNA**→
Sau-5-2	NC_002952.2	+	438463	438585	←hypo **RNA**→ ←putative DNA-binding protein
Sau-5-3	NC_002952.2	-	2882270	2882148	**RNA**→ COG3942 (COG3942)→
Sau-5-4	NC_002952.2	-	1136979	1136857	←Rpe (COG0036) **RNA**→ hypo→
Sau-1-2	NC_007622.1	-	2114388	2114266	**RNA**→
Sau-6-2	NC_002953.3	+	770241	770363	←hypo **RNA**→ ←FtsK (COG1674)
Sau-6-3	NC_002953.3	+	2519319	2519441	←hypo **RNA**→ ←AraJ (COG2814)
Sau-6-4	NC_002953.3	-	2155953	2155831	←PnbA (COG2272) **RNA**→ ABC_SMC_barmotin (cd03278)→
Sau-7-2	NC_003923.1	+	771813	771935	hypo→ **RNA**→ Rpe (COG0036)→
Sau-7-3	NC_003923.1	+	2540151	2540273	←hypo **RNA**→
Sau-7-4	NC_003923.1	-	2176848	2176726	**RNA**→ hypo→
Sau-6-5	NC_002953.3	-	1856583	1856461	←PnbA (COG2272) **RNA**→ ABC_SMC_barmotin (cd03278)→
Sau-7-5	NC_003923.1	-	1877248	1877126	←hypo **RNA**→ ←hypo
Sau-8-3	NC_002758.2	+	2597882	2598004	**RNA**→
Sau-9-3	NC_002745.2	+	2526688	2526810	tRNA_m1G_MT (pfam01746)→ **RNA**→ ←RarD (COG2962)
Sau-3-3	NZ_AAPK01000001.1	+	11544	11666	hypo (cons)→ **RNA**→ ←AraJ (COG2814)
Sau-4-3	NZ_AAPL01000042.1	+	4066	4188	**RNA**→
Sau-8-4	NC_002758.2	+	2723168	2723290	**RNA**→
Sau-9-4	NC_002745.2	+	2652742	2652864	tRNA_m1G_MT (pfam01746)→ **RNA**→ ←RarD (COG2962)
Sau-3-4	NZ_AAPK01000001.1	+	134799	134921	←DapF (COG0253) **RNA**→
Sau-4-4	NZ_AAPL01000001.1	+	115843	115965	DUF896 (pfam05979)→ **RNA**→ ←COG3607 (COG3607)
Sau-8-5	NC_002758.2	-	1142938	1142816	**RNA**→
Sau-9-5	NC_002745.2	-	1066610	1066488	COG0595 (COG0595)→ **RNA**→ ←TrkA (COG0569)
Sau-3-5	NZ_AAPK01000013.1	+	9412	9534	**RNA**→
Sau-4-5	NZ_AAPL01000011.1	-	54843	54721	**RNA**→
Sau-5-5	NC_002952.2	+	1743803	1743925	←hypo **RNA**→ ←MntH (COG1914)
Sau-5-6	NC_002952.2	-	2254246	2254124	←hypo **RNA**→ →
Sau-2-3	NC_002951.2	-	1081817	1081695	FolD (COG0190)→ **RNA**→ chitinase-related protein→
Sau-11-3	NC_007793.1	-	1058342	1058220	**RNA**→
Sau-10-3	NC_007795.1	-	978217	978095	FolD (COG0190)→ **RNA**→ hypo→
Sau-8-6	NC_002758.2	-	1116827	1116705	**RNA**→

Sau-9-6	NC_002745.2	-	1040499	1040377	←hypo **RNA**→ ←CeuB (COG4606)
Sau-3-6	NZ_AAPK01000013.1	+	35459	35581	**RNA**→
Sau-4-6	NZ_AAPL01000011.1	-	28796	28674	**RNA**→
Sau-8-7	NC_002758.2	+	2705594	2705716	**RNA**→
Sau-9-7	NC_002745.2	+	2635168	2635290	tRNA_m1G_MT (pfam01746)→ **RNA**→ ←RarD (COG2962)
Sau-3-7	NZ_AAPK01000001.1	+	117226	117348	←DapF (COG0253) **RNA**→
Sau-4-7	NZ_AAPL01000001.1	+	98270	98392	HMA (cd00371)ZntA (COG2217)→ **RNA**→ CopZ (COG2608)→
Sau-5-7	NC_002952.2	+	2550397	2550519	←NarK (COG2223) **RNA**→ COG5504 (COG5504)→
Sau-11-4	NC_007793.1	+	1873177	1873299	**RNA**→
Sau-10-4	NC_007795.1	+	1776851	1776973	AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→ **RNA**→
Sau-2-4	NC_002951.2	+	1850866	1850988	AICARFT_IMPCHas (pfam01808)MGS (pfam02142)→ **RNA**→
Sau-7-6	NC_003923.1	+	1842030	1842152	←ArgE (COG0624) **RNA**→ ←hypo
Sau-5-8	NC_002952.2	+	1547267	1547389	←Era (COG1159)MMR_HSR1 (pfam01926) **RNA**→ ←hypo
Sau-6-6	NC_002953.3	+	1821365	1821488	←ArgE (COG0624) **RNA**→ ←hypo
Sha-1-1	NC_007168.1	-	802593	802470	AraJ (COG2814)→ **RNA**→ ←MarR (COG1846)
Sha-1-2	NC_007168.1	-	1102516	1102393	←16S RIBOSOMAL RNA **RNA**→ ←Fur (COG0735)
Sha-1-3	NC_007168.1	+	2570385	2570508	hypo→ **RNA**→ ←hypo
Sha-1-4	NC_007168.1	-	839902	839779	←WecB (COG0381) **RNA**→ adh_short (pfam00106)→
Sha-1-5	NC_007168.1	-	482661	482538	←hypo **RNA**→ ←AlsD (COG3527)
Sha-1-6	NC_007168.1	+	1996068	1996191	←FabI (COG0623) **RNA**→ ←KefB (COG0475)TrkA (COG0569)
Sha-1-7	NC_007168.1	-	941168	941045	hypo→ **RNA**→ ←FabZ (cd01288)
Ssa-1-1	NC_007350.1	-	622311	622189	hypo→ **RNA**→ ←PTS_IIB_glc (cd00212)PtsG (COG1263)
Ssa-1-2	NC_007350.1	-	2080956	2080834	ARA1 (COG0656)→ **RNA**→ Glycos_transf_2 (pfam00535)→
Ssa-1-3	NC_007350.1	+	2125473	2125595	LysM (cd00118)COG3942 (COG3942)LysM (pfam01476)→ **RNA**→ ←PitA (COG0306)
Ssa-1-4	NC_007350.1	+	496871	496991	←TauB (COG1116) **RNA**→ ←hypo
Ssa-1-5	NC_007350.1	+	2503605	2503736	ARA1 (COG0656)→ **RNA**→ ←hypo
Sha-1-8	NC_007168.1	+	786707	786819	hypo→ **RNA**→ hypo→
Sha-1-9	NC_007168.1	-	1640536	1640425	hypo→ **RNA**→ hypo→
Sha-1-10	NC_007168.1	+	1150313	1150425	hypo→ **RNA**→ ARA1 (COG0656)→
Sha-1-11	NC_007168.1	+	618527	618639	AraJ (COG2814)→ **RNA**→ ←PnbA (COG2272)
Sha-1-12	NC_007168.1	-	2443450	2443338	LdcC (COG1982)→ **RNA**→ FtsX (pfam02687)→
Sha-1-13	NC_007168.1	-	385942	385830	←CitMHS (pfam03600) **RNA**→ ←hypo
Sha-1-14	NC_007168.1	-	2264744	2264632	PitA (COG0306)→ **RNA**→ ←LysM (cd00118)COG3942 (COG3942)LysM (pfam01476)
Sha-1-15	NC_007168.1	-	561630	561518	COG2514 (COG2514)→ **RNA**→ ←PhnD (COG3221)
Sha-1-16	NC_007168.1	-	2065977	2065865	Ndh (COG1252)→ **RNA**→ Peptidase_M17 (cd00433)→
Sha-1-17	NC_007168.1	+	448710	448825	BetA (COG2303)→ **RNA**→ XynB (COG3664)HTH_ARAC (smart00342)→
Sha-1-18	NC_007168.1	-	658143	658031	hypo→ **RNA**→ ←GpmA (COG0588)
Sha-1-19	NC_007168.1	-	2227722	2227611	TlyC (COG1253)→ **RNA**→ ARA1 (COG0656)→
Sha-1-20	NC_007168.1	+	319416	319527	PhnD (COG3221)→ **RNA**→ COG3638 (COG3638)→
Sha-1-21	NC_007168.1	+	2112362	2112462	hypo→ **RNA**→ hypo→
Sha-1-22	NC_007168.1	+	235915	236015	RarD (COG2962)→ **RNA**→ ←hypo
Sha-1-23	NC_007168.1	+	2229556	2229656	←TlyC (COG1253) **RNA**→ ←FruA (COG1299)PtsN (COG1762)PTS_IIB_fruc (pfam02379)
Sha-1-24	NC_007168.1	-	239998	239898	←hypo **RNA**→ ←TagB (COG1887)

Sha-1-25	NC_007168.1	-	2172032	2171921	COG1544 (COG1544)→ **RNA**→ SecA (COG0653)SEC-C (pfam02810)→
Sau-6-7	NC_002953.3	+	654232	654337	ArgS (COG0018)→ **RNA**→ COG2231 (COG2231)→
Sau-11-5	NC_007793.1	+	671024	671129	**RNA**→
Sau-10-5	NC_007795.1	+	603819	603924	**RNA**→ ←ElaA (COG2153)
Sau-2-5	NC_002951.2	+	692494	692599	**RNA**→ MdoB (COG1368)→
Sau-7-7	NC_003923.1	+	655540	655645	hypo→ **RNA**→ Rpe (COG0036)→
Sau-11-6	NC_007793.1	-	787220	787115	**RNA**→
Sau-10-6	NC_007795.1	-	720228	720123	ElaA (COG2153)→ **RNA**→
Sau-2-6	NC_002951.2	-	808765	808660	←COG4502 (COG4502) **RNA**→ ←HisC (COG0079)
Sau-8-8	NC_002758.2	+	801531	801636	**RNA**→
Sau-9-8	NC_002745.2	+	777338	777443	COG4502 (COG4502)→ **RNA**→ ←LCB5 (COG1597)
Sau-3-8	NZ_AAPK01000003.1	+	63146	63251	**RNA**→
Sau-4-8	NZ_AAPL01000004.1	-	22131	22026	**RNA**→
Sau-6-8	NC_002953.3	+	406548	406653	hypo→ **RNA**→ ←putative DNA-binding protein
Sau-7-8	NC_003923.1	+	407903	408008	hypo→ **RNA**→ ←hypo
Sau-5-9	NC_002952.2	+	2259183	2259288	←hypo **RNA**→ ←ATP-synt_DE_N (pfam02823)
Sau-5-10	NC_002952.2	-	2726948	2726843	←CopZ (COG2608) **RNA**→ ←HMA (cd00371)ZntA (COG2217)
Sha-1-26	NC_007168.1	+	771339	771448	COG2427 (COG2427)→ **RNA**→ LytD (COG4193)→
Sha-1-27	NC_007168.1	+	537863	537972	PTS_IIA_glc (cd00210)PTS_IIB_glc (cd00212)PtsG (COG1263)→ **RNA**→ ←Tdh (COG1063)
Sha-1-28	NC_007168.1	-	2183225	2183115	hypo→ **RNA**→ Gly_kinase (pfam02595)→
Sha-1-29	NC_007168.1	-	2644338	2644224	←COG1878 (COG1878) **RNA**→ ←Methionine_synt (pfam01717)CIMS_N_terminal_like (cd03312)
Ssa-1-6	NC_007350.1	-	1730753	1730634	PPAT_a (cd02163)→ **RNA**→ ←DUF795 (pfam05636)
Sha-1-30	NC_007168.1	+	967841	967952	FliA (COG1191)→ **RNA**→ Tex (COG2183)→
Sep-2-1	NC_006663.1	-	14712	14592	hypo→ **RNA**→ ABC_DR_subfamily_A (cd03230)→
Sha-1-31	NC_007168.1	+	2257974	2258044	CIMS_C_terminal_like (cd03311)→ **RNA**→ ←DUF985 (pfam06172)
Sha-1-32	NC_007168.1	+	145933	146003	Branch_AA_trans (pfam05525)→ **RNA**→ UbiE (COG2226)→
Sha-1-33	NC_007168.1	+	1895371	1895460	hypo→ **RNA**→ Nramp (pfam01566)→
Lre-2-1	NZ_AAOV01000001.1	+	718	787	PnuC (COG3201)→ **RNA**→ ←ARA1 (COG0656)
Lre-2-2	NZ_AAOV01000004.1	-	53960	53891	FbpA (pfam05833)DUF814 (pfam05670)→ **RNA**→ ←CarB (COG0458)CPSase_L_D3 (pfam02787)
Lre-2-3	NZ_AAOV01000006.1	-	42693	42624	COG3919 (COG3919)→ **RNA**→ Peptidase_S26 (pfam00461)→
Lre-2-4	NZ_AAOV01000025.1	-	15557	15488	**RNA**→
Lre-1-1	NZ_AAPZ01000007.1	+	23274	23343	COG3919 (COG3919)→ **RNA**→ Peptidase_S26 (pfam00461)→
Lre-1-2	NZ_AAPZ01000009.1	+	38129	38198	COG5549 (COG5549)→ **RNA**→ ←ZntA (COG2217)
Lre-1-3	NZ_AAPZ01000004.1	-	26250	26181	PspC (COG1983)→ **RNA**→ ←Transposase_9 (pfam01548)Transposase_20 (pfam02371)
Sha-1-34	NC_007168.1	-	710586	710488	TagB (COG1887)→ **RNA**→ ←Acetyltransf_1 (pfam00583)
Sha-1-35	NC_007168.1	+	1400040	1400149	RpmG (COG0267)→ **RNA**→ 5-FTHF_cyc-lig (pfam01812)→
Spn-2-1	NC_003028.1	+	1340542	1340648	←RimL (COG1670) **RNA**→ ←NadE (COG0171)
Spn-2-2	NZ_AAGY02000113.1	-	5545	5439	**RNA**→
Sha-1-36	NC_007168.1	+	2548059	2548170	←COG3153 (COG3153) **RNA**→ ←PhnF (COG2188)
Sha-1-37	NC_007168.1	+	1096325	1096425	MdlB (COG1132)→ **RNA**→ ←COG4129 (COG4129)
Lre-2-5	NZ_AAOV01000041.1	+	13371	13483	hypo (cons)→ **RNA**→ ←hypo (cons)
Lre-2-6	NZ_AAOV01000036.1	-	15859	15747	RpsA (COG0539)→ **RNA**→ COG1160 (COG1160)→
Spn-1-1	NC_003098.1	+	1271252	1271372	←RimL (COG1670) **RNA**→ ←NadE (COG0171)

Lre-1-4	NZ_AAPZ01000004.1	+	31872	32005	rve (pfam00665)→ **RNA**→ ←Transposase_9 (pfam01548)Transposase_20 (pfam02371)
Lre-1-5	NZ_AAPZ01000013.1	-	33113	33001	←MrcB (COG0744) **RNA**→ ←ArgS (COG0018)
Lre-2-7	NZ_AAOV01000004.1	-	53960	53848	FbpA (pfam05833)DUF814 (pfam05670)→ **RNA**→ ←CarB (COG0458)CPSase_L_D3 (pfam02787)
Lre-1-6	NZ_AAPZ01000037.1	-	5235	5123	PemK (pfam02452)→ **RNA**→ ←hypo
Lre-2-8	NZ_AAOV01000011.1	-	2891	2778	Gly_kinase (pfam02595)→ **RNA**→ hypo→
Lre-2-9	NZ_AAOV01000010.1	-	2671	2560	COG3001 (COG3001)→ **RNA**→ ←Dtyr_deacylase (cd00563)
Lre-2-10	NZ_AAOV01000006.1	+	1455	1567	AsnS (COG0017)→ **RNA**→ ←EAL (cd01948)
Lre-2-11	NZ_AAOV01000003.1	-	9042	8930	PerM (COG0628)→ **RNA**→ ←COG3382 (COG3382)
Lre-1-7	NZ_AAPZ01000003.1	-	6328	6213	←TpiA (COG0149) **RNA**→ ←SrtA (COG3764)
Lre-1-8	NZ_AAPZ01000002.1	-	39785	39670	ApbA (COG1893)→ **RNA**→ ←Uup (COG0488)
Lre-1-9	NZ_AAPZ01000016.1	+	12635	12745	←COG0062 (COG0062) **RNA**→ ←ErfK_YbiS_YhnG (pfam03734)
Lre-2-12	NZ_AAOV01000017.1	-	1140	1028	HINT_subgroup (cd01277)→ **RNA**→ ←SurA (COG0760)
Lre-1-10	NZ_AAPZ01000018.1	+	11003	11115	ABC_MTABC3_MDL1_MDL2 (cd03249)→ **RNA**→ ←COG1765 (COG1765)
Lre-1-11	NZ_AAPZ01000020.1	+	23426	23538	←hypo (cons) **RNA**→ ←hypo (cons)
Lre-2-13	NZ_AAOV01000017.1	+	24380	24492	←Tra8 (COG2826) **RNA**→ Glutaminase (pfam04960)→
Lre-2-14	NZ_AAOV01000026.1	-	4700	4587	DegQ (COG0265)→ **RNA**→ PgpA (COG1267)→
Lre-2-15	NZ_AAOV01000001.1	+	61513	61625	GldA (COG0371)→ **RNA**→ ←DUF915 (pfam06028)
Lre-1-12	NZ_AAPZ01000016.1	-	18411	18299	GldA (COG0371)→ **RNA**→ ←DUF915 (pfam06028)
Lre-2-16	NZ_AAOV01000060.1	-	5015	4903	←COG3830 (COG3830) **RNA**→ ←ARA1 (COG0656)
Lre-1-13	NZ_AAPZ01000069.1	+	5451	5563	COG2848 (COG2848)→ **RNA**→ ←Transposase_20 (pfam02371)
Lre-2-17	NZ_AAOV01000001.1	+	88090	88223	ThrS (COG0441)→ **RNA**→ ←COG2270 (COG2270)
Lre-2-18	NZ_AAOV01000025.1	-	15557	15424	**RNA**→
Lre-1-14	NZ_AAPZ01000009.1	+	38129	38262	COG5549 (COG5549)→ **RNA**→ ←ZntA (COG2217)
Lre-1-15	NZ_AAPZ01000007.1	+	23274	23407	COG3919 (COG3919)→ **RNA**→ Peptidase_S26 (pfam00461)→
Lre-1-16	NZ_AAPZ01000031.1	-	17594	17461	RhaT (COG0697)→ **RNA**→ UPF0236 (pfam06782)→
Lre-1-17	NZ_AAPZ01000025.1	-	14089	13956	←PseudoU_synth_EcTruB (cd02573) **RNA**→ ←RbfA (COG0858)
Lre-2-19	NZ_AAOV01000012.1	+	17456	17588	←PseudoU_synth_EcTruB (cd02573) **RNA**→ ←RbfA (COG0858)
Sha-1-38	NC_007168.1	+	2134361	2134471	← **RNA**→ ←SmpB (COG0691)
Sep-2-2	NC_002976.3	+	312121	312238	COG3949 (COG3949)→ **RNA**→ Aes (COG0657)→
Sep-1-1	NC_004461.1	+	419003	419120	**RNA**→
Ssu-1-6	NZ_AAFA02000044.1	+	11502	11623	beta_clamp (cd00140)→ **RNA**→ LCB5 (COG1597)→
Ssu-1-7	NZ_AAFA02000003.1	-	49970	49849	GidB (COG0357)→ **RNA**→ ←CSD (pfam00313)
Ssu-1-8	NZ_AAFA02000034.1	-	13769	13649	Peptidase_M29 (pfam02073)→ **RNA**→ COG2261 (COG2261)→
Ssu-1-9	NZ_AAFA02000040.1	+	6049	6169	DnaJ (COG0484)→ **RNA**→ LplA (COG0095)→
Ssu-1-10	NZ_AAFA02000003.1	-	61546	61426	←PriA (COG1198) **RNA**→ ←RpoZ (COG1758)
Ssu-1-11	NZ_AAFA02000051.1	+	6544	6664	**RNA**→
Ssu-1-12	NZ_AAFA02000091.1	-	5755	5635	Acetyltransf_1 (pfam00583)→ **RNA**→ ←HMG-CoA_reductase_classII (cd00644)
Ssu-1-13	NZ_AAFA02000150.1	-	2679	2558	hypo→ **RNA**→ HisS (COG0124)→
Ssu-1-14	NZ_AAFA02000032.1	-	15069	14949	**RNA**→ ←MscL (pfam01741)
Ssu-1-15	NZ_AAFA02000054.1	-	7664	7543	Rfe (COG0472)→ **RNA**→ SufC (COG0396)→
Ssu-1-16	NZ_AAFA02000001.1	-	5925	5808	PBPb (smart00062)→ **RNA**→ ←GlgB (COG0296)
Ssu-1-17	NZ_AAFA02000091.1	+	4201	4321	PksG (COG3425)→ **RNA**→ HMG-CoA_reductase_classII (cd00644)→
Ssu-1-18	NZ_AAFA02000049.1	+	507	626	COG2246 (COG2246)→ **RNA**→ ←Rbn (COG1295)

Ssu-1-19	NZ_AAFA02000002.1	+	19861	19979	←COG1193 (COG1193) **RNA**→ ←Colicin_V (pfam02674)
Ssu-1-20	NZ_AAFA02000059.1	-	9360	9242	COG2141 (COG2141)→ **RNA**→ ←COG3335 (COG3335)
Ssu-1-21	NZ_AAFA02000018.1	+	14581	14681	COG1079 (COG1079)→ **RNA**→ ←COG3547 (COG3547)Transposase_20 (pfam02371)
Lre-2-20	NZ_AAOV01000016.1	-	5380	5307	ATP-synt_DE_N (pfam02823)→ **RNA**→ MurA (COG0766)→
Lre-2-21	NZ_AAOV01000038.1	-	2197	2119	AAL_decarboxy (pfam03306)→ **RNA**→ PurE (COG0041)→
Sag-5-2	NZ_AAJQ01000032.1	-	13163	13049	←COG1247 (COG1247) **RNA**→ COG1476 (COG1476)→
Sag-2-2	NC_004116.1	-	40430	40306	←NanE (COG3010) **RNA**→
Sag-1-2	NZ_AAJO01000054.1	+	1255	1379	**RNA**→
Sag-3-2	NZ_AAJP01000052.1	+	5507	5631	←PurF (COG0034) **RNA**→ ←PurL (COG0046)PurL (COG0047)
Ssu-1-22	NZ_AAFA02000103.1	+	1904	2014	←Adenine_glyco (pfam03352) **RNA**→ ←RuvA (COG0632)
Sth-3-1	NC_006448.1	+	1151674	1151793	←DHDPS (cd00950) **RNA**→ ←Asd (COG0136)
Sth-1-1	NC_006449.1	+	1149545	1149664	**RNA**→
Sth-2-1	NZ_AAGS01000021.1	+	13313	13432	←DHDPS (cd00950) **RNA**→ ←Asd (COG0136)
Lre-2-22	NZ_AAOV01000048.1	-	3412	3278	hypo→ **RNA**→ ←Pgi (COG0166)
Smu-1-2	NC_004350.1	+	1243949	1244087	←hypo **RNA**→ ←COG4947 (COG4947)
Sha-1-39	NC_007168.1	+	2199672	2199818	PTR2 (COG3104)→ **RNA**→ LCB5 (COG1597)→
Ssa-1-7	NC_007350.1	-	2428806	2428635	←DUF867 (pfam05908) **RNA**→ ←Glyoxalase (pfam00903)

cd00118 Lysin domain, found in a variety of enzymes involved in bacterial cell wall degradation.

cd00140 Beta clamp domain.

cd00210 PTS_IIA, PTS system, glucose/sucrose specific IIA subunit.

cd00212 PTS_IIB, PTS system, glucose/sucrose specific IIB subunit.

cd00318 Phosphoglycerate kinase (PGK) is a monomeric enzyme which catalyzes the transfer of the high-energy phosphate group of 1,3-bisphosphoglycerate to ADP, forming ATP and 3-phosphoglycerate.

cd00371 Heavy-metal-associated domain (HMA) is a conserved domain of approximately 30 amino acid residues found in a number of proteins that transport or detoxify heavy metals, for example, the CPx-type heavy metal ATPases and copper chaperones.

cd00433 Cytosol aminopeptidase family, N-terminal and catalytic domains.

cd00563 D-Tyrosyl-tRNAtyr deacylases; a class of tRNA-dependent hydrolases which are capable of hydrolyzing the ester bond of D-Tyrosyl-tRNA reducing the level of cellular D-Tyrosine while recycling the peptidyl-tRNA; found in bacteria and in eukaryotes but not in archea; beta barrel-like fold structure; forms homodimers in which two surface cavities serve as the active site for tRNA binding

cd00644 Hydroxymethylglutaryl-coenzyme A (HMG-CoA) reductase (HMGR), class II, prokaryotic enzyme is a homodimer.

cd00950 Dihydrodipicolinate synthase (DHDPS) is a key enzyme in lysine biosynthesis.

cd01277 HINT (histidine triad nucleotide-binding protein) subgroup: Members of this CD belong to the superfamily of histidine triad hydrolases that act on alpha-phosphate of ribonucleotides.

cd01288 FabZ is a 17kD beta-hydroxyacyl-acyl carrier protein (ACP) dehydratase that primarily catalyzes the dehydration of beta-hydroxyacyl-ACP to trans-2-acyl-ACP, the third step in the elongation phase of the bacterial/ plastid, type II, fatty-acid biosynthesis pathway.

cd01840 yrhL-like subfamily of SGNH-hydrolases, a diverse family of lipases and esterases.

cd01948 EAL domain.

cd02000 Thiamine pyrophosphate (TPP) family, E1 of PDC_ADC_BCADC subfamily, TPP-binding module; composed of proteins similar to the E1 components of the human pyruvate dehydrogenase complex (PDC), the acetoin dehydrogenase complex (ADC) and the branched chain alpha-keto acid dehydrogenase/2-oxoisovalerate dehydrogenase complex (BCADC).

cd02163 Phosphopantetheine adenylyltransferase (PPAT) is an essential enzyme in bacteria that catalyses a rate-limiting step in coenzyme A (CoA) biosynthesis, by transferring an adenylyl group from ATP to 4'-phosphopantetheine, yielding dephospho-CoA (dPCoA).

cd02573 PseudoU_synth_EcTruB: Pseudouridine synthase, Escherichia coli TruB like.

cd03230 This family of ATP-binding proteins belongs to a multisubunit transporter involved in drug resistance (BcrA and DrrA), nodulation, lipid transport, and lantibiotic immunity.

cd03249 MTABC3 (also known as ABCB6) is a mitochondrial ATP-binding cassette protein involved in iron homeostasis and one of four ABC transporters expressed in the mitochondrial inner membrane, the other three being MDL1(ABC7), MDL2, and ATM1.

cd03278 Barmotin is a tight junction-associated protein expressed in rat epithelial cells which is thought to have an important regulatory role in tight junction barrier function.

cd03311 CIMS - Cobalamine-independent methonine synthase, or MetE, C-terminal domain_like.

cd03312 CIMS - Cobalamine-independent methonine synthase, or MetE, N-terminal domain_like.

cd03313 Enolase: Enolases are homodimeric enzymes that catalyse the reversible dehydration of 2-phospho-D-glycerate to phosphoenolpyruvate as part of the glycolytic and gluconeogenesis pathways.

COG0017 Aspartyl/asparaginyl-tRNA synthetases [Translation, ribosomal structure and biogenesis]

COG0018 Arginyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]

COG0022 Pyruvate/2-oxoglutarate dehydrogenase complex, dehydrogenase (E1) component, eukaryotic type, beta subunit [Energy production and conversion]

COG0034 Glutamine phosphoribosylpyrophosphate amidotransferase [Nucleotide transport and metabolism]

COG0036 Pentose-5-phosphate-3-epimerase [Carbohydrate transport and metabolism]

COG0041 Phosphoribosylcarboxyaminoimidazole (NCAIR) mutase [Nucleotide transport and metabolism]

COG0046 Phosphoribosylformylglycinamidine (FGAM) synthase, synthetase domain [Nucleotide transport and metabolism]

COG0047 Phosphoribosylformylglycinamidine (FGAM) synthase, glutamine amidotransferase domain [Nucleotide transport and metabolism]

COG0062 Uncharacterized conserved protein [Function unknown]

COG0079 Histidinol-phosphate/aromatic aminotransferase and cobyric acid decarboxylase [Amino acid transport and metabolism]

COG0085 DNA-directed RNA polymerase, beta subunit/140 kD subunit [Transcription]

COG0095 Lipoate-protein ligase A [Coenzyme metabolism]

COG0097 Ribosomal protein L6P/L9E [Translation, ribosomal structure and biogenesis]

COG0124 Histidyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]

COG0136 Aspartate-semialdehyde dehydrogenase [Amino acid transport and metabolism]

COG0148 Enolase [Carbohydrate transport and metabolism]

COG0149 Triosephosphate isomerase [Carbohydrate transport and metabolism]

COG0151 Phosphoribosylamine-glycine ligase [Nucleotide transport and metabolism]

COG0166 Glucose-6-phosphate isomerase [Carbohydrate transport and metabolism]

COG0171 NAD synthase [Coenzyme metabolism]

COG0190 5,10-methylene-tetrahydrofolate dehydrogenase/Methenyl tetrahydrofolate cyclohydrolase [Coenzyme metabolism]

COG0236 Acyl carrier protein [Lipid metabolism / Secondary metabolites biosynthesis, transport, and catabolism]

COG0253 Diaminopimelate epimerase [Amino acid transport and metabolism]

COG0265 Trypsin-like serine proteases, typically periplasmic, contain C-terminal PDZ domain [Posttranslational modification, protein turnover, chaperones]

COG0267 Ribosomal protein L33 [Translation, ribosomal structure and biogenesis]

COG0296 1,4-alpha-glucan branching enzyme [Carbohydrate transport and metabolism]

COG0306 Phosphate/sulphate permeases [Inorganic ion transport and metabolism]

COG0357 Predicted S-adenosylmethionine-dependent methyltransferase involved in bacterial cell division [Cell envelope biogenesis, outer membrane]

COG0371 Glycerol dehydrogenase and related enzymes [Energy production and conversion]

COG0381 UDP-N-acetylglucosamine 2-epimerase [Cell envelope biogenesis, outer membrane]

COG0396 ABC-type transport system involved in Fe-S cluster assembly, ATPase component [Posttranslational modification, protein turnover, chaperones]

COG0441 Threonyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]

COG0458 Carbamoylphosphate synthase large subunit (split gene in MJ) [Amino acid transport and metabolism / Nucleotide transport and metabolism]

COG0472 UDP-N-acetylmuramyl pentapeptide phosphotransferase/UDP-N-acetylglucosamine-1-phosphate transferase [Cell envelope biogenesis, outer membrane]

COG0475 Kef-type K+ transport systems, membrane components [Inorganic ion transport and metabolism]

COG0484 DnaJ-class molecular chaperone with C-terminal Zn finger domain [Posttranslational modification, protein turnover, chaperones]

COG0488 ATPase components of ABC transporters with duplicated ATPase domains [General function prediction only]

COG0508 Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase (E2) component, and related enzymes [Energy production and conversion]

COG0539 Ribosomal protein S1 [Translation, ribosomal structure and biogenesis]

COG0569 K+ transport systems, NAD-binding component [Inorganic ion transport and metabolism]

COG0588 Phosphoglycerate mutase 1 [Carbohydrate transport and metabolism]

COG0595 Predicted hydrolase of the metallo-beta-lactamase superfamily [General function prediction only]

COG0623 Enoyl-[acyl-carrier-protein] [acyl-carrier-protein]

COG0624 Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase and related deacylases [Amino acid transport and metabolism]

COG0628 Predicted permease [General function prediction only]

COG0632 Holliday junction resolvasome, DNA-binding subunit [DNA replication, recombination, and repair]

COG0653 Preprotein translocase subunit SecA (ATPase, RNA helicase) [Intracellular trafficking and secretion]

COG0656 Aldo/keto reductases, related to diketogulonate reductase [General function prediction only]

COG0657 Esterase/lipase [Lipid metabolism]

COG0691 tmRNA-binding protein [Posttranslational modification, protein turnover, chaperones]

COG0697 Permeases of the drug/metabolite transporter (DMT) superfamily [Carbohydrate transport and metabolism / Amino acid transport and metabolism / General function prediction only]

COG0716 Flavodoxins [Energy production and conversion]

COG0735 Fe2+/Zn2+ uptake regulation proteins [Inorganic ion transport and metabolism]

COG0744 Membrane carboxypeptidase (penicillin-binding protein) [Cell envelope biogenesis, outer membrane]

COG0760 Parvulin-like peptidyl-prolyl isomerase [Posttranslational modification, protein turnover, chaperones]

COG0766 UDP-N-acetylglucosamine enolpyruvyl transferase [Cell envelope biogenesis, outer membrane]

COG0858 Ribosome-binding factor A [Translation, ribosomal structure and biogenesis]

COG1012 NAD-dependent aldehyde dehydrogenases [Energy production and conversion]

COG1063 Threonine dehydrogenase and related Zn-dependent dehydrogenases [Amino acid transport and metabolism / General function prediction only]

COG1079 Uncharacterized ABC-type transport system, permease component [General function prediction only]

COG1116 ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component [Inorganic ion transport and metabolism]

COG1132 ABC-type multidrug transport system, ATPase and permease components [Defense mechanisms]

COG1159 GTPase [General function prediction only]

COG1160 Predicted GTPases [General function prediction only]

COG1191 DNA-directed RNA polymerase specialized sigma subunit [Transcription]

COG1193 Mismatch repair ATPase (MutS family) [DNA replication, recombination, and repair]

COG1198 Primosomal protein N' (replication factor Y) - superfamily II helicase [DNA replication, recombination, and repair]

COG1247 Sortase and related acyltransferases [Cell envelope biogenesis, outer membrane]

COG1249 Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component, and related enzymes [Energy production and conversion]

COG1252 NADH dehydrogenase, FAD-containing subunit [Energy production and conversion]

COG1253 Hemolysins and related proteins containing CBS domains [General function prediction only]

COG1263 Phosphotransferase system IIC components, glucose/maltose/N-acetylglucosamine-specific [Carbohydrate transport and metabolism]

COG1267 Phosphatidylglycerophosphatase A and related proteins [Lipid metabolism]

COG1295 Predicted membrane protein [Function unknown]

COG1299 Phosphotransferase system, fructose-specific IIC component [Carbohydrate transport and metabolism]

COG1368 Phosphoglycerol transferase and related proteins, alkaline phosphatase superfamily [Cell envelope biogenesis, outer membrane]

COG1454 Alcohol dehydrogenase, class IV [Energy production and conversion]

COG1476 Predicted transcriptional regulators [Transcription]

COG1544 Ribosome-associated protein Y (PSrp-1) [Translation, ribosomal structure and biogenesis]

COG1597 Sphingosine kinase and enzymes related to eukaryotic diacylglycerol kinase [Lipid metabolism / General function prediction only]

COG1670 Acetyltransferases, including N-acetylases of ribosomal proteins [Translation, ribosomal structure and biogenesis]

COG1674 DNA segregation ATPase FtsK/SpoIIIE and related proteins [Cell division and chromosome partitioning]

COG1758 DNA-directed RNA polymerase, subunit K/omega [Transcription]

COG1762 Phosphotransferase system mannitol/fructose-specific IIA domain (Ntr-type) [Carbohydrate transport and metabolism / Signal transduction mechanisms]

COG1765 Predicted redox protein, regulator of disulfide bond formation [Posttranslational modification, protein turnover, chaperones]

COG1835 Predicted acyltransferases [Lipid metabolism]

COG1846 Transcriptional regulators [Transcription]

COG1878 Predicted metal-dependent hydrolase [General function prediction only]

COG1887 Putative glycosyl/glycerophosphate transferases involved in teichoic acid biosynthesis TagF/TagB/EpsJ/RodC [Cell envelope biogenesis, outer membrane]

COG1893 Ketopantoate reductase [Coenzyme metabolism]

COG1914 Mn2+ and Fe2+ transporters of the NRAMP family [Inorganic ion transport and metabolism]
COG1982 Arginine/lysine/ornithine decarboxylases [Amino acid transport and metabolism]
COG1983 Putative stress-responsive transcriptional regulator [Transcription / Signal transduction mechanisms]
COG2141 Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase and related flavin-dependent oxidoreductases [Energy production and conversion]
COG2153 Predicted acyltransferase [General function prediction only]
COG2183 Transcriptional accessory protein [Transcription]
COG2188 Transcriptional regulators [Transcription]
COG2217 Cation transport ATPase [Inorganic ion transport and metabolism]
COG2223 Nitrate/nitrite transporter [Inorganic ion transport and metabolism]
COG2226 Methylase involved in ubiquinone/menaquinone biosynthesis [Coenzyme metabolism]
COG2231 Uncharacterized protein related to Endonuclease III [DNA replication, recombination, and repair]
COG2246 Predicted membrane protein [Function unknown]
COG2261 Predicted membrane protein [Function unknown]
COG2270 Permeases of the major facilitator superfamily [General function prediction only]
COG2272 Carboxylesterase type B [Lipid metabolism]
COG2303 Choline dehydrogenase and related flavoproteins [Amino acid transport and metabolism]
COG2374 Predicted extracellular nuclease [General function prediction only]
COG2427 Uncharacterized conserved protein [Function unknown]
COG2514 Predicted ring-cleavage extradiol dioxygenase [General function prediction only]
COG2608 Copper chaperone [Inorganic ion transport and metabolism]
COG2814 Arabinose efflux permease [Carbohydrate transport and metabolism]
COG2826 Transposase and inactivated derivatives, IS30 family [DNA replication, recombination, and repair]
COG2848 Uncharacterized conserved protein [Function unknown]
COG2962 Predicted permeases [General function prediction only]
COG3001 Uncharacterized protein conserved in bacteria [Function unknown]
COG3010 Putative N-acetylmannosamine-6-phosphate epimerase [Carbohydrate transport and metabolism]
COG3104 Dipeptide/tripeptide permease [Amino acid transport and metabolism]
COG3153 Predicted acetyltransferase [General function prediction only]
COG3201 Nicotinamide mononucleotide transporter [Coenzyme metabolism]
COG3221 ABC-type phosphate/phosphonate transport system, periplasmic component [Inorganic ion transport and metabolism]
COG3335 Transposase and inactivated derivatives [DNA replication, recombination, and repair]
COG3382 Uncharacterized conserved protein [Function unknown]
COG3425 3-hydroxy-3-methylglutaryl CoA synthase [Lipid metabolism]
COG3464 Transposase and inactivated derivatives [DNA replication, recombination, and repair]
COG3527 Alpha-acetolactate decarboxylase [Secondary metabolites biosynthesis, transport, and catabolism]
COG3547 Transposase and inactivated derivatives [DNA replication, recombination, and repair]
COG3607 Predicted lactoylglutathione lyase [General function prediction only]
COG3638 ABC-type phosphate/phosphonate transport system, ATPase component [Inorganic ion transport and metabolism]
COG3664 Beta-xylosidase [Carbohydrate transport and metabolism]
COG3764 Sortase (surface protein transpeptidase) [Cell envelope biogenesis, outer membrane]
COG3830 ACT domain-containing protein [Signal transduction mechanisms]
COG3919 Predicted ATP-grasp enzyme [General function prediction only]
COG3942 Surface antigen [General function prediction only]
COG3949 Uncharacterized membrane protein [Function unknown]
COG4108 Peptide chain release factor RF-3 [Translation, ribosomal structure and biogenesis]
COG4120 ABC-type uncharacterized transport system, permease component [General function prediction only]
COG4129 Predicted membrane protein [Function unknown]
COG4193 Beta- N-acetylglucosaminidase [Carbohydrate transport and metabolism]
COG4502 Uncharacterized protein conserved in bacteria [Function unknown]
COG4606 ABC-type enterochelin transport system, permease component [Inorganic ion transport and metabolism]
COG4947 Uncharacterized protein conserved in bacteria [Function unknown]
COG5504 Predicted Zn-dependent protease [Posttranslational modification, protein turnover, chaperones]
COG5549 Predicted Zn-dependent protease [Posttranslational modification, protein turnover, chaperones]
pfam00106 short chain dehydrogenase.
pfam00313 'Cold-shock' DNA-binding domain.
pfam00364 Biotin-requiring enzyme.
pfam00410 Ribosomal protein S8.
pfam00461 Signal peptidase I.

pfam00535	Glycosyl transferase.
pfam00583	Acetyltransferase (GNAT) family.
pfam00665	Integrase core domain.
pfam00903	Glyoxalase/Bleomycin resistance protein/Dioxygenase superfamily.
pfam01476	LysM domain.
pfam01548	Transposase.
pfam01566	Natural resistance-associated macrophage protein.
pfam01717	Methionine synthase, vitamin-B12 independent.
pfam01741	Large-conductance mechanosensitive channel, MscL.
pfam01746	tRNA (Guanine-1)-methyltransferase.
pfam01808	AICARFT/IMPCHase bienzyme.
pfam01812	5-formyltetrahydrofolate cyclo-ligase family.
pfam01926	GTPase of unknown function.
pfam02073	Thermophilic metalloprotease (M29).
pfam02142	MGS-like domain.
pfam02371	Transposase IS116/IS110/IS902 family.
pfam02379	PTS system, Fructose specific IIB subunit.
pfam02452	PemK-like protein.
pfam02595	Glycerate kinase family.
pfam02674	Colicin V production protein.
pfam02687	Predicted permease.
pfam02786	Carbamoyl-phosphate synthase L chain, ATP binding domain.
pfam02787	Carbamoyl-phosphate synthetase large chain, oligomerisation domain.
pfam02810	SEC-C motif. [HC]
pfam02823	ATP synthase, Delta/Epsilon chain, beta-sandwich domain.
pfam03306	Alpha-acetolactate decarboxylase.
pfam03352	Methyladenine glycosylase.
pfam03600	Citrate transporter.
pfam03734	ErfK/YbiS/YcfS/YnhG.
pfam04960	Glutaminase.
pfam05525	Branched-chain amino acid transport protein.
pfam05636	Protein of unknown function (DUF795).
pfam05670	Domain of unknown function (DUF814).
pfam05833	Fibronectin-binding protein A N-terminus (FbpA).
pfam05908	Protein of unknown function (DUF867).
pfam05979	Bacterial protein of unknown function (DUF896).
pfam06028	Bacterial protein of unknown function (DUF915).
pfam06172	Protein of unknown function (DUF985).
pfam06782	Uncharacterised protein family (UPF0236).
smart00062	Bacterial periplasmic substrate-binding proteins; bacterial proteins, eukaryotic ones are in PBPe
smart00342	helix_turn_helix, arabinose operon control protein;

Duplicate sequences: the following putative homologs are not shown in the alignment because their sequences are identical to a homolog already shown: Lre-1-1, Lre-1-14, Lre-1-2, Lre-2-3, Lre-2-4, Sag-1-1, Sag-1-2, Sag-3-1, Sag-3-2, Sag-5-1, Sag-7-1, Sau-10-1, Sau-10-2, Sau-10-3, Sau-10-4, Sau-10-5, Sau-10-6, Sau-11-5, Sau-2-1, Sau-2-2, Sau-2-4, Sau-2-5, Sau-2-6, Sau-3-2, Sau-3-3, Sau-3-4, Sau-3-5, Sau-3-6, Sau-3-7, Sau-3-8, Sau-4-1, Sau-4-2, Sau-4-3, Sau-4-4, Sau-4-5, Sau-4-6, Sau-4-7, Sau-4-8, Sau-5-2, Sau-5-3, Sau-5-4, Sau-6-2, Sau-6-3, Sau-6-4, Sau-7-1, Sau-7-2, Sau-7-3, Sau-7-4, Sau-7-5, Sau-7-6, Sau-7-7, Sau-8-1, Sau-8-2, Sau-8-6, Sau-9-1, Sau-9-2, Sau-9-3, Sau-9-4, Sau-9-5, Sau-9-6, Sau-9-7, Sau-9-8, Sep-1-1, Sha-1-12, Sha-1-13, Sha-1-15, Sha-1-4, Spn-2-2, Spy-1-1, Spy-1-2, Spy-1-3, Spy-1-4, Spy-11-2, Spy-12-1, Spy-12-2, Spy-2-1, Spy-2-2, Spy-2-3, Spy-3-1, Spy-3-2, Spy-5-1, Spy-5-3, Spy-6-1, Spy-6-2, Spy-6-3, Spy-7-1, Spy-7-2, Spy-7-3, Spy-9-1, Ssu-1-12, Sth-1-1

alignment positions 1 · · · 160

```
Ssu-1-1    UCGA..CGCAG..UAGGCGUCA.GC.AAAUC.ACU........CUUCGCAUUGCU...................................................C..GA.GUU.CUUUGGACGCCCUUGCGA.........GGGAA............
Spy-8-1    UCUG..CGUCG..UAGGCGUCU.GAUGC................UAAA..................................................................GC.ACCUAGUCGCCCUGACGU.........GGGUU............
Spy-4-1    UCUG..CGUCG..UAGGCGUCU.GAUGC................UAAA..................................................................GC.ACCUAGUCGCCCUGACGU.........GGGUU............
Spy-11-1   UCUG..CGUCG..UAGGCGUCU.GAUGC................UAAA..................................................................GC.ACCUAGUCGCCCUGACGU.........AGGUU............
Spy-10-1   UCUG..CGUCG..UAGGCGUCU.GAUGC................UAAA..................................................................GC.ACCUAGUCGCCCUGACGU.........GGGUU............
Sag-8-1    UCUG..CGUCG..UAGGCGUCU.GGUGC................UAAA..................................................................GC.ACCUAGUCGCCCUGACGU.........GGGUG............
Sag-6-1    UUUG..CGUCG..UAGGCGUCU.GCUGC................UAAA..................................................................GC.ACCUAGUCGCCCUGACGU.........GGGUG............
Sag-4-1    UCUG..CGUCG..UAGGCGUCU.GCUGC................UAAA..................................................................GC.ACCUAGUCGCCCUGACGU.........GGGUG............
Sag-2-1    UCUG..CGUCG..UAGGCGUCU.GCUGC................UAAA..................................................................GC.ACCUAGUCGCCCUGACGU.........GGGUA............

           ......<<<<<....<<<<.<<.<..<<<<<.<.........................................................................>..>>.>>>....>>>.>>>>.>>>>>...........<<<.............
           ......???2?....2??2.??.2..?????.?.........................................................................?..??.???....2??.2??2.?2???...........?2?.............
           ••••--YGCAG--URG•UGAYU-GR∘Y••••-•••∘∘∘○----•••••••R••Y•---------------------------------------------------Y--••-••R-•YCUARUCA•CYUUGC•R---------GGGYR------------
```

```
Spy-8-2    UCUUAGCGUCG..UAGACGUCU.GAACUUUC.CUC........AGUUUUUGUUUC...................................................G..GA.AAG.UCCUAGUCGUCCUGACGU.........GGGUG............
Spy-9-2    UCUUAGCGUCG..UAGACGUCU.GAACUUUC.CUC........AAUUUUUGUUUC...................................................G..GA.AAG.UCCUAGUCGUCCUGACGU.........GGGUG............
Spy-5-2    UCUUAGCAUCG..UAGACGUCU.GAACUUUC.CUC........AAUUUUUGUUUC...................................................G..GA.AAG.UCCUAGUCGUCCUGACGU.........GGGUG............
Spy-10-2   UCUUAGCGUCG..UAGACGUCU.GAACUUUC.UUC........AAUUUUUGUUUC...................................................G..GA.AAG.UCCUAGUCGUCCUGACGU.........GGGUG............
Spy-4-2    UCUUAGCAUCG..UAGACGUCU.GAACUUUC.UUC........AAUUUUUGUUUC...................................................G..GA.AAG.UCCUAGUCGUCCUGACGU.........GGGUG............
Spy-8-3    UUAG..CGUCG..UAGACGUCU.GAACUUUC.CUC........AGUUUUUGUUUC...................................................G..GA.AAG.UCCUAGUCGUCCUGACGU.........GGGUG............
Ssu-1-2    UGUU..CGUCG..UAAACGUCU.GAACUUUC.UCU........GGCCUUCGGUCA...................................................G..GA.AAG.UCCUAGUCGUUCUGACGA.........GGGUG............
Ssu-1-3    CCAA..CGUCG..UAAACGUCU.GGCCUUUC.UUC........GACGAUUUGUCA...................................................A..GA.AAG.UCCUAGUCGUUCUGACGA.........GGGUG............
Smu-1-1    UAUG..C..UG..UAGAUGUCU.GAACUUUC.UAC........GCCAGUAGGCUA...................................................A..GA.AAG.UCCUAGUCGUCCUGAUG..........GGGUG............
Ssu-1-4    ACUG..CGUCG..UAGACAACU.GUACUUUC.UUCGGUU....UC.....ACCUG...................................................U..GA.AAA.UACUAGUUGUCUUGACGA.........GGGUG............
Ssu-1-5    CAUA..CGCAG..UAGCUGAAU.GGCAGUUC.ACU........UUGCUACGCUUG...................................................U..GA.CCA.ACCUAUUCAGCCUUGCGG.........GGGUG............
Spy-8-4    UUUG..CGUCG..UAGCUGACU.GGAUGUUC.ACUCCUU....UU......GUCG...................................................U..GA.CCA.ACCUAGUCAGCCUGACGG.........AGGUG............
Spy-12-3   UUUG..CAUCG..UAGCUGACU.GGAUGUUC.ACUCCUU....UU......GUCG...................................................U..GA.CCA.ACCUAGUCAGCCUGACGG.........AGGUG............
Spy-3-3    UUUG..CGUCG..UAGCUGACU.GGAUGUUC.ACUCCUU....UU......GUCG...................................................U..GA.CCA.ACCUAGUCAGCCUGACGG.........AGGUG............
Spy-11-3   UUUG..CGUCG..UAGCUGACU.GGAUGUCC.ACUCCUU....UU......GUCG...................................................U..GA.CCA.ACCUAGUCAGCCUGACGG.........AGGUG............
Spy-9-3    UUUG..CGUCG..UAGCUGACU.GGAUGUUC.ACUCCUU....UU......GUCG...................................................U..GA.CCA.ACCUAGUCAGCCUGACGG.........GGGUG............
Spy-4-3    UUUG..CGUCG..UAGCUGACU.GGAUGUUC.ACUCCUU....UU......GUCG...................................................U..GA.CCA.AGCUAGUCAGCCUGACGG.........AGGUG............
Spy-10-3   UUUG..CGUCG..UAGCUGACU.GGAUGUUC.ACUCCUU....UU......GUCG...................................................U..GA.CCA.ACCUAGUCAGCCUGACGG.........AGGUG............
Sau-6-1    UUGA..CGCAG..UGGUUGACU.GGUCAUCC.AAU........GG.....AGAAU...................................................U..GC.CUG.ACCUAGUCAACUUUGCGG.........GGGAA............
Sau-3-1    UUGA..CGCAU..UGGUUGACU.GGUCAUCC.AAU........GG.....AGAAU...................................................U..GU.CUG.ACCUAGUCAACUUUGCGG.........GGGAA............
Sau-11-1   UUGA..CGCAG..UGGUUGACU.GGUCAUCC.AAU........GG.....AGAAU...................................................U..GU.CUG.ACCUAGUCAACUUUGCGG.........GGGAA............
Sau-5-1    UUGA..CGCAG..UGGUUGACU.GGUCAUCC.AAU........GG.....AGAAA...................................................U..GC.CUG.ACCUAGUCAACUUUGCGG.........GGGAA............
Sau-11-2   UAUU..GGCAG..UAGUUGACU.GAACGAAA.AUGCGCU....UGUAACAAGCUU...................................................U..UU.UCAAUUCUAGUCAACCUUGCUG.........GGGUG............
Sau-1-1    UAUU..GGCAG..UAGUUGACU.GAACGAAA.AUGCGCU....UGUAACAAGCUU...................................................U..UU.UCAAUUCUAGUCAACCUUGCCG.........GGGUG............
Sau-1-2    UAUU..GGCAG..UAGUUGACU.GAACGAAA.AUGCGCU....UGUAACAAGCUU...................................................U..UU.UCAAUUCUAGUCAACCUUGCCG.........GGGUG............
Sau-6-5    UAUU..GGCAG..UAGUUGACU.GAACGAAA.AUGCGCU....UGUAACAAGCUU...................................................U..UU.UCAAUUCUAGUCAACCUUGCCG.........GGGUG............
Sau-8-3    UAUU..GGCAG..UAGUUGACU.GAACGAAA.AUGCGCU....UGUAACAAGCUU...................................................U..UU.UCAAUUCUAGUCAACCUUGCCG.........GGGUG............
Sau-8-4    UAUU..GGCAG..UAGUUGACU.GAAUGAAA.AUGCGCU....UGUAACAAGCUU...................................................U..UU.UCAAUUCUAGUCAACCUUGCCG.........GGGUG............
Sau-8-5    UAUU..GGCAG..UAGUUGACU.GAACGAAA.AUGCGCU....UGUAACAAGCUU...................................................U..UU.UCAAUUCUAGUCAACCUUGCCG.........GGGUG............
Sau-5-5    UAUU..GGCAG..UAGUUGACU.GAACGAAA.AUGCACU....UGUAACAAGCUU...................................................U..UU.UCAAUUCUAGUCAACCUUGCCG.........GGGUG............
Sau-5-6    UAUU..GGCAG..UAGUUGACU.GAACGAAA.AUGCGCU....UGUAACAAGCUU...................................................U..UU.UCAAUUCUAGUCAACCUUGCCG.........GGGUG............
Sau-2-3    UAAU..GGCAG..UAGUUGAGU.GAACGAAA.AUUCGCU....UGUAACAAGCUU...................................................U..UU.UCAAUUCUAGUCAACCUUGCCG.........GGGUG............
Sau-11-3   UAAU..GGCAG..UAGUUGACU.GAACGAAA.AUUCGCU....UGUAACAAGCUU...................................................U..UU.UCAAUUCUAGUCAACCUUGCCG.........GGGUG............
Sau-8-7    UAAU..GGCAG..UAGUUGACU.GAACGAAA.AUGCGCU....UGUAACAAGCUU...................................................U..UU.UCAAUUCUAGUCAACUUUGCCG.........GGGUG............
Sau-5-7    UAUU..GGCAG..UAGUUGACU.GAACGAAA.AUACGCU....UGUAACAAGCUU...................................................U..UU.UCAAUUCUAGUCAACCUUGCCG.........GGGCG............
Sau-11-4   UAUU..GGCAG..UAGUUGACU.GAAUGAAA.AUACGCU....UGUAACAAGCUU...................................................U..UU.UCAAUUCUAGUCAACCUUGCCG.........GGGUG............
Sau-5-8    UAUU..GGCAG..UAGUUGACU.GAAUGAAA.AUGCGCU....UGCAACAAGCUU...................................................U..UU.UCAACUCUAGUCAACCUUGCCG.........GGGUG............
Sau-6-6    UAUU..GGCAG..UAGUUGACU.GAAUGAAA.AUACGCUU...GUAACAAGCUUU...................................................U..UU.UCAAUUCUAGUCAACCUUGCCG.........GGGUG............
Sha-1-1    UUUU..GUCAG..UAGCUGACU.GAUUUGAAAAUGCGCU....UGAAUCAAGCUU...................................................U.UUU.CAA.ACCUAGUCACCCUUGCCG.........GGGUA............
Sha-1-2    UUUU..GUCAG..UAGCUGACU.GAUUUGAAAAUGCGCU....UGAAUCAAGCUU...................................................U.UUU.CAA.ACCUAGUCAUCCUUGCCG.........GGGUA............
Sha-1-3    UUUU..GUCAG..UAGCUGACU.GAUUUGAAAAUGCGCU....UGAAUCAAGCUU...................................................U.UUU.CAA.ACCUAGUCAUCCUUGCCG.........GGGUA............
Sha-1-5    UUUU..GUCAG..UAGCUGACU.GAUUUGAAAAUGCGCU....UGAAUCAAGCUU...................................................U.UUU.CAA.ACCUAGUCAUCCUUGCCG.........GGGUA............
Sha-1-6    UUUU..GUCAG..UAGCUGACU.GAUUUGAAAAUGCGCU....UGAAUCAAGCUU...................................................U.UUU.CAA.ACCUAGUCAUCCUUGCCG.........UGGUA............
Sha-1-7    UUUU..GUCAG..UUGCUGACU.GAUUUGAAAAUGCGCU....UGAAUCAAGCUU...................................................U.UUU.CAA.ACCUAGUCAUCCUUUCCG.........GGGUA............
Ssa-1-1    UUUA..GGCAG..UAGAUGACU.GAAGUGAAAAUGCGCU....UAUAUCAAGCUA...................................................U.UUU.CAA.UCCUAGUCAUCCUUGCCG.........GGGUG............
Ssa-1-2    UUUA..GGCAG..UAGAUGACU.GAAUUGAAAAUACGCU....UAUAUCAAGCUU...................................................U.UUU.CAA.UCCUAGUCAUCCUUGCCG.........GGGUU............
Ssa-1-3    UUUA..GGCAG..UAGAUGACU.GAAUUAAAAAUUCGCU....UAUAUCAAGCUU...................................................U.UUU.CAA.UCCUAGUCAUCCUUGCCG.........AGGUA............
Ssa-1-4    AAAA..GGCAG..UAGAUGACU.GGAUUGAA.UUGCGCU....UAUAUUAAGCUU...................................................A..UU.CAA.CCCUAGUCAUCCUUGCCG.........GGGUG............
Ssa-1-5    AUUA..AGCAG..UAGAUGACU.GAAUUGAAAAUACCCU....UGUAUCAAGCUU...................................................U.UUU.CAA.UUCUAGUCAUCCUUGCCUUGAUUGCGUAGGUA............
Sha-1-8    GAUU..CGCAG..UAGCUGACU.GAACUGAGAAGACACU....UAAAUCAAGUUU...................................................U.CCU.CAG.UUCUGGUCAUUCUUGUAG.........GGGUG............
Sha-1-9    AAUU..CACAG..UAGCUGACU.GAACUGAGAAGGCGCU....UAAAUCAAGCUU...................................................U.UCU.CAG.UUCUAGUCAUCCUUGCAG.........GGUUG............

           ......<<<<<....<<<<.<<.<..<<<<<.<.........................................................................>..>>.>>>....>>>.>>>>.>>>>>...........<<<.............
           ......???2?....2??2.??.2..?????.?.........................................................................?..??.???....2??.2??2.?2???...........?2?.............
           ●●●●--YGCAG--URG●UGAYU-GR○Y●●●●-●●●○○○○----●●●●●●●R●●Y●---------------------------------------------------Y--●●-●●R-●YCUARUCA●CYUUGC●R---------GGGYR------------
```

```
Sha-1-10  AAUU..CGCAG..UAGCUGACU.GAACUGAAAAUGCGCU....UAACUCAAGCUU..................................................U.UUU.CAG.UUCUAGUCAUCCUUGCGG.........GGGUG............
Sha-1-11  AAUU..CGCAG..UAGCUGACU.GAACUGAAAAUGCGCU....UAAAUCAAGCUU..................................................U.UUU.CAG.UUCUAGUCAUCCUUGCGG.........GGGUG............
Sha-1-14  AAUU..CGCAG..UAGCUGACU.GAACUGAGAAGGCGCU....UAAAUCAAGCUU..................................................U.UCU.CAG.UUCUAGUCAUCCUUGCGG.........GGGUG............
Sha-1-16  AAUU..CGCAG..UAGCUGACU.GAACUGAAAAAGCACA....UAAAUCAAGCUU..................................................U.UCU.CAU.UUCUAGUUAUCCUUAUGG.........GGGUG............
Sha-1-17  AAUU..UACAG..UAGCUGACU.GAAUUGAGAAGGCGUU....UAAAUCAAGCUU..................................................U.UCU.UAG.UUCUAGUUAUCCUGCGGGGG.......GAGUG............
Sha-1-18  UAUU..GGCAG..UAGUUGAUU.AAUUUGAAAAUGCGCU....UAAAUCAUGCUU..................................................U.UUU.CAA.ACCUAGUCAUCCUUGCCG.........GGAUG............
Sha-1-19  AACU..UGCAU..AAGCUGACU.GAACUGAGAAGGCGCUC...UAUAUGAAGUGU..................................................U.UCC.CAG.UUCUUUCCAACUUUG.AG.........GGCUG............
Sha-1-20  AAUU..CGCAG..UAGCUGACU.GAACUGAG.AUAGCUCU...UAAAUCUAGCUU..................................................U..CU.CAG.UUCUUGUGUUCCUUGCGG.........GGGUA............
Sha-1-21  AAUU..CGUAG..UAGCUGACU.GAACUGAA.AAGGUGCUU..AAAAUCAAGCUU..................................................U..UU.CAG.UUGUAGUCAUCCUUGCGA.........GUGAG............
Sha-1-22  AAUU..CGCAG..UAGCUGACU.GAACUGAGAAGGCGCC....UAAAUUAAGCUU..................................................U.UCU.CAG.UUCUAGUCAUCCUUGCGG.........GCGGG............
Sha-1-23  AAUU..CUCAG..UAGCUGACU.GAACUGAAAAGGAGCU....UAAAUCAAGCUU..................................................U.UUC.CAG.UUCUAGUCAUCCUUGCGG.........GCGGG............
Sha-1-24  AAUU..CGCAG..UAGCUGACU.GAACUGAAAAGGCGCU....UAAAUCAAGCUU..................................................U.UUC.CAG.UACUAGUCAUCCUUGUGG.........GCGGG............
Sha-1-25  AAUU..UGCAG...AUAUAACU.AAGCUGAGAAGGUAAU....AAUAAUAAAAUA..................................................C.ACU.CAG.UUAUAGUUAUCCUUAUAA.........GGGUG............
Sau-6-7   UAAU..GGCAG..UAGUUGACU.GAACGAAA.AUUCGCU....UGUAACAAGCUU..................................................U..UU.UCAAUUCUAGUCAACCUUGCCG.........GCGGG............
Sau-11-6  UAUU..GGCAG..UAGUUGACU.GAACGAAA.AUGCGCU....UGUAACAAGCUU..................................................U..UU.UCAAUUCUAGUCAACCUUGCCG.........GCGGG............
Sau-8-8   UAUU..GGCAG..UAGUUGACU.GAACGAAA.AUGCGCU....UGUAACAAGCUU..................................................U..UU.UCAAUUCUAGUCAACCUUGCCG.........GCGGG............
Sau-6-8   UAUU..GGCAG..UAGUUGACU.GAACGAAA.AUGCGCU....UGUAACAAGCUU..................................................U..UU.UCAAUUCUAGUCAACCUUGCCG.........GCGGG............
Sau-7-8   UAUU..GGCAG..UAGUUGACU.GAACGAAA.AUGCGCU....UGUAACAAGCUU..................................................U..UU.UCAAUUCUAGUCAACCUUGCCG.........GGGGG............
Sau-5-9   UAUU..GGCAG..UAGUUGACU.GAACGAAA.AUGCGCU....UGUAACAAGCUU..................................................U..UU.UCAAUUCUAGUCAACCUUGCCG.........GUGGG............
Sau-5-10  UAUU..GGCAG..UAGUUGACU.GAACGAAA.AUACGCU....UGUAACAAGCUU..................................................U..UU.UCAACUCUAGUCAACCUUGCCG.........GCGGG............
Sha-1-26  AAUA..UGCAG..UAGAUGACU.GAAUUGAG.AAGGCACU...UAUUCUAAACUU..................................................A..CU.UAG.UUCAAGUCAUCCUA..AG.........GGGUG............
Sha-1-27  UGUU..GGCAG..UCGAUGACU.GAAUUGAA.AAGGUGU....UGUCUCAAGGU...................................................U..UU.CAA.UUCUAAUCCUCUUUGCCA.........UGGCG............
Sha-1-28  AAAG..CGCAG..UAGCUGACA.AAUCUGAG.AGACUGGC...CACACAAUGGCU..................................................U..CU.CAG.CUCUAGUCAUCCAUGCAG.........GAGUG............
Sha-1-29  UUAU..GGCAG..UGGCUGACU.GAACCGACUCAUCGCUU...UAGGAAAGCUC...................................................G.CUU.CAG.UUCUAGUCAUCUGUGCCG.........GGGCC............
Ssa-1-6   GUUA..CACAG..UAGCUGACU.GAAGCUAA.AUGCAC.....ACAAGCAUGUUU..................................................U..UU.AGCAUUCUAGUCAGCCUUGUGG.........GGGCG............
Sha-1-30  CAGU..CGCAA..UUGUUGCAU.GUAUUGAU.AAGAAGUU...CAACUAGUAGUU..................................................U..UU.CAG.GUCUAGUCAUCUUUGUGG.........GGGUG............
Sep-2-1   UAUA..GACAG..UAGAUGACU.GAAUUAAA.AUGCUCU....UGUAUCAAGAUU.................................................CU..UU.UAA.UUCUAGUCAUUAUUGUUG.........GUGCA............
Sha-1-31  UAUU..GGCAG..AAGUUGACC.GAUUUAAA.AAUGCGC....UUUUAACAAGCUUU................................................U..UU.UAA.UUCUAAUCAACCUUGCCU..........................
Sha-1-32  UAUU..GGCAG..UAGUUGACU.GAAUUAAA.AUGCACUU...GUAACAAGCAUCU.................................................U..UU.UAA.UUCUAGUCAACCUUACCU..........................
Sha-1-33  UAUU..GGCAG..UAGUUGACU.GAUUUAAAAAUGCACU....UGUAACAAGCUU..................................................U..UU.UAA.UUCUAGUCAUCCUUGCCU.........G.GUG............
Lre-2-1   AAUA..CGCAG..UACAACAAU.GGCCUCUA.ACGUCCGG...UAACACCGAACG..................................................U..UA.GAG.GCCUAUUGUUGCUUGCGG..........................
Lre-2-2   AAUA..CGCAG..UACAACAAU.GGCCUCUA.ACGUCCG....GCAAAACCGAACG.................................................U..UA.GAG.GCCUAUUGUUGCUUGCGG..........................
Lre-1-3   AAUA..CGCAG..UACAACAAU.GGCCUCUA.ACGUCCG....GCAACACCGAACG.................................................U..UA.GAG.GCCUAUUGUUGCUUGCGA..........................
Sha-1-34  UAUU..GGCAG..UAGUUGACU.GAAUUAAA.AAACGCA....CUUGUAAAAAUU..................................................U..UU.UAA.UUCUAGUCAACCUUACCA.........GGGGA............
Sha-1-35  AAUU..UGCAG..UAGCCGACU.UGAUAGUU.AAACGC.....UAGAGUUGCGUU..................................................U..UU.CGG.CUCAAGUCAUCCUUACAG.........GGGCA............
Spn-2-1   UUGA..CGCAG..UUGCUGCCU.GGUUUGAA.AUGCGC.....UCUAGCAAGCUU..................................................U..UU.CAA.ACCUAGUCAUCGUUGCAG.........GGGUA............
Sha-1-36  UAUU..GGUAG..UAGUUGACU.AAAUUAAA.ACUACACU...UAUAACAAGCAU..................................................U..UU.UAA.UUUUAGUCAAUCUUGCCA.........GGGCA............
Sha-1-37  UAUU..GGCAG..UAGUUGACU.GGCUAAAA.AUGCACU....UGUAACAAUCUU..................................................U..UU.UAA.UUCUAGUCAACCUUGCCG.........GAGUG............
Lre-2-5   AAAA..CGCAG..UAGCUGGCU.GG.AUUCU.GAACGUUGA..UAAAUCAACGUC..................................................C..AG.AAC.ACCUAGCCAUCCUUGCGG.........GGGUG............
Lre-2-6   GUAG..CGCGG..UAGCUGGCU.GG.AUUCC.GAACGUUGA..UAAAUCAACGUC..................................................C..GG.AAC.GCCUAGCCAUCCUCGCGG.........GGGUG............
Spn-1-1   UUGA..CGCAG..UUGCUGCAU.GGUUUGAA.AUGCGCUC...UAGCAAUCUU....................................................U..UU.CAA.ACCUAGUCAUCGUUGCAG.........GGGUA............
Lre-1-4   AAUA..CGCAG..UACAACAAU.GGCCUCUA.ACGUCCGG...CAACACCGAACG..................................................U..UA.GAG.GCCUAUUGUUGCUUGCGG.........GAGCU............
Lre-1-5   AAUA..CGCAG..UACAACAAU.GGCCUCCA.ACGUCCGG...CAAAACCGAACA..................................................U..UG.GAG.GCCUAUUGUUGCUUGCGG.........GGGCU............
Lre-2-7   AAUA..CGCAG..UACAACAAU.GGCCUCUA.ACGUCCGG...CAAAACCGAACG..................................................U..UA.GAG.GCCUAUUGUUGCUUGCGG.........GGGCU............
Lre-1-6   AAUA..CGCAG..UACAACAAU.GGCCUCUA.AUGUCCGG...UAAAACCGAACG..................................................U..UA.GAG.GCCUAUUGUUGCUUGCGG.........GGGCU............
Lre-2-8   AAUA..CGCAG..UACAACAAUAGGCCUCUA.ACGUCCGG...CAAAACCGAGCG..................................................U..UA.GAG.GCCUAUUGUUGCUUGCGG.........GGGCU............
Lre-2-9   AAUA..CGCAG..UACAACAAU.GGCCUCUA.AUGUCCG....GCAAACCGAACG..................................................U..UA.GAG.GCCUAUUGUUGCUUGCGG.........GGGCU............
Lre-2-10  AAUA..CGCAG..UACAACAAU.GGCCUCUA.AUGUCCGG...CAAAGCCGAACG..................................................U..UA.GAG.GCCUAUUGUUGUUUGCGG.........GGGCU............
Lre-2-11  AAUA..CGCAG..UACAAUAAU.GGCCUCCA.ACGUCCGG...CAAAACCGAAUA..................................................U..UG.GAG.GCCUAUUAUUGCUUGCGG.........GGGCU............
Lre-1-7   AAUA..CGCAG..UACAACAAU.GGCCUCCA.ACGUCCGGCUCAAAAGCCGAACA..................................................U..UG.GAG.GCCUAUUGUUGCUUGCGG.........GGGCU............
Lre-1-8   AAUA..CGCAG..UACAACAAU.GGCCUCCA.ACGUCCGGCAAAAAAGCCGAACA..................................................U..UG.GAG.GCCUAUUGUUGCUUGCGG.........GGGCU............

          ......<<<<<....<<<<.<<.<..<<<<<.<........................................................................>..>>.>>>....>>>.>>>>.>>>>>...........<<<.............
          ......???2?....2??2.??.2..?????.?........................................................................?..??.???....2??.2??2.?2???...........?2?.............
          ●●●●--YGCAG--URG●UGAYU-GR◦Y●●●●-●●●◦◦◦○----●●●●●●●R●●Y●--------------------------------------------------Y--●●-●●R-●YCUARUCA●CYUUGC●R---------GGGYR------------
```

```
Lre-1-9   AAUA..CGCAG..UACAACAAU.GGCCUCUA.ACGUCC.....GGAAACCGAACG...................................................C..UA.AAG.GCCUAUUGUUGCUUGCGG.........AGGCU............
Lre-2-12  AUAA..CGCAG..UACAACAAU.GGCCUCCA.ACGCCCGG...UAAAACCGAACG...................................................U..UG.GAG.GCCUAUUGUUGCUUGCGG.........GGGCU............
Lre-1-10  AUGG..CGCAG..UACACCAAU.GGCCUCCA.ACGCCCGG...UAAAACCGAAUG...................................................U..UG.GAG.GCCUAUUGUUGCUUGCGG.........GGGCU............
Lre-1-11  AUGG..CGCAG..UACACCAAU.GGCCUCCA.ACGCCCGG...UAAAACCGAACG...................................................U..UG.GAG.GCCUAUUGUUGCUUGCGG.........GGGCU............
Lre-2-13  AUGG..CGCAG..UACAACAAU.GGCCUCCA.ACGCCCGG...UAAAACCGAACG...................................................U..UG.GAG.GCCUACUGUUGCUUGCGG.........GGGCU............
Lre-2-14  AUGG..CGCAG..UACACCAAU.GGCCUCCA.AUGCCCGG...UAAAACCGAACG...................................................U..UG.GAG.GCCUAUUGUUGCUUGCGG.........GGGCU............
Lre-2-15  ACAG..CGCAG..UACAACAAU.GGCCUUCA.GUGUCCGG...UAAAGCCGAACA...................................................C..UG.AAG.GCCUAUUGUUGUUUGCGG.........GGGCU............
Lre-1-12  ACAG..UGCAG..UACAACAAU.GGCCUUCA.GUGUCCGG...UAAAGCCGAACG...................................................C..UG.AAG.GCCUAUUGUUGUUUGCGG.........GGGCU............
Lre-2-16  ACAG..CGCAG..UACAACAAU.GGCCUUCA.GCGUCCGG...CAAAGUCGAACA...................................................C..UG.AAG.GCCUAUUGUUGCUUGCGG.........GGGCU............
Lre-1-13  ACAG..CGCAG..UACAACAAU.GGCCUUCA.GCGUCUGG...CAAUAUCGAACG...................................................C..UG.AAG.GCCUAUUGUUGCUUGCGG.........GAGCU............
Lre-2-17  AAUA..CGCAG..UACAACAAU.GGCCUCUA.AUGUCCGG...CAAAACCGAACG...................................................U..UA.GAG.GCCUAUUGUUGCUUGCGA.........GGGCUCCCAGAGGCUAG
Lre-2-18  AAUA..CGCAG..UACAACAAU.GGCCUCUA.ACGUCCGG...CAAAACCGAACG...................................................U..UA.GAG.GCCUAUUGUUGCUUGCGG.........GGGCUCCCAGAGGCUAG
Lre-1-15  AAUA..CGCAG..UACAACAAU.GGCCUCUA.ACGUCCGG...CAAAACCGAACG...................................................U..UA.GAG.GCCUAUUGUUGCUUGCGG.........GGGCUCCCAGAGGCUAG
Lre-1-16  AAUA..CGCAG..UACAACAAU.GGCCUCCA.AUGUCCGG...CAAAACCGAACA...................................................U..UG.GAG.GCCUAUUGUUGCUUGCGG.........GGGCUCCCAGAGGCUAG
Lre-1-17  AUAA..CGUAG..UACAACAAU.GGCCUCCA.UCGCUCGG...UAAAACCGAACG...................................................U..UG.GAG.GCCUAUUGUUGCUUGCGG.........GAGCU............
Lre-2-19  AUAA..CGCAG..UACAACAAU.GGCCUCCA.UCGCUCGG...UAAAACCGAACG...................................................U..UG.GAG.GCCUAUUGUUGCUUGCGG.........GAGCU............
Sha-1-38  AAUA..CGUAG..UGACUGACU.GAACUAAG.AAAGCAAU...AUAUAAAGCAAU...................................................U..CU.UAG.UUCUAGUCAUUCAUACUA.........AAGUG............
Sep-2-2   UAUG..UUCAUCCGAGUUGACU.AAGUUAAAGAUGCACGCU..UAUAUUAAAUAA...................................................U..UU.UAA.AUUGAGUCAGCUUUG.AG.........GGGUG............
Ssu-1-6   UCAG..CGCAG..UGGUUGAUU.GG.CAGAU.UUGUUCGUG..UUUUACACUCCA...................................................A..AU.CUG.ACCUAAUCAACUGUGCGG.........GGGCG............
Ssu-1-7   UCAG..CGCAG..UGGUUGAUU.GG.CAGAU.UUGUUCGUG..UUUUACACUCCA...................................................A..AU.CUG.ACCUAAUCAACUGUGCGG.........AGGCG............
Ssu-1-8   UCAG..CGCAG..UGGUUGAUU.GG.CAGAU.UUGUUCGUG..UUUUGCACUCCA...................................................A..AU.CUG.ACCUAAUCAACUGUGCGG.........GGGUG............
Ssu-1-9   UCAG..CGCAG..UGGUUGAUU.GG.CAGAU.UUGUUCGUG..UUUUGCACUCCA...................................................A..AU.CUG.GCCUAAUCAACUGUGCGG.........GGGUG............
Ssu-1-10  UCAG..CGCAG..UGGUUGAUU.GG.CAGAU.UUGUUCGUG..UUUUGCACUCCA...................................................A..AU.CUA.GCCUAAUCAACUGUGCGG.........GGGUG............
Ssu-1-11  UCAG..CGCAG..UGGUUGAUU.GG.CAGAU.UUGUUCGUG..UUUUGCACUCCA...................................................A..AU.CUG.GCCUAAUCAACUGUGCGG.........GGGUG............
Ssu-1-13  UCAG..CGCAG..UGGUUGAUU.GG.CAGAU.UUGUUCGUG..UUUCACACUCCA...................................................A..AU.CUG.ACCUAAUCAACUGUGCGGG........GGGUG............
Ssu-1-14  UCAG..CGCAG..UGGUUGAUU.GG.CAGAU.UUGUUCGUG..UUUCAUACUCCA...................................................A..AU.CUG.ACCUAAUCAACUGUGCGG.........GGGUG............
Ssu-1-15  UCAG..CGCAG..UGGUUGAUU.GG.CAGAU.UUGUUCGUG..UUUUACACUCCA...................................................A..AU.CUG.ACCUAAUCAACUGUGCGG.........GGGUG............
Ssu-1-16  UCAA..CGUG.....GUUGAUU.GG.CAGAU.UUGUUCGUG..UUUCACACUCCA...................................................A..AU.CUG.ACCUAAUCAACUGUGCGG.........GGGUG............
Ssu-1-17  UCAG..CGCAG..UGGUUGAUU.GG.CAGAU.UUAUUCGUGU.UACCACACUCUA...................................................A..AU.CUG.ACCUAAUCAACUGUGCGG.........GGGCG............
Ssu-1-18  UCAG..CGCAG..UGGUUGAUU.GG.CAGAU.UUGUUCGCG..UUUCACACUCCA...................................................A..AU.CUG.GCCUAAUCAACUGUGCGG.........GGGUG............
Ssu-1-19  UCAG..CGCAG..UGGUUGAUU.GG.CAGAU.UUGUUCGUG..UUCCACACUCCA...................................................A..AU.CCG.ACCUAAUCAACUGUGCGG.........GGGU.............
Ssu-1-20  UCAG..CGCAG..UGGUUGAUU.GGAAGAUU.UGUUCGU....GUUUUACACUCC...................................................A..CA.UCU.UCCUAAUCAACUGUGCGG.........GGGUG............
Ssu-1-21  UCAG..CGCAG..UGGUUGAUU.GG.CAGAU.UUGUUCGUG..UUUUACACUCCA...................................................A..AU.CUG.ACCUAAUCAACUGUGCGA.........GGGGG............
Lre-2-20  AAAC..CACAG..UAGAUG........................CUUAGC..........................................................................CAUCCUUGCGG.........GGGUG............
Lre-2-21  AAUG..CUGAA..UAGA..........................ACACCUAACCAU..............................................................CC.....AUCCUUC.GG.........GGGUG............
Sag-5-2   AAGA..CACAG..UGGUUGAGU.GGGUUCUA.AUUCGUUG...AUAAAUCAAGUU...................................................U..UA.CAG.CCCUACUCAACUGUGCGG.........AGGUG............
Sag-2-2   AAAA..CGUAG..GGGUUGAGU.GGGCUCUA.AUUCGCUG...AUAAAUCAACUC...................................................U..UA.CAG.CCCCACUCAACUGUGCG..........GAGGG............
Ssu-1-22  AUAA..CGCAG..UUGACAUCU.GGCAAAAG.GCUU.......CGUCAAAACUCG...................................................C..CU.UUU.ACCUAGUUGUCAUUGCGG.........AGGUC............
Sth-3-1   AUGG..CGCAG..UGGUUAAUU.GGAA..AC.AUUCAU.....GUUUACAACAUU...................................................U..GU.ACU.UCCUAAUCAACUGUUCGG.........GGGCG............
Sth-2-1   AUGG..CGCAG..UGGUUAAUU.GGAAACAU.UCAU.......GUUUACAACAUU...................................................C..GU.ACU.UCCUAAUCAACUGUUCGG.........GGGCG............
Lre-2-22  AAAA..CGCAG..UAGAUGGCU.GAUUAUGA.ACGUUGA....UAAAUCAGCGUC.............................CAUAAUGCCCCAAUAGGCUAGCU..UC.GCG.UCGUAGCCAUCCUUGCAG.........GGGUG............
Smu-1-2   UUGA..CGCAG..UGGUUUAUU.GGAAGUCCUUAUCCCUUGC.UACGCAAGUGAU...................................................AGCGG.ACA.UCCUAAUAAACUGUGCGG.........AGGUG............
Sha-1-39  UAUU..GGCAG..UAGUUGACU.GAAUUAAAAAUGCACU....UGUAACUAGCUU...................................................U.UUU.UAA.UUCUAGUCAACCUUGCCG.........GGGCG............
Ssa-1-7   UUUA..GGCAG..UAGAUGACU.GAAUUGAA.AAUGUAA....AGUUUCACUGAUUAAUUCAUUAAGUUAGUGAGACUUAUGCAUGAGCUUUAGCUCAGGUAAACAU..UU.CAA.UCCUAGUCAUCCUUGCCG.........GGGUG............
          ......<<<<<....<<<<.<<.<..<<<<<.<.........................................................................>..>>.>>>....>>>.>>>>.>>>>>...........<<<.............
          ......???2?....2??2.??.2..?????.?.........................................................................?..??.???....2??.2??2.?2???...........?2?.............
          ●●●●--YGCAG--URG●UGAYU-GR○Y●●●●-●●●○○○○----●●●●●●●R●●Y●---------------------------------------------------Y--●●-●●R-●YCUARUCA●CYUUGC●R---------GGGYR------------
```

alignment positions 161 · · · 268

```
..........<<<.....<<<<<<.............................................................>>>>>>..>>>...>>>......
..........???.....??2???.............................................................???2??..???...?2?......
---------●RR●-●AY-RAA○UC○----------------------------------R●A--------------------○○○GA●UUY-●GYY-YYRCY●--YY●
```

```
Ssu-1-1     .........AGAC.GAC.GAAAUCAAAUCAACUUUGUUG....................AUU....................ACUGAUUUC.UGUC.UUUCCG..CUA
Spy-8-1     .........CGUC.UUC.GAAAUC...................................AAA.......................GAUUUC.GGAC.UUACCU..CAU
Spy-4-1     .........CGUC.UUC.GAAAUC...................................AAA.......................GAUUUC.GGAC.UUACCG..CAU
Spy-11-1    .........CGUC.UUC.GAAAUC...................................AAA.......................GAUUUC.GGAC.UUACCG..CAU
Spy-10-1    .........CAUC.UUC.GAAAUC...................................AAA.......................GAUUUC.GGAC.UUACCG..CAU
Sag-8-1     .........CAUC.UUC.GAAAUC...................................AAA.......................GAUUUC.GGAC.UUACCG..CAU
Sag-6-1     .........CGUC.UUC.GAAAUC...................................AAA.......................GAUUUC.GGAC.UUACCG..CAU
Sag-4-1     .........CGUC.UUC.GAAAUC...................................AAA.......................GAUUUC.GGAC.UUACCG..CAU
Sag-2-1     .........CGUC.UUC.GAAAUC...................................AAA.......................GAUUUC.GGAC.UUACCG..CAU
Spy-8-2     .........CGUC.AAC.GAAAUC...................................AAA.......................GAUUUC.UGAC.UUACCG..CAA
Spy-9-2     .........CGUC.AAC.GAAAUC...................................AAA.......................GAUUUC.UGAC.UUACCG..CAA
Spy-5-2     .........CGUC.AAC.GAAAUC...................................AAA.......................GAUUUC.UGAC.UUACCG..CAA
Spy-10-2    .........CGUC.AAC.GAAAUC...................................AAA.......................GAUUUC.UGAC.UUACCG..CAA
Spy-4-2     .........CGUC.AAC.GAAAUC...................................AAA.......................GAUUUC.UGAC.UUACCG..CAA
Spy-8-3     .........CGUC.AAC.GAAAUC...................................AAA.......................GAUUUC.UGAC.UUACCG..CAA
Ssu-1-2     .........CGUC.UGC.GAAAUC...................................AAA.......................GACUUC.GGAC.UUACCG..CUU
Ssu-1-3     .........CGUC.AAC.AAAAUC...................................AAA.......................GAUUUU.GGAC.UUACCG..CUU
Smu-1-1     .........CGUC.AGC.GAAAUC...................................AAA.......................GUUUUC.UGAU.UUACCGCCCAU
Ssu-1-4     .........CGUC.UAC.GAAAUC...................................AAA.......................GAUUUC.GGAC.UUACCG..CUU
Ssu-1-5     .........CGUC.UAC.GGAAUC...................................AUA.......................GAUUCC.UGAC.UUACCG..CCU
Spy-8-4     .........CGAC.GAC.GAAAUCGCAUUCUAC..........................GAAUG..................ACCGAUUUC.UGUC.UUACCG..UCU
Spy-12-3    .........CGAC.GAC.GAAAUCGCAUUCUAC..........................GAAUG..................ACCGAUUUC.UGUC.UUACCG..UCU
Spy-3-3     .........CGAC.GAC.GAAAUCGCAUUCUAC..........................GAAUG..................ACCGAUUUC.UGUC.UUACCG..CCU
Spy-11-3    .........CGAC.GAC.GAAAUCGCAUUCUAC..........................GAAUG..................ACCGAUUUC.UGUC.UUACCG..UCU
Spy-9-3     .........CGAC.GAC.GAAAUCGCAUUCUAC..........................GAAUG..................ACCGAUUUC.UGUC.UUACCG..UCU
Spy-4-3     .........CGAC.GAC.GAAAUCACAUUCUAC..........................GAAUG..................ACCGAUUUC.UGUC.UUACUG..CAA
Spy-10-3    .........CGAC.GAC.GAAAUCGCAUUCUAC..........................GAAUG..................ACCGAUUUC.UGUC.UUACUG..CAA
Sau-6-1     .........AUUC.UAA.GCAACCU..................................AGA....................UAAGGUUCG.AGAA.UUUCUC..CCU
Sau-3-1     .........AUUC.UAA.GCAACCU..................................AGA....................UAAGGUUCC.AGAA.UUUCUC..CCU
Sau-11-1    .........AUUC.UAA.GCAACCU..................................AGA....................UAAGGUUCC.AGAA.UUUCUC..CCU
Sau-5-1     .........AUUC.UAA.GCAACCA..................................AGA....................UAUGGUUCC.AGAA.UUUCUC..CCU
Sau-11-2    .........GGAC.GAC.GAAAUAAAUUUUGCGA.........................AAA....................UAUCAUUUC.UGUC.CCACUC..CCU
Sau-1-1     .........GGAC.GAC.GAAAUAAAUUUUGCGA.........................AAA....................UAUCAUUUC.UGUC.CCACUC..CCU
Sau-1-2     .........GGAC.GAC.GAAAUAAAUUUUGCGA.........................AAA....................UAUCAUUUC.UGUC.CCACUC..CCA
Sau-6-5     .........GGAC.GAC.GAAAUAAAUUUUGAGA.........................AAA....................UAUCAUUUC.UGUC.CCACUC..CCU
Sau-8-3     .........GGAC.GAC.GAAAUAAAUUUUUCGA.........................AAA....................UAUCAUUUC.UGUC.CCACUC..CCA
Sau-8-4     .........GGAC.GAC.GAAAUAAAUUUUGCGA.........................AAA....................UAUCAUUUC.UGUC.CCACUC..CCA
Sau-8-5     .........GGAC.GAC.GAAAUAAAUUUUGAGA.........................AAA....................UAUCAUUUC.UGUC.CCACUC..CCG
Sau-5-5     .........GGAC.GAC.GAAAUAAAUUUUGCGA.........................AAA....................UAUCAUUUC.UGUC.CCACUC..CCU
Sau-5-6     .........GGAC.GAC.GAAAUAAAUUUUGCGA.........................AAA....................AAUCAUUUC.UCUC.CCACUC..CCA
Sau-2-3     .........GGAC.GAC.GAAAUAAAUUUUACGA.........................AAA....................UAUCAUUUC.UGUC.CCACUC..CCU
Sau-11-3    .........GGAC.GAC.GAAAUAAAUUUUACGA.........................AAA....................UAUCAUUUC.UGUC.CCACUC..CCU
Sau-8-7     .........GGAC.GAC.GAAAUAAAUUUUACGA.........................AAA....................UAUCAUUUC.UGUC.CCACUC..CCU
Sau-5-7     .........GGAC.GAC.GAAAUAAAUUUUUCGA.........................AAA....................UAUCAUUUC.UGUC.CCACUC..CCC
Sau-11-4    .........GGAC.GAC.GAAAUAAAUUUUGCUA.........................AAA....................UAUGAUUUC.UGUC.CCACUC..CCU
Sau-5-8     .........GGAC.AAC.GAAAUAAAUUUUGUAA.........................AAA....................UAUCAUUUC.UGUC.CCACUC..CCA
Sau-6-6     .........GGAC.GAC.GAAAUAAAUUUUGCUA.........................AAA....................UAUGAUUUC.UGUC.CCACUC..CCU
Sha-1-1     .........UGAC.AAC.GAAAUAAAUUUUGAGA.........................AAA....................UAUUAUUUC.UGUC.CCACUC..CCU
Sha-1-2     .........GGAC.AAC.GAAAUAAAUUUUGAGA.........................AAA....................UAUUAUUUC.UGUC.CCACUC..CCA
Sha-1-3     .........GGAC.AAC.GAAAUAAAUUUUGAGA.........................AAA....................UAUUAUUUC.UGUC.CCACUC..CCU
Sha-1-5     .........GGAC.UAC.GAAAUAAAUUUUGAGA.........................AAA....................UAUUAUUUC.UGUC.CUACUC..CCU

            ..........<<<.....<<<<<<.............................................................>>>>>>..>>>...>>>......
            ..........???.....??2???.............................................................???2??..???...?2?......
            ---------●RR●-●AY-RAA∘UC∘----------------------------------R●A--------------------∘∘∘GA●UUY-●GYY-YYRCY●--YY●
```

```
Sha-1-6   .........GUAC.AAC.GAAAUAAAUUUUGAGA.........................AAA....................UAUUAUUUC.UGUC.CCACUC..CCA
Sha-1-7   .........GGCC.AAC.GAAAUAAAUUUUGAGA.........................AAA....................UAUUAUUUC.UGUC.CUACUC..CCA
Ssa-1-1   .........GGAC.CAC.GAAAUCUCUAUUAGA..........................AAA....................UUUGAUUUC.UGUC.CCACUC..CCU
Ssa-1-2   .........GGAC.UAC.GAAAUCUCUAUUAAA..........................AAA....................UUUGAUUUC.UGUC.CCACUC..CCU
Ssa-1-3   .........GGAA.UAC.GAAAUCUCUAUUAGA..........................AAA....................UUUGAUUUC.UGUC.CCACUC..CCG
Ssa-1-4   .........GGAC.UAC.GAAAUCUUUAUUAAA..........................ACA....................AAUGAUUUC.UGUC.CCUCUC..CCU
Ssa-1-5   .........GGAC.UAC.AAAAUCUCUAUUAGA..........................AAA....................UUGGAUUUU.UGUC.CCUCUC..CGU
Sha-1-8   .........GGAC.GAC.GAAUUU...................................AAA.....................AAGAAUUC.UGUC.CCAGUC..CUU
Sha-1-9   .........AGAC.GAC.GAAUUUA..................................AAA.......................GAAUUC.UGUC.CCACUC..CCA
Sha-1-10  .........GGAC.GAC.GAAUUU...................................AAA.....................AAGAAUUC.UGUC.CCACUC..CCU
Sha-1-11  .........GGAC.GAC.GAAUUU...................................AAA.....................AAGAAUUC.UGUC.CCACUC..CCU
Sha-1-14  .........GGAC.GAC.GAAUUUU..................................AAA......................AGAAUUC.UGUC.CCACUC..CCU
Sha-1-16  .........GAAC.AAC.GAAUUC...................................AUU.....................GUGAAUUC.UGUU.CCACUC..CUU
Sha-1-17  .........GGAC.AAU.GAAUUC...................................AAA.....................AUGAAUUC.UGAC.UCACUC..CCC
Sha-1-18  .........GGAC.CAC.GAAUUU...................................AUU.....................UUAAUUUC.UGUC.CCGCUC..CCA
Sha-1-19  .........GGAC.AAC.GAAUUCA..................................AAA.......................GAAUUC.UGUC.CCAGUC..CUU
Sha-1-20  .........AGAA.AAC.GAAUUCGU.................................UAA.......................CAAUUC.UGUC.UCACGU..CUU
Sha-1-21  .........GCCC.AAG......C...................................AAA.......................G......AGAA.UUUCAU..UAA
Sha-1-22  .........GCCC.CAG......C...................................AAA.......................G......AGAA.UUUCAU..CAA
Sha-1-23  .........GCCC.CAA......C...................................AAA.......................G......AGAA.UUUCAC..UAA
Sha-1-24  .........GCCC.CAA......C...................................AAA.......................G......AGAA.UUUCAC..UAA
Sha-1-25  .........GGAC.GAU.GAAUUC...................................AAA.....................AUGAAUUC.UGUU.CCACUC..UUU
Sau-6-7   .........GCCC.CAA......C...................................AAA.......................GAGAAAUUGGA.UUCCCA..AUU
Sau-11-6  .........GCCC.CAA......C...................................AAA.......................GAGAAAUUGGA.UUCCCA..AUU
Sau-8-8   .........GCCC.CAA......C...................................ACA.......................GAGAAAUUGGA.UUCCCA..AUU
Sau-6-8   .........GCCC.CAA......C...................................AAA.......................GAGAAAUUGGA.UUCCCG..AUU
Sau-7-8   .........GCCC.CAA......C...................................AAA.......................GAGAAAUUGGA.UUCCCG..AUU
Sau-5-9   .........GCCC.CAA......C...................................AAA.......................GAGAAAUUGUA.UUCCCA..AUU
Sau-5-10  .........GCCC.CAA......C...................................AUA.......................GAGAAAUUGGA.UUCCCA..AUU
Sha-1-26  .........GGAU.AAU.GAAUUC...................................AAA.....................UUGAAUUC.UGUC.UCGCUC..CUU
Sha-1-27  .........GGGC.GAC.GAAUUCAG.................................AAA.......................GAAUAC.UGUC.CCCCUU..AUU
Sha-1-28  .........G.GC.AAU.GAAUUCU..................................AGA......................UGAAUUC.CGUG.UCACUC..CUG
Sha-1-29  .........GUAC.GAC.GAAAUC...................................ACA....................AGUGAUUUC.UGUA.CGGCUC..CCU
Ssa-1-6   .........GGAC.AAC.GAAAGCAUUAUUAA...........................AAA....................GCUGAUUUC.UGUC.CCGGUC..CCU
Sha-1-30  .........GGAC.AAU.GAAUUCA..................................CUA......................UGAAUUC.UAUC.CGACUC..CUU
Sep-2-1   .........UGAC.UAC.AAAAUCUUUUUAUA...........................AAA....................AUAAAAUUU.UACC.UCUCUU..CAU
Sha-1-31  ............................................................................................................
Sha-1-32  ............................................................................................................
Sha-1-33  ..........GGC.GGG............................................................................GCC...ACAG..CAA
Lre-2-1   ............................................................................................................
Lre-2-2   ............................................................................................................
Lre-1-3   ............................................................................................................
Sha-1-34  .........UGGG.GCC..........................................CCG....................AAC........CCC.AACUCC..CCU
Sha-1-35  .........AGAC.UAU.GAAUCC...................................AUA.....................AUGGAUUC.UGUC.UAGCUC..CUU
Spn-2-1   .........AGAC.GAU.GAAAUC...................................GAA.......................CUCUUCGAGUU...ACUG..AAU
Sha-1-36  .........GAAC.AAU.GAAUUCAU.................................AAA.......................GAUUAC.UGUC.CUACUC..CUC
Sha-1-37  .........G............UCGGGGUCAG...........................CAA....................AGAGA............AUUU..CAU
Lre-2-5   .........GGAC.UAA.AAACUC...................................GCA....................AGCGAGUUU.AGUC.CCACUC..CCA
Lre-2-6   .........GGAA.GAC.GAAGUC...................................GCA....................AGCGACUUC.UUUC.CCACUC..CCU
Spn-1-1   .........AGAC.GAU.GAAAUCGAACUCUUCG.........................AGU....................UACCGAAUC.UGCC.UCAUUU..UCU
Lre-1-4   .........C.....CC.AGAGGCUAGCUUCGCGUCGAAAAACGAAGUC..........ACA....................AGUGACUUU.UGUC.UCGCUC..CAC

          ..........<<<.....<<<<<<.............................................................>>>>>>..>>>...>>>......
          ..........???.....??2???.............................................................???2??..???...?2?......
          ---------●RR●-●AY-RAA∘UC∘----------------------------------R●A--------------------∘∘∘GA●UUY-●GYY-YYRCY●--YY●
```

```
Lre-1-5    .........CGAA.AAC.GAAGUC...................................ACA....................AGUGACUUC.UGUC.UCGCUC..CCU
Lre-2-7    .........CGAA.AAC.GAAGUC...................................ACA....................AGUGACUUC.UGUC.UCGCUC..CCU
Lre-1-6    .........CGAA.AAC.GAAGUC...................................ACA....................AGUGACUUC.UGUC.UCGCUC..CCU
Lre-2-8    .........CGAA.AAC.GAAGUC...................................ACA....................AGUGACUUC.UGUC.UCGCUC..CCU
Lre-2-9    .........CGAA.AAC.GAAGUC...................................ACA....................AGUGACUUC.UGUC.UCGCUC..CCU
Lre-2-10   .........CGAA.AAC.GAAGUC...................................ACA....................AGUGACUUC.UGUC.UCGCUC..CCU
Lre-2-11   .........CGAA.AAC.GAAGUC...................................ACA....................AGUGACUUC.UGUC.UCGCUC..CCA
Lre-1-7    .........CGAA.AAC.GAAGUC...................................ACA....................AGUGACUUC.UGUC.UCGCUC..CCU
Lre-1-8    .........CGAA.AAC.GAAGUC...................................ACA....................AGUGACUUC.UGUC.UCGCUC..CCU
Lre-1-9    .........CGAA.AAC.GAAGUC...................................ACA....................AGUGACUUC.UGUC.UCGCUC..CCU
Lre-2-12   .........CGAA.AAC.GAAGUC...................................ACA....................AGUGAAUUC.UGUC.UCGCUC..CCA
Lre-1-10   .........CGAA.AAC.GAAGUC...................................ACG....................AGUGACUUC.UGUC.UCGCUC..CCU
Lre-1-11   .........CGAA.AAC.GAAGUC...................................ACA....................AGUGACUUC.UGUC.UCGCUC..CCA
Lre-2-13   .........CGAA.AAC.GAAGUC...................................ACG....................AGUGACUUC.UGUC.UCGCUC..CCA
Lre-2-14   .........CGAA.AAACGAAGUC...................................ACG....................AGUGACUUC.UGUC.UCGCUC..CCU
Lre-2-15   .........CGAA.AAC.GAAGUC...................................ACA....................AGUGACUUC.UGUC.UCGCUC..CCU
Lre-1-12   .........CGAA.AAC.GAAGUC...................................ACA....................AGUGACUUC.UGUC.UCGCUC..CCU
Lre-2-16   .........CGAA.AAC.GAAGUC...................................ACG....................AGUGACUUC.UGUC.UCGCUC..CCU
Lre-1-13   .........CGAA.AAC.GAAGUC...................................ACG....................AGUAACUUC.UGUC.UCGCUC..CUA
Lre-2-17   CUUCGCGUCGAAA.AAC.GAAGUC...................................ACA....................AGUGACUUC.UGUC.UCGCUC..CCU
Lre-2-18   CUUCGCGUCGAAA.AAC.GAAGUC...................................ACA....................AGUGACUUC.UGUC.UCGCUC..CCA
Lre-1-15   CUUCGCGUCGAAA.AAC.GAAGUC...................................ACA....................AGUGACUUC.UGUC.UCGCUC..CCU
Lre-1-16   CUUCGCGUCGAAA.AAC.GAAGUC...................................ACA....................AGUGACUUC.UGUC.UCGCUC..CCU
Lre-1-17   .........C.AC.A...GAGGCUAGCUUCGCGACGAAAAACGAAGUC...........ACA....................AGUGGCUUC.UGUC.UCGCUC..CCU
Lre-2-19   .........C.AC.A...GAGGCUAGCUUUGCGUCGAAAACGAACUC............ACA....................AGUGACUUC.UGUC.UCGCUC..CCC
Sha-1-38   .........AAAU.AAC.ACAAUCG..................................AAA.......................GAAUCC.UACU.UCACUC..UCU
Sep-2-2    .........GGAC.UUC.GAAUUCAU.................................CAA....................AAUGGAUUC.UGUC.CUACCC..CUU
Ssu-1-6    .........GGAA.GAC.GAACUCUUUUUUUGA..........................CCA....................GUCGAGUUC.UUUC.CCGCUC..CCU
Ssu-1-7    .........GGAA.GAC.GAACUCUUUUUUUGA..........................CCA....................GUCGAGUUC.UUUC.CCGCUC..CCU
Ssu-1-8    .........GGAA.GAC.GAACUCUUUUUGAC...........................CUA....................GUCGAGUUC.UUUC.CCACUC..CCA
Ssu-1-9    .........GGAA.GAC.GAACUCUUUUUAUC...........................UUG....................GUCGAGUUC.UUUC.CCACUC..CCC
Ssu-1-10   .........GGAA.GAC.GAACUCUUUUUAUC...........................UUG....................GUCGAGUUC.UUUC.CCACUC..CCC
Ssu-1-11   .........GGAA.GAC.GAACUCUUUUUAUC...........................UUG....................GUCGAGUUC.UUUC.CCACUC..CCA
Ssu-1-13   .........GGAA.GAC.GAACUCUUUUU..............................AGACUG.................GUCGAGUUC.UUUC.CCACUC..CCU
Ssu-1-14   .........GGAA.GAC.GAACUCUUUUU..............................AGACUG.................GUCGAGUUC.UUUC.CCACUC..CCA
Ssu-1-15   .........GGAA.AAC.GAACUCUUCUAUGG...........................AUG....................GUCGAGUUC.UUUCCCACUG..CCA
Ssu-1-16   .........GGAA.GAC.GAACUCUUUUUUGA...........................CUA....................GUUGAGUUC.UUUC.CCACUC..CCU
Ssu-1-17   .........GGAA.GAC.GAACUCUUUUCUA............................UCA....................GUCAAGUUC.UUUC.CCGCUC..CCC
Ssu-1-18   .........AGAA.GAC.GAACUCAUUUUUG............................AAA....................AUAGAGUUC.UUUC.UCACUC..CCA
Ssu-1-19   ..........GGG.AGU.GAAACAGUCUGGGG...........................AUA....................GACUGUUUC.AGCU.CAACAA..CUU
Ssu-1-20   .........GGAA.GAC.GAACUCUCUAUGA............................UUA....................UUUGAGUUC.UGUC.CCACUC..CCA
Ssu-1-21   .........C.....GG.GAAGAC...................................GAA.......................CUCUUU.U......UCUA..CCA
Lre-2-20   .........GGAC.UA..AAACUU...................................ACA....................AGCGAGUUU.AGUC.UCACCC..CUU
Lre-2-21   .........AGAC.UGA.AAACUC...................................GCA....................AGUGAGUUC.AGUC.UCACUC..UCU
Sag-5-2    .........GGAGUGAA.ACAGUCUGGGG..............................AUA.......................GACUGU.U.UC.AGGUCA..GCA
Sag-2-2    .........GGAC.GAC.AAAAUCGAAUUCCUUUGGA......................AUU....................ACCGAUUUU.UGUC.CCACUC..UCA
Ssu-1-22   .........UCAC.AAC.GAAAUUUCUAG..............................CAA.......................AAUUUC.UGUU.CGGCCG..UCU
Sth-3-1    .........AAAC.UAC.GAAAUC...................................AAACUCUUCGAGUU.........AUUGAUUUC.UGUU.UCGCUC..CCU
Sth-2-1    .........AAAC.UAC.GAAAUC...................................AAACUCUUCGAGUU.........ACUGAUUUC.UGUU.UCGCUC..CCU
Lre-2-22   .........GGAC.UAA..AAACUC..................................GAA....................ACAAGUUUAUAGUC.CCACUC..CCU
Smu-1-2    .........GGAC.AAC.AAAAUCG..................................AAAUUGCCAUUUCAUAACGAUUUACCGAUUUU.UGUU.CCACUC..UCA

           ..........<<<.....<<<<<<.............................................................>>>>>>..>>>...>>>......
           ..........???.....??2???.............................................................???2??..???...?2?......
           ---------●RR●-●AY-RAA◦UC◦----------------------------------R●A--------------------◦◦◦GA●UUY-●GYY-YYRCY●--YY●
```

```
Sha-1-39  .........AUGG.GAC.CCUAGCAAAGAGAAUUUCACAAAGAAAUUCUACAAGCUAAGCAA.......................GCUGGGGAUCG.GGGCCC..CAG
Ssa-1-7   .........GGAC.UAC.GAAAUCUCUAUUAGA..........................AAA....................UUUGAUUUC.UGUC.CCACUC..CCA
          ..........<<<.....<<<<<<.............................................................>>>>>>..>>>...>>>......
          ..........???.....??2???.............................................................???2??..???...?2?......
          ---------●RR●-●AY-RAA●UC○----------------------------------R●A--------------------○○○GA●UUY-●GYY-YYRCY●--YY●
```

References

[1] Jeffrey E. Barrick, K. A. Corbino, W. C. Winkler, A. Nahvi, M. Mandal, J. Collins, M. Lee, A. Roth, N. Sudarsan, I. Jona, J. K. Wickiser, and Ronald R. Breaker. New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proceedings of the National Academy of Sciences of the Unites States of America*, 101(17):6421–6426, 2004.

[2] Jakub Bielnicki, Yancho Devedjiev, Urszula Derewenda, Zbigniew Dauter, Andrzej Joachimiak, and Zygmunt S. Derewenda. *B. subtilis* ykuD protein at 2.0 Å resolution: Insights into the structure and function of a novel, ubiquitous family of bacterial enzymes. *Proteins*, 62:144–151, 2005.

[3] Ryan T Fuchs, Frank J Grundy, and Tina M Henkin. The S(MK) box is a new SAM-binding RNA for translational regulation of SAM synthetase. *Nat Struct Mol Bio*, 13:226–33, 2006.

[4] Shelley A. Haveman, Dawn E. Holmes, Yan-Huai R. Ding, Joy E. Ward, Jr. Raymond J. DiDonato, and Derek R. Lovley. *c*-type cytochromes in *Pelobacter carbinolicus*. *Appl Environ Microbiol*, 72:6980–5, 2006.

[5] Dawn E. Holmes, Swades K. Chaudhuri, Kelly P. Nevin, Teena Mehta, Barbara A. Methé, Anna Liu, Joy E. Ward, Trevor L. Woodard, Jennifer Webster, and Derek R. Lovley. Microarray and genetic analysis of electron transfer to electrodes in *Geobacter sulfurreducens*. *Env Microbiol*, 8:1805–1815, 2006.

[6] Byoung-Chan Kim, Xinlei Qian, Ching Leang, Maddalena V. Coppi, and Derek R. Lovley. Two putative c-type multiheme cytochromes required for the expression of OmcB, an outer membrane protein essential for optimal Fe(III) reduction in *Geobacter sulfurreducens*. *J Bact*, 188:3138–3142, 2006.

[7] Derek R. Lovley. Bug juice: harvesting electricity with microorganisms. *Nat Rev Microbiol*, 4:497–508, 2006.

[8] Teena Mehta, Maddalena V. Coppi, Susan E. Childers, and Derek R. Lovley. Outer membrane *c*-type cytochromes required for Fe(III) and Mn(IV) oxide reduction in *Geobacter sulfurreducens*. *Appl Environ Microbiol*, 71:8634–8641, 2005.

[9] Karin L. Meibom, Xibing B. Li, Alex T. Nielsen, Cheng-Yen Wu, Saul Roseman, and Gary K. Schoolnik. The *Vibrio cholerae* chitin utilization program. *PNAS*, 101:2524–9, 2004.

[10] Michinori Mutsuda and Masahiro Sugiura. Translation initiation of cyanobacterial rbcS mRNAs requires the 38 kDa ribosomal protein S1 but not the Shine-Dalgarno sequence: Development of a cyanobacterial in vitro translation system. *J Biol Chem*, 281:38314–21, 2006.

[11] Martin G. Reese. Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. *Comput Chem*, 26:51–6, 2001.

[12] Douglas B. Rusch, Aaron L. Halpern, Granger Sutton, Karla B. Heidelberg, Shannon Williamson, Shibu Yooseph, Dongying Wu, Jonathan A. Eisen, Jeff M. Hoffman, Karin Remington, Karen Beeson, Bao Tran, Hamilton Smith, Holly Baden-Tillson, Clare Stewart, Joyce Thorpe, Jason Freeman, Cynthia Andrews-Pfannkoch, Joseph E. Venter, Kelvin Li, Saul Kravitz, John F. Heidelberg, Terry Utterback, Yu-Hui Rogers, Luisa I. Falcón, Valeria Souza, Germán Bonilla-Rosso, Luis E. Eguiarte, David M. Karl, Shubha Sathyendranath, Trevor Platt, Eldredge Bermingham, Victor Gallardo, Giselle Tamayo-Castillo, Michael R. Ferrari1, Robert L. Strausberg, Kenneth Nealson, Robert Friedman, Marvin Frazier, and J. Craig Venter. The sorcerer ii global ocean sampling expedition: Northwest atlantic through eastern tropical pacific. *PLoS Biology*, 5:e77, 2007.

[13] J D Thompson, D G Higgins, and T J Gibson. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22(22):4673–4680, November 1994.