Predicting tRNA and tmRNA genes

Revised December 1, 2016

tRNAs are highly structurally conserved small RNA molecules. While the overall shape of tRNAs is extremely similar (necessary for the rapid interactions with the ribosomes) there are a number of ways a single-strand of RNA can fold to generate the tRNA stems, loops, and pseudoknots that interact to form the final structure. As such, no single prediction tool is accurate at predicting all tRNAs. The tRNAs found in phage genomes are more variable than canonical tRNAs found in their bacterial hosts—some match the host parameters perfectly, and some do not. These parameters are described in the "interpreting your results section" below.
We are exploring predicted tRNAs in actinobacteriophage genomes at the bench to try to understand how this variability affects tRNA expression and function in protein translation.

We synthesize the results of three programs to annotate tRNAs in our actinobacteriophage genomes:

1. Aragorn (v1.1) --during auto-annotation in DNA Master--- finds the basic tRNA genes
2. Aragorn (v1.2.38) (http://130.235.46.10/ARAGORN/) finds the basic tRNA genes, and adjusts the ends of the genes correctly.
3. tRNA ScanSE 2.0 (http://trna.ucsc.edu/tRNAscan-SE/) finds non-canonical tRNAs

Aragorn (version 1.1)

DNA Master's Auto-Annotate function runs the tRNA search tool Aragorn (v1.1), which may identify some basic tRNA genes in your genome. However, the version of Aragorn that is within DNA Master does not call all the tRNAs, nor does it correctly trim the ends of the tRNAs that it does identify. The results of Aragorn (v1.1) should be superceded by the results of the other programs.

Aragorn (version 1.2.38, as of December 2016)

Aragorn (v 1.2.38 or later) is the most robust tool for tRNA identification and for correctly identifying the ends of the tRNAs. If the results of the programs are in conflict regarding a single tRNA, its anticodon, and its coordinates, this is the tool that is most often correct.

To run:

Go to: http://130.235.46.10/ARAGORN/
Select the setting as shown in Figure 1.

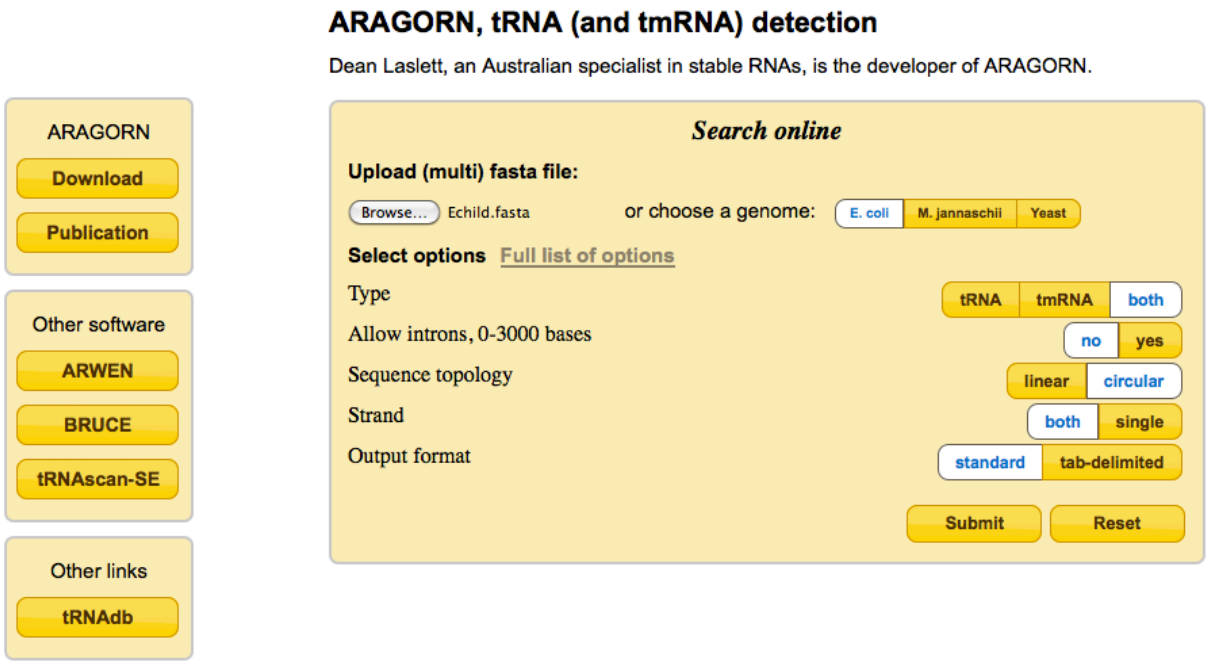


Figure 1

Setting selection process:

In the 'Upload (multi) fasta file' section, click 'Browse…' then select your phage's DNA sequence as a FASTA file.
Type: Both (tRNA & tmRNA)
Allow introns: no
Sequence topology: circular (because phage genomes circularize upon infection)
Strands: both
Output format: standard

Click the 'Submit' button.

Your results will load in a new page. The output includes the secondary structure of the tRNAs found. An example is shown in Figure 2.

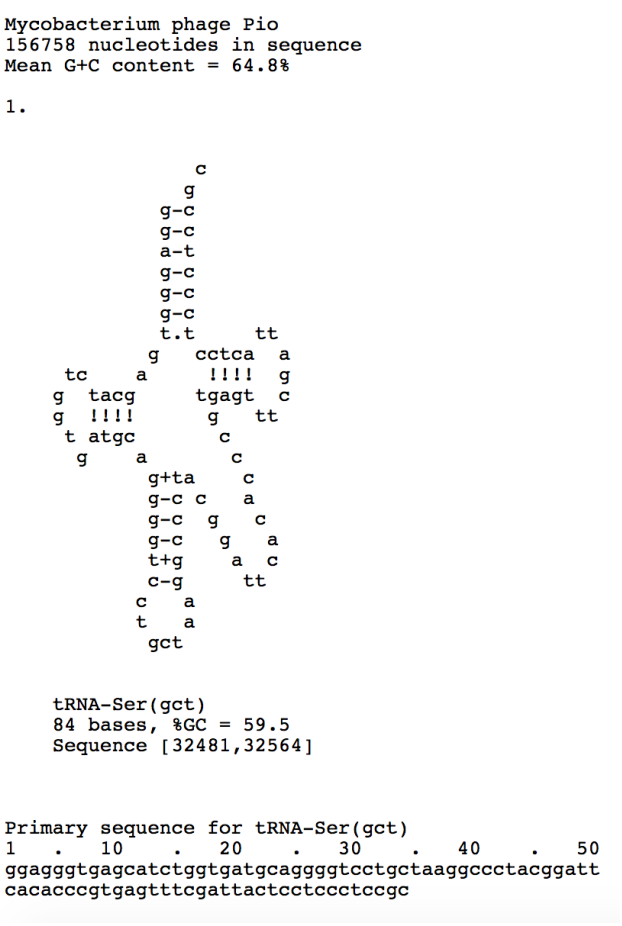


Figure 2

When using this website, please cite:
Laslett, D. & Canback, B. (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 32;11-16. PMID: 14704338

tRNAscan-SE (version 2.0)

tRNAscan-SE is better at identifying the tRNAs that do not follow the strict guidelines that bacterial tRNAs do. These tRNAs may have a slightly enlarged loop, or slightly shortened stem, or be missing a specific conserved base. Any of these flaws may prevent Aragorn from identifying them. However, it is possible to relax the parameters in tRNAscan-SE in order to identify them in our genomes. We have empirically determined the settings we think are appropriate for identifying tRNAs in our phage genomes that are likely to be expressed during infection, without including too many fasle positives. It is possible to relax the settings even further and tRNAscan-SE will find even more “tRNA-like objects” in your sequence. tRNAscan-SE is not as good as Aragorn at determining the correct ends of the tRNAs.

To run:

Go to: http://trna.ucsc.edu/tRNAscan-SE/
Select the settings as shown in Figure 3.

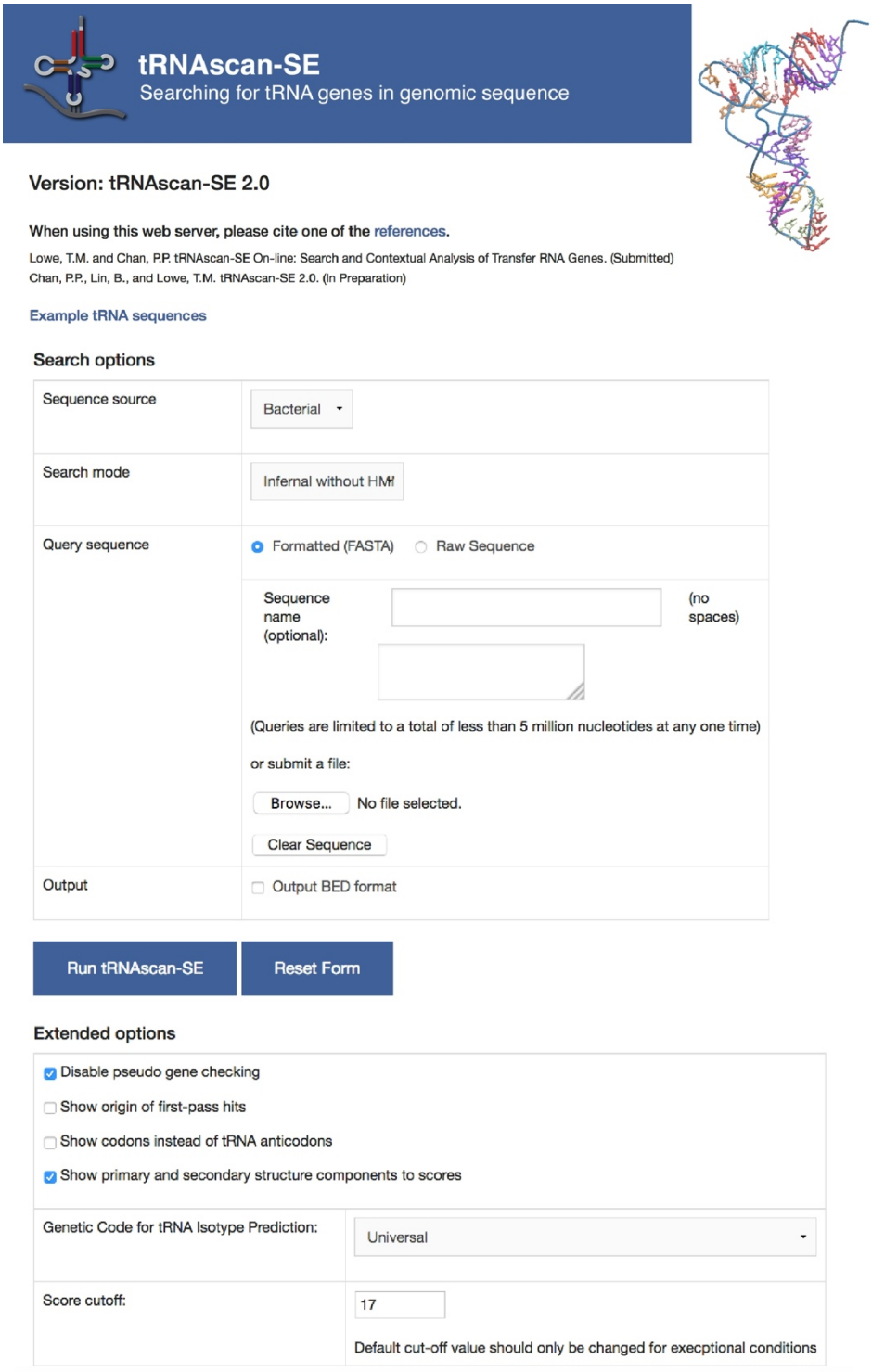


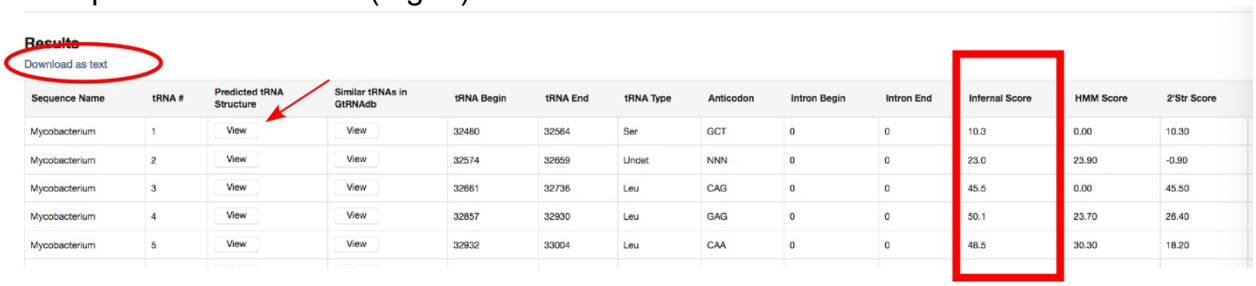
Figure 3

Setting selection process:

Sequence source: Bacterial
Search mode: Infernal without HMM
Upload your sequence: Use the .fasta file from phagesDB.
Extended Options
Check “Disable pseudo gene checking”
Check “Show primary and secondary structure components to scores”
Genetic Code for tRNA isotype Prediction: Universal
Score cutoff: 17

Click the “Run tRNAscan-SE” button.

Your results will load into a new page and have multiple sections of data, including a documentation of the settings that you have used. It is recommended that you save the complete output as an .html file. (Fig. 4)


Results
Download as text

Sequence Name	tRNA #	Predicted tRNA Structure	Similar tRNAs in GtRNAdb	tRNA Begin	tRNA End	tRNA Type	Anticodon	Intron Begin	Intron End	Infernal Score	HMM Score	2'Str Score
Mycobacterium	1	View	View	32460	32564	Ser	GCT	0	0	10.3	0.00	10.30
Mycobacterium	2	View	View	32574	32659	Undet	NNN	0	0	23.0	23.90	-0.90
Mycobacterium	3	View	View	32661	32736	Leu	CAG	0	0	45.5	0.00	45.50
Mycobacterium	4	View	View	32857	32930	Leu	GAG	0	0	50.1	23.70	26.40
Mycobacterium	5	View	View	32932	33004	Leu	CAA	0	0	48.5	30.30	18.20

Figure 4

Click on the Red Arrow as shown on Fig. 4 to view each tRNA structure:

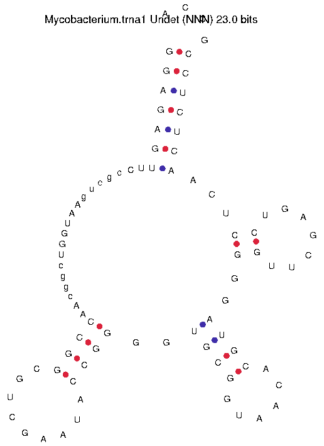


Figure 5

But also download the results as a text file (See circled text box in Fig. 4).

Sequence Name	tRNA #	tRNA Begin	Bounds End	tRNA Type	Anti Codon	Intron Bounds Begin	Intron Bounds End	Inf Score	HMM Score	2'Str Score	Pseudo	Isotype CM	Isotype Score
Mycobacterium	1	32480	32564	Ser	GCT	0	0	10.3	0.00	10.30	No	Ser	13.1
Mycobacterium	2	32574	32659	Undet	NNN	0	0	23.0	23.90	-0.90	No	Ser	30.9
Mycobacterium	3	32661	32736	Leu	CAG	0	0	45.5	0.00	45.50	No	Leu	53.9
Mycobacterium	4	32857	32930	Leu	GAG	0	0	50.1	23.70	26.40	No	Leu	58.5
Mycobacterium	5	32932	33004	Leu	CAA	0	0	48.5	30.30	18.20	No	Leu	45.4
Mycobacterium	6	94729	94802	Pro	TGG	0	0	69.4	50.10	19.30	No	Pro	92.1
Mycobacterium	7	94821	94895	Trp	CCA	0	0	46.4	35.00	11.40	No	Ser	46.7
Mycobacterium	8	94897	94983	Tyr	GTA	0	0	36.7	16.70	20.00	No	Cys	34.7
Mycobacterium	9	95797	95872	Undet	NNN	0	0	12.9	0.00	12.90	No	Leu	15.8
Mycobacterium	10	95940	95997	Undet	NNN	0	0	16.2	0.00	16.20	No	Leu	22.7
Mycobacterium	11	95998	96064	Ser	CGA	0	0	14.1	0.00	14.10	No	Lys	10.3
Mycobacterium	12	96092	96162	Sup	CTA	0	0	67.3	48.60	18.70	No	Ile	59.0
Mycobacterium	13	96315	96387	Met	CAT	0	0	71.7	41.60	30.10	No	fMet	81.9
Mycobacterium	14	96515	96585	Cys	GCA	0	0	54.0	23.20	30.80	No	Cys	67.8
Mycobacterium	15	96654	96724	Glu	CTC	0	0	50.0	16.10	33.90	No	Gln	59.1
Mycobacterium	16	96727	96799	His	GTG	0	0	64.0	41.40	22.60	No	His	90.5
Mycobacterium	17	96962	97035	Ala	TGC	0	0	57.9	45.70	12.20	No	Arg	58.8
Mycobacterium	18	97146	97223	Met	CAT	0	0	19.2	20.30	-1.10	No	Gly	2.8
Mycobacterium	19	97226	97297	Phe	GAA	0	0	70.6	53.20	17.40	No	Thr	77.3
Mycobacterium	20	97304	97376	Val	CAC	0	0	61.3	34.80	26.50	No	Ile	74.7

Figure 6

The html and text document lists are two forms of the same data. An Infernal Score of >17 is considered acceptable (Red Box in Fig. 4). This is not a strict rule, but rather a guideline to follow.

Note: We are primarily interested in the Infernal Score as a measure of the quality of the putative tRNA.

When using this web server, please cite
Lowe, T.M. and Chan, P.P. tRNAscan-SE On-line: Search and Contextual Analysis of Transfer RNA Genes. (Submitted)
Lowe, T.M. and Eddy, S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res, 25, 955-964.

Interpreting the results of all three programs

tRNA secondary structure and end determination

Manual evaluation is required to determine:

- Whether to include a tRNA prediction in your annotation
- the precise 3' end of a tRNA gene
- the anti-codon designated for the tRNA

The tRNAscan-SE provides an Infernal Score that measures the quality of the tRNA predictions. Close examination of tRNAs with scores greater than 35 shows that those tRNAs possess all of the canonical features of a tRNA.

In the canonical tRNA schematic below, the 5' end of the tRNA is a 7 base-pair segment called the Acceptor Stem. The remainder of the tRNA is depicted in the diagram; it winds all the way through three additional stem-loops of variable lengths and then back to the matching base pairs of the acceptor stem. Conserved bases are labeled in

nucleotide single-letter shorthand (G, A, or C for those specific bases; R for purines, Y for pyrimidines) at the indicated position. The tRNA algorithms score potential tRNAs based on their adherence to the conserved bases and stem-loop lengths.

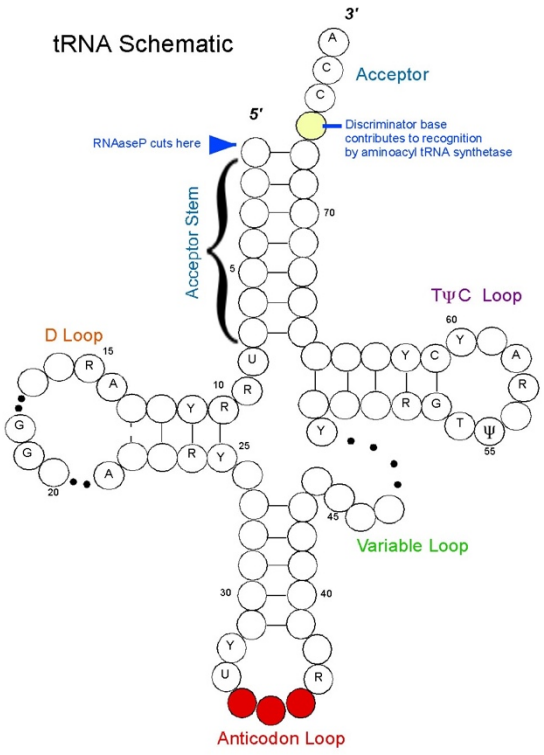


Figure 7

After the Acceptor Stem, the 3' end has up to four unpaired bases. The first is called the discriminator base, and it is a critical part of the recognition system that the tRNA synthetase uses to charge the tRNA with the correct amino acid. The discriminator base is followed by the sequence CCA.

The ends of the tRNA must be carefully checked.
Use Aragorn (v 1.2.38 or later) to call the ends.
-The acceptor stem loop must be seven base pairs.
-The CCA sequence at the 3' end must be present on the final tRNA molecule for the tRNA to be charged. Sometimes in the tRNA gene within the DNA of the genome the CCA sequence is truncated, in which case the additional part of the CCA sequence is added after transcription. **Therefore, if the 3' end of the sequence is not CCA, it should be trimmed at the first deviation from the CCA sequence, and the remainder should not be included in the gene call.** Aragorn 1.2.38 does this perfectly. This may need to be done manually for tRNAs that are only discovered using tRNAscan-SE.

The tRNA Schematic shown in Figure 7 is an adaptation of the schematic found on the Lowe website http://lowelab.ucsc.edu/tRNAscan-SE/ with review and guidance from Dr. Craig L. Peebles.

In Summary:
The phages that contain more than 1 tRNA within their genomes tend to localize the tRNAs to certain regions of the genomes (also called “tRNA clusters” in the phage tRNA literature.) It is highly unusual that a phage with multiple tRNAs will contain a sole tRNA distant genomically from all the others within its genome found by the programs, or encoded on the opposite strand as all the others, or encoded within a ORF called by GeneMark or Glimmer that has high coding potential. In general, violation of either of the two latter of the three preceding conditions is sufficient for exclusion of a potential tRNA from an annotation (we have found a single high scoring tRNA that is not part of the rest of the large cluster, however this situation is very rare). It is possible for a phage genome to have multiple tRNA clusters (for example, the Cluster C mycobacteriophages have three tRNA clusters).

The “best” tRNAs are those with a tRNAscan-SE Infernal Score higher than 35, and that are also found by web-based Aragorn. These criteria include almost all known bacterial tRNAs. Some phage tRNAs meet these standards, however, others don’t. We will include tRNAs that somewhat noncanonical components with an Infernal score of at least 17.0. Until the phage tRNAs are more extensively tested for expression and functionality in the wet lab, we will err on the side of inclusion.

In our annotations, we will include:

- All tRNAs using tRNAscan-SE 2.0 with Infernal Scores above 17.0
- All tRNAs found by Aragorn 1.2.38, even if they are not found by tRNAscan-SE 2.0 at all or have low Infernal Scores.
- The ends of the tRNA should be trimmed to match Aragorn 1.2.38 start and stop coordinates.

Entering a tRNA in DNA Master

DNA Master may have already called some of your tRNA genes using the old stand-alone version of Aragorn. If so, go to the [Feature] tab and the [[Description]] sub-tab, and enter the following information. (See Figure 8 for an example.)
Type: tRNA (not CDS)
5’ and 3’: Exact coordinates as determined above
Feature Product: “tRNA _____” (In the blank, write the amino acid 3-letter abbreviation, e.g. “Lys”.)
Feature Notes: “tRNA _____” (In the blank, write the amino acid 3-letter abbreviation followed by the anti-codon, e.g. “Lys (ttt)”.) Include the Infernal Score from tRNAscan-SE and whether it was detected by Aragorn.

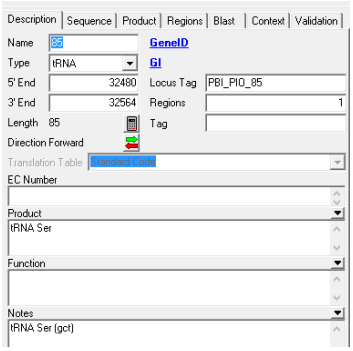


Figure 8

If you are adding a brand new tRNA, click the 'Insert' button at the bottom of the central column. Then enter in the above information in the window that opens and click 'Add Feature'. (You can leave the name blank, and it will be automatically assigned when you renumber genes, as described in Section 9.3.3.)

Identifying and annotating tmRNA genes

Description from Wikipedia:

"Transfer-messenger RNA (tmRNA) is a bacterial RNA molecule with dual tRNA-like and messenger RNA-like properties. In trans-translation, tmRNA and its associated proteins bind to bacterial ribosomes which have stalled in the middle of protein biosynthesis, for example when reaching the end of a messenger RNA which has lost its stop codon. tmRNA can recycle the stalled ribosome, add a proteolysis-inducing tag to the unfinished polypeptide, and facilitate the degradation of the aberrant messenger RNA."

The coordinates for tmRNAs can be annotated as web-based Aragorn (or the algorithm BRUCE on the Aragorn web page) calls them. Entering tmRNAs into your DNA Master annotation can be done using the same procedure as for entering tRNAs only the "Type" of feature in the should then be "tmRNA" (not CDS or tRNA).