**Finding tRNA’s**

A tRNA (transfer ribonucleic acid) gene count from a genome is a common measure provided for newly sequenced genomes. This statistic allows one to verify the validity of a genome sequence. Most, genome sequencing programs do not search for tRNAs. Therefore, specialized programs aimed at finding functional RNAs are employed to identify tRNAs within a genome. These programs include: ARAGORN, tRNA finder, and tRNA scan-se. Additionally, tRNA databases such as tRNAdb-ce can be used in conjunction with NCBI’s BLAST to identify tRNA’s within a genome. Using Microsoft Word, one can obtain a tRNA gene count from these programs’ outputs.

**ARAGORN**

*Background*

Created by Dean Laslett ARAGORN is part of a series of functional RNA gene identifying programs. It utilizes heuristic algorithms to identify tRNA sequences in a genome based on homology with previously identified tRNA consensus sequences and the ability to form the clover leaf loop secondary structure characteristic of tRNAs.

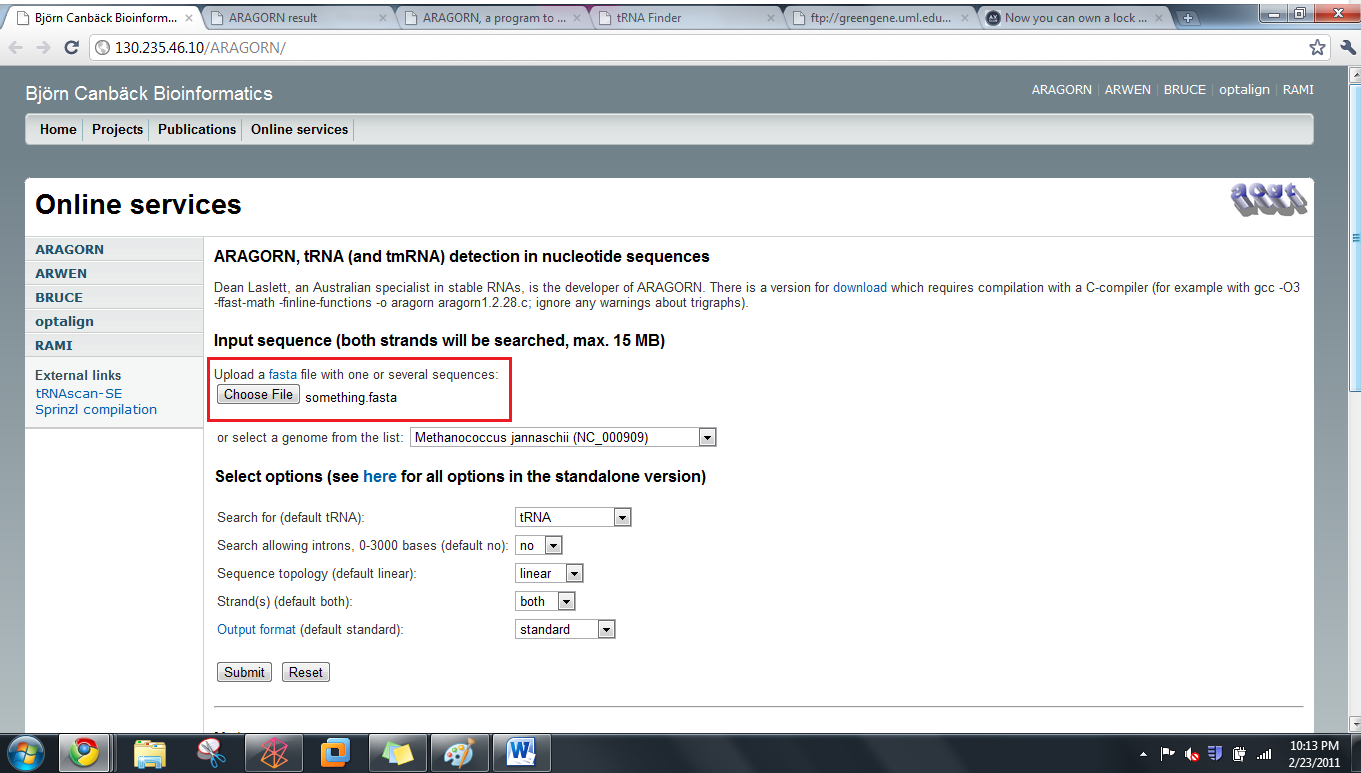
*Using ARAGORN*

The program is available for download, but it is also available as an online tool. A description of how to use the online tool is given below:

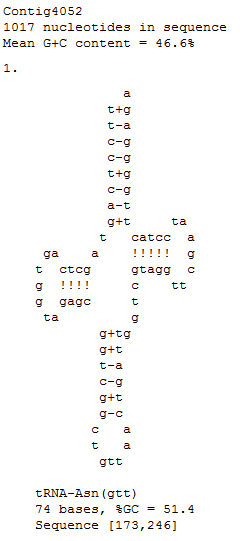
1. To access ARAGORN online use either of these links:

<http://www.acgt.se/online.html> (click on ARAGORN on left hand browser) or <http://130.235.46.10/ARAGORN/>

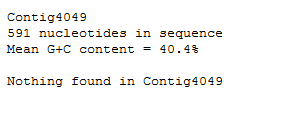
1. These links will take you to the ARAGORN homepage (shown below). Here you can upload a genome (**as a .fasta file**) and search for tRNA with various search criteria (tRNA or tmRNA, introns or no introns, linear or circular chromosome, double or single stranded, and output format)



1. Results should be organized by contigs. If there is a hit, the output will look like this:

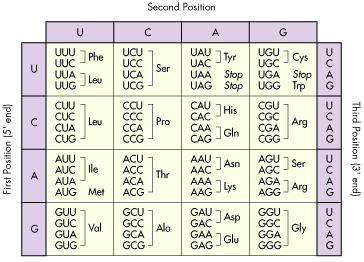


When there is no hit, the output will look like this:



*Obtaining a tRNA gene count from this output*

1. Copy the output to Microsoft Word
2. Use the search feature (ctrl + f) to find the three letter amino acid acronym e.g. Asn for each of the different type of tRNA. Add the number of hits for each acronym. This will give you a complete tRNA count. These acronyms are given in the table below:



**tRNA Finder**

*Background*

Like ARAGORN, tRNA Finder also employs various algorithms based on homology between conserved tRNA gene sequences and the ability to form tRNAs’ characteristic secondary structure. It is available online as well.

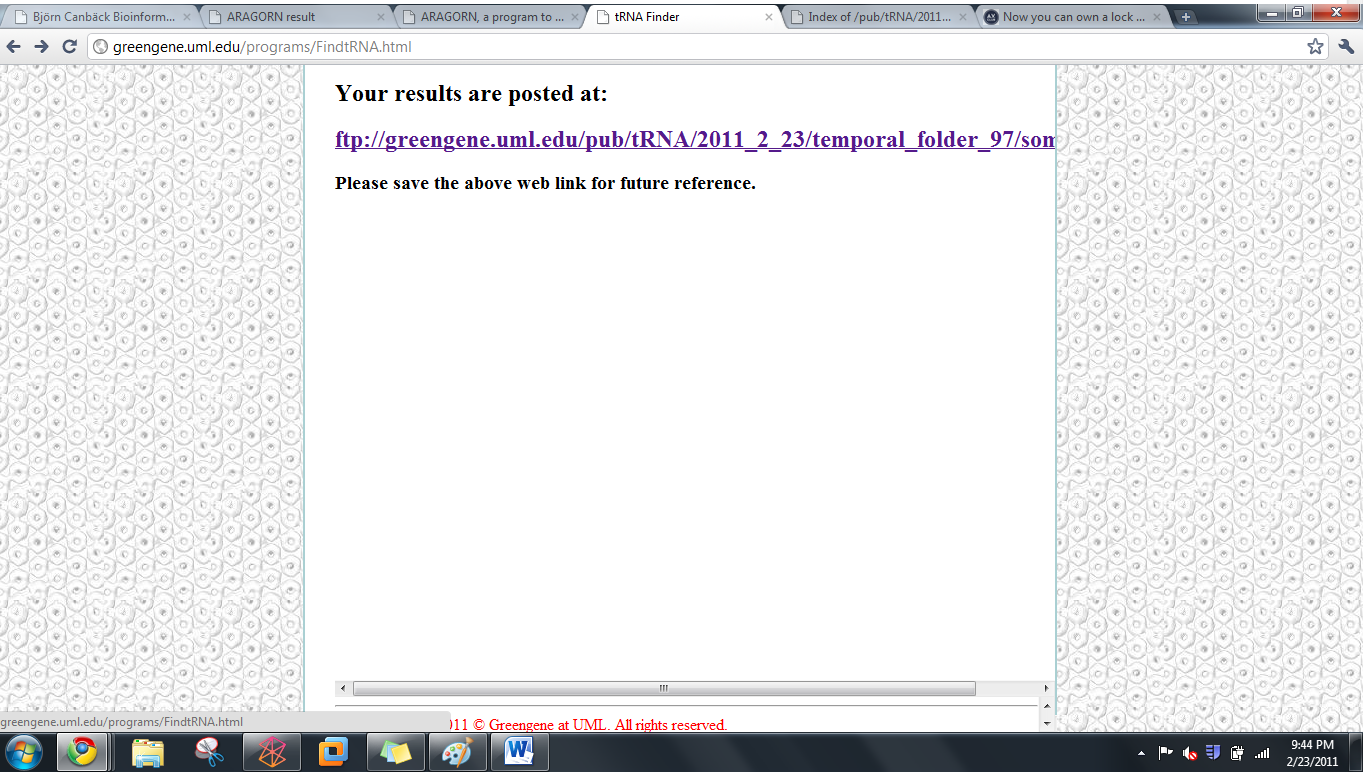
*Using tRNA finder*

tRNA finder is available as an online program as well. A description of how to use the online tool is available below:

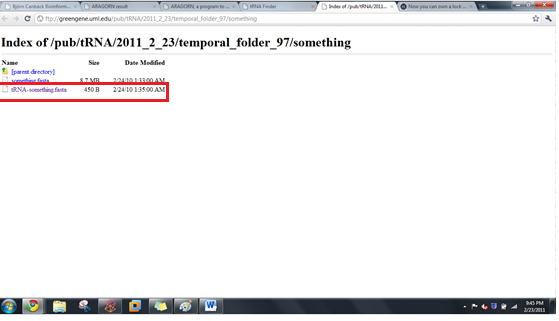
1. To access tRNA finder, use the following link: <http://greengene.uml.edu/programs/FindtRNA.html>
2. You can upload a genome as either a .fasta or .zip file by clicking the “choose file” button (in red box below).



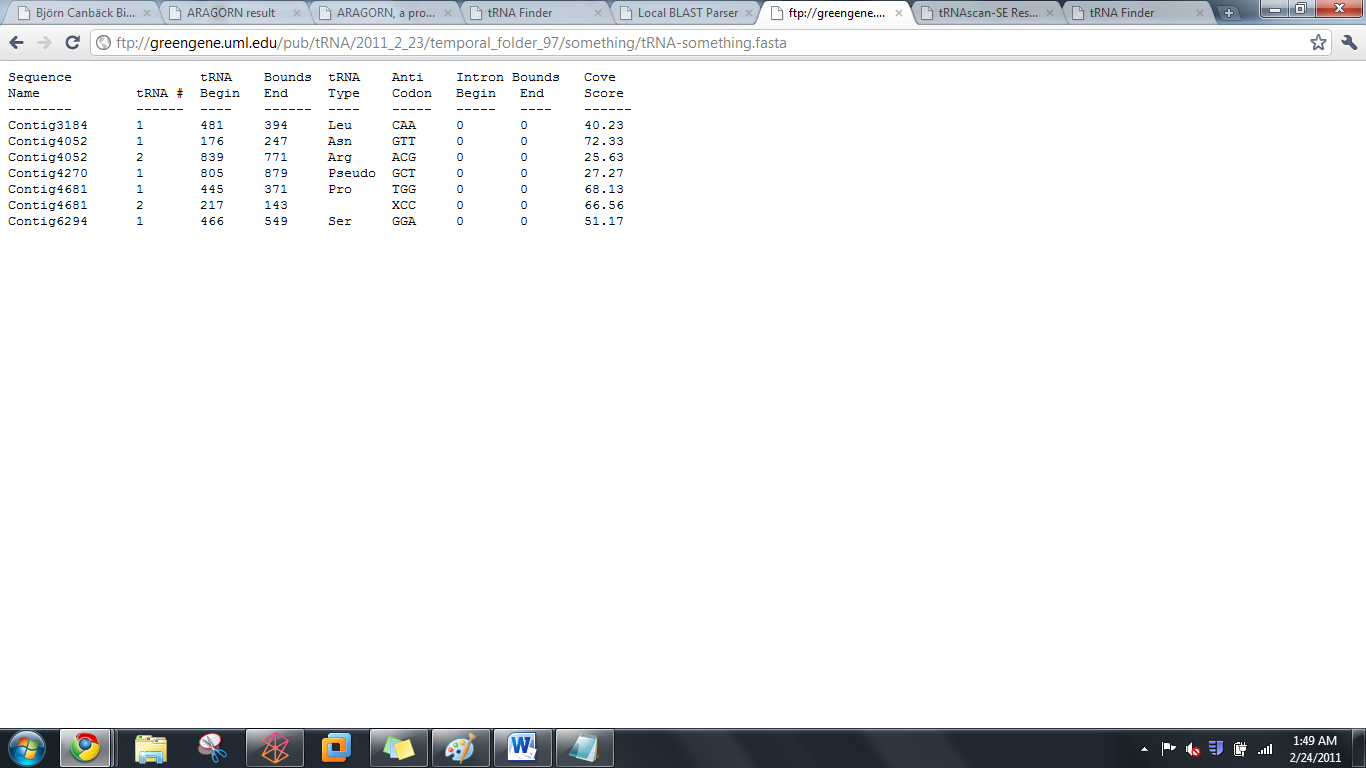
1. Clicking on the chose file button links you to a page containing a link to your results (depicted below). Click on this link.



1. Clicking on this link will direct you to a root directory page (depicted below). The bottom link will open the tRNA results.



1. The results output is organized by sequence name, number of tRNAs, location, tRNA type, anticodon, intron beginning, bounds end, and core score (a rating of how likely the identified candidate is a tRNA). It should look like this:



*Obtaining a tRNA gene count from tRNA finder*

1. Copy the output to Microsoft Word
2. Use the search feature (ctrl + f) to search for the word “contig.” The number of times the word is repeated is your tRNA gene count.

**tRNA scan-SE**

*Background*

Formed by the Lowe lab, tRNA scan-SE uses other tRNA identification software as first pass pre-filters to identify candidate tRNA gene sequences which are then analyzed by a highly selective covariance model. It has an error rate of one false positive in 15 gigabases.

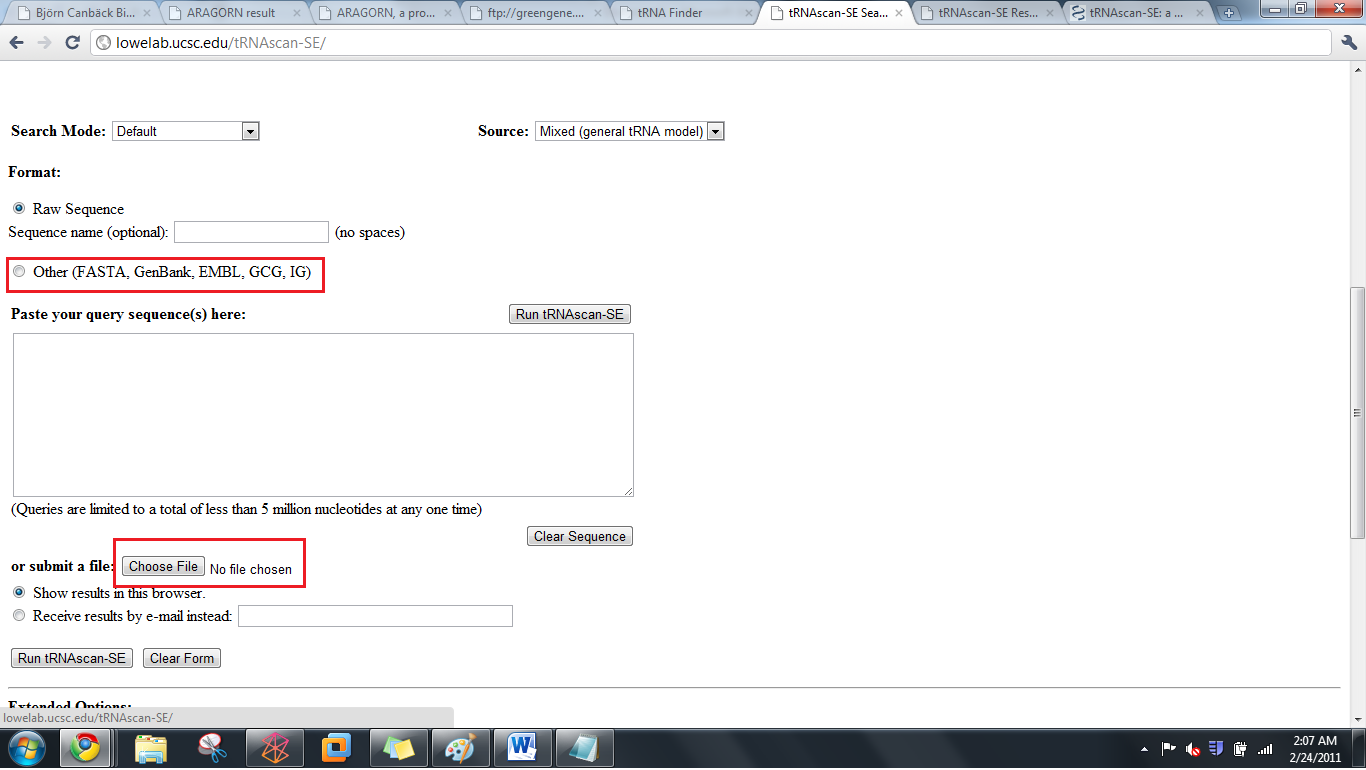
*Using tRNA scan-SE*

tRNA scan-SE can be downloaded as a program similar to BLAST. A less stable online version is available as well. It can analyze up to 15 mb worth of data. A description of how to use the online tool is provided below:

1. To access tRNA scan-SE, use the following link:

<http://lowelab.ucsc.edu/tRNAscan-SE/>

1. You can upload a genome (most likely has to be parsed to fit the size limit and avoid a browser crash) as either raw data or a .fasta file.
   1. To enter the genome as a nucleotide sequence i.e. raw data, enter the genome into the box that says “paste your query sequence(s) here”
   2. To enter the genome as a file, click on the “choose data” button and click on the upload file button (both highlighted below)



1. You will be redirected to your results which will look like this:

|  |  |
| --- | --- |
| [UCSC] | **Lowe Lab** tRNAscan-SE Search Results |

| [Lowe lab](http://lowelab.ucsc.edu/) | [tRNAscan-SE](http://lowelab.ucsc.edu/tRNAscan-SE/) | [Snoscan](http://lowelab.ucsc.edu/snoscan/) | [SnoGPS](http://lowelab.ucsc.edu/snoGPS/) | [Software](http://lowelab.ucsc.edu/software/) | [Publications](http://www.soe.ucsc.edu/~lowe/pubs/) |

**tRNAscan-SE v.1.21 (Oct 2000)**

- scan sequences for transfer RNAs

Please cite:

Lowe, T.M. & Eddy, S.R. (1997) "tRNAscan-SE: A program for

improved detection of transfer RNA genes in genomic sequence"

Nucl. Acids Res. 25: 955-964.

This program uses a modified, optimized version of tRNAscan v1.3

(Fichant & Burks, J. Mol. Biol. 1991, 220: 659-671),

a new implementation of a multistep weight matrix algorithm

for identification of eukaryotic tRNA promoter regions

(Pavesi et al., Nucl. Acids Res. 1994, 22: 1247-1256),

as well as the RNA covariance analysis package Cove v.2.4.2

(Eddy & Durbin, Nucl. Acids Res. 1994, 22: 2079-2088).

Cove-predicted tRNA secondary structures rendered by

NAVIEW (c) 1988 Robert E. Bruccoleri.

**Results**

Sequence tRNA Bounds tRNA Anti Intron Bounds Cove

Name tRNA # Begin End Type Codon Begin End Score

-------- ------ ---- ------ ---- ----- ----- ---- ------

|  |
| --- |
| Your-seq 1 178 249 Asn GTT 0 0 72.33 |
| Top of Form  Bottom of Form |
| Your-seq 2 836 763 Arg ACG 0 0 64.16 |
| Top of Form  Bottom of Form |

**Run Statistics:**

Started: Wed Feb 23 20:07:00 PST 2011

------------------------------------------------------------

Search Mode: General

Searching with: tRNAscan + EufindtRNA -> Cove

Covariance model: TRNA2.cm

tRNAscan parameters: Strict

EufindtRNA parameters: Relaxed (Int Cutoff= -32.1)

------------------------------------------------------------

First-pass (tRNAscan/EufindtRNA) Stats:

---------------

Sequences read: 1

Seqs w/at least 1 hit: 1

Bases read: 1021 (x2 for both strands)

Bases in tRNAs: 146

tRNAs predicted: 2

Av. tRNA length: 73

Script CPU time: 0.00 s

Scan CPU time: 0.00 s

Scan speed: 2042.0 Kbp/sec

First pass search(es) ended: Wed Feb 23 20:07:00 PST 2011

Cove Stats:

-----------

Candidate tRNAs read: 2

Cove-confirmed tRNAs: 2

Bases scanned by covels: 174

% seq scanned by covels: 8.5 %

Script CPU time: 0.00 s

Cove CPU time: 0.22 s

Scan speed: 790.9 bp/sec

Cove analysis of tRNAs ended: Wed Feb 23 20:07:00 PST 2011

Summary

--------

Overall scan speed: 9281.8 bp/sec

tRNAs decoding Standard 20 AA: 2

Selenocysteine tRNAs (TCA): 0

Possible suppressor tRNAs (CTA,TTA): 0

tRNAs with undetermined/unknown isotypes: 0

Predicted pseudogenes: 0

-------

Total tRNAs: 2

tRNAs with introns: 0

|

Isotype / Anticodon Counts:

Ala : 0 AGC: GGC: CGC: TGC:

Gly : 0 ACC: GCC: CCC: TCC:

Pro : 0 AGG: GGG: CGG: TGG:

Thr : 0 AGT: GGT: CGT: TGT:

Val : 0 AAC: GAC: CAC: TAC:

Ser : 0 AGA: GGA: CGA: TGA: ACT: GCT:

Arg : 1 ACG: 1 GCG: CCG: TCG: CCT: TCT:

Leu : 0 AAG: GAG: CAG: TAG: CAA: TAA:

Phe : 0 AAA: GAA:

Asn : 1 ATT: GTT: 1

Lys : 0 CTT: TTT:

Asp : 0 ATC: GTC:

Glu : 0 CTC: TTC:

His : 0 ATG: GTG:

Gln : 0 CTG: TTG:

Ile : 0 AAT: GAT: TAT:

Met : 0 CAT:

Tyr : 0 ATA: GTA:

Supres: 0 CTA: TTA:

Cys : 0 ACA: GCA:

Trp : 0 CCA:

SelCys: 0 TCA:

**Predicted tRNA Secondary Structures:**

Your-seq.trna1 (178-249) Length: 72 bp

Type: Asn Anticodon: GTT at 33-35 (210-212) Score: 72.33

\* | \* | \* | \* | \* | \* | \* |

Seq: TCCTCAGTAGCTCAGTGGTAGAGCGGTCGGCTGTTAACTGATTGGtCGTAGGTTCGAATCCTACTTGGGGAG

Str: >>>>>>>..>>>>.......<<<<.>>>>>.......<<<<<.....>>>>>.......<<<<<<<<<<<<.

Your-seq.trna2 (836-763) Length: 74 bp

Type: Arg Anticodon: ACG at 35-37 (802-800) Score: 64.16

\* | \* | \* | \* | \* | \* | \* |

Seq: GGGCCTGTAGCTCAGAGGAttAGAGCACGTGGCTACGAACCACGGTGtCGGGGGTTCGAATCCCTCCTCGCCCA

Str: >>>>..>..>>>>.........<<<<.>>>>>.......<<<<<.....>>>>>.......<<<<<<..<<<<.

*Getting a tRNA count from tRNA scan-SE output*:

1. Look at the final hit in the results section. The number associated with it will be your tRNA count.

**tRNA Database Curated Manually by Experts**

*Background*

tRNADB-CE is a database containing 927 complete and 1301 draft genomes of Bacteria and Archaea, 171 complete virus genomes, 121 complete chloroplast genomes, 12 complete eukaryote (Plant and Fungi) genomes and approximately 230 million DNA sequence entries that originated from environmental metagenomic clones. These genomes are analyzed using ARAGORN, tRNA finder, and tRNA scan-se. They have a concordance rate of 96%; the remaining 4% is analyzed and annotated by experts.

*Using tRNADB-CE*

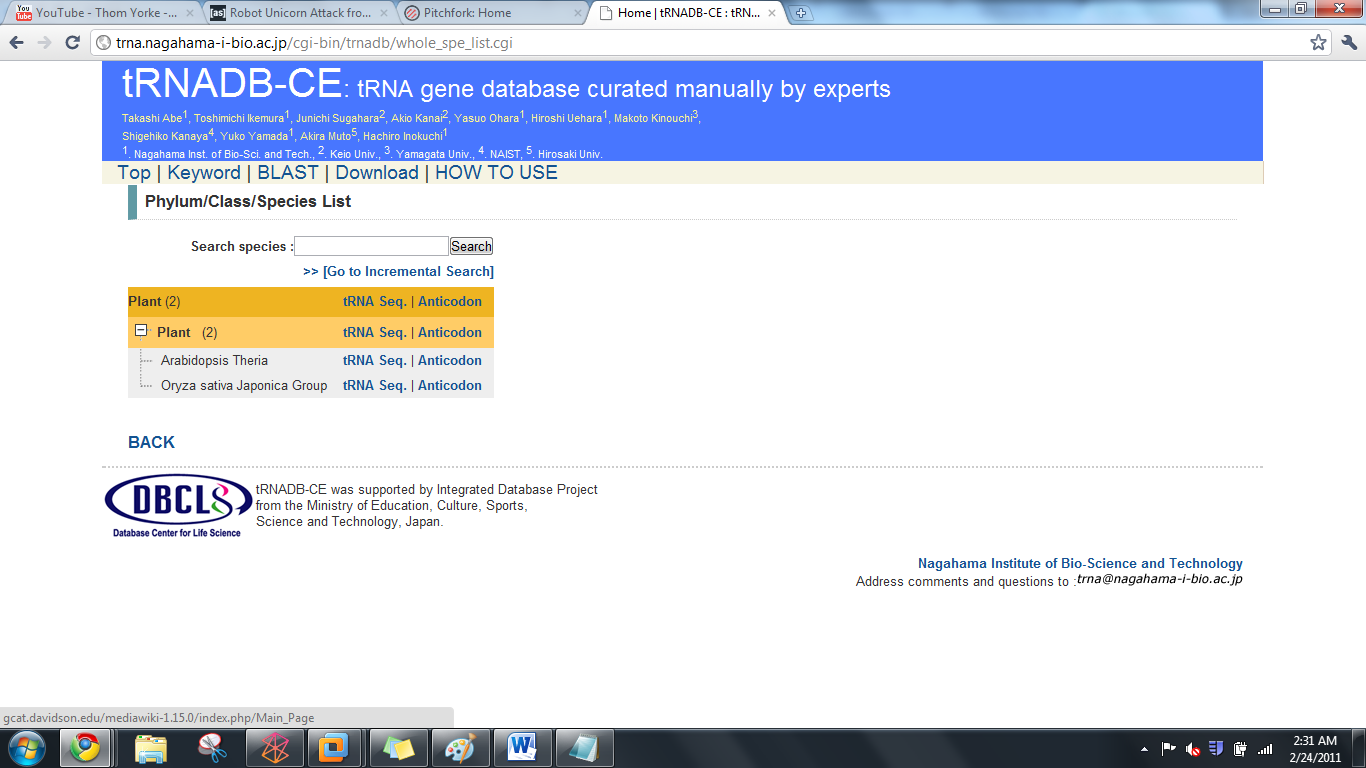
1. To access tRNADB-CE, go to the following URL:

<http://trna.nagahama-i-bio.ac.jp/cgi-bin/trnadb/index.cgi>

1. On this page, you can access which species database you want to use (depicted below):



1. Click on the desired genome (in this case, I chose plants). You will be redirected to a page that looks like this:

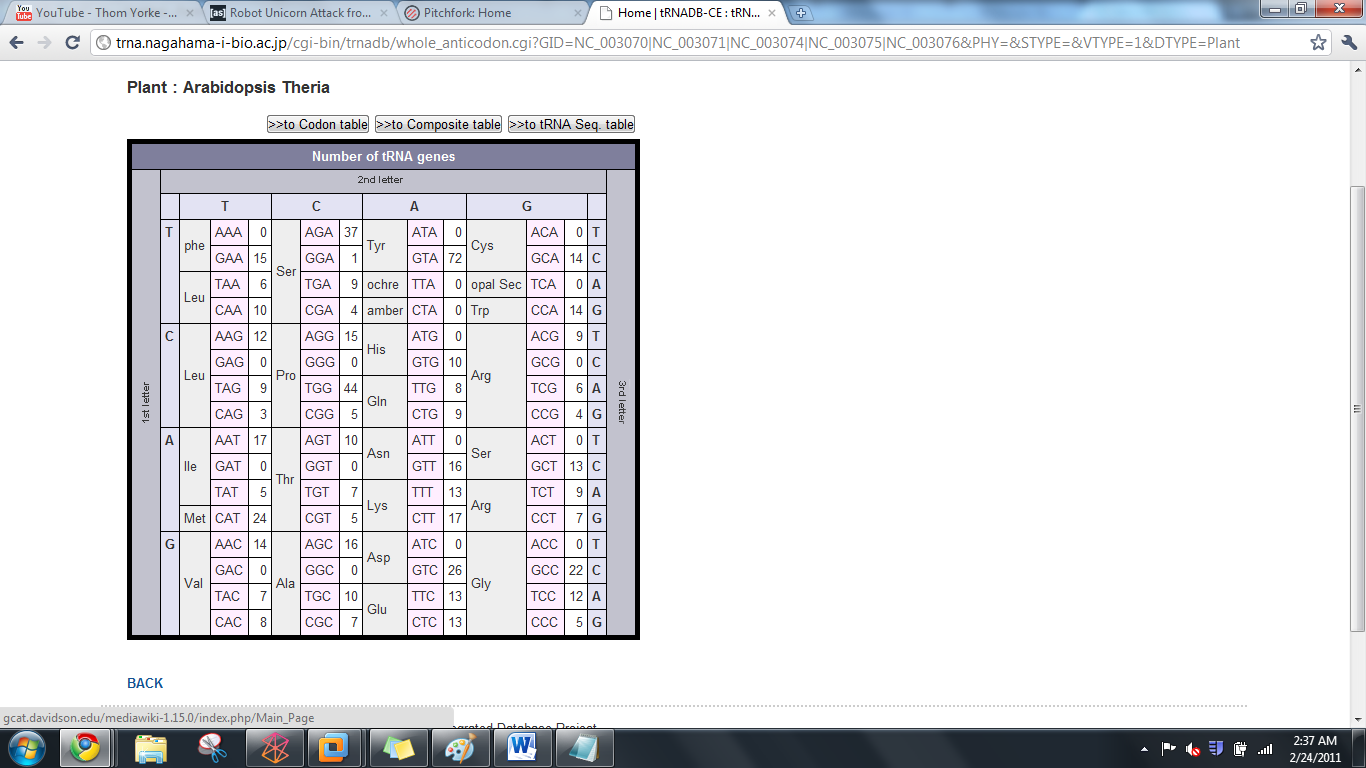


1. You can either:
   1. Click on tRNA seq to get a .fasta file containing candidate tRNA genes for your chosen species. You will be linked to a site that looks like this:



From here, click on the check all button (highlighted above) and the FASTA button (highlighted above) to download a .fasta file containing all the candidate tRNA genes your chosen species

* 1. Click on Anticodon to see how many of each type of tRNA there is in the genome of your choice. You will be linked to a site that looks like this:



*Using your fasta file with tRNA genes*

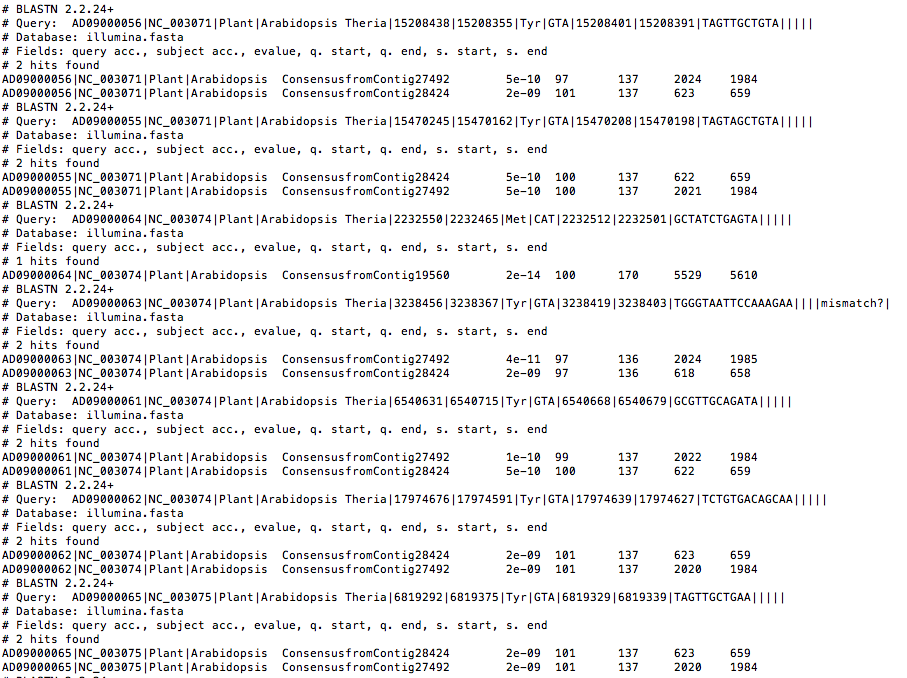
1. This file obtained in step 4a is going to be your reference. It contains the sequences for the tRNA genes in your species of choice (in this case *Arabadopsis*).
2. Using your genome of question as a database and the fasta file containing your reference tRNA genes, run BLAST. Your input commands should look like this:
   1. To make your database:

w10120:Desktop YOURUSERNAME$ /usr/local/ncbi/blast/bin/makeblastdb -title testdb -dbtype nucl -in YOUR GENOME OF QUESTION.fasta

* 1. To actually query this database using your reference tRNA gene sequence obtained from the tRNADB-CE:

w10120:Desktop YOURUSERNAME$ /usr/local/ncbi/blast/bin/blastn -query REFERENCE TRNA LIST FROM tRNADB-CE.fasta -db YOUR GENOME OF QUESTION.fasta -outfmt "7 qacc sacc evalue qstart qend sstart send"

1. You will get the following as your output:



*Obtaining a tRNA Gene Count from this Output*

1. Copy this output to Microsoft Word.
2. Use the find and replace feature (ctrl + f) to find and replace the words “0 hits found” with another word/phrase e.g. unicorn
3. Use the find feature (ctrl + f) to search for the words “hits found.” The number of times it finds this phrase will give your tRNA count.