tRNAscan-SE

by Kevin Gleason

April 28, 2015

What is tRNAscan-SE

- Peter Schattner, Angela N. Brooks and Todd M. Lowe
  - UCSC RNA Center, 2005
- Identifies 99–100% tRNA genes in DNA sequence
- Less than one false positive per 15 gigabases.
- Offered as web server and UNIX package
- Solves the issue of custom programming for RNA identification

Purpose

- Transfer RNA genes make up the single largest gene family
- Conventional gene finders do not target tRNA
  - Lack codon structure and statistical signatures of protein coding genes
- tRNA affects the Codon Bias in protein coding genes

Web Server

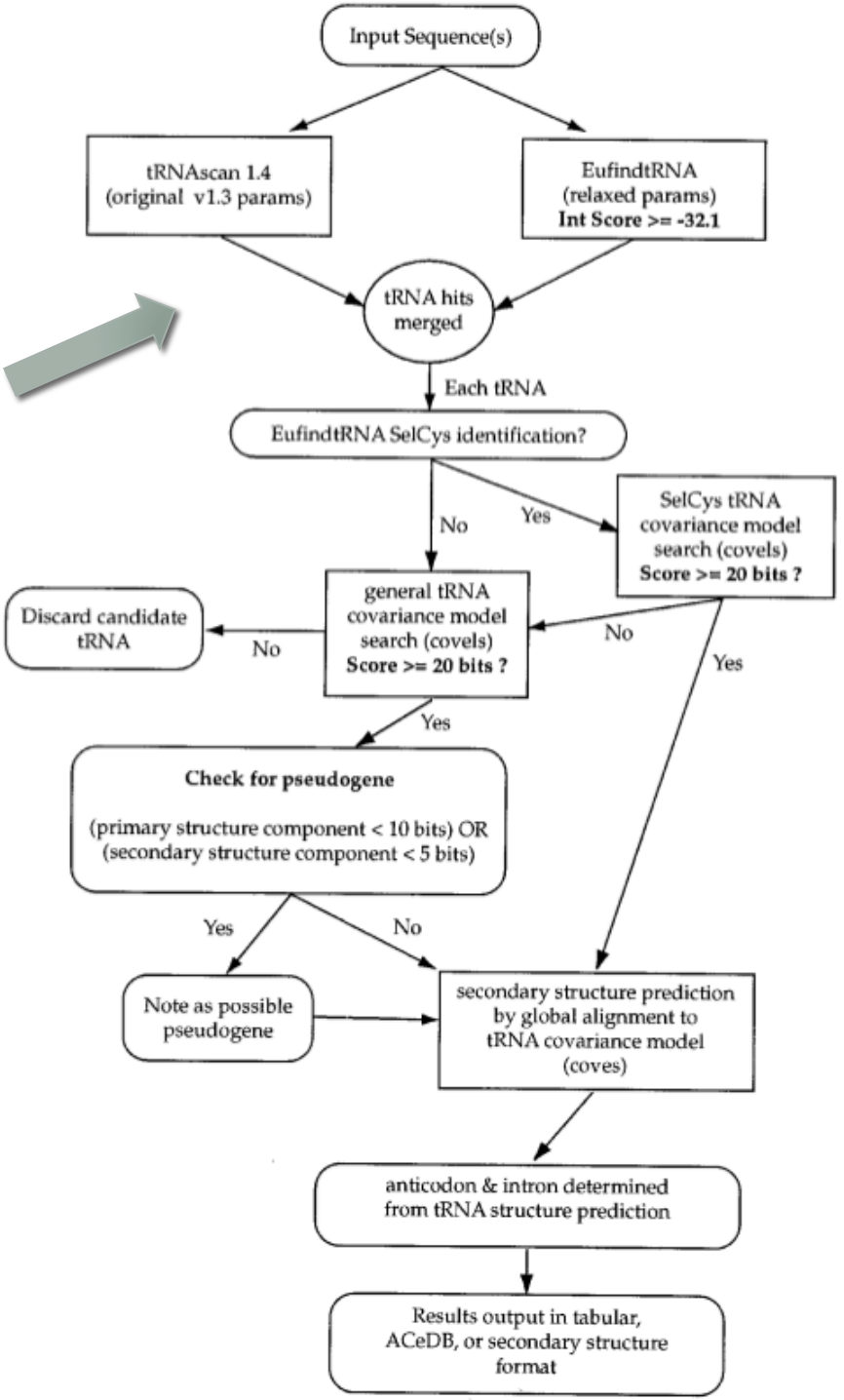
- The search mode selection
  - Determines which probabilistic model to be used in
  - Training data from selected species or phylogenetic groups (mammals, yeasts and archaea)
- Query Sequence Selection
  - Sequences to be searched

Stochastic CFG

- Watson-Crick and GU Wobble
- $S \rightarrow \lambda \mid AS \mid CS \mid GS \mid US \mid ASU \mid USA$
  $\mid CSG \mid GSC \mid GSU \mid USG \mid SS$
- Probabilities from frequencies of features in databases of RNA structure
- Can predict pairwise and nested structures, but not pseudoknots

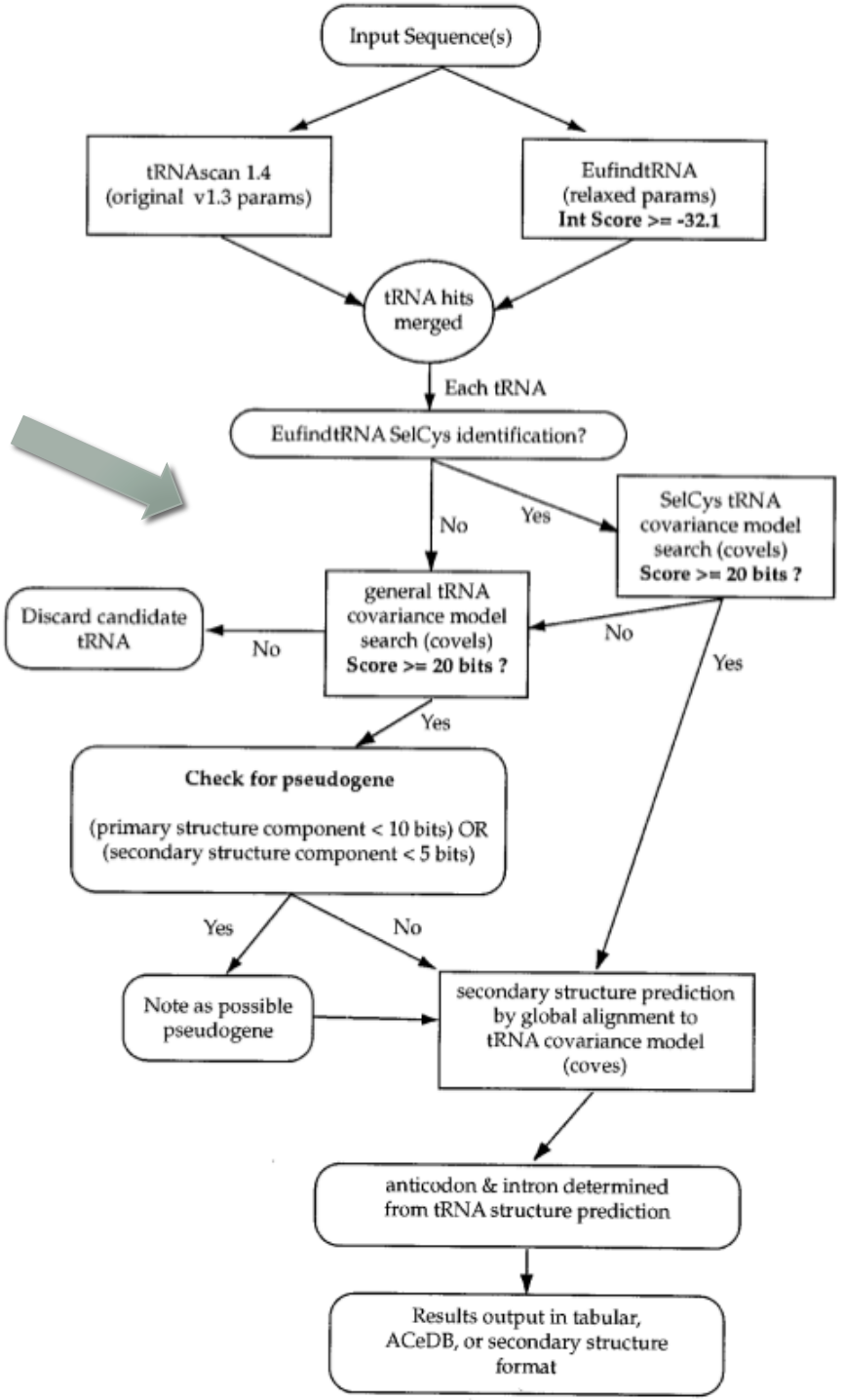
The Algorithm Phase 1

- tRNAscan and Parvesi Algorithm
- Parvesi
  - Linear sequence signals
  - RNA polymerase III promoters and terminators
  - Less specific, but predicts tRNAs that tRNAscan misses
- Results merged



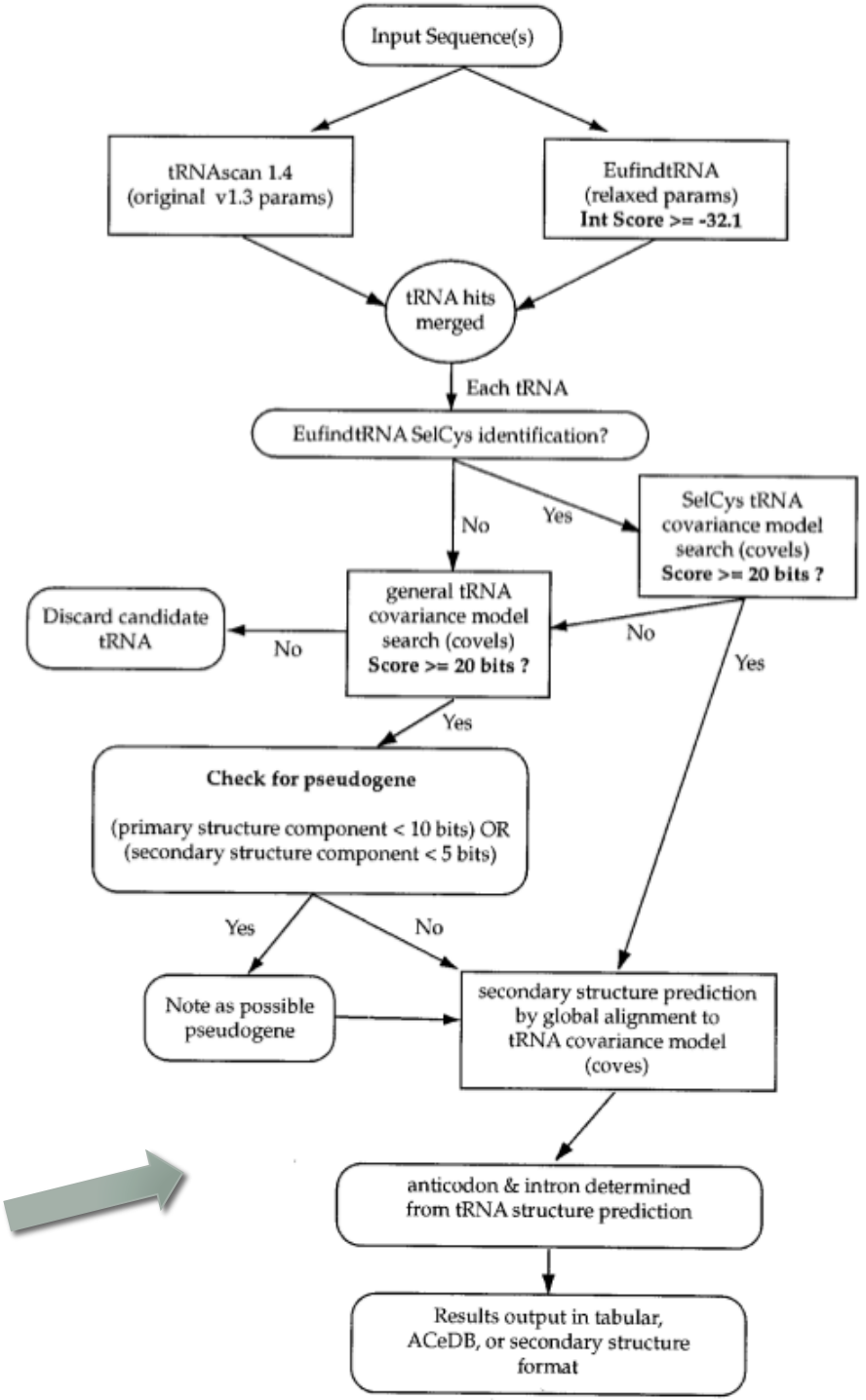
The Algorithm Phase 2

- Extracts the candidate subsequences
  - 7nt padding
- Applies a tRNA covariance model
- Structurally aligned 1415 tRNAs from modified 1993 Sprinzl database
- Discard all low scores



The Algorithm Phase 3

- tRNA with log odds scores over 20.0 bits
- Trims bounds to covels predictions
- Predicts secondary structure with coves
  - Pseudogenes filtered out
- A second profile created from HMM
  - No secondary structure is conserved
- Predictions formatted and saved



Output

- Final tRNA predictions are saved in tabular, ACeDB or secondary structure output format.
- AceDB - originally "**A C**. **e**legans **D**ata**B**ase" - is a genome database designed for handling bioinformatic data

(A) tRNAscan-SE Output

```
Sequence                        tRNA  Bounds   tRNA Anti  Intron Bounds Cove
Name                  tRNA #    Begin End      Type Codon Begin  End    Score
--------              ------    ---- ------    ---- ----- -----  ----   ------
chr6.trna18-AlaAGC      1         1    73       Ala  AGC    0      0      40.39

tRNAs decoding Standard 20 AA:            1
Selenocysteine tRNAs (TCA):               0
Possible suppressor tRNAs (CTA,TTA):      0
tRNAs with undetermined/unknown isotypes: 0
Predicted pseudogenes:                    0
                                         ---
Total tRNAs:                              1

Isotype / Anticodon Counts:
Ala    : 1     AGC: 1       GGC:         CGC:         TGC:
Gly    : 0     ACC:         GCC:         CCC:         TCC:

Predicted tRNA Secondary Structures:
chr6.trna18-AlaAGC.trna1 (1-73)   Length: 73 bp
Type: Ala   Anticodon: AGC at 34-36 (34-36)   Score: 40.39
         *    |    *    |    *    |    *    |    *    |    *    |    *    |
Seq: GGGGGATTAGCTCAAGCGGTAGGGTGCCTGCTTAGCATGCAAGAGGtAGCAGGATCGACGCCTGCATTCTCCA
Str: >>>>>>>..>>>>........<<<<.>.>>>.......<<<.<.....>>>>>.......<<<<<<<<<<<.
```

Output

Selenocysteine tRNAs (TCA)	4
Possible suppressor tRNAs (CTA,TTA)	0
tRNAs with undetermined or unknown isotypes	10
Predicted pseudogenes	224
Total tRNAs	1516

Intron Summary

tRNAs with introns	Gly GCC	Pro TGG	Val CAC	Val TAC	Arg GCG	Arg CCT	Leu CAA	Asn ATT	Lys TTT
16	1	1	1	1	1	1	8	1	1

Four Box tRNA Sets

Isotype	tRNA Count by Anticodon				Total
Ala	AGC 51	GGC 1	CGC 13	TGC 16	81
Gly	ACC	GCC 35	CCC 2	TCC 51	88
Pro	AGG 33	GGG	CGG 12	TGG 13	58
Thr	AGT 34	GGT 58	CGT 18	TGT 11	121
Val	AAC 28	GAC	CAC 19	TAC 9	56

Six Box tRNA Sets

Isotype	tRNA Count by Anticodon								Total
Ser	AGA 22	GGA 1	CGA 10	TGA 10	ACT 1	GCT 14			58
Arg	ACG 138	GCG 1	CCG 2	TCG 11			CCT 7	TCT 13	172
Leu	AAG 33	GAG	CAG 16	TAG 6			CAA 8	TAA 3	66

Two Box tRNA Sets

Isotype	tRNA Count by Anticodon				Total
Phe	AAA	GAA 23			23
Asn	ATT 12	GTT 34			46
Lys			CTT 60	TTT 70	130
Asp	ATC	GTC 49			49

Two Box & Other tRNA Sets

Isotype	tRNA Count by Anticodon				Total
Ile	AAT 32	GAT 1		TAT 18	51
Met			CAT 37		37
Tyr	ATA 2	GTA 29			31
Supres			CTA	TTA	0