## UC Santa Cruz UC Santa Cruz Previously Published Works

### Title

tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes

## Permalink

https://escholarship.org/uc/item/5fb8q49k

**Journal** Nucleic Acids Research, 44(W1)

**ISSN** 0305-1048

Authors Lowe, Todd M Chan, Patricia P

Publication Date 2016-07-08

## DOI

10.1093/nar/gkw413

## **Copyright Information**

This work is made available under the terms of a Creative Commons Attribution-NoDerivatives License, available at <u>https://creativecommons.org/licenses/by-nd/4.0/</u>

Peer reviewed

# tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes

#### Todd M. Lowe<sup>\*</sup> and Patricia P. Chan

Department of Biomolecular Engineering, University of California Santa Cruz, CA 95064, USA

Received March 13, 2016; Revised April 29, 2016; Accepted May 4, 2016

#### ABSTRACT

High-throughput genome sequencing continues to grow the need for rapid, accurate genome annotation and tRNA genes constitute the largest family of essential, ever-present non-coding RNA genes. Newly developed tRNAscan-SE 2.0 has advanced the stateof-the-art methodology in tRNA gene detection and functional prediction, captured by rich new content of the companion Genomic tRNA Database. Previously, web-server tRNA detection was isolated from knowledge of existing tRNAs and their annotation. In this update of the tRNAscan-SE On-line resource, we tie together improvements in tRNA classification with greatly enhanced biological context via dynamically generated links between web server search results, the most relevant genes in the GtRNAdb and interactive, rich genome context provided by UCSC genome browsers. The tRNAscan-SE On-line web server can be accessed at http://trna.ucsc.edu/tRNAscan-SE/.

#### INTRODUCTION

Transfer RNAs, a central component of protein translation, represent a highly complex class of genes that are ancient vet still evolving. tRNAscan-SE remains the *de facto* tool for identifying tRNA genes encoded in genomes with over 5000 citations (1), and has a wide variety of users including sequencing centers, biological database annotators, RNA biologists and computational biology researchers. Transfer RNAs predicted using tRNAscan-SE are currently available for thousands of genomes in the GtRNAdb (2,3), yet the GtRNAdb is not designed for user-driven tRNA gene detection. We previously described the tRNAscan-SE web server 11 years ago (4), a publication that has garnered over 750 citations (Google Scholar) and has been downloaded for full-text viewing more than 7500 times (Nucleic Acids Research Article Metrics). The original tRNAscan-SE web server currently receives about 1000 unique visitors a month, aiding a large swath of the research community that may not have computational resources or expertise to install and run the UNIX-based software.

Here, we describe a new web analysis server in conjunction with the development of tRNAscan-SE 2.0 (Chan et al., in preparation). The new version of tRNAscan-SE has improved covariance model search technology enabled by the Infernal 1.1 software (5); updated covariance models for more sensitive tRNA searches, leveraging a much broader diversity of tRNA genes from thousands of sequenced genomes; better functional classification of tRNAs, based on comparative analysis using a suite of 22 isotype-specific tRNA covariance models for each domain of life; and ability to detect mitochondrial tRNAs (in addition to cytosolic eukaryotic, archaeal and bacterial tRNAs), with high accuracy using new mitochondrial-specific models. In addition to new search capabilities, the web server now enables users to place their tRNA predictions in similarity context within the GtRNAdb, as well as genomic context within available UCSC genome browsers.

#### **TRNA GENE SEARCH**

Similar to the original version of the tRNAscan-SE search server (4), users may select among multiple types of tR-NAs (mixed/general, eukaryotic, bacterial, archaeal or mitochondrial) as well as different search modes to identify tRNA genes in their provided sequences. The default search mode of tRNAscan-SE 2.0 (Chan *et al.*, in preparation) utilizes the latest version of the Infernal software package (v1.1.1) (5) to search DNA sequences for tRNA-like structure and sequence similarities. The Infernal software implements a special case of profile stochastic context-free grammars called covariance models, which can be trained to have different specificities depending on the selection of structurally aligned RNA sequences which serve as training sets. tRNAscan-SE 2.0 employs a suite of covariance models in multiple analysis steps to maximize sensitivity and classification accuracy.

In an initial first-pass scan, Infernal is used with a relatively permissive score threshold (10 bits) in combination with a search model trained on tRNAs from all tRNA isotypes in order to obtain high sensitivity. The mid-level strictness filter ('--mid') is also used to accelerate search speed at minimal cost to sensitivity. The user can choose to search with a tRNA model trained on tRNAs from all domains

© The Author(s) 2016. Published by Oxford University Press on behalf of Nucleic Acids Research.

<sup>\*</sup>To whom correspondence should be addressed. Tel: +1 831 459 1511; Fax: +1 831 459 4829; Email: lowe@soe.ucsc.edu

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

of life ('Mixed'), or preferably, select from one of three domain-specific models trained on tRNAs exclusively from Archaea, Bacteria or Eukarya. A second-pass scan of individual candidates detected in the first-pass also uses Infernal (5) but with a higher score threshold to increase selectivity, no acceleration filter to increase alignment accuracy and multiple isotype-specific covariance models to better determine isotype identity. These changes in the tRNAscan-SE 2.0 software and search models produce slightly different bit scores relative to tRNAscan-SE 1.2.1 (1), although the relative rankings of previously detected tRNAs should largely be unchanged. In order to provide backward comparisons when needed, researchers can select a 'legacy' search mode which uses the tRNAscan-SE 1.2.1 search software. For users who need maximum search sensitivity for low-scoring tRNA-like sequences (e.g. tRNA-derived SINEs or pseudogenes), we include the search mode option 'Infernal without HMM' which is only recommended for very short sequence queries due to the slow search speed.

#### FUNCTIONAL CLASSIFICATION USING ISOTYPE-SPECIFIC COVARIANCE MODELS

Sequence and structure-based determinants, both positive and negative, are used by aminoacyl-tRNA synthetases to establish tRNA identity, and have been characterized in a number of model species (6). Previously, we used the anticodon exclusively to predict tRNA isotype because the anticodon can be easily identified within a tRNA gene candidate and nearly always gives unambiguous identification of the tRNA isotype. However, there exist cases of 'chimeric' tRNAs in which point mutation(s) in the anticodon sequence could result in 'recoding' events (7,8). In this case, the body of the tRNA contains identity elements recognized by one type of tRNA synthetase, yet the altered anticodon 'reads' the mRNA codon corresponding to a different amino acid. The decoding behavior of such tRNAs could theoretically be modulated by post-transcriptional anticodon modifications that either preserve or alter the genetic code. The full biological scope and significance of chimeric tRNAs is not well understood because of the historical lack of large-scale, systematic detection methods, but it is now possible and valuable to detect these potentially important chimeric tRNAs which could result in tissue or condition-specific recoding of proteins.

Thus, we developed a new multi-model annotation strategy for tRNAscan-SE 2.0 where, after establishing tRNA gene coordinates and predicting function by anticodon, we also analyze the gene prediction with a full set of isotype-specific covariance models, in a strategy similar to TFAM (9). These 20+ models are built by simply sub-dividing the original tRNA training set into subgroups of the universal 20 amino acids (Ala, Arg, Cys, etc), plus one for initiator/formyl-methionine (iMet/fMet), one to identify prokaryotic Ile tRNAs genomically encoded with a CAT anticodon, and one to recognize selenocysteine tRNAs. Each of these subgroups forms the basis for 20+ models for each domain (22 for eukaryotes, 23 for bacteria, 23 for archaea). Now, alongside the predicted isotype based on the anticodon, the highest scoring isotype-specific model is also reported; any disagreement between the two functional prediction methods is reported for closer user inspection. There may be insufficient data to establish the true tRNA identity when there is disagreement because tRNA synthetase identity elements have only been experimentally verified in a small number of species. However, we believe this supplemental isotype-specific model analysis will enable the tRNA research community to more readily identify and experimentally investigate potential tRNA chimeras in the future. An example of this type of potential chimeric tRNA is human Val-AAC-6-1 ('GGGGGTGT AGCTCAGTGGTAGAGCGTATGCTTAACATTCAT GAGGCTCTGGGTTCGATCCCCAGCACTTCCA') that contains the anticodon AAC (in bold), but scores highest against the eukaryotic tRNA isotype model for alanine.

#### EXPLORING TRNA CONTEXT: GTRNADB AND LO-CUS CONTEXT

The resulting tRNAs identified in the user's sequence can be searched against the GtRNAdb (3), yielding links to identical or highly similar tRNA genes found in the database (Figure 1). If a UCSC Genome Browser (10,11) exists for any identical or close tRNA matches, a direct link is provided to those matches, enabling the user to examine its genomic context and any additional information available in genome browser tracks. In the example given (Figure 1), 'MySeq3' contains one tRNA prediction yielding an Infernal score of 71.6 bits, as scored by an 'all-isotype' eukaryotic tRNA model. The identified tRNA has one intron and is predicted to be charged with serine based on the CGA anticodon inferred from the predicted tRNA secondary structure. Upon comparison to the full suite of specialized/isotype-specific models, the highest scoring model in second-pass analysis corresponded to tRNA-Serine, at 117.9 bits. tRNAs will usually get a higher score against their true isotypespecific model because specialized models are not 'diluted' by tRNA sequence features found only in other isotypes. Selecting the first button in each row visualizes the predicted secondary structure, while the second button executes a fast sequence similarity search to find identical or very similar tRNAs in the GtRNAdb. The figure shows perfect matches to Ser-CGA tRNAs in several Saccharomyces species; upon selecting the 'View' link for Saccharomyces cerevisiae Ser-CGA-1-1, its individual gene page displays a wealth of information, including upstream and downstream genomic sequences, atypical features (U51:U63), 13 RNA modifications previously characterized, a multiple sequence alignment with all other Ser tRNA genes in this species (not shown) and tRNA-seq expression profiles for mature and pre-tRNAs mapped to this locus (not shown). Finally, links from either the GtRNAdb gene page, or the list of perfect matching hits shows the tRNA gene in the interactive UCSC Genome Browser with tracks that show the level of multi-genome conservation among six other yeast species, the positions of previously noted modifications, tRNA-seq expression data and the 'SGD Genes' track, aiding further exploration of this gene's biological context.

#### Results

Download as text

lame	tRNA #	tRNA Structure	tRNAs in GtRNAdb	tRNA Begin	tRNA End	tRNA Type	Anticodon	Intron Begin	Intron End	Infernal Score	Pseudo	lsotype Model	Isotype Score
lySeq1	1	View	View	13	85	Thr	TGT	0	0	78.2	No	Thr	93.6
lySeq2	1	View	View	6	79	Arg	тст	0	0	75.1	No	Arg	89.2
lySeq3	1	View	View	14	114	Ser	CGA	51	69	71.6	No	Ser	117.9
U G A G	Ltma1 Sor (CGA) G C C G C U C C G U C G U G U A U U G	71.6 bits	a c	Your tF MySeq3.tr GGCACTA Sumn Perfect # tR	RNAscan-R na1 TGGCCGAG1 nary matching NA	SE Hit: Iggttaaggca GtRNAdb s	GAGAGACTCGAAT	GGAATAAAAA	GTTCGGCTAT	CTCTTGGGCTC	racccacact	GGTTCAAATCO	CTGCTGGTG View In GtRNAd
U A A G	, o	G		1 Sa	ccharomvce	s cerevisiae	tRNA-Ser-CGA-1	-1		101 bi	0 101/10	)1 (100%)	View
A.,	G G G	с с с		2 5	ccharomyce	s sp boulard	ii ATCC MYA-79	6 tRNA-Ser-C	GA-1-1	101 bi	0 101/10	01 (100%)	View
G	с С	c <sup>•</sup> u		-	loonaronnyoo.	5_5P_boulard		0_111174 001-0		101.0	5 101/10	/1 (10070)	VICW
• Infron	charomyc	ces_cerevisi	ae_tRNA-Ser	-CGA-1-1 (	tRNAscan-	SE ID: chi	-III.trna4)	Gene	tRNA-Ser	-CGA-1-1	×		
• intron 1. Sac chr View i	charomyo III:2279	ces_cerevisi 942-228042 ( 3b	ae_tRNA-Ser +) Ser (CGA	-CGA-1-1 ( ) 101 bp	tRNAscan-	SE ID: chi	TIII.trna4)	Gene: Over Orga	tRNA-Ser view	-CGA-1-1 Saccharomy	ces cerevisiae		
<pre>Intron I. Sac chr View i View i</pre>	charomyc III:2279 n GtRNAc n UCSC (	Des_cerevisi 942-228042 ( db Genome Brows	ae_tRNA-Ser +) Ser (CGA er	-CGA-1-1 ( ) 101 bp	tRNAscan-	SE ID: chi	III.trna4)	Gene: Over Orga Locu	t <b>RNA-Ser</b> view nism s	-CGA-1-1 Saccharomy chrlll:227942	ces cerevisiae -228042 (+) View	in Genome Browse	r
l. Sac chr View i Ident	charomyc TII:2279 n GtRNAc n UCSC C	Des_cerevisi 042-228042 ( db Genome Brows 101/101 (10	ae_tRNA-Ser +) Ser (CGA er 0%)	-CGA-1-1 ( ) 101 bp	tRNAscan-	SE ID: chr	-III.trna4)	Gene: Over Orga Locu	t <b>TRNA-Ser</b> view nism s JAdb Gene Symb	-CGA-1-1 Saccharomy chrill:227942 ol tRNA-Ser-CO	ces cerevisiae -228042 (+) View 3A-1-1	in Genome Browse	r
I. Sac chr View i View i Ident	charomyc III:2279 n GtRNAc n UCSC C :ities =	bes_cerevisi 942-228042 ( lb senome Brows 101/101 (10	ae_tRNA-Ser +) Ser (CGA er 0%)	-CGA-1-1 ( ) 101 bp	tRNAscan-	SE ID: chi	-III.trna4)	Gene: Over Orga Locu GIRM TRIMA	t RNA-Ser view nism s VAdb Gene Symb kscan-SE ID	-CGA-1-1 Saccharomy chrill:227942 ol tRNA-Ser-Ct chrill.trna4	ces cerevisiae -228042 (+) View 3A-1-1	in <u>Ge</u> nome Browse	r
<pre>I. Sac chr View i View i Ident Query: Sbjct:</pre>	ccharomyc TII:2279 .n GtRNAc .n UCSC C .ities = .1 ggg .1 ggg	bes_cerevisi 442-228042 ( ib Benome Brows 101/101 (10 cactatggccga ulliiiiiiiii	ae_tRNA-Ser +) Ser (CGA er 0%) gtggttaaggc 	-CGA-1-1 ( ) 101 bp gagagactcg 	tRNAscan-	SE ID: chr aaaagttcgc 	GUIL.trna4)	Gene: Over Orga Locu GRR HRNA GRR Logg	view niam 3 IAdb Gene Symb Ascan-SE ID IAdb 2009 cy Name and Sco	-CGA-1-1 Saccharomy chrill:227942 ol tRNA-Ser-CG chrill.tma4 chrill.tma4-S	ces cerevisiae -228042 (+) View 3A-1-1 erCGA (75.88 bits	in <u>Ge</u> nome Browse	ŕ
<pre>I. Sac chr View i View i Ident Query: Sbjct:</pre>	charomyc TII:2279 In GtRNAc In UCSC ( Lities = 1 ggc 1 ggc	bes_cerevisi 942-228042 ( lb eenome Brows 101/101 (10 pactatggccga pactatggccga	ae_tRNA-Ser +) Ser (CGA er 0%) gtggttaaggc. gtggttaaggc.	-CGA-1-1 ( ) 101 bp gagagactog u          gagagactog	tRNAscan- aatggaata           aatggaata	SE ID: chr aaaagttcgg           aaaagttcgg	getatet 60	Gene: Over Orga Locu Grifin HRM/ Grifin Lega Antic	tRNA-Ser view nism s Addb Gene Symb uscan-SE ID JAdb 2009 ccy Name and Sco citedt fRNA isotyp codon	CGA-1-1 Saccharomy chrill:227942 ol RNA-Ser-CC chrill.tma4- ree chrill.tma4-S e/ Ser CGA	ces cerevisiae -228042 (+) View 3A-1-1 erCGA (75.88 bits	in <u>Ge</u> nome Browse	r
<pre>I. Sac chr View i View i Ident Query: Sbjct: Query: Sbjct:</pre>	ccharomy III:2275 In GtRNAG IN UCSC C C Cities = 1 ggc 1 ggc 1 ggc 1 ggc 1 ggc 1 ggc 1 ggc	es_cerevisi 442-228042 ( benome Brows 101/101 (10 cactatggccga ullililili) cactatggccga cggctctgccc ullililii	ae_tRNA-Ser +) Ser (CGA er 0%) gtggttaaggc gtggttaaggc gcgctggttca. 	-CGA-1-1 ( ) 101 bp           gagagactcg gagagactcg aatcctgctg 	aatggaata          aatggaata gtqtcg 10 	SE ID: chr aaaagttogo            aaaagttogo 1 1	GUIL.trna4)	Gene: Over Orga Locu GRN HRNA GRN Loga Prad Antic Top i Scor	tRNA-Ser view nism 3 Adb Gene Symb secan-SE ID Adb 2009 cy Name and Soc icted tRNA lsotyp: odon	-CGA-1-1 Saccharomy chrill:227942 ol tRNA-Ser-Cd chrill.tma4- chrill.tma4-S e/ Ser CGA Best Ser (117.9 b)	ces cerevisiae -228042 (+) View 3A-1-1 erCGA (75.88 bits ts) / Leu (78.5 bits	in <u>G</u> enome Browse	r
<pre>Item I. Sac chr View i View i Ident Query: Sbjct: Query: Sbjct:</pre>	charomyc III:2275 In GtRNAc In UCSC C Lities = 1 ggc 1	bes cerevisi 942-228042 ( db Benome Brows 101/101 (10 nactatggccga 11111111111 actatggccga cgggctctgccc cgggctctgccc	ae_tRNA-Ser +) Ser (CGA er 0%) gtggttaaggc JIIIIIIIIII gtggttaaggc gcgctggttca. JIIIIIIIIII gcgctggttca.	-CGA-1-1 ( ) 101 bp           gagagactcg gagagactcg aatcctgctg aatcctgctg aatcctgctg	tRNAscan- lillilli aatggaata lillilli aatggaata gtgtcg 10 lillill gtgtcg 10	SE ID: chr aaaagttcgg            aaaagttcgg 1 1	gotatot 60	Gene: Over Orga Locu GRR 1RNZ GRR Loga Pred Antici Top 1 Soco Pred Top 1	tRNA-Ser view Addb Gene Symb ecan-SE ID Addb Gene Symb ecan-SE ID Addb 2009 cy Name and Sco cy Name and Sco ted NRAM Second Scoring / Second ing Isotype Model	CGA-1-1 Saccharomy chrill:27942 ol tRNA-Ser-CG chrill.tma4 chrill.tma4-se e/ ser CGA Best Ser (117.9 b) And Consistent	ces cerevisiae -228042 (+) View JA-1-1 erCGA (75.88 bits hs) / Leu (78.5 bits	in Genome Browse	r.
<pre>l. Sac chr View i View i Ident Query: Sbjct: Sbjct:</pre>	ccharomyc III:2279 n GtRNAC n UCSC C ittles = 1 ggc 1 ggc 1 ggc 61 ctt     61 ctt	bes_cerevisi 942-228042 ( benome Brows 101/101 (10 pactatggccga cgggctctgccc cgggctctgccc	ae_tRNA-Ser +) Ser (CGA er 0%) gtggttaaggcc gtggttaaggcc gcgctggttcaa gcgctggttcaa	-CGA-1-1 ( ) 101 bp gagagactcg           gagagactcg aatoctgotg aatoctgotg	tRNAscan- lillilli aatggaata gtgtcg 10 lilli gtgtcg 10	SE ID: chr aaaagttogg           aaaagttogg 1 1	gotatot 60	Gene: Over Orga Locu GRIN HRN/A GRIN/A GRIN HRN/A GRIN/A GRIN HRN/A GRIN HRN/A GRIN/A GRIN/A GRIN/A GRIN/A GRIN/A GRIN/A GRIN/A GRIN/A GRIN/A GRIN/A GRIN/A GRIN/A GRIN/A GRIN/A GRIN/A GRN	t RNA-Ser view niam s Add Gene Symb acan-SE ID Add 2009 Vante and Soc Scoring / Second gi Sotype Mode icted Anticodon a sotype Model icted Anticodon a sotype Model	CGA-1-1 Saccharomy chrill:227942 ol RNA-Ser-Ci chrill:tma4 chrill:tma4-S er cGA Best Ser CGA Best Ser (117.9 bi nd Consistent der 1 out of 1	Les cerevisiae -228042 (+) View -3A-1-1 erCGA (75.88 bits ts) / Leu (78.5 bits	in Genome Browse	r
<pre>1. Sac chr View i View i Ident Query: Sbjct: Query: Sbjct:</pre>	charomy III:2275 In GtRNAG UCSC C C C C C C C C C C C C C C C C C C	bes_cerevisi 042-228042 ( benome Brows 101/101 (10 cactatggccga ullililili gggctctgccc ullililili gggctctgccc	ae_tRNA-Ser +) Ser (CGA er 0%) gtggttaaggc gcgctggttca jijijjjtaaggc gcgctggttca jijijjjtaaggc ycstggttca	-CGA-1-1 () ) 101 bp	tRNAscan- aatggaata         aatggaata gtqtcg 10       gtqtcg 10 sr_Apr201:	SE ID: chr aaaagttcg           aaaagttcgg 1 1 1 1 <b>1/sacCer3)</b>	cIII.trna4) getatet 60        getatet 60 Sctatet 60	Gene: Over Orga Locu GRR HRNA HRNA GRR HRNA GRR HRNA HRNA HRNA HRNA HRNA HRNA HRNA HR	tRNA-Ser view nism s Adb Gene Symb scan-SE ID Adb 2009 cy Name and Scc Idda TRNA isotype dodon Scoring / Sacond ng isotype Model acted Athle Mode Scoring / Sacond ng isotype Model acted Athle Model Corport Model acted Athle Athle Scoring / Sacond ng isotype Model acted Athle Model Scoring / Sacond ng isotype Model acted Athle Model Athle Athle Scoring / Sacond ng isotype Model acted Athle Athl	-CGA-1-1 Saccharomy chrill:227942 ol tRNA-Ser-Cd chrill.tma4- ser CGA Ser CGA Best Ser (117.9 bl nd Consistent der 1 out of 1	ces cerevisiae -228042 (+) View 3A-1-1 erCGA (75.88 bits ts) / Leu (78.5 bits CAAATCGAAAT /	in Genome Browsee	T AATAAC
<pre>I. Sac chr View i View i Ident Query: Sbjct: Query: Sbjct:</pre>	charomyc III:2275 In GtRNAc In UCSC C C C C C C C C C C C C C C C C C C	bes_cerevisi 942-228042 ( bb Benome Brows 101/101 (10 pactatggccga cgggctctgccc cgggctctgccc pactatggccga cgggctctgccc cgggctctgcc cgggctctg	ae_tRNA-Ser +) Ser (CGA er 0%) gtggttaaggc gtggttaaggc gcgctggttca.             gcgctggttca. 	-CGA-1-1 () ) 101 bp gagagactcg            gagagactcg aatcctgctg aatcctgctg aatcctgctg 011 (SacCe	tRNAscan- aatggaata         aatggaat gtqtog 10       gtgtog 10 sr_Apr201 noul 15s 3s	SE ID: chr aaaagttcgg            1 1 1 1/sacCer3)	cIII.trna4) yetatot 60        gotatot 60 Assembly	Gene: Over Orga Locu GRR HRM2 GRR HRM2 GRR HRM2 GRR Pred Antic Top J Scor Pred Pred Rank Upst Seq. Intro	tRNA-Ser view nism a Addb Gene Symb scan-SE ID Addb Z009 cy Name and Scc tiedu FNA Isotype dodo Scoring / Sacond Scoring / Scoring / Scoring / Sacond Scoring /	-CGA-1-1 Saccharomy chrill:227942 ol tRNA-Ser-CG chrill.tma4- see chrill.tma4-s e/ ser CGA Best Ser (117.9 bi nd Consistent der 1 out of 1 tm CCATAATTT Ser-Se (2278)	ces cerevisiae -228042 (+) View 3A-1-1 erCGA (75.88 bits ts) / Leu (76.5 bits CAAATCGAAAT / '9-227997)	in <u>G</u> enome Browse	AATAAC
<pre>* nten  1. Sac chr View i View i Ident Query: Sbjct: Query: Sbjct: SC Genome </pre>	charomyc III:2275 In GtRNAc III:2275 Charles = Charles =	bes_cerevisi 942-228042 ( bb Benome Brows 101/101 (10 pactatggccga cgggctctgccga cgggctctgccc er on S. cerev (( ) ))) belowed base ( ))) ( )))) ( ))) ( )))) ( ))))))))))	ae_tRNA-Ser +) Ser (CGA er 0%) gtggttaaggc gtggttaaggc ggggtggttaaggc ggggtggttaa gcgctggttca. juiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii	-CGA-1-1 () ) 101 bp gagagactog gagagactog aatcotgotg aatcotgotg aatcotgotg 011 (SacCe 10x base 2007 htems	tRNAscan- aatggaata         aatggaat gtqtog 10       gtqtog 10       gtqtog 10	SE ID: chr aaaagttcgg            1 1 1 1 1 1 1 1 1 1 1 1 1	ciII.trna4) (ctatot 60 ())))) (ctatot 60 ()) ()) ()) ()) ()) ()) ()) ()) ()) ()	Gene: Over Orga Locu GUR HRM2 GUR HRM2 GUR Pred Antic Top J Scor Pred Pred Antic Seq. Intro Pess Seq.	tRNA-Ser view nism a Addb Gene Symb scan-SE ID Addb Z009 cy Name and Scottient Control (Second Control) Scotring / Second Ing Isotype Model ted Anticodon a Isotype An	-CGA-1-1 Saccharomy chrill:227942 ol tRNA-Ser-CG chrill.tma4 er chrill.tma4-s e / Ser CGA Best Ser (117.9 bi nd Consistent der 1 out of 1 im CCATAATTT 38-56 (2278) No	ces cerevisiae -228042 (+) View 3A-1-1 erCGA (75.88 bits ts) / Leu (78.5 bits CAAATCGAAAT / '9-227997)	in Genome Browse	AATAAC
<pre>* ##en 1. Sac chr View i View i Ident Query: Sbjct: Query: Sbjct: SCGenome ####################################</pre>	charomyc III:2275 In GtRNAc III:2275 Charles = Charles =	es cerevisi 942-228042 ( b Benome Brows 101/101 (10 pactatggccga actatggccga agggctctgccc er on S. cerev 881-228.092 202 b 881-228.092 202 b 19 000000000000000000000000000000000000	ae_tRNA-Ser +) Ser (CGA er 0%) gtggttaaggc gtggttaaggc ggggtggttaaggc ggggtggttaaggc gcgctggttca. juiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii	-CGA-1-1 () ) 101 bp gagagactog gagagactog aatcotgotg aatcotgotg aatcotgotg aatcotgotg aatcotgotg aatcotgotg	tRNAscan- aatggaata         aatggaata gtgtog 10       gtgtog 10 	SE ID: chr aaaagttcgg            1 1 1 1 1 1 1 1 1 1 1 1 1	ctatot 60        cctatot 60 Sesembly	Gene: Over Orga Gurn IRM2 Gurn Locu Gurn IRM2 Gurn Pred Antic Top J Scor Pred Pred Antic Sequ Intro Poss Gene:	tRNA-Ser view nism standard Gene Symb scan-SE ID Addb Gene Symb scan-SE ID Addb 2009 cy Name and Scot tied IRNA Isotype odon Isotype Model tied Anticodon a sotype Model tied Anticodon a	-CGA-1-1 Saccharomy chrill:227942 ol tRNA-Ser-CG chrill.tma4- ser chrill.tma4- ser CGA Best Ser (117.9 bi nd Consistent der 1 out of 1 im CCATAATTT 38-56 (2278) No 7.1.8	Ces cerevisiae -228042 (+) View 3A-1-1 erCGA (75.88 bits ts) / Leu (78.5 bits CAAATCGAAAT / '9-227997)	in Genome Browse	AATAAC
<pre>* nten  1. Sac chr View i View i Ident Query: Sbjct: Query: Sbjct: SCGenome </pre>	charomyc III:2275 In GtRNAC III:2275 Charles = Charles =	es cerevisi p42-228042 ( b Benome Brows 101/101 (10 pactatggccga cgggctctgccga cgggctctgccc er on S. cerev er on S. cerevision 101/202 (202 b) 101/202 (202 b) 101/2	ae_tRNA-Ser +) Ser (CGA er 0%) gtggttaaggc gtggttaaggc gcgctggttca. juiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii	-CGA-1-1 ( ) 101 bp gagagactog           gagagactog aatcotgotg aatcotgotg aatcotgotg 011 (SacCe 10x base 2007 htems HI/sacers) ovii t	tRNAscan-	SE ID: chr aaaagttcgg aaaagttcgg 1 1 1 1 1/sacCer3)	gotatot 60        octatot 60 Assembly	Gene: Over Orga Locu GRR 1RN2 GRR 1RN2 GRR 1RN2 GRR 1RN2 GRR 1RN2 GRR 1RN2 GRR 1RN2 GRR 1 G 1 G	tRNA-Ser view inism s 4Adb Gene Symb scan-SE ID 4Adb 2009 cy Name and Sco ted 1RNA Isotyp adon (Scoring / Second isotype Model ted 1RNA Isotyp adon (Scoring / Second a) bible Pseudogene s Score re IRNA Score (Score	-CGA-1-1 Saccharomy chrill:227942 ol tRNA-Ser-CG chrill.tma4- chrill.tma4- e / Ser CGA Best Ser (117.9 bi nd Consistent der 1 out of 1 Im CCATAATTT S8-56 (2279) No 71.8 65.1	ces cerevisiae -228042 (+) View JA-1-1 erCGA (75.88 bits ts) / Leu (78.5 bits CAAATCGAAAT / '9-227997)	in Genome Browse	AATAAC
<pre>* men 1. Sac chr View i View i Ident Query: Sbjct: Query: Sbjct: SCGenomu Humov Festition Scale uere: Listen Listen</pre>	charomyc III:2275 In GtRNAc III:2275 itties = 1 ggc 1 ggc 61 ctt 61 ctt e Browsc chrili:227,	es cerevisi p42-228042 ( b Benome Brows 101/101 (10 pactatggccga actatggccga cgggctctgccc er on S. cerev er on S. cerevisi seggctctgccc construction construct	ae_tRNA-Ser +) Ser (CGA er 0%) gtggttaaggc juiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii	-CGA-1-1 ( ) 101 bp ( ) gagagactcg gagagact gagagactcg gagagactcg gagagactcg gagagactcg gagagactcg gagagactcg gagagact gagagactcg gagac gagac gagagactcg gagagactcg gagac gagac gagagactcg gagac gaga gagac ga ga gagac ga ga ga ga ga ga ga ga ga ga ga ga ga	tRNAscan-	SE ID: chr aaaagttogg 1 1 1/sacCer3)	gotatot 60         Gotatot 60 Assembly	Gene: Over Orga Locu Grinh Loga Pred Antic Top : Sor Pred Top : Rank Upst Sor Pres Gene:	tRNA-Ser view Addb Gene Symb acan-SE ID Addb Cene Symb acan-SE ID Addb 2009 Scoring / Second ing Isotype Model total RNA Isotyp odon Scoring / Second ing Isotype Model of IRNA Isodeco ream / Jownstrea ence n Score re re RNA Score i Score	CGGA-1-1 Saccharomy chrill:227942 Saccharomy chrill:227942 Saccharomy chrill:tma4-S chrill:tma4-S e f Sac CGA Gas	CAAATCGAAAT / 19-227997)	in Genome Browse	AATAAC
I. Sac chr View i View i Ident Query: Sbjct: Sbjct: Sbjct: Sbjct:	charomyc III:2275 In GtRNAc III:2275 itites = 1 ggc 1 ggc 61 ctt 61 ctt e Browsc chrili:227,	er on S. cerevisi aggetetgece er on S. cerevisi senter of S. cerevisi aggetetgece er on S. cerevisi setter of S. cerevisi aggetetgece setter of S. cerevisi setter of S. cerevisi setter of S. cerevisi setter of S. cerevisi setter of S. cerevision setter of S. cerevision	ae_tRNA-Ser +) Ser (CGA er 0%) gtggttaaggc gtggttaaggc ggggtggttaaggc ggggtggttaa gcgctggttca visiae Apr. 22 zoom in 16 av for extended on sear for extende	-CGA-1-1 ( ) 101 bp ( ) gagagactog gagagactog gagagactog gagagactog gadagacto	tRNAscan-	SE ID: chr	rIII.trna4) gotatot 60 UIUII octatot 60 Assembly	Gene: Over Orga Locu Grin tRNA Grin Top : Scor Pred Antici Top : Scor Pred Top : Ranh Uppt Sequ Intro Press Gene:	tRNA-Ser view niam Addb Gene Symb ascan-SE ID Addb 2009 Scoring / Second ing isotype Model cited RNA isocyp odon Scoring / Second ing isotype Model of IRNA isodeco ream / Downstreat ence isocype Model of tRNA isodeco ream / Downstreat ence isocype Model isocre isocype Model of tRNA isodeco ream / Downstreat ence isocype Model isocre isotype Model isocre isotype Model isocre isotype Model isocre isotype Model isotype Model isoty	CGA-1-1           Saccharomy           chrill:227942           ol         KNA-Ser-CG           chrill:tma4-2           chrill:tma4-2           e/         Ser CGA           e/         Ser CGA           Best         Ser (117.9 bill)           nd         Consistent           der         1 out of 1           um         CATAATTT           S8-56 (22797           No         Ser-GGA           der         1.0.1 of 1           um         CATAATTT           S8-56 (22797         No           1000000000000000000000000000000000000	Des cerevisiae -228042 (+) View 3A-1-1 erCGA (75.88 bits ts) / Leu (78.5 bits CAAATCGAAAT / '9-227997)	in <u>Ge</u> nome Browse	AATAAC

Figure 1. Example tRNAscan-SE Search and Contextual Analysis. The *Saccharomyces cerevisiae* tRNA-Ser<sup>CGA</sup> is analyzed using the tRNAscan-SE web server in default eukaryotic search mode. The red arrows show the analysis path from viewing the predicted tRNA results to finding the matching tRNA gene in GtRNAdb (3), to exploring the tRNA gene in context with tRNA modifications and gene expression data in the UCSC Genome Browser (10).

#### **FUTURE DEVELOPMENT**

As the technology for sequencing and assembling genomes continues to improve, we anticipate that the demand to identify and annotate tRNA genes in new, complete genomes will continue to accelerate. Accordingly, we plan to produce a tRNA gene set 'completeness' report using phylogenetic patterns of tRNA set composition observed across all genomes represented in the GtRNAdb. Noting potential 'missing' or 'surplus' tRNA gene decoding potential will be useful to assess genome quality and completeness, as well as recognizing potential genome assembly errors or sequencing contamination. A second planned capability is metagenomic analysis of sequencing data containing an unknown mix of species. By doing comparative analysis using a suite of isotype-specific and phylum-specific covariance models (under development now), we hope to offer a tRNA identification and phylogenetic classification service as part of the tRNAscan-SE web server. Finally, with the increase in knowledge of functional tRNA fragments, we plan to offer detection and systematic classification of fragments in the context of full-length tRNA genes.

#### ACKNOWLEDGEMENT

We would like to thank Lowe Lab members Brian Lin, Allysia Mak and Aaron Cozen for their work in development of the new covariance models for tRNAscan-SE 2.0, as well as their assistance in extensive testing and feedback on the web server interface.

#### FUNDING

National Human Genome Research Institute, National Institutes of Health [HG006753-02 to T.L.]. Funding for open access charge: NHGRI/NIH [HG006753-02]; University of California, Santa Cruz department chair research stipend. *Conflict of interest statement*. None declared.

#### REFERENCES

- 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.
- Chan, P.P. and Lowe, T.M. (2009) GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.*, 37, D93–D97.
- 3. Chan, P.P. and Lowe, T.M. (2016) GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.*, **44**, D184–D189.
- Schattner, P., Brooks, A.N. and Lowe, T.M. (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.*, 33, W686–W689.
- Nawrocki, E.P. and Eddy, S.R. (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, 29, 2933–2935.
- Giege, R., Sissler, M. and Florentz, C. (1998) Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.*, 26, 5017–5035.
- Perry, J., Dai, X. and Zhao, Y. (2005) A mutation in the anticodon of a single tRNAala is sufficient to confer auxin resistance in Arabidopsis. *Plant Physiol.*, **139**, 1284–1290.
- 8. Kimata, Y. and Yanagida, M. (2004) Suppression of a mitotic mutant by tRNA-Ala anticodon mutations that produce a dominant defect in late mitosis. J. Cell Sci., **117**, 2283–2293.
- 9. Ardell, D.H. and Andersson, S.G. (2006) TFAM detects co-evolution of tRNA identity rules with lateral transfer of histidyl-tRNA synthetase. *Nucleic Acids Res.*, **34**, 893–904.
- Speir, M.L., Zweig, A.S., Rosenbloom, K.R., Raney, B.J., Paten, B., Nejad, P., Lee, B.T., Learned, K., Karolchik, D., Hinrichs, A.S. *et al.* (2016) The UCSC Genome Browser database: 2016 update. *Nucleic Acids Res.*, 44, D717–D725.
- Chan, P.P., Holmes, A.D., Smith, A.M., Tran, D. and Lowe, T.M. (2012) The UCSC Archaeal Genome Browser: 2012 update. *Nucleic Acids Res.*, 40, D646–D652.