

# Diversity of tRNA genes in eukaryotes

Jeffrey M. Goodenbour and Tao Pan<sup>1,\*</sup>Department of Human Genetics and <sup>1</sup>Department of Biochemistry and Molecular Biology, 929 East 57th street, Chicago, IL 60637, USA

Received August 1, 2006; Revised September 13, 2006; Accepted September 19, 2006

## ABSTRACT

We compare the diversity of chromosomal-encoded transfer RNA (tRNA) genes from 11 eukaryotes as identified by tRNAScan-SE of their respective genomes. They include the budding and fission yeast, worm, fruit fly, fugu, chicken, dog, rat, mouse, chimp and human. The number of tRNA genes are between 170 and 570 and the number of tRNA isoacceptors range from 41 to 55. Unexpectedly, the number of tRNA genes having the same anticodon but different sequences elsewhere in the tRNA body (defined here as tRNA isodecoder genes) varies significantly (10–246). tRNA isodecoder genes allow up to 274 different tRNA species to be produced from 446 genes in humans, but only up to 51 from 275 genes in the budding yeast. The fraction of tRNA isodecoder genes among all tRNA genes increases across the phylogenetic spectrum. A large number of sequence differences in human tRNA isodecoder genes occurs in the internal promoter regions for RNA polymerase III. We also describe a systematic, ligation-based method to detect and quantify tRNA isodecoder molecules in human samples, and show differential expression of three tRNA isodecoders in six human tissues. The large number of tRNA isodecoder genes in eukaryotes suggests that tRNA function may be more diverse than previously appreciated.

## INTRODUCTION

Transfer RNA (tRNA) consists of 75–95 nt and is ubiquitous in all organisms. All tRNAs are characterized by a secondary structure made up of three hairpin loops and a terminal helical stem (cloverleaf) which fold into an L-shaped tertiary structure. The main functional regions in tRNA are the anticodon triplets which read the messenger RNA (mRNA) codons and the 3' CCA nucleotides where an amino acid cognate to the tRNA is attached. Codon degeneracy for the 20 amino acids requires up to 5 tRNAs with distinct

anticodons (tRNA isoacceptors) to read codons for each amino acid. There are 21 isoacceptor families, one for each amino acid and one for seleno-cysteine. An isoacceptor family may consist of one tRNA member, e.g. tRNA<sup>Trp</sup>, to five tRNA members, e.g. tRNA<sup>Leu</sup> (1).

tRNA isoacceptors have been the main targets for biological, biochemical and computational studies of tRNA function. In bacteria and yeast, the abundance of tRNA isoacceptors correlates with the codon usage of highly abundant proteins (2–8). The abundance of tRNA isoacceptors in bacteria and yeast can be approximated by the number of tRNA genes in the genomes. Annotation of tRNA genes in sequenced genomes by tRNAScan-SE shows that the number of tRNA isoacceptors ranges from high 20s to low 50s (9,10).

Even before genome sequencing, it was already known by direct RNA sequencing that in *Escherichia coli* K12, the number of different tRNA species is greater than the total number of isoacceptors (11). tRNA<sup>ThrI</sup> and tRNA<sup>Thr3</sup> have the same anticodon GGU, but they differ at four residues in the TΨC stem and at nucleotide #9. tRNA<sup>Val2A</sup> and tRNA<sup>Val2B</sup> have the same anticodon GAC, but they differ at four residues in the acceptor stem and two residues in the TΨC stem. tRNA<sup>TyrI</sup> and tRNA<sup>TyrII</sup> have the same anticodon GUA, but they differ at two residues in the long variable stem-loop. All together, the total number of tRNA species in *E.coli* K12 on the basis of RNA sequencing was 44 among 41 tRNA isoacceptors. The genome sequence of *E.coli* K12 (MG1655 strain) shows three additional tRNA genes containing sequence differences in the same isoacceptor: one tRNA<sup>Leu</sup>(CAG) gene, which differs at one residue in the long variable stem from three other tRNA<sup>Leu</sup>(CAG) genes; one tRNA<sup>fmet</sup> gene, which differs at one residue in the short variable loop from three other tRNA<sup>fmet</sup> genes; and tRNA<sup>Ile</sup>(CAU), which differs at two residues in the acceptor stem (12). In all these cases, the sequence changes perfectly maintain base pairing and the conserved structural features of tRNA. All together in *E.coli* MG1655, the possible number of tRNA species that have the same anticodon but different sequences in the tRNA body is 6 among 86 annotated tRNA genes (6/86 = 7% of tRNA genes).

A well-studied example in tissue-specific expression of distinct tRNAs having the same anticodon is derived from the silkworm, *Bombyx mori* (13–15). Two tRNA<sup>Ala</sup>(AGC) products have been identified, one present in all tissue

\*To whom correspondence should be addressed. Tel: +1 773 702 4179; Fax: +1 773 702 0439; Email: taopan@uchicago.edu

types and the other only in the silk gland. These two tRNAs differ by a single nucleotide at position 40 which changes the G30–C40 base pair in the anticodon stem of the constitutive tRNA<sup>Ala</sup> to G30–U40 wobble for the silk gland specific tRNA<sup>Ala</sup>.

Previous tRNA sequencing work shows that at least two distinct tRNAs with the same anticodon is expressed in humans (16). The third nucleotide in tRNA<sup>Gln</sup>(UUG) was found to be either a C or a U.

This work describes a surprisingly large diversity of tRNA genes among 11 species of eukaryotes from budding yeast to human. Their genomes contain 170–570 tRNA genes and 41–55 tRNA isoacceptors. However, the number of tRNA genes having the same anticodon but different sequences in the tRNA body is as low as 3.6% (10/275) and as high as 55% (246/451). Furthermore, the number and percentage of such tRNA genes follow the phylogenetic arrangement of these 11 organisms.

Our previous experience with tRNA microarrays (17,18) suggests that the sensitivity and reproducibility in the detection and quantification of tRNA can be significantly improved by using hybridization probes that are complementary to the entire tRNA. Of course, the application of full-length hybridization probes does not allow the discrimination of single nucleotide differences in tRNA. Here, we describe a systematic method that utilizes the advantage of full-length hybridization but still allows discrimination of single nucleotide differences in tRNA.

## MATERIALS AND METHODS

### tRNA sequence analysis

tRNA sequences were obtained in the FASTA format from <http://lowelab.ucsc.edu/GtRNAdb/>. tRNA sequences at this site have been annotated according to anticodon, genome location and quality score by tRNAScan-SE (9,10), which mines completed genome sequencing data for tRNAs. Complete genome tRNA sets were processed by alignment in CLUSTAL X (19) to distinguish repeated sequences. For the six species analysis of tRNA<sup>Set</sup>(AGA), phylogenetic distances were determined by CLUSTAL X following alignment and drawn using NJplot (20,21).

Following raw data collection, human tRNA sequences were manually curated to determine the extent of variation in 11 discrete tRNA regions. Primary sequence data were analyzed for correct tRNA structure dependent upon base pairing of acceptor, D, anticodon and TΨC stems. Variation in stem regions was coded to indicate whether or not acceptable base pairing (Watson–Crick plus G–U) is maintained. Among the human tRNA sequence set, just three sequences with greater than two pairing errors were found and excluded from our analysis.

### Ligation procedure using 30mer model RNA

A model 30mer RNA oligonucleotide for testing the ligation method was designed to contain minimal secondary structure when the nucleotide at the 15th position is U (Table 1). Four chemically synthesized 30mer RNAs (Dharmacon Research, Lafayette, CO) have the same sequence at all positions except

**Table 1.** Model RNA, floaters and anchors used in this work

	Anchor (5'→3')	Floater (5'→3')	Z <sup>a</sup>
Set I: 5'Z	TACCAGGC ATTGCTC	pZGCGTTA CAGCGGAT <sup>b</sup>	dA,dC,dG,dT,rA, rC,rG,rU,Am,Cm,Gm,Um
Set II: 3'Z	pGCGTTAC AGCGGAT	TACCAGGC ATTGCTCZ	dA,dC,dG,dT,rA, rC,rG,rU,Am,Cm,Gm,Um
Set III: Z+3	pGTTACAG CGGAT	TACCAGGC ATTGCTCZGC	dA,dC,dG,dT

Model RNA sequence (X = A, C, G, U): 5'-AUCCGCGUAACGCXGAG-CAAUGCCUGGUA

<sup>a</sup>d = deoxy; r = ribo; m = 2'-O-methyl.

<sup>b</sup>Z = nucleotide with variable base or backbone composition.

the 15th nucleotide. All 28 floater and 3 anchor oligonucleotides were synthesized by Integrated DNA Technologies (IDT, Coralville, IA).

The ligation reaction was performed at 30°C for up to 1 h in 0.15 μM 30mer RNA, 0.5 μM floater oligonucleotide, 0.38 μM 5'-<sup>32</sup>P-labeled anchor oligonucleotide in 66 mM Tris–HCl, pH 7.6, 6.6 mM MgCl<sub>2</sub>, 10 mM dithiothreitol (DTT), 66 μM ATP, 15% DMSO and 0.125–0.25 U/μl T4 DNA ligase (USB, Cleveland, OH). The ligation product was separated on 15% polyacrylamide gel containing 7 M urea and quantified by phosphorimaging (Fuji Medicals).

### Ligation procedure using human tissue samples

Total tRNA was isolated from total RNA samples on 8% polyacrylamide gels containing 7 M urea. The tRNA bands were visualized by UV shadowing, excised from the gel and soaked for 1.5 h at room temperature in 50 mM KOAc/200 mM KCl, pH 7. Gel slices were then removed by filtration and the tRNA precipitated by ethanol, vacuum dried and dissolved in water at 10–15 ng/μl. Total RNA samples include HeLa (isolated using RNawiz kit from Ambion Inc., Austin, TX) and 6 human tissues purchased from Stratagene (La Jolla, CA): liver (#735017), brain (#735006), uterus (#735042), ovary (#735260), vulva (#735067) and testes (#735064).

The ligation reaction using the purified tRNA mixture from human samples was conducted under different conditions for hybridization and ligation. The hybridization reaction containing 100 ng total tRNA, 0.01 μM yeast tRNA<sup>Phe</sup> standard, 0.13 μM 5'-<sup>32</sup>P-labeled anchor oligonucleotide, 0.13 μM floater, 20 mM Tris–HCl, pH 7.6 and 50 mM NaCl ran for 1 h at 62°C followed by 1 h at 52°C to accommodate for differential efficiencies of the oligonucleotide substrates for ligation. Ligation was then performed at 37°C for 2 h after the addition of a 2× ligation mixture containing 132 mM Tris–HCl, pH 7.6, 13.2 mM MgCl<sub>2</sub>, 20 mM DTT, 132 μM ATP, 30% DMSO and 1 U/μl T4 DNA ligase to the hybridization mixture. Reactions were treated with 0.1 U/μl RNase H (Epicentre Technologies, Madison, WI) and 0.1 U/μl Calf Intestine Alkaline Phosphatase (Roche Applied Sciences, Indianapolis, IN) for 10 min at 37°C to reduce background. The ligation reaction mixture was stopped by the addition of an equal volume of 9 M urea/50 mM EDTA, boiled for 2 min and then loaded on 12% polyacrylamide gels containing 7 M urea.

## RESULTS

## Diversity of tRNA isodecoder genes

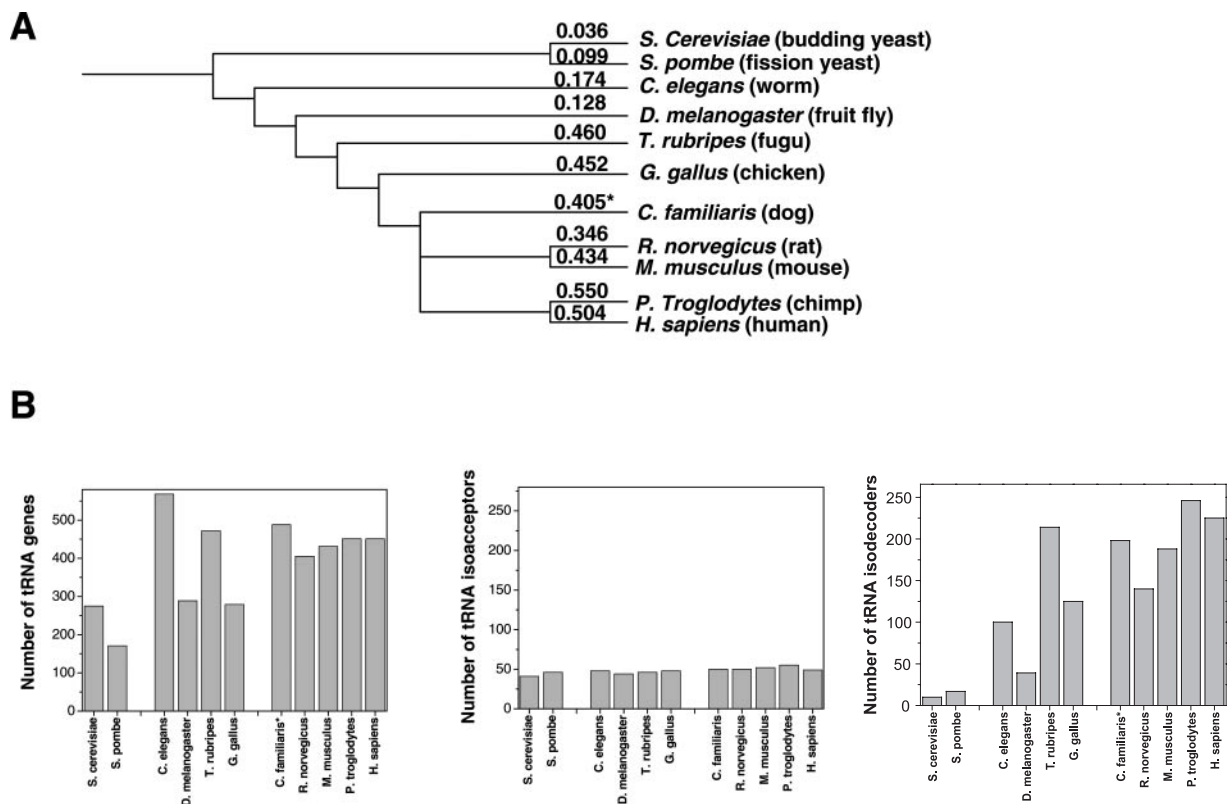
tRNAScan-SE is among the most successful programs to identify non-coding RNAs in genome sequences (9,10). tRNAScan-SE first utilizes tRNAscan1.4 and EufindtRNA to search for conserved tRNA sequences at defined positions, then evaluates co-variance in conserved tRNA sequence and secondary structure. The algorithm is capable of detecting 99–100% of tRNAs with a very low error rate (one false positive per 15 GB).

Eukaryotic genomes generally contain several hundred tRNA genes as predicted by tRNAScan-SE (<http://lowelab.ucsc.edu/GtRNAb/>). tRNAScan-SE is also capable of distinguishing tRNA pseudogenes, which range from ~170 in the human genome to ~22 000 in the mouse genome. There are, however, significant outliers. *Danio rerio* (zebra fish) has ~6000 predicted tRNA genes. *Canis familiaris* (dog) contains ~400 genes for tRNA<sup>Lys</sup> with anticodon CTT, which are excluded from our analysis to avoid unnecessary bias.

We chose to focus on tRNA sequences from 11 eukaryotic genomes as they represent a wide range in the phylogenetic tree and encompass many model organisms (Figure 1 and Table S1). These 11 genomes have predicted tRNA gene counts from 171 (fission yeast) to 568 (worm). The number of tRNA isoacceptors among these 11 species range from 41 (budding yeast) to 55 (chimp).

What is remarkable and not predicted before genome sequencing, however, are the numbers of tRNA genes having the same anticodon sequence but differences elsewhere in the tRNA body (Figure 1). We tentatively use the nomenclature of ‘tRNA isodecoder gene’ to describe these tRNA sequences. tRNA isodecoders have the same anticodon sequence (hence they decode the same codon), i.e. they belong to the same isoacceptor class, but have sequence differences elsewhere in the tRNA body. One tRNA sequence within each isoacceptor class, generally the one with the highest gene copy number, is arbitrarily designated as the majority member. The number of tRNA isodecoder genes within an isoacceptor class is the count of distinct tRNA sequences within this class excluding the majority member. For example, the human tRNA<sup>Arg</sup>(ACG) isoacceptor class has two different sequences with four and three gene copies. They differ by a single nucleotide at position #50. The four gene-copy tRNA<sup>Arg</sup>(ACG) is the majority member and has the sequence of U50, and the three gene-copy tRNA<sup>Arg</sup>(ACG) is classified as the isodecoder gene and has the sequence of C50. The number of tRNA isodecoder genes is therefore one for the tRNA<sup>Arg</sup>(ACG) isoacceptor class. By this account, the total number of different tRNA gene sequences in these 11 genomes is the number of isoacceptors (i.e. from 41 to 55) plus the number of isodecoders (i.e. from 10 to 246).

The fraction of tRNA isodecoder genes (the sum of all isodecoder genes divided by the total number of tRNA

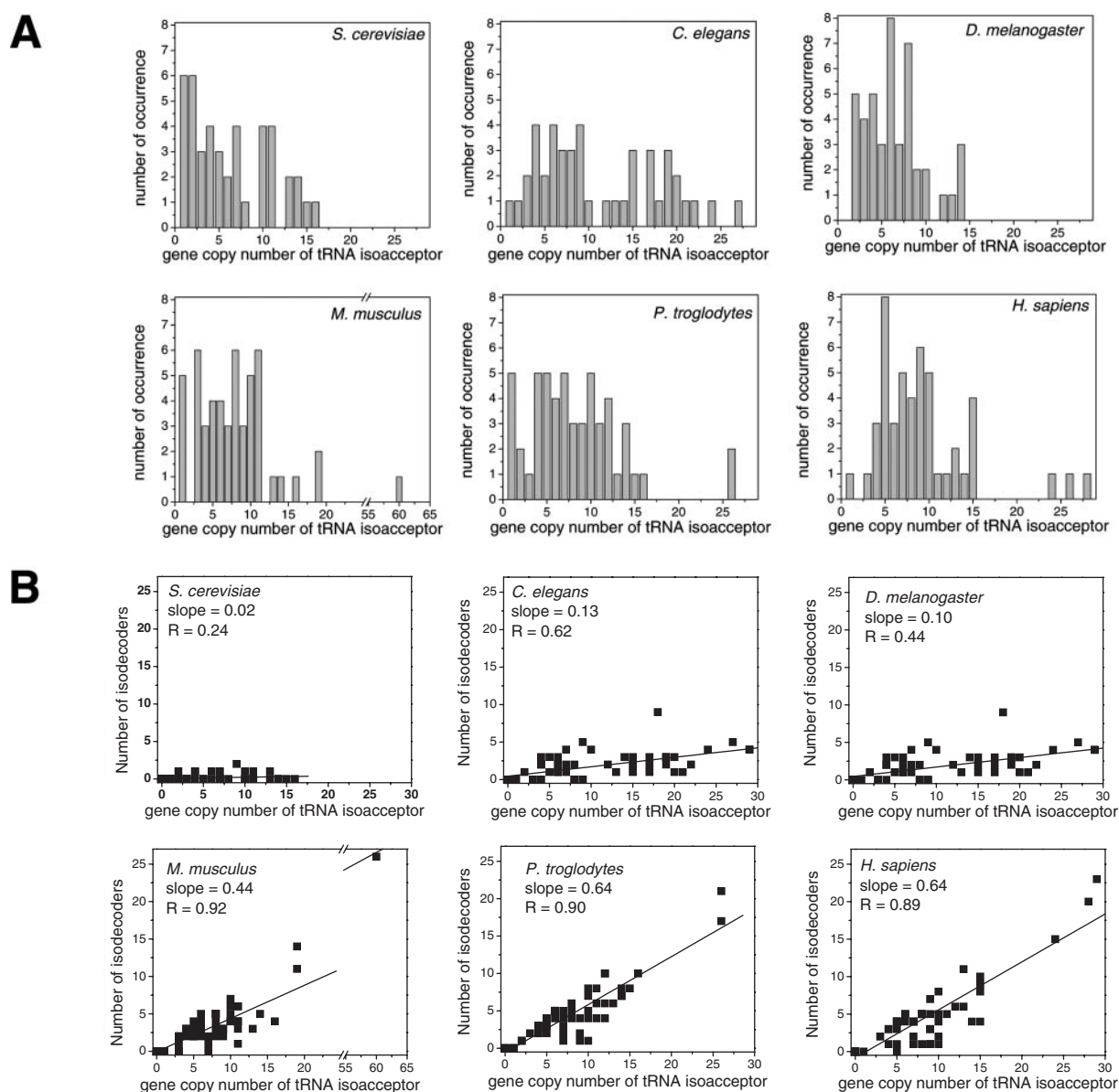


**Figure 1.** tRNA genes and isodecoder genes in 11 eukaryotes. (A) Cladogram of the organisms based on the NCBI taxonomy browser (40,41) which include two single cell yeast, worm, fruit fly, fugu, chicken and five mammals, dog, rat, mouse, chimp and human. The fraction of tRNA isodecoder genes among all tRNA genes is indicated. \*: ~400 genes in the tRNA<sup>Lys</sup>(CTT) isoacceptor class in the dog genome are not included in this count. (B) The number of tRNA genes (left), isoacceptors (middle) and isodecoders (right) in these organisms.

genes) has distinct groupings among these 11 species when plotted on a cladogram (Figure 1A). This fraction is <10% in the budding and fission yeast, 12–18% in fruit fly and worm, and increases to 35–46% in fugu, chicken, dog, rat and mouse. The fraction is highest among the two primates where >50% of tRNA genes are isodecoder genes. This phylogenetic grouping indicates that the diversity of tRNA isodecoder genes cannot be simply derived from inaccuracies in genome sequencing (a small number of them may be attributed to lower sequencing accuracy in some genomes). The fraction of tRNA isodecoder genes corresponding to the phylogenetic grouping of these organisms may suggest that they perform some kind of heretofore under-appreciated functions. It may also be a result of genome expansion.

We analyzed the sequence features of tRNA isodecoder genes further in six commonly studied species: budding yeast, worm, fruit fly, mouse, chimp and human (Figures 2 and 3). The number of tRNA isoacceptors range from 41 to 55. These isoacceptors occur between 1 and 60 times in the genome (Figure 2A). Budding yeast and fruit fly have relatively few tRNA genes (270–290) and the number of occurrences for each gene is relatively low. Worm has a high number of tRNA genes (568) and the number of occurrences is broadly distributed. A few isoacceptors in mammals have high copy numbers that distinguish them from the other isoacceptors.

The number of tRNA isodecoder genes varies from very low (10 in yeast) to very high (225–246) in chimp and



**Figure 2.** Gene copy numbers of tRNA isoacceptors versus the number of occurrence or the number of isodecoders. (A) Plot of the gene copy number of tRNA isoacceptors and the number of occurrence for each isoacceptor class. (B) Plot of the gene copy number of tRNA isoacceptors and the number of tRNA isodecoders for each isoacceptor class. A good linear correlation ( $R$ -value between 0.89 and 0.92) exists between the gene copy number of tRNA isoacceptors and the number of tRNA isodecoders in the three mammals.

human. The number of tRNA isodecoder genes in mammals has a good linear correlation ( $R$ -value of 0.89–0.92 and slope of 0.44–0.66) to the gene copy number of their corresponding isoacceptors (Figure 2B). The highest slope possible in this plot would be 1.0 when every tRNA gene is unique, after subtracting the majority member of isoacceptor classes. A slope of 0.64 shows that the bulk of human and chimp tRNA genes is unique. As for the non-mammal species, a linear correlation has significantly lower  $R$ -values (0.24–0.64) and smaller slopes (0.02–0.13). This result suggests that the evolutionary appearance of tRNA isodecoder genes in non-mammals may be less directed than in mammals.

The same description of tRNA isodecoder genes can be applied to bacterial tRNA genes in species with sequenced genomes (Supplementary Table S2 and Supplementary Figure S1). As described in the Introduction, the number of tRNA isodecoder genes in *E.coli* K12 is 6 among 86 genes ( $6/86 = 7\%$ ). Among the 139 bacterial genome sequences, the number of isodecoder genes range from 0 to 26 and the fraction of tRNA isodecoder genes range from 0 to 0.30. A great majority of species cluster in the lower regime of the tRNA gene-isodecoder gene plot (Supplementary Figure S1).

An in-depth sequence analysis of the tRNA<sup>Ser</sup>(AGA) isoacceptor class among the six eukaryotic organisms is shown in Figure 3. This isoacceptor is chosen on the basis of simplicity of comparison as well as the number of isodecoder genes in each species. This tRNA isoacceptor has 11 gene copies in yeast, 15 copies in worm, 8 copies in fruit fly and mouse, and 9 copies in chimp and human. The yeast genes have 2 sequence variants, 10 being the same plus 1 distinct isodecoder gene. Worm has 3 sequence variants, 13 being the same plus 2 isodecoders. Fruit fly, mouse and chimp have two isodecoders each and human has three isodecoders. The yeast tRNA sequences are noticeably different from all others. All tRNA genes in worm and fruit fly are clustered together among themselves. The mammalian sequences cluster more closely according to their isodecoder genes than to their species. In fact, the majority sequence (with six copies each) of these mammalian species is identical.

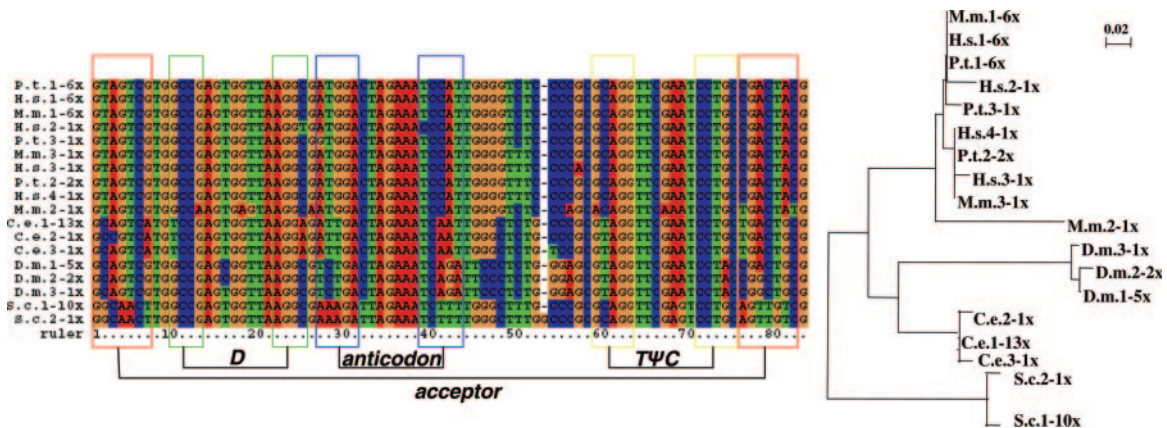
Most sequence changes in the tRNA<sup>Ser</sup>(AGA) isodecoder genes do not alter the secondary or tertiary structure of

tRNA. The fruit fly isodecoder genes involve an A–U to G–U pair change in the acceptor stem and C-to-U in the variable region of the D-loop. Sequence change in one worm isodecoder gene is an A–U to G–U in the stem of the long variable loop. The major sequence changes in the mammalian isodecoder genes involve A–U to G–U or G–U to G–C changes in various stems or a U-to-C change in an unpaired region in the variable loop. The two exceptions are a worm isodecoder gene (Ce2), which changes an A–U to C\*U mismatch in the acceptor stem, and a mouse isodecoder gene (Mm2), which changes a G–C to A\*C in the TΨC stem.

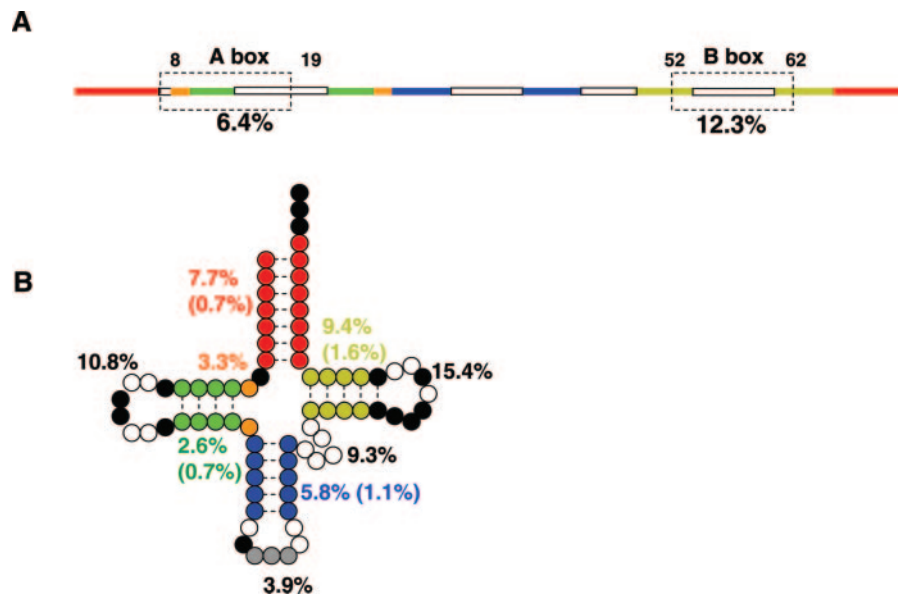
### Human tRNA isodecoder genes

We further analyzed the locations of sequence changes in human tRNA isodecoder genes in detail (Figure 4). Eukaryotic tRNA genes are transcribed by RNA polymerase III, and a portion of the Pol-III promoter is within a tRNA gene (22,23). The internal promoters constitute two discrete regions corresponding to nt 8–19 (box A) and 52–62 (box B) of a tRNA. Nucleotides 8, 14, 18 and 19 in box A, and nt 53–56, 58 and 61 in box B are highly conserved among all tRNAs because of tRNA tertiary structure. Hence, only ~7 nt within box A and 6 nt within box B are variable. Human tRNA genes vary at 6.4% of these variable nt in box A and 12.3% in box B (Figure 4A). These sequence differences may lead to differential tRNA expression in human tissues or developmental stages.

Sequence changes in tRNA isodecoder genes can also be divided into nine regions according to tRNA secondary structure (Figure 4B). The number of sequence changes is determined by comparison of isodecoders to the majority variant. Frequency of sequence changes is the number of changes in each of the nine regions divided by the total number of nucleotides surveyed in that region. The largest frequency of sequence changes is among the three non-conserved residues in the TΨC loop: among the 675 nt at these positions, 104 nt are counted as different in isodecoders. Therefore, 15.4% of these three nucleotides vary. The next most variable region is the D-loop (10.8%) among positions



**Figure 3.** Comparative sequence analysis of the tRNA<sup>Ser</sup>(AGA) isoacceptor family across six species. *S.c.*: budding yeast; *C.e.*: worm; *D.m.*: fruit fly; *M.m.*: mouse; *P.t.*: chimpanzee; *H.s.*: human. ‘-Nx’ indicates the gene copy number. For the non-mammalian species, the tRNA sequence variants are most similar within the same organism. The tRNA sequence variants are conserved across mammalian species. Each phylogenetic branch has unique sequence signatures (e.g. for the 2–71 bp, yeast sequence is GU; worm and fruit fly sequence is CG; and mammal sequence is TA).



**Figure 4.** Frequency of human isodecoder gene variations. Percentages indicate observed changes in each region divided by the total number of nucleotides assessed in that region. (A) Percent sequence variations in the A and B boxes which correspond to internal promoter regions of RNA polymerase III. (B) Percent sequence variations in nine regions according to the tRNA secondary structure. Invariant and anticodon nucleotides in all tRNAs are shown as filled black or gray circles. Percentages in parentheses (stems only) indicate sequence changes that result in non-Watson-Crick and non-GU base pairs.

15, 16–17 (variable from 1 to 3 nt) and 20 (variable from 1 to 3 nt). These high frequency regions overlap with the A and B boxes that constitute the internal promoters for Pol-III transcription. Sequence changes in the stems are between 2.6 and 9.4%. More than four-fifths of sequence changes in the stems follow the rules of Watson-Crick base pairing and G-U wobble. Of the remaining one-fifth of sequence changes that disrupt Watson-Crick or G-U pairing, 42% (30/72) are A-C pairs. A-C pairs in RNA helices have  $pK_a$  values of 6.0–6.5, and protonated A-C pairs are structurally analogous to and as stable as G-U pairs (24,25). A-C pairs in tRNA stems have been found to be functional in some bacterial tRNAs (26). The function of tRNA isodecoders containing A-C pairs may depend on local pH which can vary among subcellular environments.

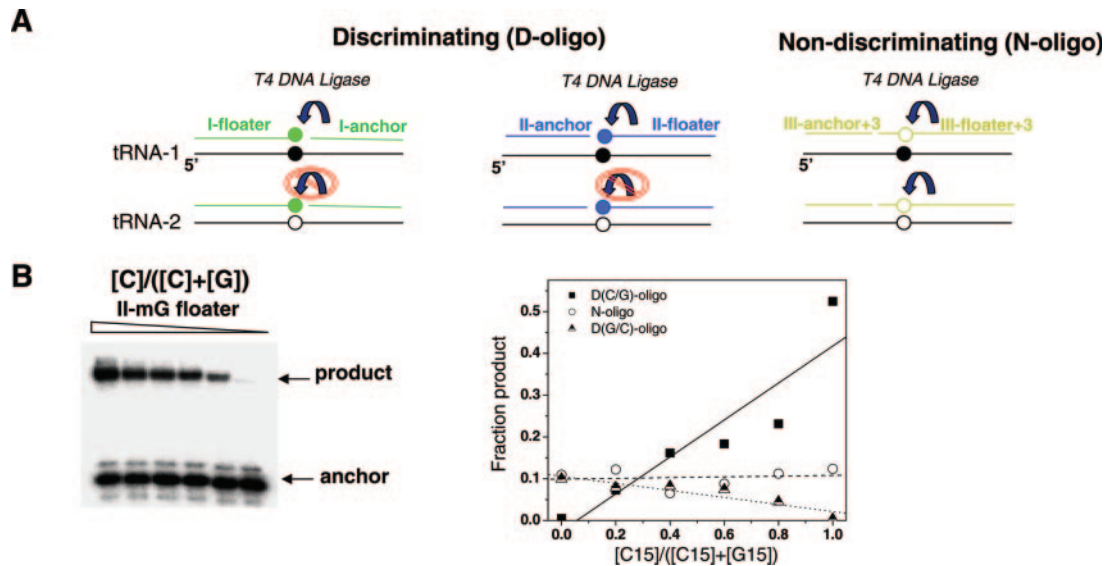
#### An experimental method to distinguish tRNA isodecoders

Experimental methods used to analyze the expression of RNA transcripts are generally based on hybridization differences of complementary oligonucleotide probes, primer extension using a mixture of deoxy and dideoxynucleotide triphosphates, and RT-PCR using primers that allow differential extension by reverse transcriptase. These methods work well when the RNA transcript is not very structured and post-transcriptional modifications do not impede hybridization or extension by the reverse transcriptase. Using purified tRNA mixture from HeLa cells, we tried to measure the expression of tRNA isodecoders by (i) differential hybridization of complementary oligonucleotides followed by RNase H cleavage; (ii) primer extension with up to 3× deoxynucleotide triphosphates and 1× dideoxynucleotide triphosphate; (iii) RT-PCR using primers with different 3' terminal nucleotides. Although HeLa tRNAs can sometimes

be detected by at least one of these methods, the result was either poorly reproducible or had very low sensitivity (data not shown). The primary problems of using these standard methods for eukaryotic tRNA appear to be derived from the extensive tRNA structure and the presence of tRNA modifications that interfere with hybridization/primer extension.

We devised a systematic method to distinguish tRNA isodecoder products that differ by a single nucleotide. The method is based on enzymatic ligation of two oligonucleotides using tRNA as template (Figure 5), similar to those described for the analysis of mRNA transcripts (27,28). In order to quantify the relative amount of two tRNA isodecoder products, two different types of oligonucleotide pairs are needed as ligation substrates. The first pair is only efficiently ligated using one of the two tRNA isodecoder templates. This pair of oligonucleotides is designated as discriminating (D-oligo), and there are two different D-oligos for each tRNA isodecoder pair. The second oligonucleotide pair is efficiently ligated for both tRNA isodecoders. This type is designated as non-discriminating (N-oligo), and there is one N-oligo for each tRNA isodecoder pair. The amount of ligation product using the D-oligo corresponds to  $[tRNA-1]/[tRNA-2]$  or  $[tRNA-2]/[tRNA-1]$ , whereas the amount of ligation product using the N-oligo corresponds to  $[tRNA-1]+[tRNA-2]$ . These data points together determine the relative amount of tRNA isodecoder pairs between two samples or may even be used to characterize the amount of tRNA isodecoder pairs in the same sample.

In order to find D-oligos for tRNA isodecoder pairs with a single nucleotide difference, we first determined the ligation efficiency using four model 30mer RNAs (Table 1). These RNAs have identical sequences except at the 15th position which is A, C, G or U. The ligation efficiency using these 30mer RNA templates is examined using 28 custom ordered oligonucleotide pairs in three configurations (Figure 5A



**Figure 5.** Detection of single nucleotide change in a model 30mer RNA by ligation. (A) The basic strategy. RNA oligonucleotides with a single nucleotide difference are used as templates for the ligation of two complementary oligonucleotides by T4 DNA ligase. To find the optimal 'solution' for each sequence at the 15th position (open and filled black circles), 28 oligonucleotides containing different sequences and backbone modifications at the ligation junction are tested. These oligos are named 'floaters' to distinguish them from the other oligonucleotide substrate that are the same in each set of the ligation reactions ('anchors'). The ligation junction is located 3' to the 15th position (set I), 5' to the 15th position (set II) and 3 nt away from the 15th position (set III). (B) Ligation reaction using a defined mixture of C15 and G15 30mer RNAs (left). The particular floater used (II-mG) is selected as the best solution that prefers C15 over G15 after screening 24 floaters in separate experiments. (Right) Quantification of the ligation products using the D-oligos (II-mG and II-dC floaters) and the N-oligo (III-dC+3) for C15 versus G15.

and Table 1). Set I and set II oligo pairs have the ligation junction at the 3' and 5' side of the 15th nucleotide in the RNA, respectively. Set III oligo pairs have the ligation junction displaced 3 nt downstream of the 15th nucleotide. Each set has one identical 'anchor' oligonucleotide substrate and 12 each (set I and II) or 4 (set III) different 'floater' oligonucleotide substrates. The floater oligonucleotides from sets I and II have different sequences or backbone modifications (Table 1). The floater oligonucleotides in set III have different sequences and the same 2' deoxy backbone.

Our results show that these 28 oligo pairs are sufficient to provide unique D and N-oligos for each of the 6 nt pairs in the RNA (Table 2). Under the standard ligation condition, the discrimination factors for the D-oligos are between 5-fold and 108-fold which should be sufficient for the discrimination of single nucleotide changes in tRNA isodecoder pairs. The amount of ligation product has a linear correlation with the known mixture of two model RNAs using the D-oligos and has little dependence using the N-oligo (e.g. C15 and G15 shown in Figure 5B).

### tRNA isodecoder in human samples

We next applied the D- and N-oligo solutions from model RNA studies to human samples to demonstrate the feasibility of this ligation approach for the analysis of biological RNAs (Figure 6). Probes for three tRNA isodecoder pairs were designed for (i) tRNA<sup>Pro</sup>(CGG) U39 versus C39; (ii) tRNA<sup>Ala</sup>(CGC) A42 versus G42 and (iii) tRNA<sup>Arg</sup>(UCG) A51G52 versus C51A52, in addition to probes for yeast tRNA<sup>Phe</sup> standard (Supplementary Figure S2). To facilitate detection, the length of the oligonucleotide substrates is

**Table 2.** D- and N-oligos for nucleotide pairs

Nucleotide pair	D-oligos	N-oligo
A and C	I-dT (A/C <sup>a</sup> ; 5 <sup>b</sup> ), II-mG (C/A, 48), RF (A/C, 1.4 <sup>c</sup> )	III-dG+3
A and G	II-rU (A/G; 41), II-dC (G/A, 27), RF (A/G, 20)	III-dC+3
A and U	II-mU (A/U; 34), I-dA (U/A, 18), RF (A/U, 21)	III-dA+3
C and G	II-mG (C/G; 108), II-dC (G/C, 30), RF (C/G, 5)	III-dC+3
C and U	II-mU (C/U; 6), I-dA (U/C, 10), RF (U/C, 6)	III-dG+3
G and U	II-dC (G/U; 107), I-dA (U/G, 91), RF (U/G, 1.8)	III-dG+3

<sup>a</sup>A/C = prefers A, C/A = prefers C and so on.

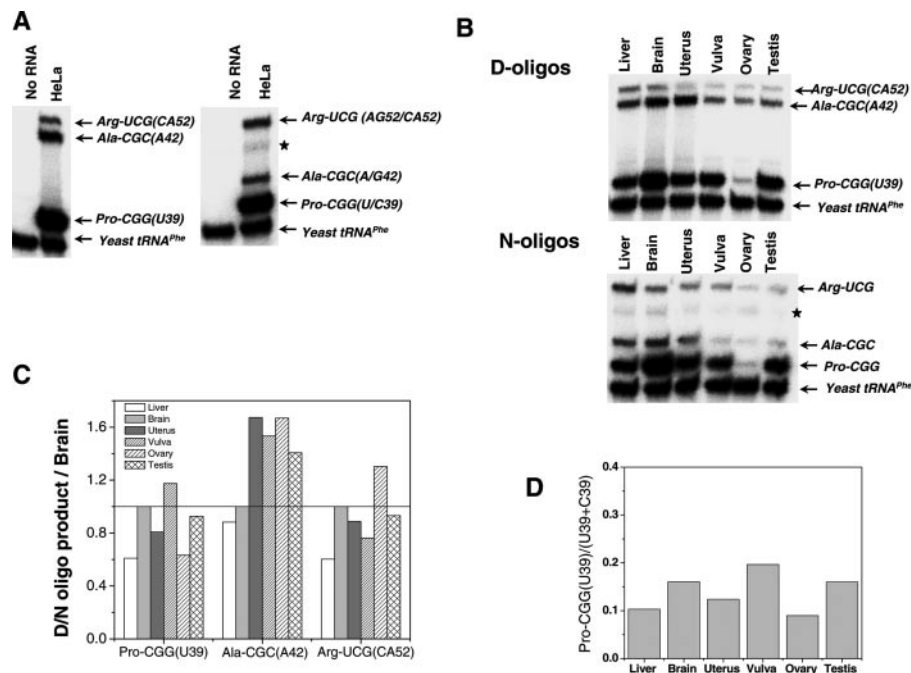
<sup>b</sup>x-Fold = amount of differential ligation product when using this D-oligo at 30°C and 0.125 U/μl T4 DNA ligase for 1 h (A+G; A+U; C+G; G+U) or 0.25 U/μl T4 DNA ligase for 0.5 h (A+C, C+U).

<sup>c</sup>RF: reaction factor = ratio of maximum amount of ligation product using specific D-oligos.

designed such that their reaction products differ by at least 5 nt. This way, the analysis of all four tRNAs can be carried out in a single ligation reaction. This length difference is achieved by the extension of a string of deoxy-A residues at the 5' end of the oligo substrates where necessary (Supplementary Figure S2).

When these oligo pairs are ligated using the total tRNA mixture from HeLa, varying amounts of ligation products are obtained (Figure 6A). The identity of these ligation products are confirmed by carrying out the ligation separately with each oligo pair (data not shown). This result shows that the ligation strategy to study tRNA isodecoder products works for biological RNA samples as well as the model RNA.

We then used these oligo pairs to compare the amount of the corresponding tRNA isodecoder products in the total



**Figure 6.** Detection of tRNA isodecoder distribution in human samples. (A) Simultaneous detection of three tRNA isodecoder pairs plus yeast tRNA<sup>Phe</sup> standard in a total tRNA mixture from HeLa. Asterisk (\*) indicates a ligation product derived from the mixture of N-oligos without tRNA (also present weakly in the 'No RNA' lane). (B) Simultaneous detection of three tRNA isodecoder pairs in a total tRNA mixture from six human tissues. (C) Relative amount of tRNA isodecoder after normalization to the total amount of its corresponding tRNA isodecoder pair in each tissue as compared to brain. (D) 'Absolute' ratio of tRNA<sup>Pro</sup>(CGG)-U39 in each tissue obtained from the ligation reactions using both D-oligos for tRNA<sup>Pro</sup>(CGG)-U39 and tRNA<sup>Pro</sup>(CGG)-C39.

tRNA mixture from six human tissues (Figure 6B). Total tRNAs from these tissues were first purified on a denaturing gel. To control for potential RNAs from different tissues that may alter the ligation efficiency, a constant, known amount of yeast tRNA<sup>Phe</sup> is included in every ligation reaction as control. The amount of the ligation product using the N-oligos shows that among tRNA<sup>Pro</sup>(CGG) and tRNA<sup>Ala</sup>(CGC), brain has the most and ovary and vulva have the least of these tRNAs. The brain sample produces more D-oligo products for tRNA<sup>Pro</sup>(CGG) and tRNA<sup>Ala</sup>(CGC) as well.

The ratio of the D-oligo product divided by the N-oligo product for the same tissue after normalization to the yeast tRNA<sup>Phe</sup> standard can be used to compare the relative amount for one particular tRNA isodecoder in each tissue (Figure 6C). This analysis shows that although the total amount of these tRNAs can be significantly different in these tissues, the relative amount is all within 2-fold to that in brain.

We also attempted to determine the 'absolute' ratio of tRNA<sup>Pro</sup>(CGG)-U39 and C39 using their corresponding D-oligo pairs, i.e. one prefers U over C and the other prefers C over U (D1 and D2 in Supplementary Figure S2). tRNA<sup>Pro</sup>(CGG) can be detected in each tissue using D1 or D2-oligo pairs. However, the relative ligation efficiency for the U-preferring D-oligo is six times greater than the C-preferring D-oligo when using the model 30mer RNA template. Assuming this relative reaction factor (U/C = 6; Table 2) is the same for tRNA<sup>Pro</sup>(CGG), the fraction of tRNA<sup>Pro</sup>(CGG)-U39 is then obtained according to  $[D(U39)\text{-product}]/([D(U39)\text{-product}] + 6 \times [D(C39)\text{-product}])$

(Figure 6D). This fraction of tRNA<sup>Pro</sup>(CGG)-U39 is between 0.09 and 0.20 among these tissues, and the remaining fraction is presumably tRNA<sup>Pro</sup>(CGG)-C39. Three of the four tRNA<sup>Pro</sup>(CGG) genes have C39 and one has U39. Hence, a completely unbiased expression should generate a fraction of 0.25. This fraction is close to that obtained from vulva, but markedly higher than that from ovary.

## DISCUSSION

This work shows that eukaryotic tRNA genes derived from genome sequences are more complex than previously appreciated. A surprise finding is the prevalence of tRNA isodecoder genes, defined as tRNA genes having the same anticodon but different sequences elsewhere in the tRNA body. More than 50% of human and chimp tRNA genes are isodecoders. Most sequence changes in tRNA isodecoder genes follow sequence constraints for the secondary and tertiary structure of tRNA. The phylogenetic relationship of tRNA isodecoder genes in 11 eukaryotes suggests that some may perform unique functions in organisms belonging to the same phylogenetic branch. The relatively high percentage of sequence changes in the internal RNA polymerase III promoter regions suggest that differential control for the expression of some isodecoder genes is feasible.

Pressing questions on tRNA isodecoders clearly are about their functions. The increase in the tRNA gene number and subsequent divergence is likely the result of vertebrate genome expansion. It is possible that they are derived from neutral drift and do not perform a distinct function. On the



other hand, evolved variations may confer some unique role yet to be determined. Unfortunately, few efforts have been devoted to study the potential functions of eukaryotic tRNA isodecoders. The benefit of having a collection of tRNA isodecoders in translation is unclear at this time. One can envision that tRNA isodecoders may be more harmful than useful in translation. In tRNA charging, each aminoacyl-tRNA synthetase identifies a set of positive sequence/structural determinants in the tRNA to ensure charging with the correct amino acid (29,30). *In vivo*, each tRNA also carries negative determinants to prevent mis-charging by the 19 other tRNA synthetases (31,32). When sequence changes occur in the tRNA body, there is a chance that the resulting tRNA isodecoder may become mis-charged more frequently. Some tRNA synthetases have exquisite editing mechanisms to eliminate mis-charged tRNAs, although these mechanisms have been studied primarily on correct tRNA isoacceptors charged with wrong amino acids (33,34). Mis-charging of tRNA isodecoders may not be subject to these editing mechanisms because they are the result of charging the right amino acid to the wrong tRNA.

Some bacterial tRNAs regulate transcription through attenuation or anti-termination mechanisms (35,36). In *Bacillus subtilis*, several uncharged tRNAs derived from perturbation of its cognate amino acid metabolism directly interact with the 5'-untranslated region in certain mRNA operons. The tRNA-mRNA interaction helps control transcription of amino acid biosynthesis enzymes or aminoacyl-tRNA synthetases. A subset of eukaryotic tRNA isodecoders may also interact with mRNAs for translational regulation.

Some bacterial tRNAs are used independently of translation as a source of activated amino acids in cell wall biosynthesis, N-terminal attachment to proteins, or substrates for enzymes in intermediary metabolism (37). A subset of eukaryotic tRNA isodecoders may also be used for similar purposes and function in non-translation activities (38).

Retroviral reverse transcription requires tRNA primers, e.g. tRNA<sup>Lys3</sup>(UUU) for HIV-1 (39). Only 6 of the 15 human tRNA<sup>Lys</sup>(UUU) genes have the complete complementary primer binding site for the HIV-1 genomic RNA. A subset of tRNA isodecoder may be selected to be optimal primers for retroviral life cycle.

Speculative roles of tRNA isodecoders are manifold. This work also describes a systematic ligation-based method to detect and quantify relative amounts of tRNA isodecoders in biological samples. This method has single nucleotide resolution and is best used to distinguish pairs of tRNA isodecoders. The method is also able to simultaneously analyze many tRNA isodecoders in different isoacceptor classes. One can envision the adaptation of this method to a microarray format in order to study the expression of tRNA isodecoder pairs from every isoacceptor family.

We now have the ability to examine tRNA isodecoder pairs in biological samples, and we also have the ability to measure cumulative expression from multiple tRNA isodecoders of an isoacceptor family using microarrays (T.P., unpublished data). The combination of these methods should be sufficient to initiate functional studies of tRNA isodecoders in a high-throughput manner. The best target for such studies may be mammals where the number of tRNA isodecoder genes is plentiful and some tRNA

isodecoders may perform functions that are uniquely mammalian.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Dr J. Yewdell for stimulating discussions and one reviewer for insightful comments. This work was supported in part by a CCSG pilot project at the University of Chicago and the NIH training grant (T32 GM007197). Funding to pay the Open Access publication charges for this article was provided by NIH.

*Conflict of interest statement.* None declared.

## REFERENCES

1. Sprinzl, M. and Vassilenko, K.S. (2005) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.*, **33**, D139–D140.
2. Ikemura, T. and Ozeki, H. (1983) Codon usage and transfer RNA contents: organism-specific codon-choice patterns in reference to the isoacceptor contents. *Cold Spring Harb. Symp. Quant. Biol.*, **47**, 1087–1097.
3. Ikemura, T. (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.*, **2**, 13–34.
4. Sharp, P.M. and Li, W.H. (1987) The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.*, **15**, 1281–1295.
5. Dong, H., Nilsson, L. and Kurland, C.G. (1996) Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J. Mol. Biol.*, **260**, 649–663.
6. Kanaya, S., Yamada, Y., Kudo, Y. and Ikemura, T. (1999) Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene*, **238**, 143–155.
7. Fuglsang, A. (2003) Codon optimizer: a freeware tool for codon optimization. *Protein Expr. Purif.*, **31**, 247–249.
8. Withers, M., Wernisch, L. and dos Reis, M. (2006) Archaeology and evolution of transfer RNA genes in the *Escherichia coli* genome. *RNA*, **12**, 933–942.
9. 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.
10. 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.
11. Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A. and Steinberg, S. (1998) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.*, **26**, 148–153.
12. Blattner, F.R., Plunkett, G., III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F. et al. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science*, **277**, 1453–1474.
13. Hagenbuchle, O., Larson, D., Hall, G.I. and Sprague, K.U. (1979) Primary transcription product of a silkworm alanine tRNA gene: identification of *in vitro* sites of initiation, termination and processing. *Cell*, **18**, 1217–1229.
14. Sprague, K.U. (1995) Transcription of eukaryotic tRNA genes. In Soll, D. and RajBhandary, U. (eds), *tRNA: Structure, Biosynthesis, and Function*. ASM Press, Washington DC, pp. 31–50.
15. Ching, O.Y., Martinez, M.J., Young, L.S. and Sprague, K.U. (2000) TATA-binding protein–TATA interaction is a key determinant of differential transcription of silkworm constitutive and silk gland-specific tRNA(Ala) genes. *Mol. Cell. Biol.*, **20**, 1329–1343.

16. Harada,F., Matsubara,M. and Kato,N. (1989) Nucleotide sequences of two glutamine tRNAs from HeLa cells. *Nucleic Acids Res.*, **17**, 8371.
17. Dittmar,K.A., Mobley,E.M., Radek,A.J. and Pan,T. (2004) Exploring the regulation of tRNA distribution on the genomic scale. *J. Mol. Biol.*, **337**, 31–47.
18. Dittmar,K.A., Sorensen,M.A., Elf,J., Ehrenberg,M. and Pan,T. (2005) Selective charging of tRNA isoacceptors induced by amino-acid starvation. *EMBO Rep.*, **6**, 151–157.
19. Thompson,J.D., Gibson,T.J., Plewniak,F., Jeanmougin,F. and Higgins,D.G. (1997) The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, **25**, 4876–4882.
20. Saitou,N. and Nei,M. (1987) The neighbor-joining method—a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**, 406–425.
21. Perriere,G. and Gouy,M. (1996) WWW-Query: an on-line retrieval system for biological sequence banks. *Biochimie*, **78**, 364–369.
22. Paule,M.R. and White,R.J. (2000) Survey and summary: transcription by RNA polymerases I and III. *Nucleic Acids Res.*, **28**, 1283–1298.
23. Geiduschek,E.P. and Kassavetis,G.A. (2001) The RNA polymerase III transcription apparatus. *J. Mol. Biol.*, **310**, 1–26.
24. Cai,Z. and Tinoco,I., Jr (1996) Solution structure of loop A from the hairpin ribozyme from tobacco ringspot virus satellite. *Biochemistry*, **35**, 6026–6036.
25. Schroeder,S.J., Burkard,M.E. and Turner,D.H. (1999) The energetics of small internal loops in RNA. *Biopolymers*, **52**, 157–167.
26. McClain,W.H. (2006) Surprising contribution to aminoacylation and translation of non-Watson–Crick pairs in tRNA. *Proc. Natl Acad. Sci. USA*, **103**, 4570–4575.
27. Nilsson,M., Antson,D.O., Barbany,G. and Landegren,U. (2001) RNA-templated DNA ligation for transcript analysis. *Nucleic Acids Res.*, **29**, 578–581.
28. Nilsson,M., Barbany,G., Antson,D.O., Gertow,K. and Landegren,U. (2000) Enhanced detection and distinction of RNA by enzymatic probe ligation. *Nat. Biotechnol.*, **18**, 791–793.
29. Giege,R., Sissler,M. and Florentz,C. (1998) Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.*, **26**, 5017–5035.
30. Ibba,M. and Soll,D. (1999) Quality control mechanisms during translation. *Science*, **286**, 1893–1897.
31. Normanly,J. and Abelson,J. (1989) tRNA identity. *Annu. Rev. Biochem.*, **58**, 1029–1049.
32. McClain,W.H. (1993) Rules that govern tRNA identity in protein synthesis. *J. Mol. Biol.*, **234**, 257–280.
33. Schimmel,P. and Ribas de Pouplana,L. (2001) Formation of two classes of tRNA synthetases in relation to editing functions and genetic code. *Cold Spring Harb. Symp. Quant. Biol.*, **66**, 161–166.
34. Geslain,R. and Ribas de Pouplana,L. (2004) Regulation of RNA function by aminoacylation and editing? *Trends Genet.*, **20**, 604–610.
35. Yanofsky,C. (2000) Transcription attenuation: once viewed as a novel regulatory strategy. *J. Bacteriol.*, **182**, 1–8.
36. Henkin,T.M. and Yanofsky,C. (2002) Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. *Bioessays*, **24**, 700–707.
37. Soll,D. (1993) Transfer RNA: an RNA for all seasons. In Gesteland,R.F. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 157–184.
38. Karakozova,M., Kozak,M., Wong,C.C., Bailey,A.O., Yates,J.R., III, Mogilner,A., Zebroski,H. and Kashina,A. (2006) Arginylation of beta-actin regulates actin cytoskeleton and cell motility. *Science*, **313**, 192–196.
39. Litvak,S., Sarih-Cottin,L., Fournier,M., Andreola,M. and Tarrago-Litvak,L. (1994) Priming of HIV replication by tRNA(Lys3): role of reverse transcriptase. *Trends Biochem. Sci.*, **19**, 114–118.
40. Wheeler,D.L., Chappay,C., Lash,A.E., Leipe,D.D., Madden,T.L., Schuler,G.D., Tatusova,T.A. and Rapp,B.A. (2000) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **28**, 10–14.
41. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J., Rapp,B.A. and Wheeler,D.L. (2000) GenBank. *Nucleic Acids Res.*, **28**, 15–18.