Gene Expression

$stAI_{calc}$: tRNA Adaptation Index Calculator based on Species-Specific weights

Renana Sabi[1,*], Renana Volvovitch Daniel[1,*] and Tamir Tuller[1,2,+]

[1] Department of Biomedical Engineering, Tel Aviv University, Ramat Aviv 69978, Israel

[2] The Sagol School of Neuroscience, Tel Aviv University, Ramat Aviv 69978, Israel

*Equal contribution

+To whom correspondence should be addressed

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXX

ABSTRACT

Summary:

The tRNA Adaptation Index (tAI) is a tRNA-centric measure of translation efficiency which includes weights that take into account the efficiencies of the different wobble interactions. To enable the calculation of the index based on a species-specific inference of these weights, we created the $stAI_{calc}$. The calculator includes optimized tAI weights for 100 species from the three domains of life along with a standalone software package that optimizes the weights for new organisms.

The tAI with the optimized weights should enable performing large scale studies in disciplines such as molecular evolution, genomics, systems biology and synthetic biology.

Availability and Implementation: The calculator is publicly available at http://tau-tai.azurewebsites.net/

Contact: tamirtul@post.tau.ac.il (TT)

1 INTRODUCTION

mRNA translation is one of the central stages of protein biosynthesis in which ribosomes translate a poly-nucleotide sequence of mRNA codons into a poly-amino acid sequence. As the ribosome encounters a new codon, it halts until a complementary tRNA anti-codon is presented and delivers its amino acid to the growing polypeptide chain. The codon-anticodon pairing is enabled by both, the canonical Watson-Crick interactions and wobble interactions (Crick, 1966).

The tRNA Adaptation Index (tAI) is a measure of translational efficiency which takes into account the intracellular concentration of tRNA molecules and the efficiencies of each codon-anticodon pairing (dos Reis, et al., 2004). Specifically, the index includes weights, originally derived based on gene expression in *S. cerevisiae*, that quantify the efficiency of each wobble interaction. The efficiencies of these interactions are expected to vary across the tree of life, thus, the tAI weights related to the wobble interactions are expected to be different for different organisms. In (Sabi and Tuller, 2014) an organism-specific approach for inferring the tAI wobble weights was suggested; the inference was based on the fact that highly expressed genes tend to have higher codon usage bias (CUB) and higher adaptation to the tRNA pool (Sabi and Tuller, 2014). The tAI with the optimized weights was shown to improve predictions of protein abundance in various non-fungal organisms.

The advantages of the tAI with the optimized weights are threefold. First, the tAI is a biophysical measure of adaptation of codons to the intracellular tRNA pool. Indeed, in the recent years, the tAI (and variants of this measure) has been widely used to study fundamental questions related to gene expression regulation, molecular evolution, and more (Brockmann, et al., 2007; Ciandrini, et al., 2013; Dana and Tuller, 2014; dos Reis, et al., 2004; Gingold and Pilpel, 2011; Goodman, et al., 2013; Man and Pilpel, 2007; Novoa and de Pouplana, 2012; Pechmann and Frydman, 2013; Sharp, et al., 2005; Tuller, et al., 2010; Tuller, et al., 2011).

Second, since the wobble weights are inferred individually for each specific organism, tAI calculated based on these weights has the potential to outperform tAI based on the yeast-weight in predicting protein levels/translation efficiency in non-fungi (Sabi and Tuller, 2014).

Third, the organism-specific weights are inferred based *solely on the coding sequence of the organism* and thus can be calculated for many organisms for which expression data is not available. Specifically today there are thousands of organisms with sequenced genomes and only a few dozen organisms with large scale gene expression measurements.

Here we present the $stAI_{calc}$, an online tool that calculates tAI with optimized wobble weights based on the approach of (Sabi and Tuller, 2014). The tool provides the tAI value of the sequence as well as the normalized tAI values of each coding codon in the selected genome. In addition, it includes an option to download a standalone software package of $stAI_{calc}$ that optimizes the weights

for new organisms not included in the list. Implementation examples and application description appear below.

2 APPLICATION DESCRIPTION

The usage of $stAI_{calc}$ is simple and convenient. Entering the $stAI_{calc}$ webpage, the user should select an organism from a list of optional organisms and provide the query genes' coding sequences in a FASTA format file. For organisms not included in the list, a standalone software package is available for download.

By the tAI, the estimated translation efficiency of the translated codon depends on the intracellular concentration of the isoacceptor tRNA and the efficiency of the codon-anticodon pairing. Specifically, the translational efficiency of the *i*-th codon (out of 61 possible codons) is given by:

$$W_i = \sum_{j=1}^{n_i} (1 - s_{ij}) \cdot tGCN_{ij}$$

Where n_i is the number of tRNA types/anticodons that pair with the *i*-th codon, *tGCNij* is the tRNA gene copy number (Copy numbers can be retrieved, for example, from the genomic tRNA database at http://gtrnadb.ucsc.edu (Chan and Lowe, 2009). As explained in the user guide tRNA levels can be rather used if available), and *Sij* is a value between 0 and 1 representing a constraint on the pairing between the *i*-th codon and the *j*-th tRNA. Normalized weights (w_i) are obtained from the *Wi* by dividing each *Wi* by the maximum value among all codons. *Sij* weight closer to 0 represent more efficient wobble interaction (Figure 2A). Zero *Wi* weights are set to the mean w_i.

In (Sabi and Tuller, 2014), the *Sij* weights were optimized based on a hill-climbing approach with multiple initial points that aims at maximizing the correlation between tAI and a measure of codon usage bias (inspired by the relative CUB, (Roymondal, et al., 2009)). Statistics regarding the similarity/difference between the resultant optimized weights and the original yeast weights are shown in Figure 1. As can be seen, while there is a relation/positive correlation with the original yeast weights they are typically not extremely similar (median correlation is 0.5).

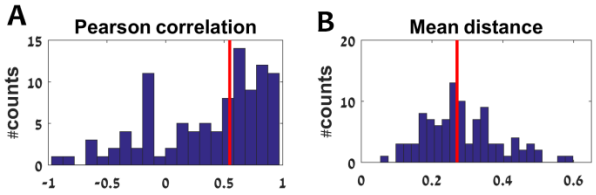

Figure 1. Comparison between the yeast and the species-specific *Sij* vectors. The histograms of the 100 species (blue) and median values (red) are shown. **(A)** Pearson correlations between yeast and species-specific *Sij*. **(B)** The mean distance between the yeast and the species-specific *Sij* vectors; calculated by taking the mean of the sum of the absolute differences between each pair of corresponding components in the two vectors (thus, the maximum distance is 1 and the minimum is 0).

Illustration of wobble interactions and optimized weights is described in figure 2.A. tAI based on these species-specific weights are expected to correlate with protein levels in organisms where CUB is related to the amount of adaptation to the tRNA pool (for example, Figure 2B). The online tool includes a total of 100 different species from the three domains of life. The specific genome source and GenBank record used for the inference in (Sabi and Tuller, 2014) are included within the name of each organism in the list. The coding sequences of the species used for the *Sij* weights inference were retrieved from the NCBI (http://ftp.ncbi.nih.gov/genomes/). tRNA copy numbers were obtained from the Genomic tRNA Database (http://gtrnadb.ucsc.edu/). For organisms not included in the list, a standalone $stAI_{calc}$ package is available for download. This package also provides the final correlation between the optimized tAI and CUB for the user evaluation. A detailed user guide for the online tool and the standalone package is also provided in the website.

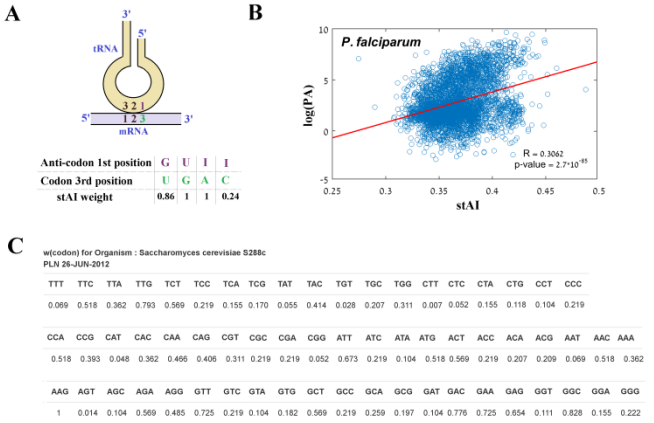

Figure 2. A. An illustration of wobble interaction between the third position of a mRNA codon (green) and the first position of a tRNA anti-codon (purple). The four possible wobble couplings and their tAI weights (*Sij*) in *S. cerevisiae* are summarized below. The I:U pairing is referred as a canonical interaction (thus its s-value is 0). **B**. Spearman's rank correlation between tAI and protein abundance (PA) in *Plasmodium falciparum*. PA measurements were retrieved from the protein abundance database at http://pax-db.org/ (Wang, et al., 2012). **C**. Normalized tAI weights of each codon in *S.cerevisiae* based on the optimized *Sij* as appear online in $stAI_{calc}$.

3 SMALL-SCALE EXAMPLE

The following example describes the implementation of the online $stAI_{calc}$ for a *S. cerevisiae* gene. At the calculator first page, we select from the list the organism *Saccharomyces cerevisiae* S288c. A FASTA formatted text file containing the gene coding sequence must be provided. For example:

> YAR035C-A

ATGAGATTAAATTACAGTCGCTGCTATTATAGTAGCCAGCGGCGGCGCCAGTCACTTCC

AAAGCGTTTTCCTCTTATTTAG

The tAI value of this gene is 0.1198. Below the value, the relative stAI weights (w_i) in *S. cerevisiae* are listed (Figure 2C). The output is shown on the screen and can be exported into a text file.

ACKNOWLEDGEMENT

R.S. is supported in part by the Edmond J Safra Bioinformatics center at Tel Aviv University.

Conflict of Interest: none declared.

REFERENCES

Brockmann, R.*, et al.* (2007) Posttranscriptional expression regulation: what determines translation rates?, *PLoS Comput Biol*, **3**, e57.

Chan, P.P. and Lowe, T.M. (2009) GtRNAdb: a database of transfer RNA genes detected in genomic sequence, *Nucleic acids research*, **37**, D93-D97.

Ciandrini, L., Stansfield, I. and Romano, M.C. (2013) Ribosome traffic on mRNAs maps to gene ontology: genome-wide quantification of translation initiation rates and polysome size regulation, *PLoS Comput Biol*, **9**, e1002866.

Crick, F.H. (1966) Codon—anticodon pairing: the wobble hypothesis, *Journal of molecular biology*, **19**, 548-555.

Dana, A. and Tuller, T. (2014) The effect of tRNA levels on decoding times of mRNA codons, *Nucleic acids research*, **42**, 9171-9181.

dos Reis, M., Savva, R. and Wernisch, L. (2004) Solving the riddle of codon usage preferences: a test for translational selection, *Nucleic acids research*, **32**, 5036-5044.

Gingold, H. and Pilpel, Y. (2011) Determinants of translation efficiency and accuracy, *Molecular systems biology*, **7**, 481.

Goodman, D.B., Church, G.M. and Kosuri, S. (2013) Causes and effects of N-terminal codon bias in bacterial genes, *Science*, **342**, 475-479.

Man, O. and Pilpel, Y. (2007) Differential translation efficiency of orthologous genes is involved in phenotypic divergence of yeast species, *Nature genetics*, **39**, 415-421.

Novoa, E.M. and de Pouplana, L.R. (2012) Speeding with control: codon usage, tRNAs, and ribosomes, *Trends in Genetics*, **28**, 574-581.

Pechmann, S. and Frydman, J. (2013) Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding, *Nature structural & molecular biology*, **20**, 237-243.

Roymondal, U., Das, S. and Sahoo, S. (2009) Predicting gene expression level from relative codon usage bias: an application to Escherichia coli genome, *DNA research*, **16**, 13-30.

Sabi, R. and Tuller, T. (2014) Modelling the Efficiency of Codon–tRNA Interactions Based on Codon Usage Bias, *DNA Research*, **21**, 511-526.

Sharp, P.M.*, et al.* (2005) Variation in the strength of selected codon usage bias among bacteria, *Nucleic acids research*, **33**, 1141-1153.

Tuller, T.*, et al.* (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation, *Cell*, **141**, 344-354.

Tuller, T.*, et al.* (2011) Composite effects of gene determinants on the translation speed and density of ribosomes, *Genome Biol*, **12**, R110.

Wang, M.*, et al.* (2012) PaxDb, a database of protein abundance averages across all three domains of life, *Molecular & Cellular Proteomics*, **11**, 492-500.