Biochemical and Biophysical Research Communications xxx (2018) 1–6

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Codon usage revisited: Lack of correlation between codon usage and the number of tRNA genes in enterobacteria

Joaquín Rojas [a], Gabriel Castillo [a], Lorenzo Eugenio Leiva [a], Sara Elgamal [b], Omar Orellana [a], Michael Ibba [b], Assaf Katz [a, *]

[a] Programa de Biología Celular y Molecular, ICBM, Facultad de Medicina, Universidad de Chile, Santiago, 8380453, Chile
[b] Department of Microbiology and The Center for RNA Biology, Ohio State University, Columbus, OH, 43210, USA

ARTICLE INFO

Article history:
Received 17 May 2018
Accepted 25 May 2018
Available online xxx

Keywords:
tRNA
Codon usage
Efficiency of translation
Enterobacteria
Salmonella enterica
Escherichia coli

ABSTRACT

It is widely believed that if a high number of genes are found for any tRNA in a rapidly replicating bacteria, then the cytoplasmic levels of that tRNA will be high and an open reading frame containing a higher frequency of the complementary codon will be translated faster. This idea is based on correlations between the number of tRNA genes, tRNA concentration and the frequency of codon usage observed in a limited number of strains as well as from the fact that artificially changing the number of tRNA genes alters translation efficiency and consequently the amount of properly folded protein synthesized. tRNA gene number may greatly vary in a genome due to duplications, deletions and lateral transfer which in turn would alter the levels and functionality of many proteins. Such changes are potentially deleterious for fitness and as a result it is expected that changes in tRNA gene numbers should be accompanied by a modification of the frequency of codon usage. In contrast to this model, when comparing the number of tRNA genes and the frequency of codon usage of several *Salmonella enterica* and *Escherichia coli* strains we found that changes in the number of tRNA genes are not correlated to changes in codon usage. Furthermore, these changes are not correlated with a change in the efficiency of codon translation. These results suggest that once a genome gains or loses tRNA genes, it responds by modulating the concentrations of tRNAs rather than modifying its frequency of codon usage.

1. Introduction

Proteins, which are essential for the physiology of all organisms, are coded in nucleic acids. In order to translate the genetic message contained in nucleic acids into a polypeptide, organisms use a code where each three contiguous nucleotides (a codon) are translated to a specific amino acid. tRNAs are essential for this process. In one side of the folded tRNAs there is a loop that contains a three nucleotide sequence called the anticodon, that can interact specifically with the complementary codons in mRNAs during translation. The 3′-extreme of the tRNA is able to carry an amino acid that can be transferred to a nascent peptide. This reaction is catalyzed by the ribosome that additionally ensures a correct matching between the codon in the mRNA and the anticodon in tRNA [1–3].

Although there are some exceptions, in most organisms and growth conditions each codon codes for a single amino acid with high precision [4–7]. Nevertheless, the standard genetic code is redundant with 61 codons that code for only 20 canonical amino acids. Thus, while Met and Trp are coded by a single codon, Arg, Leu and Ser are coded by as many as 6 different codons [8]. Thanks to wobble interactions -where the ribosome allows interactions between non Watson-Crick base pairs at the third codon position- a tRNA can recognize several codons. This allows one tRNA to decode several codons (coding for the same amino acid) and also a single codon to be decoded by several tRNAs (carrying the same amino acid). There has been a long debate regarding the role of such redundancy in protein synthesis. One of the most accepted ideas is that different codons coding for the same amino acid will be translated at a different speed. Thus, highly expressed proteins will require their genes to be coded mostly by codons that are efficiently translated which are expected to correlate with high cellular levels of the corresponding tRNAs. This idea is based on the fact that translation elongation speed depends on the concentration of aa-tRNAs [8]. Also, it has been shown that in rapidly replicating

* Corresponding author. Programa de Biología Celular y Molecular, ICBM, Facultad de Medicina, Universidad de Chile, Santiago, 8380453, Chile.
E-mail address: akatz@med.uchile.cl (A. Katz).

https://doi.org/10.1016/j.bbrc.2018.05.168
0006-291X/

organisms, there is a positive correlation between tRNA concentration (or the number of tRNA genes, see below) and the usage of the codons they decode in highly expressed genes [8,9]. Consistent with gene copy number effects, artificially changing the concentration of tRNAs in an organism [10] or the frequency of codons in a gene [11,12] alters gene expression and cell fitness [13]. Additionally, usage of infrequent codons that are expected to be translated at slower rates has been associated with modulation of processes coupled to translation such as protein folding or secretion [3,8,11,14]. In fact, changing codon usage in genes may induce the production of incorrectly folded proteins [3,11,14].

Bacteria can easily obtain genetic material through lateral transfer, a process that is essential for a rapid adaptation of the genome to new environments. Many of the mechanisms involved in the transfer of genetic material involve the transfer of tRNA genes. For instance, viral genomes, genomic islands and plasmids have been shown to carry one or several tRNA genes [15–18]. Internal recombination of different areas of a genome can also duplicate or produce the loss of genes coding for tRNAs [18]. Thus, the number of genes coding for tRNAs in bacterial species can change in a single generation. Concentration of tRNAs has been shown to be correlated to the number of genes coding for them, at least in bacteria such as *Escherichia coli* [19,20]. Thus, changes in the number of tRNA genes are expected to alter the cellular concentration of the corresponding tRNAs. As discussed above, this is expected to modify the speed of translation of some of the codons altering the levels and functionality of at least part of the proteome.

McDonald et al. have shown that changes in the number of tRNA genes derived from lateral transfer events correlate with codon usage of accompanying genes acquired by such events in genomes of *Escherichia coli* and *Shigella*. Potentially, these correlations allow a more efficient translation of the newly acquired genes [18]. We have found previously a similar trend while analyzing codon usage and tRNA gene content of an integrative conjugative element present in the genome of the chemolithoautotrophic bacterium *Acidithiobacillus ferrooxidans* [15,21]. Nevertheless, recent reports have shown that most tRNAs in this mobile element are expressed at very low levels [22], questioning their ability to improve gene expression. If it is common that changes in tRNA gene copy numbers produce only small changes in tRNA levels, then we would expect these have minor effects on gene expression and, consequently, on the frequency of codon usage in most genes of the genome. To test this hypothesis, we studied the relationship between the number of tRNA genes and the frequency of codon usage in genomes of enterobacteria. We have selected as models two well studied enterobacteria, *Escherichia coli* and *Salmonella enterica*. Our results indicate that changes in the number of genes coding for tRNAs have only minor effects on codon usage of both the whole genome and highly expressed genes. Correspondingly, fusions of codons translated by these tRNAs to *gfp*, have little effect on its translation efficiency. These results suggest that expression of tRNA genes acquired by lateral transfer in enterobacteria is rapidly adapted to the requirements of the host genome.

2. Materials and methods

2.1. Selection of analyzed genomes

RefSeq versions of genomes from *E. coli* and *Salmonella* were downloaded from NCBI ftp site. Only genomes annotated as being at “Chromosome” or “Complete Genome” levels were used for further analyses. In order to reduce excessive sequence

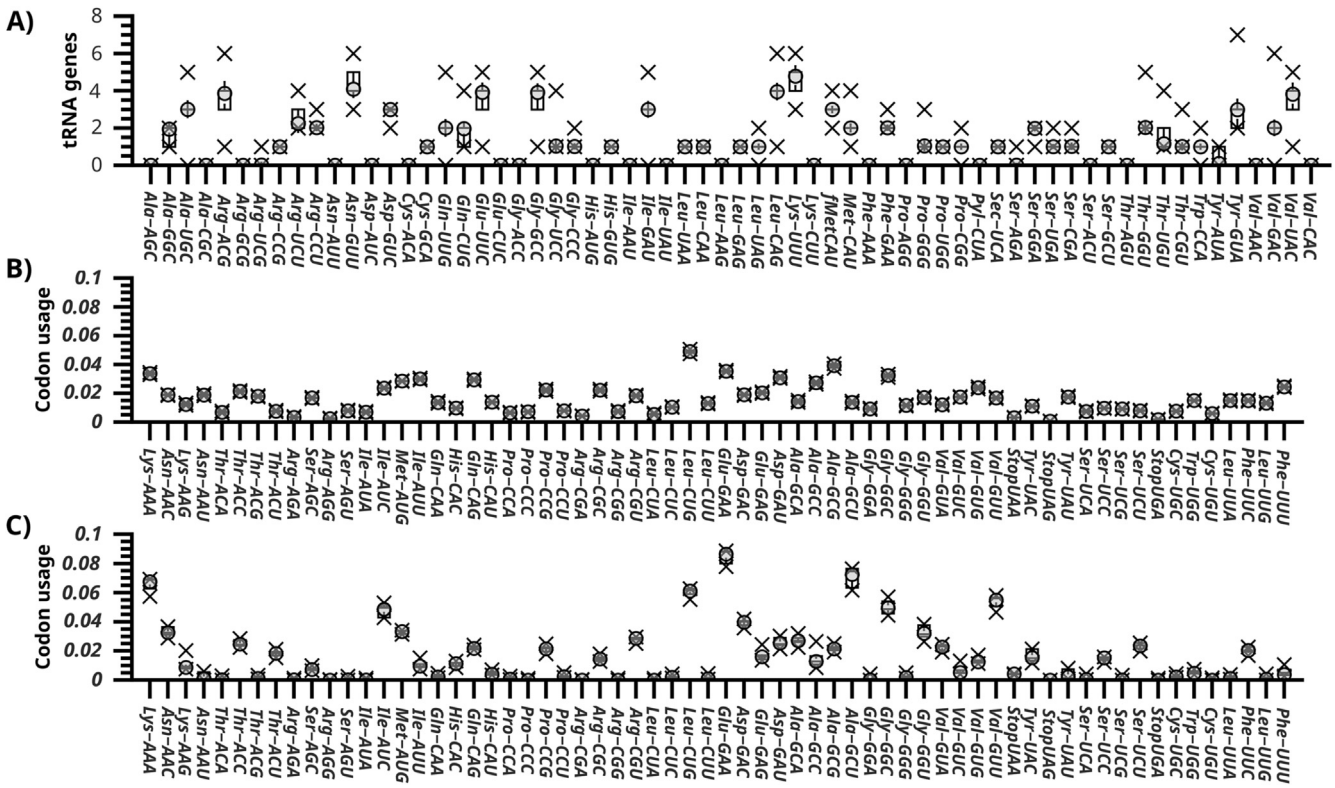

Fig. 1. tRNA gene copy numbers and frequency of codon usage in *Salmonella enterica*. The box plots show A) the average of tRNA gene copy numbers for each possible anticodon and B) the average of frequency of codon usage in all or C) a subset of highly expressed genes in 139 *Salmonella enterica* strains. In all graphs an “X” symbol indicates the most extreme values, a circle indicates the mean value and the horizontal lines of the box indicate limits where 5, 50 and 95% of the data is contained. Whiskers indicate standard deviations.

Please cite this article in press as: J. Rojas, et al., Codon usage revisited: Lack of correlation between codon usage and the number of tRNA genes in enterobacteria, Biochemical and Biophysical Research Communications (2018), https://doi.org/10.1016/j.bbrc.2018.05.168