Gene 745 (2020) 144640

Contents lists available at ScienceDirect

Gene

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

Research paper

The combined influence of codon composition and tRNA copy number regulates translational efficiency by influencing synonymous nucleotide substitution

Manish P. Victor[a], Debarun Acharya[b,*], Sandip Chakraborty[a], Tapash C. Ghosh[a,c,*]

[a] Division of Bioinformatics, Bose Institute, Kolkata, West Bengal, India
[b] Department of Microbiology, Bose Institute, Kolkata, West Bengal, India
[c] Department of Microbiology, Raiganj University, Raiganj, West Bengal, India

ARTICLE INFO

Keywords:
Chaperone
Co-adaptation index
Codon usage bias
mRNA abundance
Nucleotide substitution
Protein abundance
tRNA

ABSTRACT

Codon usage bias is an important genomic phenomenon, where highly expressed genes use optimal codons for smoother translation with high yield, facilitated by the cognate tRNAs. Here, we presented the tRNA co-adaptation index (co-AI) by correlating tRNA gene copy number and codon composition in *Saccharomyces cerevisiae*. We observed that this co-AI is positively correlated with protein abundance and translation rate. Considering nucleotide substitutions, co-AI influences synonymous substitutions more than gene expression and protein abundance, the most important determinants of evolutionary rate. Co-AI correlates positively with mRNA secondary structure stability and mRNA half-life, which may lead to protein accumulation under high co-AI. However, the highly expressed proteins encoded by high co-AI genes are assisted by molecular chaperones to attain their proper functional conformation and prevent accumulation.

1. Introduction

The universal genetic code is redundant, where 61 codons bind to their respective complementary (cognate) aminoacylated tRNAs and code for the 20 standard amino-acids, a phenomenon also known as the 'degeneracy of genetic code'. Leaving aside methionine and tryptophan, the remaining 18 amino-acids are encoded by more than one codon. These different codons coding for the same amino-acids are termed as synonymous codons. Point mutations generating synonymous codons are silent because both the wild type and the mutants produce identical proteins. tRNAs ensure the correct interpretation of the genetic code to produce the desired polypeptide and prevent premature translation (Blanchet et al., 2018). However, due to the degeneracy of the genetic code and the wobble base pairing, all the 61 anticodons may not be required to form the desired polypeptide. In such cases, near-cognate tRNAs are used. However, in a lesser extent, translational errors also incorporate non-cognate aminoacylated tRNAs, affecting translational fidelity (Blanchet et al., 2018). The copy numbers of tRNAs complementary to each codon for each amino acid vary between species, which are well documented in GtRNAdb database (Chan and Lowe, 2015). According to this database, there are 275 *Saccharomyces cerevisiae* tRNA genes decoding the standard 20 amino acids, with a maximum of 16 tRNA genes for $tRNA^{Gly}_{GCC}$ and $tRNA^{Asp}_{GTC}$, and a minimum of zero tRNA genes. For the latter group, there is no cognate tRNA and near-cognate or non-cognate tRNAs should be preferred. This is proved to be an important factor hindering the translational efficiency and growth rate in bacteria (Du et al., 2017). Moreover, not all synonymous codons show equal occurrences in mRNAs. Although the genetic code remains conserved amongst organisms, codon usage preferences differ between organisms (Grantham et al., 1980; Ikemura, 1985; Chen et al., 2004). The variation in the codon usage pattern and the bias for any codon has two general classes of explanation, namely the selectionist and the neutralist perspectives. Support for the selection theory arises from the observation of a strong correlation between levels of gene expression and codon usage bias across large scale gene expression data (Duret, 2000; Castillo-Davis and Hartl, 2002; Ghaemmaghami et al., 2003; Goetz and Fuglsang, 2005). Codon usage bias also plays an essential role in the accuracy of translation (Akashi, 1994; Stoletzki and Eyre-Walker, 2006). Both the notions indicate that the highly expressed genes, as well as evolutionarily conserved and functionally important

Abbreviations: co-AI, co-adaptation index; mRNA, messenger RNA; tRNA, transfer RNA

* Corresponding authors at: Department of Microbiology, Bose Institute, Kolkata, West Bengal, India (D. Acharya) and Division of Bioinformatics, Bose Institute, Kolkata, West Bengal, India (T.C. Ghosh).

E-mail addresses: debarun@jcbose.ac.in (D. Acharya), tapash@jcbose.ac.in (T.C. Ghosh).

https://doi.org/10.1016/j.gene.2020.144640
Received 23 January 2020; Received in revised form 29 March 2020; Accepted 31 March 2020
Available online 01 April 2020

genes need to be heavily biased. In such genes with high codon usage bias the more frequently used codons are recognized by more abundant tRNA molecules (Ikemura, 1985). This would in-turn result in a heightened speed of translation elongation by reducing the tRNA screening time during the translation process (Curran and Yarus, 1989; Sørensen et al., 1989). Additionally, an abundant tRNA will rarely lead to incorrect incorporation of an amino-acid during the rapid translation process. The neutral theory, however, explains codon usage bias as a consequence of the mutational differences as some codons are more mutable and tend to have lower equilibrium frequencies. This theory very well explains the codon usage bias differences amongst organisms when compared for the GC content of the genes (Kanaya et al., 2001; Knight et al., 2001; Chen et al., 2004). The fact of a strong relationship between gene expression level and codon usage bias is contested by the differences in the mutational biases in genes transcribed at different levels (Francino and Ochman, 2001). Looking at the exon structure for codons in *Drosophila melanogaster* and *Caenorhabditis elegans* most of the optimal codons contain either cytosine or guanine at the third position (GC3) (Duret and Mouchiroud, 1999; Duret, 2002), making the above proposition not to be completely fit in the determination of the codon usage pattern. However, while studying their intron structure, abundance of cytosine and guanine isn't seen, raising queries that these stretches of genes should be affected by the same transcription coupled mutational process at synonymous sites (Duret and Mouchiroud, 1999; Duret, 2002). With both factors *i.e.* selection and mutation involved in the phenomenon of codon usage bias, the current accepted model is mutation-selection- drift balance model (Bulmer, 1991; Duret, 2002) which propounds the incorporation of major and minor codons due to mutational and selection pressure along with genetic drift.

To understand the molecular evolution, nucleotide substitution rate measures like the dN (non-synonymous substitution per nonsynonymous sites) and dS (synonymous substitution per synonymous sites) have been found to be effective in calculating the rate of evolution. These calculations are helpful in understanding the codon usage changes over time in organisms. Studies with *S. cerevisiae* showed gene expression as a dominant factor in determining the codon usage over evolutionary time-span (Drummond et al., 2005a). Protein expression can viewed from the perspectives of translational robustness and translational accuracy represented by dN and dS, which entails that if both are low then the accuracy with robustness is achieved, leading to the proper folding of a protein (Wilke and Drummond, 2006). Hence, the rate of change in dN will be highly constrained and incorporation of the non-synonymous codons is prevented in functionally important genes. Additionally, the abundance of desired aminoacylated tRNA molecules along with lower dS reduces the chance of recognition by near-cognate tRNA, resulting in a greater fidelity of translation.

In all of the previous studies the codon usage bias has been accounted cumulatively to mRNA abundance, its stability, protein abundance and its accuracy. It has also been shown that the existing codon composition is optimized for each gene for their required expression in *S. cerevisiae*, as the computationally generated mRNAs that are highly biased failed to increase the expression level (Victor et al., 2019), suggesting that only biased codon composition is insufficient to explain the scenario. In the present study we have tried to understand the influence of tRNA gene copy number on the mRNA abundance, protein abundance and translation rate. As the functionality of a protein depends on its proper folding besides its appropriate abundance (dosage), we tried to explore how the combined effect of codon usage bias and tRNA copy number influence protein folding to cope up with the elevated translation rate, using the co-adaptation index that represents the correlation between the proportion of each of the 61 codons in a gene and the tRNA gene copy number for that codon. From our study we found that codon usage bias mainly influences the synonymous substitution to adapt with the cognate tRNA pool which increases transcription efficiency by controlling the mRNA stability and local ribosome traffic and finally influences the protein abundance in highly expressed genes.

2. Results and discussion

2.1. Co-adaptation of codon and cognate tRNA influence protein abundance and mRNA stability

Generally, the codon-usage pattern is biased in almost all organisms, which controls the translation speed (Zhao et al., 2017) and ensures optimal expression level (Victor et al., 2019). The highly expressed genes show a more biased preference towards optimal codons to achieve their elevated expression level required for the cells (Du et al., 2017). However, analyzing *Saccharomyces cerevisiae* genes revealed that the codon composition of genes are optimized for the desired expression and using biased codons failed to elevate the latter (Victor et al., 2019), as to achieve this elevated translation speed, they must ensure ample amount of aminoacylated tRNAs and ribosomes to be present within the cell. However, GtRNAdb suggests that the tRNA genes for all 61 codons are not present in *Saccharomyces cerevisiae* genome (Chan and Lowe, 2015), while many of the tRNAs containing same anticodon have multiple gene copies or duplicates. The usage of codons complementary to such high copy number tRNA anticodons may increase translational efficiency. Thus, genes required to express at a higher level may show a preference towards such codons. In this study, we explored how tRNA gene pool is correlated with codon pool in *Saccharomyces cerevisiae*, and explored its underlying evolutionary significance. We correlated each tRNA copy number in *S. cerevisiae* for 61 amino acids with the frequency of each of the 61 codons to obtain the tRNA co-adaptation Index (co-AI) for each gene. We observed that the co-adaptation index of genes is positively correlated with mRNA abundance (Spearman $\rho = 0.262$, $P = 3.89 \times 10^{-81}$, $N = 5135$) and protein abundance (Spearman $\rho = 0.306$, $P = 1.59 \times 10^{-111}$, $N = 5135$). The result indicates that the highly expressed genes use cognate tRNA adapted codons in higher frequency and the protein abundance depends on the copy number of cognate tRNAs. As mRNA abundance is also correlated to protein abundance, we also did a partial correlation between co-AI and protein abundance controlling for mRNA abundance and observed a significant positive correlation ($\rho = 0.107$, $P = 1.56 \times 10^{-14}$, $N = 5135$), indicating tRNA co-adaptation influences protein abundance independent of mRNA abundance. We further divided the data of co-AI into five classes with high (Category A) to low (Category E) co-AI (see Materials and methods) and observed that the protein abundance decreases gradually from category A to E ($P = 3.66 \times 10^{-116}$, $N = 5135$, Kruskal-Wallis test) (Table 1). It suggests that co-AI strongly affects the translational efficiency because tRNA co-adaptation helps in optimizing the translation rate for highly expressed genes, as they require lower search time for the tRNA during translation (Ikemura, 1985; Duret, 2000). To strengthen this claim further, we divided all *S. cerevisiae* genes into two groups (Low ENC and High ENC) based on their effective number of codons (ENC), where a low ENC value represents high codon usage bias (see Materials and methods). We found that within the same ENC group, the protein abundances decrease from co-AI category A to E ($P_{Low\_ENC} = 4.66 \times 10^{-9}$, $N_{Low\_ENC} = 409$; $P_{High\_ENC} = 5.93 \times 10^{-47}$, $N_{High\_ENC} = 4726$, Kruskal-Wallis test) (Fig. 1). This proves that although genes with high codon usage bias are highly expressed, codon usage bias alone is not sufficient to increase the protein abundance; rather the tRNA copy number coalesces with it to elevate the protein abundance.

An mRNA molecule which is highly stable houses a greater number of ribosomes during the translation process and results in a higher efficiency under the presence of abundant cognate-tRNA molecules (Mao et al., 2014). We also observed a negative correlation between co-AI and mRNA folding free energy (Spearman $\rho = -0.242$, $P = 2.39 \times 10^{-69}$, $N = 5135$), indicating mRNA secondary structure stability increases with co-AI, as lower mRNA folding free energy

Table 1
Comparison of genomic, transcriptomic and proteomic features of genes in different co-adaptation index categories.

Correlation Group	Number of genes	Protein abundance	mRNA Folding free energy	dN	dS	dN/dS	Half life	Proteins/sec	%chaperone-assisted proteins
Category A (co-AI > 0.6)	179	1580.123	−6.557	0.051	0.795	0.080	36.136	1.765	84.127
Category B (0.60 ≥ co-AI > 0.5)	610	524.157	−6.216	0.071	1.047	0.076	26.245	0.825	78.802
Category C (0.5 ≥ co-AI > 0.4)	1427	146.918	−5.703	0.095	1.256	0.083	16.890	0.239	67.946
Category D (0.4 ≥ co-AI > 0.3)	1745	60.797	−5.552	0.108	1.312	0.090	13.241	0.142	62.599
Category E (co-AI ≤ 0.3)	1174	45.471	−5.519	0.115	1.264	0.099	11.543	0.169	59.149

indicates higher stability in mRNA secondary structure. When we compared the mRNA folding free energy in the five categories, we observed that it becomes less negative from category A to E ($P = 1.50 \times 10^{-90}$, N = 5135), indicating mRNAs of high co-AI genes are structurally more stable than the low co-AI groups (Table 1).

2.2. Co-adaptation of codon and cognate tRNA influence synonymous substitution

In general, the codon usage pattern has seen to be determined by an equilibrium among mutation, genetic drift, and natural selection (Akashi and Eyre-Walker, 1998; Akashi, 2001). In a number of cases, natural selection influences the codon usage bias and gene expression (mRNA abundance and protein abundance) (Duret, 2002). This is prevalent in the unicellular organisms where gene expression level plays a major role in codon usage determination (Drummond et al., 2006). The highly expressed genes tend to show a lower evolutionary change (Drummond et al., 2005b). In this study, we observed that along with gene expression and protein abundance, co-adaptation index (co-AI) also shows a negative correlation with protein evolutionary rates (dN/dS) (Spearman $\rho = -0.136$, $P < 1.0 \times 10^{-6}$, N = 4176). When considering the different categories, we observed that average dN, dS and dN/dS increases from Category A to E ($P_{dN} = 2.56 \times 10^{-60}$, $P_{dS} = 1.69 \times 10^{-39}$, $P_{dN/dS} = 8.87 \times 10^{-16}$, N = 4091, Kruskal-Wallis test) (Fig. 2, Table 1), suggesting that co-AI may also control nucleotide substitution rates and the genes adapted to the tRNA tend to evolve slower. However, this co-adaptation should have more impact on synonymous substitutions. Moreover, it is a well established fact that the protein evolution is mainly guided by gene expression (Drummond et al., 2006). Therefore, in category A of our dataset, the reduced evolutionary rate may result from their higher expression than other groups. Also, when synonymous and nonsynonymous substitutions are considered separately, codon usage bias reduces the rate of nucleotide substitution at synonymous site (Sharp and Li, 1987). To compare the effect of co-AI and gene expression on nucleotide substitution, we performed a principal component regression analysis in all the components dN, dS and dN/dS taking tRNA correlation (co-AI), gene expression and protein abundance as variables. We found that the variation in dN and dN/dS is mainly influenced by gene expression and protein abundance, but the variation in dS is strongly influenced by tRNA correlation (Fig. 3). This observation is also consistent in all subgroups.

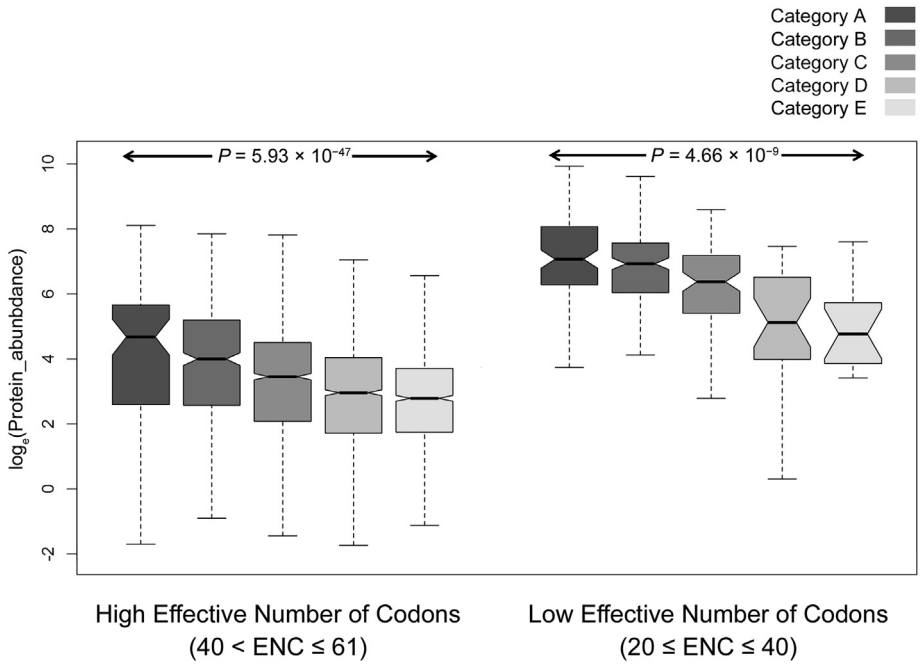

Fig. 1. Boxplot representing the log protein abundance values of each co-AI categories within high- and low ENC (effective number of codons) group of *Saccharomyces cerevisiae*genes.

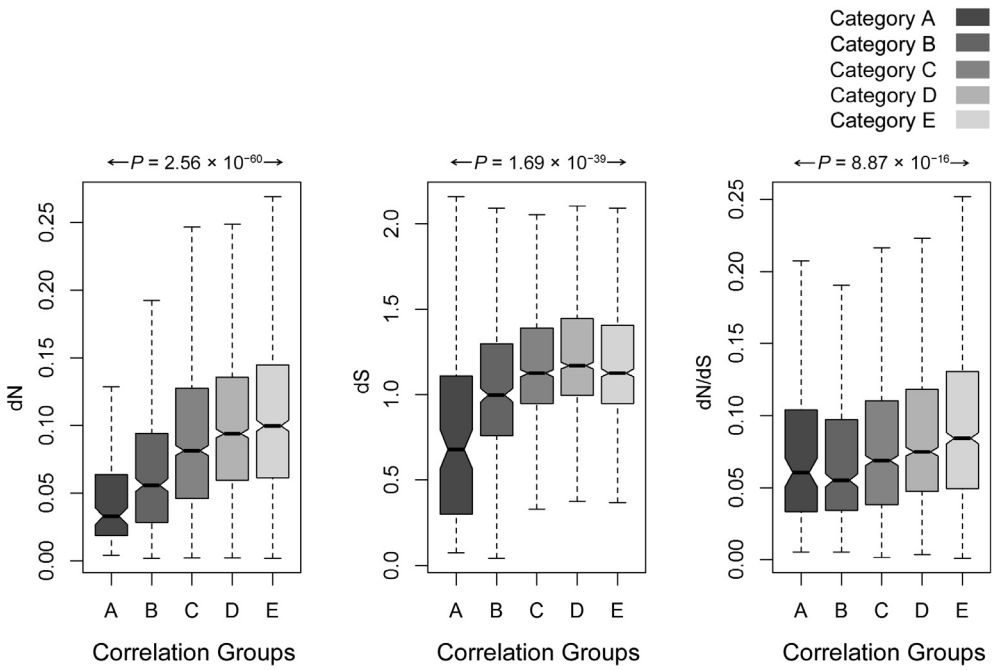

Fig. 2. The nucleotide substitution rates of *Saccharomyces cerevisiae* genes within different co-AI categories, using *Saccharomyces bayanus* 1:1 orthologs.

2.3. Folding of proteins is essential in addition to their appropriate abundance

We observed positive correlation between co-adaptation index (co-AI) and translation rate (protein/sec) (Spearman $\rho = 0.158$, $P = 3.31 \times 10^{-28}$, N = 4968). The evidence became more significant when we compared the category A with the category E of co-adapted genes [Average transcription $rate_{CategoryA} = 1.765$ proteins/sec, $N_{CategoryA} = 167$, Average transcription $rate_{CategoryE} = 0.169$ proteins/sec, $N_{CategoryE} = 1100$, Mann-Whitney U Test, $P = 5.74 \times 10^{-28}$]. The translation rate is more than ten times higher in highly co-adapted genes of category A. This is also consistent with a previous study which suggested that the rate of translation elongation is higher for genes utilizing optimal or preferred codons (Yu et al., 2015). Our results clearly suggest that the tRNA pool is oriented in a way to facilitate maximum expression of the genes for which the cellular demand is high. Similarly, an appropriate abundance of protein is essential to keep the cells functioning properly (Torrent et al., 2018). However, proteins must be folded to their functional conformation to perform their cellular functions. A misfolded protein is either non-functional, that increases the energy cost of the cells to refold it or is toxic to the cells which can lead to its malfunctioning (Chen et al., 2011). Genes with high co-AI exhibit very high level of mRNA abundance and protein abundance. The genes in the category A also have a longer mRNA half-life which gradually decreases in rest of the categories ($P = 1.05 \times 10^{-114}$, Kruskal-Wallis test) (Table 1). Now, with the appropriate mRNA abundance and longer mRNA half life leads to increased protein abundance, but if the proteins remains unfolded or gets misfolded then the system needs to do away with it. Therefore, with faster translation rate, faster folding should be a necessary requirement for highly expressed genes. However, protein folding may be intrinsic (depending on the proteins' sequence) or chaperone-assisted, depending on the folding ability inherently associated with the protein molecules due to their amino-acid composition (Raineri et al., 2010).

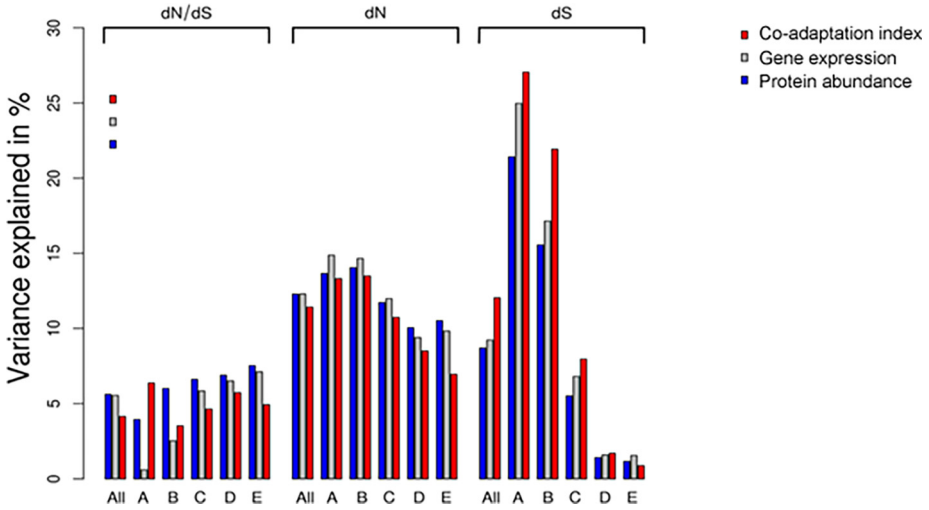

Fig. 3. Principal component regression analysis to analyze the effect of three principal components (co-adaptation index, mRNA abundance and protein abundance) on nonsynonymous (dN), synonymous (dS) nucleotide substitutions and evolutionary rate (dN/dS).

The proportion of chaperone-assisted proteins was found to be higher in category A, which gradually decreases in other categories with decreasing co-AI (Table 1). This suggests that the chaperones play an important role to meet the increased requirement of faster folding in genes with high co-AI, which are translationally more efficient. Together, these results suggest that co-AI is an important factor controlling mRNA structural stability, and genes with high co-AI are highly expressed as they have higher cognate tRNA abundance. These genes are well adapted to the existing tRNA pool, and must be evolutionarily more conserved, to make full use of it. The protein sequence of the high co-AI groups revealed that they have longer mRNA half-lives ($P = 1.05 \times 10^{-114}$, Kruskal-Wallis test), besides a high translation rate that may make them more vulnerable to aggregation. Therefore they are more often assisted by chaperones than the low co-AI groups, to ensure rapid folding and to prevent protein accumulation.

3. Conclusion

In this study, we explored the influence of tRNAs and codon composition on the protein abundance, using *Saccharomyces cerevisiae* as a model organism. We observed that the correlation of tRNA copy number and codon usage patterns for individual genes (co-AI) have a wide range of distribution, where the group of genes having high co-AI values shows a higher expression. We speculate that this is due to the availability of ample amount of desired tRNAs during translation for genes with high co-AI. For the genes that show a lower co-AI also produce mRNAs with lower folding stability that can finally affect the protein abundance. When we examined the nucleotide substitution patterns of genes, we observed low synonymous (dS) and nonsynonymous (dN) nucleotide substitutions for genes with high co-AI. A principal component regression analysis suggests that although the gene expression is the major controlling factor of dN, the dS is influenced more by co-AI values. From this study it can be said that the combined influence of codon composition of a gene and the copy number of different tRNAs facilitates faster recognition of codons by the tRNA molecules, resulting in more efficient translation elongation due to less ribosomal delay leading to elevated protein abundance. Such high protein abundance may lead to misfolding or protein accumulation, which can be prevented by the higher chaperone interaction of proteins encoded by genes with high co-AI.

4. Materials and methods

4.1. Data collection

We downloaded the *Saccharomyces cerevisiae* ORFs from SGD (https://www.yeastgenome.org/) (Cherry et al., 2012). All erroneous sequences were removed by CodonW. The mRNA molecules per cell were collected from genome-wide expression analysis data by Holstege et al. (Holstege et al., 1998) and the protein abundance data were collected from Arava et al. (Arava et al., 2003). Proteins with no expression level information or zero expression level were omitted from study. We considered only the longest transcript as a representative of that particular gene when more than one transcript is available for each gene. Finally, we obtained the data for 5135 *S. cerevisiae* genes with protein abundance data, which we used for our analyses. We downloaded *S. cerevisiae* protein-protein interactions from BioGRID database (version 3.4.152) (Chatr-Aryamontri et al., 2017). The chaperone proteins were obtained from Gong et al. (Gong et al., 2009). Proteins interacting with at least one chaperone protein were considered as chaperone-client and those without any known interaction with any of the chaperone proteins were considered as non-client proteins of chaperone.

4.2. Calculation of co-adaptation index (co-AI) and effective number of codons (ENC)

The copy numbers of *Saccharomyces cerevisiae* tRNA molecules containing anticodons for each of the 61 codons were obtained from GtRNAdb 2.0 release 17 (Chan and Lowe, 2015). The frequency of each of the 61 codons for each *S. cerevisiae* gene were computed using EMBOSS (Rice et al., 2000). For finding out the co-AI we calculated the Pearson correlation coefficient between the frequencies of the 61 codons for each individual gene with the tRNA gene copy number for that codon in *S. cerevisiae*. Then we divided the *S. cerevisiae* genes into five categories in the decreasing order of this correlation: Category A (co-AI $>$ 0.6), Category B (0.60 $\geq$ co-AI $>$ 0.5), Category C (0.5 $\geq$ co-AI $>$ 0.4), Category D (0.4 $\geq$ co-AI $>$ 0.3) and Category E (co-AI $\leq$ 0.3) with decreasing order of co-AI. The effective number of codons (ENC, ranges 20 $\leq$ ENC $\leq$ 61) were calculated using CodonW (Peden, 1999), where a lower ENC corresponds to higher codon usage bias. We divided our data in two groups based on ENC values: Low ENC (20 $\leq$ ENC $\leq$ 40) and High ENC (40 $<$ ENC $\leq$ 61).

4.3. Calculation of evolutionary rate

For the calculation of evolutionary rate, we identified 1:1 orthologous protein pairs of *Saccharomyces cerevisiae* and *S. bayanus* using reciprocal protein BLAST. We allowed similarity $\geq$ 80%, overlap $\geq$ 80%, e-value $<$ 10^{-5} and gap $<$ 5% as the cut off and identified 4503 orthologous pairs. We calculated the synonymous substitution rate (dS), non-synonymous substitution rate (dN) and dN/dS by Yang-Nielsen method using PAML package (Yang, 2007) and in-house PERL script. We controlled the mutation saturation by discarding all dS values $\geq$ 3 (Acharya and Ghosh, 2016).

4.4. Folding free energy and mRNA half life

The folding free energy for NATIVE and all the sequences were calculated using the RNAfold attribute from Vienna software package2 (Lorenz et al., 2011). A window of 40 nucleotides and a step size of 1 nucleotide is used to calculate the folding free energy following Park et al. (Park et al., 2013). The predicted local folding free energies were averaged to obtain the folding free energy for the entire mRNA sequence.

4.5. Statistical analyses

The statistical calculations were carried out using IBM SPSS 22. All the correlations reported in this study are Spearman correlation, unless otherwise stated. In all statistical analyses, we used 95% level of confidence as a measure of significance.

CRediT authorship contribution statement

MPV, DA and SC conceptualized the work. MPV and SC curated and analysed the data. MPV, DA and TCG wrote the manuscript. All authors read and approved the final version of the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2020.144640.

References

Acharya, D., Ghosh, T.C., 2016. Global analysis of human duplicated genes reveals the relative importance of whole-genome duplicates originated in the early vertebrate evolution. BMC Genom. 17, 1–14.

Akashi, H., 1994. Synonymous codon usage in Drosophila melanogaster – natural-selection and translational accuracy. Genetics 136, 927–935.

Akashi, H., 2001. Gene expression and molecular evolution. Curr. Opin. Genet. Dev. 11, 660–666.

Akashi, H., Eyre-Walker, A., 1998. Translational selection and molecular evolution. Curr. Opin. Genet. Dev. 8, 688–693.

Arava, Y., Wang, Y., Storey, J.D., Liu, C.L., Brown, P.O., Herschlag, D., 2003. Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U.S.A 100, 3889–3894.

Blanchet, S., Cornu, D., Hatin, I., Grosjean, H., Bertin, P., Namy, O., 2018. Deciphering the reading of the genetic code by near-cognate tRNA. Proc. Natl. Acad. Sci. 115, 3018–3023.

Bulmer, M., 1991. The selection-mutation-drift theory of synonymous codon usage. Genetics 129, 897–907.

Castillo-Davis, C.I., Hartl, D.L., 2002. Genome evolution and developmental constraint in Caenorhabditis elegans. Mol. Biol. Evol. 19, 728–735.

Chan, P.P., Lowe, T.M., 2015. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. Nucleic Acids Res. 44, D184–D189.

Chatr-Aryamontri, A., Oughtred, R., Boucher, L., Rust, J., Chang, C., Kolas, N.K., O'Donnell, L., Oster, S., Theesfeld, C., Sellam, A., Stark, C., Breitkreutz, B.J., Dolinski, K., Tyers, M., 2017. The BioGRID interaction database: 2017 update. Nucleic Acids Res. 45, D369–D379.

Chen, B., Retzlaff, M., Roos, T., Frydman, J., 2011. Cellular strategies of protein quality control. Cold Spring Harbor Perspect. Biol. 3, a004374.

Chen, S.L., Lee, W., Hottes, A.K., Shapiro, L., McAdams, H.H., 2004. Codon usage between genomes is constrained by genome-wide mutational processes. Proc. Natl. Acad. Sci. U.S.A 101, 3480–3485.

Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., Fisk, D.G., Hirschman, J.E., Hitz, B.C., Karra, K., Krieger, C.J., Miyasato, S.R., Nash, R.S., Park, J., Skrzypek, M.S., Simison, M., Weng, S., Wong, E.D., 2012. Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res. 40, D700–D705.

Curran, J.F., Yarus, M., 1989. Rates of aminoacyl-tRNA selection at 29 sense codons in vivo. J. Mol. Biol. 209, 65–77.

Drummond, D.A., Bloom, J.D., Adami, C., Wilke, C.O., Arnold, F.H., 2005. Why highly expressed proteins evolve slowly. Proc. Natl. Acad. Sci. 102, 14338–14343.

Drummond, D.A., Raval, A., Wilke, C.O., 2006. A single determinant dominates the rate of yeast protein evolution. Mol. Biol. Evol. 23, 327–337.

Du, M.-Z., Wei, W., Qin, L., Liu, S., Zhang, A.-Y., Zhang, Y., Zhou, H., Guo, F.-B., 2017. Co-adaption of tRNA gene copy number and amino acid usage influences translation rates in three life domains. DNA Res. 24, 623–633.

Duret, L., 2000. tRNA gene number and codon usage in the C. elegans genome are co-adapted for optimal translation of highly expressed genes. Trends Genet. 16, 287–289.

Duret, L., 2002. Evolution of synonymous codon usage in metazoans. Curr. Opin. Genet. Dev. 12, 640–649.

Duret, L., Mouchiroud, D., 1999. Expression pattern and surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and Arabidopsis. Proc. Natl. Acad. Sci. 96, 4482–4487.

Francino, M.P., Ochman, H., 2001. Deamination as the basis of strand-asymmetric evolution in transcribed Escherichia coli sequences. Mol. Biol. Evol. 18, 1147–1150.

Ghaemmaghami, S., Huh, W., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., Weissman, J.S., 2003. Global analysis of protein expression in yeast. Nature 425, 737–741.

Goetz, R.M., Fuglsang, A., 2005. Correlation of codon bias measures with mRNA levels: analysis of transcriptome data from Escherichia coli. Biochem. Biophys. Res. Commun. 327, 4–7.

Gong, Y., Kakihara, Y., Krogan, N., Greenblatt, J., Emili, A., Zhang, Z., Houry, W.A., 2009. An atlas of chaperone–protein interactions in Saccharomyces cerevisiae: implications to protein folding pathways in the cell. Mol. Syst. Biol. 5.

Grantham, R., Gautier, C., Gouy, M., Mercier, R., Pave, A., 1980. Codon catalog usage and the genome hypothesis. Nucl. Acids Res. 8 197-197.

Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., Young, R.A., 1998. Dissecting the regulatory circuitry of a eukaryotic genome. Cell 95, 717–728.

Ikemura, T., 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. 2, 13–34.

Kanaya, S., Kinouchi, M., Abe, T., Kudo, Y., Yamada, Y., Nishi, T., Mori, H., Ikemura, T., 2001. Analysis of codon usage diversity of bacterial genes with a self-organizing map (SOM): characterization of horizontally transferred genes with emphasis on the E. coli O157 genome. Gene 276, 89–99.

Knight, R.D., Freeland, S.J., Landweber, L.F., 2001. A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. Genome Biol. 2 (research0010), 1.

Lorenz, R., Bernhart, S.H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., Hofacker, I.L., 2011. ViennaRNA Package 2.0. Algorithms Mol Biol 6, 26.

Mao, Y., Liu, H., Liu, Y., Tao, S., 2014. Deciphering the rules by which dynamics of mRNA secondary structure affect translation efficiency in Saccharomyces cerevisiae. Nucleic Acids Res. 42, 4813–4822.

Park, C., Chen, X., Yang, J.R., Zhang, J., 2013. Differential requirements for mRNA folding partially explain why highly expressed proteins evolve slowly. Proc. Natl. Acad. Sci. U.S.A. 110, E678–E686.

Peden, J.F., 1999. Analysis of Codon Usage, Dept of Genetics. University of Nottingham.

Raineri, E., Ribeca, P., Serrano, L., Maier, T., 2010. A more precise characterization of chaperonin substrates. Bioinformatics 26, 1685–1689.

Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the European molecular biology open software suite. Elsevier current trends.

Sharp, P.M., Li, W.H., 1987. The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Mol. Biol. Evol. 4, 222–230.

Sørensen, M.A., Kurland, C., Pedersen, S., 1989. Codon usage determines translation rate in Escherichia coli. J. Mol. Biol. 207, 365–377.

Stoletzki, N., Eyre-Walker, A., 2006. Synonymous codon usage in Escherichia coli: selection for translational accuracy. Mol. Biol. Evol. 24, 374–381.

Torrent, M., Chalancon, G., de Groot, N.S., Wuster, A., Babu, M.M., 2018. Cells alter their tRNA abundance to selectively regulate protein synthesis during stress conditions. Sci. Signal. 11, eaat6409.

Victor, M.P., Acharya, D., Begum, T., Ghosh, T.C., 2019. The optimization of mRNA expression level by its intrinsic properties—insights from codon usage pattern and structural stability of mRNA. Genomics.

Wilke, C.O., Drummond, D.A., 2006. Population genetics of translational robustness. Genetics 173, 473–481.

Yang, Z., 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24.

Yu, C.-H., Dang, Y., Zhou, Z., Wu, C., Zhao, F., Sachs, M.S., Liu, Y., 2015. Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. Mol. Cell 59, 744–754.

Zhao, F., Yu, C.-H., Liu, Y., 2017. Codon usage regulates protein structure and function by affecting translation elongation speed in Drosophila cells. Nucleic Acids Res. 45, 8484–8492.