

Original Research Article

<https://doi.org/10.20546/ijcmas.2020.911.234>

Synonymous Codon Usage Bias Factors Affecting Chloroplast Genome of Grape Wine *Vitis vinifera*

Farshad Talat¹ and S.S. Udikeri^{2*}

¹Seed and Plant Improvement Research Department, West Azarbaijan Agricultural and Natural Resources Research and Education Center, AREEO, Urmia, Iran

²University of Agricultural Sciences, Dharwad, India

*Corresponding author

ABSTRACT

In order to optimize protein production in expression systems, analysis of codon usage patterns is very important. In this regard, chloroplast has special importance due to its small size and high copy number of genome. *Vitis vinifera* is the most important fruit species in the modern world. In this research, the complete nucleotide sequence of the chloroplast genome in *Vitis vinifera* was studied, and then analyzed using CodonW software. Correspondence analysis and method of effective number of codon as Nc-plot were conducted to analyze synonymous codon usage. According to the corresponding analysis, codon bias in the chloroplast genome of *Vitis vinifera* is related to their gene length, mutation bias, gene hydropathy level of each protein, gene function and selection or gene expression only subtly affect codon usage. This study provides insights into the molecular evolution studies.

Keywords

Vitis, Chloroplast, Codon

Article Info

Accepted:

15 October 2020

Available Online:

10 November 2020

Introduction

The genetic code uses 64 codons to represent the 20 standard amino acids and the translation termination signal. Each codon is recognized by a subset of a cell's transfer ribonucleotide acid molecules (tRNAs), and with the exception of a few codons that have been reassigned in some lineages (Osawa and Jukes, 1989; Osawa *et al.*, 1990) the genetic

code is remarkably conserved, although it is still in a state of evolution (Osawa *et al.*, 1992).

Generally, the alternative synonymous codons for any amino acid are not randomly used (Ghosh, 2000; Grantahn *et al.*, 1980). Studies of the synonymous codon usage reveal information on molecular evolution of individual genes, which provides data to

improve gene recognition algorithms, are utilized to design DNA primers, and detects horizontal transfer events (Fickett, 1982; Peng *et al.*, 2007).

Several factors such as directional mutational bias (Onofio and Bernardi, 1992; Gupta and Ghosh 2001), translational selection (Hou and Yang 2003), secondary structure of proteins (Gu *et al.*, 2004; Sharp and Li, 1987), replicational and transcriptional selection, gene length (Ranjan *et al.*, 2007; Gupta *et al.*, 2004), gene expression level (Carlini *et al.*, 2001), gene density (Kahali *et al.*, 2007; Versteeg *et al.*, 2004), and environmental factors (Lynn *et al.*, 2002) have been reported to influence the codon usage in various organisms. Codon usage variation is represented by two major paradigms. Codon usage is determined by either mutational bias or natural selection (.The unified theory for codon usage has not been provided so far (Chen *et al.*, 2011).

The chloroplast genome has long been a focus of research in plant molecular evolution due to its small size, high copy number, conservation and extensive characterization at the molecular level. More recently, with technical advances in DNA sequencing, the number of completely sequenced chloroplast genomes has grown rapidly ([http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid = 2759&opt=plastid#pageTop](http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=2759&opt=plastid#pageTop)). Since 2006, more than twenty chloroplast genomes have been sequenced every year. The translation machinery in chloroplasts is known to be structurally similar to those in prokaryotes, leading to the suggestion that the translation mechanism and patterns of codon usage in chloroplasts might be similar to those in *Escherichia coli* (Sugiura, 1992). However, most studies in plants were mainly focused on codon bias in nuclear genomes. Nevertheless, Morton found that the asymmetry of two

strands of the chloroplast genome from *Euglena gracilis* was the major factor contributing to codon bias (Morton, 1999). In addition, it was considered that context dependent mutation played some roles in shaping codon usage of the chloroplast genomes of grass species (Morton 2003), although there was also evidence that the codon usage of certain chloroplast genes was influenced by selection (Morton, 1998).

The primary cultivated grape species, *Vitis vinifera*, is one of around 60 *Vitis* species. *Vitis vinifera* is considered to be one of the major fruit crops in the world based on hectares cultivated and economic value.

The complete chloroplast genome sequence of *Vitis vinifera* has been determined (Jansen *et al.*, 2006). Therefore it is of interest to understand the factors that shape codon usage in this species. In this study, the analysis of codon usage bias in chloroplast genome of *Vitis vinifera* is reported using methods of multivariate statistical analysis and correlation analysis are determined.

Materials and Methods

Sequence data

The complete chloroplast genome sequence from *Vitis vinifera* (NC_007957.1) was obtained from Gene Bank of the National Center for Biotechnology Information (NCBI). Genes having more than 100 codons and without any intermediate stop codon were selected for the current study. A total of genes were combined for codon usage analysis.

Analysis

Numbers of codons (NCs), Relative synonymous Codon Usage (RSCU), and GC composition of codons were calculated for each gene. The analysis was carried out by

CODONW 1.4 (<http://codonw.sourceforge.net/>).

Nc, the “effective number of codons” used in a gene measures the bias away from equal usage of codons within synonymous groups (26). Nc can take values from 20 to 61, when only one codon or all synonyms in equal frequencies were used per amino acid, respectively. Nc appears to be a good measure of general codon usage bias (Wright, 1990; Cameron and Aguad’e, 1998). The sequences in which Nc values are <30 are highly expressed while those with >55 are poorly expressed genes (Sharp and Cowe, 1991).

Relative Synonymous Codon Usage (RSCU), is defined as the ratio of the observed frequency of codons to the expected frequency if all the synonymous codons for those amino acids are used equally (Sharp and Li, 1987). RSCU is used to observe the synonymous codon usage variation among the genes.

The Codon Adaptation Index (CAI) was used to estimate the extent of bias toward codons that were known to be preferred in highly expressed genes. A CAI value is between 0 and 1.0, and a higher value means a stronger codon bias and a higher expression level. G+C value is the frequency of nucleotides that are guanine or cytosine.

Hydropathicity value is the General Average Hydropathicity or GRAVY score, for the hypothetical translated gene product. It is calculated as the arithmetic mean of the sum of the hydropathic indices of each amino acid. Length value is equivalent to the length of one gene.

NC-plot. Wright (30) suggested the NC-plot (NC plotted against GC3s) as part of a general strategy to investigate the patterns of synonymous codon usage. Genes whose

codon choice is constrained only by a G+C mutation bias will lie on or just below the curve of the predicted values (Wright, 1990).

COA analysis

Correspondence analysis (COA) has become the method of choice for multivariate statistical analysis of codon usage (Maria 2001).

Results and Discussion

The size of the *Vitis vinifera* chloroplast genome is 160,928 bp. The overall GC content of the chloroplast genome is 37.40%. Coding regions make up 57.08% of the chloroplast genome (49.94% protein-coding genes, 6.34% RNA genes) and non-coding regions, which contain intergenic spacer (IGS) regions and introns, comprising 43.28%. Among the full 131 coding genes of the *Vitis vinifera* chloroplast genome, we identified 79, 30 and 4 protein-coding, transfer RNA, and ribosomal RNA genes, respectively.

First of all we observe the Nc-plot distribution, which ENC and GC3s values were calculated (Fig. 1). ENC values vary from 35.72 to 59.36 with a mean of 50.22 and standard deviation of 5.56. The heterogeneity of codon usage was further confirmed from the GC3s values ranging from 17% to 58% with a mean of 41% and standard deviation of 10%. Wright suggested that a plot of NC versus GC3s could be used effectively to explore the codon usage variation among the genes (Duret 2000). He argued that the comparison of actual distribution of genes, with the expected distribution under no selection could be indicative if codon usage bias of genes have some other influences other than compositional constraints. If the codon usage bias is completely dictated by GC3s the values of NC should fall on the

expected curve between GC3s and NC-plot of the *Vitis vinifera* chloroplast genome shown in Figure 1. NC values which lie below the expected curve, indicating that these genes have additional codon usage bias apart from compositional bias (Fig. 1).

According to *Vitis vinifera* chloroplast gene function, 54 sequences can be classified six categories: The number of the first classified gene is 13, which encoding ribosomal protein. The *rpl* and *rps* genes encode large and small subunit ribosomal protein; the number of the second classified genes is 15, including of *psa* gene, *psb* gene, *atp* gene, *pet* gene and *rbcL* gene; The third category is a conservative gene: *yfc*. The fourth category is translation apparatus genes, including the *rpo* gene of the

RNA polymerase gene family. The fifth type is the miscellaneous proteins gene, for example *accD*; the sixth category is unknown function and hypothetical protein gene.

Figure 1 shows the diversity among genes in terms of RSCU. In the leftmost of first axis, genes with the greatest codon bias are located, and those with lowest codon bias are found in the rightmost of the axis. Figure 2, shows the plots of ENC values towards GC3 values; given that none of these points have been located on the curve, consequently codon selection of them are not limited to mutation bias of GC3. On the other hand, the codon usage of the points below the curve is dependent from compositional constraint.

Figure.1 Effective number of codons (NC) used in each gene plotted against GC content at synonymously variable third position of codons (GC3s).The continuous curve plots the relationship between NC and GC3s in the absence of selection.

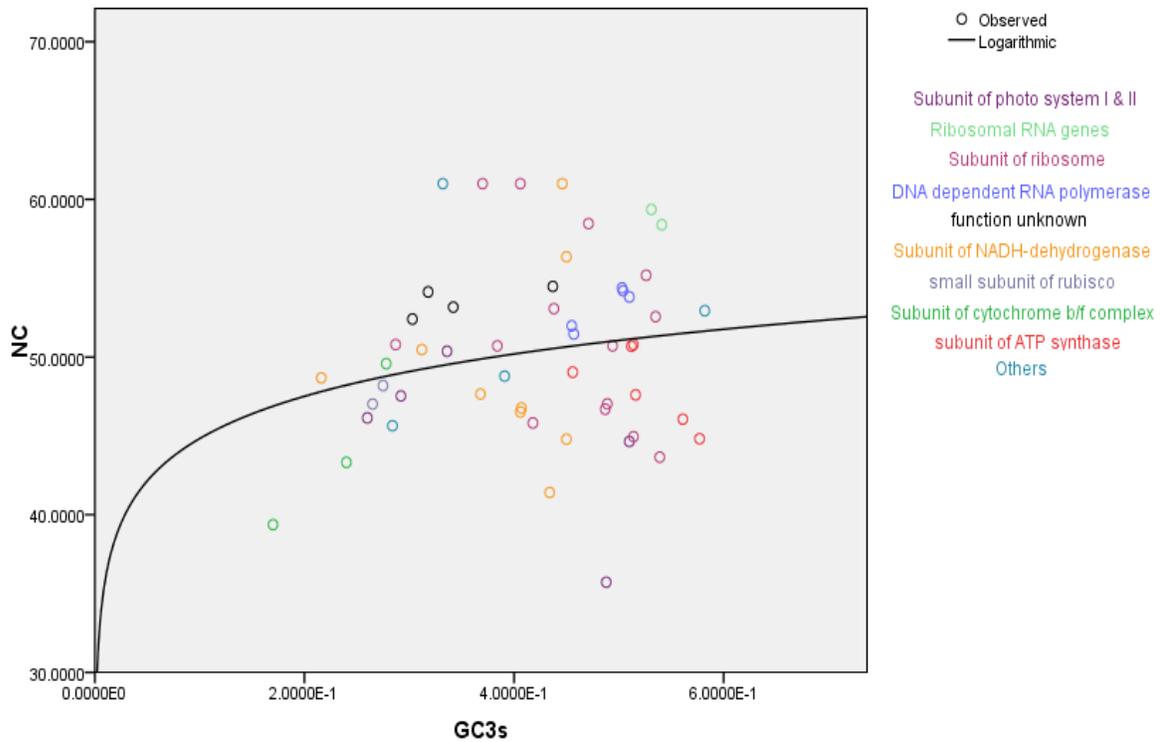
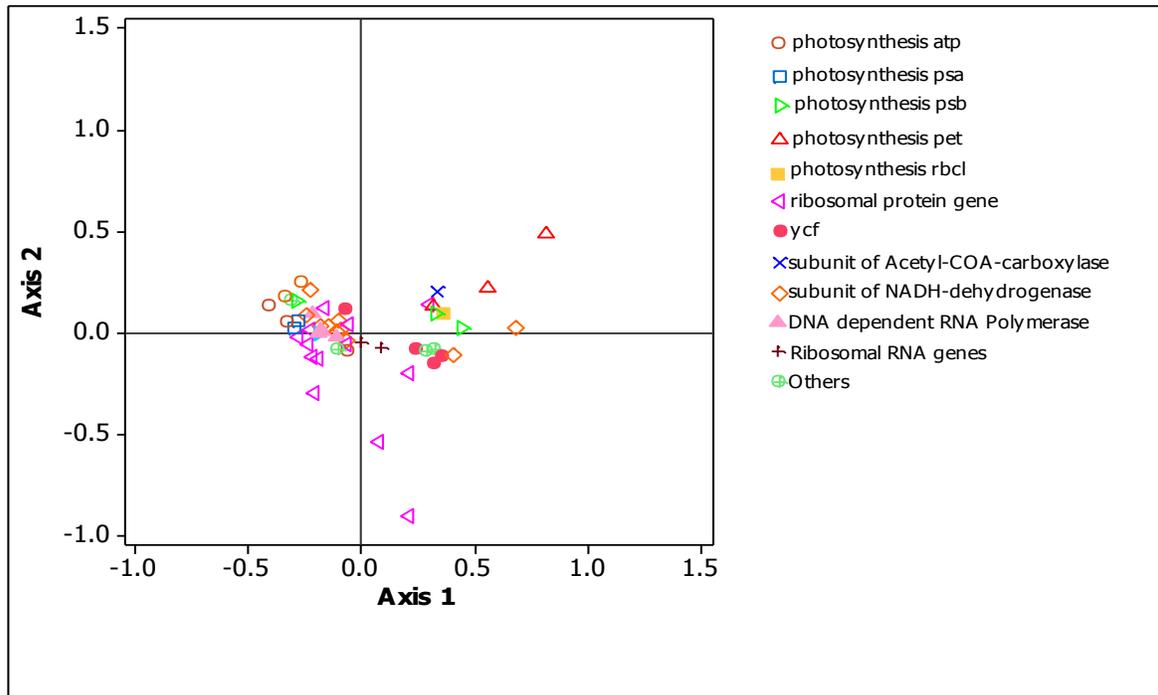


Figure.2 Correspondence analysis of the relative synonymous codon usage in 54 genes from chloroplast genome of *Vitis vinifera*



It is thought that optimal codons help to achieve faster translation rates and higher accuracy. As a result of these factors, translational selection is expected to be stronger in highly expressed genes. Several earlier discussions of plant codon usage focused on the differences between codon choice in plant nuclear genes and in chloroplasts (Mario *et al.*, 2004). Chloroplasts differ from the nuclear genome of higher plants in that they encode only 30 tRNA species. Since chloroplasts have restricted their tRNA genes, the use of preferred codons by chloroplast encoded proteins appears more extreme. However, a positive correlation has been reported between the level of isoaccepting tRNA for a given amino acid and the frequency with which this codon is used in the chloroplast genome (Bulmer, 1991). The codon usage patterns of chloroplast genes more conserved in GC content and influenced by translation level. Maybe chloroplast and nuclear genes shared particularly different

features of codon usage and evolutionary constraints. In the future, further research on a comparative analysis of codon bias and factors in shaping the codon usage patterns among mitochondrion, chloroplast and nuclear genes may help clarify the relationship between nuclear genomes and mitochondria, chloroplast genome in *Vitis vinifera*. This perhaps will provide detailed research about endosymbiont hypothesis and a framework from which to build more robust models to improve our understanding of not only molecular evolution in general, but also how we interpret molecular data for reconstructing phylogenies.

In this study, we presented evidence suggesting that codon bias in the chloroplast genome of *Vitis vinifera* is closely related to their gene length, mutation bias, gene hydrophathy level of each protein, gene function and selection or gene expression. Why it is related with gene length? We argue

that it is consequence of selection to maximize translational speed, minimize the costs of proofreading, or maximize the accuracy of translation; by using codons which match common tRNAs or which bind the tRNA efficiently, it is thought that the time to find and bind the correct tRNA is minimized, along with the probability of misincorporating an incorrect tRNA. Both missense and processivity errors can potentially lead to a positive correlation between synonymous codon bias and gene length. This study has provided a basic understanding of the mechanisms for codon usage bias, which could be useful in further studies of their molecular evolution, cloning and heterologous expression of its chloroplast genetic engineering.

References

- Bulmer, M. The selection-mutation-drift theory of synonymous codon usage. *Genetics*. 1991. 129. 897–907.
- Carlini, D. B., Chen, Y., Stephan, W. The relationship between third-codon position nucleotide content, codon bias, mRNA secondary structure and gene expression in the drosophilid alcohol dehydrogenase genes Adh and Adhr. *Genetics*. 2001. 159: 623–33.
- Chen, X., Xiaoning, C., Quanzhan, C., Hongxia, Z., Yao, C., Ailing, B., Factors affecting synonymous codon usage bias in chloroplast genome of *Oncidium Gower Ramsey*. *J. Evol. Bioinform.*. 2011. 7: 271-278.
- Comeron, J, M., and Aguadé, M., An evaluation of measures of synonymous codon usage bias. *J. Mol. Evol.* 1998. 47. 268–274.
- Duret., L., tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet.* 2000. 16: 287–289.
- Fickett, J,W., Recognition of protein coding regions in DNA sequences. *Nucleic Acids Res.* 1982. 10. 5303–5318.
- Ghosh T. Studies on codon usage in *Entamoeba histolytica*. *Int J Parasitol.* 2000. 30: 715–22.
- Grantham, R., Gautier, C., and Gouy, M., Codon catalog usage and the genome hypothesis. *Nucleic Acids Res.* 1980. 8. 49–62.
- Greenacre, M, J. Theory and applications of correspondence analysis. Academic Press: London; 1984.
- Gu, W., Zhou, T., Ma, J., Sun, X., and Lu, Z., Analysis of synonymous codon usage in SARS Coronavirus and other viruses in the Nidovirales, *Virus Res*, 2004. 101. 155–161.
- Gupta, S. K., and Ghosh, T.C., Gene expressivity is the main factor in dictating the codon usage variation among the genes in *Pseudomonas aeruginosa*,” *Gene*. 2001. 273. 63–70.
- Gupta,S, K., Bhattacharyya, T,K., and Ghosh, T,C., Synonymous codon usage in *Lactococcus lactis*: mutational bias versus translational selection, *J. Biomol. Str. & Dyn.* 2004. 21, 527–536.
- Hou, Z C., and Yang, N., Factors affecting codon usage in *Yersinia pestis*,” *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao Shanghai.* 2003. 5: 580–586.
- Jansen, R, K., Kaittanis, C., Sasaki, C., Lee, S., Tomkins, J., Andrew, J, A., and Daniell, H., Phylogenetic analyses of *Vitis* (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among rosids. *BMC Evol. Biol.* 2006. 6. 32. <https://doi.org/10.1186/1471-2148-6-32>
- Kahali, B., Basak, S., and Ghosh, T, C., Reinvestigating the codon and amino acid usage of *S. cerevisiae* genome: a new insight from protein secondary

- structure analysis. *Biochem Biophys Res Commun*. 2007. 354. 693–699.
- Lynn, D.J., Singer, G, A, C., and D. A. Hickey, D.A., Synonymous codon usage is subject to selection in thermophilic bacteria *Nucleic Acids Res*. 2002. 30. 4272–4277.
- Maria, D, E., Synonymous codon usage in bacteria. *Curr Issues Mol Biol*. 2001. 3. 91–97.
- Mario dos Reis., Renos, S., and Lorenz, W. Solving the riddle of codon usage preferences: A test for translational selection. *Nucleic Acids Res*. 2004. 17. 5036–5044.
- Morton, B.R., Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages. *Mol Evol*. 1998. 46, 449–459.
- Morton, B.R., Strand asymmetry and codon usage bias in the chloroplast genome of *Euglena gracilis*. *Proc Natl Acad Sci U S A*. 1999. 96. 5123–5128.
- Morton, B.R., The role of context-dependent mutations in generating compositional and codon usage bias in grass chloroplast. *DNA J Mol Evol*. 2003. 56. 616–629.
- Onofrio, G.D., Bernardi, G., A universal compositional correlation among codon positions,” *Gene*, 1992. 110, 81–88.
- Osawa, S., Muto, A., T. H. Jukes, T.H., Ohama, T., Evolutionary changes in the genetic code. In Proceedings of the Royal Society Biology. 1990. 19–28
- Osawa, S., and Jukes, T.H., 1989. Codon reassignment (codon capture) in evolution, *J. Mol.Evol*. 28., 271–278,
- Osawa, S., Jukes, T.H., Watanabe A and Muto, A., Recent evidence for evolution of the genetic code, *Microbiol. Rev*. 1992. 56, 229–264.
- Peng, J., Xiao, S., Zuhong, Lu., Analysis of Synonymous Codon Usage in *Aeropyrum pernix* K1 and Other Crenarchaeota Microorganisms. *JGG*. 2007. 34, 275–84.
- Ranjan, A., Vidyarthi, A.S., and Poddar, R., Evaluation of codon bias perspectives in phage therapy of *Mycobacterium tuberculosis* by multivariate analysis,” *In Silico Biol* 2007. 7, 423–431.
- Sharp, P, M, and Cowe, E., Synonymous codon usage in *Saccharomyces cerevisiae*,” *Yeast*. 1991. 7. 657–678.
- Sharp, P.M., and W. H. Li, W.H., The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications, *Nucleic Acids Res*. 1987. 15, 1281–1295.
- Sugiura, M. The chloroplast genome. *Plant Mol Biol*. 1992. 19. 149–168.
- Versteeg, R., Van Schaik, B.D., Van Batenburg., The human transcriptome map reveals extremes in gene density, intron length, GC content and repeat pattern for domains of highly and weakly expressed genes. *Genome Res*. 2003. 13, 1998–2004.
- Wright, F., The ‘effective number of codons’ used in a gene, *Gene*. 1990. 87, 23–29.

How to cite this article:

Farshad Talat and Udikeri, S.S. 2020. Synonymous Codon Usage Bias Factors Affecting Chloroplast Genome of Grape Wine *Vitis vinifera*. *Int.J.Curr.Microbiol.App.Sci*. 9(11): 1971-1977. doi: <https://doi.org/10.20546/ijcmas.2020.911.234>