Molecular Neurobiology (2019) 56:1737–1748
https://doi.org/10.1007/s12035-018-1173-y

Codon Usage Pattern of Genes Involved in Central Nervous System

Arif Uddin[1] · Supriyo Chakraborty[2]

Received: 10 January 2018 / Accepted: 1 June 2018 / Published online: 19 June 2018

Abstract

Codon usage bias (CUB) is the non-uniform usage of synonymous codons in which some codons are more preferred to others in the transcript. Analysis of codon usage bias has applications in understanding the basics of molecular biology, genetics, gene expression, and molecular evolution. To understand the patterns of codon usage in genes involved in the central nervous system (CNS), we used bioinformatic approaches to analyze the protein-coding sequences of genes involved in the CNS. The improved effective number of codons (ENC) suggested that the overall codon usage bias was low. The relative synonymous codon usage (RSCU) revealed that the most frequently occurring codons had a G or C at the third codon position. The codons namely TCC, AGC, CTG, CAG, CGC, ATC, ACC, GTG, GCC, GGC, and CGG (average RSCU > 1.6) were over-represented. Both mutation pressure and natural selection might affect the codon usage pattern as evident from correspondence and parity plot analyses. The overall GC content (59.93) was higher than AT content, i.e., genes were GC-rich. The correlation of GC12 with GC3 suggested that mutation pressure might affect the codon usage pattern.

Keywords Central nervous system · Codon usage bias · Mutation pressure · RSCU

Background

Neurodegenerative diseases (ND) occur when a nerve cell in the central nervous system or peripheral nervous system loses its function over time and ultimately the cell dies. ND affect millions of people over the globe, among which Alzheimer's disease and Parkinson's disease are more prevalent. Reports have suggested that over five million Americans suffer from Alzheimer's disease, and at least 500,000 Americans suffer from Parkinson's disease. The risk of ND increases gradually and its incidence is dramatically related to age. Neurodegeneration is the process of neuropathological conditions and brain aging. It is well known that brain pathology and neurodegenerative diseases are the most important causes of death all over the world. The neurodegenerative disorders such as Parkinson disease (PD), Alzheimer's disease (AD), dementia, cerebrovascular impairment, and seizure disorders have been accounted for the major health problem in the twenty-first century. Neurodegenerative disorders are caused by the defects in some of the genes. AD is a progressive neurodegenerative disorder that accounts for a vast majority of age-related dementia and is known to be one of the most serious health problems in the modern world. AD is characterized by cognitive demur and the accumulation of Aβ deposits and neurofibrillary tangles in the brain. Genetically, the mutations in three genes (i.e., APP, PSEN1, and PSEN2) have been shown to cause AD [1]. Fronto temporal dementia (FTD) with parkinsonism brings about a mutation in the microtubule-associated protein tau [2]. PD is the second most common neurodegenerative disease of adult onset, characterized by a severe loss of dopaminergic neurons in the substantia nigra region of the brain and cytoplasmic inclusions. The mutation in α-synuclein leads to Parkinson disease (PD). The nigra region and cytoplasmic inclusions consist of insoluble protein aggregates in the form of Lewy bodies, causing difficulty in the progressive movement, namely the classic triad of tremor, bradykinesia, and rigidity. The PD occurs at an average age of 50 to 60 years [3–5]. The mutations of five genes have now been shown to cause parkinsonism in early onset such as α-

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12035-018-1173-y) contains supplementary material, which is available to authorized users.

✉ Arif Uddin
arif.uddin29@gmail.com

✉ Supriyo Chakraborty
supriyoch_2008@rediffmail.com

[1] Department of Zoology, Moinul Hoque Choudhury Memorial Science College, Algapur, Hailakandi, Assam 788150, India

[2] Department of Biotechnology, Assam University, Silchar, Assam 788011, India

synuclein (*SNCA* or PARK1) [6]; parkin (PRKN or PARK2) [7]; DJ-1 (DJ1 or PARK7) [8]; PTEN-induced putative kinase I (PINK1 or PARK6) [9]; and leucine-rich repeat kinase 2 or dardarin (LRRK2 or PARK8) [10]. The Lewy body dementia (LBD) is the second most common type of dementia in the aged people, perhaps accounts for 15% of all dementia [11] which is characterized by progressive cognitive impairment, including fluctuating course, visual hallucinations, and parkinsonism. On the other hand, Lewy bodies are also frequently found in classic AD patients with mutations in APP, PSEN1, and PSEN2 [12]. Huntington disease (HD) is due to degeneration of neurons in the basal ganglia and then in cortical regions which are characterized by involuntary movements (chorea), psychiatric symptoms, and dementia. Approximately 90% of HD cases are hereditary and transmitted in an autosomal dominant fashion. The HD gene was the first autosomal disease located in the chromosome 4q16. The HD is due to a defect in a single gene, i.e., huntingtin, although such defects only account for 50% of the cases [13].

Amino acids are the building blocks of proteins. Each amino acid is encoded by a codon, a sequence of three nitrogenous bases. Except for methionine and tryptophan, all the amino acids are encoded by more than two codons, which are known as synonymous codons. The phenomenon of synonymous codon usage is not uniform, i.e., some codons are more preferred to others. It is known as codon usage bias (CUB) and found to be species specific [14, 15]. CUB is determined by either compositional constraints under mutation pressure or natural selection, and it is a unique characteristic of organisms from prokaryotes to eukaryotes [16]. An earlier study suggested that the CUB in mammals has been caused by the difference in isochores or variation in tRNA pool of the cell [17, 18]. Various factors have been attributed to cause the synonymous codon usage bias such as protein secondary structure, expression level of the gene, gene function, and translational selection [19, 20]. The codon usage bias is positively correlated with gene expression level, but is inversely associated with the rate of synonymous substitution in human [21]. The diverse factors which affect the evolution of CUB have been extensively studied in various organisms [22, 23]. Earlier reports suggested that highly expressed genes have stronger CUB which may possibly be due to the selection pressure influencing on those genes (Ikemura 1985). However, the degree of selection varies which depends on the amino acids that exhibit stronger bias. Mutation pressure is the major factor contributing to codon usage variation in various prokaryotes [19] and in many mammals [24]. However, translational selection plays the important role in codon usage bias in *Drosophila* [25] and in some plants [26].

Analysis of CUB acquires importance in the heterologous gene expression, prediction of the expression level of genes, design of degenerate primers as well as in the prediction of gene functions. The studies of CUB have focused on model organisms namely *Drosophila*, *Caenorhabditis*, *Arabidopsis*, *Giardia lamblia*, *Entamoeba histolytic*, and *Saccharomyces cerevisiae* [17]. However, no work was reported on nucleotide composition and codon usage bias of genes associated with the central nervous system (CNS). This study sheds insight into the factors which influence the codon usage bias. This study also elucidates the over-represented and under-represented codons which help increase or decrease the gene expression level.

Methodology

Sequence Retrieval

The coding sequences (cds) of the genes (in FASTA format) associated with the central nervous system were retrieved from the National Center for Biotechnology Information (NCBI) GenBank database (http://www.ncbi.nlm nih.gov). In these analyses, we used only the coding sequences of 52 genes that are devoid of any unknown base (S1). To minimize sampling errors and improve the quality of sequences, coding sequences without correct initiation and termination codons or with internal termination codons were avoided.

Compositional Properties

The compositional properties such as the overall frequency of occurrence of the nucleotides (A, C, T, and G %) and the frequency of each nucleotide at the third site of the synonymous codons (A3, C3, T3, and G3%) were calculated for the genes associated with CNS. The start and stop codons along with the codons for met and trp were excluded from the analysis.

Relative Synonymous Codon Usage

RSCU of a codon is the ratio of the observed frequency of a synonymous codon to the sum of all the synonymous codons encoding the same amino acid multiplied by degeneracy level [27]. The synonymous codons with RSCU values > 1.6 and < 0.6 were referred as over-represented and under-represented codons, respectively [28].

GC3 Analysis

It is a useful index for evaluating the degree of base composition bias, which indicates the frequency of either a guanine or cytosine at the third codon position of synonymous codons for an amino acid.

Improved Effective Number of Codons and Effective Number of Codon Prime (ENC′)

The effective number of codon (ENC) is used to quantify the codon usage bias of the cds of interest, independent of the gene length and the number of amino acids [29]. The ENC values range from 20 for a cds showing high codon usage bias to 61 for cds showing no bias. When the ENC value of a cds is less than or equal to 35, the cds is said to have a significant codon usage bias [30]. It is a non-directional measure of codon usage bias, i.e., higher ENC value means lower codon usage bias and vice versa. The ENC of a cds was calculated using the following formula:

$$ENC = 2 + \frac{9}{F_2} + \frac{1}{F_3} + \frac{5}{F_4} + \frac{3}{F_6}$$

where Fk (k = 2, 3, 4, 6) is the mean of Fk values for the k-fold degenerate amino acids.

The mathematical formula used in calculating the ENC gives inappropriate weightage to the amino acids of low abundance and the magnitude of error also varies according to degeneracy level of codons. So, two parameters viz. improved ENC and ENC prime were developed by Satapathy et al. 2017, which measure the codon usage bias of a cds more accurately [31]. In this study, improved ENC and ENC prime were estimated using online tool (http://agnigarh.tezu.ernet.in/~ssankar/cub.php).

Competition Adaptation Index

Competition adaptation index (CompAI) is a measure of translational speed of a coding sequence. It is based on the tRNA gene copy number (tGCN) of a species. CompAI value varies from 0 (lowest translation rate) to 1 (highest translation rate). It takes into account the cognate as well as the near-cognate anticodons of the corresponding codons [32]. When compAI equals zero, it indicates the highest competition in translation and hence reveals the lowest translation rate of the coding sequence. But when the compAI value is 1, it means the lowest competition for translation of the coding sequence and therefore the coding sequence is translated at the highest rate. Therefore, compAI is used as a measure of translation rate of the coding sequences.

Correspondence Analysis

Correspondence analysis (COA) is a multivariate statistical method that has been used in our analysis to explore the major trends in codon usage patterns of coding sequences of genes associated with CNS using RSCU values of codons. The coding sequences of the genes were represented as a 59-dimensional vector, and each dimension corresponded to the RSCU value of each of 59 codons used to minimize the effect of amino acid composition on codon usage [33, 34].

PR2-Bias Plot Analysis

The Parity Rule 2 (PR2) bias plot was drawn using the value of AT-bias [A/(A + T)] as the ordinate and GC-bias [G/(G + C)] as the abscissa [35]. The center of the plot, i.e., 0.5, is the place where A = T and G = C (PR2), indicating no bias between the two complementary strands of DNA resulting from mutation and selection rates (substitution rates) [36].

Statistical Analysis

Correlation analysis was performed to detect and quantify the relationship between overall nucleotide composition and its composition at the third codon position. All statistical analyses were carried out using the statistical software SPSS 16.0 for windows. Nucleotide compositional features were estimated by using an in-house perl script.

Results

Codon Usage Bias of Genes Associated with Central Nervous System

The improved ENC of genes associated with human central nervous system ranged from 31.34 to 52.01, with a mean of 40.53. But the ENC prime ranged from 43.43 to 55.4 with a mean value of 50.07. These results suggested low codon usage bias of these genes [31, 37].

Relative Synonymous Codon Usage

To understand the pattern of non-uniform usage of synonymous codons in these genes, relative synonymous codon usage (RSCU) of individual codons was calculated. Additionally, the RSCU values of 59 sense codons also support the conclusion that the genes associated with central nervous system showed weak codon bias. Nearly half of the codons (24/59) were frequently used as shown in Fig. 1. More frequent codons ended with G or C. Among these, the codons TCC, AGC, CTG, CAG, CGC, ATC, ACC, GTG, GCC, GGC, and CGG showed (average RSCU > 1.6) strong usage bias (over-represented) (Fig. 1**)** while 20 codons were under-represented (average RSCU < 0.6). It was also evident from the Fig. 1 that the trend of codon usage differed among genes which strongly support our hypothesis that the pattern of synonymous codon usage was different among the genes associated with CNS [28, 38].

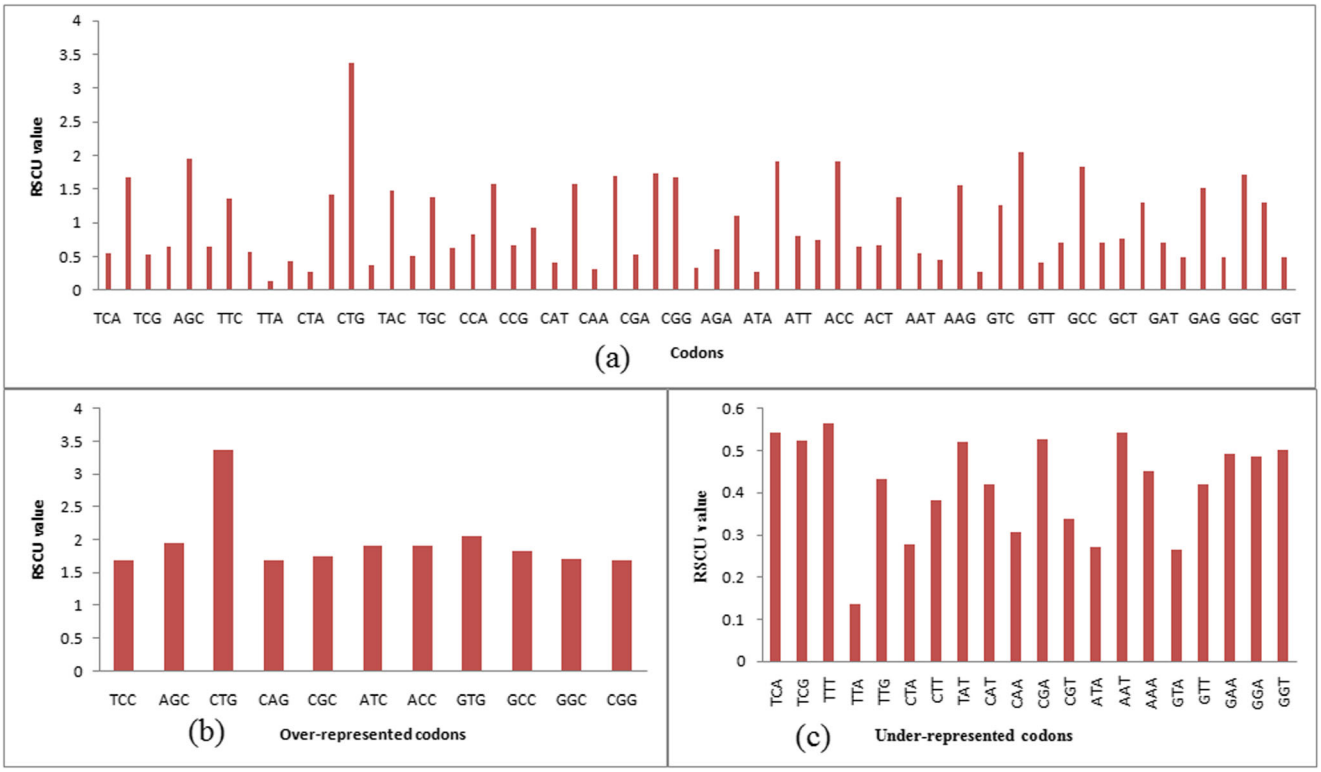

Fig. 1 **a** Frequency of 59 codons. **b** Over-represented codons of genes associated with CNS. **c** Under-represented codons of genes associated with CNS

Correspondence Analysis

Correspondence analysis was performed on the RSCU values of codons in the coding sequences of genes involved in CNS. All the genes showed scattered distribution which indicated that codon usage pattern was different in CNS genes. The axis 1 of the COA contributed 36.47% and axis 2 contributed 9.24% of the total variation, respectively (Fig. 2). Axis 1 represented the major factor that affected codon usage bias and the distribution density of G/C ended codons was closer to axis 1 than that of codons ending with A/T. These results suggested that nucleotide composition of G/C ended codons influenced the codon usage bias more than A/T ended codons [39]. Further, some of the genes were in a discrete distribution, suggesting that other factors such as natural selection might have affected the codon usage bias of the genes [40].

Parity Plot

If only mutation pressure influences the codon usage bias of a gene, GC and AT should be used equally in 4-fold degenerate codon families while natural selection would not necessarily cause the equal use of GC and AT [35]. We, therefore, analyzed the associations between the purine (A and G) and the pyrimidine (C and T) content using A3/A3 + T3 as ordinate and G3/G3 + C3 as abscissa in the 4-fold degenerate codon families. We found that A and T (G and C) were not used proportionally in these degenerate codons (Fig. 3), which revealed that both mutation pressure and natural selection might influence the codon usage bias [41, 42].

Nucleotide Composition of Genes Involved in CNS

We analyzed the nucleotide composition of coding sequences of genes associated with CNS to understand the compositional properties which greatly influenced the codon usage bias [25]. We found the nucleobase C (31.16%) and G (29.38%) occurred more frequently than nucleobase A (21.71%) and T (17.73%). The same trend was found for nucleotide composition at the 3rd codon position. The nucleobase C occurred most frequently at the third codon position (38.63%) and A occurred the least frequently (12.05%). The overall nucleotide composition and the composition at the third codon position in the coding sequences of genes associated with CNS suggested that compositional constraint might influence the codon usage pattern of these genes. The average GC content of the genes was 59.93%, i.e., genes were GC-rich, which was different from GC content at the first (61.85%), second (44.22%), and third (73.73%) position of codons. The GC content at the third position was higher than GC content at the first and second codon position and the greatest difference of GC content was found between second and third codon position (Fig. 4**).**

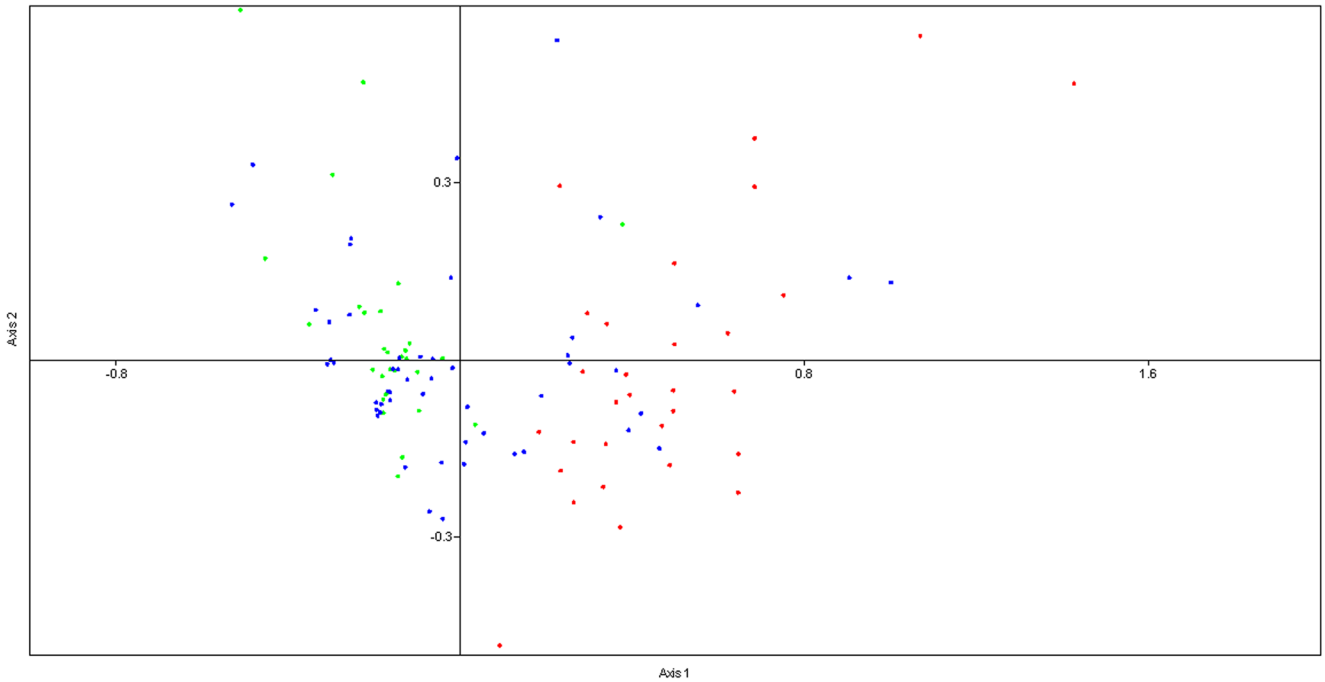

Fig. 2 Correspondence analysis of genes involved in CNS using RSCU values. Red color indicates AT-ending codons, diamond color indicates GC-ending codons, and blue indicates different genes. Axis 1 contributes 36.47% and axis 2 contributes 9.24% of the total variation

Relationship of Compositional Properties

We compared the correlation between general nucleotide composition (A, T, G, C, GC) and nucleotide composition at the third codon position (A3, T3, G3, C3, GC3) using the Pearson's product moment method to understand whether the evolution of codon usage bias in the coding sequences of genes associated with CNS had been driven by mutation pressure alone or by both mutation pressure and translational selection. We found a significant correlation between overall nucleotide composition and its third codon position (Table 1) which suggested that both mutation pressure and natural selection might affect the codon usage bias of these genes [37].

Further, we found significant positive correlation between GC12 and GC3 ($r = 0.412^{**}$, $p < 0.01$) while significant negative correlation between ENC and GC3 ($r = -876^{**}$, $p < 0.05$), which suggest that mutation pressure might influence the codon usage bias [28, 42].

Regression Analysis Among Compositional Properties

We performed regression analysis between homogeneous overall nucleotide composition and its third codon position

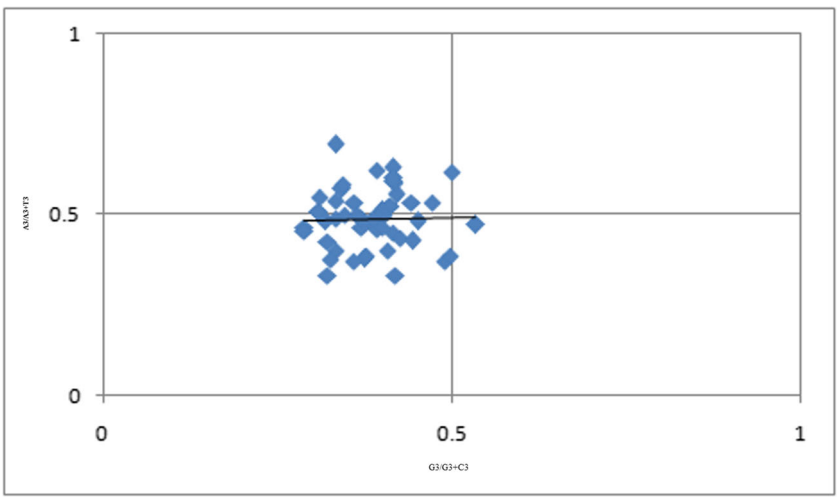

Fig. 3 Parity plot of genes in 4-fold degenerate codon family

Fig. 4 Distribution of overall GC content and GC content at the first, second, and third position of codon

to explore the magnitude of contribution of each base for mutational pressure. From the Fig. 5, it was evident that G, G3, T, and T3 were responsible for higher mutational pressure than C, C3, A, and A3.

Relationship Between Improved ENC and Compositional Properties

We found significant positive correlation between codon usage bias (ENC) and different nucleobases namely A ($r = 0.498^{**}$, $p < 0.01$), T ($r = 0.752^{**}$, $p < 0.01$), A3 ($r = 0.771^{**}$, $p < 0.01$), and T3 ($r = 0.888^{**}$, $p < 0.01$) while significant negative correlation was found between ENC and nucleobase viz. G ($r = -0.773^{**}$, $p < 0.01$, C ($r = -0.60^{**}$, $p < 0.01$), G3($r = -0.791^{**}$, $p < 0.01$ and C3 ($r = -0.736^{**}$, $p < 0.01$) (Table 2). From the Fig. 6, it was evident that the regression coefficient was negative between ENC and nucleobase G, C, G3, and C3 while it was positive between ENC and nucleobase A, T, A3, and T3. Since ENC is a non-directional measure of codon usage bias (i.e., high value of ENC indicates low codon usage bias), the negative regression coefficient between ENC and nucleobase G, C, G3, and C3 indicated that two nucleobases G and C positively influenced the codon usage bias [43].

Table 1 Overall nucleotide composition and composition at the third codon position

	A3	T3	G3	C3	GC3
A	0.585^{**}	0.560^{**}	-0.327^{*}	-0.715^{**}	-0.602^{**}
T	0.633^{**}	0.804^{**}	-0.776^{**}	-0.553^{**}	-0.761^{**}
G	-0.778^{**}	-0.748^{**}	0.875^{**}	0.529^{**}	0.802^{**}
C	-0.581^{**}	-0.730^{**}	0.367^{**}	0.833^{**}	0.694^{**}
GC	-0.805^{**}	-0.892^{**}	0.709^{**}	0.849^{**}	0.895^{**}

$^{*}p < 0.05$, $^{**}p < 0.01$

Relationship Between Codon Usage Bias and Various Skews

The GC skews in most of the genes were negative whereas AT skews were positive (S2). This result indicated that asymmetrical nucleotide composition between the two strands of DNA, one with an abundance of C over G and the other with an abundance of A over T [44].

We performed correlation analysis between ENC and each skew to understand the effect of skewness on codon usage bias. We found highly significant positive correlation between ENC and purine skew ($r = 0.675^{**}$, $p < 0.01$), ENC and pyrimidine skew ($r = 0.795^{**}$, $p < 0.01$), ENC and amino skew ($r = 0.580^{**}$, $p < 0.01$), ENC and keto skew ($r = 0.803^{**}$, $p < 0.01$ which suggested that these skews might affect the codon usage of the genes [38].

Effect of Competition Adaptation Index on Codon Usage Bias

CompAI is a measure of translational speed of a coding sequence and its value ranges from 0 to 1. When compAI value is zero, it indicates the lowest translation rate of the coding sequence whereas the compAI value 1 suggests the highest translation rate [32]. We estimated compAI for coding sequences of genes involved in CNS and found the average value was 0.190, which suggested that translation rate of the coding sequence was low. The values of compAI ranged from 0.145 to 0.230. Further, we performed correlation analysis between ENC and compAI to understand whether compAI had any effect on codon usage bias. We found highly significant positive correlation between ENC and compAI ($r = -0.357^{**}$, $p < 0.01$), which suggested that translation speed of a coding sequence might influence the codon usage bias of genes.

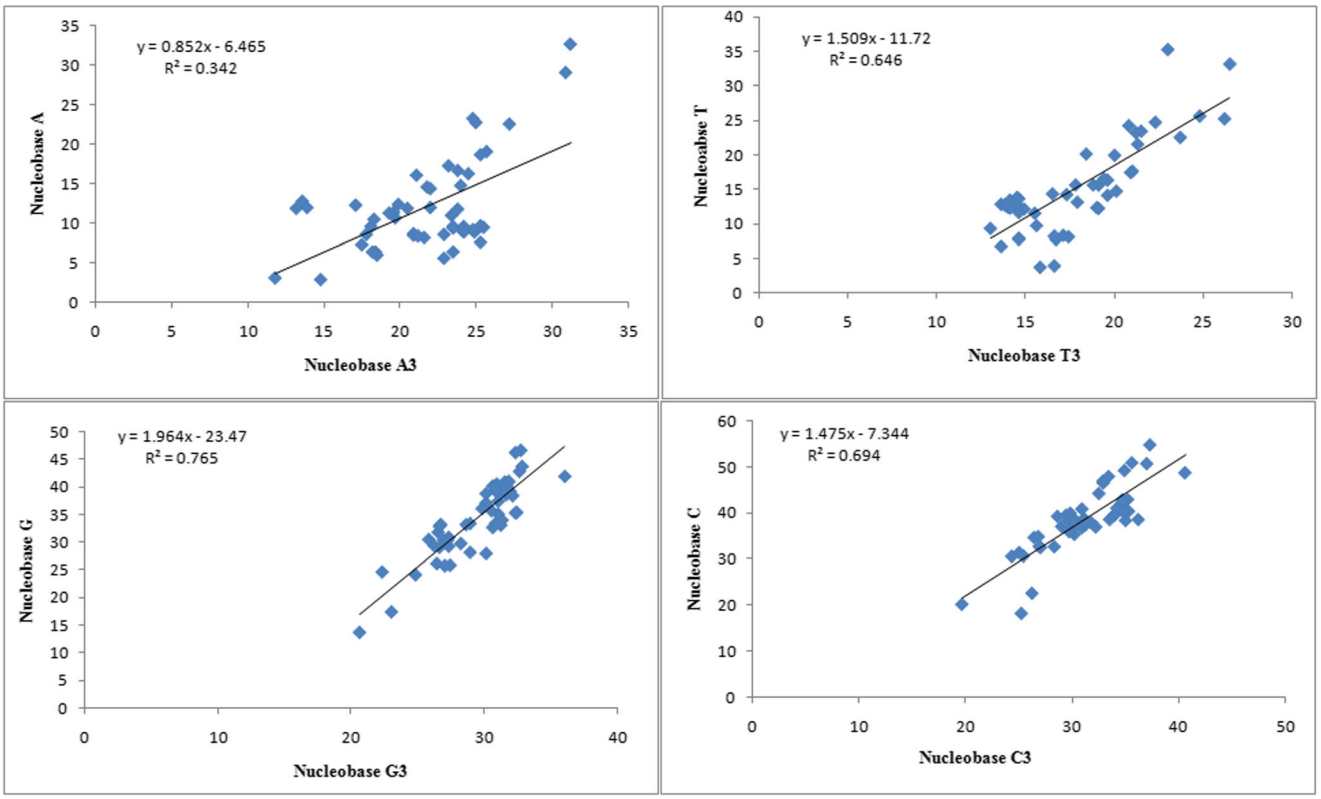

Fig. 5 Regression analysis between overall nucleotide composition and nucleotide composition at the third codon position

Mutational Responsive Index and Translational Selection (P2)

The average MRI and P2 values of genes involved in CNS were 0.45 and 0.06, respectively. The positive value of MRI indicates directional mutation pressure while negative MRI value indicates that translational selection acts on the genes [45]. Our finding suggested that directional mutation might affect the codon usage bias of the genes while the role of translational selection was very low [46, 47].

Comparison of Codon Usage Bias with Phylogenetic Analysis

Multiple sequence alignment analysis was done using Clustal X2, and then, the phylogenetic tree was constructed using MEGA 6 (Fig. 7a). The phylogenetic tree was compared with ENC distribution (Fig. 7b) and GC content at the third codon position (Fig. 7c). The phylogenetic tree was performed using a neighbor-joining method based on the coding sequences. We observed that some of the genes showed close evolutionary relationship, and the difference in their corresponding ENC and GC3 values was relatively small. This suggests that the closer the evolutionary relationship of genes, the more similar their codon usage bias supporting the result of Zhao et al. [48].

Discussion

Neurodegenerative diseases are now very common in human beings. These diseases are caused by several factors namely genetic and environmental factors, food habits etc. Many diseases are due to mutation of genes which are involved in the development of the central nervous system. Recent research mostly focuses on generating animal model of neurodegenerative diseases and a little work on codon usage bias was done for genes involved in AD. But no work was reported regarding the important genes involved in the normal function of CNS. Codon usage bias is an important phenomenon, and it exists in a wide variety of organisms, ranging from prokaryotes to eukaryotes. Among all the theories proposed to describe the origin of CUB, neutral theory and selection-mutation-drift

Table 2 Correlation between codon usage bias and nucleotide composition

	A	T	G	C	A3	T3	G3	C3
ENC	0.498**	0.752**	−0.773**	−0.600**	0.771**	0.888**	−0.791**	−0.736**

*$p < 0.05$, **$p < 0.01$

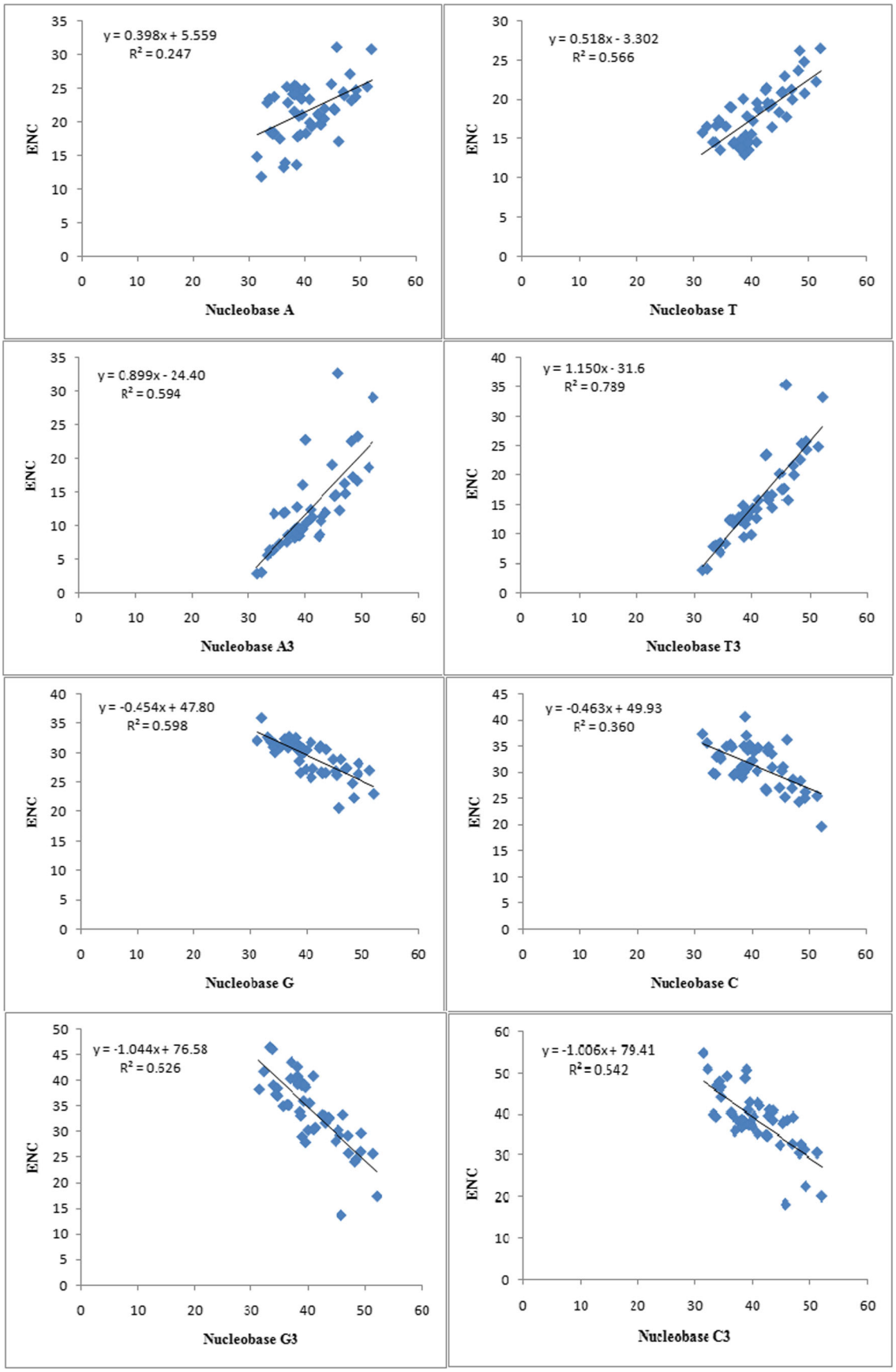

Fig. 6 Regression analysis between ENC and various nucleobases

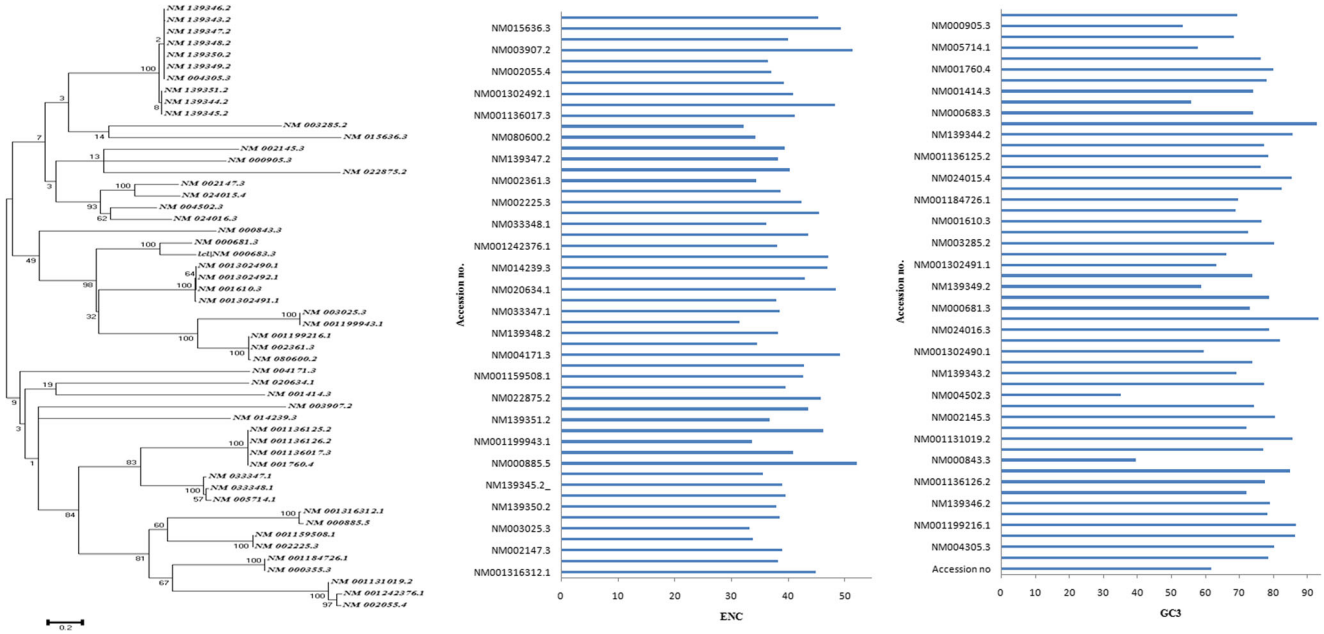

Fig. 7 Comparative analysis of phylogenetic tree, distribution of ENC and GC3 content of genes involved in CNS. **a** Phylogenetic tree. **b** Distribution of ENC. **c** GC3 content in CNS genes

balance model are the most important ones. But after the completion of the whole genome sequencing of many organisms, these two theories proved insufficient to explain the phenomenon of CUB [42]. Various other factors were proposed to affect CUB which includes GC-content [49], gene length [50], RNA structure [51], protein structure [52], hydrophobicity and aromaticity of the encoded proteins [53], and environmental stress [54] etc. This study would give insight into the molecular details of important genes involved in the normal function of CNS.

In this study, the mean improved ENC was 40.53 (i.e., higher than 35) indicating a relatively low codon usage bias. The ENC value of albumin superfamily varied from 51.65 to 56.62 [55]. The ENC value in *B. mori* ranged from 30.06 to 61.00, with an average of 53.12 [56]. The ENC values in four species of *Bungarus* were 35–60 [47]. The ENC values for CYB gene in different species of aves and mammals were 59.66 and 58.33, respectively, which indicated low codon usage bias [38]. Furthermore, the ENC values of the CHIKV genomes ranged from 54.55 to 56.41 with an average of 55.56. These results indicate low codon usage bias of these genes [28]. The presence of low codon usage bias might be helpful for efficient replication in vertebrates with different cell types having different codon preferences [25].

The bar diagram using RSCU values of codons displayed a remarkable difference of codon usage in genes. More frequent codons (RSCU > 1) ended with G or C. Correspondence analysis is a multivariate statistical method used to analyze the pattern of codon usage variation in genes and to distribute the codons in axis 1 and axis 2 with these trends [34]. The distribution density of G/C ended codons was closer to axis 1 than AT ended codons, but some genes were in discrete distribution in our current study. Wei et al. analyzed CUB in mitochondrial DNA of *B. mori* and found that the contribution of axis 1 was 12.07% and axis 2 8.64% to total variation. In their work, the positions of the genes were close to axes with a concentrate distribution, suggesting that the compositional features for mutation bias might correlate to the CUB. Further, a few genes were in a discrete distribution, suggesting that natural selection might also influence the codon usage bias [40].

The GC content plays an important role in CUB. It may influence the bendability, thermostability, and convertability of B form of DNA to Z form of DNA. The GC content is also involved in the active process of transcription because it has the capability to keep the coding region of DNA in an open chromatin state [57]. Previous studies found that highly expressed genes may have low mutation rates due to DNA repair mechanisms [58]. The synonymous codons generally differ at the third codon position and the GC3 (guanine and cytosine at the third position of codon) is a good indicator of the extent of synonymous CUB [59]. Previous studies reported that genes with higher GC3 content tend to get more methylated leading to mutation as compared to the genes with low GC3 content [60]. It was also reported that GC3 acts as an isochore marker but the association between GC3 and the GC content of the flanking regions is still doubtful [61].

The compositional property of nucleobase is an important feature of genes which affects the codon usage bias. The GC-rich organisms like some bacteria, fungi, *Oryza sativa*, *Triticum aestivum*, *Hordium vulgare* tend to use G or C at the third codon position. But AT-rich organisms such as

Plasmodium falciparum and *Onchocerca volvulus* are more likely to use A or T ar third codon position. In our current analysis, the percentage of nucleobase C and G was higher than A and T and the average GC content of these coding sequences was 59.93%, i.e., transcripts were GC-rich. In support of our study on CNS genes, the overall GC content (53.04%) of CYP genes was higher than the AT content in human [62]. The overall GC (65.2%) was higher than AT (34.8%) content in GATA2 gene across mammals [63]. It was found that the occurrence of the nucleobase C at the third codon position was the highest but that of T was the lowest, i.e., GC content was found to be higher than AT content in the genome analysis of rabbit [64]. Previous studies revealed that a close relationship exists between nucleotide compositional properties and gene function [65]. In our current study, most of the genes showed different nucleotide composition suggesting variation in their biological properties. In support to our findings, earlier studies also found that ALB and AFP, two members of albumin superfamily, showed similar pattern of nucleotide distribution indicating that those genes resemble in their structures and biological functions. But AFM and VDBP genes, although grouped in the albumin superfamily, showed differential compositional properties suggesting difference in their biological functions in contrast to the other members of the group [55] .

The protein translation is a biological process which is prone to errors. The effect of translation error resulting from CUB is based on its effect on protein function as well as its frequency [66]. Both missense errors and nonsense errors are referred to as translation errors. The missense errors incorporate wrong amino acids in the growing peptide chain while nonsense errors cause premature termination of a growing polypeptide chain. Some of the researchers considered that selection against missense errors may cause codon usage bias, and as a result, translation accuracy is maintained [24, 67–70]. Previous studies reported that the synonymous codons for an amino acid vary in translation error proneness. The translation error rate of a codon also depends on the relative abundance of its cognate and near cognate tRNAs [71].

Two major evolutionary forces namely mutation pressure and natural selection are considered to influence the CUB of a genome. The causes of mutational bias are non-uniform DNA repair, non-random replication errors, and chemical decay of nucleotides [72]. The neutral mutation bias typically acts on DNA sequences at the third position of codon altering their compositional properties but do not affect the functionality of proteins since amino acid composition and their sequence in protein remains unaltered. In our current study, significant correlation between overall nucleotide composition and its composition at third codon position suggested that both mutation pressure and natural selection might influence the codon usage bias of these genes [37]. In our earlier study, we found significant correlation between them for CYB gene [38], genes of *Bungarus* species [47], and SPANX gene [73] etc. Similar result was also found for ND2 gene [74] and for mitochondrial genes of *B. mori* [40].

Acknowledgements The authors are thankful to Assam University, Silchar, Assam, India, for providing necessary facilities.

Funding Unfunded: The work was not supported by any national or international organization.

Compliance with Ethical Standards

The study is based on DNA sequence-based analysis. Ethical clearance is not applicable.

Conflict of Interest The authors declare that they have no conflict of interests.

References

1. Risch NJ (2000) Searching for genetic determinants in the new millennium. Nature 405:847–856
2. Bertram L, Tanzi RE (2005) The genetic epidemiology of neurodegenerative disease. J Clin Invest 115:1449–1457
3. Tanner CM, Ottman R, Goldman SM, Ellenberg J, Chan P et al (1999) Parkinson disease in twins: an etiologic study. Jama 281: 341–346
4. Maher N, Golbe L, Lazzarini A, Mark M, Currie L, Wooten GF, Saint-Hilaire M, Wilk JB et al (2002) Epidemiologic study of 203 sibling pairs with Parkinson's disease the GenePD study. Neurology 58:79–84
5. de la Fuente-Fernandez R (2003) A note of caution on correlation between sibling pairs. Neurology 60:1561–1561
6. Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, Pike B, Root H et al (1997) Mutation in the α-synuclein gene identified in families with Parkinson's disease. Science 276: 2045–2047
7. Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, Minoshima S, Yokochi M, Mizuno Y et al (1998) Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature 392:605–608
8. Bonifati V, Rizzu P, van Baren MJ, Schaap O, Breedveld GJ, Krieger E, Dekker MC, Squitieri F et al (2003) Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. Science 299:256–259
9. Valente EM, Abou-Sleiman PM, Caputo V, Muqit MM, Harvey K, Gispert S, Ali Z, del Turco D et al (2004) Hereditary early-onset Parkinson's disease caused by mutations in PINK1. Science 304: 1158–1160
10. Zimprich A, Biskup S, Leitner P, Lichtner P, Farrer M, Lincoln S, Kachergus J, Hulihan M et al (2004) Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron 44:601–607
11. McKeith IG, Galasko D, Kosaka K, Perry E, Dickson DW et al (1996) Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB) report of the consortium on DLB international workshop. Neurology 47:1113–1124
12. Hamilton RL (2000) Lewy bodies in Alzheimer's disease: A neuropathological review of 145 cases using α-Synuclein immunohistochemistry. Brain Pathol 10:378–384

13. Tanzi RE, McClatchey AI, Lamperti ED, Villa-Komaroff L, Gusella JF, et al. (1988) Protease inhibitor domain encoded by an amyloid protein precursor mRNA associated with Alzheimer's disease.
14. Bennetzen JL, Hall B (1982) Codon selection in yeast. J Biol Chem 257:3026–3031
15. Plotkin JB, Robins H, Levine AJ (2004) Tissue-specific codon usage and the expression of human genes. Proc Natl Acad Sci U S A 101:12588–12591
16. Duret L (2002) Evolution of synonymous codon usage in metazoans. Curr Opin Genet Dev 12:640–649
17. Bernardi G, Olofsson B, Filipski J, Zerial M, Salinas J, Cuny G, Meunier-Rotival M, Rodier F (1985) The mosaic genome of warm-blooded vertebrates. Science 228:953–958
18. Dittmar KA, Goodenbour JM, Pan T (2006) Tissue-specific differences in human transfer RNA expression. PLoS Genet 2:e221
19. Gupta S, Ghosh T (2001) Gene expressivity is the main factor in dictating the codon usage variation among the genes in Pseudomonas aeruginosa. Gene 273:63–70
20. Liu Q (2006) Analysis of codon usage pattern in the radioresistant bacterium Deinococcus radiodurans. Biosystems 85:99–106
21. Urrutia AO, Hurst LD (2001) Codon usage bias covaries with expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection. Genetics 159:1191–1199
22. Ikemura T (1982) Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes: Differences in synonymous codon choice patterns of yeast and Escherichia coli with reference to the abundance of isoaccepting transfer RNAs. J Mol Biol 158:573–597
23. Moriyama EN, Powell JR (1998) Gene length and codon usage bias in Drosophila melanogaster, Saccharomyces cerevisiae and Escherichia coli. Nucleic Acids Res 26:3188–3193
24. Akashi H (1994) Synonymous codon usage in Drosophila melanogaster: Natural selection and translational accuracy. Genetics 136:927–935
25. Jenkins GM, Holmes EC (2003) The extent of codon usage bias in human RNA viruses and its evolutionary origin. Virus Res 92:1–7
26. Knight RD, Freeland SJ, Landweber LF (2001) A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. In: Genome biology 2: research0010
27. Sharp PM, 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
28. Butt AM, Nasrullah I, Tong Y (2014) Genome-wide analysis of codon usage and influencing factors in chikungunya viruses. PLoS One 9:e90905
29. Wright F (1990) The 'effective number of codons' used in a gene. Gene 87:23–29
30. Comeron JM (2004) Selective and mutational patterns associated with gene expression in humans influences on synonymous composition and intron presence. Genetics 167:1293–1304
31. Satapathy SS, Sahoo AK, Ray SK, Ghosh TC (2017) Codon degeneracy and amino acid abundance influence the measures of codon usage bias: Improved Nc ($\hat{N}c$) and ENCprime ($\hat{N}'c$) measures. Genes Cells 22:277–283
32. Dilucca M, Cimini G, Semmoloni A, Deiana A, Giansanti A (2015) Codon Bias patterns of E. coli's interacting proteins. PLoS One 10: e0142127
33. Grantham R, Gautier C, Gouy M (1980) Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. Nucleic Acids Res 8:1893–1912
34. Shields DC, Sharp PM (1987) Synonymous codon usage in Bacillus subtilis reflects both translational selection and mutational biases. Nucleic Acids Res 15:8023–8040
35. Sueoka N (1995) Intrastrand parity rules of DNA base composition and usage biases of synonymous codons. J Mol Evol 40:318–325
36. Sueoka N (1999) Two aspects of DNA base composition: G+ C content and translation-coupled deviation from intra-strand rule of a= T and G= C. J Mol Evol 49:49–62
37. Zhang Z, Dai W, Dai D (2013) Synonymous codon usage in TTSuV2: Analysis and comparison with TTSuV1. PLoS One 8: e81469
38. Uddin A, Chakraborty S (2015) Synonymous codon usage pattern in mitochondrial CYB gene in pisces, aves, and mammals. Mitochondrial DNA: 1–10.
39. Zhou T, Gu W, Ma J, Sun X, Lu Z (2005) Analysis of synonymous codon usage in H5N1 virus and other influenza a viruses. Biosystems 81:77–86
40. Wei L, He J, Jia X, Qi Q, Liang Z, Zheng H, Ping Y, Liu S et al (2014) Analysis of codon usage bias of mitochondrial genome in Bombyx mori and its relation to evolution. BMC Evol Biol 14:262
41. Uddin A, Chakraborty S (2016) Codon usage trend in mitochondrial CYB gene. Gene 586:105–114
42. Yang X, Luo X, Cai X (2014) Analysis of codon usage pattern in Taenia saginata based on a transcriptome dataset. Parasit Vectors 7: 1–11
43. Behura SK, Severson DW (2012) Comparative analysis of codon usage bias and codon context patterns between dipteran and hymenopteran sequenced genomes. PLoS One 7:e43111
44. Beletskii A, Bhagwat AS (2001) Transcription-induced cytosine-to-thymine mutations are not dependent on sequence context of the target cytosine. J Bacteriol 183:6491–6493
45. Gatherer D, McEwan NR (1997) Small regions of preferential codon usage and their effect on overall codon bias-the case of the plp gene. IUBMB Life 43:107–114
46. Sur S, Sen A, Bothra AK (2007) Mutational drift prevails over translational efficiency in Frankia nif operons. Indian J Biotechnol 6:321–328
47. Chakraborty S, Nag D, Mazumder TH, Uddin A (2017) Codon usage pattern and prediction of gene expression level in Bungarus species. Gene 604:48–60
48. Zhao Y, Zheng H, Xu A, Yan D, Jiang Z, Qi Q, Sun J (2016) Analysis of codon usage bias of envelope glycoprotein genes in nuclear polyhedrosis virus (NPV) and its relation to evolution. BMC Genomics 17:677
49. Hey J, Kliman RM (2002) Interactions between natural selection, recombination and gene density in the genes of Drosophila. Genetics 160:595–608
50. 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
51. Hartl DL, Moriyama EN, Sawyer SA (1994) Selection intensity for codon bias. Genetics 138:227–234
52. Orešič M, Dehn MH, Korenblum DH, Shalloway DH (2003) Tracing specific synonymous codon–secondary structure correlations through evolution. J Mol Evol 56:473–484
53. Romero H, Zavala A, Musto H (2000) Codon usage in chlamydia trachomatis is the result of strand-specific mutational biases and a complex pattern of selective forces. Nucleic Acids Res 28:2084–2090
54. Goodarzi H, Torabi N, Najafabadi HS, Archetti M (2008) Amino acid and codon usage profiles: Adaptive changes in the frequency of amino acids and codons. Gene 407:30–41
55. Mirsafian H, Mat Ripen A, Singh A, Teo PH, Merican AF, Mohamad SB (2014) A comparative analysis of synonymous codon usage Bias pattern in human albumin superfamily. Sci World J 2014:1–7
56. Jia X, Liu S, Zheng H, Li B, Qi Q, Wei L, Zhao T, He J et al (2015) Non-uniqueness of factors constraint on the codon usage in Bombyx mori. BMC Genomics 16:356

57. Schwartz S, Meshorer E, Ast G (2009) Chromatin organization marks exon-intron structure. Nat Struct Mol Biol 16: 990–995
58. Hoeijmakers JH (2001) Genome maintenance mechanisms for preventing cancer. Nature 411:366–374
59. Shen W, Wang D, Ye B, Shi M, Ma L, Zhang Y, Zhao Z (2015) GC3 biased gene domains in mammalian genomes. In: Bioinformatics: btv329, vol 31, pp. 3081–3084
60. Tatarinova TV, Alexandrov NN, Bouck JB, Feldmann KA (2010) GC3 biology in corn, rice, sorghum and other grasses. BMC Genomics 11:308
61. Aota S-i, Ikemura T (1986) Diversity in G+ C content at the third position of codons in vertebrate genes and its cause. Nucleic Acids Res 14:6345–6355
62. Malakar AK, Halder B, Paul P, Chakraborty S (2016) Cytochrome P450 genes in coronary artery diseases: Codon usage analysis reveals genomic GC adaptation. Gene 590:35–43
63. Mazumder TH, Uddin A, Chakraborty S (2016) Transcription factor gene GATA2: Association of leukemia and nonsynonymous to the synonymous substitution rate across five mammals. Genomics 107:155–161
64. FADIEL A (2003) GENOME ANALYSIS OF GENBANK KNOWN RABBIT.
65. Garcia JA, Fernández-Guerra A, Casamayor EO (2011) A close relationship between primary nucleotides sequence structure and the composition of functional genes in the genome of prokaryotes. Mol Phylogenet Evol 61:650–658
66. Shah P, Gilchrist MA (2010) Effect of correlated tRNA abundances on translation errors and evolution of codon usage bias. PLoS Genet 6:e1001128
67. Drummond DA, Wilke CO (2009) The evolutionary consequences of erroneous protein synthesis. Nat Rev Genet 10:715–724
68. Akashi H (2001) Gene expression and molecular evolution. Curr Opin Genet Dev 11:660–666
69. Arava Y, Boas FE, Brown PO, Herschlag D (2005) Dissecting eukaryotic translation and its control by ribosome density mapping. Nucleic Acids Res 33:2421–2432
70. Stoletzki N, Eyre-Walker A (2007) Synonymous codon usage in Escherichia coli: Selection for translational accuracy. Mol Biol Evol 24:374–381
71. Kramer EB, Farabaugh PJ (2007) The frequency of translational misreading errors in E. Coli is largely determined by tRNA competition. Rna 13:87–96
72. Kaufmann WK, Paules RS (1996) DNA damage and cell cycle checkpoints. FASEB J 10:238–247
73. Choudhury MN, Chakraborty S (2015) Codon usage pattern in human SPANX genes. Bioinformation 11:454–459
74. Uddin A, Mazumder TH, Choudhury MN, Chakraborty S (2015) Codon bias and gene expression of mitochondrial ND2 gene in chordates. Bioinformation 11:407–412