Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Unravelling Codon Usage Bias in the VTE4 Gene governing γ- tocopherol methyltransferase in the Oilseed Brassica

**Vijay Kamal Meena**
Agriculture University Jodhpur
**Kapil Choudhary**
Agriculture University Jodhpur
**Subhash Chand**
subhashchand5415@gmail.com

Indian Grassland and Fodder Research Institute
**Prashant Yadav**
Directorate of Rapeseed-Mustard Research
**Bharath Kumar Alam**
Central Research Institute for Dryland Agriculture
**Ashish Sheera**
Sri Karan Narendra Agriculture University, Jobner
**Rakesh Choudhary**
Rani Lakshmi Bai Central Agricultural University
**Harsh Vardhan Singh Shekhawat**
Agriculture University Jodhpur
**Jitendar Kumar Sharma**
Agriculture University Jodhpur
**Lekha Meel**
Agriculture University Jodhpur
**Devendra Kumar Yadava**
Agriculture University Jodhpur

**Research Article**

**Keywords:** Brassicas, VTE4, α-tocopherol content, parity rule-2, neutrality plot, codon usage bias

**Posted Date:** May 4th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-9409680/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# Unravelling Codon Usage Bias in the *VTE4* Gene governing γ-tocopherol methyltransferase in the Oilseed Brassica

**Vijay Kamal Meena[1,3], Kapil Choudhary[1], Subhash Chand[2,3*], Prashant Yadav[4], Bharath Kumar Alam[5], Ashish Sheera[6], Rakesh Choudhary[7], Harsh Vardhan Singh Shekhawat[1], Jitendar Kumar Sharma[1], Lekha Meel[1], Devendra Kumar Yadava[3]**

[1]Agriculture University Jodhpur, Rajasthan, India
[2]ICAR- Indian Grassland and Fodder Research Institute, Jhansi, India
[3]Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi, India
[4]ICAR-Directorate of Rapeseed-Mustard Research, Bharatpur, Rajasthan, India;
[5]ICAR-Central Research Institute for Dryland Agriculture Hyderabad, Telangana, India
[6]Sri Karan Narendra Agriculture University, Jobner, Rajasthan, India
[7]Rani Lakshmi Bai Central Agricultural University, Jhansi, India

*** Correspondence:**
Corresponding Author
subhashchand5415@gmail.com

**Abstract**

Codon usage bias (CUB) is a phenomenon that shows variation both within the genes of a species and across different species, wherein particular codons are preferred over their synonymous counterparts. Genomic nucleotide configuration provides insights into the molecular basis of genes and their evolutionary relationships in distinct plant species. In this study, the *VTE4* gene, responsible for enhancing the nutritional quality and oil stability of oilseed *Brassica*, sequences from six *Brassica* species were extracted through the NCBI GenBank database. Subsequently, various CUB-related parameters, including nucleotide composition (AT and GC content), relative synonymous codon usage (RSCU), effective number of codons (ENC), frequency of optimal codons (Fop), relative codon usage bias (RCBS), neutrality plot (GC12 vs. GC3), parity rule-2 [(A3/(A3 + T3) vs. (G3/(G3 + C3)], and correspondence analysis, were assessed to analyse codon biasness within the U's triangle of Brassica species. The results showed AT bias across the *Brassica* species and moderate codon usage frequency for specific amino acids based on RSCU values in the *VTE4* gene. An evolutionary study confirmed a similarity in codon usage preference among species clustered together. The elevated ENC value, along with low Fop and RSCU values, indicated a low level of gene expression and a moderate bias in codon usage preference within the *Brassica* genus. Moreover, the results from the neutrality plot, parity rule, and correspondence analysis revealed that natural selection pressure had a more significant influence on shaping codon usage bias (CUB) in the *VTE4* gene than mutation pressure. This research contributes to understanding codon optimization, enhancing exogenous gene expression, and advancing transgenic engineering for improved α-tocopherol content profiling in Brassica species to enhance seed oil quality.

Keywords: Brassicas; *VTE4*; **α**-tocopherol content; parity rule-2; neutrality plot; codon usage bias

## 1 Introduction

The genetic code refers to the specific order of nucleotides found in both DNA and RNA molecules and plays a critical role in determination of the exact sequence of amino acids within a protein. Within genetic code, there are 61 codons responsible for encoding specific amino acids, and an additional 3 codons act as a translation termination signal. It is worth noting that majority of amino acids are represented by more than one codon, since only 20 different amino acids are needed for the creation of proteins [1–3]. For instance, Met and Trp each have a single codon, while some amino acids have varying numbers of codons: Phe, Tyr, Cys, His, Lys, Asn, Gln, Glu, and Asp have two codons each; Ile has three; Gly, Pro, Val, Ala, and Thr have four; and Ser, Leu, and Arg have six codons each[4, 5]. Synonymous codons represent identical amino acids but exhibit varying patterns of utilization across diverse species and even within the genes of a single species [6]. Codon usage bias (CUB) characterizes the non-uniform preference for specific synonymous codons within a genome. Nevertheless, the selection of synonymous codons for a given amino acid is not purely arbitrary in individual genes or species [7]. Moreover, certain codons exhibit higher frequency than their synonymous counterparts in genes, leading to bias in their usage. These frequently employed codons are referred to as preferred codons. Early research on CUB primarily focused on two genetic model systems: *Escherichia coli* and *Saccharomyces cerevisiae* [6, 8–10]. Furthermore, CUB is frequently regarded as encompassing the combined influences of mutation pressure, natural selection, and genetic drift [11, 12]. Moreover, CUB has proven to be highly effective in various applications, including exogenic gene expressions [13], elucidating the underlying causes of speciation [14], predicting gene functions [15], and estimating genetic divergence within species [16, 17]. These applications have been successfully applied to a range of organisms, spanning lower organisms like *E. coli* [15], *S. cerevisiae* [15], and DENV (dengue) virus [16] as well as higher plants such as *Arabidopsis thaliana*[18], *Brassica rapa*[19], and *Cucumis sativus* [20]. Hence, it is crucial to unravel the genetic differences between different species and enhance the effective expression of foreign genes in *Brassica* plants through CUB.

The *VTE4* (γ-tocopherol methyltransferase) gene plays a pivotal role for enhancing nutritional quality and oil stability of oilseed *Brassica* and act as a key genetic determinant for the oil quality improvement [21–26]. Like *FAE1* and *FAD2* genes [6, 27], *VTE4* has a significant impact on the fatty acid profile of Brassica oilseeds[28]. *VTE4* is essential for biosynthesis pathway of α-tocopherols, which are vital antioxidants and a crucial component of vitamin E in plant tissues [24, 29, 30]. Further, α-tocopherols contribute to the oxidative stability of oils, preventing the formation of undesirable compounds during storage and processing[31, 32] In oilseed *Brassica*, where the oil quality is of utmost importance, a higher expression of the *VTE4* gene leads to increased α-tocopherol levels in the oil [33, 34]. This not only enhances the oil's shelf life but also boosts its nutritional value, making it a more desirable choice for consumers[32, 35, 36] Therefore, the *VTE4* gene is a significant target for genetic improvement strategies in oilseed *Brassica*, contributing to its overall quality and market competitiveness in the oilseed industry.

In the current study, we have elucidated the pivotal role of the *VTE4* gene in oilseed *Brassica* species. Further, investigated codon bias and base composition dynamics associated with *VTE4* across a diverse set of *Brassica* species, including diploids like *B. rapa* (AA), *B. nigra* (BB), *B. oleracea* (CC), as well as amphidiploids like *B. juncea* (AABB), *B. carinata* (BBCC), and *B. napus* (AACC) and discussed the genetic factors contributing to vitamin E biosynthesis and overall plant health. Additionally, parameters such as GC content, ENC, Fop, codon bias index, amino acid length, protein hydropathicity, aromaticity score, RCBS, and RCA were studied to provide meaningful perspectives into the utilization patterns of the *VTE4* gene, its structural and functional attributes, and its evolutionary significance within the context of oilseed *Brassica* species. Research findings would help to enhance the varietal development program in oilseed *Brassica* with improved vitamin E content, thereby contributing to the diversification and nutritional quality of oilseed *Brassica* crops.

## 2. Materials and Methods

### *2.1. Sequence Data Source*

The coding sequences (CDS) of the VTE4 gene were retrieved exclusively from the Brassica Database (BRAD; http://brassicadb.cn/) using gene-specific identifiers. Corresponding CDS, genomic DNA, and protein sequences were obtained through the BRAD Gene Sequence retrieval system. All analysed sequences represent

annotated genes from publicly available reference genome assemblies of Brassica species, including *B. rapa, B. nigra, B. oleracea, B. juncea, B. napus, and B. carinata* (U's triangle). Gene annotations, species identification, and sequence validation were based on curated datasets available within the database. As this study relied entirely on publicly available genomic data, no plant materials were collected, and no experimental procedures, ethical approvals, or permissions were required. Further details of the analysed genes, including gene IDs, sequence length, and chromosomal positions, are provided in **Table 1**.

**Table 1.** *VTE4* gene coding sequences (CDS) in *Brassica* species: gene ID, genome, length, chromosomal position, and coded amino acid numbers.

| SN | Brassica species | Gene ID | Genome | Length (bp) | Chromosome | Amino Acid |
|---|---|---|---|---|---|---|
| **1** | *B. juncea* cv. Varuna | *BjuVA02G17730* | AABB | 1044 | A02 | 347 |
| **2** | *B. juncea* cv. Varuna | *BjuVA02G28700* | AABB | 1044 | A02 | 347 |
| **3** | *B. juncea* cv. Varuna | *BjuVB05G49330* | AABB | 1044 | B05 | 347 |
| **4** | *B. napus* cv. Darmor | *A02p23940_BnaDAR* | AACC | 1044 | A02 | 347 |
| **5** | *B. napus* cv. Darmor | *C02p21050_BnaDAR* | AACC | 1047 | C02 | 348 |
| **6** | *B. napus* cv. Darmor | *C02p37750_BnaDAR* | AACC | 1044 | C02 | 347 |
| **7** | *B. rapa* cv. Chiifu | *Bra008507* | AA | 927 | A02 | 308 |
| **8** | *B. nigra* cv. Sangam | *BniB032508_TA* | BB | 1044 | B03 | 347 |
| **9** | *B. oleracea* cv. HDEM | *BolC2t09793H* | CC | 1044 | C02 | 347 |
| **10** | *B. carinata* cv. sxm20200214 | *Bc_Predicted* | BBCC | 1059 | C03 | 352 |

### *2.2. Base Composition and Codon Preference*

The program, CUSP [39] was employed to determine the GC content (GC1, GC2, and GC3) at the first, second, and third positions within codons. Subsequently, the average GC1 and GC2 values within the coding sequences of both genes were combined to calculate GC12. Additionally, various codon usage metrics including CodonW1.4 [40] was used to calculate the following: the count of synonymous codons (L_sym), amino acid sequence length (L_aa), GRAVY index, aromaticity score (Aromo), Effective Number of Codons (ENC), Frequency of Optimal Codons (Fop values), Codon Bias Index (CBI), and Relative Synonymous Codon Usage (RSCU). . The nucleotide composition of GC3s, T3, C3, A3, and G3 at the third position of synonymous codons was also analyzed.

### *2.3. Relative Synonymous Codon Usage (RSCU)*

The theoretical observation value represents the point where the frequency of synonymous codon usage remains consistent, signifying no codon bias. Conversely, RSCU reflects the value between the actual observed frequency and the theoretical expectation. An RSCU of 1 indicates no codon bias, while an RSCU >1 suggests that a specific codon is used more often compared to its synonymous counterparts[41]

$$\mathrm{RSCU}_{ij} = \frac{\mathrm{X}_{ij}}{\frac{1}{n_{ij}} \sum_{j=1}^{ni} \mathrm{X}_{ij}}$$

Here, '$n_i$' represents the count of synonymous codons associated with the $i^{th}$ amino acid, and '$X_{ij}$' represents the frequency at which the $j^{th}$ codon occurs for the $i^{th}$ amino acid.

*2.4. ENC-GC3 Plot Analysis*

The ENC is a valuable metric for assessing CUB within genes and genomes. The ENC value falls within the range of 20 to 61. A gene with an ENC value of 20 demonstrates a strong bias, implying that it predominantly employs a single codon to represent each amino acid. Conversely, an ENC value of 61 suggests no bias, meaning that all synonymous codons are equally employed to encode a particular amino acid. In general, a lower ENC number signifies a more pronounced CUB [42]. Additionally, when the ENC score is 35 or lower, it is considered to indicate significant usage bias in the gene [42]. Another informative parameter, GC3s, measures the proportion of G and C nucleotides at the third position within codons, excluding those encoding Methionine (AUG) and Tryptophan (UGG). GC3s serve as a noteworthy indicator of nucleotide composition bias. To explore the sources of CUB variation, researchers utilize the ENC-GC3s plot. The predicted ENC value for each gene is calculated using Wright's formula [42]:

$$\mathrm{ENC} = 2 + \mathrm{GC3}\frac{29}{\mathrm{GC3}^2 + (1 - \mathrm{CG3})^2}$$

When G + C compositional restrictions are the sole driving factor, genes tend to align closely with or near the expected curve. However, if CUB is mainly driven by G + C compositional constraints but is also affected by additional factors unrelated to these constraints, the gene's position on the curve will be significantly lower than expected [43].

*2.5. PR2-Bias Plot*

The PR2-bias plot is a graphical representation using a two-dimensional chart. It utilizes A3/ (A3 + T3) on the horizontal axis and G3/ (G3 + C3) on the vertical axis to depict the nucleotide compositions found at the third position of codons. This information helps to assess the influence of mutation pressure and natural selection by examining the bias towards AT and GC content [44]. This plot becomes particularly meaningful when evaluating the amino acids encoded by four codon groups within a gene. These amino acids include Gly (GGA, GGU, GGG, GGC), Ala (GCA, GCU, GCG, GCC), Thr (ACA, ACU, ACG, ACC), Pro (CCA, CCU, CCG, CCC), Ser (UCA, UCU, UCG, UCC), Arg (CGA, CGU, CGG, CGC), Leu (CUA, CUU, CUG, CUC), and Val (GUA, GUU, GUG, GUC) [44]. The arrangement of points around the center of the plot, where A equals to T and G equals to C, indicates the extent and direction of deviation in base composition. When the mean values of A3/ (A3 + T3) and G3/ (G3 + C3) are greater than 0.5, it suggests a bias towards A and G, while values less than 0.5 indicate a bias towards T and C. Moreover, if A and T (or G and C) exhibit a proportional distribution, it implies that mutation pressure predominantly influences CUB. Conversely, if there is a disproportional distribution, it indicates the simultaneous impact of natural selection and other factors on CUB [44–46].

*2.6. Neutrality Plot*

The neutrality plot serves as a quantitative analytical approach for investigating how mutation pressure and natural selection impact CUB, with the results graphed against the GC12 and GC3 content, as determined using the method outlined by Sueoka [47]. A slope of 0 signifies no discernible influence of directional mutation pressure (indicating strong selective constraints), whereas a slope of 1 implies complete neutrality.

*2.7. Frequency of Optimal Codon (Fop)*

Optimal codons are specific codon sequences that exhibit the highest frequency of occurrence within a given species. Furthermore, Fop serves as a commonly employed metric for assessing gene CUB [48]. Fop, in this context, represents the proportion of optimal codons among all synonymous codons found in a particular species [48]. Fop values vary from 0.36, indicating genes with comparable CUB, to 1, signifying genes with a pronounced and distinct CUB [49].

$$Fop = \frac{1}{N}\sum_{i} syn(i)n_i$$

In this context, '$n_i$' represents the quantity of codons within the $i^{th}$ gene, 'N' denotes the overall count of codons in that gene, and 'syn(i)' signifies the quantity of codons that code for the same amino acid as codon 'i' [50].

*2.8. Evaluation and Analysis of Gene Expression*

RCBS measures the codon bias within a gene and plays a crucial role in evaluating gene expression. It can estimate codon usage bias even in the absence of a reference dataset [51]. To calculate the RCBS value for each sequence of the *VTE4* gene, the formula developed by Fox and Erill [52] was employed.

### *2.9. Clustering Based on Codon Usage Bias*

R software (version 4.3.1) was used to cluster six VTE4 genes [70] based on CUB. Relative codons were variables in the codon usage frequency analysis, while the gene was considered a factor. CUB analysis was performed using the RSCU values of 59 synonymous codons, removing the three stop codons (UAA, UAG, and UGA), as well as AUG (Met) and UGG (Trp).

### 2.10. Codon Adaptation Index (CAI)

CAI values serve as a common metric for evaluating the extent of preference for specific codons in highly expressed genes. These values range from 0 to 1, with higher values indicating a stronger bias towards codon usage and correspondingly higher gene expression levels. In this study, CAI values were calculated for genes encoding ribosomal proteins in *Brassica* species [42]. Ribosomal protein-coding genes from respective Brassica genomes were used as reference sets for CAI calculation, as these genes are generally considered constitutively and highly expressed. Reference codon frequencies were obtained from BrassicaDB and processed using CAIcal. CAI offers an estimate of gene expression levels by considering the idea that translational selection plays a role in optimizing gene sequences based on their expression levels. Additionally, aromaticity values in this context indicate the amount of aromatic amino acids (Trp, Tyr, and Phe ) within the resulting gen output.

### *2.11. Correspondence Analysis (COA)*

Correspondence analysis has found extensive application in the study of codon usage variation amongst the related genes [20, 53]. This analytical technique is employed to mitigate biases at the amino acid level, focusing on RSCU values for each codon. In this multidimensional approach, the genes under investigation are represented in a 59-dimensional space (excluding tryptophan, methionine, and stop codons), and the principal correspondence components are used to illustrate dissimilarities among orthologs [53, 54]. The analysis is based on a matrix of codon usages, which is then reduced to two dimensions: X (representing the total number of orthologs) and Y (representing codon usage values). Moreover, the visual output helps to eliminate extraneous noise, providing a simplified representation of complex data [48].

### *2.12. Statistical Analysis*

We employed R software to create a heat map displaying the Pearson correlation coefficient between codon usage, GC3 content, and RSCU values across various *Brassica* species. For this investigation, the hierarchical clustering method [55] was used. In addition, various software tools viz., Microsoft Excel, MEGA11 (phylogenetic analysis was performed using MEGA11 with the Neighbor-Joining method under the Maximum Composite Likelihood model. Bootstrap support was calculated using 1000 replications.), CUSP (for GC1 and GC2 analysis), CAIcal (for Codon Adaptation Index calculation) [56], and CodonW (http://codonw.sourceforge.net/) for correspondence analysis, GRAVY (Grand Average of Hydropathicity), Aromo analysis, and other relevant analyses [57] were utilized.

## **3. Results**

### *3.1. VTE4 Gene Codon Usage Biases in Brassica Species*

CUB provides insights into the preferential utilization of codons within the genome [58]. A heatmap was created to gain a clearer understanding of variations in codon usage and base composition, especially the GC content at the third base position (Figure 1). The analysis showed a strong positive correlation between GC3 and codons that end with C/T in the *VTE4* gene. In *VTE4*, the codons CTC, TCC, TGC, CGT, and TCT were the most commonly used, while codons ending with G/A (GTG and ACG) were also notably prevalent (**Figure 1**). Consequently, it is probable to suggest that GC3 preferences may influence codon usage in the *VTE4* gene.

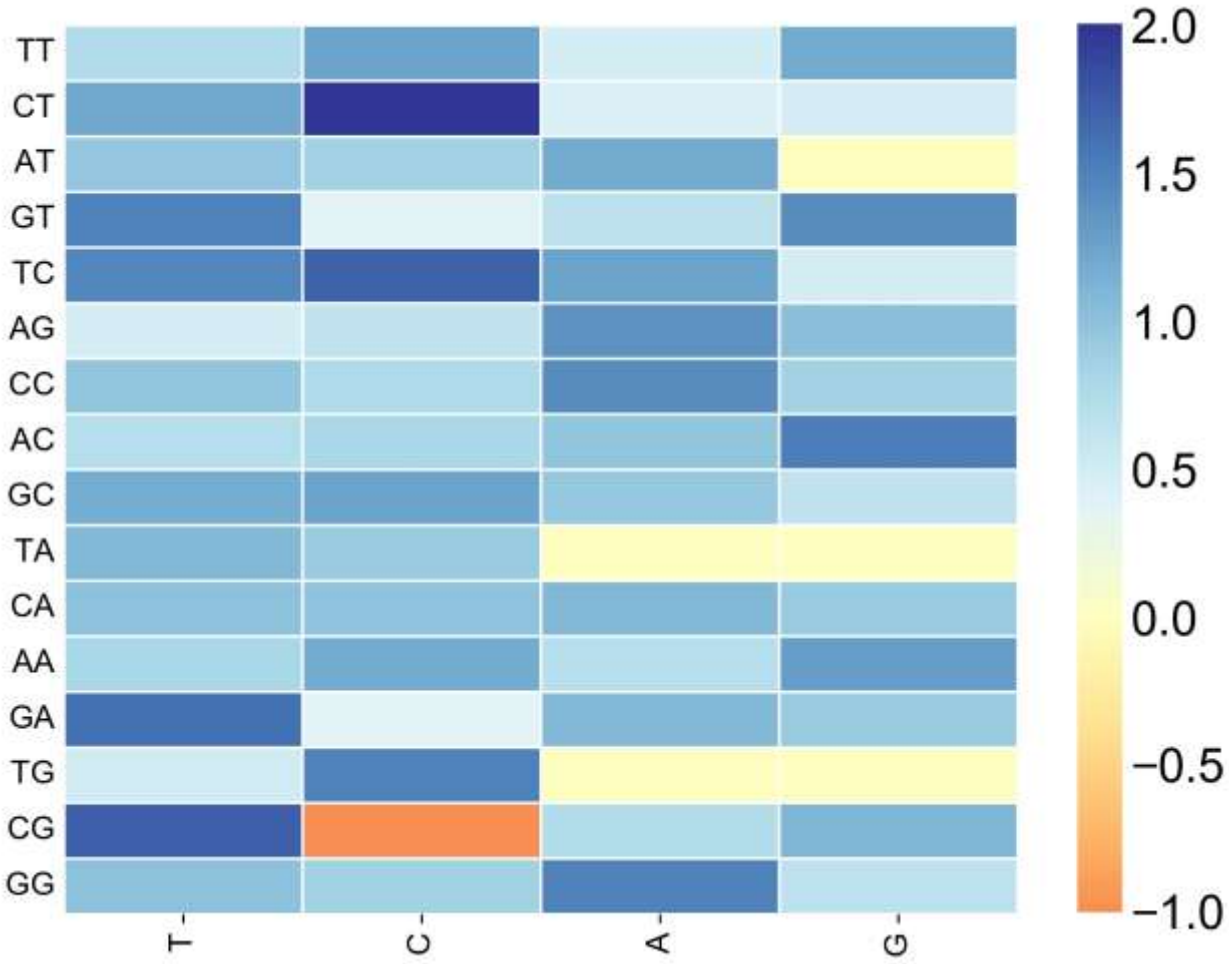



**Figure_1.** Heat map of codon values for the *VTE4* gene's third nucleotide position. The stop codons (TGA, TAG, TAA) and non-degenerate codons (ATG, TGG) that were totally absent from *VTE4* genes in all six *Brassica* species are represented by the white blocks. A strongly negative correlation is represented by the orange color, whereas a substantially positive correlation is indicated by the color dark blue.

### 3.2. *Preference of the G/C-Ended VTE4 Gene Codon in Brassica Species*

In this investigation, we analysed the coding sequence of *VTE4* gene in six *Brassica* species (**Table 2**). The results revealed that the nucleotide G had the highest average content (268.6), followed by A (263.6), T (257.5), and C (244.4) in *VTE4*. The average AT content was 50.40%, while the GC content was 49.60%. This indicates a relatively higher frequency of AT usage compared to GC in the *VTE4* gene coding sequences across the species. Specifically, the GC content at the codon's first and second nucleotide positions (GC12) was 49.36%, lower than AT12 (50.64%). Examining the third nucleotide position of the codon, G3 had the highest frequency (91.8), followed by C3 (87.5), A3 (84.5), and T3 (80.9). Interestingly, codons ending with G/C were slightly preferred over those ending with A/T in the *VTE4* gene coding sequence across these species. Of 18 optimal codons, nine ended with A/T, and the other nine ended with G/C. Notably, the codon CGC was not preferred (**Figure 2**).

**Table 2**. Nucleotide profiling of the *VTE4* gene's coding sequences in *Brassica* species.

| S.N.[A] | length | A | T | G | C | A3 | T3 | G3 | C3 | AT (%) | GC (%) | GC1 (%) | GC2 (%) | AT3 (%) | GC3 (%) | GC12 (%)[2^] | AT12 (%)[2^] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1.** | 1044 | 258 | 262 | 275 | 249 | 80 | 81 | 96 | 91 | 49.81 | 50.19 | 54.02 | 45.69 | 49.14 | 50.86 | 49.86 | 50.14 |
| **2.** | 1044 | 270 | 261 | 266 | 247 | 88 | 80 | 92 | 88 | 50.86 | 49.14 | 52.87 | 45.11 | 50.57 | 49.43 | 48.99 | 51.01 |
| **3.** | 1044 | 267 | 260 | 271 | 246 | 87 | 83 | 91 | 87 | 50.48 | 49.52 | 53.74 | 44.83 | 50.00 | 50.00 | 49.28 | 50.72 |
| **4.** | 1044 | 269 | 258 | 268 | 249 | 87 | 82 | 92 | 87 | 50.48 | 49.52 | 52.87 | 45.69 | 50.00 | 50.00 | 49.28 | 50.72 |
| **5.** | 1047 | 263 | 259 | 275 | 250 | 83 | 82 | 96 | 88 | 49.86 | 50.14 | 53.87 | 45.56 | 49.00 | 51.00 | 49.71 | 50.29 |
| **6.** | 1044 | 266 | 262 | 267 | 249 | 87 | 82 | 90 | 89 | 50.57 | 49.43 | 53.16 | 45.69 | 50.57 | 49.43 | 49.43 | 50.57 |
| **7.** | 927 | 241 | 232 | 251 | 203 | 74 | 69 | 83 | 83 | 51.02 | 48.98 | 53.07 | 44.66 | 50.81 | 49.19 | 48.87 | 51.13 |
| **8.** | 1044 | 268 | 260 | 270 | 246 | 87 | 83 | 91 | 87 | 50.57 | 49.43 | 53.45 | 44.83 | 50.00 | 50.00 | 49.14 | 50.86 |
| **9.** | 1044 | 267 | 262 | 266 | 249 | 88 | 82 | 89 | 89 | 50.67 | 49.33 | 53.16 | 45.69 | 50.86 | 49.14 | 49.43 | 50.57 |
| **10.** | 1059 | 267 | 259 | 277 | 256 | 84 | 85 | 98 | 86 | 49.67 | 50.33 | 53.82 | 45.33 | 48.16 | 51.84 | 49.58 | 50.42 |
| **Mean** | 1034.1 | 263.6 | 257.5 | 268.6 | 244.4 | 84.5 | 80.9 | 91.8 | 87.5 | 50.40 | 49.60 | 53.40 | 45.31 | 49.91 | 50.09 | 49.36 | 50.64 |
| **SD1^** | 35.98 | 8.20 | 8.61 | 6.97 | 14.06 | 4.27 | 4.16 | 4.04 | 2.01 | 0.44 | 0.44 | 0.41 | 0.40 | 0.84 | 0.84 | 0.29 | 0.29 |

A: confer with Table 4 to determine *Brassica* species' serial numbers; GC12 (%)2^: average of the percentage of GC content at the first and second nucleotide positions in the codon; AT12 (%)2^: average of the percentage of AT content at the first and second nucleotide positions in the codon; SD1^: standard deviation.

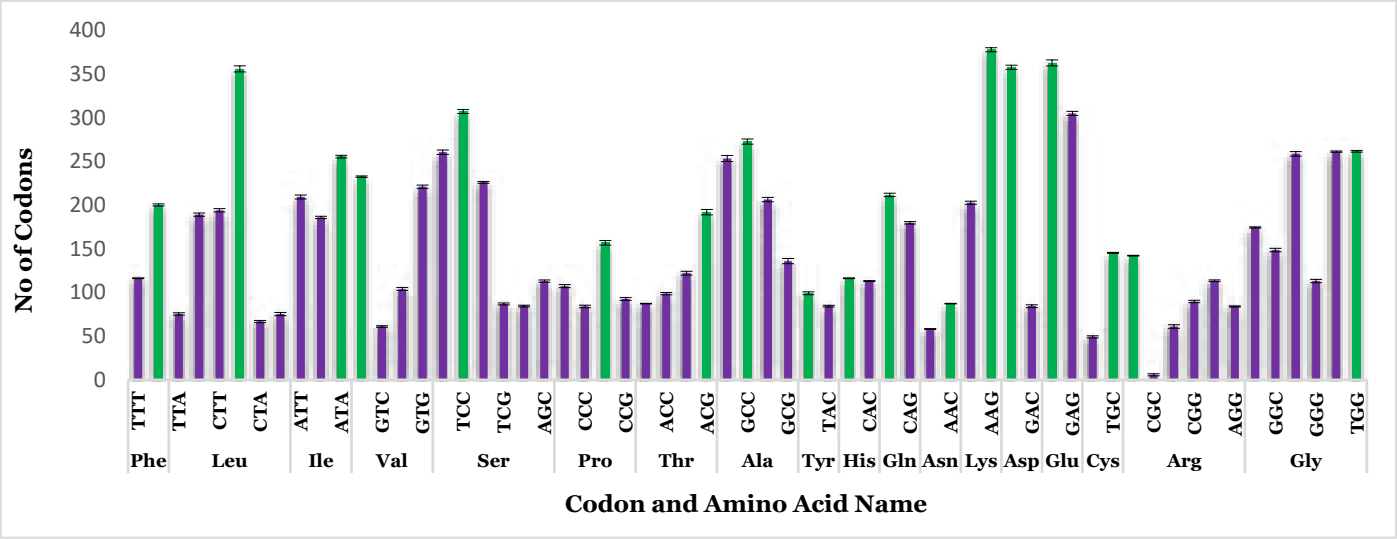



**Figure_2.** Comprehensive optimal and non-optimal codons frequency of *VTE4* gene used in *Brassica* species. Optimal and non-optimal codons with matching amino acids are represented by the colors green and blue, respectively.

### *3.3. Relative Synonymous Codon Usage of VTE4 Gene*

For the *VTE4* gene, an RSCU value above one indicates a high relative usage frequency, while a value greater than 1.60 signifies a strong preference. Additionally, 28 codons exhibited an RSCU value exceeding 1, demonstrating excessive per se frequency across various *Brassica* species. Of them, 15 codons showed a notable preference for those ending in A/T (**Table 3**). In contrast, codons ending in G/C (13) were almost equally preferred among the six *Brassica* species. The RSCU cluster analysis indicated that certain codons are favored in *VTE4* genes across these species (**Figure 3**). Specifically, codon TTG and AGA had RSCU values exceeding 1.6 in *Bra008507*, while CTC was consistent in all *Brassica* species. Moreover, TCC appeared prominently in *BjuVB05G49330, A02p23940_BnaDAR, C02p21050_BnaDAR, C02p37750_BnaDAR, BniB032508_TA, BolC2t09793H,* and CCA in *C02p37750_BnaDAR, Bra008507,* and *BolC2t09793H*. Additionally, the codon ACG showed a strong preference in *BjuVA02G17730, C02p21050_BnaDAR, Bra008507,* and *Bc_Predicted*, whereas GAT was prominent in all species except *Bc_Predicted*. Lastly, TGC was favoured in *BjuVA02G28700, A02p23940_BnaDAR*, and *Bra008507*; and CGT was prevalent in all species except *BjuVB05G49330* and *BniB032508_TA*. Notably, codon GGA displayed a strong preference in *C02p21050_BnaDAR* and *Bc_Predicted*.

**Table 3**. Complete relative synonymous codon usage (RSCU) for the *VTE4* gene among Brassica species.

| Amino acid | Codon | No [(A)] | RSCU [(B)] | Amino acid | Codon | No | RSCU |
|---|---|---|---|---|---|---|---|
| Phe | TTT | 40 | 0.7356 | Ala | GCT | 87 | 1.1651* |
| | TTC | 69 | 1.2644* | | GCC | 94 | 1.2561* |
| Leu | TTA | 26 | 0.4727 | | GCA | 71 | 0.9519 |
| | TTG | 65 | 1.1914* | | GCG | 47 | 0.6263 |
| | CTT | 67 | 1.2158* | Tyr | TAT | 34 | 1.0762* |
| | CTC | 123 | 2.228* | | TAC | 29 | 0.9238 |
| | CTA | 23 | 0.4189 | His | CAT | 40 | 1.0143* |
| | CTG | 26 | 0.4728 | | CAC | 39 | 0.9857 |
| Ile | ATT | 72 | 0.9641 | Gln | CAA | 73 | 1.0807* |
| | ATC | 64 | 0.8572 | | CAG | 62 | 0.9193 |
| | ATA | 88 | 1.1787* | Asn | AAT | 20 | 0.8 |
| Val | GTT | 80 | 1.504* | | AAC | 30 | 1.2* |
| | GTC | 21 | 0.3942 | Lys | AAA | 70 | 0.699 |
| | GTA | 36 | 0.6736 | | AAG | 130 | 1.301* |
| | GTG | 76 | 1.4285* | Asp | GAT | 123 | 1.6191* |
| Ser | TCT | 90 | 1.4469* | | GAC | 29 | 0.3809 |
| | TCC | 106 | 1.7097* | Glu | GAA | 125 | 1.0849* |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | TCA | 78 | 1.2591* | | GAG | 105 | 0.9151 |
| | TCG | 30 | 0.4829 | Cys | TGT | 17 | 0.4996 |
| | AGT | 29 | 0.4727 | | TGC | 50 | 1.5004* |
| | AGC | 39 | 0.6291 | Arg | CGT | 49 | 1.7239* |
| Pro | CCT | 37 | 0.9751 | | CGC | 2 | 0.0632 |
| | CCC | 29 | 0.7634 | | CGA | 21 | 0.7325 |
| | CCA | 54 | 1.4233* | | CGG | 31 | 1.0877* |
| | CCG | 32 | 0.8384 | | AGA | 39 | 1.3758* |
| Thr | ACT | 30 | 0.7022 | | AGG | 29 | 1.0171* |
| | ACC | 34 | 0.7886 | Gly | GGT | 60 | 1.0043* |
| | ACA | 42 | 0.9808 | | GGC | 51 | 0.8527 |
| | ACG | 66 | 1.5288* | | GGA | 89 | 1.4899* |
| | | | | | GGG | 39 | 0.653 |

No (A), a gene's overall number of favored codons. (B) Average RSCU values determined using synonymous codon use frequencies for a gene. *In the genome, codons with RSCU values (>1.0) are favored.

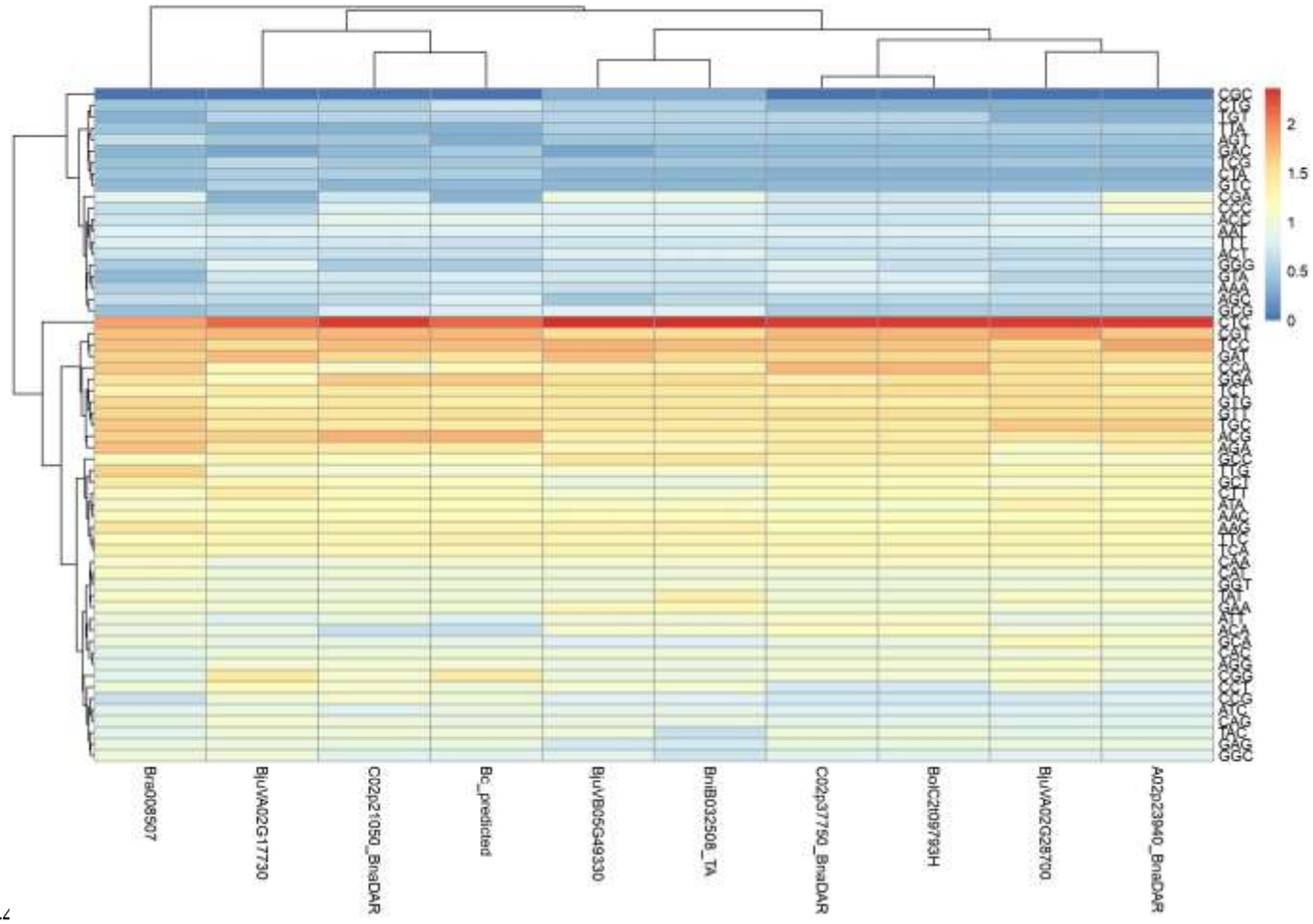



**Figure_3.** Grouping of each codon's relative synonymous codon usage (RSCU) values in the *VTE4* gene among *Brassica* species. The RSCU value of a codon (shown in rows) for every gene in each of the six *Brassica* species (represented in the column) is denoted by each block on the map. Different RSCU values are indicated by the color intensity on the block: light red to dark blue RSCU < 1.60 and intensity toward dark red RSCU > 1.60.

### *3.4. System Relationship of Codon Usage Patterns of VTE4 Gene*

The neighbor-joining method was used to construct neighboring evolutionary trees of various *Brassica* species (**Figure 4**). The findings emphasized that genes exhibiting closer evolutionary relationships among species displayed notably similar patterns in codon usage. The *BjuVA02G28700* gene was closer to *A02P23940BnaDAR*, while *C02P37750BnaDAR* was also nearer to *BolC2t09793H*. *BjuVB05G49330* was in closer proximity to *BniB032508TA* compared to *BjuVA02G17730*. *BC_Predicted* (*BrCarinata*) exhibited a closer association with *C02P21050BnaDAR*, while *Bra008507* was distinctly grouped in a separate clade. Codon usage patterns showed clustering broadly consistent with known Brassica genomic relationships; however, codon usage alone is insufficient for phylogenetic inference.

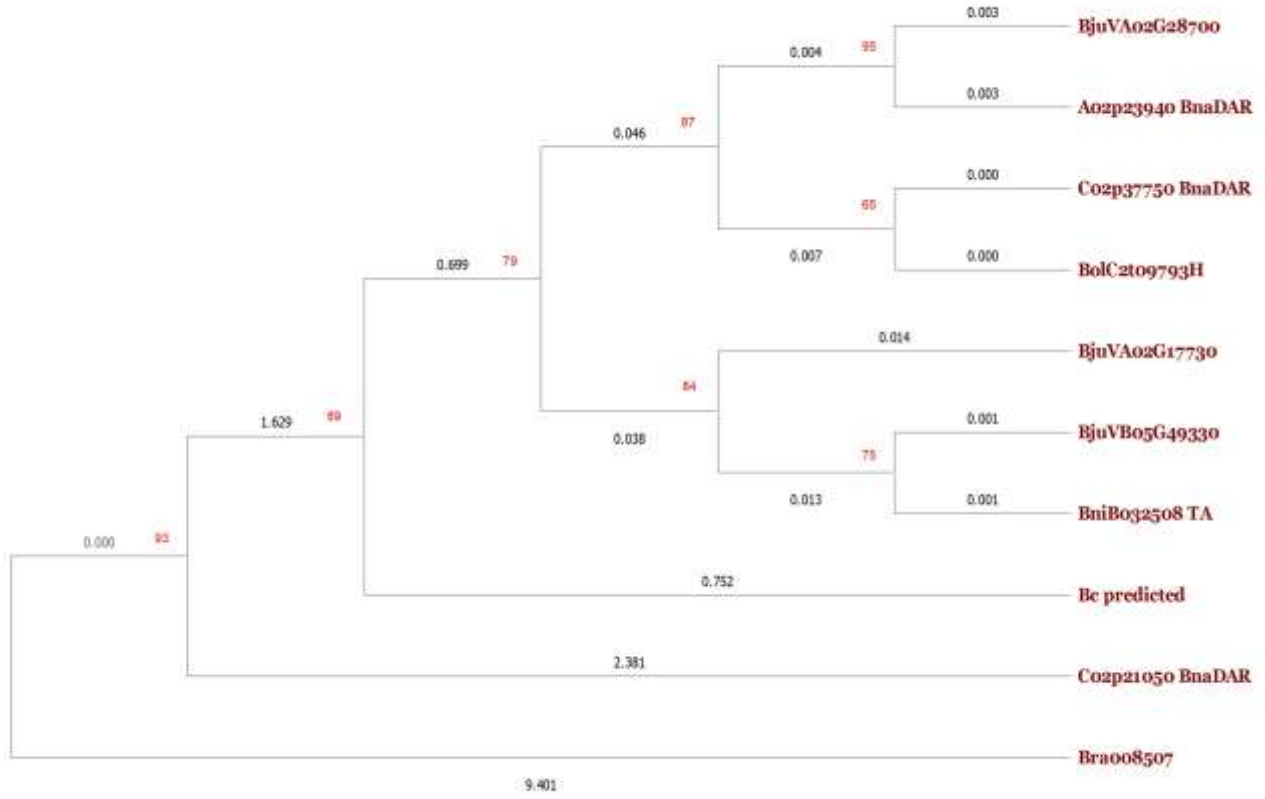



**Figure_4.** Phylogenetic tree based on 10 homologs of *VTE4* gene was constructed using the Neighbor-Joining (NJ) method, with variance estimation performed using bootstrap with 1000 replications. Numbers above branches are bootstrap values. Higher values indicate strong support for these nodes and vice-versa for lower values. The tree was inferred using the Maximum Composite Likelihood (MCL) model, providing insight into the evolutionary relationships among the analyzed species.

*3.5. Influence from Selection Pressure of VTE4 Gene in Different Brassica Species*

With an average of 59.875, the ENC value for the *VTE4* gene analysis ranged from 58.96 to 60.82, suggesting that there is not much diversity in codon use across *Brassica* species. The GC3 values varied from 0.464 to 0.492, averaging 0.475 **(Table 4)**. Fop and GC3 showed a substantial negative relationship (Pearson r = -0.6766, $p<0.05$). In contrast, Fop (r = -0.06838) and GC3 (r = 0.300203) both exhibited non-significant correlations with the ENC value (**Table 4, Figure 5**). The RCBS values ranged from 0.166 to 0.179, with an average of 0.1738 (Table 3; Figure 6). CAI values varied between 0.128 and 0.135, averaging 0.131, while RCA ranged from 0.477 to 0.502, with an average of 0.4895. Additionally, CAI values were between 0.207 and 0.215, with an average of 0.2143, and CBI values spanned from 0.024 to 0.054, averaging 0.0397. Notably, RCBS showed non-significant correlations with RCA, CAI, and CBI, whereas a significant positive correlation was found between RCA and CBI (r = 0.906253). The aromaticity values ranged from 0.0747 to 0.0844, with an average of 0.0764. The length of amino acids varied across *Brassica* species, from 308 (in *Bra008507*) to 352 (in *BrCarinata*), while the length of synonymous codons ranged from 291 (in *Bra008507*) to 333 (in *BrCarinata*). Notably, the RCBS values suggested that the *VTE4* gene, when expressed at lower levels, preferred specific codons.

280

281

282 **Table 4**. Codon uses bias indices for the *VTE4* gene across the six *Brassica* species.

| Title | RCBS | RCA | CAI | CBI | Fop | ENC | GC3 | GC | L_sym | L_aa | The Highest RSCU | Gravy | Aromo |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***BjuVA02G17730*** | 0.166 | 0.49 | 0.215 | 0.024 | 0.401 | 60.24 | 0.483 | 0.503 | 329 | 347 | CTC (L) | -0.07983 | 0.074928 |
| ***BjuVA02G28700*** | 0.177 | 0.48 | 0.209 | 0.029 | 0.407 | 58.96 | 0.468 | 0.493 | 329 | 347 | CGT (R) | -0.10144 | 0.07781 |
| ***BjuVB05G49330*** | 0.174 | 0.49 | 0.213 | 0.041 | 0.413 | 60.64 | 0.474 | 0.497 | 329 | 347 | TCC (S) | -0.10634 | 0.074928 |
| ***A02p23940_BnaDAR*** | 0.177 | 0.477 | 0.207 | 0.038 | 0.41 | 59.31 | 0.474 | 0.497 | 329 | 347 | GAT (D) | -0.14611 | 0.074928 |
| ***C02p21050_BnaDAR*** | 0.179 | 0.488 | 0.215 | 0.026 | 0.403 | 59.85 | 0.485 | 0.503 | 330 | 348 | ACG (T) | -0.08707 | 0.074713 |
| ***C02p37750_BnaDAR*** | 0.175 | 0.492 | 0.214 | 0.054 | 0.419 | 59.86 | 0.468 | 0.496 | 329 | 347 | GTT (V) | -0.10951 | 0.074928 |
| ***Bra008507*** | 0.175 | 0.496 | 0.223 | 0.051 | 0.426 | 59.56 | 0.464 | 0.491 | 291 | 308 | TGC (C) | -0.11526 | 0.084416 |
| ***BniB032508_TA*** | 0.169 | 0.488 | 0.212 | 0.041 | 0.413 | 60.82 | 0.474 | 0.496 | 329 | 347 | GGA (G) | -0.10749 | 0.074928 |
| ***BolC2t09793H*** | 0.177 | 0.492 | 0.214 | 0.054 | 0.419 | 59.62 | 0.465 | 0.495 | 329 | 347 | TCT (S) | -0.10951 | 0.074928 |
| ***BrCarinata*** | 0.169 | 0.502 | 0.221 | 0.039 | 0.411 | 59.89 | 0.492 | 0.505 | 333 | 352 | GTG (V) | -0.10909 | 0.076705 |

283 Relative codon bias usage (RCBS), relative codon adaptation (RCA), codon adaptation index (CAI), codon bias index (CBI), and optimal codon frequency (Fop) The effective number of codons is known
284 as ENC. L_aa—Amino acid length; L-sym—Number of synonymous codons; Aromo-Aromaticity score; GRAVY—Grand average of hydropathicity.

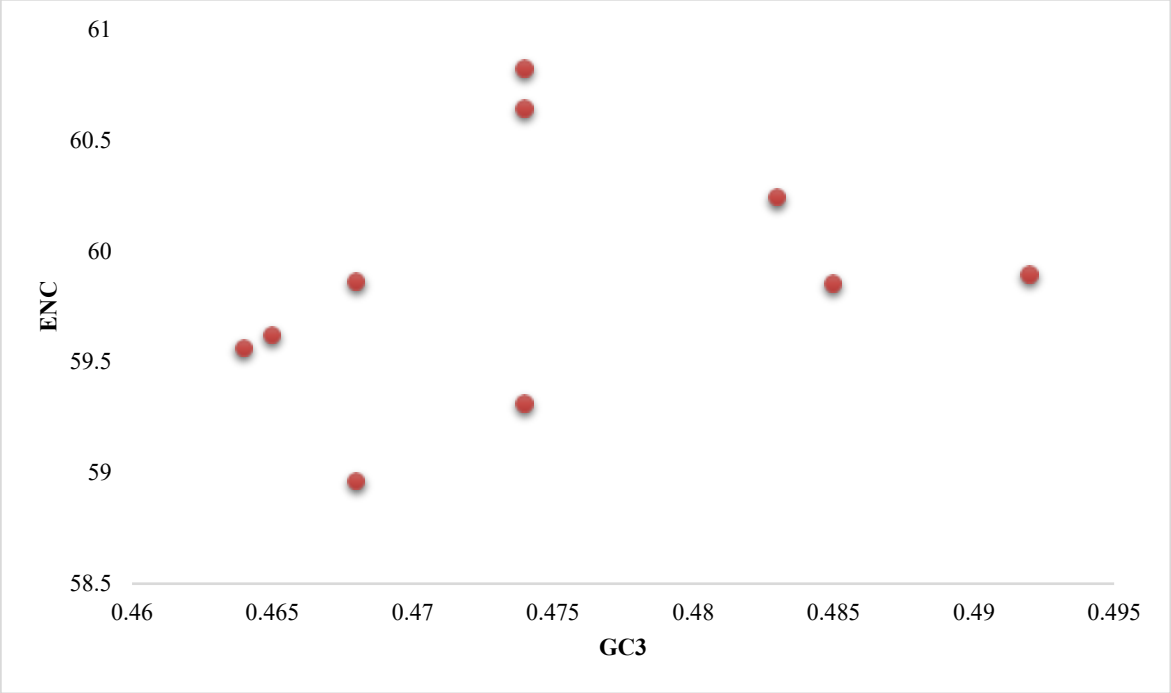



**Figure_5.** Relationship between ENC and GC3s values for *VTE4* gene among *Brassica* species. The ENC and GC3 values of the *VTE4* gene's coding sequences are shown by orange dots.

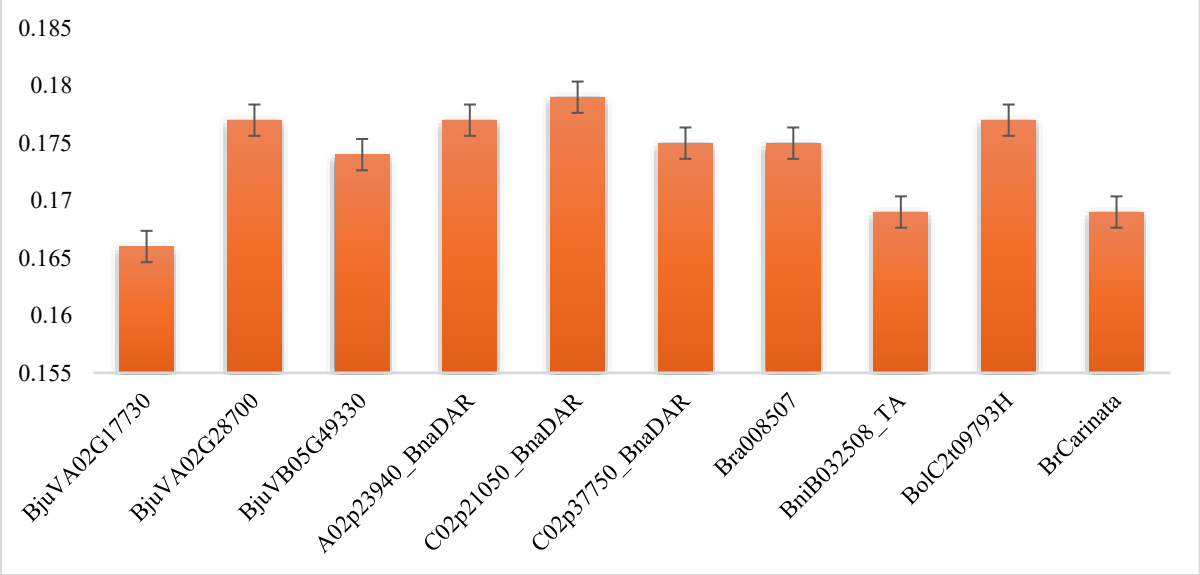



**Figure 6.** Allocation of relative synonymous codon usage (RCBS) for *VTE4* genes across six *Brassica* species.

*3.6. Neutrality Plot*

Neutrality plot analysis was investigated to study selection pressure and mutational impact on codon usage in the genome. The neutrality plot was constructed by plotting the GC content at the first and second positions of codons (GC12) against the GC content at the third position (GC3). Gene *VTE4* exhibited a significant positive correlation ($p$<0.05) with a Pearson correlation coefficient (r) of 0.7096 and an R-squared value ($R^2$) of 0.5036. The regression slope between GC12 and GC3 was moderate, measuring 0.2346. The moderate slope (0.2346) suggests that both mutation pressure and selective constraints contribute to codon usage variation in *VTE4*. Biased gene conversion may also influence GC content variation

and should be considered as a potential confounding evolutionary force. Although selection appears to play a role, mutation bias cannot be excluded as a contributing factor.3 (**Figure 7**).

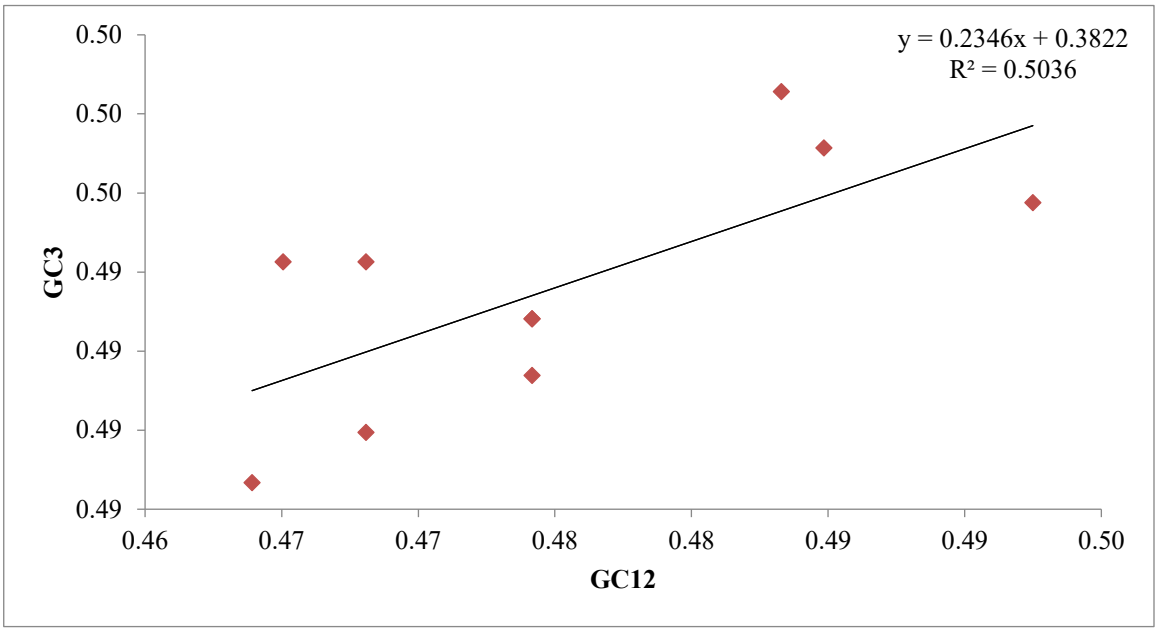



**Figure_7.** Neutrality plot (GC12 vs. GC3) for *VTE4* gene. The data points are represented as orange diamonds, and the black line represents the linear regression fit to the data.

*3.7. PR-2 Bias Plot*

The genetic makeup of the genome is significantly influenced by mutation and selection pressure. In the PR-2 bias plot, if the values accumulate around the midpoint (0.5), it suggests the absence of any notable bias or a dominant occurrence of mutations. In this study, the *VTE4* gene exhibited a much closer positioning to the center, indicating that natural selection has a more prominent influence on codon usage in *VTE4* than mutation pressure (**Figure 8**). It also underscores a biased utilization of A/G at synonymous codon positions within this gene across the *Brassica* species.

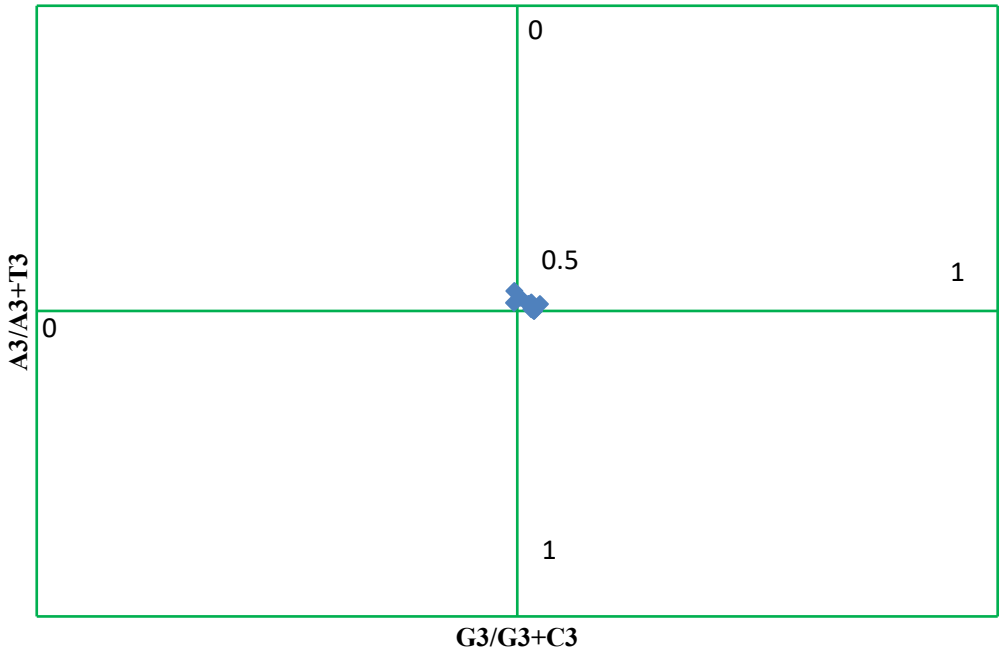


**Figure 8.** PR2 (parity rule 2) bias plot for *VTE4* gene.

### *3.8. Correspondence Analysis*

Correspondence analysis (COA) was employed to explore the variation in codon usage among species within the U's category for the *VTE4* gene. Notably, the first principal axis accounted for a substantial 39.15% of the overall variance, while the second principal axis explained a lesser 23.55% of the variance (**Figure 9A**). The *VTE4* orthologs were primarily categorized into two distinct groups along the horizontal axis (**Figure 9B**). Notably, two orthologs, namely *BjuVB05G49330* and *BniB032508_TA*, were positioned at the far right of axis-1 and displayed a pronounced codon bias, while the remaining orthologs were situated to the left (both upper and lower) of axis-1, indicating a weaker bias in their codon usage. Moreover, the COA analysis revealed an interesting pattern with respect to the nucleotide composition of codons. Codons ending with GC were clustered near the origin, whereas codons ending with AT were more widely dispersed from the origin (**Figure 9C**).

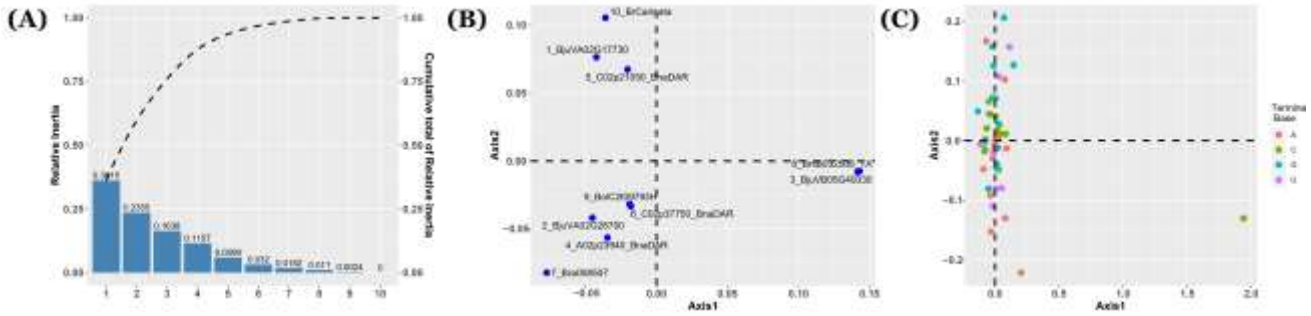



**Figure 9.** In (A), various primary axes are revealed by correspondence analysis using RSCU value. The cumulative sum of the inertia described by the nine axes is indicated by the dotted line. In (B), the RSCU values serve as the basis for the ortholog distribution. In (C), where A, U, G, and C were colored differently in the correspondence analysis that showed the distribution of synonymous codons that terminated with distinct nucleotides.

## 4. Discussion

The genomic CUB is influenced by various genomic components, including genetic code, length of gene, recombination frequency, GC content, and gene expression levels [51, 53]. Additionally, codon bias can be shaped by factors such as mutation rates, selection pressures, replication and transcription and translation preferences, and hydrophobicity of proteins [59, 60]. Both mutation and selection significantly influence codon usage, spanning from unicellular prokaryotes to multicellular eukaryotes [61, 62]. In the case of multicellular eukaryotic model organisms like *Caenorhabditis elegans* and *Drosophila melanogaster*, selection pressures play a key role in determining codon usage [63]. Conversely, in viruses such as the Parvoviridae family, codon usage is influenced by both mutation and selection pressure [64]. Each gene exhibits distinct codon frequencies in the realm of plants, even for the same amino acid. The nucleotide composition within codons further impacts CUB [65]. The interaction between mutation and reverse mutation frequencies in codon base composition influences nucleotide preferences in genomic DNA [66]. Generally, mutations at the third nucleotide position of codons do not alter the corresponding amino acid during protein synthesis, and the GC content is essential for indicating mutational trends. As a result, the third nucleotide of codons is better equipped to withstand mutational pressures, making GC3 a useful parameter for examining codon usage bias (CUB). Although the third codon position is often considered relatively neutral, synonymous sites may be subject to translational selection, mRNA stability constraints, and regulatory influences, which limit strict neutrality assumptions.

In this investigation, the termination of most codons within the U's triangle *Brassica* species was characterized by ending with G/C. The overall AT content in the *VTE4* gene surpassed the GC content across the various species. While overall nucleotide composition showed slight AT predominance, specific codon-level preferences revealed enrichment of certain GC-ending codons. This indicates that local codon preference may not strictly mirror overall base composition, reflecting interplay between compositional bias and translational selection. Prior studies have consistently reported a

prevalence of A/T endings in codons, and researchers have previously disclosed an elevated AT to-GC-content ratio in the genomes of plant species across different genes[59, 67] Likely, *Fop* analysis revealed a predominance of A/T termination in optimal codons within the *VTE4* gene. Of 59 synonymous codons, 28 exhibited RSCU values exceeding one in *VTE4*, indicating a high frequency of utilization across six *Brassica* species. Orthologs originating from the A genomes of *B. juncea* and *B. napus* exhibited greater similarity to the C genome of *B. napus* and *B. oleracea*. In contrast, orthologs of the B genome in *B. juncea* were found to be more closely related to the B genome of *B. nigra*. The A genomes of *B. rapa* and *B. carinata* displayed the greatest dissimilarity compared to all other orthologs. Previous research has indicated that genes sharing similar functions exhibit analogous patterns in codon usage[59, 68] Consequently, all the examined *Brassica* genotypes displayed a distinctive interrelationship among themselves regarding the synonymous codon usage patterns for the *VTE4* gene.

A higher ENC value indicates a lower gene expression level in *Brassica* species. In this investigation, all six *Brassica* species exhibited an ENC value exceeding 58, suggesting a reduced gene expression level. Codon usage preferences vary among different crop species, and individual genes within a plant species exhibit distinct codon preferences[59] The study highlighted a significant negative correlation between ENC and GC3, as well as between ENC and Fop, while a positive correlation was found between Fop and GC3. These results contrast with earlier studies[59, 69] which reported a positive association between ENC and GC3, emphasizing that codon usage bias is influenced by the third nucleotide position, particularly G/C, in the genome. Additionally, the Fop values indicated low levels of codon usage bias in the *VTE4* gene across various *Brassica* species.

The RCBS values of the gene serve as a metric for assessing the effective index of gene expression, primarily influenced by the preference for gene base composition [42]. In this study, the *VTE4* gene exhibited a decreased RCBS value, suggesting reduced gene expression associated with lower codon preference. These results align with earlier research that also documented reduced RCBS values in the *AP2* gene[59]The CAI quantifies how well the codon usage in a gene aligns with the codon usage observed in highly expressed genes. In other words, CAI serves as a predictor for the expression level of a gene[41]In this investigation, the CAI values were identified to be low (<0.223). The CBI assesses the degree to which a gene employs a specific set of optimal codons [70]. Our results demonstrated a positive correlation between CAI and RCA, while CBI exhibited a positive correlation with Fop. The observed low CAI and CBI values for the *VTE4* gene indicate a diminished CUB across species, a factor critical in influencing gene expression levels. Similar low CAI and CBI values were observed in *A. ceylanicum*, underscoring the importance of optimal codon usage for higher gene expression [71].

The Aromaticity score reflects the frequency of aromatic amino acids in the hypothetical translated gene product and serves as an indicator of amino acid usage [72]. Differences in the amino acid composition of the gene in question may lead to variations in aromaticity. Aromaticity exhibited a negative correlation with amino acid length and the number of synonymous codons, emphasizing the influence of aromaticity on the *VTE4* gene's protein characteristics across six *Brassica* species. The neutrality plot provided conclusive evidence that selection pressure and mutation are pivotal factors influencing the preferential use of codons in the *VTE4* gene. Selection pressure plays a pivotal role when the correlation between GC12 and GC3 tends to be low and the regression slope approaches zero. Conversely, for mutation-driven dynamics, the correlation is expected to be high, and the regression slope differs from zero [73]. Previous studies reported a minimal correlation and regression slope between GC12 and GC3 in *A. ceylanicum* and *Brassica* genes, indicating that mutation significantly affected CUB in the genome [6, 71] Conversely, mitochondrial genes have shown a higher correlation [74]. It is important to emphasize that codon usage indices such as CAI, ENC, RCBS, and Fop primarily reflect synonymous codon preference and potential translational adaptation rather than direct measurements of transcriptional expression levels. Similarly, protein-level parameters including aromaticity and GRAVY represent intrinsic physicochemical properties of the encoded protein and do not directly determine gene expression. However, amino acid composition may impose structural, functional, or metabolic constraints that indirectly influence translational efficiency and codon selection over evolutionary time. Therefore, observed associations between codon usage metrics and protein characteristics should be interpreted as indirect evolutionary correlations rather than causal determinants of gene expression.

Selection pressure or mutational constraints can lead to a preferential utilization of nucleotides in synonymous codon positions [75]. In this study, the PR2-bias plot illustrated an imbalance between the proportions of A and T, as well as in G and C, at the third nucleotide position of codons within the gene. Notably, the prevalence of T3 and C3 exceeded A3 and G3. Additionally, the distribution of AT and GC was not uniform around the center, suggesting that factors such as mutation

and selection pressure significantly influenced CUB in genes. Comparable observations have been documented in *Brassica* species [6]. The COA analysis effectively differentiated the examined genes based on their codon usage variation and also shed light on the impact of nucleotide composition at each codon. Two notable orthologs, specifically *BjuVB05G49330* and *BniB032508_TA*, exhibited a prominent codon bias as they were positioned at the far right of axis-1. In contrast, the remaining orthologs on both the upper and lower sides of axis-1 displayed a weaker bias in codon usage. Additionally, the COA analysis unveiled an intriguing pattern regarding the nucleotide composition of codons. Codons terminating with GC were closely clustered around the origin, while those ending with AT were dispersed farther from the origin. This observation highlights the significant role of mutations and other factors in influencing the codon usage bias for the *VTE4* gene, aligning with similar findings reported in *A. ceylanium* genes, *Mycobacterium tuberculosis* coding genome, *Jatropha curcas* genes and *Brassica FAE* and *FAD* genes [6, 71, 73, 76].

## 5. Conclusions

Codon usage bias (CUB) is profoundly determined by both mutational and natural selection pressures, influencing the arrangement of nucleotide blueprint throughout the genomic DNA of living species. Understanding CUB is crucial for unraveling the gene evolution mechanisms, functional adaptability, and survival tactics. Our investigation focused on the *VTE4* gene, revealing an AT bias and moderate GC bias across six *Brassica* species. The study indicates that both selection pressure and mutation play pivotal roles in shaping CUB for the *VTE4* gene. Previous studies in plant species have also highlighted the impact of selection pressure and mutation on codon usage patterns, shedding light on the evolutionary relationships among genes based on CUB. Codon usage patterns showed clustering broadly consistent with known Brassica genomic relationships; however, codon usage alone is insufficient for phylogenetic inference. As this study focuses on a single gene (*VTE4*), the findings may refaromalect gene-specific evolutionary patterns rather than genome-wide trends across Brassica species. Furthermore, the study suggests that *Brassica* species's seed oil quality can be improved through the presence of natural mutations promoting high tocopherol levels at the genomic level. Manipulating codon usage in oilseed *Brassica* breeding holds promise for developing high-quality varieties tailored to market demand, thereby promoting sustainable diversification in oilseed crops.

## Declarations

### Abbreviations

CUB: Codon Usage Bias
RSCU: Relative Synonymous Codon Usage
ENC: Effective Number of Codons
Fop: Frequency of Optimal Codons
RCBS: Relative Codon Bias Score
CAI: Codon Adaptation Index
CBI: Codon Bias Index
COA: Correspondence Analysis

### Ethics approval and consent to participate

Not applicable. This study did not involve human participants, animals, or direct plant sampling.

### Consent for publication

Not applicable.

### Availability of data and materials

The datasets analysed during the current study are publicly available in the Brassica Database (BRAD; http://brassicadb.cn/). All VTE4 gene sequences were retrieved using gene-specific identifiers (listed in Table 1) through the BRAD Gene Sequence portal (http://brassicadb.cn/#/GeneSequence/). The corresponding coding sequences (CDS), genomic DNA, and protein sequences can be accessed by querying these gene IDs. These sequences are derived from publicly available annotated genome assemblies and do not have individual GenBank accession numbers. No new sequencing data were generated in this study.

### Competing Interests

The authors declare that they have no competing interests.

**Funding**

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

**Authors' contributions**

**Vijay Kamal Meena**: Conceptualization, data analysis, manuscript writing

**Kapil Choudhary**: Data curation and analysis

**Subhash Chand**: Supervision and manuscript editing

**Prashant Yadav, Bharath Kumar Alam, Ashish Sheera**: Data validation and interpretation

**Rakesh Choudhary, Harsh Vardhan Singh Shekhawat, Jitendar Kumar Sharma, Lekha Meel**: Technical support and review

**Devendra Kumar Yadava**: Overall guidance and critical revision

All authors read and approved the final manuscript.

**Acknowledgements**

The first author is grateful to the NCBI for accessing the genomic data and to Agriculture University, Jodhpur for providing scientific inputs during the research.

**Reference**

1. Smolskaya S, Andreev Y. Site-Specific Incorporation of Unnatural Amino Acids into Escherichia coli Recombinant Protein: Methodology Development and Recent Achievement. Biomolecules. 2019;9:255. https://doi.org/10.3390/biom9070255.

2. Bailey SF, Alonso Morales LA, Kassen R. Effects of Synonymous Mutations beyond Codon Bias: The Evidence for Adaptive Synonymous Substitutions from Microbial Evolution Experiments. Genome Biol Evol. 2021;13. https://doi.org/10.1093/gbe/evab141.

3. Lateef OM, Akintubosun MO, Olaoba OT, Samson SO, Adamczyk M. Making Sense of “Nonsense” and More: Challenges and Opportunities in the Genetic Code Expansion, in the World of tRNA Modifications. Int J Mol Sci. 2022;23:938. https://doi.org/10.3390/ijms23020938.

4. Bailey J. Nucleosides, Nucleotides, Polynucleotides (RNA and DNA) and the Genetic Code. . In: In Inventive Geniuses Who Changed the World. Cham, Switzerland: Springer; 2022.

5. Iriarte A, Lamolle G, Musto H. Codon Usage Bias: An Endless Tale. J Mol Evol. 2021;89:589–93. https://doi.org/10.1007/s00239-021-10027-z.

6. Chaudhary R, Chand S, Alam BK, Yadav P, Meena VK, Patel MK, et al. Codon Usage Bias for Fatty Acid Genes FAE1 and FAD2 in Oilseed Brassica Species. Sustainability. 2022;14:11035. https://doi.org/10.3390/su141711035.

7. Jou WM, Haegeman G, Ysedaert M, Fiers W. Nucleotide Sequence of the Gene Coding for the Bacteriophage MS2 Coat Protein. Nature. 1972;237:82–8. https://doi.org/10.1038/237082a0.

8. Grantham R, Gautier C, Gouy M, Jacobzone M, Mercier R. Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res. 1981;9:213–213. https://doi.org/10.1093/nar/9.1.213-b.

9. Plotkin JB, Robins H, Levine AJ. Tissue-specific codon usage and the expression of human genes. Proceedings of the National Academy of Sciences. 2004;101:12588–91. https://doi.org/10.1073/pnas.0404957101.

10. Begum NS, Chakraborty S. Influencing elements of codon usage bias in Birnaviridae and its evolutionary analysis. Virus Res. 2022;310:198672. https://doi.org/10.1016/j.virusres.2021.198672.

11. Zhang Z, Li J, Cui P, Ding F, Li A, Townsend JP, et al. Codon Deviation Coefficient: a novel measure for estimating codon usage bias and its statistical significance. BMC Bioinformatics. 2012;13:43. https://doi.org/10.1186/1471-2105-13-43.

12. Hershberg R, Petrov DA. Selection on Codon Bias. Annu Rev Genet. 2008;42:287–99. https://doi.org/10.1146/annurev.genet.42.110807.091442.

13. Kane JF. Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. Curr Opin Biotechnol. 1995;6:494–500. https://doi.org/10.1016/0958-1669(95)80082-4.

14. Ahn I, Jeong B-J, Bae S-E, Jung J, Son HS. Genomic Analysis of Influenza A Viruses, including Avian Flu (H5N1) Strains. Eur J Epidemiol. 2006;21:511–9. https://doi.org/10.1007/s10654-006-9031-z.

15. Lin K. Conserved codon composition of ribosomal protein coding genes in Escherichia coli, Mycobacterium tuberculosis and Saccharomyces cerevisiae: lessons from supervised machine learning in functional genomics. Nucleic Acids Res. 2002;30:2599–607. https://doi.org/10.1093/nar/30.11.2599.

16. Zhou M, Guo J, Cha J, Chae M, Chen S, Barral JM, et al. Non-optimal codon usage affects expression, structure and function of clock protein FRQ. Nature. 2013;495:111–5. https://doi.org/10.1038/nature11833.

17. Zhou J, Li X, Lan X, Han S-Y, Wang Y, Hu Y, et al. The genetic divergences of codon usage shed new lights on transmission of hepatitis E virus from swine to human. Infection, Genetics and Evolution. 2019;68:23–9. https://doi.org/10.1016/j.meegid.2018.11.024.

18. Chiapello H, Lisacek F, Caboche M, Alain Hénaut. Codon usage and gene function are related in sequences of Arabidopsis thaliana. Gene. 1998;209:GC1–38. https://doi.org/10.1016/S0378-1119(97)00671-9.

19. Srivastava S, Chanyal S, Dubey A, Tewari AK, Taj G. Patterns of Codon Usage Bias in WRKY Genes of Brassica rapa and Arabidopsis thaliana. Journal of Agricultural Science. 2019;11:76. https://doi.org/10.5539/jas.v11n4p76.

20. Nie X, Deng P, Feng K, Liu P, Du X, You FM, et al. Comparative analysis of codon usage patterns in chloroplast genomes of the Asteraceae family. Plant Mol Biol Report. 2014;32:828–40. https://doi.org/10.1007/s11105-013-0691-z.

21. Fritsche S, Wang X, Jung C. Recent advances in our understanding of tocopherol biosynthesis in plants: An overview of key genes, functions, and breeding of vitamin E improved crops. Antioxidants. 2017;6. https://doi.org/10.3390/antiox6040099.

22. Fritsche S, Wang X, Nichelmann L, Suppanz I, Hadenfeldt S, Endrigkeit J, et al. Genetic and functional analysis of tocopherol biosynthesis pathway genes from rapeseed (Brassica napus L.). Plant Breeding. 2014;133:470–9. https://doi.org/10.1111/pbr.12190.

23. Goffman FD, Becker HC. Genetic variation of tocopherol content in a germplasm collection of Brassica napus L. Euphytica. 2002;125:189–96. https://doi.org/10.1023/A:1015858618480.

24. Endrigkeit J, Wang X, Cai D, Zhang C, Long Y, Meng J, et al. Genetic mapping, cloning, and functional characterization of the BnaX.VTE4 gene encoding a γ-tocopherol methyltransferase from oilseed rape. Theoretical and Applied Genetics. 2009;119:567–75. https://doi.org/10.1007/s00122-009-1066-6.

25. Meena VK, Taak Y, Chaudhary R, Chand S, Patel MK, Muthusamy V, et al. Deciphering the Genetic Inheritance of Tocopherols in Indian Mustard (Brassica juncea L. Czern and Coss). Plants. 2022;11:1779.

26. Meena VK, Taak Y, Chaudhary R, Chand S, Patel MK, Saroj R, et al. Development of functional markers for high α-tocopherol content in Indian mustard [Brassica juncea (L.) Czern. and Coss] employing genetic and molecular analysis of VTE4 gene. Indian Journal of Genetics and Plant Breeding. 2024;84:551-`560. https://doi.org/10.31742/ISGPB.84.4.5.

27. Saini N, Singh N, Kumar A, Vihan N, Yadav S, Vasudev S, et al. Development and validation of functional CAPS markers for the FAE genes in. Breed Sci. 2016;66:831–7. https://doi.org/10.1270/jsbbs.16132.

28. Rauf S, Fatima S, Ortiz R. Modification of Fatty Acid Profile and Oil Contents Using Gene Editing in Oilseed Crops for a Changing Climate. GM Crops Food. 2023;:1–12. https://doi.org/10.1080/21645698.2023.2243041.

29. Fritsche S, Wang X, Nichelmann L, Suppanz I, Hadenfeldt S, Endrigkeit J, et al. Genetic and functional analysis of tocopherol biosynthesis pathway genes from rapeseed (Brassica napus L.). Plant Breeding. 2014;133:470–9. https://doi.org/10.1111/pbr.12190.

30. Wang X, Zhang C, Li L, Fritsche S, Endrigkeit J, Zhang W, et al. Unraveling the Genetic Basis of Seed Tocopherol Content and Composition in Rapeseed (Brassica napus L.). PLoS One. 2012;7:8–14. https://doi.org/10.1371/journal.pone.0050038.

31. Shahidi F, de Camargo A. Tocopherols and Tocotrienols in Common and Emerging Dietary Sources: Occurrence, Applications, and Health Benefits. Int J Mol Sci. 2016;17:1745. https://doi.org/10.3390/ijms17101745.

32. Ali E, Hussain S, Hussain N, Kakar KU, Shah JM, Zaidi SHR, et al. Tocopherol as plant protector: an overview of Tocopherol biosynthesis enzymes and their role as antioxidant and signaling molecules. Acta Physiol Plant. 2022;44:20. https://doi.org/10.1007/s11738-021-03350-x.

33. Niu Y, Zhang Q, Wang J, Li Y, Wang X, Bao Y. Vitamin E synthesis and response in plants. Front Plant Sci. 2022;13:994058.

34. Zhang H, Shi Y, Sun M, Hu X, Hao M, Shu Y, et al. Functional Differentiation of BnVTE4 Gene Homologous Copies in α-Tocopherol Biosynthesis Revealed by CRISPR/Cas9 Editing. Front Plant Sci. 2022;13. https://doi.org/10.3389/fpls.2022.850924.

35. Yusuf MA, Gupta PK, Sarin NB. Alpha-tocopherol value added Brassica juncea : fortified human nutrition and abiotic stress tolerance to the plant. 1993;1301:5–8.

36. Liao P, Chen X, Wang M, Bach TJ, Chye ML. Improved fruit α-tocopherol, carotenoid, squalene and phytosterol contents through manipulation of Brassica juncea 3-HYDROXY-3-METHYLGLUTARYL-COA SYNTHASE1 in transgenic tomato. Plant Biotechnol J. 2018;16:784–96. https://doi.org/10.1111/pbi.12828.

37. Yim WC, Swain ML, Ma D, An H, Bird KA, Curdie DD, et al. The final piece of the Triangle of U: Evolution of the tetraploid Brassica carinata genome. Plant Cell. 2022;34:4143–72. https://doi.org/10.1093/plcell/koac249.

38. Chand S, Patidar OP, Chaudhary R, Saroj R, Chandra K, Meena VK, et al. Rapeseed-Mustard Breeding in India: Scenario, Achievements and Research Needs. In: Islam, A. MH and MH, editor. In Brassica Breeding and Biotechnology. London, UK: IntechOpen; 2021. p. 174. https://doi.org/10.5772/intechopen.96319.

39. Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics. 2000;16:276–7. https://doi.org/10.1016/S0168-9525(00)02024-2.

40. Peden JF. Analysis of Codon Usage. University of Nottingham; 1999.

41. Sharp PM, Li WH. An evolutionary perspective on synonymous codon usage in unicellular organisms. J Mol Evol. 1986;24:28–38. https://doi.org/10.1007/BF02099948.

42. Wright F. The ‘effective number of codons’ used in a gene. Gene. 1990;87:23–9. https://doi.org/10.1016/0378-1119(90)90491-9.

43. Majeed A, Kaur H, Bhardwaj P. Selection constraints determine preference for A/U-ending codons in Taxus contorta. Genome. 2020;63:215–24. https://doi.org/10.1139/gen-2019-0165.

44. Sueoka N. Intrastrand parity rules of DNA base composition and usage biases of synonymous codons. J Mol Evol. 1996;42:323–323. https://doi.org/10.1007/BF02198860.

45. Sueoka N. Translation-coupled violation of Parity Rule 2 in human genes is not the cause of heterogeneity of the DNA G+C content of third codon position. Gene. 1999;238:53–8. https://doi.org/10.1016/S0378-1119(99)00320-0.

570 46. Sueoka N, Kawanishi Y. DNA G+C content of the third codon position and codon usage biases of human genes.
571 Gene. 2000;261:53–62. https://doi.org/10.1016/S0378-1119(00)00480-7.

572 47. Sueoka N. Directional mutation pressure and neutral molecular evolution. Proceedings of the National Academy of
573 Sciences. 1988;85:2653–7. https://doi.org/10.1073/pnas.85.8.2653.

574 48. Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. Mol Biol Evol. 1985;2:13–34.
575 https://doi.org/10.1093/oxfordjournals.molbev.a040335.

576 49. Stenico M, Lloyd AT, Sharp PM. Codon usage in Caenorhabditis elegans: delineation of translational selection and
577 mutational biases. Nucleic Acids Res. 1994;22:2437–46. https://doi.org/10.1093/nar/22.13.2437.

578 50. Lavner Y, Kotlar D. Codon bias as a factor in regulating expression via translation rate in the human genome. Gene.
579 2005;345:127–38. https://doi.org/10.1016/j.gene.2004.11.035.

580 51. Karlin S, Mrázek J. What Drives Codon Choices in Human Genes? J Mol Biol. 1996;262:459–72.
581 https://doi.org/10.1006/jmbi.1996.0528.

582 52. Fox JM, Erill I. Relative Codon Adaptation: A Generic Codon Bias Index for Prediction of Gene Expression. DNA
583 Research. 2010;17:185–96. https://doi.org/10.1093/dnares/dsq012.

584 53. Zhang R, Zhang L, Wang W, Zhang Z, Du H, Qu Z, et al. Differences in Codon Usage Bias between Photosynthesis-
585 Related Genes and Genetic System-Related Genes of Chloroplast Genomes in Cultivated and Wild Solanum Species. Int
586 J Mol Sci. 2018;19:3142. https://doi.org/10.3390/ijms19103142.

587 54. Andargie M, Congyi Z. Genome-wide analysis of codon usage in sesame (Sesamum indicum L.). Heliyon.
588 2022;8:e08687. https://doi.org/10.1016/j.heliyon.2021.e08687.

589 55. Team RDC. R: A Language and Environment for Statistical Computing. 2019.

590 56. Puigbò P, Bravo IG, Garcia-Vallve S. CAIcal: A combined set of tools to assess codon usage adaptation. Biol Direct.
591 2008;3:38. https://doi.org/10.1186/1745-6150-3-38.

592 57. Peden JF. Analysis of Codon Usage. University of Nottingham; 1999.

593 58. Hassan S, Mahalingam V, Kumar V. Synonymous Codon Usage Analysis of Thirty Two Mycobacteriophage
594 Genomes. Adv Bioinformatics. 2009;2009:1–11. https://doi.org/10.1155/2009/316936.

595 59. Wu Y qing, Li Z yuan, Zhao D qiu, Tao J. Comparative analysis of flower-meristem-identity gene APETALA2 (AP2)
596 codon in different plant species. J Integr Agric. 2018;17:867–77. https://doi.org/10.1016/S2095-3119(17)61732-5.

597 60. Butt AM, Nasrullah I, Tong Y. Genome-wide analysis of codon usage and influencing factors in chikungunya viruses.
598 PLoS One. 2014;9. https://doi.org/10.1371/journal.pone.0090905.

599 61. Gouy M, Gautier C. Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res. 1982;10:7055–
600 74.

601 62. Bragg JG, Quigg A, Raven JA, Wagner A. Protein elemental sparing and codon usage bias are correlated among
602 bacteria. Mol Ecol. 2012;21:2480–7. https://doi.org/10.1111/j.1365-294X.2012.05529.x.

603 63. Vicario S, Moriyama EN, Powell JR. Codon usage in twelve species of Drosophila. BMC Evol Biol. 2007;7:1–17.
604 https://doi.org/10.1186/1471-2148-7-226.

605 64. Shi SL, Jiang YR, Liu YQ, Xia RX, Qin L. Selective pressure dominates the synonymous codon usage in
606 parvoviridae. Virus Genes. 2013;46:10–9. https://doi.org/10.1007/s11262-012-0818-6.

607 65. Wong GKS, Wang J, Tao L, Tan J, Zhang J, Passey DA, et al. Compositional gradients in Gramineae genes. Genome
608 Res. 2002;12:851–6. https://doi.org/10.1101/gr.189102.

66. Xu C, Cai X, Chen, Hongxia Zhou Q, Cai Y, Ben A. Factors Affecting synonymous codon Usage Bias in chloroplast Genome of Oncidium Gower Ramsey. Evolutionary Bioinformatics. 2011;:271–8. https://doi.org/10.4137/EBO.S8092.

67. Wu Y, Zhao D, Tao J. Analysis of Codon Usage Patterns in Herbaceous Peony (Paeonia lactiflora Pall.) Based on Transcriptome Data. Genes (Basel). 2015;6:1125–39. https://doi.org/10.3390/genes6041125.

68. Tatarinova T V, Alexandrov NN, Bouck JB, Feldmann KA. GC3 biology in corn, rice, sorghum and other grasses. BMC Genomics. 2010;11:308. https://doi.org/10.1186/1471-2164-11-308.

69. Yang Z, Nielsen R. Mutation-Selection Models of Codon Substitution and Their Use to Estimate Selective Strengths on Codon Usage. Mol Biol Evol. 2008;25:568–79. https://doi.org/10.1093/molbev/msm284.

70. Bennetzen JL, Hall BD. Codon selection in yeast. Journal of Biological Chemistry. 1982;257:3026–31. https://doi.org/10.1016/s0021-9258(19)81068-2.

71. Singh R. Analysis of synonymous codon usage bias in Ancylostoma ceylanicum. Gene Rep. 2021;24:101290. https://doi.org/10.1016/j.genrep.2021.101290.

72. Nayak KC. Comparative study on factors influencing the codon and amino acid usage in Lactobacillus sakei 23K and 13 other lactobacilli. Mol Biol Rep. 2012;39:535–45. https://doi.org/10.1007/s11033-011-0768-4.

73. Gun L, Yumiao R, Haixian P, Liang Z. Comprehensive analysis and comparison on the codon usage pattern of whole mycobacterium tuberculosis coding genome from different area. Biomed Res Int. 2018;2018. https://doi.org/10.1155/2018/3574976.

74. Deb B, Uddin A, Mazumder GA, Chakraborty S. Analysis of codon usage pattern of mitochondrial protein-coding genes in different hookworms. Mol Biochem Parasitol. 2018;219:24–32. https://doi.org/10.1016/j.molbiopara.2017.11.005.

75. Błazej P, Mackiewicz D, Wnetrzak M, Mackiewicz P. The impact of selection at the amino acid level on the usage of synonymous codons. G3: Genes, Genomes, Genetics. 2017;7:967–81. https://doi.org/10.1534/g3.116.038125.

76. Wang Z, Wang G, Cai Q, Jiang Y, Wang C, Xia H, et al. Genomewide comparative analysis of codon usage bias in three sequenced Jatropha curcas. J Genet. 2021;100:20. https://doi.org/10.1007/s12041-021-01271-9.