

Mutational Pressure Drives Evolution of Synonymous Codon Usage in Genetically Distinct *Oenothera* plastomes

Rahul R. Nair ¹, Nimal T. Raveendran ¹, Vijaya R. Dirisala ¹, Manivasagam B Nandhini ², Thilaga Sethuraman ², Thirupati C. Venkateswarulu ¹, Ganesh Doss ^{2,*}

¹Department of Biotechnology, Vignan University, Vadlamudi P.O., Guntur District, Andhra Pradesh, INDIA, Pin: 522213

²Department of Plant Biotechnology, School of Biotechnology, Madurai Kamaraj University, Palkalai Nagar, Madurai District, Tamil Nadu, INDIA, Pin: 625021

Corresponding author: Ganesh Doss, Department of Plant Biotechnology, Madurai Kamaraj University, Madurai District, Tamil Nadu, Pin 625021, INDIA. Tel: +91-9486284825, E-mail: ganeshdsneha@yahoo.co.in

Received: February 10, 2014; **Revised:** September 26, 2014; **Accepted:** October 14, 2014

Background: Most of the amino acids are encoded by more than one codon, termed as synonymous codons. Synonymous codon usage is not random as it is unique to species. In each amino acid family, some synonymous codons are preferred and this is referred to as synonymous codon usage bias (SCUB). Trends associated with evolution of SCUB and factors influencing its diversification in plastomes of genetically distinct *Oenothera* plastomes have not been investigated so far.

Objectives: In the present study, major forces that shape SCUB in *Oenothera* plastomes and putative preferred codons in the protein coding genes (PCG) of plastomes were identified.

Materials and Methods: To unravel various features of SCUB across selected *Oenothera* plastomes, commonly used codon usage indices such as relative synonymous codon usage (RSCU), synonymous codon usage order (SCUO), effective number of codons (ENC) and codon adaptation index (CAI) were calculated. Correspondence analysis (COA) on RSCU was performed to identify various characteristics of SCUB across different PCG in *Oenothera* plastomes. Spearman's rank correlation analysis was adopted to correlate nucleotide contents, codon usage indices and major axes of COA to find out critical parameters in shaping SCUB.

Results: Mutational bias due to compositional constraints played crucial role in shaping SCUB as T₃ and GC₃ contents were in strong negative correlation with all axes of COA. Nevertheless, significant negative correlations between axis 1 and 3 with ENC and CAI respectively, in all species, and narrow distribution of GC contents in neutrality plot, indicate the role of natural selection. Hydrophathy score of proteins was found to be influencing SCUB in *O. glazioviana* as it showed strong negative correlation with axis 2.

Conclusion: We concluded that mutational pressure coupled with weak selection influenced SCUB in the examined plastomes of *Oenothera*. In addition, all examined species of *Oenothera* exist as disjunct populations in different parts of North America and these populations might have experienced genetic drift as random mutations in small populations that have been fixed over a period of time.

Keywords: Genetic drift; Mutational pressure; *Oenothera*; Selection; Synonymous codon usage bias

1. Background

Most amino acids are encoded by multiple triplets of nucleotides (*i.e.*, synonymous codons) that differ in the third codon position (1) or rarely in the second codon position (2). Usage of synonymous codons for a given amino acid is not at equal frequencies both within and between genomes (2-5). Different organisms exhibit species specific preference towards a subset of synonymous codons for coding particular amino acids (3, 6). This non random usage of synonymous codons is referred as SCUB and is an essential characteristic of both prokaryotic and eukaryotic genomes (2).

Though synonymous mutations are generally neu-

tral or silent due to no change in amino acid sequence, SCUB are reported to have profound effects on gene expression and function (7-10). Population genetic studies suggested that mutational biases due to nucleotide compositional constraints or weak selection on specific codons might be contributing to SCUB (10-12). Many studies confirmed that SCUB is higher in genomic regions on which substantial purifying selection acts at amino acid level (10-13). Highly expressed genes experience stronger SCUB than genes with low expression (4). Moreover, evolutionarily conserved protein coding regions show stronger SCUB (14). However, the role of various physiological

processes that contribute to SCUB in protein coding regions of genomes has remained elusive (10).

Contributing major forces on SCUB fall into three categories: (i) nucleotide compositional constraints (15), (ii) translational elongation rate optimization by natural selection, and (iii) a balance between mutational pressure, selection and genetic drift in a finite population (11, 16, 17). Other contributing factors include interaction between codons and anticodons (18), efficacy of replication (19) and usage of codon pairs (20).

A plastid genome of higher plants comprises of 80 protein coding genes (PCGs), 4 rRNA genes and 30 tRNA genes (21). Since most proteins in the chloroplast are essential for photosynthesis, protein coding regions of chloroplast are highly conserved in higher plants, although a few exceptions exist (22, 23). In plastidic genome, mutational pressure favours high representation of A/T and it appears to be the major factor shaping SCUB (24 - 26). However, codon usage of *psbA* gene is found to be highly correlated with the corresponding t-RNA population in the chloroplast, indicating the possible influence of selection for translational efficiency on *psbA*. Previous studies on SCUB and variety of factors influencing its diversification in chloroplast genome, revealed that even though mutational bias is predominant, selection on codon usage cannot be nullified. A recent finding suggested that intron evolution and DNA methylation could be considered as potential factors that frame SCUB in land plants (27).

The angiosperm genus *Oenothera* (Family: Onagraceae) is commonly distributed in South Africa (28) and North America (29). Genus *Oenothera* is considered to be well suited for understanding the various molecular aspects of speciation process as it is amongst well-characterized plant genera (30). The genus *Oenothera* has been regarded as an ideal model to study evolution of plant genomes (particularly plastids), since substantial information about its systematics and genetics are available (30). In *Oenothera*, plastome-genomes are highly incompatible. However, fertile plants have been evolved due to (i) the exchange of plastids and nuclei between species, and (ii) the exchange of individual chromosome or complete haploid set between species (30).

2. Objectives

Plastid genomes of 5 *Oenothera* sp. have been completely sequenced (30) and revealed that all plastomes are genetically distinct due to wide nucleotide substitution, small insertions, deletions and repetitions (30). In addition, phylogenetic analysis proved that

these plastomes differ from common ancestral origin of vascular plants by a 56 kb inversion within the large single copy region (30). These findings suggested *Oenothera* plastomes as most suitable candidates to explore SCUB as well as the factors influencing its diversification. Basic features of molecular evolution can be identified by determining evolutionary patterns at synonymous sites in codons. Hence, in the present study, major objectives were i) investigation of trends associated with synonymous codon usage in 5 distinct *Oenothera* plastomes to obtain an insight into the major forces that shape SCUB and ii) identification of putative preferred codons in the PCG of plastomes that helps to optimize heterologous gene expression. Correlation analysis of various codon usage indices provided a better understanding about the pattern of SCUB in the plastomes. Identification of putative optimal codons would certainly pave the way for developing transplastomic *Oenothera* sp. for enabling evolutionary biologists to study underlying molecular mechanisms behind plant genome evolution.

3. Materials and Methods

3.1. Sequence Data and Nucleotide Compositions

Complete nucleotide sequences of all 5 plastomes of *Oenothera* sp. viz., *Oenothera argillicola*, *Oenothera biennis*, *Oenothera elata*, *Oenothera glazioviana* and *Oenothera parviflora* were retrieved from National Centre for Biotechnology Information (NCBI) website and details were presented (Table 1). PCGs of each genome were extracted and coding sequences (CDS) that contain less than 300 codons were excluded in order to avoid sampling errors. Integrity of the CDS was evaluated by examining the presence of initiation and termination codons at appropriate places without any internal stop codons. Duplicate sequences were removed from the dataset. Thus, final data set for analysis contained 54 CDS for *O. argillicola*, *O. biennis*, *O. elata* and *O. parviflora* and 53 CDS for *O. glazioviana*.

Overall and local nucleotide compositions (*i.e.*, nucleotide contents at 1st, 2nd and 3rd codon positions) were calculated for each CDS. Spearman's rank correlation analysis was used to reveal the correlations between overall and silent base contents such as A₃, T₃, G₃ and C₃ to unravel intrinsic properties of SCUB.

3.2. Indices of Codon Usage

3.2.1. Relative Synonymous Codon Usage (RSCU)

RSCU value of each codon was calculated to study the trend associated with SCUB in PCGs of *Oenothera*

Table 1. *Oenothera* species and their average local GC composition and ENC values

Organism	GenBank Accession number	Number of PCGs	Average GC ₁	Average GC ₂	Average GC ₃	Average ENC
<i>Oenothera argillicola</i>	EU262887.1	54	47.92	39.93	31.69	48.91
<i>Oenothera biennis</i>	EU262889.1	54	47.85	39.61	31.57	48.92
<i>Oenothera elata</i>	AJ271079.3	54	47.85	40.01	31.56	48.90
<i>Oenothera glazioviana</i>	EU262890.1	53	48.06	39.97	31.70	48.93
<i>Oenothera parviflora</i>	EU262891.1	54	47.85	40.01	31.56	48.90

plastomes. RSCU value has been extensively used to study codon usage of PCGs in various genomes as it is independent of amino acid composition. RSCU values were calculated using the following equation (31).

RSCU:

$$\frac{\text{Observed frequency of a codon}}{\text{Expected frequency of the particular codon provided}}$$

If RSCU value of a particular codon is greater than 1, it indicates the biased codon usage (31).

3.2.2. Effective Number of Codons (ENC)

ENC of particular gene has been widely used in codon usage research to measure the extent of SCUB of that particular gene (32). ENC can vary from 20 (strictly biased; 1 codon for 1 amino acid) to 61 (no bias; all synonymous codons are used equally for each amino acid family). It is considered as an effective method to measure SCUB because it is independent of gene length (32). Preference towards particular codons in each amino acid family due to either selection or mutational pressure reduces the value of ENC. If ENC value of a gene is 35 or less, that particular gene can be considered as highly biased and vice versa. Expression levels of highly biased genes have been considered as high.

Expected ENC value under no selection can be calculated for any value of GC₃ as per the equation (32)

$$E(ENC) = 2 + s + \{29/[s^2 + (1-s)^2]\}$$

Where $s = GC_3$

A plot between calculated ENC value of each CDS and its corresponding GC₃ value was developed for all *Oenothera* sp. to provide an understanding of the influence of GC compositional constraints in shaping SCUB. If majority of genes are grouped on or just fall below the left/right hand side of the expected GC₃ curve, GC₃ compositional constraints will be suggested as the major force that determine SCUB (32). If

majority of genes are grouped considerably below the expected GC₃, selection may be the significant force in shaping SCUB (32).

3.2.3. Codon Adaptation Index (CAI)

CAI is used to measure the extent of SCUB towards a subset of codons in each amino acid family of a given gene on the basis of preferred codons (translationally optimal) (33) in highly expressed genes such as ribosomal proteins and translational elongation factors. CAI is a good indicator of the level of expression as it takes into account all 59 synonymous codons in a quantitative manner (33). CAI value of a gene may vary from 0 to 1, a lower value indicates less SCUB (low expression level) and higher value close to 1 indicates higher SCUB (high expression level) for a given gene. In the present study, ribosomal protein coding genes of each *Oenothera* species were used as the reference set of highly expressed genes for finding out CAI values for corresponding species (34).

$$CAI = \exp 1/L \sum_{n=1}^L \ln w_n \quad (33)$$

Where w_n = relative adaptness of nth codon, L = number of codons

3.2.4. Synonymous codon usage order (SCUO)

SCUO is used for quantitative evaluation of relationship between GC composition at each codon position and SCUB for a gene and it is computed as per the equation (35)

$$SCUO = 1 + p/2 \log_2 (p/2) + \left\{ \frac{(1-p)}{2} \right\} \log_2 \left\{ \frac{(1-p)}{2} \right\}$$

Tukey test was used to analyse the differences in SCUB within genomes and the Wilcoxon two-sample test was used to compare the SCUB across five plastomes.

3.3. Correspondence Analysis (COA)

COA was performed to study the various characteristics of SCUB across different PCG in each *Oenothera plastome* (36) based on RSCU values (37, 38). All PCGs were plotted in a 59 dimensional vector space based on the usage of 59 synonymous codons. Each PCG is regarded as a 59 dimensional vector and RSCU value of each codon is represented as a dimension (39). Major variations in the trend associated with synonymous codon usage were explained by the first axis with subsequent axes explaining diminishing amounts of variance (40). Spearman's rank correlation analysis was used to reveal correlations between various codon usage indices described above and major axes of COA as this method of correlation is independent of any kind of distributional assumptions (41).

3.4. Identification of Putative Optimal Codons

To identify putative optimal codons/ preferred codons, 10% of PCG located on both extremes left and right of axis 1 of COA were chosen to form 2 data sets (42). For each of the 59 synonymous codons, Chi square test was applied to the 2×2 table that was constructed from the above 2 data sets. First row of the table contains observed frequencies of a codon and the second row contains total frequency of other synonymous alternatives of that particular codon (41).

3.5. Cluster Analysis

Cluster analysis on RSCU values (39) was performed in order to understand the grouping of *Oenothera* species according to the codon usage. In cluster analysis, a 5×59 matrix was generated in which rows and columns correspond to pooled RSCU values of 59 codons and five *Oenothera* species, respectively. Clustering of *Oenothera* species occurred based on the RSCU values by unweighted pair group average clustering using Euclidean distance.

3.6. Bioinformatic and Statistical Softwares

Total base compositions and base compositions at each codon positions were calculated by using MEGA version 5.2.2 (43). Dambe version 5.3.31 (44) was used to find out RSCU values. Online version of codonW (45; <http://mobyli.pasteur.fr/cgi-bin/portal.py>) was used to estimate ENC, hydrophathy score (a number indicating hydrophobic/hydrophilic properties of side chain of an amino acid) and aromaticity (frequency of aromatic amino acids) values. CAI values were calculated by using CAI calculator 2 (46). CodonO (47) was employed to compute SCUO

values (35). All kinds of statistical analysis including correspondence analysis and cluster analysis were carried out using PAST version 2.12 (48) and the significance was measured at 5% level.

4. Results

4.1. Intrinsic Properties of Synonymous Codon Usage

Total and synonymous nucleotide contents were estimated. In the plastomes of *O. argillicola*, AT content was higher than GC content. Among the silent base contents, viz., A_3 , T_3 , G_3 , and C_3 (A , T , G and C contents at 3rd codon position), T_3 was noted to be higher than all others with a mean and standard deviation (SD) of 36.64 and 4.78, respectively. Lowest nucleotide content at silent site was noted to be C_3 with a mean and SD of 14.54 and 3.36, respectively. Spearman's rank correlation analysis revealed strong positive correlations between A and A_3 , T and T_3 , G and G_3 , and C and C_3 . Whereas significant negative correlations were observed between other heterogeneous nucleotide contents (Table 2). Strong negative correlation between A and T_3 and vice versa suggested the possible influence of AT at silent sites (AT_3) in shaping SCUB of PCGs in *O. argillicola* plastomes. Additionally, high positive correlations of GC_3 with G , C and GC contents indicated GC compositional constraints might also be present. However, no correlation existed between GC_3 and any of A/T contents. These complex correlations revealed that nucleotide compositional constraints play a crucial role in framing SCUB across PCGs in *O. argillicola* plastomes. Similar patterns of correlations were identified in other 4 *Oenothera* plastomes examined (Table 2).

4.2. GC Composition Influences on SCUB

GC composition has been regarded as an important force that shape codon and amino acid usages (49). Total GC content and GC composition at 3 codon positions of all selected PCGs of *Oenothera* plastomes were calculated and a dot plot was produced with respective SCUO values (Figure 1). Strong linear but negative correlations were found between SCUO and variables such as GC_3 ($r = -0.495$, $p < 0.01$), GC_1 ($r = -0.442$, $p < 0.01$) and GC ($r = -0.353$, $p < 0.01$). Among these variables, dependency of SCUO on GC_3 ($SCUO = 0.008 (GC_3) + 0.481$) was noticed as stronger as revealed by the highest correlation between them. To study the influence of overall GC on local compositions, linear correlation analysis was performed for GC with all local GC contents. GC was linearly corre-

Table 2. Spearman's rank correlation analysis between overall base composition and base composition at silent site in plastomes of five *Oenothera* species

	Axes	A ₃	T ₃	G ₃	C ₃	GC ₃
<i>O. argillicola</i>	A	0.699**	-0.689**	0.126	-0.192	-0.127
	T	-0.508**	0.801**	-0.282*	-0.114	-0.245
	G	-0.172	-0.253	0.468**	0.253	0.527**
	C	-0.049	-0.138	-0.168	0.492**	0.345*
	GC	-0.167	-0.246	0.246	0.423**	0.562**
<i>O. biennis</i>	A	0.719**	-0.689**	0.097	-0.191	-0.145
	T	-0.519**	0.808**	-0.270*	-0.104	-0.235
	G	-0.165	-0.242	0.470**	0.219	0.497**
	C	-0.056	-0.154	-0.202	0.497**	0.338*
	GC	-0.182	-0.231	0.240	0.405**	0.548**
<i>O. elata</i>	A	0.711**	-0.691**	0.100	-0.194	-0.133
	T	-0.513**	0.810**	-0.271*	-0.101	-0.238
	G	-0.175	-0.253	0.488**	0.226	0.511**
	C	-0.053	-0.158	-0.211	0.503**	0.330*
	GC	-0.186	-0.240	0.246	0.411**	0.549**
<i>O. glazioviana</i>	A	0.704**	-0.689**	0.104	-0.159	-0.112
	T	-0.526**	0.818**	-0.289*	-0.109	-0.246
	G	-0.145	-0.269	0.496**	0.194	0.498**
	C	-0.062	-0.169	-0.191	0.525**	0.356**
	GC	-0.172	-0.261	0.274**	0.394**	0.561**
<i>O. parviflora</i>	A	0.711**	-0.691**	0.100	-0.194	-0.133
	T	-0.513**	0.810**	-0.271*	-0.101	-0.238
	G	-0.175	-0.253	0.488**	0.226	0.511**
	C	-0.053	-0.158	-0.211	0.503**	0.330*
	GC	-0.186	-0.240	0.246	0.411**	0.511**

lated with GC₁ ($r = 0.840$, $p < 0.01$), with GC₂ ($r = 0.613$, $p < 0.01$) and with GC₃ ($r = 0.560$, $p < 0.01$). Whereas GC₃ was in high correlation with GC₁ ($r = 0.471$, $p < 0.01$) but not with GC₂ ($r = -0.138$). Similarly, GC₁ was also found not correlated with GC₂. Similar pattern of linear correlations were also observed in other *Oenothera* species. These results suggested that overall GC content, GC₁ and GC₃ influenced SCUB in all examined *Oenothera* plastomes. Thus, mutational pressure has significant role in dictating SCUB across PCGs in *Oenothera* plastomes. Difference in SCUB among five species of *Oenothera* was compared by Wilcoxon two-sample test and the result was indicative of no significant difference in SCUB between any two species.

4.3. Features of Overall and Strand Specific Relative Synonymous Codon Usage

Overall and strand specific synonymous codon

usage were examined (Table 3). In 18 synonymous families of amino acid, A and T ending codons were used more frequently than G and C ending codons, indicating an AT-rich nature of plastomes. Most of the 3, 4 and 6 fold degenerate amino acid families were observed to use T ending codons except Gly and Arg. Strand specific codon usage bias was observed for 6 fold degenerate amino acid Arg in all *Oenothera* sp. except *O. elata*. For Arg, codon usage was biased towards CGT in minus strand for all species whereas in plus strand, codon usage was biased towards CGA in *O. biennis*, *O. glazioviana* and *O. parviflora*. However, both CGT and CGA were used at equal frequencies to code Arg in minus strand encoded genes. Four fold degenerate amino acid Val used GTA most often in all plus strand encoded genes whereas all minus strand encoded genes used GTT most frequently (Strand specific codon usage). Chi-square analysis on codon count of 10% genes distributed on extreme

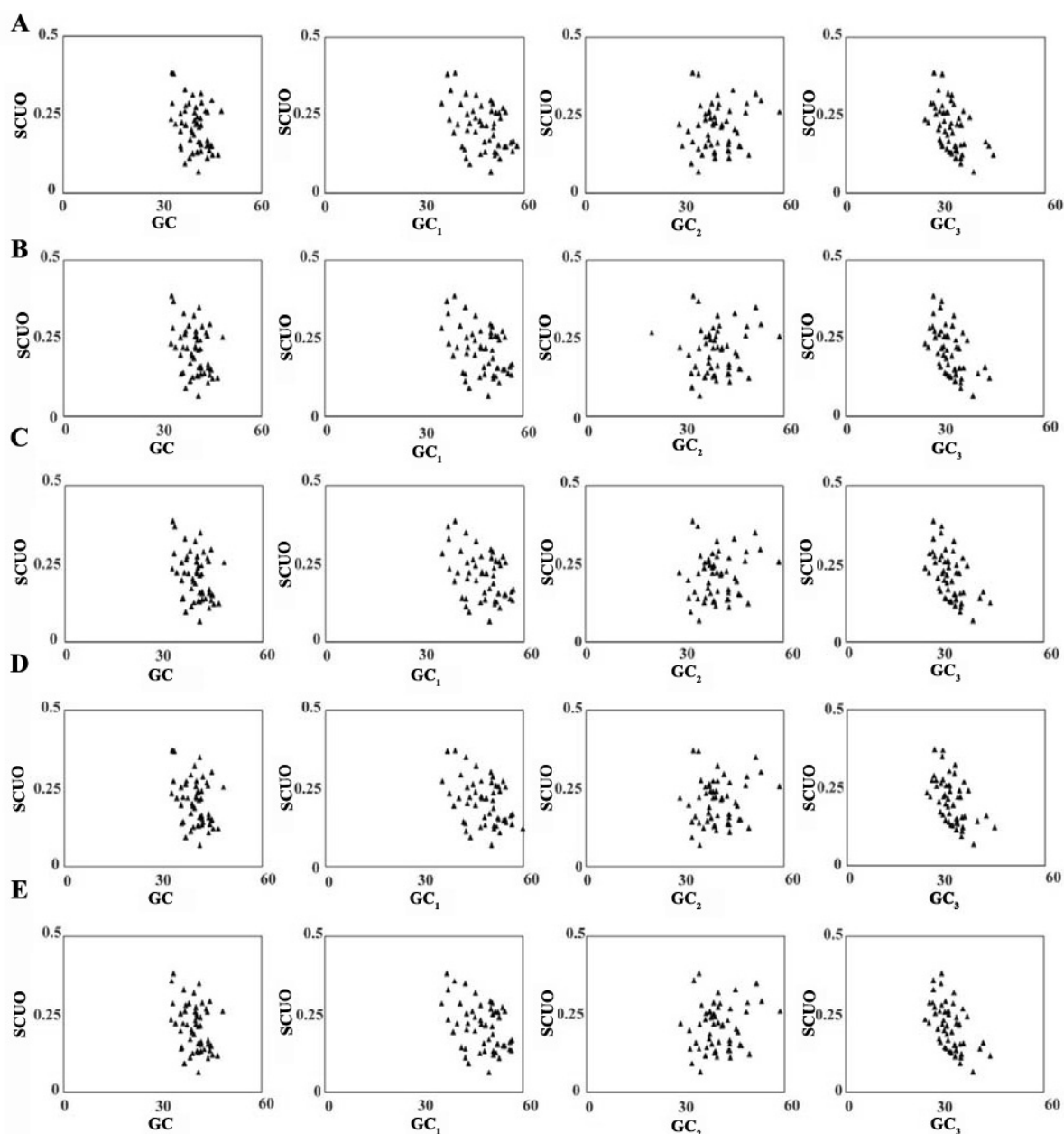


Figure 1. Relationship of GC composition and SCUB. SCUB of A: *O. argillicola* B: *O. biennis* C: *O. elata* D: *O. glazioviana* and E: *O. Parviflora*

left and extreme right of axis 1 revealed 5 statistically over represented codons (*i.e.*, putative optimal codons) in *O. argillicola* (*i.e.*, GCT, GAA, CAT, AAT and CCT), 1 in *O. biennis* (CGA), 4 in *O. elata* (*i.e.*, TGT, AAT, GTT and GTA), 7 in *O. glazioviana* (*i.e.*, GAT, TTT, CAT, AAT, CCT, CGT and TCT) and 2 in *O. parviflora* (*i.e.*, GCT and GTA). All putative optimal codons used A/T ending codons only.

4.4. Quantification of SCUB

ENC has been used as a reliable tool in SCUB

analysis as it is effective for short genes and for skewed usage of amino acids (32). ENC value of a gene clearly demonstrates SCUB in a range from extreme bias to minimal bias. Plotting ENC values of genes against corresponding GC₃ values displays major characteristics of synonymous codon usage patterns of PCGs in a genome. In this study, majority of protein coding genes were grouped on the left hand side of the expected GC₃ curve in all chosen *Oenothera* sp. (Figure 2). Hence, GC₃ compositional constraints might influence SCUB across PCGs in

Oenothera plastomes. However, some genes were located considerably below the expected GC_3 curve indicating the possible influence of some other force such as natural selection in framing SCUB. No significant correlation was observed between GC_3 and GC_{12} in neutrality plot (Figure 3). This suggested that intragenomic GC mutational bias on GC content at all codon position is low, which in turn indicates high conservation of GC content. Furthermore, narrow distribution of GC contents was observed in neutrality plot, revealing the role of selection in framing SCUB. Association between A, T and G, C was analyzed using parity rule 2 (PR2) bias plot and noticed that A and T

contents were used more proportionally than G and C contents (Figure 4).

4.5. Various Factors Affecting SCUB

The COA on RSCU values of PCGs in 5 *Oenothera* plastomes was carried out and positions of PCGs along first 4 axes are given in Figure 5. The first 4 axes accounted for 34.59%, 34.76%, 34.66%, 35.57% and 34.49% of total variation in *O. argillicola*, *O. biennis*, *O. elata*, *O. glazioviana* and *O. parviflora*, respectively. No single major explanatory axis was found to detail variations in all the chosen plastomes. Significant negative correlations were found between

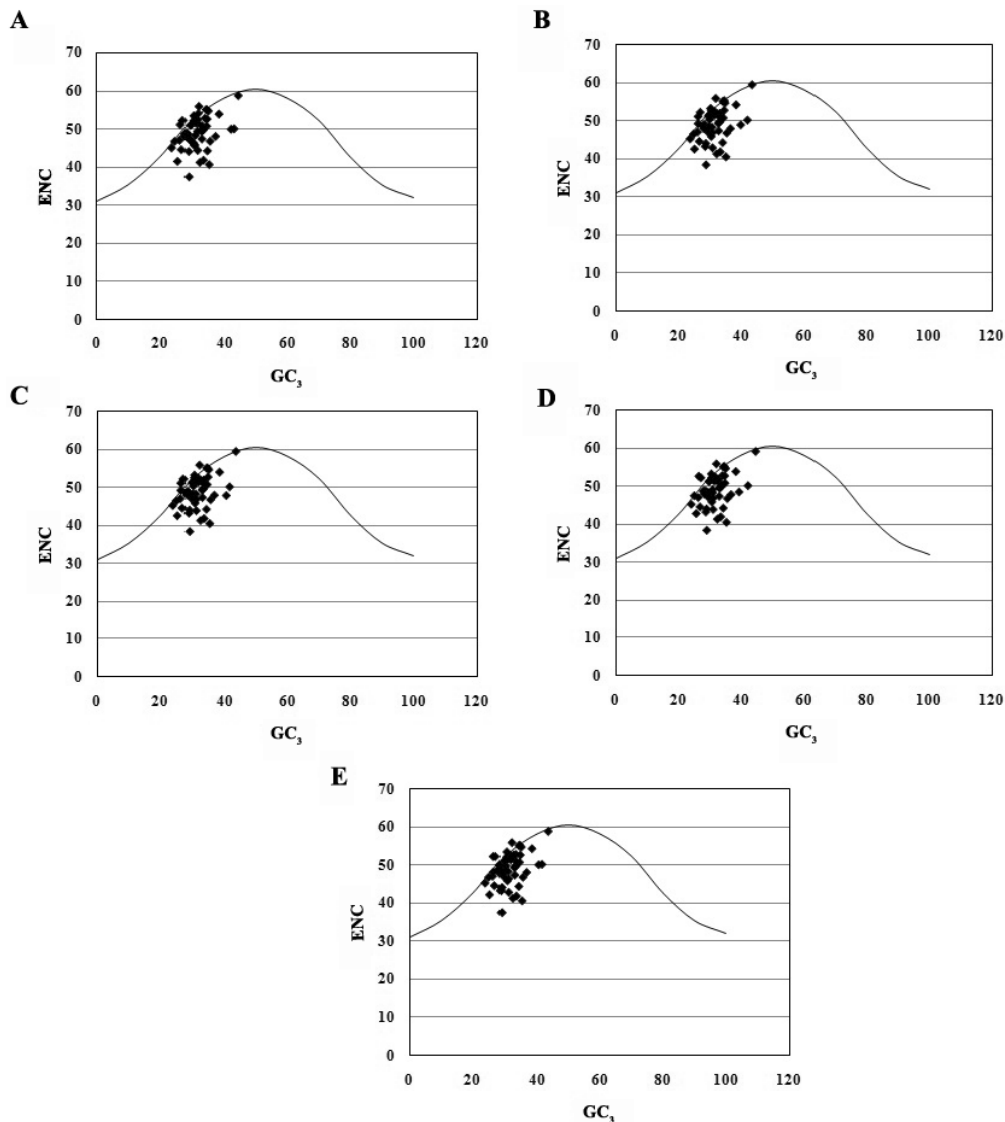


Figure 2. Quantification of SCUB. ENC Vs GC_3 plot of A: *O. argillicola* B: *O. biennis* C: *O. elata* D: *O. glazioviana* and E: *O. Parviflora*

axis 1 and indices, indicating gene expression levels such as ENC and CAI in all chosen species (Table 4). Axis 4 was in high negative correlation with ENC in *O. argillicola* and *O. parviflora*. Another index of level of gene expression, viz., CAI was positively correlated with axis 4 of *O. argillicola* and was negatively corre-

lated with axis 2 of *O. glazioviana*. Significant correlations of various axes of COA with gene expression indices such as ENC and CAI suggested the influence of gene expression levels in the SCU variation across PCGs in *Oenothera* plastomes. No correlations were observed between any of the 4 axes of COA and gene

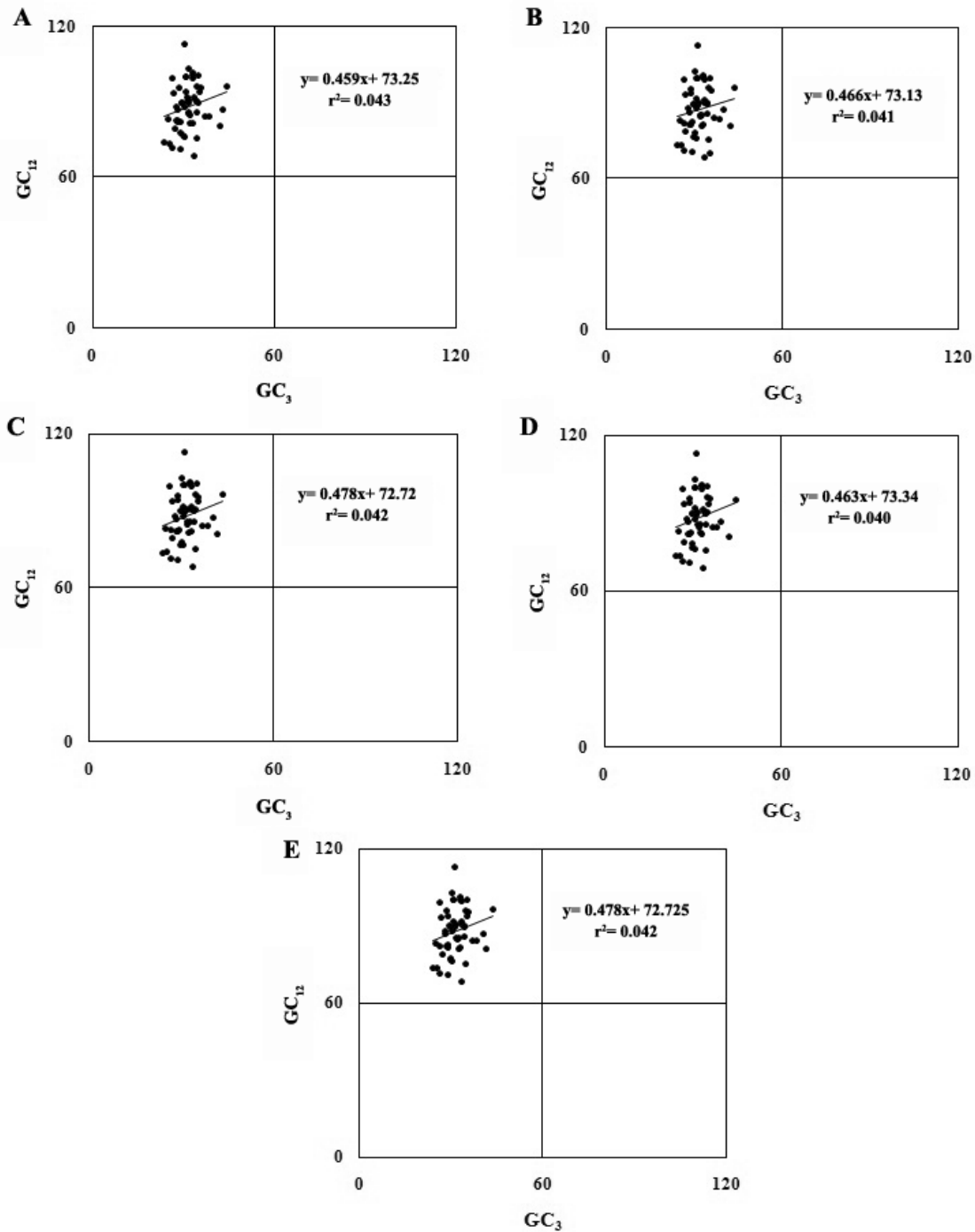


Figure 3. Correlation analysis between GC₁₂ and GC₃. Neutrality plot of A: *O. argillicola* B: *O. biennis* C: *O. elata* D: *O. glazioviana* and E: *O. Parviflora*

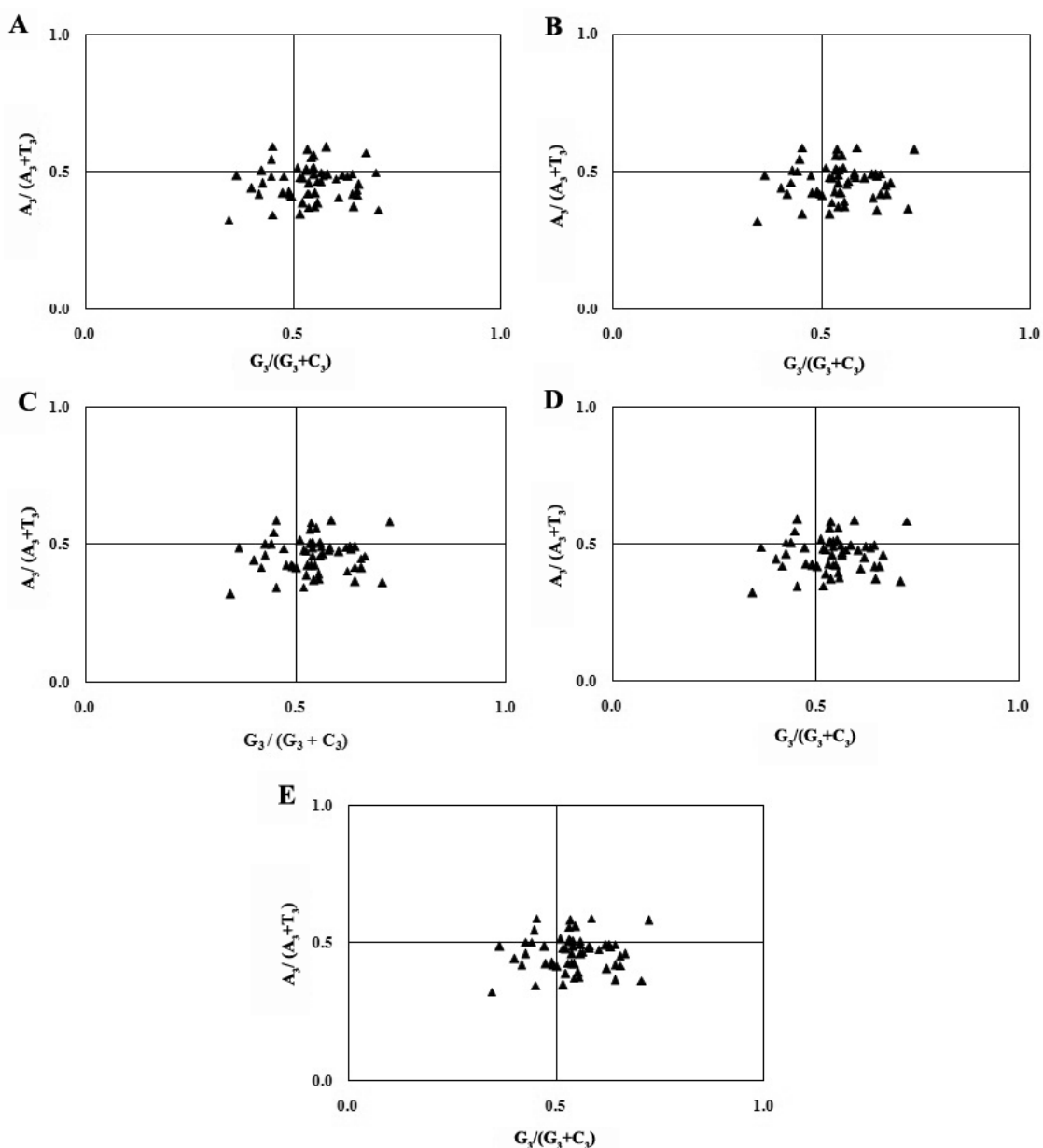


Figure 4. Parity rule 2 bias plot. PR2 bias plot of A: *O. argillicola* B: *O. biennis* C: *O. elata* D: *O. glazioviana* E: *O. Parviflora*

length or aromaticity. In *O. glazioviana*, hydrophathy score was in significant negative correlation with axis 2, but no such correlation was observed in any other *Oenothera* sp. Strong negative correlation between T_3 content and axis 2 in all *Oenothera* sp. indicated its high influence on SCU variations. Interestingly, GC_3 content was in strong negative correlation with axis 3 in all chosen species and was in positive correlation with axis 4 in *O. elata*. This suggested that GC_3 and T_3 influences SCU variations considerably in all exam-

ined PCGs of *Oenothera* plastomes. Correlation analysis between first 4 axes of COA and RSCU value of 59 synonymous codons revealed certain significant negative correlations in all chosen species, *i.e.*, axis 1 with GCC, TGC, GAT, GGG, CAT, AAT, CCC, CCG, AGA and TCG and axis 2 with TGC, GGT, CAT, CTT, TTG, AGA, CGT and GTT (Table 5). Though correlations existed between other 2 axes (*i.e.*, axes 3, 4) and RSCU value of certain codons, it was observed to be species specific. These results pointed out that mutational pressure

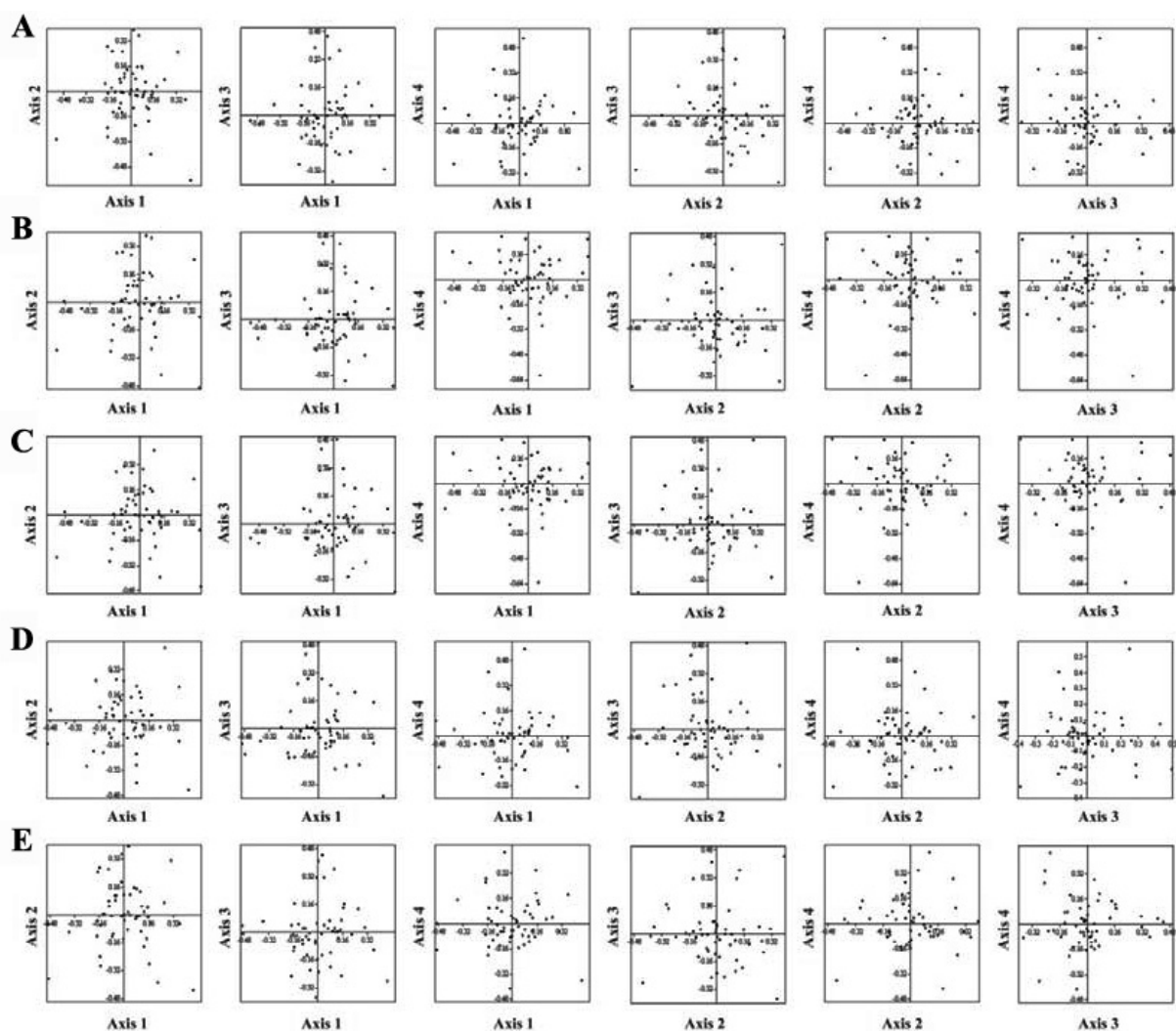


Figure 5. Variations in codon usage. Correspondence analysis of A: *O. argillicola* B: *O. biennis* C: *O. elata* D: *O. glazioviana* and E: *O. parviflora*

combined with weak selection might be acting on the PCGs of all *Oenothera* plastomes to cause SCUB. Cluster analysis revealed no major differences in synonymous codon usage across genetically distinct *Oenothera* plastomes as all *Oenothera* species formed only one cluster (Figure 6).

5. Discussion

All preferred codons were found to use A/T ending codons in *Oenothera* plastomes as plastid chromosomes are AT rich (5, 40). Mutational pressure towards or against GC composition determine the ATGC compositions of a genome (40, 50). In all examined *Oenothera* plastomes, AT₃ (AT content at silent sites) is expected to be an important factor in SCU variation

across PCGs. However, strong positive correlations existed between GC₃ and individual G/C contents. This suggested that GC₃ may also be considered as one of the possible factors. This can be explained by extremely low GC₃ that influences SCU considerably (32). Therefore, high AT₃ (~ 68.10%) and low GC₃ (~ 31.70%) can be regarded as the major factors behind SCU variation in *Oenothera* plastomes similar to what has already been reported in *Coffea arabica* (5), *Populus alba* (51), and in both *Nicotiana tabacum* and *Oryza sativa* (26). Moreover, point mutations, repetitions, insertions/ deletions and inversions were reported to contribute to base compositional changes in *Oenothera* (30). The impact of these mutations may reflect in SCU variations across PCGs in *Oenothera*

Table 4. Spearman's rank correlation analysis between four axes of COA and various codon usage indices

	Axes	A ₃	T ₃	G ₃	C ₃	GC ₃	ENC	CAI	Gravy	Aromaticity	Length
<i>O. argillicola</i>	Axis 1	-0.121	0.162	-0.157	0.216	0.109	-0.274*	-0.101	0.156	-0.038	-0.012
	Axis 2	0.247	-0.458**	-0.004	0.337*	0.192	0.150	-0.172	-0.195	-0.205	0.069
	Axis 3	0.122	0.111	-0.197	-0.094	-0.302*	-0.149	0.138	0.199	0.055	-0.078
	Axis 4	0.095	0.095	-0.414**	0.104	-0.194	-0.377**	0.286*	0.262	-0.024	-0.148
<i>O. biennis</i>	Axis 1	-0.051	0.111	-0.181	0.209	0.053	-0.325*	-0.073	0.151	-0.042	-0.028
	Axis 2	0.251	-0.458**	-0.045	0.370**	0.215	0.159	-0.211	-0.150	-0.191	0.056
	Axis 3	0.130	0.089	-0.127	-0.158	-0.332*	-0.127	0.164	0.197	-0.004	-0.064
	Axis 4	-0.197	0.062	0.062	0.192	0.266	-0.074	-0.140	0.074	0.144	0.057
<i>O. elata</i>	Axis 1	-0.060	0.112	-0.173	0.220	0.073	-0.319*	-0.082	0.137	-0.061	-0.008
	Axis 2	0.272*	-0.475**	-0.028	0.333*	0.220	0.187	-0.198	-0.211	-0.239	0.047
	Axis 3	0.158	0.047	-0.113	-0.114	-0.296*	-0.078	0.143	0.148	-0.008	-0.047
	Axis 4	-0.186	0.036	0.094	0.178	0.279*	-0.029	-0.157	0.041	0.155	0.054
<i>O. glazioviana</i>	Axis 1	-0.006	0.089	-0.187	0.194	0.024	-0.281*	-0.056	0.081	-0.105	-0.064
	Axis 2	0.232	-0.502**	0.055	0.340*	0.289*	0.251	-0.272*	-0.276*	-0.217	0.085
	Axis 3	0.251	-0.018	-0.124	-0.155	-0.364**	-0.086	0.184	0.096	-0.074	-0.056
	Axis 4	0.105	0.053	-0.278*	0.025	-0.212	-0.245	0.264	0.146	-0.074	-0.052
<i>O. parviflora</i>	Axis 1	-0.037	0.158	-0.163	0.101	-0.045	-0.363**	0.056	0.206	-0.053	-0.057
	Axis 2	0.258	-0.392**	-0.091	0.330*	0.180	0.106	-0.217	-0.091	-0.150	0.107
	Axis 3	0.081	0.095	-0.042	-0.186	-0.297*	-0.048	0.115	0.134	0.047	-0.091
	Axis 4	-0.042	0.081	-0.332*	0.240	-0.041	-0.288*	0.178	0.221	0.015	-0.031

Table 5. Major codons having significant correlations with first two axes of COA in all examined *Oenothera* species

Codon	<i>O. argillicola</i>		<i>O. biennis</i>		<i>O. elata</i>		<i>O. glazioviana</i>		<i>O. parviflora</i>	
	Axis 1	Axis 2	Axis 1	Axis 2	Axis 1	Axis 2	Axis 1	Axis 2	Axis 1	Axis 2
GCC	-0.578**	0.116	-0.540**	0.268*	-0.525**	0.312*	-0.459**	0.368**	-0.590**	0.150
TGC	-0.465**	-0.350**	-0.520**	-0.313*	-0.521**	-0.298*	-0.566**	-0.220	-0.497**	-0.316*
GAT	-0.317*	-0.217	-0.387**	-0.193	-0.409**	-0.195	-0.423**	-0.202	-0.340*	-0.157*
GGG	-0.345*	0.279*	-0.316*	0.259	-0.298*	0.276*	-0.228	0.323*	-0.389**	0.184
CAT	-0.361**	-0.306	-0.383**	-0.331*	-0.383**	-0.303*	-0.484**	-0.252	-0.330*	-0.346*
AAT	-0.301*	-0.078	-0.311*	-0.029	-0.349**	0.019	-0.358**	0.048	-0.308*	0.009
CCC	-0.401**	0.264	-0.426**	0.309*	-0.402**	0.376**	-0.332	0.454**	-0.509**	0.171
CCG	-0.359**	-0.085	-0.371**	-0.009	-0.368**	0.017	-0.318*	0.072	-0.332*	-0.158
AGA	-0.270*	-0.312*	-0.206	-0.301*	-0.188	-0.226	-0.189	-0.131	-0.253	-0.350**
TCG	-0.500**	0.158	-0.464**	0.100	-0.477**	0.119	-0.480**	0.119	-0.476**	0.008
GGT	0.070	-0.304*	0.101	-0.300*	0.080	-0.295*	0.055	-0.316*	0.171	-0.326*
CTT	-0.038	-0.302*	-0.035	-0.271*	-0.031	-0.247	0.058	-0.218	0.019	-0.393**
TTG	-0.054	-0.509**	-0.041	-0.481**	-0.043	-0.483**	-0.097	-0.434**	-0.006	-0.530**
CGT	0.114	-0.767**	0.075	-0.768**	0.050	-0.750**	-0.082	-0.765**	0.158	-0.765**
GTT	-0.217	-0.325*	-0.188	-0.285*	-0.190	-0.295*	-0.118	-0.302*	-0.107	-0.321*

plastomes.

Influence of GC composition on SCU was further elucidated by correlation analysis between SCUO and GC composition at each codon positions. Apparent linear relationship was found between overall GC content, GC₁ and GC₃ of all examined PCGs. As observed in grass models (40), we herein noticed that GC₃ was

the dominant factor in framing SCUB in *Oenothera* sp. This result suggested mutational pressure as significant driving force of SCU variations in *Oenothera* sp. ENC Vs GC₃ plot also confirmed the role of GC composition on SCUB as most of the PCGs lie on or just below the expected curve. However, grouping of some genes considerably below the expected curve points

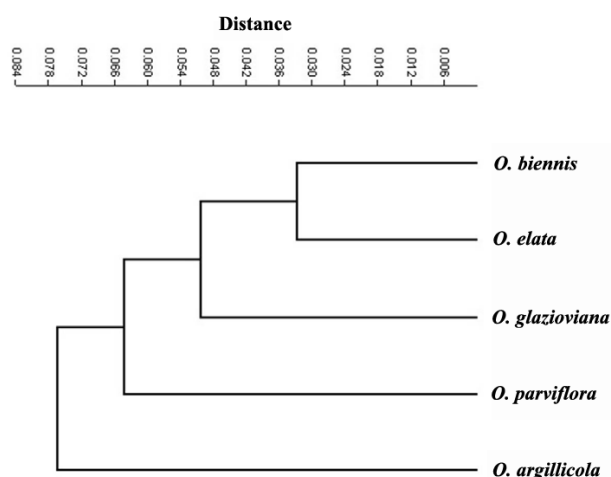


Figure 6. Cluster analysis. Dendrogram showing the extent of divergence of RSCU values of five *Oenothera* plastomes

out the influence of weak selection. In addition, neutrality plot showed narrow distribution of GC and no correlation was found between GC₃ and GC₁₂. Slope of the GC₁₂ Vs GC₃ plot was close to 0, indicating the role of specific evolutionary pressure (*i.e.*, selection pressure) in shaping SCUB. Thus, selection against mutational pressure may be acting on the PCGs at *Oenothera* plastomes, and intragenomic GC mutational bias on GC content was small similar to other plastomes (52). In a single stranded DNA, Chargaff's 2nd parity rule states that A almost equals T and G almost equals C (53, 54). However, PR2 bias plot analysis confirmed the deviation from Chargaff's 2nd parity rule in organellar DNA as A and T contents were used more proportionally than G and C contents in *Oenothera* plastomes.

Strand specific codon usage bias was observed in *Oenothera* plastomes: 6- and 4-fold degenerate amino acid for Arg and Val, respectively. This may be due to the intrinsic efficiencies of individual codons (55) and may not be correlated with translation efficiencies. Though 5 species of *Oenothera* are closely related, number of putative optimal codons varied for each species (*i.e.*, 1 to 7). All optimal codons used A/T at their ending position as observed in *C. arabica* (5), *T. aestivum* and *H. vulgare* (40). Thus mutational bias can be regarded as a major factor for SCU variations in *Oenothera* plastomes (24). If other selection pressures are absent, this mutational bias towards A/T ending codons would certainly increase the RSCU value of

synonymous/T ending codons to more than 1 (40).

The COA on RSCU values of *Oenothera* plastomes revealed no single major explanatory axis to explain the total variations. This pointed out that apart from the 2 major forces behind SCUB, *viz.*, mutational bias and natural selection, some other factors may be acting on the PCGs to cause SCUB. Similar observation was found in pooid grass models (40). CAI and ENC values have been proven as reliable indices for measuring the level of gene expression (32, 42). First axis of COA was in significant negative correlation with all examined species of *Oenothera* ($P < 0.01$). This suggested clearly that gene expression level also has considerable influence in SCU variation across PCGs. Influence of gene expression levels on SCUB in plant genes was recently reported in *Zea mays* (56) and also in *Oncidium ramsey* (34). Significant negative correlations of axis 2 with T₃, and axis 3 with GC₃ in all *Oenothera* species suggest the influence of T₃ and GC₃ contents in framing SCUB. Though no correlation was observed between length/ aromaticity and various axes of COA in all species, hydropathy score was in significant negative correlation with second axis in *O. Glazioviana*. This suggests the role of hydrophobic character of proteins in SCU variations as observed in *O. ramsey* (34) and in grass models (40). Moreover, certain codons were found to have significant negative correlation with axes 1 and 2 in all species. Among them, more than 60% of codons contained pyrimidine at the 3rd positions. These results suggested that mutational pressure combined with weak selection dictates SCU in all examined PCGs of *Oenothera* plastomes.

All examined plastomes belong to the subsection *Euoenothera* (*biennis* group) (29). Interestingly, high degree of phenotypic variation was observed among members of *biennis* group across various disjunct populations in different places of North America (29). Thus, small disjunct populations of *Oenothera* sp. are expected to experience genetic drift since random mutations in small population lead to random fixation over a period of time (57). Unexpected evolutionary changes are considered as a result of random process such as genetic drift rather than natural selection (59).

We conclude that the present finding would certainly facilitate studies on plant genome evolution as *Oenothera* sp. are considered to be suitable for studying compartmental co-evolution (30). Moreover, putative optimal codons for each species were identified and those codons can be used for optimization of heterologous gene expression by introducing point mutations (56).

Acknowledgements

There is no acknowledgement

Author Contributions

RRN and GD conceptualized the study. RRN and NTR contributed equally to this study, both of them equally carried out most of the experiments and wrote the manuscript. VRD, MBN and GD have critically revised the manuscript and the experimental design. MBN, TS and TCV helped in experiments. All the authors have read and approved the final manuscript.

Funding

No financial support from any agencies.

Financial Disclosure

The authors declare that they have no competing interests.

References

- Sharp PM, Emery LB, Zeng K. Forces that influence the evolution of codon bias. *Philos Trans R Soc Lond B Biol Sci.* 2010;**365**:1203-1212.doi:10.1098/rstb.2009.0305
- Ermolaeva MD. Synonymous codon usage in bacteria. *Curr Iss Mol Biol.* 2001;**3**:91-97.
- Grantham R, Gautier C, Gouy M, Mercier R, Pavé A. Codon catalog usage and the genome hypothesis. *Nucleic Acids Res.* 1980;**8**:49-62.doi:10.1093/nar/8.1.197-c
- Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol.* 1985;**2**:13-34.
- Nair RR, Nandhini MB, Monalisha E, Murugan K, Sethuraman T, Nagarajan S, Rao NS, Ganesh D. Synonymous codon usage in chloroplast genome of *Coffea arabica*. *Bioinformatics.* 2012;**8**:1096-1104.doi:10.6026/97320630081096
- Morton BR. Rates of synonymous substitution do not indicate selective constraints on the codon bias of the *psbA* gene. *Mol Biol Evol.* 1997;**14**:412-419.
- Parmley JL, Hurst LD. How do synonymous mutations affect fitness? *Bio Essays.* 2007;**29**:515-519.doi:10.1002/bies.20592
- Hershberg R, Petrov DA. Selection on codon bias. *Annu Rev Genet.* 2008;**42**:287-299.doi:10.1146/annurev.genet.42.110807.091442
- Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet.* 2011;**2**:32-42.doi:10.1038/nrg2899
- Agashe D, Gomez NCM, Drummond DA, Marx CJ. Good Codons, Bad Transcript: Large Reductions in Gene Expression and Fitness Arising from Synonymous Mutations in a Key Enzyme. *Mol Biol Evol.* 2013;**30**:549-560.10. doi:10.1093/molbev/mss273
- Bulmer M. The selection-mutation-drift theory of synonymous codon usage. *Genetics.* 1991;**129**:897-907.
- 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**:568-579.doi:10.1093/molbev/msm284
- Carlini DB, Chen Y, Stephan W. The relationship between third-codon position nucleotide content, codon bias, mRNA secondary structure and gene expression in the drosophilid alcohol dehydrogenase genes *Adh* and *Adhr*. *Genetics.* 2001;**159**:623-633.
- Akashi H. Inferring weak selection from patterns of polymorphism and divergence at silent sites in *Drosophila* DNA. *Genetics.* 1995;**139**:1067-1076.
- Bernardi G, Bernardi G. Compositional constraints and genome evolution. *J Mol Evol.* 1986;**24**:1-11.doi:10.1007/BF02099946
- Sharp PM, Stenico M, Peden JF, Lloyd AT. Codon usage: mutational bias, translational selection or both? *Biochem Soc Transact.* 1993;**21**:835-841.doi:10.1042/bst0210835
- Akashi H. Codon bias evolution in *Drosophila*. Population genetics of mutation-selection drift. *Gene.* 1997;**205**:269-278. doi:10.1016/S0378-1119(97)00400-9
- Kurland CG. Major codon preference theme and variations. *Biochem Soc Trans.* 1993;**21**:841-846.doi:10.1042/bst0210841
- Deschavanne I, Filipinski J. Correlation of GC content with replication timing and repair mechanisms in weakly expressed *E. coli* genes. *Nucleic Acids Res.* 1995;**23**:1350-1353.doi:10.1093/nar/23.8.1350
- Irwin B, Heck JD, Hatfield GW. Codon pair utilization biases influence translational elongation step times. *J Biol Chem.* 1995;**270**:22801-22806.doi:10.1074/jbc.270.39.22801
- Clegg MT, Learn GH, Golenberg EM. *Molecular evolution of chloroplast DNA*. In: Selander RK, Clark AG, Whittam TS (Eds). *Evolution at the Molecular Level*. Sunderland, Sinauer Publishers, England, 1991;PP.135-149.
- Downie SR, Palmer JD. *Use of chloroplast DNA rearrangements in reconstructing plant phylogeny*. In: Soltis, P.S., Soltis DE, Doyle, JJ, (Eds.). *Molecular Systematics of Plants*, Chapman and Hall Publishers, New York, 1992;PP.14-35.
- Wolfe KH, Morden CW, Ems SC, Palmer JD. Rapid evolution of the plastid translational apparatus in a nonphotosynthetic plant: loss or accelerated sequence evolution of tRNA and ribosomal protein genes. *J Mol Evol.* 1992;**35**:304-317.doi:10.1007/BF00161168
- Morton BR. Chloroplast DNA codon use: evidence for selection at the *psbA* locus based on tRNA availability. *J Mol Evol.* 1992;**37**:273-280.doi:10.1007/BF00175504
- Morton BR. Codon use and the rate of divergence of land plant chloroplast genes. *Mol Biol Evol.* 1994;**11**:231-238.
- Morton BR. Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages. *J Mol Evol.* 1998;**46**:449-59.doi:10.1007/PL00006325
- Qin Z, Cai Z, Xia G, Wang M. Synonymous codon usage bias is correlative to intron number and shows disequilibrium among exons in plants. *BMC Genomics.* 2013;**14**:1-11.doi: 10.1186/1471-2164-14-56
- Frean M, Balkwill K, Gold C, Burt S. The expanding distributions and invasiveness of *Oenothera* in southern Africa. *S Afr J Bot.* 1997;**63**:449-458.
- Cleland RE. The evolution of the North American *oenotheras* of the "biennis" group. *Planta.* 1958;**51**:378-398.
- Greiner S, Wang X, Rauwolf U, Silber MV, Mayer K, Meurer J, Haberer G, Herrmann RG. The complete nucleotide sequences of the five genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: I. Sequence evaluation and plastome evolu-

- tion. *Nucleic Acids Res.* 2008;**36**:2366-2378.doi:10.1093/nar/gkn08
31. Sharp PM, Li WH. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for rare codon. *Nucleic Acid Res.* 1986;**14**:7737-7749.doi:10.1093/nar/14.19.7737
 32. Wright F. The "effective number of codons" used in a gene. *Gene.* 1990;**87**:23-29.doi:10.1016/0378-1119(90)90491-9
 33. Sharp PM, Li WH. The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acid Res.* 1987;**15**:1281-1295.doi:10.1093/nar/15.3.1281
 34. Xu C, Cai X, Chen Q, Zhou H, Cai Y, Ben A. Factors Affecting Synonymous Codon Usage Bias in Chloroplast Genome of *Oncidium Gower Ramsey*. *Evol Bioinform.* 2011;**7**:271-278.doi:10.4137/EBO.S8092
 35. Wan XF, Xu D, Kleinhofs A, Zhou J. Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes. *BMC Evol Biol.* 2004;**4**.doi:10.1186/1471-2148-4-19.
 36. Perriere G, Thioulouse J. Use and misuse of correspondence analysis in codon usage studies. *Nucleic Acids Res.* 2002;**30**:4548-4555.doi:10.1093/nar/gkf565
 37. Wang HC, Hickey DA. Rapid divergence of codon usage patterns within the rice genome. *BMC Evol Biol.* 2007;**7**.doi:10.1186/1471-2148-7-S1-S6.
 38. Liu Q, Feng Y, Xue Q. Analysis of factors shaping codon usage in the mitochondrion genome of *Oryza sativa*. *Mitochondrion.* 2004;**4**:313-320.doi:10.1016/j.mito.2004.06.003
 39. Roychoudhury S, Mukherjee D. A detailed comparative analysis on the overall codon usage pattern in herpesviruses. *Virus Res.* 2010;**148**:31-43.doi:10.1016/j.virusres.2009.11.018
 40. Sablok G, Nayak KC, Vazquez F, Tatarinova TV. Synonymous codon usage, GC3, and evolutionary patterns across plastomes of three pooid model species: Emerging grass genome models for monocots. *Mol Biotechnol.* 2011;**49**:116-128.doi:10.1007/s12033-011-9383-9
 41. Zhou M, Li X. Analysis of synonymous codon usage patterns in different plant mitochondrial genomes. *Mol Biol Rep.* 2009;**36**:2039-2046.doi:10.1007/s11033-008-9414-1
 42. Gupta SK, Bhattacharyya TK, Ghosh TC. Synonymous codon usage in *Lactococcus lactis*: mutational bias versus translational selection. *J Biomol Struct Dyn.* 2004;**21**:527-536.doi:10.1080/07391102.2004.10506946
 43. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 2011;**28**:2731-2739.doi:10.1093/molbev/msr121
 44. Xia X. DAMBE5: A comprehensive software package for data analysis in molecular biology and evolution. *Mol Biol Evol.* 2013;**30**(7):1720-1728.doi:10.1093/molbev/mst064
 45. Peden JF. Analysis of Codon Usage, Ph.D. Thesis, University of Nottingham, 1999.
 46. Wu G, Culley DE, Zhang W. Predicted highly expressed genes in the genomes of *Streptomyces coelicolor* and *Streptomyces avermitilis* and the implications for their metabolism. *Microbiol.* 2005;**151**:2175-2187.doi:10.1099/mic.0.27833-0
 47. Angellotti MC, Bhuiyan SB, Chen G, Wan XF. CodonO: codon usage bias analysis within and across genomes. *Nucleic Acids Res.* 2007;**35**:132-136.doi:10.1093/nar/gkm392
 48. Knight RD, Freeland SJ, Landweber LF. Rewiring the keyboard: Evolvability of the genetic code. *Nat Rev Genet.* 2001;**2**:49-58.doi:10.1038/35047500
 49. Hammer Q, Harper DAT, Ryan PD. PAST: Paleontological statistics software package for education and data analysis. *Palaeontol Electron.* 2001;**4**:1-9.
 50. Sueoka N. On the genetic basis of variation and heterogeneity of DNA base composition. *Proc Natl Acad Sci USA.* 1962;**48**:582-592.
 51. Meng Z, Wei L, Xia L. Analysis of synonymous codon usage in chloroplast genome of *Populus alba*. *J Forestry Res.* 2008;**19**:293-297.doi:10.1007/s11676-008-0052-1
 52. Liu Q, Xue Q. Comparative studies on codon usage patterns of chloroplasts and their host nuclear genes in four plant species. *J Genet.* 2005;**84**:55-62.doi:10.1007/BF02715890
 53. Rudner R, Karkas JD, Chargaff E. Separation of *B. subtilis* DNA into complementary strands, 3. Direct analysis. *Proc Natl Acad Sci USA.* 1968;**60**:921-922.
 54. Nikolaou C, Almirantis Y. Deviations from Chargaff's second parity rule in organellar DNA Insights into the evolution of organellar genomes. *Gene.* 2006;**381**:34-41.doi:10.1016/j.gene.2006.06.010
 55. Nakamura M, Sugiura M. Translation efficiencies of synonymous codons for arginine differ widely and are not correlated with codon usage in chloroplasts. *Gene.* 2011;**472**:50-54.doi:10.1016/j.gene.2010.09.008
 56. Liu H, He R, Zhang H, Huang Y, Tian M, Zhang J. Analysis of synonymous codon usage in *Zea mays*. *Mol Biol Rep.* 2010;**37**:677-684.doi:10.1007/s11033-009-9521-7
 57. Kimura M. The neutral theory of molecular evolution and the world view of the neutralists. *Genome.* 1989;**31**:24-31.
 58. Johnson MTJ, Vellend M, Stinchcombe JR. Evolution in plant populations as a driver of ecological changes in arthropod communities. *Philos Trans R Soc Lond B Biol Sci.* 2009;**364**:1593-1505.doi:10.1098/rstb.2008.0334