ISSN 2816-8089

RESEARCH ARTICLE

Analysis for Molecular Distinction in the Chloroplast DNA Sequences of *Gymnospora montana* (Celastraceae) and *Belanites aegyptiaca* (Balanitaceae) from Semi-arid Area

Jagdishchandra K. Monpara[1,2], Kiran S. Chudasama[1], Vrinda S. Thaker[1*]

[1]*Plant Biotechnology and Genetic Engineering Laboratory, Department of Biosciences, Saurashtra University, Rajkot- 360 005, Gujarat, India.*
[2]*Vimal Research Society for Agro-Biotech and Cosmic Power, 80 feet road, Aji area, Rajkot 360 003, Gujarat, India.*

Abstract

Two medicinally important plants *Gymnospora montana* (Celastraceae) and *Belanites aegyptiaca* (Balanitaceae) (belongs to order Celastrales and Zygophyllales, respectively) showed marked similarity in their cpDNA sequences. However, earlier reports showed *B. aegyptiaca* in the family Zygophyllaceae. The detailed nucleotide analysis may help to understand evolution of the plastomes of these families and therefore, detail analysis of their cpDNA sequences is performed for codon use bias and its index, relative synonymous codon uses value (RSCU), effective number of codons (ENC), GC content of the gene and frequencies of the nucleotides G+C at various positions in synonymous codon were calculated and compared it with *Tribulus terrestris.* Length of the gene and ENC showed close relationship which suggests that longer genes have less codon bias. The codons for leucine, isoleucine and serine were most abundant in the studied plant species. The correlation analysis suggested that codon usage patterns in both cp genomes appear due to the different forces; natural selection, mutation pressure, GC content of gene and protein length. Their role in the gene evolution process is discussed.

Key Words: *cpDNA; Codon usage bias; Evolulation; G. montana; B. aegyptiaca*

***Corresponding Author**: Vrinda S. Thaker, Plant Biotechnology and Genetic Engineering Laboratory, Department of Biosciences, Saurashtra University, Rajkot- 360 005, Gujarat, India; E-mail: thakervs@gmail.com*

***Received Date**: January 06, 2023, **Accepted Date**: February 06, 2023, **Published Date**: February 14, 2023*

***Citation:** Monpara JK, Chudasama KS, Thaker VS. Analysis for Molecular Distinction in the Chloroplast DNA Sequences of Gymnospora montana (Celastraceae) and Belanites aegyptiaca (Balanitaceae) from Semi-arid Area. Int J Bioinfor Intell Comput. 2023;2(1):82-98.*

1. Introduction

Codons play a crucial role in the process of genetic information transmission from mRNA to protein. Codons for each amino acid must be accurately identified to ensure that genetic information is correctly expressed. In genetic code, more than one codon often encodes the same amino acid and is known as synonymous codons for a specific amino acid. Except two amino acids methionine and tryptophan, remaining all other amino acid are encoded by 2-6 synonymous codons [1]. The unequal use of synonymous codons is known as codon usage bias (CUB). Synonymous codons are used with unequal frequencies in genomes, gene of same species genome and within a single gene [2]. Codon usage patterns can be affected by natural selection, mutation pressure, recombination rate, translational selection, gene length, tRNA abundance, GC composition protein secondary structures, length of the intron, environmental stress and during the process of the genome and gene evolution [3-5]. Codon usage plays an important role during the chloroplast genome evolution [6].

Chloroplasts are subcellular organelles essential for the photosynthesis and metabolism of green plants. Compared to the nuclear genome, the chloroplast genome, possesses many characteristics, including simple, small size, and highly conserved. It is widely used in research such as identification, phylogenetics and adaptive analysis [7]. In recent years, with the development of high throughput sequencing technology for chloroplast genomes, a large number of chloroplast genomes have been sequenced, and the codon usage bias of chloroplast genomes has become possible. Therefore, the cp genome is a special tool for plant systems evolution research [8].

In the present study, two medicinally important plants cp genomes i.e., *Balanites aegyptiaca* and *Gymnosporia montana* are selected. All parts of plant i.e., root, stem bark, leaves, and fruit pulp have possessed medicinal properties. These plants are widespread in the semi-arid regions of India and taxonomically belong to the Eudicot clade (APG III, 2009). The medicinal plant *B. aegyptiaca* is placed in family Zygophyllaceae by some taxonomists [9] and it is placed in the separate family Balanitaceae because of its distinguishing morphological and anatomical characters [10]. *G. montana* (Celastraceae) is also possess pharmaceutical important compounds [11] and showed high similarity with the cp genome of *B. aegyptiaca*.

Complete cpDNA sequence of members of Zygophyllaceae and allied plant species showed marked similarity. For example, in nucleotide pattern. *G. montana* was observed in the Zygophyllaceae clade instead Celastraceae [12] (Figure 1).

Both these plants are taxonomically belonging to closely related families and hence this study is aimed to explore and understood the organelle's genome structure organization and evolution of these plants. In this work, we performed a comparative analysis on CUB of protein coding genes in cp genomes of *B. aegyptiaca* with *G. montana* and *Tribulus terrestris*. We report an analysis of codon usage and find the key factors that shape codon patterns in protein-coding genes of the selected plants cp genome.

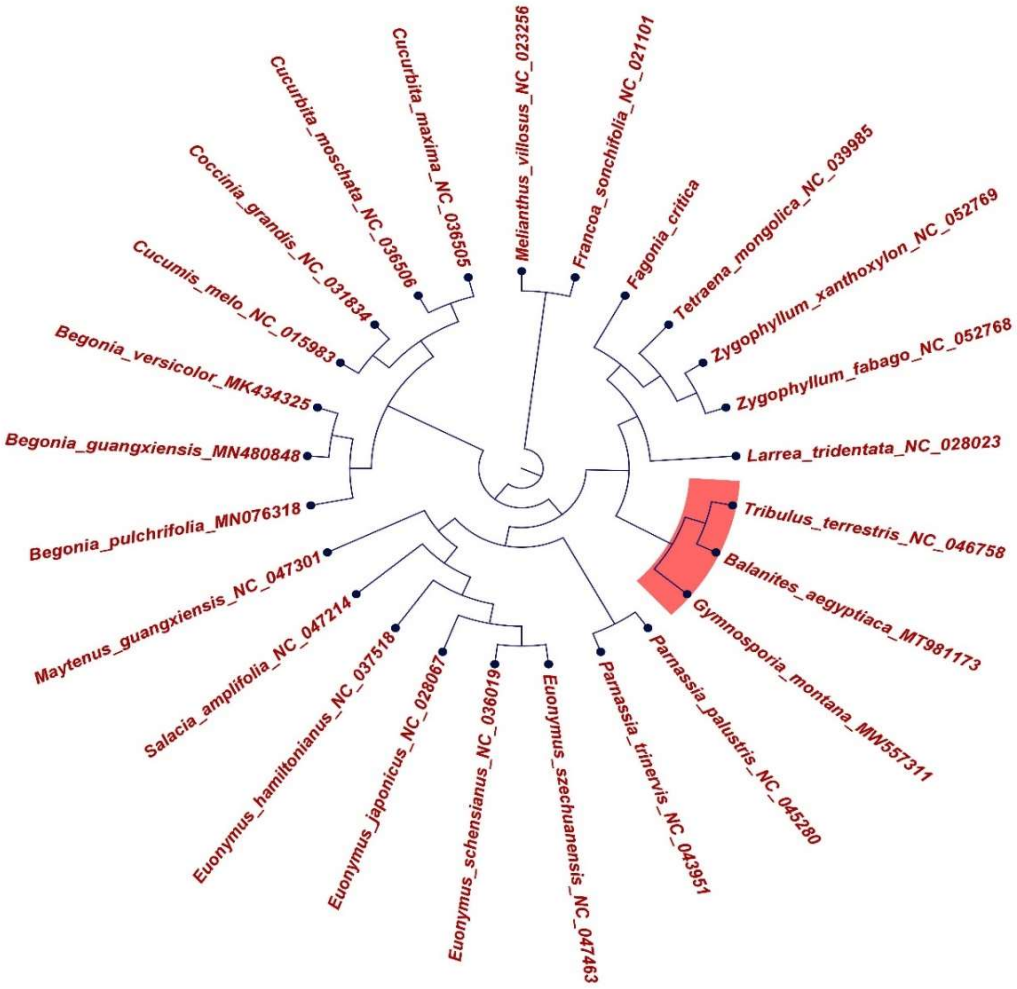

Figure 1: *Phylogenetic position of B. aegyptiaca, and G. montana by maximum likelihood analysis of twenty-five completed chloroplast genomes of Rosids order.*

2. Materials and Method

2.1. Plant material, chloroplast DNA extraction, sequencing of genomes

Approximately 100g fresh leaves of *B. aegyptiaca* and *G. montana*, were collected from the Botanical Garden of Saurashtra University campus, Rajkot, Gujarat, India and kept in dark for 48 hours at 4°C to decrease starch level. Chloroplast DNA isolation was performed according to Bhatt and Thaker [13]. DNA was checked through 1% agarose gel electrophoresis. The quality and concentration of the DNA was confirmed by measuring the optical density at 260/280 nm. DNA was quantified by the spectrophotometric method using Microplate reader (μ Quant, Bio-Tek Instruments, USA). This cpDNA was sequenced using a high throughput Ion torrent genome machine with an Ion torrent server (torrent suite v3.2). cp genome was assembled using CLC Genomic workbench v9.5.64 (CLC bio, QIAGEN). The assembly of *B.*

aegyptiaca and *G. montana* was performed in the reference-assisted mode using the *Tribulus terrestris* (NC_046758) and Parnassia trinervis (NC_043951) as the reference genome, respectively. Complete chloroplast genomes were annotated using the online program Dual Organellar GenoMe Annotator (DOGMA) [14] and (CpGAVAS) Chloroplast Genome Annotation Visualization Analysis [15]. The circular map of cp genome was prepared using the CpGAVAS. The LSC, SSC, and IR values were measured using the CpGAVAS and Vmatch software tool with default parameter [16]. Finally, the complete cp genome of *B. aegyptiaca* and *G. montana* was submitted to the NCBI GenBank database (Accession Number: MT981173, MW557311).

2.2. Codon usage bias and related indices analysis

The number of codon usage indicators was estimated via the software DnaSP 6. and MEGA7, including the relative synonymous codon usage value (RSCU), codon bias index (CBI), the effective number of codons (ENC), G+C content of the gene (GC), the frequency of the nucleotides G+C at the 1st, 2nd and 3rd position of synonymous codons (GC3s).

2.3. Codon usage

Codon usage was determined for all protein-coding genes of *T. terrestris*, *B. aegyptiaca*, and *G. montana*. The relative synonymous codon usage (RSCU) values and codon usage were determined with MEGA7 [17]. Amino acid frequency was also calculated and expressed in percentage. RSCU, the observed frequency of codon divided by the frequency expected, is an important indicator of CUB [18]. RSCU values are close to 1.0 when all synonymous codons are used equally without any bias, while they are >1 when synonymous codons are used more frequently than expected and <1 when synonymous codons are used less frequently than expected [19].

2.4. ENC-plot (ENC versus GC3s)

ENC-plot mapping analysis of *B. aegyptiaca*, and *G. montana* was employed to analyze and determine the crucial factors influencing the codon usage bias. The ENC value range from 20 (each amino acid uses only one synonymous codon) to 61 (Each synonymous codon is equally used), which is inversely proportional to the codon bias GC3s, value refers to the ratio of G and C content at the third position of one codon. The codon usage pattern across genes was examined by a plot of ENC versus GC3s. The ENC plot reflects the relationship of the ENC values against the GC3S values. ENC values are located on or near the expected curve, when mutation pressure plays a key role in the formation of codon usage patterns. Conversely, when the use of a codon is constrained by natural selection, the ENC value will be well below the prospective curve [20].

2.5. Analysis of neutrality plot

A neutrality plot (GC12 vs. GC3) of *B. aegyptiaca*, and *G. montana* was performed to investigate the extent of influence between mutation pressure and natural selection on the patterns of codon usage. The slope of the plot regression was zero indicating no effects of directional mutation pressure (complete selective constraints). The Slope1 depicted that the codon usage

bias is completely affected by directional mutation pressure representing complete neutrality [21,22].

3. Results and Discussion

3.1. Chloroplast genome features of *B. aegyptiaca* and *G. montana*

The Genome sequencing and assembly of Ion Torrent PGM sequencing produced >50X of data for the chloroplast genomes of *B. aegyptiaca* and *G. montana*. The complete circular cp genome sequence of *B. aegyptiaca* and *G. montana* is 15,3425bp and 14,5309bp in length respectively (Figure 2). The *B. aegyptiaca* and *G. montana* genome present a common quadripartite structure of similar size to the majority of angiosperm species and consists of two copies of IR regions (24805/ 23454 bp) that are separated by the LSC region (86565/77654bp) and the SSC region (17415/20755bp) Table 1.

Table 1: *Comparative studies on genomic data of B. aegyptiaca, G. montana, T. terrestris and P. trinervis.*

Feature	*Balanites aegyptiaca*	*Tribulus terrestris*	*Gymnosporia montana*	*Parnassia trinervis*
Entire chloroplast genome size (bp)	153425	158184	145309	153590
LSC region (bp)	86565	88878	77654	69455
SSC region (bp)	17415	17622	20755	18287
IRa region (bp)	24805	25842	23454	25582
IRb region (bp)	24805	25842	23454	25582
No. of genes	132	130	126	133
No. of Proteins	91	84	90	87
No. of tRNA	33	37	28	37
No. of rRNA	8	8	8	8
No. of genes with introns	15	15	12	11
GC content (%)	35.8%	35.8%	37.3%	37.0%
GC content for gene sequences	38.6%	39.2%	39.2%	39.8%
GC content for coding sequences	37.1%	37.3%	37.6%	38.1%
GC content for rRNA genes	53.2%	55.3%	54.9%	54.7%
GC content for tRNA genes	43.7%	53.2%	52.9%	53.3%
NCBI accession numbers	MT981173	NC_046758	MW557311	NC_043951

3.2. Analysis of synonymous codon usage patterns

In our study, the codon usage bias in the plastome was computed using the protein-coding gene sequence of *B. aegyptiaca* and *G. montana* and further compared with the *T. terrestris* member of Zygophyllaceae family cp genome. These protein gene sequences encoded 26691 (*B. aegyptiaca*), 24181 (*G. montana*), and 26558 (*T. terrestris*) codons, including the stop codon. The relative synonymous codon usages (RSCU) of all three cp genomes are presented (Figure

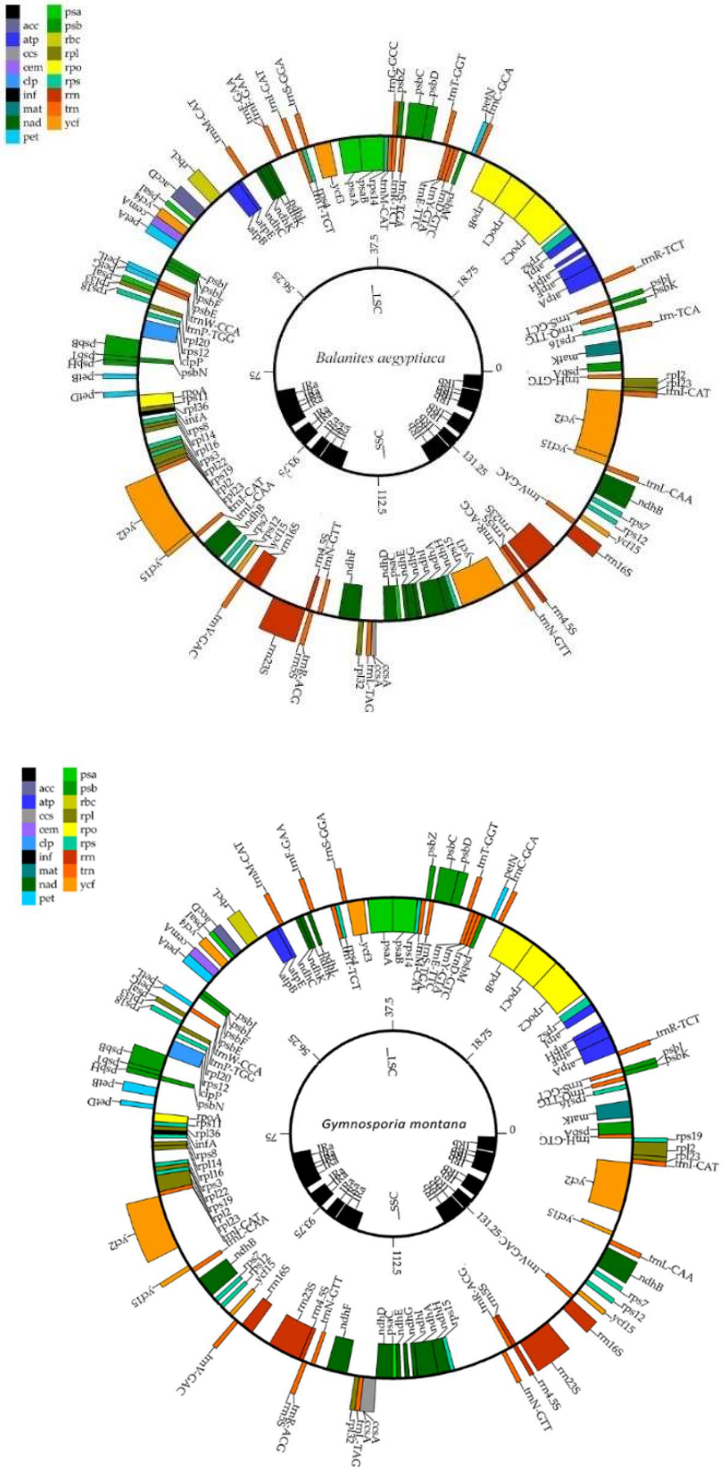

Figure 2: *Genome map of (A) B. aegyptiaca (B) G. montana chloroplast prepared using CPGAVAS server Genes shown outside are transcribed in the clockwise direction while genes shown on the inside of the circle are transcribed in the anti-clockwise direction. Legend indicates the functional group to which each gene belongs.*

3). The codon AGA(R) encoding arginine exhibited the highest RSCU value in the range of 1.78-2.2 in all plants. The RSCU value of >1 was in twenty-nine, thirty-one and thirty codons codon usage bias in *B. aegyptiaca*, *G. montana* and *T. terrestris* cp genome, respectively. Out of these 29 codons, twenty-eight were A+U ending codons. Only one codon, UUG, has a RSCU >1 value in *B. aegyptiaca* and *T. terrestris*, while in *G.montana* out of 31, three codons, UUG, UAG, AGG have RSCU>1 value. Conversely, the G+C ending codons exhibited the opposite pattern (RSCU<1), representing that they are less common in this cp genome. There are only two codons, methionine (AUG) and tryptophan (UGG), amino acids with no codon bias. The data presented in Figure 3 illustrates that the RSCU value increases with the number of codons that code for a specific amino acid. High codon preference, especially a strong A or U bias in codon usage is reported in many angiosperm land plants Forsythia suspense, Aquilaria sinensis, Euphorbiaceae of chloroplast genomes [23-26].

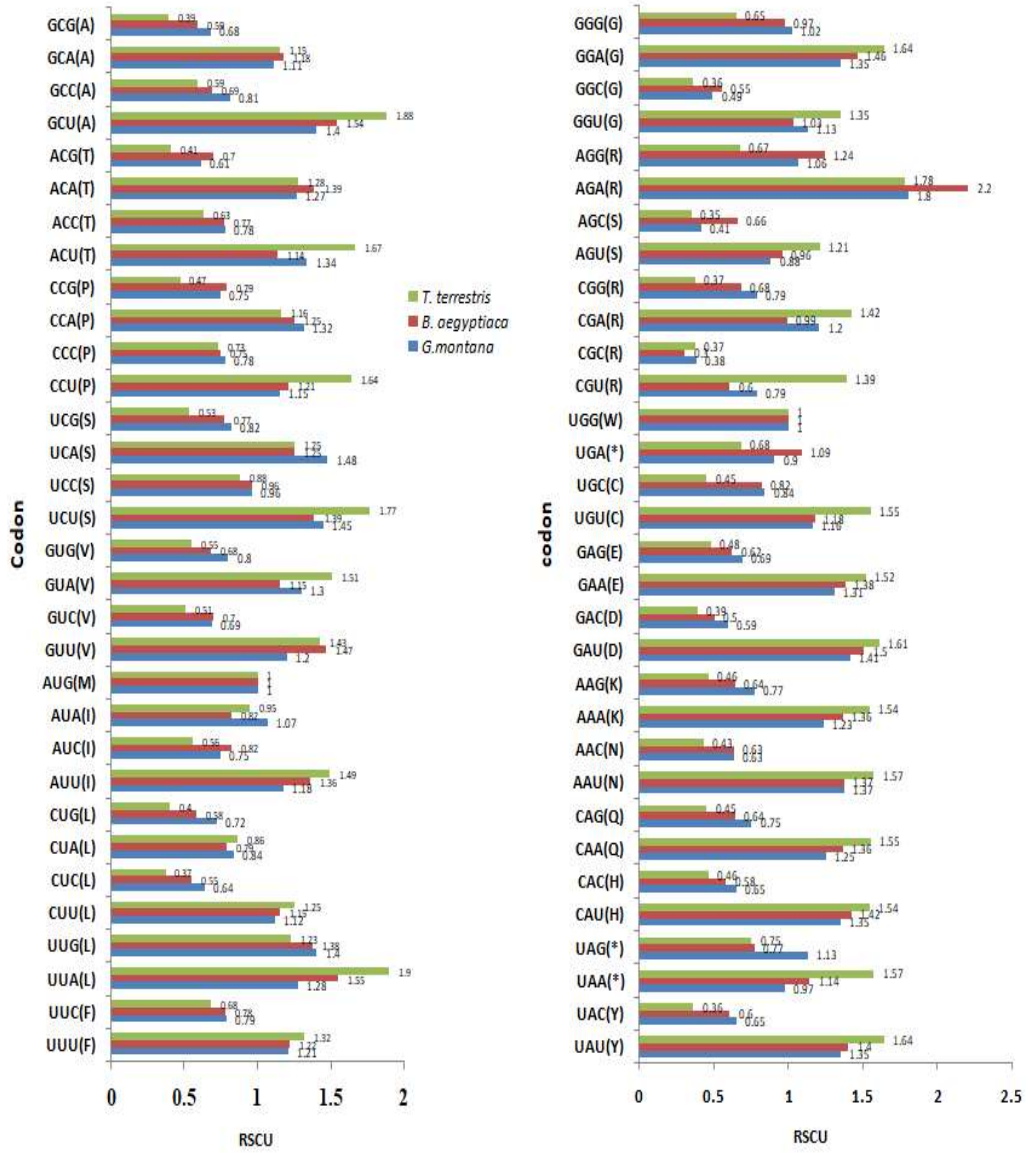

Figure 3: *Overall RSCU values of codons in G. montana, B. aegyptiaca and T. terrestris cp genomes.*

The distribution of 20 amino acids and stop codons in all the protein coding genes of all three cp genomes are presented in Figure 4. Some slight variation in amino acid distribution was observed in all three genomes. The codons for leucine, isoleucine and serine codons were the most abundant, whereas the codon coding for tryptophan and cysteine is the least number in all three genomes. Leucine and isoleucine are the most commonly observed amino acids in the proteins of chloroplast genomes of many angiosperms plants [27,28].

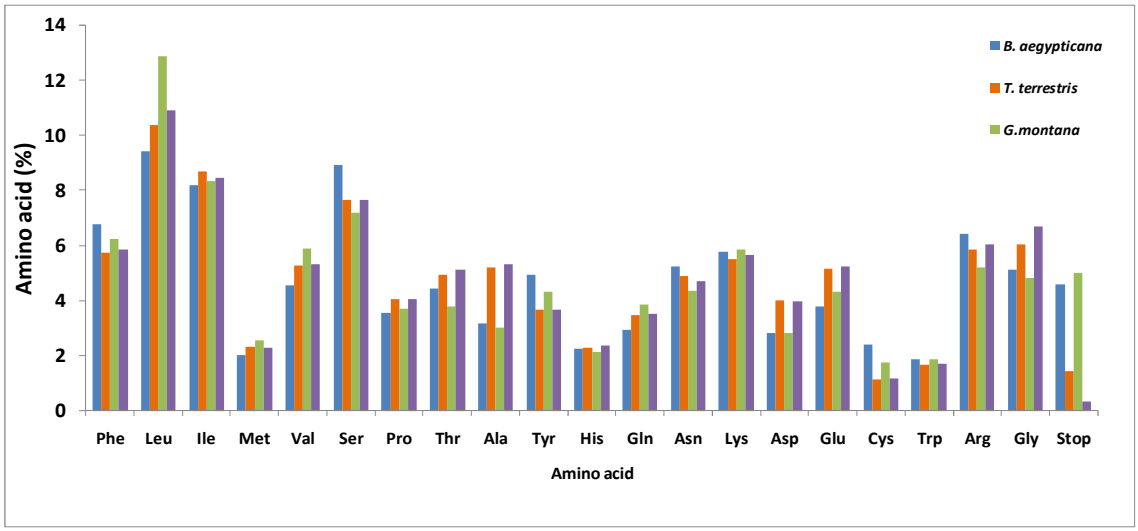

Figure 4: *Composition of amino acid codons in B. aegyptiaca, T. terrestris, and G. montana.*

The Codon bias index (CBI) measures the extent that which preferred codons are used in a gene [29]. In this work, the CBI of genes for *G. montana*, *B. aegyptiaca*, and *T. terrestris* genomes is calculated. The codon bias index was observed in the range 0.177-0.828, 0.175-0.828, and 0.228-0.878 for *G. montana*, *B. aegyptiaca*, and *T. terrestris*, respectively (Figure 5). The CBI is an overall measure of codon bias for the entire gene [30]; the CBI ranges from 0 (no codon bias in the gene) to 1 (maximum codon bias).

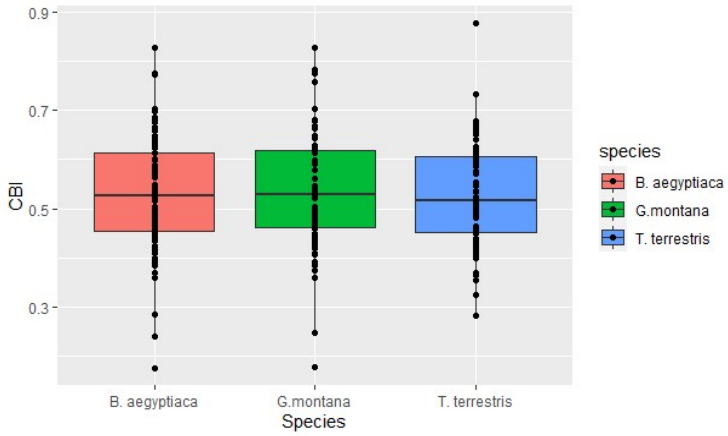

Figure 5: *CBI value in cp genome.*

We also measured the GC content at different codon positions (second, third, and all) in the CDSs (Figure 6). The *G. montana* GC content was observed in the range 0.293-0.478, 0.265-

0.587, and 0.102-0.533 for GCc, GC2, and GC3, respectively. A similar pattern was observed for *B. aegyptiaca* 0.266-0.469 (GCc), 0.219-0.587 (GC2), and 0.102-0.583 (GC3). Some minor variation was observed for GC content of *T. terrestris* in the range 0.285-0.469 (GCc), 0.258-0.565 (GC2), and 0.133-0.415 (GC3). The GC content of the second and third codon positions increased to a lesser extent. Similar patterns of changes in all three codon positions were observed in all three cp genomes.

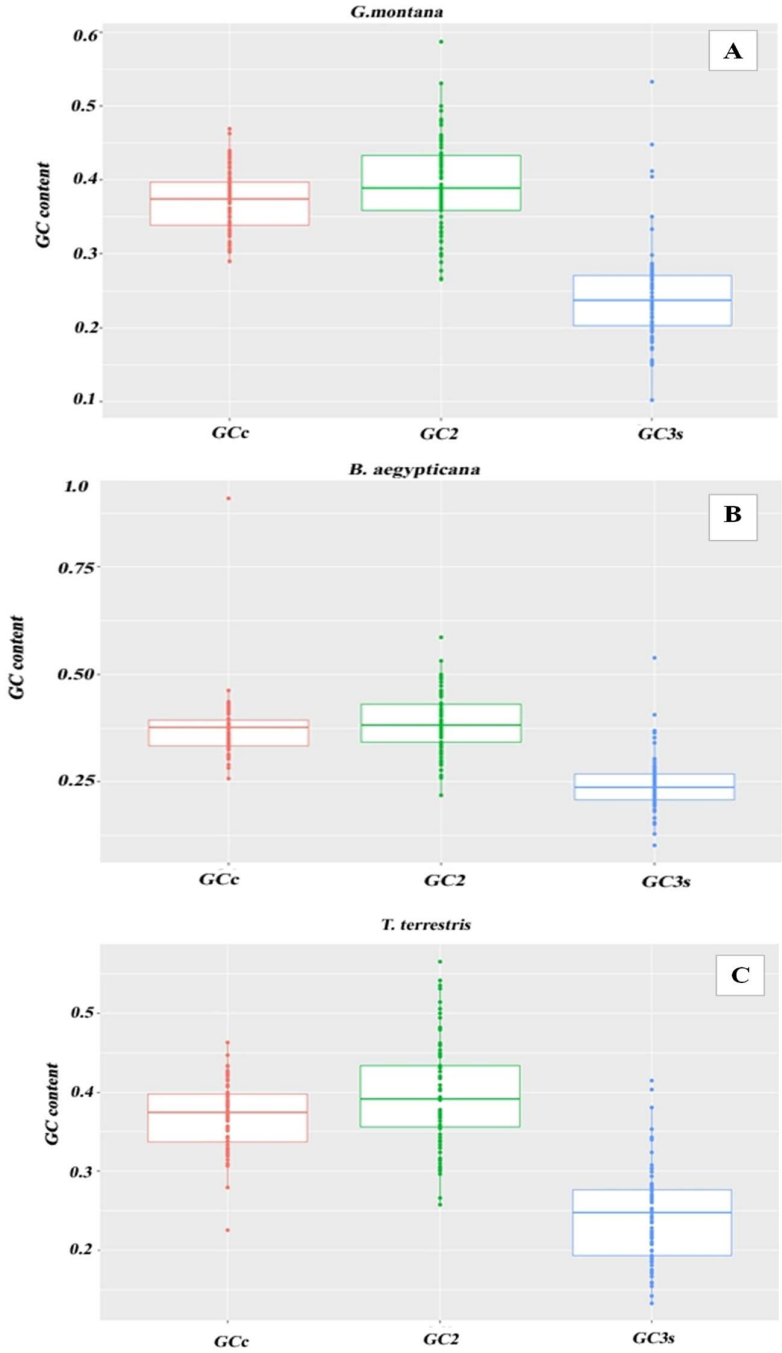

Figure 6: *Box plot of GC contents variation in different codon positions (2nd, 3rd and all (overall cp genome)) (A) G. montana (B) B. aegyptiaca (C) T. terrestris.*

ENC is an important index to measure the codon usage bias in the genome. ENC plays a key role in their codon usage profile. ENC value of individual genes of the cp genome showed much less variability of bias, with a value of 28-61 for all three species. To better understand the relations between *B. aegyptiaca* and *G. montana* and *T. terrestris* gene composition and codon usage bias, the ENC–GC3 scatter diagram was constructed (Figure 7). ENC plot was widely used to investigate the determinants of the codon usage variation among genes in the cp genome. The solid curve represents the expected position of CDSs of the cp genome, whose codon usage was only formed due to the GC3s. Each point represents one coding gene in *B. aegyptiaca* and *G. montana* and *T. terrestris* cp genome. The distributions of ENC and GC3s of chloroplast genomes of all plant species were similar (Figure 7). In our work, it was observed in Belanitaceae and Celastraceae plants that few genes were situated on or close to the standard curve, which suggests that genes are under extreme mutation pressure. Half of the genes were positioned below the expected curve, indicating the gene expression is subject to natural selection. Analysis of the ENC-GC3 plot indicates that codon usage bias of the plants of cp genome was affected by the combined effect of mutation pressure and natural selection. Earlier, many researchers suggested the codon usage bias of the chloroplast genome of Morus notabilis [22], Brassica campestris [31], Poaceae [32], and Asteraceae [33] were influenced by the effect of both mutation pressure, natural selection and other factors. Mutation pressure and natural selection have been considered major evolutionary forces that influenced the codon usage patterns of genes [30,34].

The influence of mutation pressure and natural selection on *G. montana*, B. *aegyptiaca*, and *T. terrestris*, codon usage was evaluated using a neutrality analysis (Figure 8). The neutrality plot was drawn from the regression analysis of GC12 as Y-axis and GC3 as X-axis. In the plot, each point represented one coding gene. In the neutrality plot analysis (Figure 8), a positive correlation was observed between the GC12 and GC3 values of (r2=0.023, *G. montana*), (r2=0.026, *B. aegyptiaca*) and (r2=0.002, *T. terrestris*). The slopes of the linear regression were 0.15833, 0.1763, and 0.04008 for *G. montana*, *B. aegyptiaca*, and *T. terrestris* coding sequences (Figure 8), respectively. These results indicate that mutation pressure accounted for 15.8%, 14.7%, and 4.0% of the selection force for the *G. montana*, *B. aegyptiaca*, and *T. terrestris* for the coding sequence, whereas natural selection accounted for 84.28%, 82.37%, and 96% respectively. Our results indicate that GC12 versus GC3 content is due to both mutation pressure and natural selection [35]. Our results suggest that natural selection played a major role in influencing the codon usage pattern in the cp genome of *G. montana* and *B. aegyptiaca*, while mutation pressure had a minor role during the evolution process. Earlier, many authors reported that natural selection is the dominant force rather than mutation pressure in evolution [36-39].

Generally, it has been observed that codon usage bias is affected by gene length. We have determined a correlation analysis between ENC and protein length (Figure 9). The majority of genes were distributed between ENC values of 35 and 55. In *G. montana*, positive linear correlations were found between protein length and ENC ($r=0.247$, $P<0.001$) in the *B. aegyptiaca* ($r=0.193$, $P<0.001$) and *T. terrestris* ($r=0.153$, $P<0.001$) correlation was similar. The results indicated that protein length shaped codon usage in *G. montana*, B. *aegyptiaca*, and T. *terrestris*, the longer genes had a lower degree of codon bias. Overall results observed that longer protein length has weaker codon usage bias, while short protein did not affect codon usage. This result indicates that the codon usage pattern in *G. montana*, *B. aegyptiaca*, and *T. terrestris* might be shaped by other selection constraints. Earlier studies have shown that gene length is positive and negatively affects codon usage bias. A negative correlation was

observed between gene length and codon usage pattern in cucumber, melon, and Arabidopsis thaliana, whereas a positive correlation was observed in S. indicum, chayote, and wax gourd [40-42]. Translational selection is attributed to both positive and negative correlations. This codon usage patterns research not only provides information about the variation in the cp genome and evolutionary mechanisms but also contributes to investigating the factors that force *G. montana* and *B. aegyptiaca* evolution.

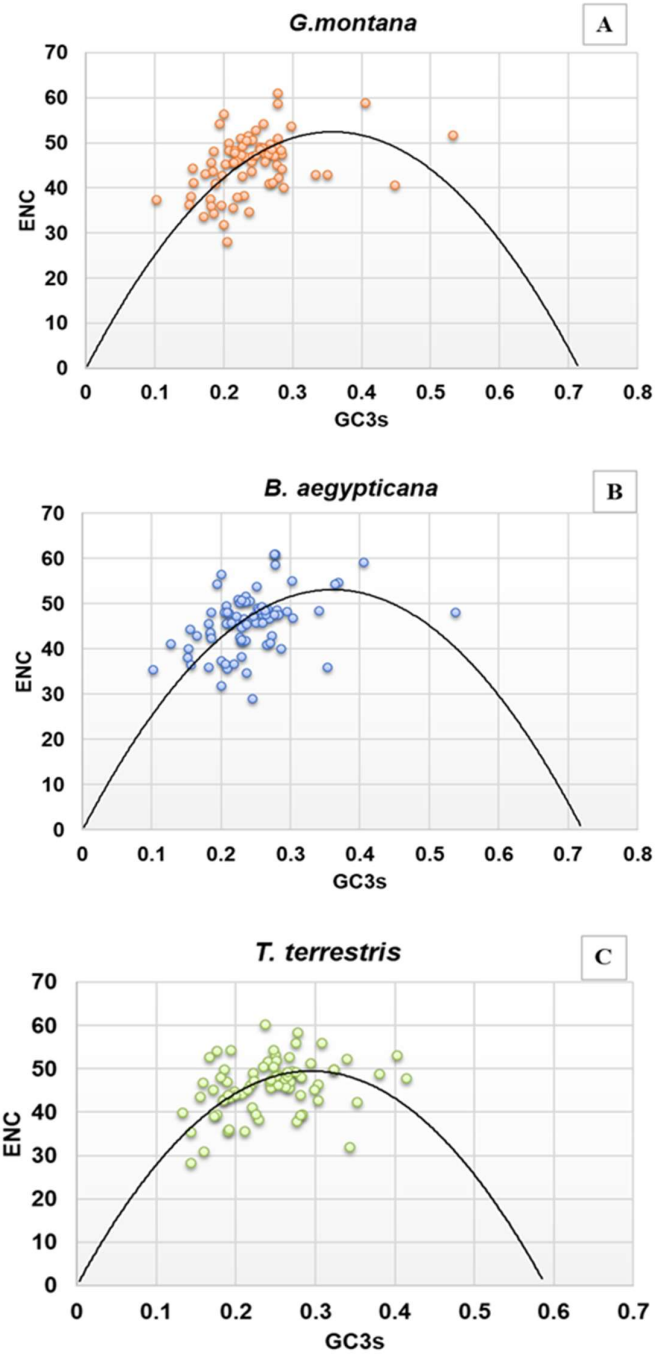

Figure 7: *Distribution of Effective number of codons (ENC)-plots of the chloroplast genomes. Solid lines indicate the expected ENC values (A) G. montana (B) B. aegyptiaca (C) T. terrestris.*

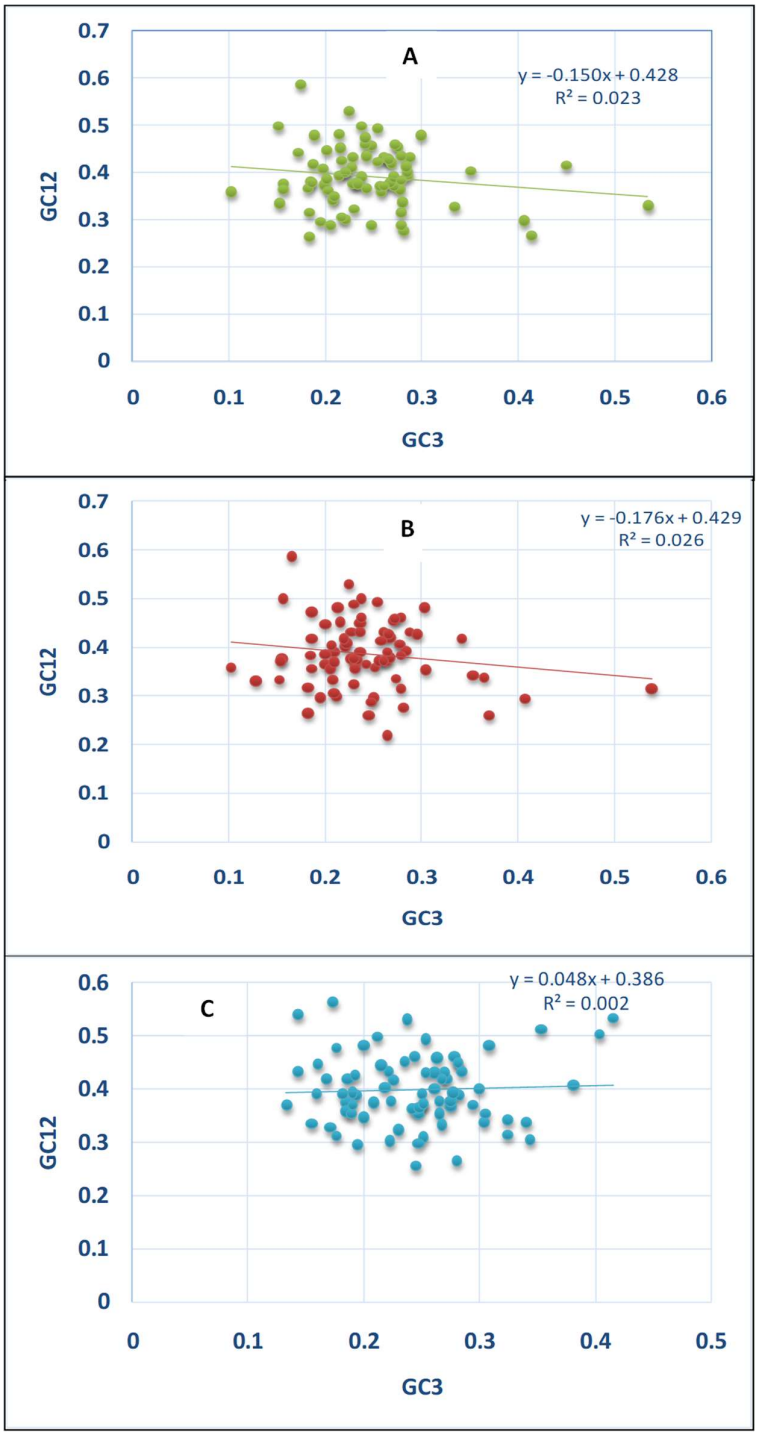

Figure 8: *Neutrality plots for chloroplast genes in (A) G. montana (B) B. aegyptiaca (C) T. terrestris.*

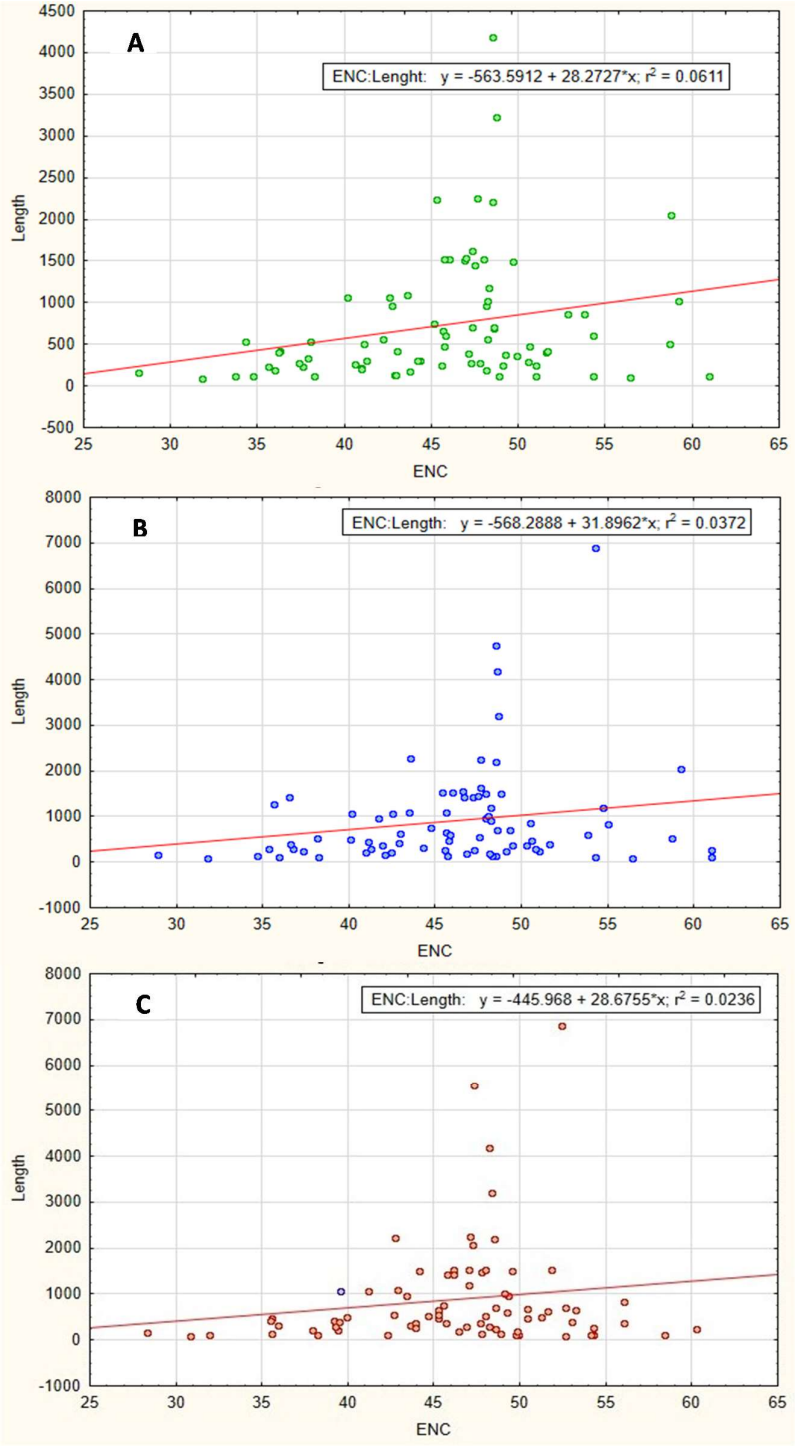

Figure 9: *Plot of ENC value variation against different protein length (A) G. montana (B) B. aegyptiaca (C) T. terrestris.*

4. Conclusion

This is the first report to study codon usage in Zygophyllaceae and Celastraceae plastomes. The correlation analysis suggested that codon usage patterns in both cp genomes appear due to the different forces, natural selection, mutation pressure, GC content of gene and protein length. Natural selection might have played a major role, while mutation pressure had a minor role in shaping the CUB of cp gene in both plants. The study of CUB and factors affecting it will provide a better understanding of the evolutionary history of both plants and, also improve the expression efficiency of exogenous genes by selecting the specific codon for transgenic research and gene editing of medicinally important plants.

Acknowledgement

First author is thankful to Vimal Research Society for Agro-Biotech & Cosmic Powers, Rajkot for Financial support, and Department of Biosciences, Saurashtra University, Rajkot, Gujarat, India for providing lab facilities.

References

1. Quax TE, Claassens NJ, Söll D, et al. Codon bias as a means to fine-tune gene expression. Mol Cell. 2015;59:149-61.

2. Mazumder TH, Chakraborty S. Gaining insights into the codon usage patterns of TP53 gene across eight mammalian species. PLoS One. 2015;10:e0121709.

3. Mitra S, Ray SK, Banerjee R. Synonymous codons influencing gene expression in organisms. Res Repo Biochem. 2016;6:57-65.

4. Sheng J, She X, Liu X, et al. Comparative analysis of codon usage patterns in chloroplast genomes of five Miscanthus species and related species. PeerJ. 2021;9:e12173.

5. Zhang Y, Shen Z, Meng X, et al. Codon usage patterns across seven Rosales species. BMC Plant Biol. 2022;22:65.

6. Sablok G, Chen TW, Lee CC, et al. ChloroMitoCU: Codon patterns across organelle genomes for functional genomics and evolutionary applications. DNA Res. 2017;24:327-32.

7. Dobrogojski J, Adamiec M, Luciński R. The chloroplast genome: a review. Acta Physiol Plant. 2020;42:98.

8. Iriarte A, Lamolle G, Musto H. Codon usage bias: an endless tale. J Mol Evol. 2021;89:589-93.

9. Narayana L, Satyanarayana P, Radhakrishnaiah H. Systematic position of Balanitaceae. In: Bilgrami K, Dogra J (eds), Phytochemistry and Plant Anatomy. CBS Publications, Delhi, India. 1990.

10. Figueiredo E, Smith GF. Plants of Angola/Plantas de Angola. South African National Biodiversity Institute, Pretoria, South Africa. 2008.

11. Dhuru B, Jayswal P, Sharma M, et al. In vitro antibacterial activity of *Gymnosporia montana* against some bacterial strains. Int J Univers Pharm Life Sci. 2012;2:12-20.

12. Monpara JK. Comparative biotechnological and bioinformatics studies on chloroplast DNA of some selected medicinal plants. Ph. D thesis, Saurashtra University, Gujarat, India. 2022.

13. Bhatt PP, Thaker VS. The complete chloroplast genome of *Cenchrus ciliaris* (Poaceae). Mitochondrial DNA B Resour. 2018;3:674-5.

14. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. Bioinformatics. 2004;20:3252-5.

15. Liu C, Shi L, Zhu Y, et al. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. BMC Genom. 2012;13:1-7.

16. Abouelhoda MI, Kurtz S, Ohlebusch E. Replacing suffix trees with enhanced suffix arrays. J Discrete Algorithms. 2004;2:53-86.

17. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33:1870-4.

18. Sharp PM, Li WH. An evolutionary perspective on synonymous codon usage in unicellular organisms. J Mol Evol. 1986;24:28-38.

19. Gupta SK, Bhattacharyya TK, Ghosh TC. Synonymous codon usage in *Lactococcus lactis*: mutational bias versus translational selection. J Biomol Struct Dyn. 2004;421:527-35.

20. Wright F. The 'effective number of codons' used in a gene. Gene. 1990;87:23-9.

21. Sueoka N. Directional mutation pressure and neutral molecular evolution. Proc Natl Acad Sci USA. 1988;85:2653-7.

22. Wen Y, Zou Z, Li H, et al. Analysis of codon usage patterns in *Morus notabilis* based on genome and transcriptome data. Genome. 2017;60:473-84.

23. Wang W, Yu H, Wang J, et al. The complete chloroplast genome sequences of the medicinal plant *Forsythia suspensa* (Oleaceae). Int J Mol Sci. 2017;18:2288.

24. Wang Y, Zhan DF, Jia X, et al. Complete chloroplast genome sequence of *Aquilaria sinensis* (Lour.) Gilg and evolution analysis within the Malvales order. Front Plant Sci. 2016;7:280.

25. Wang Z, Xu B, Li B, et al. Comparative analysis of codon usage patterns in chloroplast genomes of six Euphorbiaceae species. PeerJ. 2020;8:e8251.

26. Gichira AW, Avoga S, Li Z, et al. Comparative genomics of 11 complete chloroplast genomes of Senecioneae (Asteraceae) species: DNA barcodes and phylogenetics. Bot Stud. 2019;60:17.

27. Lee SR, Kim K, Lee BY, et al. Complete chloroplast genomes of all six Hosta species occurring in Korea: molecular structures, comparative, and phylogenetic analyses. BMC Genom. 2019;20:833.

28. Henriquez CL, Ahmed I, Carlsen MM, et al. Molecular evolution of chloroplast genomes in Monsteroideae (Araceae). Planta. 2020;251:72.

29. Bahiri-Elitzur S, Tuller T. Codon-based indices for modeling gene expression and transcript evolution. Comput Struct Biotechnol J. 2021;19:2646-63.

30. Morton BR. The role of context-dependent mutations in generating compositional and codon usage bias in grass chloroplast DNA. J Mol Evol. 2003;56:616-29.

31. Paul P, Malakar AK, Chakraborty S. Compositional bias coupled with selection and mutation pressure drives codon usage in *Brassica campestris* genes. Food Sci Biotechnol. 2018;27:725-33.

32. Zhang Y, Nie X, Jia X, et al. Analysis of codon usage patterns of the chloroplast genomes in the Poaceae family. Aust J Bot. 2012;60:461-70.

33. Nie X, Deng P, Feng K, et al. Comparative analysis of codon usage patterns in chloroplast genomes of the Asteraceae family. Plant Mol Biol Rep. 2014;32:828-40.

34. Yengkhom S, Uddin A, Chakraborty S. Deciphering codon usage patterns and evolutionary forces in chloroplast genes of *Camellia sinensis* var. assamica and *Camellia sinensis* var. sinensis in comparison to *Camellia pubicosta*. J Integr Agric. 2019;18:2771-85.

35. Hebert PD, Cywinska A, Ball SL, et al. Biological identifications through DNA barcodes. Proc Biol Sci. 2003;270:313-21.

36. Liu Q, Xue Q. Comparative studies on codon usage pattern of chloroplasts and their host nuclear genes in four plant species. J Genet. 2005;84:55-62.

37. Li X, Song H, Kuang Y, et al. Genome-wide analysis of codon usage bias in *Epichloe festucae*. Int J Mol Sci. 2016;17:1138.

38. Mazumdar P, Binti Othman R, Mebus K, et al. Codon usage and codon pair patterns in non-grass monocot genomes. Ann Bot. 2017;120:893-909.

39. Anwar AM, Aljabri M, El-Soda M. Patterns of genome-wide codon usage bias in tobacco, tomato and potato. Biotechnol Biotechnol Equip. 2021;35:657-64.

40. Duret L, Mouchiroud D. Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and Arabidopsis. Proc Natl Acad Sci. 1999;96:4482-7.

41. Niu Y, Luo Y, Wang C, et al. Deciphering codon usage patterns in genome of *Cucumis sativus* in comparison with nine species of Cucurbitaceae. J Agron. 2021;11:2289.

42. Andargie M, Congyi Z. Genome-wide analysis of codon usage in sesame (*Sesamum indicum* L.). Heliyon. 2022;8:e08687.