

Article

Analysis of Codon Usage Patterns of Six Sequenced *Brachypodium distachyon* Lines Reveals a Declining CG Skew of the CDSs from the 5'-ends to the 3'-ends

Jianyong Wang ¹, Yujing Lin ² and Mengli Xi ^{1,*}

¹ Key Laboratory of Forest Genetics and Biotechnology of Ministry of Education, Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, China; wangjy@njfu.edu.cn

² Shanghai Center for Plant Stress Biology and Center for Excellence in Molecular Plant Sciences, University of Chinese Academy of Sciences, Shanghai 200032, China; yjlin@psc.ac.cn

* Correspondence: ximenglinjfu@126.com

Abstract: *Brachypodium distachyon*, a new monocotyledonous model plant, has received wide attention in biological research due to its small genome and numerous genetic resources. Codon usage bias is an important feature of genes and genomes, and it can be used in transgenic and evolutionary studies. In this study, the nucleotide compositions and patterns of codon usage bias were calculated using Codon W. Additionally, an ENC plot, Parity rule 2 and correspondence analyses were used to explore the major factors influencing codon usage bias patterns. The numbers of hydrogen bonds and skews were used to analyze the GC trend in the 5'-ends of the coding sequences. The results showed that minor differences in the codon usage bias patterns were revealed by the ENC plot, Parity rule 2 and correspondence analyses. The analyses of the CG-skew and the number of hydrogen bonds showed a declining trend in the number of cytosines at the 5'-ends of the CDSs (from the 5'-ends to the 3'-ends), indicating that GC may play a major role in codon usage bias. In addition, our results laid a foundation for the study of codon usage bias patterns in *Brachypodium* genus and suggested that the GC plays a major role in determining these patterns.

Keywords: *Brachypodium distachyon*; codon usage bias; CG-skew



Citation: Wang, J.; Lin, Y.; Xi, M. Analysis of Codon Usage Patterns of Six Sequenced *Brachypodium distachyon* Lines Reveals a Declining CG Skew of the CDSs from the 5'-ends to the 3'-ends. *Genes* **2021**, *12*, 1467. <https://doi.org/10.3390/genes12101467>

Academic Editor:
Heinz Himmelbauer

Received: 5 July 2021
Accepted: 20 September 2021
Published: 23 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

B. distachyon, a new monocotyledonous model plant, has received wide attention in genetics and evolution research due to its small genome, short life cycle, easy transformation, self-proliferation and numerous genetic resources [1]. In recent decades, scientists have collected many inbred lines from Asia and Europe [2]. Since *B. distachyon* is self-fertile and not domesticated, researchers support the unaltered view that intra-species natural diversity among inbred lines depends on mutation and selection pressure, which can be reflected in the pattern of codon usage [3].

Owing to the redundancy of genetic codons, synonymous mutations have no effect on protein function or organismal fitness; therefore, they were initially considered “silent mutations” [4–7]. The sequencing of the organisms revealed that synonymous codons have a non-random distribution among genes or genomes, which is called codon usage bias; it is caused by a range of factors, including GC content, mutation, selection and codon position [8–11]. Codon usage bias influences transcription [12,13], protein translation [14,15] and gene expression levels [16,17]. Thus, the codon usage bias pattern provides important information for reconstruction vectors and for the improvement of expression levels, as well as for the revelation of plant–pathogen coevolutionary events and the improvement of host resistance [18–20]. Previous research has found that codon usage, frequently of high- and low-GC-content genes, is different in *B. distachyon* Bd21, and 27 codons were identified

as optimal for evolution studies [8]. However, it is unclear whether the codon usage pattern is same in the other *B. distachyon* inbred lines. For example, Bd1-1, Bd18-1, Bd2-3, Bd29-1, Bd3-1 and Bd30-1 are all members of the “core inbred line set”, an important set of *B. distachyon* genetic stocks, which are often used to develop single seed descent-derived inbred lines. Hence, the objective of this study was to gather meaningful information on codon usage bias that could be used for the analysis of the evolution of *Brachypodium* in the post-genomic era.

In this study, we compared the codon usage bias patterns among six *B. distachyon* inbred lines, using overall GC content, the effective number of codons (ENC) and relative synonymous codon usage (RSCU). A neutrality plot analysis was performed and the optimal codons were determined. In addition, we investigated the relationships between the number of hydrogen bonds, the AT-skew, the CG-skew and the codon positions. The results provided insights into genetic modification and genome evolution that can be utilized to improve the genetic transformation efficiency and to understand the genetic variation of *Brachypodium*.

2. Materials and Methods

2.1. Coding Sequence (CDS) Data

The complete genomes of six *B. distachyon* lines, including Bd1-1, Bd18-1, Bd2-3, Bd29-1, Bd3-1 and Bd30-1 [3], were download from BrachyPan (<https://phytozome-next.jgi.doe.gov/brachypan>, accessed on 4 March 2021). In order to avoid sampling errors, all the CDSs were extracted in accordance with the following considerations: (1) the CDSs’ length should be the fold of three; (2) the CDSs should not be less than 300 bp; (3) the CDSs should contain the initiation codon (ATG) and termination codons (TAA, TAG or TGA); (4) the CDSs should not have one or more intermediate stop codons; and (5) the CDSs should be encoded by the longest primary transcripts [11]. All the filtered CDSs were used to analyze codon usage bias with Codon W 1.4.2 (<https://sourceforge.net/projects/codonw/>, accessed on 13 March 2021).

2.2. ENC

The ENC may be estimated by the degree of deviation among the synonymous codon usage patterns of the six genomes. Its value ranges between 20 (strong bias of codons) and 61 (all codons used randomly) [21,22]. The ENC value depends on GC3s, another index of codon usage that refers to the G-to-C ratio at the third position of one codon. The expected values were calculated using following formula:

$$ENC = 2 + S + \frac{29}{S^2 + (1 - S)^2}$$

where S represents the GC3s value [22].

2.3. Parity Rule (PR2)

The PR2 plot is used to assess the influence of mutation pressure, natural selection or both on codon bias [23]. There is no deviation in the DNA under mutation and selection conditions when the content of $A = T$ and $G = C$. The $A3s/(A3s + T3s)$ and $G3s/(G3s + C3s)$ were calculated and plotted to illustrate the relationship between mutation and selection pressures on each gene [24].

2.4. RSCU

The RSCU value, as proposed by Sharp and Li in 1986, is the ratio of the observed number of codons to the expected number of codons, and it was calculated using the following formula [25,26]:

$$RSCU = \frac{g_{ij}}{\sum_j n_i g_{ij}} n_i$$

where g_{ij} represents the frequency of the i th codon for the j th amino acid and n_i represents n_i kinds of synonymous codons encoding the same amino acid. *RSCU* values of more than, less than and equal to 1.0 indicated positive, negative and no bias, respectively, in codon usage. Thus, codons with *RSCU* values > 1.0 were used more frequently than other codons.

2.5. Correspondence Analysis (COA)

The COAs of codon usage in the nuclear genomes of six *B. distachyon* lines were used to study the relationships between codon usage and factors based on *RSCU* in Codon W 1.4.2. In the COA analysis, the *RSCU* values of 59 codons, excluding ATG, TGG and three termination codons (TAA, TAG and TGA), were placed in a multi-dimensional space along 59 orthogonal axes to explain the data variation, and the correlation analysis between axis 1 and codon usage indices, including GC content, GC3s and ENC, were carried out in R (Version 4.0.2) [27].

2.6. Codon Position and Hydrogen Bonds

Villada's method, which has been used in bacteria and archaea, was applied to analyze the relationship between codon position and hydrogen bonds in *B. distachyon* [28]. First, nucleotides in the CDSs of each genome were arranged in a matrix, with CDSs as rows and codons as columns. Next, the sequences were treated using Biostrings (<https://bioconductor.org/packages/release/bioc/html/Biostrings.html>, accessed on 3 April 2021) and SeqinR [29] to ensure that all the CDSs were left aligned from the 5'-end. Subsequently, the nucleotide composition was converted into the number of hydrogen bonds on the basis of adenine (thymine) having two hydrogen bonds and guanine (cytosine) having three hydrogen bonds. Finally, the mean number of hydrogen bonds at each codon position in the genomes of the six *B. distachyon* lines was calculated using Hmisc (<https://rdocumentation.org/packages/Hmisc/versions/4.5-0>, accessed on 3 April 2021), with a 95% confidence interval based on 1000 nonparametric bootstrap replicates.

2.7. CG-Skew and AT-Skew Values

The CG-skew and AT-skew values of the sequences from the six genomes were calculated in a sliding window of 100 bp with 1-bp steps, using the following formulae [30]:

$$\text{CG-skew} = \frac{C - G}{C + G} \quad \text{and} \quad \text{AT-skew} = \frac{A - T}{A + T},$$

where A, T, G and C represent the total numbers of these bases in the 100-bp windows.

2.8. Optimal Codons

To determine the optimal codons, 5% of the total genes from each of the six *B. distachyon* lines located at the extreme right and left of axis 1 were regarded as high- and low-expression datasets, using the two-way chi-square contingency test, respectively. The codons were defined as optimal if they showed a significantly greater frequency of usage ($p < 0.01$) in the high-expression datasets than in the low-expression datasets [8].

3. Results

3.1. Nucleotide Base Composition in Genomes of Six *B. distachyon* Lines

To avoid sampling errors, six *B. distachyon* genome coding sequences (CDSs), Bd1-1, Bd18-1, Bd2-3, Bd29-1, Bd3-1 and Bd30-1, were filtered, resulting in 37261, 41810, 38148, 35932, 30621 and 36985 datasets, respectively. The nucleotide composition and codon usage bias indices of the filtered CDSs were estimated by Codon W 1.4.2, which included the frequencies of adenine, guanine, cytosine and thymine bases at the third positions of the codons (A3s, G3s, C3s and T3s, respectively), the GC content at the third positions (GC3s), the number of synonymous codons (L_sym), the total number of amino acids (L_aa), the general average hydrophobicity of the proteins (GRAVY) and the aromaticity of the proteins (Aromo), as shown using means and standard deviations of all the genes in

Table 1. For the nucleotide composition, there were tiny differences at the third positions of the codons among the six inbred lines, conforming to $C > G > T > A$, which suggested that a similar codon composition was present in *B. distachyon*. The GC3s' values were greater than 0.55, indicating that the six lines tended to use codons with G and C at the third codon positions. In Bd3-1, the G3s, C3s and GC3s were less than in the other five lines (Figure 1a).

Table 1. The information pertaining to the six *B. distachyon* lines.

	Bd1-1	Bd18-1	Bd2-3	Bd29-1	Bd30-1	Bd3-1
Origin	Turkey	Turkey	Iraq	Ukraine	Spain	Iraq
Habit	winter	winter	spring	winter	winter	spring
Genome size	183 MB	269 MB	269 MB	253 MB	195 MB	170 MB
All GC *	45%	46%	46%	46%	45%	44%
Before filtered	42227	46621	45459	44278	42026	37234
After filtered	37261	41810	38148	35932	36985	30621
T3s	0.27 ± 0.13	0.27 ± 0.13	0.27 ± 0.13	0.28 ± 0.13	0.27 ± 0.13	0.30 ± 0.12
C3s	0.38 ± 0.14	0.38 ± 0.14	0.38 ± 0.14	0.37 ± 0.13	0.38 ± 0.14	0.35 ± 0.13
A3s	0.22 ± 0.11	0.22 ± 0.11	0.22 ± 0.11	0.23 ± 0.10	0.22 ± 0.11	0.24 ± 0.10
G3s	0.36 ± 0.10	0.36 ± 0.10	0.36 ± 0.09	0.35 ± 0.09	0.36 ± 0.10	0.35 ± 0.09
GC3s	0.60 ± 0.19	0.60 ± 0.18	0.59 ± 0.18	0.58 ± 0.18	0.60 ± 0.19	0.56 ± 0.17
L_sym	433.07 ± 303.94	439.45 ± 312.49	425.25 ± 302.44	421.47 ± 308.73	434.01 ± 305.26	424.25 ± 317.25
L_aa	449.64 ± 314.48	456.34 ± 323.33	441.66 ± 312.92	437.91 ± 319.52	450.68 ± 315.96	440.81 ± 328.13
Gravy	-0.27 ± 0.37	-0.27 ± 0.37	-0.27 ± 0.37	-0.27 ± 0.37	-0.27 ± 0.37	-0.28 ± 0.37
Aromo	0.08 ± 0.03	0.08 ± 0.03	0.08 ± 0.03	0.08 ± 0.03	0.08 ± 0.03	0.08 ± 0.03

* All GC represents the GC content of the genome.

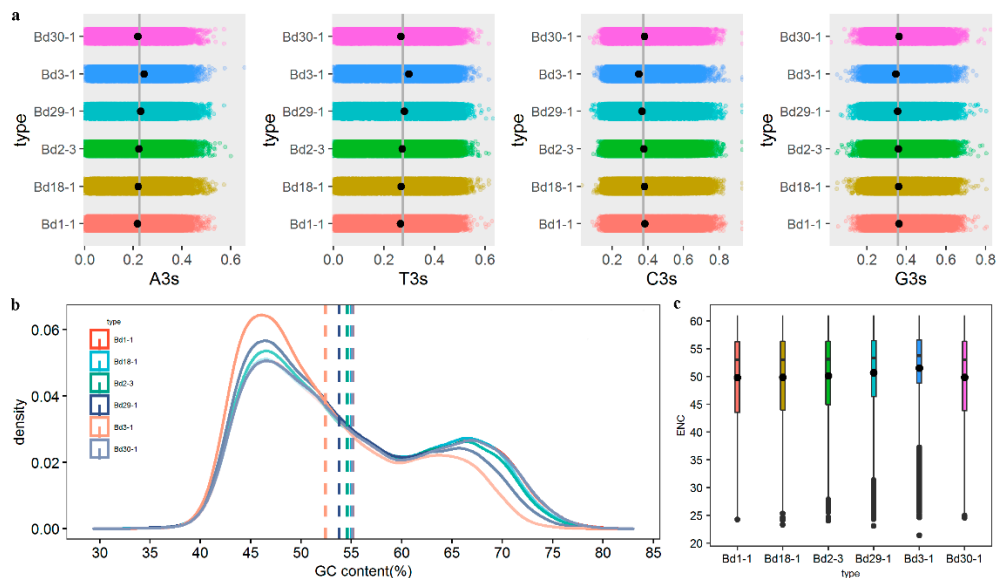


Figure 1. The nucleotide compositions and codon usages of six *B. distachyon* lines. (a) The distribution of four bases at the third positions of the codons; (b) The distribution of genes with different GC contents; (c) The box plot of ENC values from the six lines.

GC content, as an indicator of the strength of the mutations in a biological evolutionary process, can reflect the composition of the DNA sequence. To investigate the nucleotide compositions of the six genomes, we calculated the GC content of each gene from datasets using Codon W 1.4.2. Next, the density curves of genes from each line were plotted, as shown in Figure 1b. The bimodal distribution of the GC content had mean values of 52.44–55.18%. The maximum GC content value appeared in the Bd1-1, at 55.18%, with a variance of 9.28, and the minimum value was 52.44%, in Bd3-1. Interestingly, the bimodal distribution of the six inbred lines illustrated that there were far more genes with low GC content than with high GC content.

The effective number of codons (ENC) was used to assess the heterogeneity of uniform codon usage, which may have reflected the codons' degree of preference during the protein translation process. The ENC varied from 20 to 61. There was strong codon bias in genes with high expression levels when the ENC was less than 35, whereas an ENC greater than 50 indicated that the use of codons was random.

The ENC of the six lines were computed and a box plot was constructed to compare the differences in ENC, as shown in Figure 1c. The mean ENC ranged from 49.77 in Bd1-1 to 51.53 in Bd3-1. The ENC values from the all genomes were near 50, indicating that the codon usage bias was weak in *B. distachyon*.

3.2. ENC and GC3 Analyses

The GC3 was defined as the proportion of the GC content at the third positions of the codons to the total number of bases. ENC reflects the degree of preference for synonymous codon usage. An ENC-GC3s plot analysis is often used to analyze the relationship between nucleotide composition and codon bias. When the data point falls near the standard curve, the codon bias pattern is mainly determined by the mutation pressure rather than the external selection pressure. The similar ENC-GC3 plot distribution of the CDSs from the six lines is shown in Figure 2. Most points are in an ENC range of 25 to 61 and a GC3 range of 0.25 to 1.0. Although only a few CDSs were found close to the expected curve, a majority of the points deviated below the curve, with low ENC values (Figure 2), indicating that the codon usage biases of the six genomes were affected mainly by natural selection, mutation pressure and other factors (such as gene length and expression level). Furthermore, the codon biases were shown to have been shaped by the combined effects of mutation and selection pressures in previous studies of Poaceae [31,32], Orchidaceae [33] and Malvaceae [24].

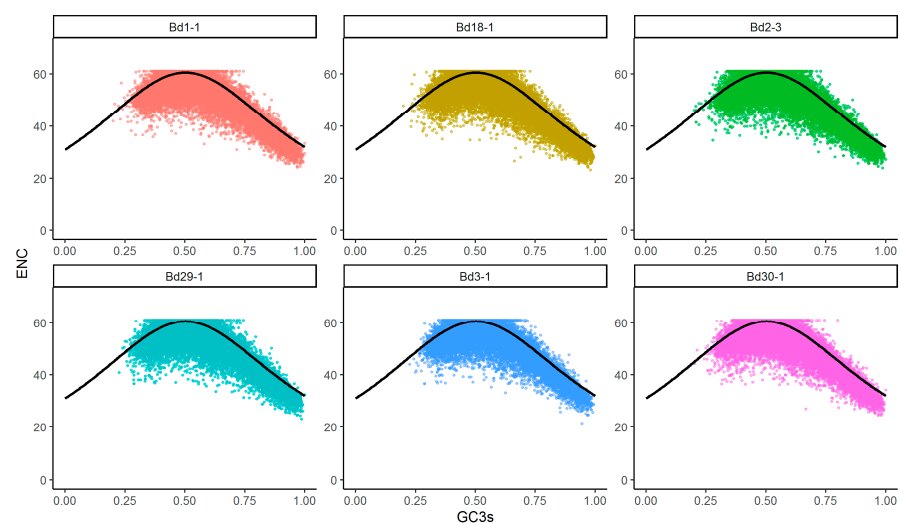


Figure 2. ENC-GC3 plots of the six *B. distachyon* genomes. The black solid lines indicate the expected ENC values.

3.3. PR2 Plot Analysis

Parity analysis is an efficient method for the assessment of the mutation pressure on the bases in the third positions of codons during the evolutionary process. In a PR2 plot, four regions are divided by two lines, which indicate $A = T$ and $G = C$. The points represent the $G3/(G3 + C3)$ and $A3/(A3 + T3)$ values of CDSs distributed around the center position of a two-dimensional scatter plot. The PR2 plots of the six *B. distachyon* lines are shown in Figure 3. The values of the AT-biases were 0.46, 0.46, 0.45, 0.45, 0.45 and 0.45 in Bd1-1, Bd18-1, Bd2-3, Bd29-1, Bd3-1 and Bd30-1, respectively. However, the values of the GC-biases were 0.49, 0.49, 0.50, 0.50, 0.51 and 0.49, respectively. Consistently, the data of the different lines revealed similar patterns on the T/C-bias at the third position of the

codons. As a whole, the unbalanced frequency between A/T and G/C suggested that the codon usage bias was not only affected by mutation but also by selection and other factors in *B. distachyon* inbred lines. Similar results in codon usage bias studies have been reported in Asteraceae [34] and Poaceae [35].

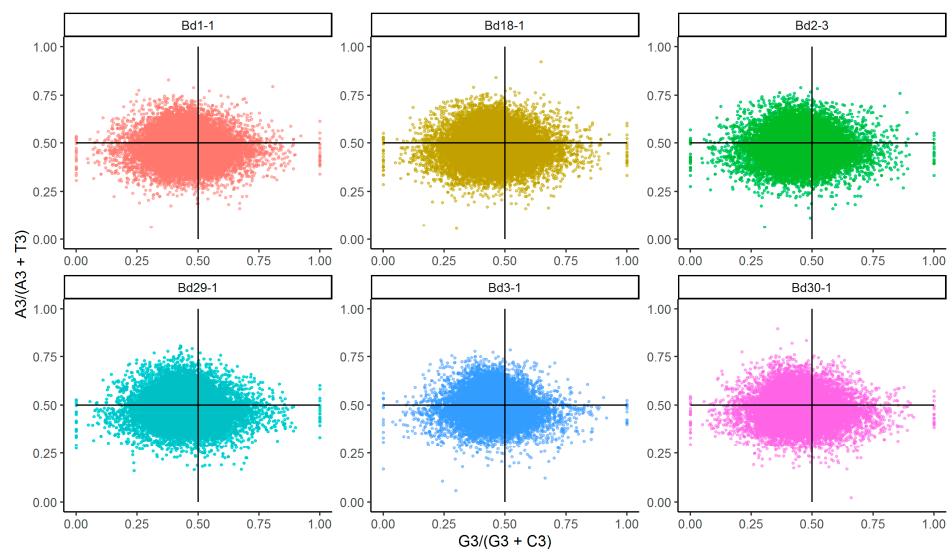


Figure 3. The Parity rule 2 (PR2) plots of the six *B. distachyon* lines.

3.4. Correspondence Analysis (COA)

To investigate the main variation of the codon usage in the nuclear genomes of the six lines, correspondence analyses of all the genes in each genome were performed based on the RSCU values of 59 encoding codons (excluding Met, Trp and the three stop codons), as calculated by Codon W (Table 2). Axis 1, the primary explanatory axis, accounted for 45.14%, 44.49%, 43.42%, 42.08%, 44.91% and 39.94% of the overall variations of codon usage for Bd1-1, Bd18-1, Bd2-3, Bd29-1, Bd30-1 and Bd3-1, respectively.

Table 2. The main variances of the four axes of the six lines.

Lines	Axis 1 (%)	Axis 2 (%)	Axis 3 (%)	Axis 4 (%)
Bd1-1	45.14	4.09	2.78	2.28
Bd18-1	44.49	4.18	2.84	2.29
Bd2-3	43.42	4.07	2.86	2.31
Bd29-1	42.08	3.99	2.95	2.40
Bd30-1	44.91	4.08	2.83	2.26
Bd3-1	39.94	3.93	3.03	2.46

The distribution of the GC content in monocotyledons genomes is divided into two regions, based on a GC content of 60% [8]. Thus, there were different kinds of genes associated with different GC contents; high-GC ($\geq 60\%$), middle-GC (40–60%) and low-GC ($\leq 40\%$) were plotted on the axes. Interestingly, the genes with low- and high-GC content were located to the left and right, respectively, in Bd1-1 and Bd2-3, but the opposite distribution pattern was found in Bd18-1, Bd29-1, Bd3-1 and Bd30-1 (Figure 4). A further analysis showed similar results between GC content and the gene position on the first axis, with an R^2 value of 0.89–0.90 (Supplementary Materials Figure S1). We also investigated the distribution of synonymous codons on Axis 1 and 2. A similar phenomenon occurred in the four-type codons, with the G/C-ending codons being distinguished clearly from A/T-ending codons on Axis 1 and pyrimidines being distinguished from purines on Axis 2 (Supplementary Materials Figure S2). These results suggested that nucleotide composition was one of the major factors in the synonymous codon usage patterns of the six *B. distachyon* inbred lines.

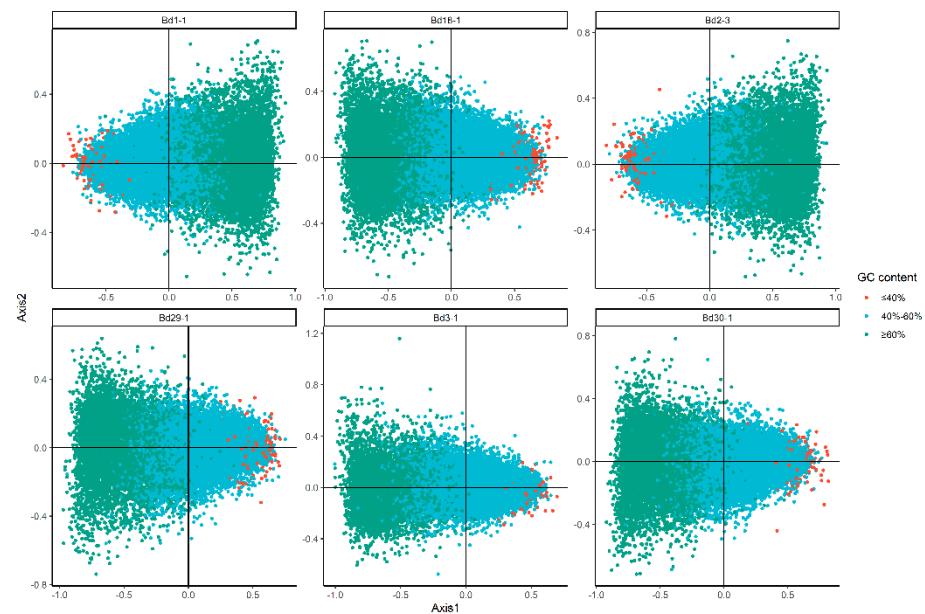


Figure 4. The correspondence analysis of the six *B. distachyon* inbred lines. The red, blue and green dots represent genes with GC content $\leq 40\%$, between 40–60% and $\geq 60\%$, respectively.

3.5. RSCU Value Analysis

The RSCU value represents a quantitative relationship between the number of codons found and expected, reflecting the usage bias of different codons encoding the same amino acid. The RSCU values of each gene in the six lines were first calculated using Codon W. Next, 50 genes in each genome were selected randomly to form a dataset for the analysis of the RSCUs of the different lines (Figure 5). There were 27 codons with RSCU > 1 ending with C/G in the six *B. distachyon* inbred lines. As shown in Figure 5, the RSCU values of almost all the codons ending with G/C were greater than those ending with A/T. We also investigated the optimal codons in the six *B. distachyon* inbred lines. The 5% genes located at the two ends of the first axis in the COA analysis were selected from each genome to represent codon bias gene datasets (high-bias and low-bias datasets). In total, 27 optimal codons, including UUC, CUC, CUG, AUC, GUC, GUG, UAC, CAC, CAG, AAC, AAG, GAC, GAG, UCC, UCG, CCC, CCG, ACC, ACG, GCC, GCG, UGC, CGC, CGG, AGC, GGC and GGG, were found in all the lines. They mainly occurred in the high-bias gene datasets, according to chi-square tests. Thus, there was a high degree of unity among the codon usage preferences of the six lines, based on their RSCU values.

3.6. Hydrogen Bonds at Different Codon Positions

The positioning of synonymous codons in CDSs can dictate protein expression via many mechanisms, such as protein cotranslational folding and local translation efficiency [36]. Single-stranded DNA is useful for RNA polymerization during denaturation because of the strong effects on protein translation at the 5'-end regions of the CDSs [28]. Thus, to understand the influence of codon position on protein translation, we investigated the relationship between the number of hydrogen bonds and codon positions at the 5'-end regions of the CDSs. The average number of hydrogen bonds gradually declined along with the codon positions of the CDSs from the 5'-ends to the 3'-ends in all the genomes, reaching lower levels (7.35–7.45) after the 300th codon of the CDSs. For example, the average number of hydrogen bonds at the 334th codon in Bd3-1 was only 7.35 (Figure 6a).

Next, the codons were divided into two types, based on the number of hydrogen bonds: cheap codons (six or seven hydrogen bonds) and expensive codons (eight or nine hydrogen bonds). The expensive codons were mostly used at the 5'-ends of the CDSs in all the genomes, and the usage frequency of the expensive codons declined as the codons' positions increased (Figure 6b). The key positions with equal usage frequencies of the

As shown in Figure 7a, a significant downward trend in the CG-skew was observed in the six *B. distachyon* lines with different slopes. The greatest slope of the CG-skew appeared in Bd29-1 and the smallest appeared in Bd2-3. We also found the G and C contents were almost equal at the 5'-end regions of CDSs in Bd3-1, whereas slightly more Cs were used than Gs in the other five lines.

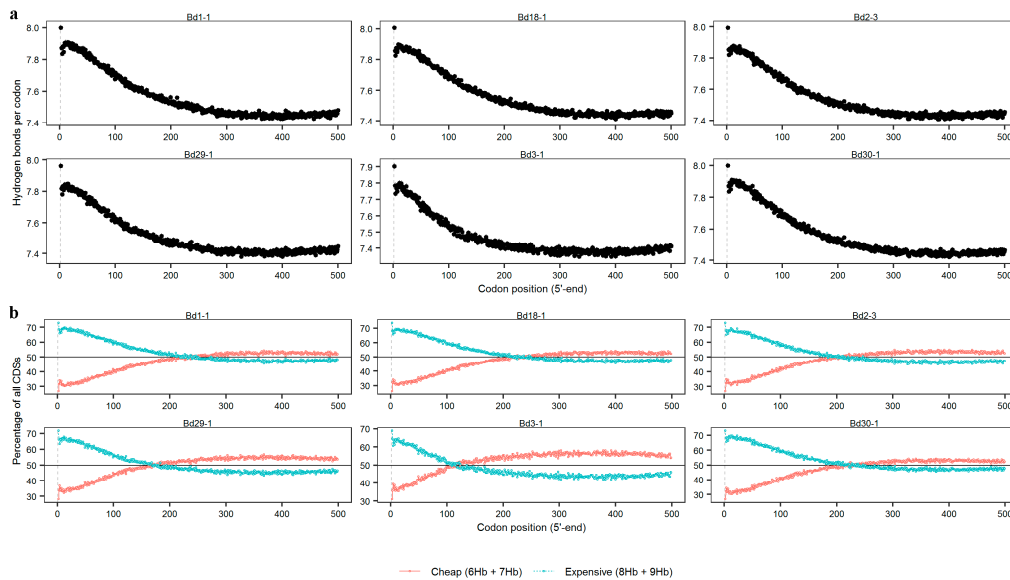


Figure 6. The relationship between the number of hydrogen bonds and codon position. (a) The hydrogen bonds per codon at the 5'-ends of the CDSs in the six lines; (b) The percentages of the two type codons at the 5'-ends of the CDSs in the six lines.

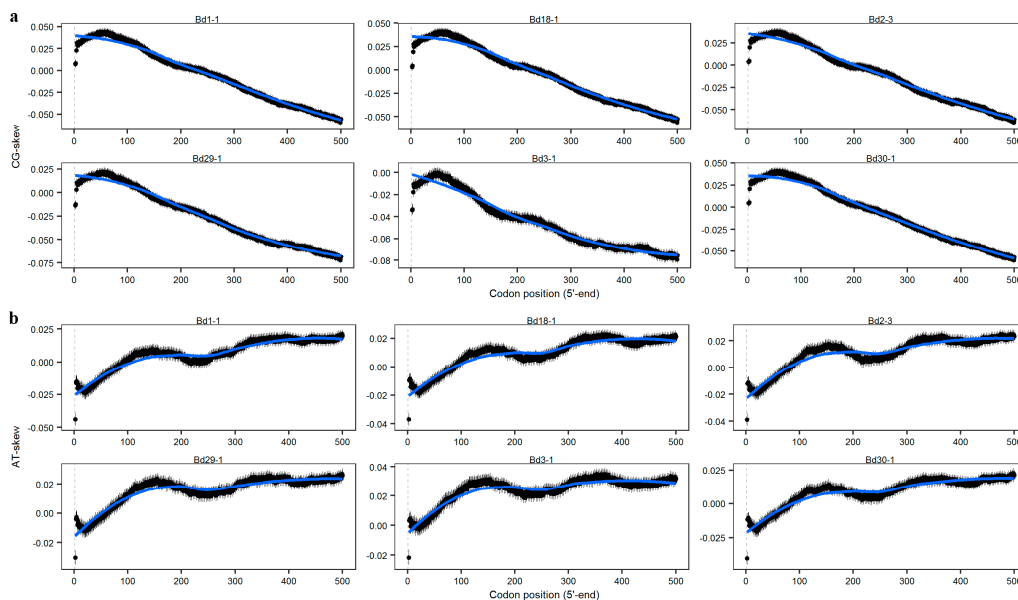


Figure 7. The relationships between both CG- and AT-skews and codon position. (a) CG-skews of nucleotides at the 5'-ends of the CDSs in the six lines; (b) AT-skews of nucleotides at the 5'-ends of the CDSs in the six *B. distachyon* lines.

Although the AT-skew presented an overall upward tendency (up-down-up), unlike the CG-skew trend, the slopes differed among the lines (Figure 7b). The greatest increase occurred in Bd1-1, and the smallest increase appeared in Bd3-1, which stabilized at approximately 0.03 after 300 bp.

These results suggest that most cases followed a similar trend, with a preference for C at the 5'-ends and for G at the 3'-ends, suggesting that this is a conserved phenomenon

among *B. distachyon* lines. However, the differences in the AT-skew trends appeared to be species-dependent, implying that they were affected by other factors.

4. Discussion

Owing to the redundancy of genetic codons, synonymous mutations in codons have no effect on protein function and organismal fitness, and therefore, were initially considered “silent mutations” [4–7]. Synonymous codons are non-randomly distributed among the genes or genomes of sequenced organisms, which is termed codon usage bias. The bias may be caused by various factors, including GC content, mutation, selection and codon position [8–11]. In this study, the CDSs from six *B. distachyon* inbred lines were used to analyze codon usage bias and the factors influencing the patterns. The bimodal distribution of the GC content and the negative correlation between the CG-skew and the number of hydrogen bonds in the codons positioned at the 5'-ends of the CDSs were detected in the six lines, indicating that GC content plays a major role in codon usage bias [37,38].

4.1. Multiple Factors Play a Role in GC Content Distribution in *B. distachyon* Inbred Lines

The GC content, as an indicator of the strength of the mutation pressure during the evolutionary process, may reflect the composition of a DNA sequence. The GC content of monocotyledonous plants has gradually increased and has been greater than that of dicotyledonous plants since their separation, approximately 200 million years ago [39]. In our study, the average GC content of the CDSs ranged from 52.53 to 57.68% in the six *B. distachyon* inbred lines. Interestingly, previous research showed a bimodal distribution of GC content in Bd21, which is similar to our results [8].

Currently, three hypotheses may be used to explain codon usage bias patterns: selection, mutation and GC-biased gene conversion [40]. The selection hypothesis states that thermal stability leads to a high GC content. However, base composition is also shaped by other factors. For instance, the specific functions of genes may determine the GC contents of gene regions, and then change the codon usage bias. The mutational theory states that GC mutations may be driven by high GC content, which, in turn, promotes the evolution of species. The GC-biased gene conversion hypothesis states that single-strand DNA could form heteroduplexes with homologous sequences during recombination. Mismatches in the heteroduplexes could be repaired by changing the nucleotide bias toward GC alleles. Muyle found that G or C is used preferentially and that GC-biased gene conversion is positively correlated with codon position in the rice genome [40]. These results indicated that the bimodal distribution of GC in Poaceae may be greatly affected by selection and GC-biased gene conversion [40]. However, we found that the bimodal pattern in Bd3-1 was different from the other five inbred lines, implying that other factors affected the pattern. A similar pattern would appear in all the lines if the recombination and GC-biased dynamics were mostly conserved; partial conservation may cause the differences in the intensity of the GC-biased gene conversion, as has been documented in polyploidy plants, such as Triticeae [41], *Gossypium* [42] and *Nicotiana* [43]. However, the *B. distachyon* inbred line Bd3-1 is a diploid plant; the pattern of GC-biased gene conversion may differ with polyploidy plants. Just as Bd3-1 is different from other lines, *B. distachyon* contains many lines that may produce more interesting results. Our results could not represent the whole of *B. distachyon* or even the *Brachypodium*, since only six lines were investigated, a very small portion. Thus, the differences in GC-biased gene conversion need to be explored during the evolution of *Brachypodium* in the future.

4.2. Hydrogen Bonding at the 5'-Ends of CDSs Is Different between Plants and Bacteria

The positioning of synonymous codons in CDSs can dictate protein expression by many mechanisms, such as protein cotranslational folding and local translation efficiency [36]. Villada found that the number of hydrogen bonds in the 5'-ends of the CDSs gradually increased in the bacteria, achieving stability after the 15th codon, with the frequency of expensive codons being greater than 50% [28]. However, a contrary result

was found in the *B. distachyon*. The number of hydrogen bonds gradually declined as the codon position increased, leading to equal frequencies between the expensive and cheap codons after the 200th to 300th codons. This trend was verified by the CG-skew. Villada implied that the coupling of transcription and translation requires more energy in the prokaryotes to avoid the premature termination of transcription when increasing the number of hydrogen bonds [28]. The *B. distachyon* species, with its high GC content, may feature different patterns, owing to the separation of translation and transcription in space and time. Interestingly, the frequency of the cheap codons was greater than that of the expensive codons after approximately the 110th codon in Bd3-1. Thus, the patterns of the two types of codons varied among all the lines, implying that other factors affected the distribution pattern. Choosing different synonymous codons could affect hydrogen bonding. Plants may use this advantage in order to smoothly reduce the energy requirement for unwinding the double-strand DNA molecule in CDSs. Just as Bd3-1 is different from other lines, there may be differences in energy requirements among lines. Hence, the problem needs to be further explored.

5. Conclusions

In this study, we systematically compared the differences in codon usage bias patterns among six *B. distachyon* lines. A declining trend in the number of cytosines in the CDSs from the 5'-ends to the 3'-ends was detected by CG-skew and the number of hydrogen bonds, indicating that GC may play a major role in codon usage bias. Further research is needed to understand the mechanism behind the GC content's effect on codon usage, using multi-omics data from more inbred lines.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/genes12101467/s1>, Figure S1: The correlations between the GC contents of genes in the six species and their positions on the first axis. Figure S2: The distributions of codons from the six lines on the two axes.

Author Contributions: J.W. and M.X. designed the research. J.W. and Y.L. conducted the experiments. J.W. and Y.L. participated in the data analysis and manuscript preparation. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by grant 31670603 from the National Natural Science Foundation of China, and funds from the Priority Academic Program Development of Jiangsu Higher Education Institutions.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: These sequence data were produced by the US Department of Energy Joint Genome Institute.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Vogel, J.P.; Garvin, D.F.; Mockler, T.C.; Schmutz, J.; Rokhsar, D.; Bevan, M.W.; Barry, K.; Lucas, S.; Harmon-Smoth, M.; Lail, K.; et al. Genome Sequencing and Analysis of the Model Grass *Brachypodium distachyon*. *Nature* **2010**, *463*, 763–768. [[CrossRef](#)]
2. Scholthof, K.B.G.; Irigoyen, S.; Catalan, P.; Mandadi, K.K. *Brachypodium*: A Monocot Grass Model Genus for Plant Biology. *Plant Cell* **2018**, *30*, 1673–1694. [[CrossRef](#)]
3. Gordon, S.P.; Contreras-Moreira, B.; Woods, D.P.; Des Marais, D.L.; Burgess, D.; Shu, S.; Stritt, C.; Roulin, A.C.; Schackwitz, W.; Tyler, L.; et al. Extensive Gene Content Variation in the *Brachypodium distachyon* Pan-genome Correlates with Population Structure. *Nat. Commun.* **2017**, *8*, 2184. [[CrossRef](#)]
4. Quax, T.E.F.; Claassens, N.J.; Söll, D.; van der Oost, J. Codon Bias as a Means to Fine-Tune Gene Expression. *Mol. Cell* **2015**, *59*, 149–161. [[CrossRef](#)]

5. Chu, D.; Wei, L. Parsing the Synonymous Mutations in the Maize Genome: Isoaccepting Mutations are More Advantageous in Regions with Codon Co-occurrence Bias. *BMC Plant Biol.* **2019**, *19*, 422. [[CrossRef](#)]
6. Wei, L. Selection on Synonymous Mutations Revealed by 1135 Genomes of *Arabidopsis thaliana*. *Evol. Bioinform.* **2020**, *16*, 1176934320916794. [[CrossRef](#)]
7. Plotkin, J.B.; Kudla, G. Synonymous but not the Same: The Causes and Consequences of Codon Bias. *Nat. Rev. Genet.* **2011**, *12*, 32–42. [[CrossRef](#)]
8. Liu, H.; Huang, Y.; Du, X.; Chen, Z.; Zeng, X.; Chen, Y.; Zhang, H. Patterns of Synonymous Codon Usage Bias in the Model Grass *Brachypodium distachyon*. *Genet. Mol. Res. GMR* **2012**, *11*, 4695–4706. [[CrossRef](#)]
9. Wang, H.; Hickey, D.A. Rapid Divergence of Codon Usage Patterns within the Rice Genome. *BMC Evol. Biol.* **2007**, *7*, S6. [[CrossRef](#)]
10. Hu, W.; Wei, S. Codon Usage Bias of the Wheat Flower Development Gene WAG-2 and Other AGAMOUS Group Genes. *J. Agric. Sci.* **2017**, *9*, 56–67. [[CrossRef](#)]
11. Mazumdar, P.; Binti Othman, R.; Mebus, K.; Ramakrishnan, N.; Ann Harikrishna, J. Codon Usage and Codon Pair Patterns in Non-Grass Monocot Genomes. *Ann. Bot.* **2017**, *120*, 893–909. [[CrossRef](#)]
12. Novoa, E.M.; de Pouplana, L.R. Speeding with Control: Codon usage, tRNAs, and Ribosomes. *Trends Genet. TIG* **2012**, *28*, 574–581. [[CrossRef](#)] [[PubMed](#)]
13. Presnyak, V.; Alhusaini, N.; Chen, Y.-H.; Martin, S.; Morris, N.; Kline, N.; Olson, S.; Weinberg, D.; Baker, K.E.; Graveley, B.R.; et al. Codon Optimality is a Major Determinant of mRNA Stability. *Cell* **2015**, *160*, 1111–1124. [[CrossRef](#)]
14. Pechmann, S.; Frydman, J. Evolutionary Conservation of Codon Optimality Reveals Hidden Signatures of Co-translational Folding. *Nat. Struct. Mol. Biol.* **2013**, *20*, 237–243. [[CrossRef](#)] [[PubMed](#)]
15. Tuller, T.; Waldman, Y.Y.; Kupiec, M.; Rupp, E. Translation Efficiency Is Determined by both Codon Bias and Folding Energy. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 3645–3650. [[CrossRef](#)]
16. Sahoo, S.; Das, S.S.; Rakshit, R. Codon Usage Pattern and Predicted Gene Expression in *Arabidopsis thaliana*. *Gene X* **2019**, *2*, 100012. [[CrossRef](#)]
17. Liu, H.; He, R.; Zhang, H.; Huang, Y.; Tian, M.; Zhang, J. Analysis of Synonymous Codon Usage in *Zea mays*. *Mol. Biol. Rep.* **2009**, *37*, 677–684. [[CrossRef](#)]
18. Thao, N.P.; Tran, L.-S.P. Enhancement of Plant Productivity in the Post-Genomics Era. *Curr. Genom.* **2016**, *17*, 295–296. [[CrossRef](#)]
19. Nguyen, T.H.; Wang, D.; Rahman, S.U.; Bai, H.; Yao, X.; Chen, D.; Tao, S. Analysis of Codon Usage Patterns and Influencing Factors in Rice Tungro Bacilliform Virus. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* **2021**, *90*, 104750. [[CrossRef](#)]
20. Chen, H.; Raffaele, S.; Dong, S. Silent Control: Microbial Plant Pathogens Evade Host Immunity without Coding Sequence Changes. *FEMS Microbiol. Rev.* **2021**, *45*, fuab002. [[CrossRef](#)]
21. LaBella, A.L.; Oplente, D.A.; Steenwyk, J.L.; Hittinger, C.T.; Rokas, A. Variation and Selection on Codon Usage Bias across an Entire Subphylum. *PLoS Genet.* **2019**, *15*, e1008304. [[CrossRef](#)]
22. Wang, Z.; Xu, B.; Li, B.; Zhou, Q.; Wang, G.; Jiang, X.; Wang, C.; Xu, Z. Comparative Analysis of Codon Usage Patterns in Chloroplast Genomes of Six *Euphorbiaceae* Species. *PeerJ* **2020**, *8*, e8251. [[CrossRef](#)] [[PubMed](#)]
23. 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*, 3574976. [[CrossRef](#)] [[PubMed](#)]
24. Wang, L.; Xing, H.; Yuan, Y.; Wang, X.; Saeed, M.; Tao, J.; Feng, W.; Zhang, G.; Song, X.; Sun, X. Genome-wide Analysis of Codon Usage Bias in Four Sequenced Cotton Species. *PLoS ONE* **2018**, *13*, e0194372. [[CrossRef](#)]
25. Sharp, P.M.; Li, W.H. An Evolutionary Perspective on Synonymous Codon Usage in Unicellular Organisms. *J. Mol. Evol.* **1986**, *24*, 28–38. [[CrossRef](#)]
26. Sharp, P.M.; Tuohy, T.M.; Mosurski, K.R. Codon Usage in Yeast: Cluster Analysis Clearly Differentiates Highly and Lowly Expressed Genes. *Nucleic Acids Res.* **1986**, *14*, 5125–5143. [[CrossRef](#)]
27. Vasanthi, S.; Dass, J.F.P. Comparative Genome-wide Analysis of Codon Usage of Different Bacterial Species Infecting *Oryza sativa*. *J. Cell. Biochem.* **2018**, *119*, 9346–9356. [[CrossRef](#)]
28. Villada, J.C.; Duran, M.F.; Lee, P.K.H. Interplay between Position-Dependent Codon Usage Bias and Hydrogen Bonding at the 5' End of ORFomes. *mSystems* **2020**, *5*, e00613–e00620. [[CrossRef](#)]
29. Charif, D.; Thioulouse, J.; Lobry, J.R.; Perrière, G. Online Synonymous Codon Usage Analyses with the ade4 and seqinR Packages. *Bioinform. Oxf. Engl.* **2005**, *21*, 545–547. [[CrossRef](#)]
30. Tanaka, T.; Koyanagi, K.O.; Itoh, T. Highly Diversified Molecular Evolution of Downstream Transcription Start Sites in Rice and *Arabidopsis*. *Plant Physiol.* **2009**, *149*, 1316–1324. [[CrossRef](#)]
31. Bomblies, K.; Doebley, J.F. Molecular Evolution of *FLORICAULA/LEAFY* Orthologs in the Andropogoneae (Poaceae). *Mol. Biol. Evol.* **2005**, *22*, 1082–1094. [[CrossRef](#)]
32. Zeng, J.; Fan, X.; Sha, L.; Kang, H.Y.; Wang, Y.; Zhang, H.; Zhou, Y. Molecular Evolution and Diversity of Dimeric α -Amylase Inhibitor Gene in *Kengyilia* Species (Triticeae: Poaceae). *Gene* **2013**, *529*, 262–268. [[CrossRef](#)]
33. 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*, EBO.S8092. [[CrossRef](#)]
34. Nie, X.; Deng, P.; Feng, K.; Liu, P.; Du, X.; You, F.M.; Weining, S. Comparative Analysis of Codon Usage Patterns in Chloroplast Genomes of the Asteraceae Family. *Plant Mol. Biol. Report.* **2014**, *32*, 828–840. [[CrossRef](#)]

35. Zhang, Y.; Nie, X.; Jia, X.; Zhao, C.; Biradar, S.S.; Wang, L.; Du, X.; Weining, S. Analysis of Codon Usage Patterns of the Chloroplast Genomes in the Poaceae Family. *Aust. J. Bot.* **2012**, *60*, 461–470. [[CrossRef](#)]
36. Walsh, I.M.; Bowman, M.A.; Soto Santarriaga, I.F.; Rodriguez, A.; Clark, P.L. Synonymous Codon Substitutions Perturb Cotranslational Protein Folding in vivo and Impair Cell Fitness. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 3528–3534. [[CrossRef](#)]
37. Aditama, R.; Tanjung, Z.; Sudania, W.; Nugroho, Y.; Utomo, C.; Liwang, T. Analysis of Codon Usage Bias Reveals Optimal Codons in *Elaeis guineensis*. *Biodivers. J. Biol. Divers.* **2020**, *21*, 5311–5337. [[CrossRef](#)]
38. Shah, S.; Chen, Y.; Bhattacharya, D.; Chan, C.X. Sex in Symbiodiniaceae Dinoflagellates: Genomic Evidence for Independent Loss of The Canonical Synaptonemal Complex. *Sci. Rep.* **2020**, *10*, 9792. [[CrossRef](#)]
39. Margis, R.; Fusaro, A.F.; Smith, N.A.; Curtin, S.J.; Watson, J.M.; Finnegan, E.J.; Waterhouse, P.M. The Evolution and Diversification of Dicers in Plants. *FEBS Lett.* **2006**, *580*, 2442–2450. [[CrossRef](#)]
40. Muyle, A.; Serres-Giardi, L.; Ressayre, A.; Escobar, J.; Glémin, S. GC-Biased Gene Conversion and Selection Affect GC Content in the *Oryza* Genus (rice). *Mol. Biol. Evol.* **2011**, *28*, 2695–2706. [[CrossRef](#)]
41. Escobar, J.S.; Cenci, A.; Bolognini, J.; Haudry, A.; Laurent, S.; David, J.; Glémin, S. An Integrative Test of the Dead-End Hypothesis of Selfing Evolution in Triticeae (Poaceae). *Evolution* **2010**, *64*, 2855–2872. [[CrossRef](#)] [[PubMed](#)]
42. Wendel, J.F.; Schnabel, A.; Seelanan, T. Bidirectional Interlocus Concerted Evolution Following Allopolyploid Speciation in Cotton (*Gossypium*). *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 280–284. [[CrossRef](#)] [[PubMed](#)]
43. Kovarik, A.; Matyasek, R.; Lim, K.Y.; Skalická, K.; Koukalová, B.; Knapp, S.; Chase, M.; Leitch, A.R. Concerted Evolution of 18S–5.8S–26S rDNA Repeats in *Nicotiana* Allotetraploids. *Biol. J. Linn. Soc.* **2004**, *82*, 615–625. [[CrossRef](#)]