



## Article

# Comparative Study on Codon Usage Patterns across Chloroplast Genomes of Eighteen *Taraxacum* Species

Yang Yang<sup>1</sup>, Xingliang Wang<sup>2</sup> and Zhenjie Shi<sup>1,3,\*</sup> <sup>1</sup> College of Life Sciences, Hainan Normal University, Haikou 571199, China; 20211071300007@hainnu.edu.cn<sup>2</sup> Harbin Academy of Agricultural Sciences, Harbin 150029, China; 13789700617@163.com<sup>3</sup> Sanya Research Institute, Nanjing Agricultural University, Sanya 572025, China

\* Correspondence: stou0516@gmail.com

**Abstract:** This study investigates codon usage bias within the chloroplast genomes of 18 *Taraxacum* species, focusing on the base composition and various metrics including GC content, Relative Synonymous Codon Usage (RSCU), Effective Number of Codons (ENc), and GC3s. Our analysis revealed a pronounced preference for A/T-ending codons across *Taraxacum* species, with GC content across the first, second, and third positions of the codons (GC1, GC2, GC3) and the average GC content consistently below 50%. A detailed examination using the RSCU metric identified 29 commonly preferred A/T-ending codons, indicating a strong codon usage bias towards these endings. Specifically, the codon for leucine (UUA) emerged as highly preferred, while the codon for serine (AGC) was least favored. Through the ENc–GC3s plot analysis, we explored the forces shaping this bias, finding evidence that both mutation pressure and natural selection significantly influence codon preference, with most coding sequences showing weak bias. The PR2 plot analysis further confirmed the role of these factors by demonstrating a higher frequency of T over A and C over G at the third codon position, pointing towards a mutation bias complemented by natural selection. Collectively, our findings highlight a consistent pattern of codon usage bias in the chloroplast genomes of *Taraxacum* species, influenced by a combination of mutation pressure, natural selection, and possibly other unidentified factors.

**Keywords:** chloroplast genome; codon; *Taraxacum*

**Citation:** Yang, Y.; Wang, X.; Shi, Z. Comparative Study on Codon Usage Patterns across Chloroplast Genomes of Eighteen *Taraxacum* Species. *Horticulturae* **2024**, *10*, 492. <https://doi.org/10.3390/horticulturae10050492>

Academic Editor: Jose V. Die

Received: 6 April 2024

Revised: 5 May 2024

Accepted: 8 May 2024

Published: 10 May 2024



**Copyright:** © 2024 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

Codons, composed of three adjacent nitrogen-containing bases on messenger RNA (mRNA), serve as the link between nucleic acids and proteins, playing a crucial role in the transfer of genetic information [1]. Genetic information carried by DNA is translated into amino acids through triplets of codons [2]. The frequency of synonymous codon usage varies among different organisms, from prokaryotes to eukaryotes, due to codon usage bias. Codon usage bias is influenced by environmental factors, base mutations, genetic drift, and gene expression levels [3]. Generally, codon usage bias in plants is affected by both natural selection and mutation pressure [4].

The genus *Taraxacum*, one of the largest groups within the *Asteraceae* family, comprises approximately 2400 species, primarily found in the temperate to subtropical regions of the Northern Hemisphere, with some species in tropical South America [5]. Additionally, many species within the *Taraxacum* genus have been found to produce a variety of secondary metabolites, which have been widely used in medicinal and industrial production across different countries and regions [6,7]. For example, *Taraxacum* species have been utilized in traditional Chinese medicine or as herbs of significant medicinal value [6]. In plant taxonomy, *Taraxacum* is one of the most challenging genera to classify within the *Asteraceae* family [8]. Accurate species identification and differentiation from closely related plants are fundamental for the industrial application of rubber-producing *Taraxacum* species [9].

However, extensive morphological variability, the occurrence of different ploidy levels, and asexual reproduction pose significant challenges to their classification [10]. With the rapid advancement of sequencing and omics technologies, chloroplast genome data for most *Taraxacum* species have been published, accelerating progress in species evolution and classification research [11]. Moreover, comparative analyses of chloroplast genomes have revealed that, while the overall genome structure is conserved, there are significant variations in genome size, gene content, and structural rearrangements across different plant lineages [12,13]. For example, a recent study examined codon usage patterns across the chloroplast genomes of eighteen *Taraxacum* species, providing insights into the evolutionary dynamics of these genomes [14]. We can theoretically improve the expression levels of heterologous genes [15]. Although codon preference has been extensively studied in various prokaryotic and eukaryotic model organisms, this phenomenon has received little attention in the *Taraxacum* species [16].

With the rapid development of transgenic technology, chloroplast gene transformation methods have been developed and validated by many researchers. However, to construct more stable transgenic systems, it is necessary to study codon usage patterns in plants. In this study, we compared and analyzed the codon usage bias in the chloroplast genomes of eighteen *Taraxacum* species and discussed the reasons for its formation. Parameters such as the GC content at three positions (GC1, GC2, GC3), Relative Synonymous Codon Usage (RSCU), and Effective Number of Codons (ENc) were calculated. By analyzing all the chloroplast genomes of eighteen *Taraxacum* species, this study will provide references for research on genetic transformation and molecular evolution (Table 1).

**Table 1.** Base composition of codons in the chloroplast genome of eighteen *Taraxacum* species.

Species	Assembly	GC%	GC1%	GC2%	GC3%	CDSs Number (Before Filtering)	CDSs Number (After Filtering)	L_aa
<i>T. amplum</i>	KX499525.1	37.9%	45.45%	37.99%	30.35%	85	52	22,534
<i>T. brevicorniculatum</i>	KX198559.1	39.0%	48.00%	40.02%	28.85%	81	48	16,259
<i>T. coreanum</i>	MN689809.1	37.9%	45.65%	37.96%	30.21%	86	55	23,369
<i>T. dealbatum</i>	CNA0052002	38.0%	45.87%	38.04%	30.12%	95	63	25,501
<i>T. erythrospermum</i>	MN689810.1	38.0%	45.62%	37.96%	30.28%	86	55	23,363
<i>T. hallaisanense</i>	MW067130.1	37.9%	45.57%	37.97%	30.12%	83	55	23,671
<i>T. kok-saghyz</i>	KX198560.1	38.8%	47.67%	39.69%	28.92%	81	50	17,132
<i>T. leucanthum</i>	CNA0052001	38.0%	45.89%	38.02%	30.08%	95	63	25,542
<i>T. longipyramidatum</i>	CNA0052004	38.0%	45.95%	38.01%	30.08%	96	63	25,368
<i>T. mongolicum</i>	KU736961.1	37.9%	45.62%	37.95%	30.23%	86	55	23,371
<i>T. monochlamydeum</i>	CNA0052005	38.0%	45.53%	37.67%	30.71%	98	65	28,484
<i>T. multiscaposum</i>	CNA0052003	38.0%	45.86%	38.02%	30.05%	95	63	25,671
<i>T. obtusifrons</i>	KX499524.1	37.9%	45.38%	37.87%	30.31%	84	51	22,258
<i>T. officinale</i>	KU361241.1	37.9%	45.65%	37.95%	30.23%	86	55	23,369
<i>T. parvulum</i>	CNA0052008	38.0%	45.86%	38.01%	30.04%	96	63	25,524
<i>T. platycarpum</i>	KU736960.1	37.9%	45.64%	37.94%	30.27%	86	55	23,369
<i>T. stenolobum</i>	CNA0052006	38.0%	45.53%	37.67%	30.71%	98	65	28,484
<i>T. xinyuanicum</i>	CNA0052007	38.0%	45.87%	38.00%	30.09%	96	63	25,694

Note: GC1, GC2, and GC3 represent the GC content at the first, second, and third position; L\_aa: the total number of amino acids.

## 2. Materials and Methods

### 2.1. Sequence Retrieval and Filtering

Complete chloroplast genome sequences and their gene annotations for 18 *Taraxacum* species were downloaded from the National Center for Biotechnology Information (NCBI) database (<https://www.ncbi.nlm.nih.gov>, (accessed on 20 March 2024)). The original number of protein-coding sequences (CDSs) for these 18 *Taraxacum* species varied, with counts as follows: 80, 84, 83, 83, 83, 85, 83, 83, 85, and 83 (Table 1).

To avoid errors in analysis, all CDSs from the chloroplast genomes of the 18 *Taraxacum* species were extracted based on the following criteria [17]:

- (1) The length of the CDS should be greater than 300 bp;
- (2) Each CDS must start with the start codon (ATG) and end with a stop codon (TAG, TGA, TAA);
- (3) The number of bases should be divisible by three;
- (4) CDSs must not include any internal stop codons or incorrect bases. The GC content at three positions (GC1, GC2, GC3) was then calculated using the CUSP program in EMBOSS explorer (<http://emboss.toulouse.inra.fr/>, (accessed on 20 March 2024)). Ultimately, a refined selection of CDSs, numbering between 48 to 65 for the 18 *Taraxacum* chloroplast genomes, was used for further analysis.

## 2.2. Analysis of Relative Synonymous Codon Usage (RSCU)

The RSCU value is the ratio of the observed frequency of usage of a codon to the expected frequency if all codons for the same amino acid were used equally [18]. RSCU is a statistical measure of the true frequency of each synonymous codon's relative frequency. An RSCU value of 1 implies that each synonymous codon is used at the same frequency; an RSCU value > 1 indicates it is used more frequently than others; an RSCU value < 1 means it is used less frequently. The RSCU values for all CDSs of the 18 *Taraxacum* species were calculated using Formula (1)

$$\text{RSCU} = \frac{x_{ij}}{\sum_j^{n_i} x_{ij}} n_i \quad (1)$$

where  $x_{ij}$  represents the frequency of codon  $j$  for encoding the  $i$ -th amino acid, and  $n_i$  stands for the number of synonymous codons for the  $i$ -th amino acid. An RSCU value of 1.0 indicates no codon usage bias; it is equally selected among all synonymous codons. When RSCU values are greater than 1.0, it is understood that the codon has a strong positive usage bias. Conversely, RSCU values below 1.0 indicate a negative usage bias.

## 2.3. Identification of Putative Optimal Codons

The Effective Number of Codons (ENc) is an important parameter for assessing the degree of codon usage bias, with values range from 20 (only one synonymous codon used to encode an amino acid) to 61 (all synonymous codons are used equally) [19], where a smaller ENc value indicates a stronger codon usage bias. The ENc values for each *Taraxacum* species were calculated using the CodonW 1.4.2 software (<http://codonw.sourceforge.net/>, (accessed on 20 March 2024)). Based on the ENc values, the chloroplast genes of each *Taraxacum* species were reordered from low to high.

The top and bottom 5% of genes were selected as high and low expression datasets, respectively, and their RSCU values were calculated using CodonW 1.4.2. Optimal codons were identified through the  $\Delta$ RSCU method which involves calculating the difference between the high and low RSCU values for each codon. A codon is defined as optimal if its  $\Delta$ RSCU value is greater than or equal to 0.08 and the absolute RSCU value in either the high or low expression dataset is greater than 1.

Based on the ENc values, the top and bottom 10% of genes were selected to establish high and low expression gene datasets. The RSCU values of codons in both expression libraries were calculated and compared using  $\Delta$ RSCU ( $\Delta$ RSCU = RSCU of high expression – RSCU of low expression), and codons satisfying RSCU > 1 and  $\Delta$ RSCU > 0.08 were then identified as optimal codons.

## 2.4. Codon Usage Bias Analysis

The GC content at the 1st (GC1), 2nd (GC2), and 3rd (GC3) positions of codons for 18 *Taraxacum* genome sequences was determined using the EMBOSS online website (<https://www.bioinformatics.nl/cgi-bin/emboss/cusp>, (accessed on 21 March 2024)). Furthermore, the Codon Usage Bias of 18 *Taraxacum* genome CDSs, including the ENc, Codon Adaptation Index (CAI), GC content at the third position of synonymous codons (GC3s), the total number of amino acids (L\_aa), and RSCU were analyzed using CodonW software (<http://codonw.sourceforge.net/>, (accessed on 21 March 2024)).

### 2.5. ENc–GC3s Plot Analysis

The ENc is used to assess codon usage bias at the genome level, with values ranging from 34.11 to 61. GC3s, the GC content at the third position of synonymous codons, is an important indicator that reveals trends in nucleotide ratio. GC3s is a crucial indicator of nucleotide composition and refers to the content of guanine (G) and cytosine (C) at the third position of codons, excluding Met and Trp. To explore factors that influence codon usage bias, an ENc plot was drawn with GC3s as the *x*-axis and ENc as the *y*-axis. The expected ENc value was calculated using Formula (2).

$$\text{ENc} = 2 + S + \frac{29}{S^2 + (1 - S)^2} \quad (2)$$

where *S* stands for GC3s. The ENc plot reveals the relationship between codon bias and base composition. If mutational pressure plays a significant role in forming codon usage patterns, ENc values will be located near or within the expected curve. When natural selection and other variables affect codon usage, ENc values will be much lower than the expected curve. If the codon usage bias is primarily influenced by mutation pressure, genes will be located near or on the standard curve. Conversely, if codon usage bias is affected by natural selection, genes will be located below the expected curve.

### 2.6. PR2 Plot Analysis

Parity Rule 2 plot analysis is commonly used to estimate the impact of mutation pressure and natural selection on codon bias as a graphical analysis revealing the base composition at the third position of each codon. We established a graph with A3/(A3 + T3) as the *y*-axis and G3/(G3 + C3) as the *x*-axis where points around the center point (A=T, G=C) reveal the degree and direction of base bias. To further analyze the usage and relationship of purines and pyrimidines at the third codon position in the genome sequences, a PR2 plot was drawn with G3/(G3 + C3) as the *x*-axis and A3/(A3 + T3) as the *y*-axis. The proportion of A, T, G, and C bases determines whether base mutations affect the variation of nucleotide bases. If the proportion of G and C (or A and T) is similar, mutation pressure fully influences the codon usage bias. The center point means that there is no bias between natural selection and mutation pressure. If genes are uniformly distributed around the center, consider that the codon bias might be entirely caused by mutational pressure. When their proportions are too different, natural selection and other variables can influence codon usage bias.

### 2.7. Neutrality Plot Analysis

Neutrality plot analysis is used to estimate the degree of impact of mutation pressure and natural selection on codon usage bias, based on the neutrality plot (GC12–GC3), where GC12 represents the average GC content of the first and second codon positions, and GC3 represents the GC content at the third codon position. A scatter plot is created with GC12 as the *y*-axis and GC3 as the *x*-axis. GC12 represents the average GC content of GC1 and GC2, with the chart having GC12 as the *y*-axis and GC3 as the *x*-axis. GC12 is the average GC content at the first and second positions of the codon. GC3 for each *Taraxacum* species chloroplast gene was calculated using a Perl script ([http://GitHub-hxiang1019/calc\\_GC\\_content](http://GitHub-hxiang1019/calc_GC_content), (accessed on 22 March 2024)). A regression slope of 0 indicates that codon bias is entirely affected by natural selection, while a significant correlation with a slope of 1 suggests that mutational pressure may be the sole driving force. A coefficient of the regression curve close to or equal to 1 indicates that mutation pressure is the main factor in codon usage bias. Conversely, a coefficient close to or equal to 0 means that natural selection is the main factor.

### 2.8. Correspondence Analysis (COA)

For the study of codon usage patterns, COA is a multivariate statistical method used to discuss the variation of RSCU and the distribution of genes in multidimensional space. In this study, 59 codons (excluding Met, Trp, and the three stop codons) were used to generate orthogonal axes to reflect the diversity of codon usage. The range of the first axis (axis 1) represents the most significant factor affecting codon usage frequency variation, with the remaining 58 axes representing progressively smaller factors. COA reveals the most variable factors affecting codon usage patterns in CDSs. The relationship between axis 1, axis 2, and codon usage indices, including GC content at the third position of synonymous codons (GC3s), Effective Number of Codons (ENc), total number of amino acids (Laa), and Codon Adaptation Index (CAI), was plotted using Origin 9.8.0 software [20] to further study components affecting codon usage bias.

### 2.9. Statistical Analysis

Codon usage frequency was calculated using the CUSP in-line program in EMBOSS (<https://www.bioinformatics.nl/cgi-bin/emboss/cusp>, (accessed on 22 March 2024)). Charts in this manuscript were completed using the R 4.3.1 programming language and OriginPro 2021 software. The entire manuscript was edited using Microsoft 365 Business Premium.

## 3. Results

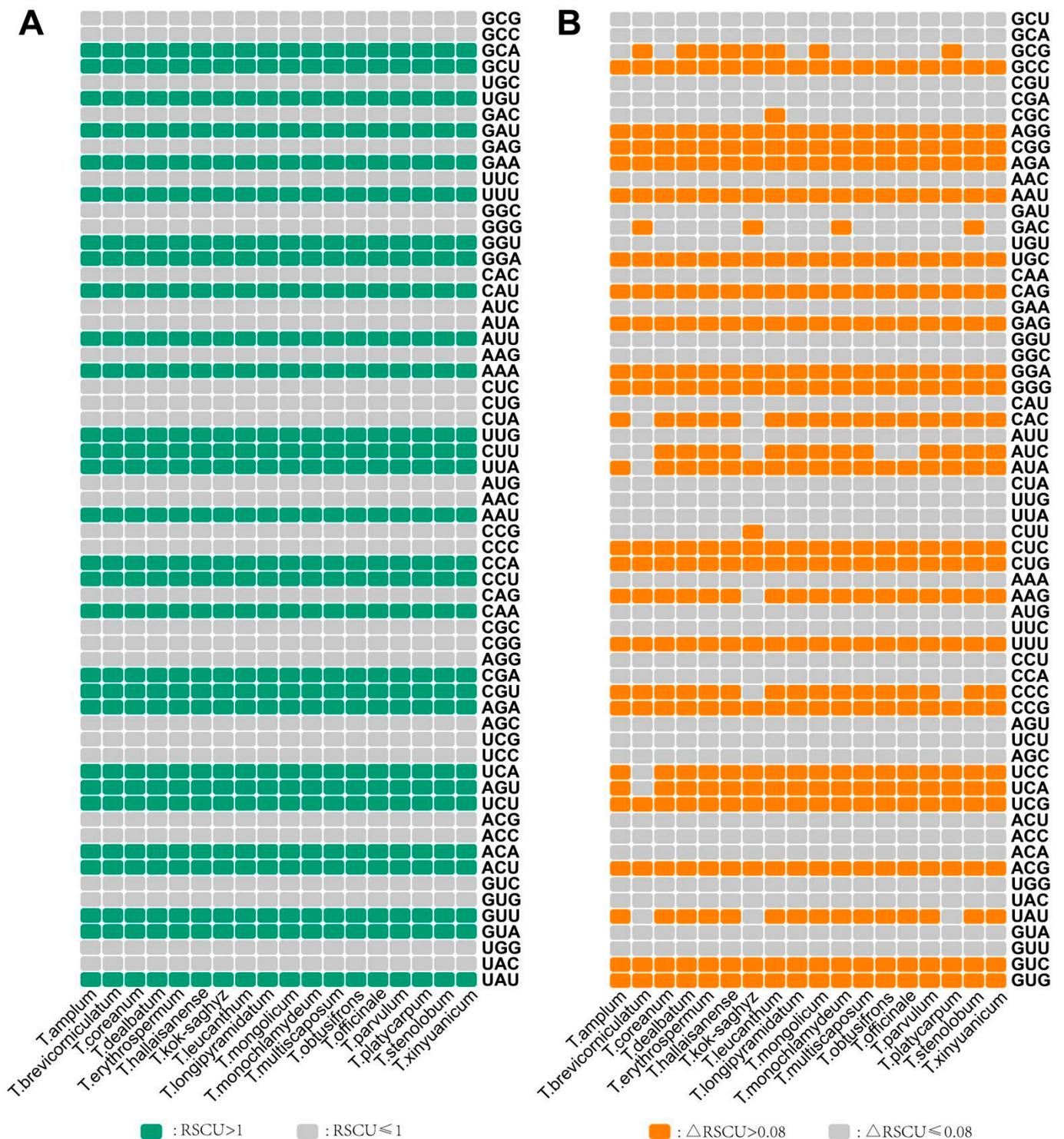
### 3.1. Characteristics of Codon Usage Bias Analysis

The characteristics of codon usage bias analysis based on the base composition in the chloroplast genomes of 18 *Taraxacum* species were studied through analysis of GC1, GC2, GC3, and the average GC content (Table 1). It was observed that in all 18 *Taraxacum* species, the frequencies of GC1, GC2, and GC3, as well as the average GC content, were all below 50%, indicating a preference for codons ending in A/T. Furthermore, the frequency of GC1 was higher than that of GC2 and GC3, with GC3 showing the lowest frequency. The average GC content of these 18 *Taraxacum* species was similar, ranging from 37.9% to 39.0%, indicating a high degree of codon preference among the 18 *Taraxacum* species. The number of protein-coding genes in the chloroplast genomes ranged from 81 to 98. After quality control, the number of genes ranged from 48 to 65.

### 3.2. RSCU and Identification of Optimal Codons

The analysis of codon usage bias was conducted through the RSCU metric. In the chloroplast genomes of 18 *Taraxacum* species, there were 29 common preferred codons with RSCU values greater than 1. This indicates a clear preference for A/T-ending codons in the chloroplast genomes of the 18 *Taraxacum* species. The obtained RSCU values ranged from 0.2837 to 2.025. Among these values, the codon for LEU, UUA showed a strong preference (RSCU > 1.8) in all 18 *Taraxacum* species, while the codon for Ser, AGC showed the lowest codon preference (RSCU < 0.4) among the 18 *Taraxacum* species. Of these codons, 28 ended in A/T, accounting for 96.55% of the total number of preferred codons, which is significantly higher than the number of G/C-ending preferred codons (Table S1). Overall, the types and numbers of preferred codons in the chloroplast genomes of the 18 *Taraxacum* species are highly similar, indicating a consistent pattern of codon usage.

Based on the criteria of RSCU > 1 and  $\Delta$ RSCU > 0.08, optimal codons were identified. The results, shown in Figure 1 and Table S2, revealed a total of 22–28 optimal codons. There were 18 optimal codons common to all 18 *Taraxacum* species.

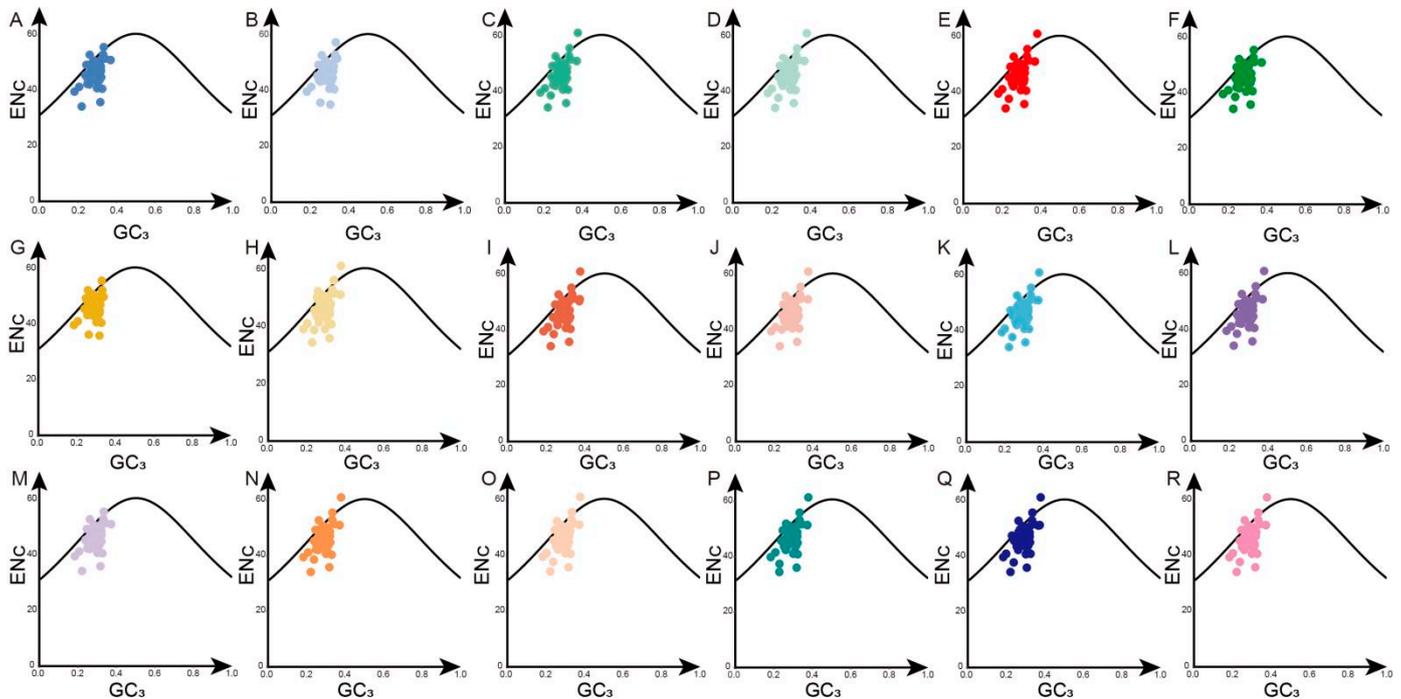


**Figure 1.** Relative synonymous codon usage (A) and optimal codon analysis (B) of eighteen *Taraxacum* species.

### 3.3. ENc Plot

The ENc and GC3s plot was used to investigate the main factors affecting codon usage bias. Figure 2 shows that the ENc-GC3 patterns of the 18 *Taraxacum* species genomes were similar. The ENc values ranged from 34.11 to 61, indicating varying codon preference trends. Using an ENc value of 35 as the criterion to distinguish between weak and strong codon usage bias, only three coding sequences exhibiting strong bias were identified (ENc < 35).

Consistent with the average ENc values (47.10–48.66), most coding sequences exhibited only weak bias. This indicates that while mutation pressure plays a role in codon usage patterns, other factors, particularly natural selection, have a significant impact.



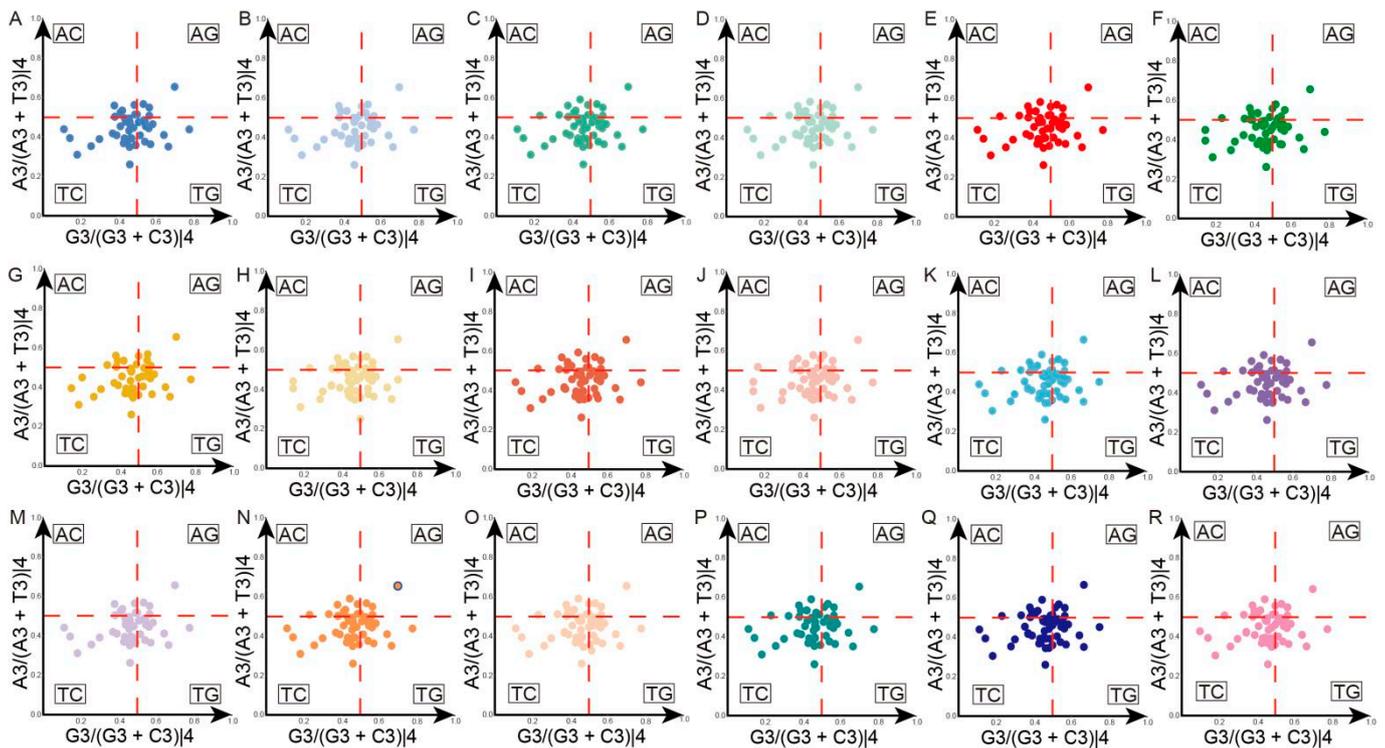
**Figure 2.** ENc map of the chloroplast genomes of eighteen *Taraxacum* species. (A) *T. amplum*; (B) *T. brevicorniculatum*; (C) *T. coreanum*; (D) *T. dealbatum*; (E) *T. erythrosperrum*; (F) *T. hallaisanense*; (G) *T. kok-saghyz*; (H) *T. leucanthum*; (I) *T. longipyramidatum*; (J) *T. mongolicum*; (K) *T. monochlamydeum*; (L) *T. multiscaposum*; (M) *T. obtusifrons*; (N) *T. officinale*; (O) *T. parvulum*; (P) *T. platycarpum*; (Q) *T. stenolobum*; (R) *T. xinyuanicum*.

### 3.4. PR2 Plot

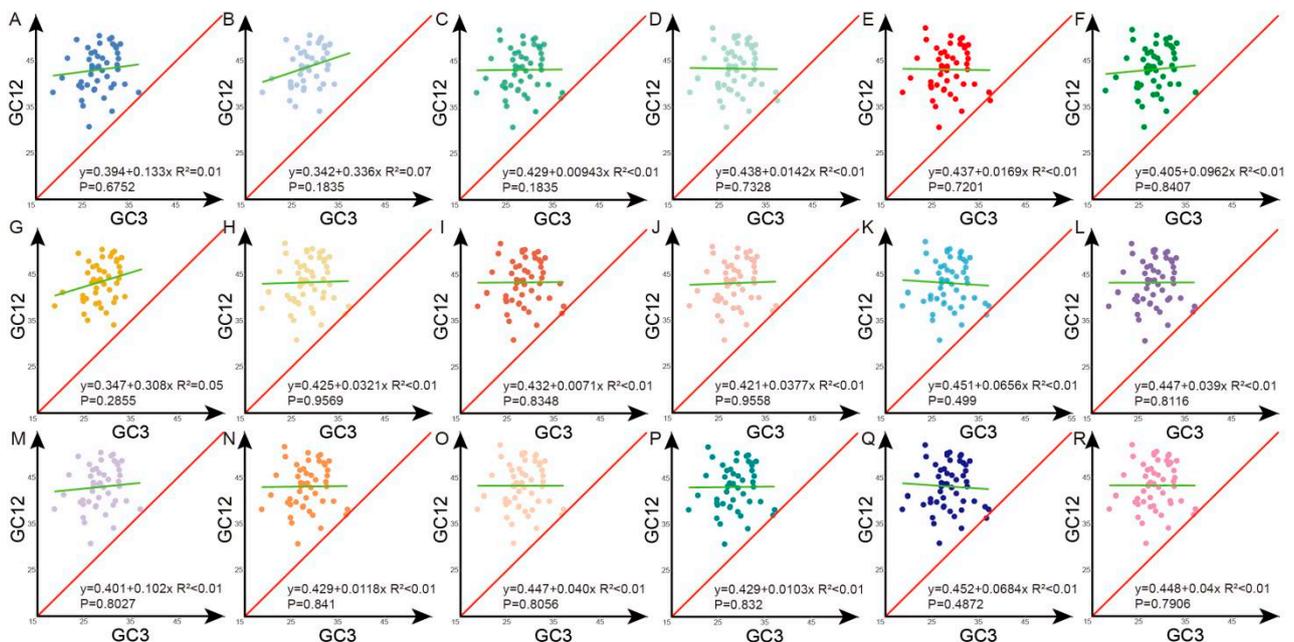
A PR2 plot was used to investigate the impact of mutation and natural selection on codon usage bias by analyzing the use of A/T and G/C at the third position of the codons. Figure 3 shows that most genes from the 18 *Taraxacum* species types were in the lower left area, indicating a higher frequency of T at the third position of the codon compared to A, and a greater frequency of C compared to G. This indicates similar third-position codon usage among the 18 *Taraxacum* species types, with influencing factors for codon bias including mutation, natural selection, and other unknown elements.

### 3.5. Neutrality Plot

Based on neutrality analysis (GC12 vs. GC3, Figure 4), the impact of mutation pressure and natural selection was quantitatively assessed. Figure 4 shows that the regression line slopes for the chloroplast genomes of the 18 *Taraxacum* species ranged from 0.0071 to 0.336, with correlation coefficients ( $r = 0.1835$  to  $0.9569$ ) indicating a weak correlation between GC12 and GC3. These results suggest that mutation pressure accounts for only 0.71% to 33.60% of the factors influencing codon usage bias in the 18 *Taraxacum* species chloroplast genomes, while natural selection and other factors account for 66.40% to 99.29%. Therefore, the influence of mutation pressure on codon bias is limited, while natural selection plays an important, if not dominant, role. These findings are consistent with the analyses of the ENc-GC3s and PR2 plots.



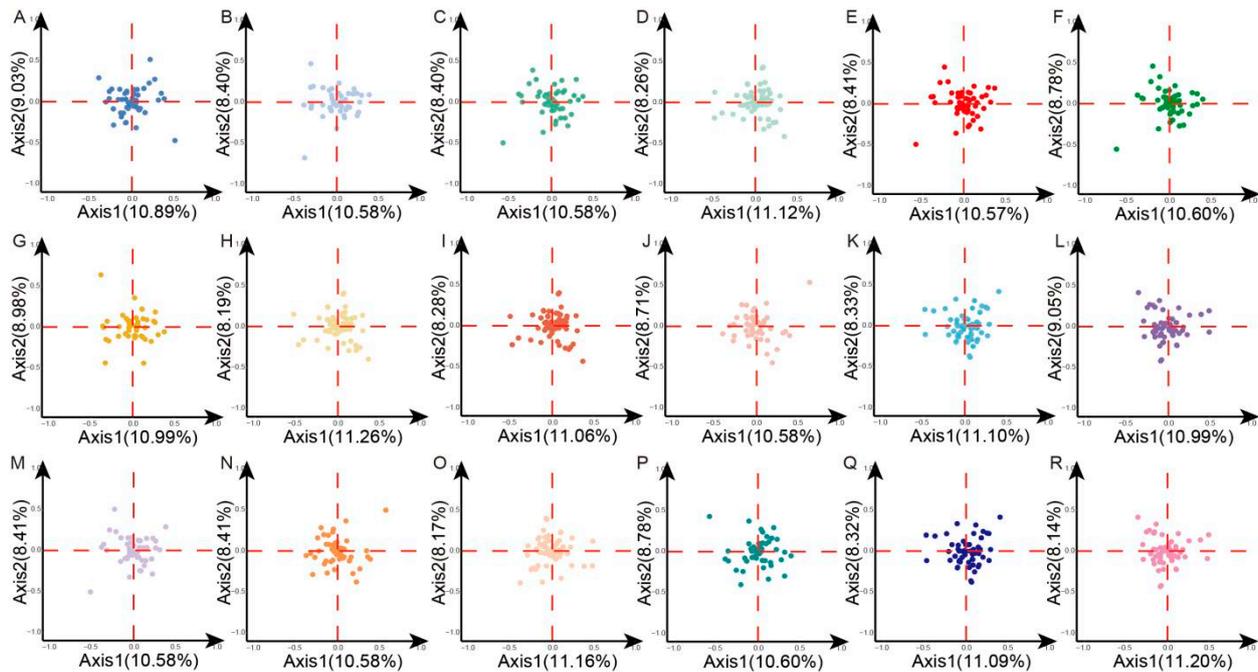
**Figure 3.** PR2 map of the chloroplast genomes of eighteen *Taraxacum* species. (A) *T. amplum*; (B) *T. brevicorniculatum*; (C) *T. coreanum*; (D) *T. dealbatum*; (E) *T. erythrospermum*; (F) *T. hallaisanense*; (G) *T. kok-saghyz*; (H) *T. leucanthum*; (I) *T. longipyramidatum*; (J) *T. mongolicum*; (K) *T. monochlamydeum*; (L) *T. multiscaposum*; (M) *T. obtusifrons*; (N) *T. officinale*; (O) *T. parvulum*; (P) *T. platycarpum*; (Q) *T. stenolobum*; (R) *T. xinyuanicum*.



**Figure 4.** Neutral map of the chloroplast genomes of eighteen species of *Taraxacum*. (A) *T. amplum*; (B) *T. brevicorniculatum*; (C) *T. coreanum*; (D) *T. dealbatum*; (E) *T. erythrospermum*; (F) *T. hallaisanense*; (G) *T. kok-saghyz*; (H) *T. leucanthum*; (I) *T. longipyramidatum*; (J) *T. mongolicum*; (K) *T. monochlamydeum*; (L) *T. multiscaposum*; (M) *T. obtusifrons*; (N) *T. officinale*; (O) *T. parvulum*; (P) *T. platycarpum*; (Q) *T. stenolobum*; (R) *T. xinyuanicum*.

### 3.6. Correspondence Analysis (COA)

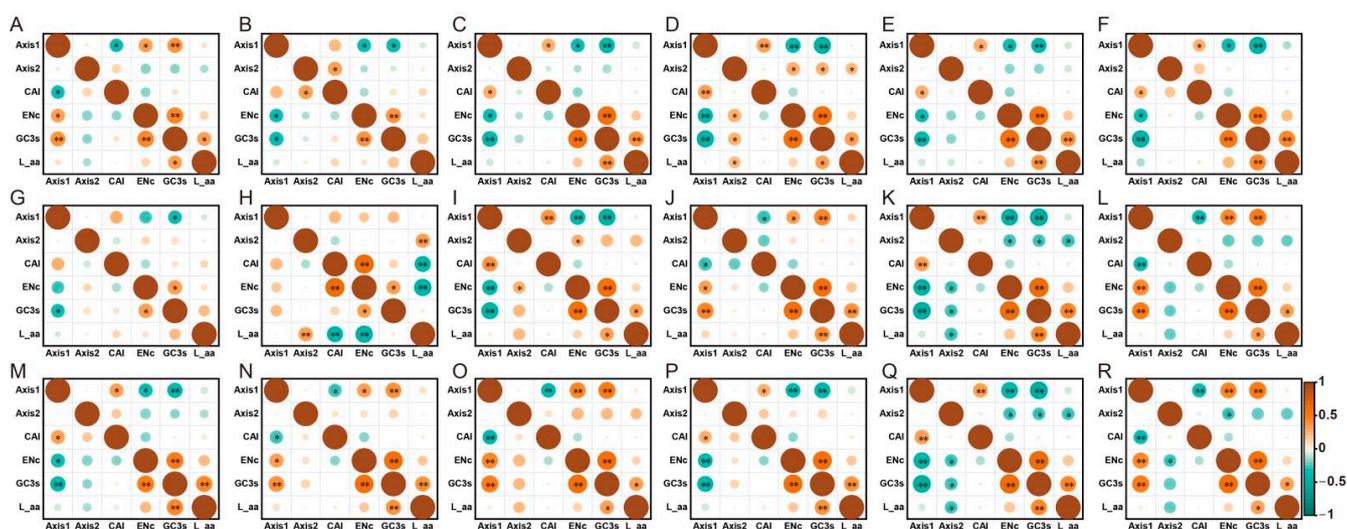
Correspondence analysis was conducted to investigate the main contributing factors to codon usage variation in the *Taraxacum* species, with a focus on RSCU values. Figure 5 shows that axis 1 explained 10.57% to 11.39% of the total variation in the chloroplast genomes of the 18 *Taraxacum* species, while axis 2 explained 8.14% to 9.05% of the variation, indicating that axis 1 and axis 2 have the most significant impact, especially axis 1. These results suggest that gene expression levels and gene length affect codon usage bias in the 18 *Taraxacum* species chloroplast genomes, with gene expression having a greater impact.



**Figure 5.** Correspondence analysis of chloroplast genomes of eighteen species of *Taraxacum*. (A) *T. amplum*; (B) *T. brevicorniculatum*; (C) *T. coreanum*; (D) *T. dealbatum*; (E) *T. erythrospermum*; (F) *T. hallaisanense*; (G) *T. kok-saghyz*; (H) *T. leucanthum*; (I) *T. longipyramidatum*; (J) *T. mongolicum*; (K) *T. monochlamydeum*; (L) *T. multiscaposum*; (M) *T. obtusifrons*; (N) *T. officinale*; (O) *T. parvulum*; (P) *T. platycarpum*; (Q) *T. stenolobum*; (R) *T. xinyuanicum*.

In these 18 chloroplast genomes of the *Taraxacum* species, no single, clear trend in codon usage preference is evident, likely due to multiple influencing factors. As shown in Figure 6, in *T. brevicorniculatum*, *T. coreanum*, *T. dealbatum*, *T. erythrospermum*, *T. hallaisanense*, *T. longipyramidatum*, *T. obtusifrons*, and *T. platycarpum*, both ENc and GC3s exhibit a negative correlation with axis 1. This indicates that higher GC content in these chloroplast genomes might suppress codon usage diversity to maintain the accuracy and stability of gene expression. In *T. amplum*, *T. mongolicum*, *T. multiscaposum*, *T. officinale*, and *T. parvulum*, the negative correlation between CAI and axis 1 suggests that a reduction in the codon adaptation index might lead to a decline in gene expression efficiency, possibly due to selection pressures in specific environments. Similarly, in *T. monochlamydeum* and *T. stenolobum*, the negative correlation between ENc and GC3s, and the correlation of these indices with L\_aa on axis 2, highlights the relationship between genomic diversity and coding efficiency. Additionally, the negative correlation between CAI and axis 1, and between ENc and axis 2 in *T. xinyuanicum*, further demonstrates the dual inhibition of codon adaptation and diversity. On the other hand, in *T. amplum*, *T. mongolicum*, *T. multiscaposum*, *T. officinale*, *T. parvulum*, and *T. xinyuanicum*, the positive correlation between ENc and GC3s with axis 1 reveals a positive relationship between GC content at the third codon position and the number of effective codons, potentially enhancing gene expression efficiency to better meet

biological functional needs. The positive correlation between CAI and axis 1 in *T. coreanum*, *T. erythrospermum*, *T. hallaisanense*, *T. monochlamydeum*, *T. obtusifrons*, *T. platycarpum*, and *T. stenolobum* indicates that these genomes may respond to functional needs by improving gene expression efficiency. In *T. brevicorniculatum* and *T. dealbatum*, CAI is positively correlated with axis 2 and axis 1, respectively. Meanwhile, the positive correlation of ENc, GC3s, and L\_aa with axis 2 in *T. dealbatum* demonstrates that these genomes improve gene expression adaptability through the combined effect of various factors. The positive correlation of L\_aa with axis 2 in *T. leucanthum* emphasizes the impact of specific amino acid numbers on gene expression, while in *T. longipyramidatum*, the positive correlation between CAI and axis 1, and between ENc and axis 2, reflects the complex interactions among these factors. The regulatory patterns exhibited by different genomes may reflect strategies through which these genomes adjust gene expression and adapt to the demands of biological function under varying ecological environments and evolutionary pressures.



**Figure 6.** Correlation analysis of axis 1, axis 2, and codon utilization index of chloroplast genomes of eighteen species of *Taraxacum*. (A) *T. amplum*; (B) *T. brevicorniculatum*; (C) *T. coreanum*; (D) *T. dealbatum*; (E) *T. erythrospermum*; (F) *T. hallaisanense*; (G) *T. kok-saghyz*; (H) *T. leucanthum*; (I) *T. longipyramidatum*; (J) *T. mongolicum*; (K) *T. monochlamydeum*; (L) *T. multiscaposum*; (M) *T. obtusifrons*; (N) *T. officinale*; (O) *T. parvulum*; (P) *T. platycarpum*; (Q) *T. stenolobum*; (R) *T. xinyuanicum*. GC3s indicates the GC content at the third codon position of synonymous codons; ENc represents the effective number of codons; CAI means the codon adaptation index; L\_aa is defined as the total number of amino acids. \* Represents  $p < 0.05$ , \*\* represents  $p < 0.01$ .

#### 4. Discussion

This study performed a comprehensive analysis of codon usage bias specifically within the chloroplast genomes of 18 *Taraxacum* species, uncovering a consistent pattern unique to these genomes. Although prior research has explored comparative genomics in the genus, our focus on chloroplast genomes provides new insights into codon usage bias across these species [21]. Basic compositional analysis revealed that these species all exhibit high A/T content and low G/C content, favoring codons ending in A/T. This pattern is also observed in other plant species. A comparison of Relative Synonymous Codon Usage (RSCU) values further revealed 29 preferred codons common across the *Taraxacum* genus, with the vast majority (96.55%) ending in A/T [18,22,23]. This phenomenon suggests an adaptive mechanism potentially driven by the chloroplast genome environment. The lower G/C content may indicate a selection pressure favoring more efficient DNA replication and transcription processes, as the synthesis of A/T pairings requires less energy than that of G/C pairings [3,24]. The preference for codons ending in A/T observed in the *Taraxacum* species may be associated with their adaptation to specific ecological environments, al-

though this relationship has not been conclusively proven [25–27]. Further research should explore how environmental conditions could influence changes in codon usage preferences and their impact on plant survival. The significant preference for certain codons (e.g., for LEU, UUA) implies a selective optimization of the translation process in chloroplasts. This optimization might be linked to the evolutionary adaptation of dandelion species to specific environmental conditions or physiological needs [28]. This concurs with the theory that codon usage bias is not merely a result of genetic drift but also reflects the balance between mutational pressure and natural selection [29–31].

This study explored various biological factors affecting codon preference in the *Taraxacum* genus, including gene expression levels, gene length, tRNA abundance, mutational tendencies, and GC content. While these factors play roles in codon usage bias, our analysis underscores the significance of mutational pressure and natural selection as dominant factors, aligning with Hanson et al. who proposed that codon usage bias might reflect a balance between coding efficiency and transcriptional accuracy [32–34]. Moreover, our findings emphasize the unique role of natural selection in the formation of codon preferences within the chloroplast genome, supporting the hypothesis of Bhattacharyya et al. which suggests that the abundance of specific tRNA molecules directly influences the preference for certain codons during translation. This is because the availability of tRNA species directly impacts the efficiency and speed of protein synthesis, as the cell preferentially uses codons that correspond to the more abundant tRNA molecules. This selective pressure leads to the evolution of codon preferences that align with the tRNA pool within the cell, optimizing the translation process and ensuring efficient protein production [34–36]. Analysis through PR2 plots and ENc-GC3s revealed that codon preferences are shaped not only by mutations but also significantly by natural selection. This resonates with the Rocha view that natural selection plays a decisive role in optimizing codon usage [37,38]. Particularly in the *Taraxacum* species, despite the influence of mutational pressure on codon usage, natural selection and other factors (such as gene expression levels and tRNA abundance) predominantly shape codon preferences [39,40]. This discovery not only highlights the importance of natural selection in determining codon usage bias but also correlates with findings in other studies, such as those on *Euphorbiaceae* and *Epimedium* species [41,42]. Furthermore, our results align with those of other species, emphasizing the widespread influence of natural selection on codon usage preferences in plant genomes. For instance, Dhindsa et al. also reported the decisive role of natural selection in optimizing codon usage across different organisms [43]. This consensus not only deepens our understanding of the mechanisms behind codon usage preference formation but also lays the foundation for further exploration of codon usage patterns under different ecological and evolutionary backgrounds.

Additionally, research on determining optimal codons provides a basis for enhancing the expression of exogenous genes in plants through codon optimization design. In our study, we identified 18 common optimal codons across 18 *Taraxacum* species genomes. This not only guides effective codon optimization but also enriches our understanding of the relationship between codon preference and gene expression.

## 5. Conclusions

Species within the dandelion genus undeniably play a crucial role in agricultural economics, global international trade, and human daily life. In this study, bioinformatics methods were used to systematically analyze the codon usage preferences of 18 dandelion genomes. The analysis revealed significant similarities in codon usage patterns among these genomes, exhibiting a preference for codons ending in A/T. This study determined that codon preference is influenced by natural selection, mutation pressure, and other factors, with natural selection being the primary determinant. The findings of this study are significant for researching the evolution of dandelions and improving the expression efficiency of exogenous genes. It provides new avenues for genetic transformation in the cultivation of dandelions. Based on this, we will continue to conduct research on

gene function verification, genetic transformation, targeted gene codon optimization, and directed transformation, making further efforts for the development of the dandelion medicinal industry.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/horticulturae10050492/s1>, Supplementary File S1: Table S1. RSCU values of codons in the chloroplast genomes of 18 *Taraxacum* species. Supplementary File S2: Table S2. RSCU and  $\Delta$ RSCU values in the high and low expression libraries in the chloroplast genomes of 18 *Taraxacum* species.

**Author Contributions:** Conceptualization, Y.Y. and X.W.; Formal analysis, Y.Y.; Resources, Y.Y.; Writing—original draft, Y.Y.; Writing—review and editing, Y.Y., X.W. and Z.S.; Visualization, Y.Y.; Project administration, Y.Y.; Funding acquisition, Z.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** The special research fund for doctoral students of Sanya YazhouBay Science and Technology City (HSPHDSRF-2022-07-002).

**Data Availability Statement:** Data are contained within the article and Supplementary Materials.

**Acknowledgments:** Thanks to the College of Life Sciences of Hainan Normal University for their help in the experiment.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

A	Adenine
T	Thymine
C	Cytosine
G	Guanine
GC1, GC2, GC3	The G + C content at the first, second, third codon positions
A3, T3, G3, C3	The content of A, T, G, and C at the third codon position
GC12	The average GC content at the first and second codon positions
RSCU	Relative synonymous codon usage
RFSC	Relative synonymous codon usage frequency
ENc	Effective number of codons
PR2	Parity Rule 2
COA	Correspondence analysis
GC3s	GC content at the third codon position of synonymous codons
L_aa	Total number of amino acids
CAI	Codon adaptation index
NCBI	National Center for Biotechnology Information

## References

- Manyanga, F.; Sithole, A. *Nucleic Acids, Structure and Function for General Biochemistry, Biology and Biotechnology*; Lulu.com: Morrisville, NC, USA, 2014.
- Zolyan, S. On the minimal elements of the genetic code and their semiotic functions (degeneracy, complementarity, wobbling). *Biosystems* **2023**, *231*, 104962. [[CrossRef](#)]
- Parvathy, S.T.; Udayasuriyan, V.; Bhadana, V. Codon usage bias. *Mol. Biol. Rep.* **2022**, *49*, 539–565. [[CrossRef](#)]
- Hu, H.; Dong, B.; Fan, X.; Wang, M.; Wang, T.; Liu, Q. Mutational bias and natural selection driving the synonymous codon usage of single-exon genes in rice (*Oryza sativa* L.). *Rice* **2023**, *16*, 11. [[CrossRef](#)]
- Lin, Y.-X.; Zhang, L.-B.; Zhang, X.-C.; He, Z.-R.; Wang, Z.-R.; Lu, S.-G.; Wu, S.-G.; Xing, F.-W.; Zhang, G.-M.; Liao, W.-B. Flora of China. *Harv. Pap. Bot.* **2013**, *13*, 301–302.
- Martinez, M.; Poirrier, P.; Chamy, R.; Prüfer, D.; Schulze-Gronover, C.; Jorquera, L.; Ruiz, G. *Taraxacum officinale* and related species—An ethnopharmacological review and its potential as a commercial medicinal plant. *J. Ethnopharmacol.* **2015**, *169*, 244–262. [[CrossRef](#)]
- Sharifi-Rad, M.; Roberts, T.H.; Matthews, K.R.; Bezerra, C.F.; Morais-Braga, M.F.B.; Coutinho, H.D.; Sharopov, F.; Salehi, B.; Yousaf, Z.; Sharifi-Rad, M. Ethnobotany of the genus *Taraxacum*—Phytochemicals and antimicrobial activity. *Phytother. Res.* **2018**, *32*, 2131–2145. [[CrossRef](#)]

8. Zeisek, V. *Taxonomic Principles, Reproductive Systems, Population Genetics and Relationships within Selected Groups of Genus Taraxacum (Asteraceae)*; Charles University: Prague, Czech Republic, 2018.
9. Zhang, Y.; Iaffaldano, B.J.; Zhuang, X.; Cardina, J.; Cornish, K. Chloroplast genome resources and molecular markers differentiate rubber dandelion species from weedy relatives. *BMC Plant Biol.* **2017**, *17*, 34. [[CrossRef](#)]
10. Hörandl, E. The classification of asexual organisms: Old myths, new facts, and a novel pluralistic approach. *Taxon* **2018**, *67*, 1066–1081. [[CrossRef](#)]
11. Tan, Y.; Cao, J.; Tang, C.; Liu, K. Advances in Genome Sequencing and Natural Rubber Biosynthesis in Rubber-Producing Plants. *Curr. Issues Mol. Biol.* **2023**, *45*, 9342–9353. [[CrossRef](#)]
12. Daniell, H.; Lin, C.-S.; Yu, M.; Chang, W.-J. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol.* **2016**, *17*, 1–29. [[CrossRef](#)]
13. Zhao, J.; Chen, H.; Li, G.; Jumaturti, M.A.; Yao, X.; Hu, Y. Phylogenetics Study to Compare Chloroplast Genomes in Four Magnoliaceae Species. *Curr. Issues Mol. Biol.* **2023**, *45*, 9234–9251. [[CrossRef](#)]
14. Ji, K.-k.; Song, X.; Chen, C.-G.; Li, G.; Xie, S.-Q. Codon usage profiling of chloroplast genome in Magnoliaceae. *J. Agric. Sci. Technol.* **2020**, *22*, 52–62.
15. Fath, S.; Bauer, A.P.; Liss, M.; Spriestersbach, A.; Maertens, B.; Hahn, P.; Ludwig, C.; Schäfer, F.; Graf, M.; Wagner, R. Multiparameter RNA and codon optimization: A standardized tool to assess and enhance autologous mammalian gene expression. *PLoS ONE* **2011**, *6*, e17596. [[CrossRef](#)]
16. Bahiri-Elitzur, S.; Tuller, T. Codon-based indices for modeling gene expression and transcript evolution. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 2646–2663. [[CrossRef](#)]
17. Wei, Z.; Zhu, S.-X.; Van den Berg, R.; Bakker, F.T.; Schranz, M.E. Phylogenetic relationships within *Lactuca* L. (Asteraceae), including African species, based on chloroplast DNA sequence comparisons. *Genet. Resour. Crop Evol.* **2017**, *64*, 55–71. [[CrossRef](#)]
18. Das, J.K.; Roy, S. Comparative analysis of human coronaviruses focusing on nucleotide variability and synonymous codon usage patterns. *Genomics* **2021**, *113*, 2177–2188. [[CrossRef](#)]
19. 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](#)]
20. OriginLab Corporation. *Origin(Pro), Version 2021*; OriginLab Corporation: Northampton, MA, USA, 2021.
21. He, M.; Han, X.; Qin, X.; Bao, J.; Li, H.; Xie, Q.; Yang, Y.; Jin, X. Comparative chloroplast genome analyses provide new insights into phylogeny of *Taraxacum* and molecular markers for distinguishing rubber producing dandelions from their weedy relatives in China. *Ind. Crop. Prod.* **2024**, *207*, 117712. [[CrossRef](#)]
22. Li, Q.; Luo, Y.; Sha, A.; Xiao, W.; Xiong, Z.; Chen, X.; He, J.; Peng, L.; Zou, L. Analysis of synonymous codon usage patterns in mitochondrial genomes of nine *Amanita* species. *Front. Microbiol.* **2023**, *14*, 1134228. [[CrossRef](#)]
23. Xu, C.; Dong, J.; Tong, C.; Gong, X.; Wen, Q.; Zhuge, Q. Analysis of synonymous codon usage patterns in seven different citrus species. *Evol. Bioinform.* **2013**, *9*, EBO-S11930. [[CrossRef](#)]
24. Kornberg, A.; Baker, T.A. *DNA Replication*; University Science Books: New York, NY, USA, 2005.
25. Li, Z.; Huang, Z.; Wan, X.; Yu, J.; Dong, H.; Zhang, J.; Zhang, C.; Wang, S. Complete chloroplast genome sequence of *Rhododendron mariesii* and comparative genomics of related species in the family Ericaceae. *Comp. Cytogenet.* **2023**, *17*, 163. [[CrossRef](#)]
26. Salih, M.; Hussein, R. *Nuclear and Chloroplast Genome Diversity in Apomictic Microspecies of Taraxacum*; University of Leicester: Leicester, UK, 2017.
27. Liu, Q.; Li, X.; Li, M.; Xu, W.; Schwarzacher, T.; Heslop-Harrison, J.S. Comparative chloroplast genome analyses of *Avena*: Insights into evolutionary dynamics and phylogeny. *BMC Plant Biol.* **2020**, *20*, 406. [[CrossRef](#)]
28. Freeland, J.R. *Molecular Ecology*; John Wiley & Sons: Hoboken, NJ, USA, 2020.
29. Plotkin, J.B.; Dushoff, J.; Desai, M.M.; Fraser, H.B. Codon usage and selection on proteins. *J. Mol. Evol.* **2006**, *63*, 635–653. [[CrossRef](#)]
30. Trotta, E. Selection on codon bias in yeast: A transcriptional hypothesis. *Nucleic Acids Res.* **2013**, *41*, 9382–9395. [[CrossRef](#)]
31. Cutter, A.D.; Wasmuth, J.D.; Blaxter, M.L. The evolution of biased codon and amino acid usage in nematode genomes. *Mol. Biol. Evol.* **2006**, *23*, 2303–2315. [[CrossRef](#)]
32. Hanson, G.; Collier, J. Codon optimality, bias and usage in translation and mRNA decay. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 20–30. [[CrossRef](#)]
33. Qian, W.; Yang, J.-R.; Pearson, N.M.; Maclean, C.; Zhang, J. Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet.* **2012**, *8*, e1002603. [[CrossRef](#)]
34. Duan, H.; Zhang, Q.; Wang, C.; Li, F.; Tian, F.; Lu, Y.; Hu, Y.; Yang, H.; Cui, G. Analysis of codon usage patterns of the chloroplast genome in *Delphinium grandiflorum* L. reveals a preference for AT-ending codons as a result of major selection constraints. *PeerJ* **2021**, *9*, e10787. [[CrossRef](#)]
35. Bhattacharyya, D.; Uddin, A.; Das, S.; Chakraborty, S. Mutation pressure and natural selection on codon usage in chloroplast genes of two species in *Pisum* L. (Fabaceae: Faboideae). *Mitochondrial DNA Part A* **2019**, *30*, 664–673. [[CrossRef](#)]
36. Wang, Z.; Cai, Q.; Wang, Y.; Li, M.; Wang, C.; Wang, Z.; Jiao, C.; Xu, C.; Wang, H.; Zhang, Z. Comparative analysis of codon bias in the chloroplast genomes of theaceae species. *Front. Genet.* **2022**, *13*, 824610. [[CrossRef](#)]
37. Zhou, Z.; Dang, Y.; Zhou, M.; Li, L.; Yu, C.-H.; Fu, J.; Chen, S.; Liu, Y. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E6117–E6125. [[CrossRef](#)] [[PubMed](#)]

38. Rocha, E.P. Codon usage bias from tRNA's point of view: Redundancy, specialization, and efficient decoding for translation optimization. *Genome Res.* **2004**, *14*, 2279–2286. [[CrossRef](#)] [[PubMed](#)]
39. Quax, T.E.; 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](#)] [[PubMed](#)]
40. Novoa, E.M.; Pavon-Eternod, M.; Pan, T.; de Pouplana, L.R. A role for tRNA modifications in genome structure and codon usage. *Cell* **2012**, *149*, 202–213. [[CrossRef](#)] [[PubMed](#)]
41. Wang, Y.; Jiang, D.; Guo, K.; Zhao, L.; Meng, F.; Xiao, J.; Niu, Y.; Sun, Y. Comparative analysis of codon usage patterns in chloroplast genomes of ten *Epimedium* species. *BMC Genom. Data* **2023**, *24*, 3. [[CrossRef](#)] [[PubMed](#)]
42. 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](#)]
43. Dhindsa, R.S.; Copeland, B.R.; Mustoe, A.M.; Goldstein, D.B. Natural selection shapes codon usage in the human genome. *Am. J. Hum. Genet.* **2020**, *107*, 83–95. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.