GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Patterns of synonymous codon usage bias in the model grass *Brachypodium distachyon*

H. Liu[1], Y. Huang[2], X. Du[1], Z. Chen[1], X. Zeng[1], Y. Chen[1] and H. Zhang[1]

[1]College of Life Sciences, Sichuan Agricultural University, Ya'an, Sichuan, China
[2]Maize Research Institute, Sichuan Agricultural University, Sichuan, China

Corresponding author: H. Zhang
E-mail: zhyu@sicau.edu.cn

Genet. Mol. Res. 11 (4): 4695-4706 (2012)
Received January 9, 2012
Accepted June 26, 2012
Published October 17, 2012
DOI http://dx.doi.org/10.4238/2012.October.17.3

ABSTRACT. *Brachypodium distachyon* has been proposed as a new model for the temperate grass because it is related to the major cereal grain species (such as wheat, barley, oat, maize, rice, and sorghum) and many forage and turf species. In this study, a multivariate statistical analysis was performed to investigate the characteristics of codon bias and the main factors affecting synonymous codon usage in *Brachypodium*. We found that low- and high-GC content genes with different codon usage occur frequently in the genome. The results of neutrality, correspondence, and correlation analyses indicated that mutational pressure and selective constraint were the main factors in shaping codon usage. Coding sequence length and the hydrophobicity of each protein were also identified as influences on codon usage bias, although their effect was minor. In addition, 27 codons, defined as "optimal codons", might provide useful information for gene engineering, gene prediction, and molecular evolution studies.

Key words: *Brachypodium distachyon*; Codon usage; Optimal codons; Correspondence analysis

INTRODUCTION

Brachypodium distachyon, a small, annual, self-pollinated grass with wide distribution in temperate areas, belongs to the same species of Pooideae as wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), oat (*Avena sativa*), and meadow, and forage grasses. *Brachypodium* has been considered a new model plant species for functional genomics research of important temperate cereals and forage grasses (Draper et al., 2001). It offers many characters as a model species, such as small genome, fewer chromosomes, reduced stature, short life cycle, and undemanding growth requirements (Draper et al., 2001). At the same time, the pathogens causing powdery mildew, stripe rust, and *Fusarium* head blight in Triticeae cereals and blast in rice can also infect *Brachypodium* and cause disease (Draper et al., 2001; Peraldi et al., 2011). These traits make *B. distachyon* a good model for researching and improving resistance to some crop diseases. Research on *Brachypodium* has made important advancements recently with respect to cytogenetics, molecular genetics, genomics, tissue culture, and gene transformation (International Brachypodium Initiative, 2010).

The genetic code uses 64 codons to encode the 20 standard amino acids and the translation termination signal. With the exception of Met and Trp, multiple codons are assigned for one amino acid. Synonymous codons (those encoding the same amino acid) are used in unequal frequencies between organisms, within genomes, and even within genes. The pattern of synonymous codon usage often varies among species, and it is an important feature of an organism. To date, the genomes of many organisms have been sequenced, so sufficient sequences are available for analyses of synonymous codon usage, and codon usage for many microorganisms, lower organisms, and higher plants and animals has been investigated. For the vast majority organisms, biased codon usage is primarily a result of mutational pressure or natural selection (Eyre-Walker, 1991; Stenico et al., 1994; Duret and Mouchiroud, 1999; Morton and Wright, 2007; Wang and Hickey, 2007; Jiang et al., 2008; Roychoudhury and Mukherjee, 2010). In plants, most studies on codon usage bias to date have focused on model plants and important food crops, such as *Arabidopsis thaliana*, *Oryza sativa*, *Zea mays*, and *T. aestivum* (Liu et al., 2004; Morton and Wright, 2007; Wang and Hickey, 2007; Zhang et al., 2007; Liu et al., 2010). For these organisms, the features of codon choice show some diversity, but the main factors influencing codon usage bias are base compositional mutation bias and selection to increase translation efficiency. In other words, the codon usage of highly expressed genes is shifted toward a more restricted set of major synonymous codons rather than to other less highly expressed genes, and the highly expressed genes display more significant variations in codon usage. Obviously, the study of codon usage bias provides a guide for selecting appropriate host expression systems and optimizing codon usage to improve the expression of exogenous genes.

As a model plant in functional genomics study, *Brachypodium* is usually used to host expression systems for investigating the function of heterologous genes. Therefore, the codon usage profile of *Brachypodium* is important and meaningful. Recently, a large number of genes have been identified from the genome sequence of *Brachypodium*, and they provide a convenient means for examining codon bias (International Brachypodium Initiative, 2010). In this study, we comprehensively analyzed the codon usage patterns of *Brachypodium* through multivariate statistical analysis, explored the key factors in shaping codon choice, and determined “optimal codons”.

MATERIAL AND METHODS

Retrieval of sequences

Publicly available *Brachypodium* complementary DNA sequences (32,255 in all) that contain complete coding sequences (CDS) were obtained from genome annotation data in a *B. distachyon* database (http://mips.helmholtz-muenchen.de/plant/brachypodium/download/index.jsp). All CDSs were extracted by writing a C program. A check of CDS applicability was performed based on the following criteria: 1) to minimize sampling error, each CDS was ≥100 codons long; 2) each CDS had the proper initial codon at the beginning and termination codon at the end and lacked an intermediate stop codon; 3) exact duplicated sequences were detected and excluded from the dataset (Zhang et al., 2007; Liu et al., 2010). The final dataset contained 24,439 CDSs.

Measures of codon usage bias

Relative synonymous codon usage (RSCU) is defined as the ratio of the observed frequency of codons to the expected frequency if all the synonymous codons for those amino acids are used equally. Two single codons for methionine (ATG) and tryptophan (TGG) and 3 stop codons (TAA, TAG, TGA) were excluded from the calculation. So the index of RSCU normalized the dataset of various amino acid compositions. The effective number of codons (ENC) is often used to measure the magnitude of codon bias for an individual gene, which is essentially independent of gene length (Wright, 1990). The values of ENC are always between 20 (for a gene with extreme bias using only one codon per amino acid) and 61 (for a gene with no bias using synonymous codons equally). In general, if the ENC value of a gene is 35 or less, that gene is considered to have a strong codon bias, whereas an ENC value of 50 or higher is generally considered low codon bias (Roychoudhury and Mukherjee, 2010). ENC values were measured for each gene in this study. It has been reported that the ENC is correlated with the nucleotide composition of genes. To study these aspects in *Brachypodium*, we calculated the frequency of G + C for all the codons in the sequence dataset (GC codon) and the first, second, and third codon positions (GC1, GC2 and GC3, respectively). GC12 was the average of GC1 and GC2, and it was used for neutrality plot analysis. The GC3s value is the frequency of G + C at the third synonymously variable coding position (excluding Met, Trp, and termination codons). To examine the influence of GC content on codon usage, the relationship of ENC and GC3s content of each gene was plotted according to the equation described by Wright (1990). The codon adaptation index (CAI) is used to predict the level of gene expression and assess the extent to which selection has been effective in molding the pattern of codon usage (Sharp and Li, 1987; Naya et al., 2001; Gupta et al., 2004). In this study, the CAI value for every gene was calculated, and the genes coding ribosomal proteins were used as reference sequence (Liu et al., 2004; Liu et al., 2010).

Correspondence analysis

Correspondence analysis is widely used to investigate major trends in the variation of codon usage among genes. To understand the codon usage variation of genes in *Brachypodium*, we used the RSCU values of genes for correspondence analysis to minimize the affect

of amino acid composition. All genes were plotted in a 59-dimensional hyperspace according to their usage of the 59 sense codons. Genes in which given codons were used in a similar fashion were close to each other on the graph. CodonW 1.4 (http://bioweb.pasteur.fr/seqanal/interfaces/codonw.html) was used to calculate the indices of codon usage and performing the correspondence analysis.

Determination of optimal codons

To define optimal codons, we used a chi-square test to examine the significance of codon usage difference between 2 datasets. The 2 groups of datasets, which were regarded as the high- and low-bias gene datasets, were 5% of the total genes located at the extreme right and left of axis 1 produced by correspondence analysis on RSCU, respectively (Liu and Xue, 2005). Codons with a frequency of usage that was significantly higher ($P < 0.01$) in high-bias genes than that in genes with low bias were defined as the optimal codons (Liu et al., 2010). SPSS 12.0 was implemented for statistical analysis.

RESULTS

Nucleotide compositional constraint analysis

To investigate the nucleotide composition of *Brachypodium* genes, we computed the GC content of the entire concatenated sequence and each gene (Kawabe and Miyashita, 2003). We concatenated all genes to one sequence, which comprised 10,908,249 codons. The GC content in 3 codon positions (GC1, GC2, and GC3) was 0.575, 0.445, and 0.612, respectively. This analysis showed that the GC content at these positions was significantly different. GC3 was higher than GC1 and GC2, and GC2 was the lowest of all 3 codon positions. The average GC content of all codons (GC_{codon}) was 0.544.

In addition, we calculated the GC content of each gene and drafted a distribution figure (Figure 1). The GC content value for the 24,439 genes ranged from 31.4 to 81.1%, with a mean value of 0.5654 and a standard deviation of 0.0909. As shown in Figure 1, a wide, distinctly bimodal distribution of GC content was present among genes, and the vertical dividing line was positioned at 60% GC content. All of the genes in the *Brachypodium* genome were divided into 2 classes: one class included low-GC genes (GC content <60%), and the other included high-GC genes (GC content ≥60%). The number of low-GC genes was greater than that of high-GC genes. The results were consistent with previous reports for rice and maize (Carels and Bernardi, 2000; Wang and Hickey, 2007; Liu et al., 2010).

To examine the relationships among the 3 positions, the indices of GC1, GC2, GC3, and GC12 were computed for each gene, and neutrality plots (GC12 *vs* GC3) were constructed (Figure 2) (Sueoka, 1988). The results showed that genes in *Brachypodium* had a wide range of GC3 (0.228-1.000), and the difference in GC3 usually reflected the neutral mutation bias, leading to different codon choice in each gene. Simultaneously, a significantly positive correlation in neutrality plots was found, indicating that the intragenomic GC mutation bias affected the GC contents similarly in all codon positions. We preliminarily inferred that the neutral mutation bias influenced codon usage.

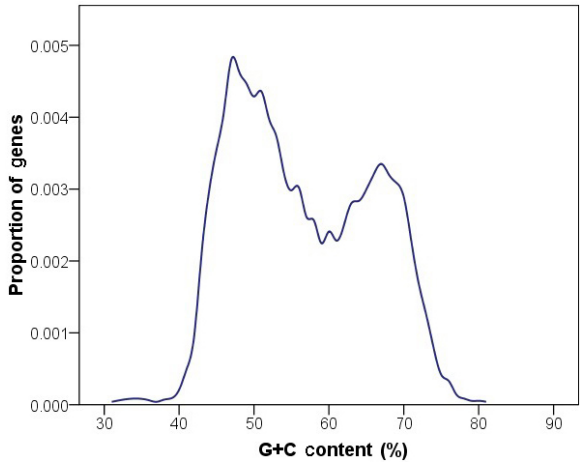

Figure 1. Distribution of genes with different GC contents in *Brachypodium*.

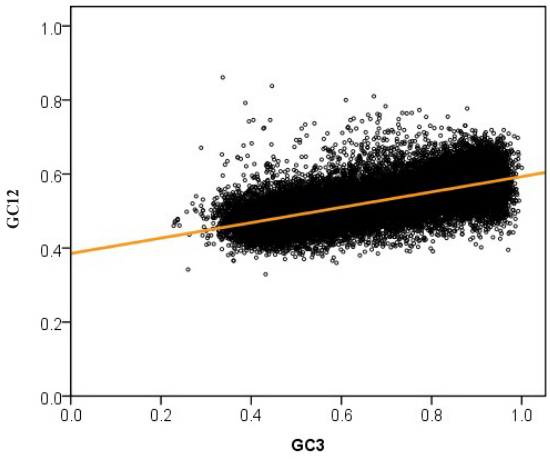

Figure 2. Neutrality plot (GC12 against GC3). The regression line is y = 0.208x + 0.385; R^2 = 0.413; OP = 0.486.

Relationship between ENC and GC3s values

Computing the ENC and GC3s values of each gene revealed that these values varied widely among genes. The ENC values ranged from 23.26 to 61, and the GC3s values ranged from 0.197 to 1. Genes with low ENC values (≤35) and strong bias totaled 3099, and 10,695 genes had high ENC values (50-61) with weak bias. The percentage of strong- and weak-bias genes was 12.68 and 43.76%, respectively. We also calculated the correlation coefficient between ENC and GC3s values. The results showed that the ENC value was strongly negatively correlated with the GC3s values of each gene ($r = -0.873$, $P < 0.0001$). These calculations suggested that genes with higher GC3s values and lower ENC values had strong bias. The results also showed that nucleotide composition mutation bias shaped codon usage.

To further investigate codon usage variation among genes with different GC content, we plotted ENC against the GC3s values of each gene (Figure 3) (Wright, 1990). The solid line (shown in blue) is the expected position of genes in which codon usage is determined only by GC3s values. A few genes were located on the reference line, which indicates that the GC3s value was the only factor determining the codon usage pattern (see Figure 3). However, most genes with ENC values lower than those expected lay well below the reference line, indicating that GC3s value was a major determinant of codon usage and that other factors independent of nucleotide composition shaped codon usage as well.

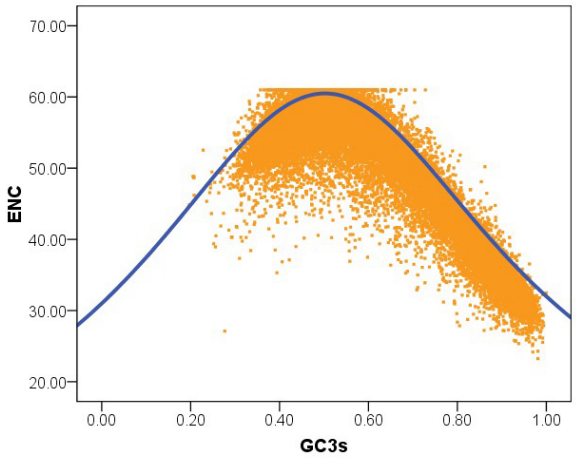

Figure 3. Distribution of effective number of codons (ENC) and GC3s of *Brachypodium* genes. The solid line (shown in blue) indicates the expected ENC value if the codon bias is only due to GC3s.

Correspondence analysis

Based on the RSCU values, the result of correspondence analysis revealed the major trends in codon usage of *Brachypodium*. The first and second axes accounted for 46.65 and 4.16% of the overall variation in the dataset, respectively. The third and remaining axes were each responsible for even smaller amounts of variation (2.76-0.6%), indicating that the first axis was the major explanatory axis for interpreting codon usage variation among genes. The gene distribution on the first and second axes is shown in Figure 4A. In this study, high-GC (≥60%) and low-GC (<60%) genes are plotted in orange and blue, respectively. Of particular interest was that the high- and low-GC content of genes separated along the first axis (see Figure 4A). Genes with low-GC content were located to the left, and genes with high-GC content were located to the right. Further calculations revealed a significant correlation ($r = 0.954$, $P < 0.0001$) between the GC content of individual genes and their positions on the first axis (Figure 4B). In addition, the gene positions on axis 1 were strongly correlated with the GC3s value ($r = 0.994$, $P < 0.0001$) and significantly negatively correlated with ENC ($r = -0.875$, $P < 0.0001$). The distribution of synonymous codons on the plane defined by the first 2 axes is displayed in Figure 4C. Distinguishing the G/C-ending codons from A/T-ending codons on the first axis was easy, as was distinguishing pyrimidines (C and T) from purines (A and G) at

codon ends on the second axis. The distribution of GC content along the first axis from low to high proved that the major factor influencing the difference in synonymous codon usage among *Brachypodium* genes was the nucleotide composition of CDSs.

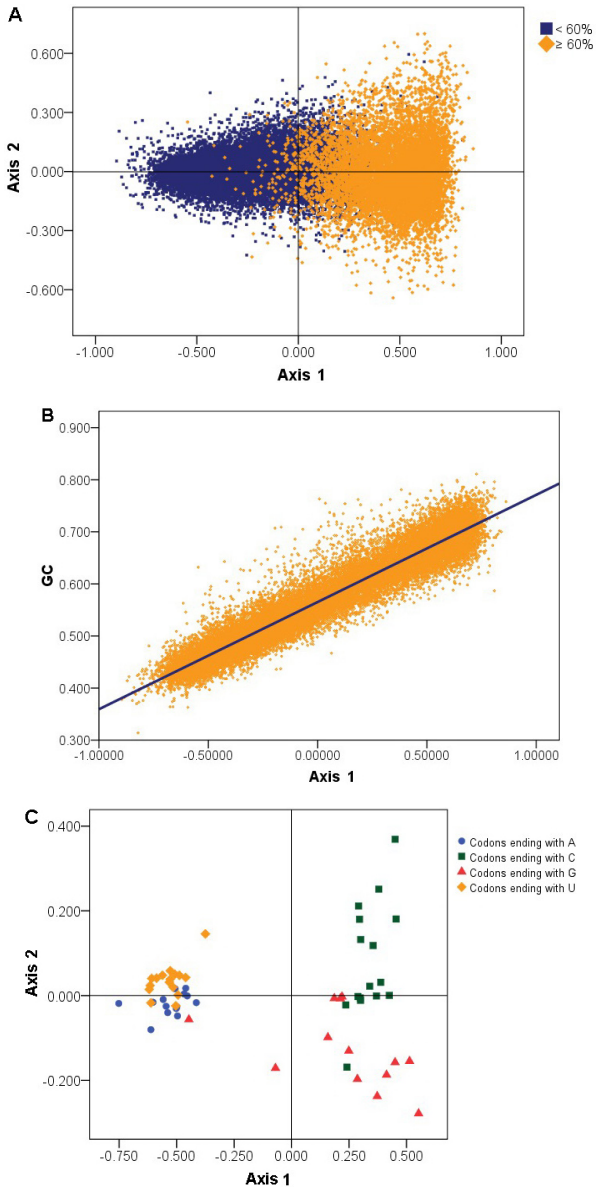

Figure 4. Correspondence analysis of relative synonymous codon usage for *Brachypodium* genes. **A.** The distribution of genes is shown along the first and second axes. Orange and blue dots indicate genes with GC content more than or equal to 60%, and less than 60%, respectively. **B.** Correlation between GC content of each gene and their position on the first axis. **C.** The distribution of codons on the same two axes as shown in Panel A. Codons ending with A, T, C, and G are shown in blue, orange, green, and red, respectively.

Gene expression level and synonymous codon usage bias

The expression level of each gene was assessed through CAI values (Naya et al., 2001; Gupta et al., 2004), which varied from 0.046 to 0.874 with a mean value of 0.3008 and a standard deviation of 0.1817. A significantly positive correlation ($r = 0.864$ and 0.815, respectively, $P < 0.0001$) was found between CAI and GC3s values and GC content. In addition, the correlation analysis showed significant correlation ($r = 0.863$ and -0.867, respectively, $P < 0.0001$) between CAI and axis 1 positions and ENC, although the values of the correlation coefficients were slightly lower than that of gene nucleotide composition. The results indicated that gene expression level shaped codon usage in *Brachypodium*, and the genes with higher expression level had a greater degree of codon usage bias and richer GC.

Relationship between CDS length, hydrophobicity of encoded protein and synonymous codon usage bias

CDS length and the hydrophobicity of the encoded proteins have been demonstrated to be correlated with codon usage in some species (Duret and Mouchiroud, 1999; Liu et al., 2004; Liu, 2006; Zhao et al., 2007). To determine whether the same relationship holds true for *Brachypodium* genes, we implemented a bivariate correlate analysis. The results showed that the 4 correlation coefficients ($r = -0.341$, -0.342, 0.253, and -0.25, respectively, $P < 0.01$) were significant between CDS length and the gene positions on axis 1, GC3s value, ENC value, and CAI. This result indicated that more biased genes with shorter CDS length and higher GC3s value and expression level were located at the right side of the first axis, and the opposite was true for longer genes. Furthermore, the correlation coefficients between the hydrophobicity of encoded proteins and the gene positions on axis 1, GC3s value, ENC value, and CAI were computed; the 4 values ($r = 0.204$, 0.201, -0.146, and 0.166, respectively, $P < 0.01$) showed significance. The findings indicated that genes encoding more hydrophobic proteins had stronger codon biases and higher GC3s values and expression levels. Therefore, we deduced that CDS length and the hydrophobicity of encoded proteins affect codon bias, but the absolute values of these correlation coefficients are low and far smaller than that of nucleotide composition and gene expression level. The results implied that nucleotide composition and gene expression level were the major source of codon usage variation, whereas CDS length and the hydrophobicity of the encoded proteins played a minor role in affecting codon usage in *Brachypodium*.

Optimal codons

The first axis was the primary explanatory axis, and was significantly correlated to GC3s, ENC, and CAI values. We selected 5% of the total genes located at the extreme right and left of axis 1, and regarded them as the high- and low-bias gene datasets. The average RSCU values of the high- and low-bias datasets are listed in Table 1. Twenty-seven codons were determined to be the optimal codons, which were significantly more frequent among the high-bias genes ($P < 0.01$) according to the chi-square test. All optimal codons ended with G or C, indicating that codon usage in *Brachypodium* was biased to G- or C-ending synonymous codons.

Table 1. Optimal codons of *Brachypodium* genes based on their relative synonymous codon usage (RSCU) values.

Amino acid	Codon	High		Low		Amino acid	Codon	High		Low	
		RSCU	Count	RSCU	Count			RSCU	Count	RSCU	Count
Phe	UUU	0.05	346	1.23	19,162	Ser	UCU	0.18	838	1.6	22,917
	UUC*	1.95	14,515	0.77	11,934		UCC*	2.48	11,553	0.64	9091
Leu	UUA	0.01	71	0.82	11,107		UCA	0.11	531	1.53	21,848
	UUG	0.15	901	1.38	18,630		UCG*	1.53	7112	0.34	4920
	CUU	0.13	765	1.55	20,975		AGU	0.04	205	1.12	16,036
	CUC*	3.31	20,256	0.61	8281		AGC*	1.65	7666	0.76	10,904
	CUA	0.06	375	0.69	9394	Pro	CCU	0.21	1177	1.57	17,619
	CUG*	2.34	14,316	0.94	12,723		CCC*	1.43	8212	0.47	5304
Ile	AUU	0.09	342	1.41	20,275		CCA	0.17	957	1.61	18,105
	AUC*	2.83	11,207	0.69	9994		CCG*	2.2	12,590	0.34	3823
	AUA	0.08	325	0.9	12,908	Thr	ACU	0.08	366	1.48	16,116
Val	GUU	0.08	627	1.67	23,371		ACC*	1.8	7914	0.65	7089
	GUC*	1.71	12,667	0.62	8719		ACA	0.1	454	1.56	17,084
	GUA	0.04	295	0.75	10,432		ACG*	2.01	8836	0.31	3381
	GUG*	2.17	16,121	0.96	13,351	Ala	GCU	0.14	1740	1.65	25,725
Tyr	UAU	0.05	235	1.29	14,121		GCC*	2.03	25,555	0.58	9103
	UAC*	1.95	9310	0.71	7839		GCA	0.13	1666	1.47	23,043
His	CAU	0.13	571	1.43	16,317		GCG*	1.7	21,305	0.3	4659
	CAC*	1.87	8183	0.57	6485	Cys	UGU	0.04	141	1.13	9191
Gln	CAA	0.12	633	1.04	19,380		UGC*	1.96	6693	0.87	7082
	CAG*	1.88	9700	0.96	17,753	Arg	CGU	0.14	657	0.86	6936
Asn	AAU	0.1	507	1.31	25,488		CGC*	2.76	12,498	0.48	3889
	AAC*	1.9	9212	0.69	13,352		CGA	0.1	432	0.65	5230
Lys	AAA	0.09	644	0.97	25,718		CGG*	1.94	8806	0.54	4342
	AAG*	1.91	13,129	1.03	27,422		AGA	0.08	376	1.94	15,571
Asp	GAU	0.13	1303	1.43	37,778		AGG	0.98	4436	1.52	12,155
	GAC*	1.87	18,449	0.57	14,962	Gly	GGU	0.11	905	1.34	18,397
Glu	GAA	0.16	1545	1.12	34,300		GGC*	2.54	21,352	0.72	9845
	GAG*	1.84	18,082	0.88	27,017		GGA	0.19	1585	1.26	17,342
							GGG*	1.16	9727	0.68	9411

Codon usage was compared using the chi-squared contingency test to identify optimal codons. Values followed by asterisks are significantly more often ($P < 0.01$).

DISCUSSION

Previous studies have shown that many factors - for instance, base composition, expression level, structure of encoded protein - affect codon usage bias (Chiapello et al., 1998; Liu et al., 2005; Mukhopadhyay et al., 2007a). Many hypotheses have been put forward to explain the origin of codon usage bias. Among them, the neutral theory and the selection-mutation-drift theory have been widely supported (Bulmer, 1991; Hershberg and Petrov, 2008). According to neutral theory, mutation occurs at degenerate coding position neutrally, which leads to random synonymous codon choice; GC and AT are used proportionally in the degenerate codon groups in genes. In the selection-mutation-drift model, codon usage patterns result from the balance in a finite population between selection favoring an optimal codon for each amino acid and mutation together with drift that allows the persistence of nonoptimal codons. So biased codon usage is primarily the result of the directional mutation pressure on DNA sequences or natural selection affecting gene translation (Bulmer, 1988; Sueoka and Kawanishi, 2000; Mitreva et al., 2006). The study of some bacteria, such as *Bacillus subtilis*, has shown that codon usage is attributable to the equilibrium between natural selection and base compositional mutation bias (Shields and Sharp, 1987; Sharp et al., 1993), and this phenomenon holds true for some plants (Morton and Wright, 2007; Liu et al., 2004, 2010). The results of recent research on viruses such as baculovi-

ruses and herpesviruses have shown that mutation bias is the main factor influencing codon usage bias (Jiang et al., 2008; Roychoudhury and Mukherjee, 2010). Consistent results have been found in mammals - for example, codon usage in humans appears to be correlated with the GC content of genes and is strongly influenced by mutation pressure (Eyre-Walker, 1991). However, for some eukaryotes, such as *Drosophila melanogaster* and *Caenorhabditis elegans*, codon usage bias is mainly caused by translational selection (Shields et al., 1988; Stenico et al., 1994). So the factors shaping codon usage are unrelated to biological evolution.

In this study, the preliminary analysis of base composition revealed that the GC3 content of the concatenated sequence was the highest and many genes with rich GC were present in the genome. These findings indicated that the codons ending with G or C were favored, and the content of G and C in various codon positions may reflect important characteristics of codon usage pattern in *Brachypodium*. Therefore, further analysis of the association between mutation pressure and codon bias was carried out. First, a neutrality plot revealed the relationship between GC12 and GC3, which may be helpful in examining the mechanism through which mutation-selection equilibrium shapes codon usage. The regression coefficient (slope) provided a measure of the relative neutrality of GC12 to GC3. If GC12 and GC3 were equally neutral, the points representing genes were distributed along the diagonal line (slope of unity). On the contrary, if GC12 was completely non-neutral, the points were distributed on the parallel lines of the abscissa (slope of zero) (Sueoka, 1988). In our analysis, GC12 was compared with GC3 and significant correlation was observed; the regression coefficient was 0.208, suggesting that both neutral mutation and selective constraint play important roles in shaping codon usage.

Second, in an ENC plot, the distance between genes and the expected line indicated whether codon usage was influenced by composition constraint or other factors (Wright, 1990). Our results showed that a few genes were located on the reference line, and most genes gathered together below the expected curve of the ENC versus the GC3s plot. So composition constraint affected codon usage to a large extent. Finally, a significant correlation among GC, GC3s values and the ENC, and the first axis positions were found, indicating that the variations in synonymous codon usage among genes were based on the nucleotide content within them (Liu et al., 2004; Liu et al., 2010). Therefore, the results suggested that mutational pressure was the main factor that determined codon usage bias in *Brachypodium*.

From the neutrality plot, we found that selective constraint shaped codon usage. To examine the effect of selective constraint, a correlation analysis was performed among CAI and GC content, GC3s value, ENC value, and first axis position. The results showed that translational selection was the other main factor affecting codon usage patterns, except for base composition. Highly expressed genes tended to use optimal codons to increase their translational accuracy and efficiency in *Brachypodium*. In addition, codon usage was affected by gene length and the hydrophobicity of encoded proteins, but the extent of the effect was far less than that of base composition and translational selection.

The grass family (Poaceae) comprises several subfamilies - including Pooideae, Ehrhartoideae, and Panicoideae - and the important food crops, wheat, rice, and maize belong to the 3 subfamilies, respectively (International Brachypodium Initiative, 2010). The Pooideae subfamily is the largest grass subfamily, involving most cool season cereal, forage, and turf grasses. *Brachypodium* is one of the members of the Pooideae subfamily, which during the evolution of the Pooideae, diverged just before the clade of the "core pooid" genera that contain the majority of important temperate cereals and forage grasses (Doust, 2007). Owing

to the close genetic relationships of these species, research using *Brachypodium* as a model system is essential for understanding and improving crops and grasses (Draper et al., 2001). Through surveys of codon usage among *Brachypodium* genes, we know the patterns and factors shaping codon usage. Similar codon usage patterns have been observed in rice, maize, and other monocot species (Liu et al., 2004; Zhang et al., 2007; Liu et al., 2010), but clear differences occur in dicot species (Murray et al., 1989; De Amicis and Marchetti, 2000). When *Brachypodium* is the host expression system, codon optimization may be unnecessary for expression of the exogenous genes originating in Poaceae. However, when the exogenous genes are from dicot species or other species with more distant genetic relationships, the codon optimization is helpful for raising the expression.

Two main definitions of optimal codon are used. One defines optimal codons as those with frequencies that increase with gene expression (Duret and Mouchiroud, 1999). The other defines optimal codons as those showing a statistically significant increase in frequency between genes with low- and high-codon usage bias (Stenico et al., 1994). To determine the optimal codon, the low and high bias or expression datasets of genes are assessed through multiple methods, such as CAI, ENC, expressed sequence tag counts, and transfer RNA gene copy number (Kawabe and Miyashita, 2003; Liu et al., 2004; Mukhopadhyay et al., 2007b). Our study evaluated one major trend in the first axis, which can account for 46.65% of the total variation and explain approximately 10 times as much of the variation as that explained by the second and subsequent axes. Moreover, the positions of the genes on axis 1 were significantly correlated with some independent measures of codon bias, such as ENC, CAI, and GC3s values. So the genes with the extreme values at 2 ends of the first axis were defined as the high- or low-bias datasets. The optimal codons determined in this study could facilitate research on the function genome and molecular evolution in the Pooideae subfamily.

ACKNOWLEDGMENTS

Research supported by the National Natural Science Foundation of China (#31171557), the Key Project of the Science and Technology Committee of Sichuan Province, China (#2008JY0097), and the Foundation of Sichuan Agricultural University.

REFERENCES

Bulmer M (1988). Are codon usage patterns in unicellular organisms determined by selection-mutation balance? *J. Mol. Biol.* 1: 15-26.

Bulmer M (1991). The selection-mutation-drift theory of synonymous codon usage. *Genetics* 129: 897-907.

Carels N and Bernardi G (2000). Two classes of genes in plants. *Genetics* 154: 1819-1825.

Chiapello H, Lisacek F, Caboche M and Henaut A (1998). Codon usage and gene function are related in sequences of *Arabidopsis thaliana*. *Gene* 209: GC1-GC38.

De Amicis F and Marchetti S (2000). Intercodon dinucleotides affect codon choice in plant genes. *Nucleic Acids Res.* 28: 3339-3345.

Doust A (2007). Architectural evolution and its implications for domestication in grasses. *Ann. Bot.* 100: 941-950.

Draper J, Mur LA, Jenkins G, Ghosh-Biswas GC, et al. (2001). *Brachypodium distachyon*. A new model system for functional genomics in grasses. *Plant Physiol.* 127: 1539-1555.

Duret L and Mouchiroud D (1999). Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 96: 4482-4487.

Eyre-Walker AC (1991). An analysis of codon usage in mammals: selection or mutation bias? *J. Mol. Evol.* 33: 442-449.

Gupta SK, Bhattacharyya TK and Ghosh TC (2004). Synonymous codon usage in *Lactococcus lactis*: mutational bias

versus translational selection. *J. Biomol. Struct. Dyn.* 21: 527-536.
Hershberg R and Petrov DA (2008). Selection on codon bias. *Annu. Rev. Genet.* 42: 287-299.
International Brachypodium Initiative (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463: 763-768.
Jiang Y, Deng F, Wang H and Hu Z (2008). An extensive analysis on the global codon usage pattern of baculoviruses. *Arch. Virol.* 153: 2273-2282.
Kawabe A and Miyashita NT (2003). Patterns of codon usage bias in three dicot and four monocot plant species. *Genes Genet. Syst.* 78: 343-352.
Liu H, He R, Zhang H, Huang Y, et al. (2010). Analysis of synonymous codon usage in *Zea mays*. *Mol. Biol. Rep.* 37: 677-684.
Liu Q (2006). Analysis of codon usage pattern in the radioresistant bacterium *Deinococcus radiodurans*. *Biosystems* 85: 99-106.
Liu Q and Xue Q (2005). Comparative studies on codon usage pattern of chloroplasts and their host nuclear genes in four plant species. *J. Genet.* 84: 55-62.
Liu Q, Feng Y, Zhao X, Dong H, et al. (2004). Synonymous codon usage bias in *Oryza sativa*. *Plant Sci.* 167: 101-105.
Liu Q, Dou S, Ji Z and Xue Q (2005). Synonymous codon usage and gene function are strongly related in *Oryza sativa*. *Biosystems* 80: 123-131.
Mitreva M, Wendl MC, Martin J, Wylie T, et al. (2006). Codon usage patterns in Nematoda: analysis based on over 25 million codons in thirty-two species. *Genome Biol.* 7: R75.
Morton BR and Wright SI (2007). Selective constraints on codon usage of nuclear genes from *Arabidopsis thaliana*. *Mol. Biol. Evol.* 24: 122-129.
Mukhopadhyay P, Basak S and Ghosh TC (2007a). Synonymous codon usage in different protein secondary structural classes of human genes: implication for increased non-randomness of GC3 rich genes towards protein stability. *J. Biosci.* 32: 947-963.
Mukhopadhyay P, Basak S and Ghosh TC (2007b). Nature of selective constraints on synonymous codon usage of rice differs in GC-poor and GC-rich genes. *Gene* 400: 71-81.
Murray EE, Lotzer J and Eberle M (1989). Codon usage in plant genes. *Nucleic Acids Res.* 17: 477-498.
Naya H, Romero H, Carels N, Zavala A, et al. (2001). Translational selection shapes codon usage in the GC-rich genome of *Chlamydomonas reinhardtii*. *FEBS Lett.* 501: 127-130.
Peraldi A, Beccari G, Steed A and Nicholson P (2011). *Brachypodium distachyon*: a new pathosystem to study *Fusarium* head blight and other *Fusarium* diseases of wheat. *BMC Plant Biol.* 11: 100.
Roychoudhury S and Mukherjee D (2010). A detailed comparative analysis on the overall codon usage pattern in herpesviruses. *Virus Res.* 148: 31-43.
Sharp PM and Li WH (1987). The codon Adaptation Index - a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15: 1281-1295.
Sharp PM, Stenico M, Peden JF and Lloyd AT (1993). Codon usage: mutational bias, translational selection, or both? *Biochem. Soc. Trans.* 21: 835-841.
Shields DC and Sharp PM (1987). Synonymous codon usage in *Bacillus subtilis* reflects both translational selection and mutational biases. *Nucleic Acids Res.* 15: 8023-8040.
Shields DC, Sharp PM, Higgins DG and Wright F (1988). "Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. *Mol. Biol. Evol.* 5: 704-716.
Stenico M, Lloyd AT and Sharp PM (1994). Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases. *Nucleic Acids Res.* 22: 2437-2446.
Sueoka N (1988). Directional mutation pressure and neutral molecular evolution. *Proc. Natl. Acad. Sci. U. S. A.* 85: 2653-2657.
Sueoka N and Kawanishi Y (2000). DNA G+C content of the third codon position and codon usage biases of human genes. *Gene* 261: 53-62.
Wang HC and Hickey DA (2007). Rapid divergence of codon usage patterns within the rice genome. *BMC Evol. Biol.* 7: S6.
Wright F (1990). The 'effective number of codons' used in a gene. *Gene* 87: 23-29.
Zhang WJ, Zhou J, Li ZF, Wang L, et al. (2007). Comparative analysis of codon usage patterns among mitochondrion, chloroplast and nuclear genes in *Triticum aestivum* L. *J. Integr. Plant Biol.* 49: 246-254.
Zhao S, Zhang Q, Chen Z, Zhao Y, et al. (2007). The factors shaping synonymous codon usage in the genome of *Burkholderia mallei*. *J. Genet. Genomics* 34: 362-372.