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

Codon usage bias analysis for the spermidine synthase gene from *Camellia sinensis* (L.) O. Kuntze

E. You, Y. Wang, Z.T. Ding, X.F. Zhang, L.L. Pan and C. Zheng

Tea Research Institute, Qingdao Agricultural University, Qingdao, Shandong, China

Corresponding author: Z.T. Ding
E-mail: dzttea@163.com

Genet. Mol. Res. 14 (3): 7368-7376 (2015)
Received August 8, 2014
Accepted January 29, 2015
Published July 3, 2015
DOI http://dx.doi.org/10.4238/2015.July.3.12

ABSTRACT. The spermidine synthase (*SPDS*) gene exists widely in all types of plants. In this paper, the codon usage of the *SPDS* gene from *Camellia sinensis* (*CsSPDS*) was analyzed. The results showed that the codon usage of the *CsSPDS* gene is biased towards the T-ended or A-ended codons, which is similar to that observed in 73 genes selected from the *C. sinensis* genome. An ENC-plot for 15 *SPDS* genes from various plant species suggested that mutational bias was the major factor in shaping codon usage in these genes. Codon usage frequency analysis indicated that there was little difference between the *CsSPDS* gene and dicot genomes, such as *Arabidopsis thaliana* and *Nicotiana tabacum*, but significant differences in codon usage were observed between the *CsSPDS* gene and monocot genomes, such as *Triticum aestivum* and *Zea mays*. Therefore, *A. thaliana* and *N. tabacum* expression systems may be more suitable for the expression of the *CsSPDS* gene.

Key words: *Camellia sinensis* (L.) O. Kuntze; Codon usage bias; *SPDS*; ENC-plot; Correlation analysis; Cluster analysis

INTRODUCTION

Following a long period of evolution, codons are used in a species-specific manner, a phenomenon known as codon usage bias. In recent years, an ever-increasing body of studies on codon usage bias has been reported for plant breeding (Kawabe and Miyashita, 2003; Ravi et al., 2008; Lei et al., 2013). Liu and Xue (2005) reported that the chloroplast genome might display particular characteristics of codon usage that are different from its host nuclear genome. Ruhfel et al. (2014) put forward that their analyses of the plastid sequence data recovered a strongly supported framework of relationships for green plants. Several articles have also reported the codon usage pattern, and the factors that shape codon usage, for *Zea mays* (Liu et al., 2010), *Silene latifolia* (Qiu et al., 2011), seven different citrus species (Xu et al., 2013), and the Asteraceae family (Nie et al., 2013). However, the exact codon usage characteristics for single genes of higher plants have not been well explored to date.

Camellia sinensis (L.) O. Kuntze is a perennial evergreen woody crop distributed from tropical to temperate regions. In recent years, research into the cold resistance of *C. sinensis* has become a key area of interest (Wang et al., 2012, 2013). Other studies have shown that polyamines (PAs) play fundamental roles in self-defense against cold stress (Sagor et al., 2013), while the spermidine synthase gene (*SPDS*) is known to be a key gene for the synthesis of PAs (Alcázar et al., 2010). To our knowledge, codon usage of the *SPDS* gene in *C. sinensis* has not been investigated in any detail. In this paper, we studied the codon usage bias of the *CsSPDS* gene and compared it to the codon usage of genes and genomes for other species. The codon usage bias analysis was performed to assist in the development of the most suitable plant expression system for *CsSPDS*.

MATERIAL AND METHODS

Sequence data

The *CsSPDS* gene (GenBank accession No. KF306297) used in this study was cloned by our laboratory. The *SPDS* sequences of another 14 plants (Table 1) were obtained from GenBank (http://www.ncbi.nlm.nih.gov). Codon usage frequencies of the genomes of 6 species were obtained from the Codon Usage Database (http://www.kazusa.or.jp/codon/). A total of 73 publicly available *C. sinensis* cDNA sequences containing complete coding sequence were screened from the GenBank nucleotide database.

Indices of codon usage bias

In order to investigate the characteristics of synonymous codon usage of different amino acid compositions in samples, relative synonymous codon usage (RSCU) of 59 informative codons (excluding Met, Trp, and the three termination codons) was computed. The RSCU value was calculated by dividing the observed codon usage by the expected value when all codons for the same amino acid are used equally (Liu et al., 2004; Guo et al., 2012; Shi et al., 2012; Pan et al., 2013). RSCU values close to 1.0 indicate that codons for the same amino acid are used equally, and RSCU values >1.0 indicate a strong bias for the corresponding codons (Sharp and Li, 1986). The effective number of codons (ENC) can be used in a gene as a simple measure of codon bias, which is the best estimator of absolute synonymous codon

usage bias. The value of ENC ranges from 20 to 61, while the larger the codon preference extent in a gene, the smaller the ENC value (Wright, 1990). The frequency of GC3s (at the third synonymously variable coding position) was calculated for all of the informative codons.

Table 1. List of accession numbers for the 15 *SPDS* genes analyzed.

Species	Accession number
Camellia sinensis (L.) O. Kuntze	KF306297
Nicotiana tabacum L.	AF321139
Oryza sativa L.	EU714031
Malus domestica Borkh.	AB072916
Solanum lycopersicum L.	NM_001247564
Zea mays (L.) subsp. *mays*	NM_001112372
Arabidopsis thaliana (L.) Heynh.	AY087226
Pisum sativum L.	AF043109
Panax ginseng C.A. Mey	GQ229380
Coffea arabica L.	AB015599
Solanum tuberosum L.	AJ345003
Eleutherococcus senticosus (Rupr. & Maxim.) Maxim.	JQ365624
Populus tomentosa Carr.	JQ002667
Ammopiptanthus mongolicus (Maxim. ex Kom.) S.H. Cheng	DQ519362
Hyoscyamus niger L.	AB006691

Cluster analysis based on usage bias and gene sequences

In the cluster analysis, 15 *SPDS* genes were clustered according to their RSCU values using Ward's method based on squared Euclidean distance (Shi et al., 2013). The formula, which calculates the Euclidean distance coefficient (D*ab*) of codon usage bias between two genes A and B, is as follows:

$$Dab = \sqrt{\sum_{i=1}^{59} (RSCUai - RSCUbi)^2} \qquad \text{(Equation 1)}$$

where, *RSCUai* is the relative synonymous codon usage frequency of codon *i* in sequence *a*. Similarly, *RSCUbi* is the relative synonymous codon usage frequency of codon *i* in sequence *b*. The calculation was continued until all the sequences were included and formed a single cluster.

Analysis tools

The online CHIPS program (codon heterozygosity in a protein coding sequence) was used to determine ENC and CUSP was used to estimate codon usage frequency (create a codon usage table) in EMBOSS (http://emboss.bioinformatics.nl/). CodonW 1.4.4 (http://bioweb.pasteur.fr/seqanal/interfaces/codonw.html) was used to estimate GC, GC3s, and RSCU. The ENC-plot, the clustering tree for RSCU, and the correlation analysis were carried out using the multi-analysis SPSS v18.0 software, and the phylogenetic tree was conducted based on coding sequences (CDs) using the MEGA4 software.

RESULTS

Comparison of codon usage bias within the *C. sinensis* genome

We performed a comparison of the codon usage patterns for the *CsSPDS* gene and 73 other genes selected from the *C. sinensis* genome using RSCU values (Figure 1). There were 22 codons whose RSCU values were higher than 1.2 in the *CsSPDS* gene and 17 codons whose RSCU values were higher than 1.2 in the *C. sinensis* genome (Figure 1). Obviously, these are the preferred codons of the *CsSPDS* gene and the 73 *C. sinensis* genome genes. Furthermore, there were 17 codons ending in A/T in the 22 preferred codons of the *CsSPDS* gene, suggesting that the *CsSPDS* gene had a bias toward the synonymous codons with A and T at the third codon position. It was clear that the codon usage bias of the *CsSPDS* gene was similar to the other 73 selected genes from the *C. sinensis* genome. Moreover, the RSCU values for 5 codons were larger than 2.0, indicating that the *CsSPDS* gene has a strong preference for these codons during translation.

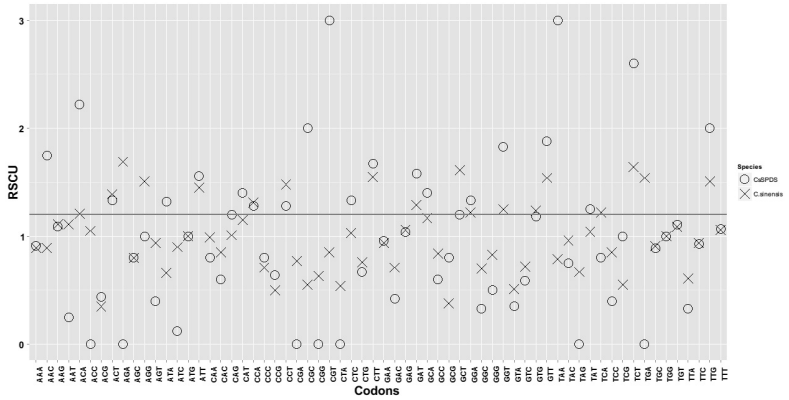

Figure 1. RSCU distribution of 59 codons in the *CsSPDS* and *Camellia sinensis* genes.

Comparisons of codon usage frequency with genomes of other species

Transgenic research often requires heterologous gene expression and selecting an appropriate expression system is very important. In this study, the frequency of codon usage by the *CsSPDS* gene was compared to that of the genomes for 6 other species: *Arabidopsis thaliana*, *Nicotiana tabacum*, *Triticum aestivum*, *Zea mays*, *Escherichia coli*, and yeast (Table 2). When compared with dicotyledons, such as *A. thaliana* and *N. tabacum*, we found that 14 ratios were greater than 2.0 and 17 ratios were less than 0.5. Meanwhile, comparisons with monocotyledons, such as *T. aestivum* and *Z. mays*, 31 ratios were greater than 2.0 and 26 ratios were less than 0.5. These results suggest that the codon preference of the *CsSPDS* gene is more similar to dicotyledons than to monocotyledons. Therefore, it can be speculated that the *CsSPDS* gene may express well in *A. thaliana* and *N. tabacum*. In addition, our results showed that the number of codons with a ratio >2 and a ratio <0.5 was 25 (*E. coli*) and 24 (yeast), respectively, indicating that the yeast expression system may be superior to the *E. coli* expression system for the *CsSPDS* gene.

Table 2. Comparison of codon usage preference between *CsSPDS* and the *SPDS* gene in other representative organisms.

Codon	Amino acid	*CsSPDS*	*A. thaliana*	*N. sylvestris*	*T. aestivum*	*Z. mays*	*E. coli*	Yeast	*SPDS/A*	*SPDS/N*	*SPDS/T*	*SPDS/Z*	*SPDS/E*	*SPDS/Y*
GCA	A	21.21	17.50	23.10	15.40	16.70	20.60	16.10	1.21	0.92	1.38	1.27	1.03	1.32
GCC	A	9.09	10.30	12.50	32.70	31.10	25.50	12.50	0.88	0.73	0.28	0.29	0.36	0.73
GCG	A	12.12	9.00	5.80	22.20	23.30	31.70	6.10	1.35	2.09	0.55	0.52	0.38	1.99
GCT	A	18.18	28.30	31.20	16.40	21.00	15.60	21.10	0.64	0.58	1.11	0.87	1.17	0.86
TGC	C	12.12	7.20	7.20	13.60	12.20	6.90	4.70	1.68	1.68	0.89	0.99	1.76	2.58
TGT	C	15.15	10.50	9.80	5.10	5.60	5.50	8.00	1.44	1.55	2.97	2.71	2.75	1.89
GAC	D	12.12	17.20	16.90	29.00	32.10	18.60	20.20	0.70	0.72	0.42	0.38	0.65	0.60
GAT	D	45.46	36.60	36.90	17.10	22.90	32.10	37.80	1.24	1.23	2.66	1.98	1.42	1.20
GAA	E	36.36	34.30	36.00	15.30	19.90	38.20	48.50	1.06	1.01	2.38	1.83	0.95	0.75
GAG	E	39.39	32.20	29.40	38.20	40.80	17.70	19.10	1.22	1.34	1.03	0.97	2.23	2.06
TTC	F	21.21	21.80	18.00	24.80	25.10	16.90	18.20	0.97	1.18	0.86	0.85	1.26	1.17
TTT	F	24.24	20.70	25.10	12.30	12.60	23.20	26.10	1.17	0.97	1.97	1.92	1.04	0.93
GGA	G	24.24	24.20	23.20	14.90	13.40	9.00	10.90	1.00	1.04	1.63	1.81	2.69	2.22
GGC	G	6.06	9.20	11.20	31.10	30.30	27.90	9.70	0.66	0.54	0.19	0.20	0.22	0.62
GGG	G	9.09	10.20	10.50	17.90	15.40	11.30	6.00	0.89	0.87	0.51	0.59	0.80	1.52
GGT	G	33.33	22.20	22.30	13.70	14.10	24.40	24.00	1.50	1.49	2.43	2.36	1.37	1.39
CAC	H	9.09	8.70	8.70	13.80	14.90	9.80	7.70	1.04	1.04	0.66	0.61	0.93	1.18
CAT	H	21.21	13.80	13.40	8.40	10.10	13.60	13.70	1.54	1.58	2.53	2.10	1.56	1.55
ATA	I	33.33	12.60	14.00	6.70	8.40	5.40	17.80	2.65	2.38	4.98	3.97	6.17	1.87
ATC	I	3.03	18.50	13.90	24.30	22.70	24.20	17.00	0.16	0.22	0.12	0.13	0.13	0.18
ATT	I	39.39	21.50	27.80	11.70	13.80	29.80	30.40	1.83	1.42	3.37	2.85	1.32	1.30
AAA	K	30.30	30.80	32.60	10.50	15.10	33.20	42.20	0.98	0.93	2.89	2.01	0.91	0.72
AAG	K	36.36	32.70	33.50	37.60	39.40	10.70	30.70	1.11	1.09	0.97	0.92	3.40	1.18
CTA	L	0.00	9.90	9.40	7.40	7.30	4.00	13.30	0.00	0.00	0.00	0.00	0.00	0.00
CTC	L	12.12	16.10	12.30	27.40	25.40	11.00	5.40	0.75	0.99	0.44	0.48	1.10	2.24
CTG	L	6.06	9.80	10.20	22.30	25.80	50.90	10.40	0.62	0.59	0.27	0.23	0.12	0.58
CTT	L	15.15	24.10	24.00	12.30	15.70	11.70	12.10	0.63	0.63	1.23	0.97	1.30	1.25
TTA	L	3.03	12.70	13.40	3.70	5.70	13.90	26.70	0.24	0.23	0.82	0.53	0.22	0.11
TTG	L	18.18	20.90	22.30	12.10	13.00	14.00	27.00	0.87	0.82	1.50	1.40	1.30	0.67
ATG	M	24.24	24.50	25.00	24.30	24.20	27.00	20.90	0.99	0.97	1.00	1.00	0.90	1.16
AAC	N	21.21	20.90	17.90	21.40	22.10	21.40	24.90	1.01	1.19	0.99	0.96	0.99	0.85
AAT	N	3.03	22.30	28.00	10.50	13.50	18.60	36.30	0.14	0.11	0.29	0.22	0.16	0.08
CCA	P	24.24	16.10	19.80	23.40	13.80	8.50	18.20	1.51	1.22	1.04	1.76	2.85	1.33
CCC	P	15.15	5.30	6.60	14.60	13.50	5.80	6.80	2.86	2.30	1.04	1.12	2.61	2.23
CCG	P	12.12	8.60	5.00	16.40	15.80	21.80	5.30	1.41	2.42	0.74	0.77	0.56	2.29
CCT	P	24.24	18.70	18.70	11.70	12.60	7.30	13.60	1.30	1.30	2.07	1.92	3.32	1.78
CAA	Q	12.12	19.40	20.70	42.20	13.20	15.00	27.50	0.62	0.59	0.29	0.92	0.81	0.44
CAG	Q	18.18	15.20	15.00	38.20	23.70	29.50	12.10	1.20	1.21	0.48	0.77	0.62	1.50
AGA	R	0.00	19.00	16.00	6.70	8.80	2.90	21.30	0.00	0.00	0.00	0.00	0.00	0.00
AGG	R	3.03	11.00	12.20	13.00	14.80	1.90	9.20	0.28	0.25	0.23	0.20	1.59	0.33
CGA	R	0.00	6.30	5.30	3.00	4.40	3.90	3.00	0.00	0.00	0.00	0.00	0.00	0.00
CGC	R	6.06	3.80	3.90	12.80	14.30	21.00	2.60	1.60	1.55	0.47	0.42	0.29	2.33
CGG	R	0.00	4.90	3.70	8.90	9.50	6.30	1.70	0.00	0.00	0.00	0.00	0.00	0.00
CGT	R	9.09	9.00	7.50	5.60	6.00	20.30	6.50	1.01	1.21	1.62	1.52	0.45	1.40
AGC	S	12.12	11.30	10.00	16.60	16.40	16.00	9.70	1.07	1.21	0.73	0.74	0.76	1.25
AGT	S	6.06	14.00	13.30	6.60	7.80	9.50	14.20	0.43	0.46	0.92	0.78	0.64	0.43
TCA	S	12.12	18.30	17.60	10.60	11.00	7.80	18.80	0.66	0.69	1.14	1.10	1.55	0.64
TCC	S	6.06	11.20	10.20	18.00	16.40	8.90	14.20	0.54	0.59	0.34	0.37	0.68	0.43
TCG	S	15.15	9.30	5.30	10.70	10.70	8.70	8.50	1.63	2.86	1.42	1.42	1.74	1.78
TCT	S	39.39	25.20	20.00	10.30	12.00	8.70	23.50	1.56	1.97	3.82	3.28	4.53	1.68
ACA	T	15.15	15.70	17.40	9.10	10.50	8.20	17.80	0.97	0.87	1.67	1.44	1.85	0.85
ACC	T	0.00	10.30	9.70	19.10	16.60	22.80	12.60	0.00	0.00	0.00	0.00	0.00	0.00
ACG	T	3.03	7.70	4.50	9.70	11.00	14.80	7.90	0.39	0.67	0.31	0.28	0.20	0.38
ACT	T	9.09	17.50	20.30	9.20	10.70	9.10	20.30	0.52	0.45	0.99	0.85	1.00	0.45
GTA	V	9.09	9.90	11.40	5.30	6.40	11.10	11.80	0.92	0.80	1.72	1.42	0.82	0.77
GTC	V	15.15	12.80	11.10	21.30	21.00	15.10	11.60	1.18	1.37	0.71	0.72	1.00	1.31
GTG	V	30.30	17.40	16.70	25.00	25.50	25.50	10.60	1.74	1.81	1.21	1.19	1.19	2.86
GTT	V	48.49	27.20	26.80	14.30	15.70	18.50	22.00	1.78	1.81	3.39	3.09	2.62	2.20
TGG	W	12.12	12.50	12.20	12.00	13.00	15.20	10.30	0.97	0.99	1.01	0.93	0.80	1.18
TAC	Y	9.09	13.70	13.50	20.60	19.40	12.10	14.60	0.66	0.67	0.44	0.47	0.75	0.62
TAT	Y	15.15	14.60	17.80	8.30	9.50	16.50	18.90	1.04	0.85	1.83	1.59	0.92	0.80
TAA	*	3.03	0.90	1.10	0.60	0.50	2.00	1.00	3.37	2.75	5.05	6.06	1.52	3.03
TAG	*	0.00	0.50	0.50	0.60	0.70	0.30	0.50	0.00	0.00	0.00	0.00	0.00	0.00
TGA	*	0.00	1.20	1.00	1.50	1.10	1.10	0.70	0.00	0.00	0.00	0.00	0.00	0.00

*Termination codons; Data underlined: there are obvious differences in the values (≤0.5, ≥2.0) among the codons of two species; *SPDS*/A, *SPDS*/N, *SPDS*/T, *SPDS*/Z, *SPDS*/E, *SPDS*/Y represents frequency ratio of *CsSPDS* gene to *Arabidopsis thaliana*, *Nicotiana tabacum*, *Triticum aestivum*, *Zea mays*, *Escherichia coli*, and yeast, respectively.

Codon usage of *SPDS* genes

ENC-plots have been widely used to investigate patterns of codon usage in genes of various species. The curve of ENC-plots shows the expected position of genes whose codon usage is only determined by variation in the GC3 content (Wright, 1990). The distribution plots of ENC versus GC3s for the *SPDS* gene among 15 species are shown in Figure 2. ENC values for these *SPDS* sequences fluctuated between 46.56 and 59.51. The ENC value for the *CsSPDS* gene was 51.03, a little lower than the mean value of 51.91. From these data, the distribution plot of ENC versus GC3s for the *SPDS* gene among 15 species indicated that only the *HnSPDS* gene was above the curve, the remainder of the genes was just below the expected curve. The relationship between ENC and GC3 was positive ($P < 0.01$), indicating that most *SPDS* genes were subject to G+C compositional constraints.

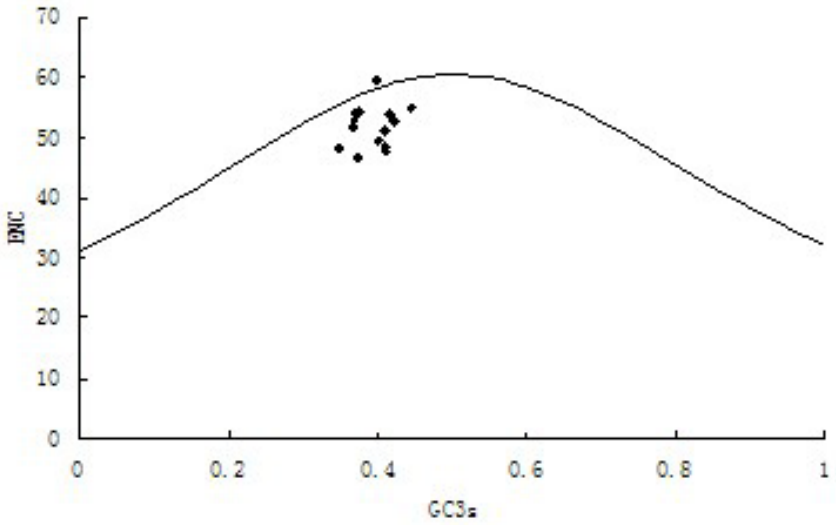

Figure 2. ENC-plot of the *SPDS* gene from 15 plant species (listed in Table 1). ENC denotes the effective number of codons of each gene and GC3s denote the G+C content on the third synonymous codon position of each gene. The solid line indicates the expected ENC value if the codon bias is only due to GC3s.

Cluster analysis of *SPDS* genes

Based on the RSCU values of the *SPDS* genes, the 15 samples were divided into two clades (Figure 3A). From the clustering results, *C. sinensis*, *O. ginseng* and *E. senticosus* were clustered together, indicating that they have similar codon usage. However, the phylogenetic analysis (Figure 3B) based on CDs of the 15 *SPDS* genes provides different clues on their relatedness. In the phylogenetic tree, *C. sinensis* was divided into one distinct terminal branch of the first clade. The length of the terminal branch of *CsSPDS* was the shortest in the first clade, indicating that the *CsSPDS* gene had had a distant genetic relationship with other *SPDS* genes in the first clade (the more distant the genetic relationship is, the bigger the expected variation in codon usage bias). Therefore, clustering based on RSCU did not accurately reflect the genetic relationship among the plants.

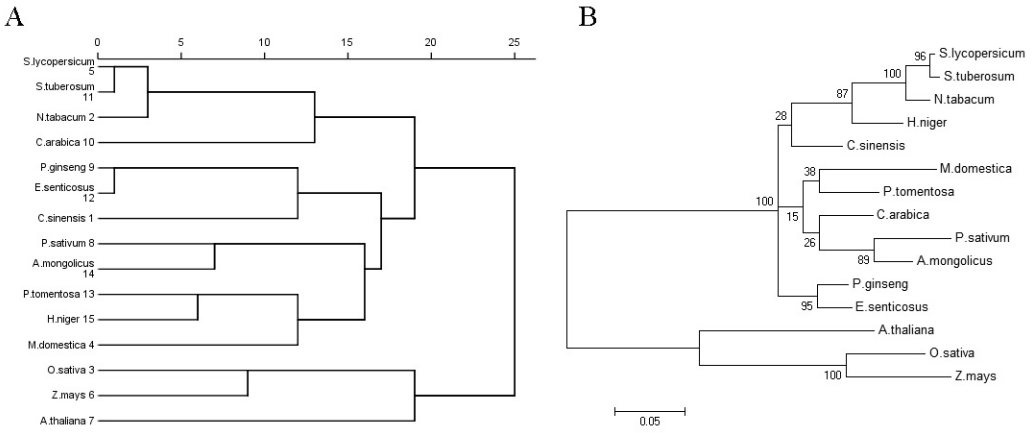

Figure 3. A. Clustering dendrogram based on RSCU values of the *SPDS* gene from 15 plant species (listed in Table 1). **B.** Phylogenetic analysis of the *SPDS* gene from 15 plant species.

DISCUSSION

Many previous studies have investigated codon usage bias in a number of eukaryotes and prokaryotes. In recent years, a number of studies have proven that following long-term evolution, species formed a set of specific codons in order to survive (Li et al., 2010; Qin et al., 2013). Many different factors were involved in codon usage bias, such as mutational bias, gene expression level, natural selection, gene length, and GC composition (Sueoka and Kawanishi, 2000; Blake et al., 2003; Lü et al., 2005; Behura and Severson, 2013). Without natural selection, special mutations would affect the base composition of sequences and they would be found at the third site of synonymous codons (Nei and Kumar, 2000). In the present study, the codon usage pattern was analyzed in *CsSPDS* to elucidate the factors responsible for variation in codon usage. We found that the *CsSPDS* gene preferred to use codons ending with A/T, the same as the other *C. sinensis* genome genes. This result suggests that mutational bias is one of the factors shaping codon usage in *CsSPDS* and in the *C. sinensis* genome genes.

An ENC-plot reflects the determinants of the codon usage variation among genes in different organisms. Wright (1990) held that if a particular gene was subject to G+C compositional constraints, it would lie on or just below the expected curve. If a gene was subject to selection for translationally optimal codons, it would lie considerably below the expected curve. From the ENC-plot, we could see that the *SPDS* genes were distributed regularly, with all the *SPDS* genes, except *HnSPDS*, lying near the solid curve on the left side of this distribution. In addition, a positive correlation was observed between ENC and GC3s. These results suggest that mutational bias is the major factor shaping codon usage in these *SPDS* genes.

Heterologous gene expression is a powerful methodology used in biotechnological processes. Codon usage has a significant impact on heterologous gene expression. In this study, we found that the diversity of codon bias between the *CsSPDS* gene and dicotyledonous plants was less than that between the *CsSPDS* gene and monocotyledonous plants. Therefore, *CsSPDS* genes may be expressed more efficiently in *A. thaliana* and *N. tabacum*. Furthermore, a number of recent experiments have demonstrated that genes optimized with host-preferred codon usage showed a higher level of expression in host cells than the wild-type genes (Grote

et al., 2005; Sau et al., 2006). Thus, some modifications of codons are necessary. This may serve as a guide for manipulating the expression of the *CsSPDS* gene.

ACKNOWLEDGMENTS

Research supported by the Livelihood Project of Qingdao city (#13-1-3-85-nsh) and the Breeding Project of Shandong Province (#2012LZ12-03). We are grateful to Associate Professor Juan Du (Life Sciences Institute, Qingdao Agricultural University) for assisting in the analysis of the results.

REFERENCES

Alcázar R, Altabella T, Marco F, Bortolotti C, et al. (2010). Polyamines: molecules with regulatory functions in plant abiotic stress tolerance. *Planta* 231: 1237-1249.

Behura SK and Severson DW (2013). Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes. *Biol. Rev. Camb. Philos. Soc.* 88: 49-61.

Blake WJ, Kærn M, Cantor CR and Collins JJ (2003). Noise in eukaryotic gene expression. *Nature* 422: 633-637.

Grote A, Hiller K, Scheer M, Münch R, et al. (2005). JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* 33: W526-W531.

Guo XL, Wang Y, Yang LC and Ding ZT (2012). Analysis of codon use features of CBF gene in *Camellia sinensis*. *Hereditas* 34: 1614-1623.

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.

Lei BB, Li SS, Liu GZ, Chen ZW, et al. (2013). Evolution of mitochondrial gene content: loss of genes, tRNAs and introns between *Gossypium harknessii* and other plants. *Plant Syst. Evol*. 299: 1889-1897.

Li ZP, Ying DQ, Li P, Li F, et al. (2010). Analysis of synonymous codon usage bias in 09H1N1. *Virol. Sin.* 25: 329-340.

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 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 and Xue Q (2004). Analysis of factors shaping codon usage in the mitochondrion genome of *Oryza sativa*. *Mitochondrion* 4: 313-320.

Lü H, Zhao WM, Zheng Y, Wang H, et al. (2005). Analysis of synonymous codon usage bias in *Chlamydia*. *Acta Biochim. Biophys. Sin.* 37: 1-10.

Nei M and Kumar S (2000). Molecular evolution and phylogenetics. Oxford University Press, Cary.

Nie X, Deng P, Feng K, Liu P, et al. (2013). Comparative analysis of codon usage patterns in chloroplast genomes of the Asteraceae family. *Plant Mol. Biol. Rep.* 32: 828-840.

Pan LL, Wang Y, Hu JH, Ding ZT, et al. (2013). Analysis of codon use features of stearoyl-acyl carrier protein desaturase gene in *Camellia sinensis*. *J. Theor. Biol.* 334: 80-86.

Qin Z, Cai Z, Xia G and Wang M (2013). Synonymous codon usage bias is correlative to intron number and shows disequilibrium among exons in plants. *BMC Genomics* 14: 56.

Qiu S, Bergero R, Zeng K and Charlesworth D (2011). Patterns of codon usage bias in *Silene latifolia*. *Mol. Biol. Evol.* 28: 771-780.

Ravi V, Khurana JP, Tyagi AK and Khurana P (2008). An update on chloroplast genomes. *Plant Syst. Evol.* 271: 101-122.

Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, et al. (2014). From algae to angiosperms - inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evol. Biol.* 14: 23.

Sagor GH, Berberich T, Takahashi Y, Niitsu M, et al. (2013). The polyamine spermine protects *Arabidopsis* from heat stress-induced damage by increasing expression of heat shock-related genes. *Transgenic Res*. 22: 595-605.

Sau K, Gupta SK, Sau S, Mandal SC, et al. (2006). Factors influencing synonymous codon and amino acid usage biases in Mimivirus. *Biosystems* 85: 107-113.

Sharp PM and Li WH (1986). An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* 24: 28-38.

Shi H, Wang Y, Yang LC and Ding ZT (2012). Analysis of codon bias of the cold regulated transcription factor *ICE1* in

tea plant. *Acta Hortic. Sin.* 39: 1341-1352.
Shi SL, Jiang YR, Liu YQ, Xia RX, et al. (2013). Selective pressure dominates the synonymous codon usage in Parvoviridae. *Virus Genes* 46: 10-19.
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 XC, Zhao QY, Ma CL, Zhang ZH, et al. (2013). Global transcriptome profiles of *Camellia sinensis* during cold acclimation. *BMC Genomics* 14: 415.
Wang Y, Jiang CJ, Li YY, Wei CL, et al. (2012). CsICE1 and CsCBF1: two transcription factors involved in cold responses in *Camellia sinensis*. *Plant Cell Rep.* 31: 27-34.
Wright F (1990). The 'effective number of codons' used in a gene. *Gene* 87: 23-29.
Xu C, Dong J, Tong C, Gong X, et al. (2013). Analysis of synonymous codon usage patterns in seven different citrus species. *Evol. Bioinform. Online* 9: 215-228.