Bioinformatic analyses of GRAS genes in *Betula kirghisorum* based on transcriptome data

C.J. Yang[1], G.Y. Li[1] and Y.L. Cui[2,3]

[1]College of Forestry, Northeast Forestry University, Harbin, China
[2]Institute of Computer Science and Information Engineering, Harbin Normal University, Harbin, China
[3]College of Mechanical and Electrical Engineering, Northeast Forestry University, Harbin, China

Corresponding author: Y.L. Cui
E-mail: cui_cui_1111@163.com

Genet. Mol. Res. 15 (3): gmr.15038321
Received December 21, 2015
Accepted May 25, 2016
Published September 23, 2016
DOI http://dx.doi.org/10.4238/gmr.15038321

ABSTRACT. The transcriptomes of salt-stressed and unstressed *Betula kirghisorum* plants were analyzed using high throughput sequencing technology. A total of 52,239,804 and 51,772,998 clean reads were obtained from the two libraries, respectively, and *de novo* assembled into 60,545 all-unigenes. A total of 39,997 unigenes were annotated using public databases. Overall, 7206 genes were differentially expressed in unigenes and were involved in 127 pathways. Thirteen transcription factor families were identified in *B. kirghisorum*, including GRAS proteins, which are plant-specific transcription factors. By using bioinformatic methods to predict and analyze physicochemical properties, structural data were obtained on the 19 potential GRAS proteins. The results revealed that these proteins are hydrophilic, with significant differences in their length and molecular weight. The main secondary structures were alpha helices and random coils. BkGRAS

proteins possess typical GRAS domains: LHR I; VHIID motif; LHR II; PFYRE motif; and SAW motif. In the majority of BkGRAS proteins, AGG, AGA, UCU, GCU, GGG, CCA, GUU, GUG, AUU, GAU, and AAG codons were used preferentially. Aside from the *BkGRAS17* gene (relative synonymous codon usage (RSCU) = 1.20), usage of the UUA codon by other *BkGRAS* genes was low (RSCU < 1.0). The effective number of codons showed that *BkGRAS* genes have low codon bias. Subcellular localization analysis that predicted these proteins are found in the nucleus, cytoplasm, or chloroplast. BkGRAS proteins were divided into six subfamilies: SCR, LISCL, SCL3, DELLA, HAM, and PAT1. These results provide important information for the further functional study of GRAS genes in *B. kirghisorum*.

Key words: Transcriptome; *Betula kirghisorum*; GRAS transcription factor; Bioinformatic analysis

INTRODUCTION

Transcription factors play important roles in plant growth and development, morphological evolution, and anti-stress process. One important kind of transcription factors, GRAS family, exists widely in plants. GRAS proteins play important roles in plant growth and development, and help plants to resist stress occurring in response to temperatures, drought, salt, and abiotic stress. The name of the gene family GRAS was designed from the first three functionally characterized members (GAI, RGA, SCR) (Di Laurenzio et al., 1996; Peng et al., 1997; Silverstone et al., 1998; Pysh et al., 1999). GRAS proteins have variable N-terminal domains and conserved C-terminal domains. The typical structure of the C-terminal includes a leucine heptads repeat I (LHR I), VHIID, leucine heptads repeat II (LHRII), PFYRE, and SAW motifs. GRAS proteins were first identified in *Arabidopsis* (Benfey et al., 1993), and were subsequently found in other plants, including rice (Li et al., 2003), tobacco (Czikkel and Maxwell, 2007), maize (Lim et al., 2000), barley (Chandler et al., 2002), petunia (Tuurman et al., 2002), cucumber (Wis'niewska et al., 2013), *Medicago truncatula* (Kim and Nam, 2013), and pine (Solé et al., 2008). Recently, GRAS proteins were also identified in *Populus*, which are woody plants (Ma et al., 2011), and named PeSCL7. Currently, 106, 34, and 60 putative GRAS genes have been identified in *Populus*, *Arabidopsis*, and rice, respectively (Liu and Widmer, 2014). Information on gene distribution, chromosomal localization, and GRAS gene expression in the plant genome will have significance for the further study of gene function and for the potential use of these genes in the improvement of abiotic stress tolerance in crops. In addition, this study would provide a basis for improving the current poor ecological environment, including sparse plants and low species diversity in areas subjected to adverse conditions.

MATERIAL AND METHODS

Culture and preparation of salt-tolerant birch materials

Betula kirghisorum was obtained from Naurzum Reserve in the northwest of Kazakhstan. A previous experiment found that plant height and ground diameter of *B.*

kirghisorum 23 increased and malonaldehyde (MDA) content decreased under $NaHCO_3$ stress (0.4%), and the salt injury index was less than in other families, indicating that members of this family have high resistance to $NaHCO_3$ stress (Yang et al., 2013). Therefore, *B. kirghisorum* 23 was used as the test materials in this study. Annual potted *B. kirghisorum* 23 plants (approximately 20 cm in high) were subjected to salt stress (0.6% $NaHCO_3$ for 48 h), and non-$NaHCO_3$ stress-treated plants were used as the control group. Total RNA was extracted from leaves using the cetyltrimethyl ammonium bromide (CTAB) method.

Transcriptome sequencing analysis

RNA was sent to BGI company, which constructed a cDNA library and performed high-throughput transcriptome sequencing. Reads of samples subjected to salt stress for 48 h and of the control sample were obtained. Raw reads were cleaned by removing the adaptor-containing sequences and low-quality reads. All clean reads were subsequently assembled using the *de novo* assembly program Trinity software (http://trinityrnaseq.sourceforge.net/).

Functional annotation of unigenes

The assembled unigenes were submitted to databases for homology and annotation comparison using the BLASTx and BLASTn algorithms (e-value < 10^{-5}), including non-redundant (NR) protein sequences, Swiss-Prot, Kyoto Encyclopedia of Genes and Genomes (KEGG), Clusters of Orthologous Groups (COG), and non-redundant nucleotide (NT) analyses.

Analysis of differential gene expression

Gene expression levels were calculated using the reads per kilobase per million reads (RPKM) method (Mortazavi et al., 2008). A rigorous algorithm was developed to identify differentially expressed genes between the two samples (control group and 0.6% $NaHCO_3$ stress-treated group) based on a previously described method (Audic and Claverie, 1997). The false-discovery rate (FDR) was controlled to determine differentially expressed genes (Benjamini and Yekutieli, 2001). In this study, FDR $\leq$ 0.001 and the absolute value of log2 ratio $\geq$1 were used as the threshold to judge the significance of differences in gene expression.

Database search

The NCBI (http://www.ncbi.nlm.nih.gov/) database was used to downloaded GRAS gene sequences of *Populus* and *Arabidopsis* from *P. trichocarpa* (JGI v3.0) and *A. thaliana* (TAIR annotation release 10).

Analysis of physicochemical properties of GRAS proteins in *B. kirghisorum* ToolProtParam (expasy.org/prot param/http://web.) was used to determine the number of amino acids, molecular weight, and isoelectric point, and to predict the hydrophilicity and instability index of the identified BkGRAS proteins.

Domain analysis of GRAS proteins in *B. kirghisorum*

In this study, full-length or domain-specific amino acid sequences of BkGRAS proteins

were aligned using MUSCLE. Sequence alignment and domain analysis of *B. kirghisorum* GRAS proteins were performed as previously described (Pysh et al., 1999).

Secondary structure prediction of GRAS proteins in *B. kirghisorum*

The secondary structure of the BkGRAS proteins was predicted in the SOPMA server (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?Page=npsa_sopma. html).

Prediction of transmembrane domains and signal peptides in *B. kirghisorum* GRAS proteins

The TMpred software (http://embnet.vital-it.ch/software/TMPRED_form.html) was used to analyze the transmembrane domain of BkGRAS proteins. The SignaIP 4.1 software (http://www.cbs.dtu.dk/services/SignaIP/) was used to determine whether these proteins contained a signal peptide. The subcellular localization was predicted with the WoLF PSORT software (http://www. genscript.com/results/142647512625789.html).

Codon preference analysis of *B. kirghisorum* GRAS proteins

The CodonW software was used to analyze the relative synonymous codon usage (RSCU), effective number of codons (ENC), and nucleotide composition (including A3s, T3s, G3s, C3s, GC, and GC3s).

Phylogenetic analysis of *B. kirghisorum* GRAS proteins

Liu and Widmer (2014) identified 93 and 33 GRAS genes in *Populus* and *Arabidopsis*, respectively, with full-length domain sequences. Therefore, we downloaded those GRAS gene sequences for *Populus* and *Arabidopsis* from the *P. trichocarpa* (JGI v3.0) and *A. thaliana* (TAIR annotation release 10) databases, respectively, and multiple-sequence alignment was performed using multiple alignment in fast Fourier transform (MAFFT) with 19 BkGRAS proteins. We constructed the phylogenetic tree for *B. kirghisorum*, *Populus*, and *Arabidopsis* based on the alignment. Finally, we modified the phylogenetic tree using iTOL: Interactive Tree Of Life (http://itol.embl.de/itol.cgi).

RESULTS

Transcriptome sequencing analysis

After libraries were constructed using the transcriptomes of salt-stressed and unstressed control plants, 56,646,418 and 55,911,538 raw reads were generated, respectively, from each library. After removing low-quality reads, a total of 52,239,804 and 51,772,998 high-quality clean reads remained, respectively. After sequence assembly, the two library obtained 66,467 and 53,517 unigenes, respectively, of which 64,494 and 51,908 were singletons. The mean length of unigenes in the two libraries was 592 and 636 bp, respectively. The total number of all-unigenes was 60,545 after the two libraries were assembled, with a mean length of 758 bp. Transcriptome yield and assembly yield statistics are shown in Table 1.

Table 1. Transcriptome and assembly quality.

	0.6% $NaHCO_3$ (stress 48 h)	Control (without stress)
Total reads	56,646,418	55,911,538
Total nucleotides (nt)	4,701,582,360	4,659,569,820
High-quality reads	52,239,804	51,772,998
High-quality reads (%)	92.22	92.59
GC percentage (%)	47.09	47.70
Number of contigs	115,514	97,241
Number of singletons	64,494	51,908
Number of clusters	1973	1609
Number of unigenes	66,467	53,517
Mean length (nt) of contigs	331	355
Mean length (nt) of unigenes	592	636

Function annotation of unigenes

Unigenes were annotated with the NR, NT, SwissProt, KEGG, COG, GO databases (Table 2). For functional annotation analysis, 37,653, 34,259, 22,763, 20,414, 13,359, and 16,586 unigenes were annotated based on the NR, NT, Swiss-Prot, KEGG, COG, and GO databases, respectively, and the total number of annotated unigenes was 39,997.

Table 2. Unigene annotation.

Sequence file	NR	NT	SwissProt	KEGG	COG	GO	All
All-Unigene.fa	37,653	34,259	22,763	20,414	13,359	16,586	39,997

NR: non-redundant protein database, NT: non-redundant nucleotide, SwissProt: Swiss-Prot database, KEGG: Kyoto Encyclopedia of Genes and Genomes, COG: Clusters of Orthologous Groups, GO: Gene Ontology, All: unigenes were annotated with the databases of NR, NT, SwissProt, KEGG, COG, and GO. Then, counted the number of unigenes annotated with each database.

Differential expression of unigenes

Genes that were differentially expressed between samples were identified by differential expression analysis of unigenes. Of the 7206 differentially expressed unigenes, 12,556 were up-regulated and 9275 were down-regulated (Figure 1).

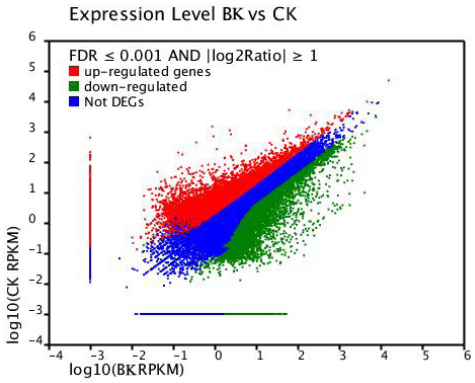

Figure 1. Differences in gene expression between stress-treated and control samples. FDR: false-discovery rate; Bk: *Betula kirghisorum*; CK: control group; RPKM: reads per kilobase per million reads.

Pathway analysis of different unigenes

In vivo, different genes coordinate to exert biological function. Pathway analysis was helpful to further understand the biological function of genes. Differentially expressed genes involved in the main biochemical metabolic and signal transduction pathways were identified through pathway significant enrichment analysis. The results are shown in **Table S1**. A total of 7206 genes were identified as differentially expressed in unigenes, which were involved in 127 pathways involved in biochemical metabolism and signal transduction. Differentially expressed unigenes were involved in metabolic pathways (1811 unigenes, accounting for 25.13% of the total differentially expressed unigenes), biosynthesis of secondary metabolites (963, 13.36%), plant-pathogen interactions (734, 10.19%), plant hormone signal transduction (531, 7.37%), glycerophospholipid metabolism (279, 3.87%), endocytosis (269, 3.73%), ribosome (263, 3.65%), RNA transport (246, 3.41%), phenylpropanoid biosynthesis (238, 3.30%), spliceosome (237, 3.29%), ether lipid metabolism (225, 3.12%), purine metabolism (191, 2.65%), starch and sucrose metabolism (188, 2.61%), protein processing in endoplasmic reticulum (179, 2.48%), and ribosome biogenesis in eukaryotes (177, 2.46%). The distribution proportion of differential expression genes was more than 2% in the above 15 pathways. Other differentially expressed genes were distributed in the remaining 112 pathways, and the proportion ranged from 0.01 to 1.85%.

Screening of the key transcription factor families

Plant transcription factors regulate the expression of a series of genes related to stress, affecting plant tolerance to adverse conditions. Transcription factors are closely associated with resistance to cold, drought, and salt stress, and to disease resistance. Plants can respond to various internal and external signals when subject to stress, and these responses involve the precise regulation of functional gene expression. When plants perceived external signals such as drought, high salt, hormone, disease or intracorporal cell development signal, the transcription factor would be activated through a series of transmission process. Transcription factors, in combination with cis-acting elements, activate RNA polymerase II transcription complexes, which initiate gene transcription, and then make corresponding responses to the above signals through gene product. Transcription factors play important regulatory roles in the plant response to adverse conditions, and accurately control the expression of downstream functional genes, allowing the plant to respond to the stress condition. Thirteen different transcription factor families (Table 3) were identified in the transcriptome of salt-stressed *B. kirghisorum*. These transcription factor families are related to the environmental-stress response. GRAS proteins are plant-specific transcription factors closely related to abiotic stress. GRAS proteins usually consist of 400-700-amino acid residues (Hirsch and Oldroyd, 2009); therefore, we studied 19 GRAS proteins containing >400-amino acid residues.

Physicochemical properties of GRAS proteins in *B. kirghisorum*

Prediction of structural domains confirmed that these genes belonged to the GRAS gene family; therefore, these genes were named *BkGRAS1* to *BkGRAS19*. Predicted physicochemical properties of BkGRAS proteins are shown in Table 4. The number of amino acids in these proteins and their molecular weights differed significantly. The size of

the deduced GRAS proteins varied greatly, ranging from 408 amino acids (BkGRAS3) to 1029 amino acids (BkGRAS5). There were 15 genes encoding more than 500 amino acids. The molecular weight varied from 45.90 kDa (BkGRAS3) to 111.93 kDa (BkGRAS5). The theoretical PI of BkGRAS proteins ranged from acidic to alkaline. The range of the predicted theoretical PI was small, from 5.00 to 7.00, for which BkGRAS14 and BkGRAS15 had the lowest values. The grand average hydropathicity (GRAVY) was predicted to be negative, indicating that these proteins are hydrophilic. If the instability index of a protein is <40, then the protein is stable, when the instability index is >40, the protein is unstable. Table 1 shows that all of the BkGRAS proteins are unstable.

Table 3. Key transcription factors.

Name	Number	Up-regulated	Down-regulated	Notes
GRAS	84	29	54	1
WRKY	77	57	19	1
Zinc finger protein	353	168	175	10
NAC	97	49	47	1
MYB	222	103	116	3
bHLH	75	26	43	6
HSF	15	9	6	
MADS	28	15	13	
ERF	42	26	14	2
bZIP	77	40	37	
AP2	119	29	63	27
ARF	85	56	22	7
HD-ZIP	19	15	4	

Number of genes with no expression variation in experiment group when compared with those in control group.

Table 4. Physicochemical properties of GRAS proteins in *Betula kirghisorum*.

Gene name	Amino acid number	Molecular weight (kDa)	Isoelectric point	GRAVY	Instability index
BkGRAS1	537	60.06	5.50	-0.231	53.07
BkGRAS2	537	60.06	5.56	-0.223	53.56
BkGRAS3	**408**	**45.90**	5.54	-0.314	**41.61**
BkGRAS4	537	60.01	5.63	-0.224	51.01
BkGRAS5	**1029**	**111.93**	5.75	-0.242	57.34
BkGRAS6	437	47.89	5.69	-0.173	**57.98**
BkGRAS7	604	65.53	5.35	-0.243	51.58
BkGRAS8	660	72.41	6.11	-0.473	52.62
BkGRAS9	654	74.35	6.43	**-0.632**	52.39
BkGRAS10	581	65.07	5.82	-0.401	54.69
BkGRAS11	610	67.75	5.73	-0.137	55.51
BkGRAS12	546	61.27	6.15	-0.342	50.20
BkGRAS13	537	58.70	5.58	**-0.103**	49.28
BkGRAS14	577	64.55	5.00	-0.355	43.98
BkGRAS15	577	64.54	**5.00**	-0.357	44.61
BkGRAS16	474	53.00	5.59	-0.155	47.81
BkGRAS17	547	60.57	5.49	-0.211	48.34
BkGRAS18	549	61.17	6.11	-0.215	53.79
BkGRAS19	446	50.37	**6.72**	-0.409	50.41

Bold: the maximum and minimum for all properties. GRAVY: grand average of hydropathicity.

Domain analysis of GRAS proteins in *B. kirghisorum*

The GRAS gene products share significant similarity throughout their C-termini. The C-terminal regions can be subdivided into five distinct sequence motifs, found in the following order: LHR I; VHIID motif; LHR II; PFYRE motif; and SAW motif. The P-N- H-D-Q-L residues are conserved in the VHIID motif. The P residue is conserved in the PFYRE motif. The SAW motif includes three pairs of conserved residues: R-E, W-G, and W-W. As in *A. thaliana*, the *B. kirghisorum* GRAS gene family included these typical GRAS domains, as shown in Figure 2.

Figure 2. Domain analysis of GRAS family members in *Betula kirghisorum*. Yellow highlights refer to conserved residues in the individual motifs.

Secondary structure prediction of GRAS proteins in *B. kirghisorum*

Methods currently used to predict the secondary structure of proteins include pROF, GoRIV, GORI, NNPREDICT, PHDsec, SSprov2.0, PSIPRED, PREDATOR, APSSP2, and SOPMA, and the number of websites used to predict secondary structures has increased gradually. In this study, the SOPMA method was used to predict the secondary structure of BkGRAS proteins. The SOPMA method combines data from five different methods to predict secondary structure. Four different secondary structure motifs were evaluated; α-helix, β-sheet (extended strand), random coil, and beta turn "h", "e", "c", and "t" represent α-helix, extended strand (β-sheet), random coil, and beta turn, respectively. The major secondary structure of BkGRAS5 and BkGRAS9 proteins was predicted to be random coils, and the remaining BkGRAS proteins possessed alpha helices (Table 5).

Table 5. Secondary structure prediction of GRAS proteins in *Betula kirghisorum*.

Protein name	Alpha helix (Hh, %)	Extended strand (Ee, %)	Beta turn (Tt, %)	Random coil (Cc, %)
BkGRAS1	45.07	13.59	8.19	33.15
BkGRAS2	43.39	14.34	8.57	33.71
BkGRAS3	47.55	17.89	9.80	24.75
BkGRAS4	44.88	13.97	8.19	32.96
BkGRAS5	32.65	16.72	7.68	**42.95**
BkGRAS6	45.31	15.56	10.3	28.83
BkGRAS7	42.38	14.57	8.61	34.44
BkGRAS8	43.94	11.36	6.82	37.88
BkGRAS9	37.46	13.76	9.33	**39.45**
BkGRAS10	40.45	17.73	9.47	32.36
BkGRAS11	41.15	15.25	7.21	36.39
BkGRAS12	47.25	12.64	9.34	30.77
BkGRAS13	48.79	12.85	8.75	29.61
BkGRAS14	42.81	16.46	7.11	33.62
BkGRAS15	42.29	16.64	7.11	33.97
BkGRAS16	50.63	14.35	8.44	26.58
BkGRAS17	45.16	15.17	9.32	30.35
BkGRAS18	43.35	12.02	6.19	38.43
BkGRAS19	43.27	16.82	10.76	29.15

Bold: indicates that the main secondary structure was a random coil (Cc) in the BkGRAS5 and BkGRAS9 proteins.

Analysis of the transmembrane structure, signal peptide, and subcellular localization of GRAS proteins in *B. kirghisorum*

Bioinformatics were used to predict and analyze the membrane structure, signal peptide, and subcellular localization of 19 BkGRAS proteins, in order to more efficiently explore their function. Typical transmembrane helical regions are composed of 20-30 hydrophobic amino acids (e.g., Leu, Ile, Val, Met, Gly, and Ala). Predictions made using the TMHMM software showed that the 19 encoded BkGRAS proteins contained no transmembrane domains. Signal peptides are located in the N-terminal of newly synthesized peptides and generally consist of 15-30-amino acid residues, in which 6-15 nonpolar amino acids are positively charged. The signal peptide determines whether a protein will be secreted, and is also related to the positioning of a protein or its nascent polypeptide in cells. Signal peptide prediction results indicated that the 19 BkGRAS proteins did not contain a signal peptide sequence, suggesting that they are not secreted. The subcellular localization can be used to predict protein function,

and the subcellular localizations of 19 BkGRAS proteins were predicted by the wolf PSORT software. The predictions are represented by a score, with the higher scores indicative of greater predictive accuracy. As shown in Table 6, BkGRAS1, 2, 5, 7, 8, 9, 13, and 17 proteins were located in the nucleus, and BkGRAS2 was localized to the cytoplasm. BkGRAS3, 4, 11, 16, and 19 were located in the cytoplasm, and BkGRAS6, 10, 12, 14, 15, and 18 were located in the chloroplast.

Table 6. Subcellular localization of GRAS proteins in *Betula kirghisorum*.

Protein name	Predicted subcellular localization
BkGRAS1	nucl: 6, cyto: 4, chlo: 2, vacu: 1
BkGRAS2	nucl: 5, cyto: 5, cysk: 2, chlo: 1
BkGRAS3	cyto: 8, mito: 4, pero: 1
BkGRAS4	cyto: 5, nucl: 4, chlo: 2, cysk: 2
BkGRAS5	nucl: 8, cyto: 4, vacu: 1
BkGRAS6	chlo: 10, nucl: 2, mito: 1
BkGRAS7	nucl: 6, cyto: 5, chlo: 2
BkGRAS8	nucl: 6, chlo: 5, mito: 2
BkGRAS9	nucl: 12, cyto: 1
BkGRAS10	chlo: 9, nucl: 2, plas: 1, E.R.: 1
BkGRAS11	cyto: 11, chlo: 3
BkGRAS12	chlo: 6, nucl: 4, cysk: 2, mito: 1.5, cyto_mito: 1.5
BkGRAS13	nucl: 8.5, nucl_plas: 5, cyto: 3, chlo: 1
BkGRAS14	chlo: 6, nucl: 5, cyto: 1, mito: 1
BkGRAS15	chlo: 6, nucl: 5, cyto: 1, mito: 1
BkGRAS16	cyto: 8, nucl: 5
BkGRAS17	nucl: 8, chlo: 3, mito: 1, plas: 1
BkGRAS18	chlo: 9, nucl: 4
BkGRAS19	cyto: 6, nucl: 5, mito: 1, pero: 1

Ncl: nucleus; Cyto: cytoplasm; Mito: mitochondria; Pero: peroxisome; Vacu: vacuole; E.R.: endoplasmic reticulum; Chlo: chloroplast; Cysk: cytoskeleton; Plas: plasma membrane.

Indices of codon usage in *B. kirghisorum* GRAS proteins

Analysis of relative synonymous codon usage

Genes from different species or within the same species exhibit various patterns of codon usage bias. Codon bias is the non-random use of synonymous codons in genes. RSCU is an index used to study variation in the usage of all synonymous codons encoding the same amino acid among genes (Sharp and Li, 1986). RSCU values were calculated as the ratio of the observed and expected frequencies of a codon (excluding methionine, tryptophan, and the three termination codons). When RSCU values are equal to 1.0, codons for the same amino acids are used equally and randomly. RSCU values greater than 1.0 indicate a positive codon usage bias for the corresponding codons, and codons with values less than 1.0 exhibit negative codon usage bias.

The RSCU value reflects codon use preference (Sharp and Li, 1986). Codon usage bias of 19 BkGRAS proteins is shown in Table 7. The RSCU value of the UUG codon was higher than 1.0 in all BkGRAS proteins, indicating that these proteins preferred the UUG codon. In most of the BkGRAS proteins, the RSCU values of AGG, AGA, UCU, GCU, GGG, CCA,

GUU, GUG, AUU, GAU, and AAG were higher than 1.0, indicating that these codons were used frequently and that there was codon bias. For the CAG codon, RSCU values of BkGRAS13, 16, and 17, proteins were less than 1.0, those for BkGRAS14 and 15 proteins were 1.0, and those for the remaining proteins were higher than 1.0, indicating that different genes within the same species and within the same gene family exhibit different usage bias for the same codon.

With the exception of BkGRAS17, the RSCU values of the UUA codon were lower than 1.0 in the remaining 18 BkGRAS proteins, indicating that the UUA codon is not favored in these proteins. The RSCU values were 1.0 in amino acids containing Ile in BkGRAS13; Asn in BkGRAS6, 14, and 15; Cys in BkGRAS3, 7, and 19; Gln in BkGRAS14 and 15; His in BkGRAS11, 14, and 15; Lys in BkGRAS12; and Phe in BkGRAS11. These results indicated that the above GRAS genes had no codon bias when encoding these amino acids.

Analysis of ENC values and GC% and GC3s% levels

The ENC or Nc was used to estimate the degree of codon bias for a single gene, ranging from 20 to 61. If the ENC value of a gene is 20, this indicates that the gene has the strongest codon usage bias and uses only one codon. If all synonymous codons of a certain gene are used equally, its ENC value is 61 (Jiang et al., 2008). In general, if the ENC value of a gene is 35 or less, that gene is thought to possess strong codon bias (Wright, 1990). As shown in Table 8, the ENC value of the genes encoding BKGRAS6 (46.16), BKGRAS16 (49.63), and BKGRAS18 (49.78) were close to 50 and for the remaining genes exceeded 50, indicating that BkGRAS genes possess low codon bias.

A3s, T3s, G3s, and C3s represent the frequency of adenine, thymine, guanine, and cytosine at the third position of synonymous codons (Li et al., 2014). The nucleotide levels of A3, T3, C3, and dG3 in BkGRAS gene-coding regions are listed in Table 8. The G3s and C3s values of the BkGRAS6, 7, 8, 11, 13, and 18 genes were greater than the values for T3s and A3s, which indicates that the third base of the synonymous codon in these six genes preferred G or C bases.

Wright (1990) reported that the ENC value was correlated with the G+C content (GC%) and the G+C content at the third base of codon (GC3s%) of the gene. Therefore, in the present study we also calculated nucleotide compositions (overall GC content especially the GC3s%) of the coding sequence of 19 BkGRAS genes using the CodonW software. As shown in Table 8, GC and GC3s values of six genes containing BkGRAS6, 7, 8, 11, 13, and 18 were greater than 0.5 in *B. kirghisorum*, indicating that these six genes are slightly GC rich and AT poor. The GC and GC3s values of the remaining BkGRAS genes were less than 0.5, which indicates that these genes preferentially used codons ending in T or A; these results are consistent with the results obtained through RSCU analysis.

Phylogenetic tree analysis of GRAS proteins in *B. kirghisorum*

B. kirghisorum and *Populus* are woody plants, and GRAS proteins in *A. thaliana* have been well studied. Therefore, in the present study, a phylogenetic analysis of GRAS proteins from *B. kirghisorum*, *Populus*, and *Arabidopsis* was performed. As shown in the phylogenetic tree (Figure 3), BkGRAS6, BkGRAS9, and BkGRAS16 proteins belong to the SCR, LISCL, and SCL3 subfamilies, respectively. BkGRAS7 and BkGRAS13 proteins belong to the DELLA subfamily. BkGRAS5, BkGRAS11, and BkGRAS18 proteins belong to the HAM subfamily, and the remaining BkGRAS proteins belong to the PAT1 subfamily.

Table 7. Relative synonymous codon usage of GRAS genes in *Betula kirghisorum*.

Amino acid	Codon	*BkGRAS*									
		1	2	3	4	5	6	7	8	9	10
Arg	CGU	0.58	0.58	0.62	0.56	0.32	0.22	0.00	0.42	0.77	0.90
	CGC	0.39	0.39	0.62	0.38	0.49	1.11	0.22	0.84	0.64	0.60
	CGA	0.39	0.39	0.21	0.38	0.65	0.67	0.89	1.40	0.13	0.15
	CGG	0.58	0.58	1.24	0.75	0.97	0.44	2.67	1.40	0.77	1.05
	AGA	2.32	2.32	2.07	2.25	1.95	1.56	0.67	0.84	1.40	1.80
	AGG	1.74	1.74	1.24	1.69	1.62	2.00	1.56	1.12	2.30	1.50
Leu	UUA	0.65	0.67	0.15	0.65	0.58	0.41	0.29	0.62	0.53	0.75
	UUG	1.75	1.78	1.54	1.75	1.38	2.03	1.74	1.14	1.37	1.29
	CUU	1.20	1.22	1.23	1.20	1.27	0.71	0.58	0.62	1.79	1.07
	CUC	0.87	0.89	0.77	0.87	1.62	1.02	1.35	1.24	0.84	0.96
	CUA	0.44	0.44	0.15	0.44	0.35	0.31	0.48	0.62	0.53	0.32
	CUG	1.09	1.00	2.15	1.09	0.81	1.53	1.55	1.76	0.95	1.61
Ser	**UCU**	1.58	1.68	1.78	1.58	2.15	0.49	1.26	0.60	1.03	1.68
	UCC	1.26	1.26	1.33	1.26	1.08	1.35	1.68	1.89	0.52	1.16
	UCA	1.58	1.58	0.67	1.58	0.87	0.00	0.53	0.26	1.76	0.84
	UCG	0.21	0.11	0.67	0.21	0.87	1.96	1.05	1.63	0.83	0.53
	AGU	0.95	0.95	0.44	0.95	0.36	0.12	0.42	0.51	0.93	0.63
	AGC	0.42	0.42	1.11	0.42	0.67	2.08	1.05	1.11	0.93	1.16
Ala	**GCU**	1.83	1.78	1.37	1.83	1.55	0.53	0.95	0.67	1.20	1.40
	GCC	0.69	0.67	0.57	0.57	0.85	1.47	1.42	1.19	0.60	0.50
	GCA	1.49	1.56	1.49	1.60	0.94	0.27	0.34	0.44	1.70	1.50
	GCG	0.00	0.00	0.57	0.00	0.66	1.73	1.29	1.70	0.50	0.60
Gly	GGU	1.07	1.10	0.17	1.07	1.23	0.53	0.80	0.74	0.86	0.39
	GGC	0.80	0.83	0.87	0.80	1.42	0.95	1.20	1.30	0.86	0.65
	GGA	0.80	0.83	1.04	0.80	0.43	0.32	0.56	0.56	1.19	1.16
	GGG	1.33	1.24	1.91	1.33	0.92	2.21	1.44	1.40	1.08	1.81
Pro	CCU	0.67	0.65	1.26	0.67	1.21	1.00	0.61	0.82	1.11	1.60
	CCC	1.47	1.42	0.63	1.47	0.74	0.60	0.73	1.36	0.78	0.57
	CCA	1.73	1.68	1.26	1.73	1.26	0.60	1.33	0.64	1.33	1.49
	CCG	0.13	0.26	0.84	0.13	0.79	1.80	1.33	1.18	0.78	0.34
Thr	ACU	1.47	1.47	1.07	1.40	1.49	0.63	0.89	0.74	1.04	1.38
	ACC	1.26	1.26	1.33	1.20	1.21	1.89	2.22	1.67	1.04	1.38
	ACA	1.05	1.05	1.60	1.00	0.74	0.21	0.44	0.00	1.22	1.08
	ACG	0.21	0.21	0.00	0.40	0.56	1.26	0.44	1.58	0.70	0.15
Val	**GUU**	1.74	1.70	1.03	1.70	1.14	0.48	0.72	0.80	1.08	1.33
	GUC	0.41	0.40	0.77	0.50	1.22	0.80	1.44	0.80	1.30	0.72
	GUA	0.51	0.50	0.52	0.50	0.57	0.16	0.10	0.00	0.54	0.51
	GUG	1.33	1.40	1.68	1.30	1.06	2.56	1.74	2.40	1.08	1.44
Ile	**AUU**	1.38	1.38	1.63	1.38	1.86	0.88	1.30	1.05	1.22	1.50
	AUC	1.15	1.15	0.88	1.15	0.66	1.41	1.30	1.65	1.00	0.75
	AUA	0.46	0.46	0.50	0.46	0.48	0.71	0.39	0.30	0.78	0.75
Asn	AAU	1.19	1.19	0.84	1.19	1.02	1.00	0.55	0.77	1.30	0.88
	AAC	0.81	0.81	1.16	0.81	0.98	1.00	1.45	1.23	0.70	1.12
Asp	**GAU**	1.83	1.75	1.28	1.75	1.27	0.86	1.06	0.48	1.35	1.24
	GAC	0.17	0.25	0.72	0.25	0.73	1.14	0.94	1.52	0.65	0.76
Cys	UGU	0.83	0.83	1.00	0.83	0.88	0.00	1.00	0.67	1.09	1.11
	UGC	1.17	1.17	1.00	1.17	1.13	2.00	1.00	1.33	0.91	0.89
Gln	CAA	0.87	0.87	0.73	0.82	0.83	0.89	0.87	0.84	0.74	0.80
	CAG	1.13	1.13	1.27	1.18	1.17	1.11	1.13	1.16	1.26	1.20
Glu	GAA	1.26	1.30	0.97	1.24	0.85	0.20	0.65	0.53	1.11	0.98
	GAG	0.74	0.70	1.03	0.76	1.15	1.80	1.35	1.47	0.89	1.02
His	CAU	1.13	1.13	1.11	1.13	0.74	0.50	0.84	0.71	1.20	1.07
	CAC	0.88	0.88	0.89	0.88	1.26	1.50	1.16	1.29	0.80	0.93
Lys	AAA	0.27	0.27	0.67	0.27	0.76	0.57	0.48	0.80	0.92	0.83
	AAG	1.73	1.73	1.33	1.73	1.24	1.43	1.52	1.20	1.08	1.17
Phe	UUU	1.11	1.11	1.54	1.11	1.14	0.13	0.21	0.81	1.03	1.45
	UUC	0.89	0.89	0.46	0.89	0.86	1.88	1.79	1.19	0.97	0.55
Tyr	UAU	1.22	1.22	0.83	1.22	0.89	0.25	0.33	0.50	0.95	1.07
	UAC	0.78	0.78	1.17	0.78	1.11	1.75	1.67	1.50	1.05	0.93

Continued on next page

Table 7. Continued.

Amino acid	Codon	*BkGRAS*								
		11	12	13	14	15	16	17	18	19
Arg	CGU	0.67	0.71	1.20	0.34	0.34	0.30	1.03	0.34	0.69
	CGC	0.89	0.18	1.20	0.34	0.34	0.00	0.21	0.69	0.51
	CGA	0.89	0.71	0.40	0.51	0.51	0.30	0.62	1.03	0.34
	CGG	0.89	1.41	2.20	1.54	1.54	0.60	1.24	1.20	0.86
	AGA	0.22	1.24	0.60	1.71	1.71	2.70	1.03	1.20	1.71
	AGG	2.44	1.76	0.40	1.54	1.54	2.10	1.86	1.54	1.89
Leu	UUA	0.27	0.35	0.31	0.60	0.60	0.52	1.20	0.00	0.31
	UUG	1.09	1.62	1.14	1.32	1.32	2.26	1.64	1.53	1.69
	CUU	2.00	1.27	0.83	1.80	1.80	1.48	1.85	0.81	0.92
	CUC	1.55	0.58	2.28	0.72	0.72	0.52	0.22	2.34	0.92
	CUA	0.55	0.81	0.10	0.84	0.72	0.35	0.44	0.31	0.31
	CUG	0.55	1.38	1.34	0.72	0.84	0.87	0.65	1.02	1.85
Ser	**UCU**	1.70	1.38	0.65	1.53	1.53	1.36	1.63	1.27	1.69
	UCC	1.43	1.13	2.61	0.61	0.61	0.82	0.81	1.50	0.94
	UCA	1.04	1.00	0.52	1.42	1.42	2.05	1.83	0.46	0.94
	UCG	0.78	0.50	1.17	0.41	0.41	0.27	0.00	1.27	0.38
	AGU	0.39	0.75	0.52	0.81	0.81	0.41	0.71	0.58	0.56
	AGC	0.65	1.25	0.52	1.22	1.22	1.09	1.02	0.92	1.50
Ala	**GCU**	1.47	1.60	1.20	0.98	0.98	1.05	1.64	1.45	1.50
	GCC	0.84	0.71	1.47	1.17	1.17	1.05	1.13	1.53	0.88
	GCA	0.77	1.51	0.27	1.85	1.85	1.58	0.82	0.26	1.25
	GCG	0.91	0.18	1.07	0.00	0.00	0.32	0.41	0.77	0.38
Gly	GGU	0.89	0.85	0.84	1.16	1.16	1.69	0.63	0.00	0.77
	GGC	1.04	0.85	1.37	0.77	0.77	0.15	0.38	2.26	0.46
	GGA	0.30	1.21	0.63	0.90	0.90	0.77	1.50	0.87	0.92
	GGG	1.78	1.09	1.16	1.16	1.16	1.38	1.50	0.87	1.85
Pro	CCU	0.85	1.28	0.65	1.60	1.55	1.91	1.47	1.02	1.38
	CCC	1.00	0.64	1.16	0.53	0.65	0.35	0.27	1.21	0.55
	CCA	1.38	1.60	0.39	1.73	1.68	1.22	2.00	0.74	1.66
	CCG	0.77	0.48	1.81	0.13	0.13	0.52	0.27	1.02	0.41
Thr	ACU	1.33	1.28	0.63	1.64	1.71	2.00	1.33	0.70	0.75
	ACC	1.71	1.28	1.26	0.55	0.38	0.86	1.00	2.61	1.00
	ACA	0.38	1.12	0.63	1.64	1.71	1.14	1.33	0.00	1.75
	ACG	0.57	0.32	1.47	0.18	0.19	0.00	0.33	0.70	0.50
Val	**GUU**	1.00	1.57	0.47	1.88	1.88	1.60	1.89	1.47	1.06
	GUC	0.80	0.57	1.41	0.35	0.35	0.40	0.67	1.68	0.71
	GUA	0.60	0.29	0.35	0.71	0.71	0.60	0.33	0.11	0.71
	GUG	1.60	1.57	1.76	1.06	1.06	1.40	1.11	0.74	1.53
Ile	**AUU**	1.50	1.50	1.00	1.10	1.10	1.22	1.20	0.68	1.50
	AUC	0.70	0.75	1.00	0.90	0.90	0.78	1.00	2.18	1.00
	AUA	0.80	0.75	1.00	1.00	1.00	1.00	0.80	0.14	0.50
Asn	AAU	1.13	1.14	0.76	1.00	1.00	1.50	1.05	0.50	0.82
	AAC	0.87	0.86	1.24	1.00	1.00	0.50	0.95	1.50	1.18
Asp	**GAU**	1.31	1.13	0.92	1.40	1.40	1.47	1.36	0.74	1.13
	GAC	0.69	0.87	1.08	0.60	0.60	0.53	0.64	1.26	0.87
Cys	UGU	0.67	0.88	0.80	1.33	1.33	1.67	1.23	0.80	1.00
	UGC	1.33	1.13	1.20	0.67	0.67	0.33	0.77	1.20	1.00
Gln	CAA	0.73	0.88	1.10	1.00	1.00	1.48	1.12	0.77	0.82
	CAG	1.27	1.12	0.90	1.00	1.00	0.52	0.88	1.23	1.18
Glu	GAA	0.81	0.97	0.79	1.20	1.20	1.13	1.11	0.97	0.91
	GAG	1.19	1.03	1.21	0.80	0.80	0.87	0.89	1.03	1.09
His	CAU	1.00	0.62	0.95	1.00	1.00	1.09	0.88	0.71	1.29
	CAC	1.00	1.38	1.05	1.00	1.00	0.91	1.13	1.29	0.71
Lys	AAA	0.50	1.00	1.09	0.92	0.92	1.11	0.84	0.94	0.76
	AAG	1.50	1.00	0.91	1.08	1.08	0.89	1.16	1.06	1.24
Phe	UUU	1.00	0.73	0.10	1.57	1.57	1.29	1.11	0.42	1.29
	UUC	1.00	1.27	1.90	0.43	0.43	0.71	0.89	1.58	0.71
Tyr	UAU	0.89	1.05	0.15	0.88	0.88	1.64	1.29	0.29	0.92
	UAC	1.11	0.95	1.85	1.13	1.13	0.36	0.71	1.71	1.08

"1-19" represent the "*BkGRAS1* to *BkGRAS19*" genes, respectively. Underlined values represent codons for which the RSCU (relative synonymous codon usage) values of all genes were higher than 1.0. Bold text indicates that RSCU values were higher than 1.0 in most *BkGRAS* genes. Red values indicate that there was no codon usage bias in certain amino acids of certain genes.

Table 8. Base composition and effective codon number of GRAS genes in *Betula kirghisorum*.

Gene name	ENC	A3s	T3s	G3s	C3s	GC3s	GC
BkGRAS1	52.13	30.4%	40.3%	28.2%	25.4%	42.1%	45.9%
BkGRAS2	52.35	30.7%	40.2%	27.7%	25.6%	41.9%	45.9%
BkGRAS3	55.73	26.9%	34.4%	39.0%	26.7%	50.4%	49.1%
BkGRAS4	52.76	30.1%	39.8%	28.7%	25.5%	42.6%	46.1%
BkGRAS5	55.97	24.0%	38.1%	30.6%	30.8%	48.8%	49.8%
BkGRAS6	46.16	13.3%	16.2%	52.9%	39.7%	75.2%	58.5%
BkGRAS7	50.81	17.2%	22.9%	42.8%	40.4%	66.8%	57.1%
BkGRAS8	53.00	17.4%	20.8%	45.0%	39.9%	68.5%	57.8%
BkGRAS9	58.58	32.0%	36.9%	33.7%	26.6%	45.7%	46.3%
BkGRAS10	56.64	28.7%	36.9%	33.1%	26.3%	46.4%	48.1%
BkGRAS11	55.81	22.5%	36.1%	36.2%	30.6%	52.2%	51.0%
BkGRAS12	57.76	30.5%	35.5%	31.9%	28.8%	47.0%	47.9%
BkGRAS13	51.96	17.2%	23.2%	37.6%	43.8%	66.2%	58.2%
BkGRAS14	54.74	35.7%	40.7%	25.6%	24.1%	38.5%	45.0%
BkGRAS15	54.64	35.5%	40.7%	25.9%	24.1%	38.7%	45.1%
BkGRAS16	49.63	35.7%	40.8%	30.3%	18.9%	38.1%	43.0%
BkGRAS17	54.51	33.1%	40.3%	27.1%	23.7%	39.8%	45.5%
BkGRAS18	49.78	17.2%	24.1%	32.8%	48.5%	65.9%	56.0%
BkGRAS19	56.96	28.5%	34.0%	36.7%	26.7%	49.2%	49.4%

Bold represents genes that prefer codons ending with G and C.

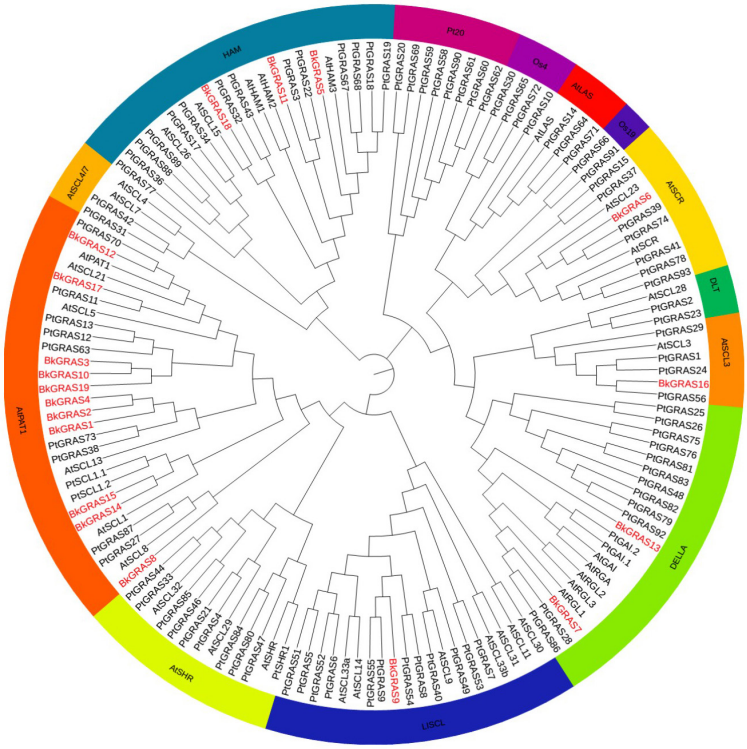

Figure 3. Combined phylogenetic analysis of GRAS proteins from *Betula kirghisorum*, *Populus*, and *Arabidopsis*.

DISCUSSION

In the present study, approximately 66,467 and 53,517 unigenes were obtained from libraries generated from salt-stressed and unstressed plants, respectively, in which there were 64,494 and 51,908 singletons. The mean length of unigenes in the two libraries was 592 and 636 bp, respectively. When the libraries were assembled, we obtained 60,545 all-unigenes, with a mean length of 758 bp. Analysis of differential expression of unigenes indicated that 12,556 and 9275 identified genes were up- or down-regulated, respectively, in samples stressed for 48 h or unstressed. A total of 7206 differentially expressed genes were differentially expressed in unigenes, and these genes were involved in 127 pathways. Most differential expression genes involved in various biological pathways, such as metabolism, signal transduction, and transcriptional regulation. DEGseq can also be applied to identify differential expression of exons or pieces of transcripts. Users can define their own 'genes' and compare the expression difference of these 'genes' using DEGseq by simply providing their own annotation files (Wang et al., 2010). As shown in the **Table S1**, the number of DEG genes increased significantly in the metabolic pathways (1811, 25.13%), biosynthesis of secondary metabolites (963, 13.36%), plant-pathogen interactions (734, 10.19%), and plant hormone signal transduction (531, 7.37%) after salt stress treatment, indicating that these genes might be specifically expressed during salt stress and might play an important role in adaptation to salt stress. In addition, 13 transcription factor families related to adverse stress conditions were identified. Transcription factors are widely studied in the modern biological sciences and represent an important aspect of functional genomics research. GRAS transcription factors play important roles in plant growth, development, and signal transduction, and in biological and abiotic stress responses. Intensively studying the characteristics, classification, structural domain, chromosome distribution, and phylogenetic tree of GRAS proteins in plants helps to further understand and study their function. Recently, increasing number of studies on the GRAS gene family in plants has been published. For example, the GRAS gene family has been reported in *Arabidopsis*, rice (Tian et al., 2004), grape (Sun et al., 2011), cabbage (Song et al., 2014), physic nut (*Jatropha curcas* L.) (Wu et al., 2015), and tobacco (Chen et al., 2015); however, there have been no reports on the GRAS gene family in *B. kirghisorum*. Therefore, this study investigated 19 *BkGRAS* genes obtained through transcriptome sequencing. Our result indicated that both the number of amino acids and the molecular weight of the 19 BkGRAS proteins varied significantly, with the number of amino acids ranging from 408 to 1029 and the molecular weight from 45.90 to 111.93 kDa. The isoelectric points of these proteins ranged from acidic to alkaline, and all proteins were unstable and hydrophilic. The 19 BkGRAS proteins possessed no transmembrane domains, were not secreted, and possessed secondary structures consisting mainly of alpha helices and random coils. Domain analysis indicated that all proteins included typical GRAS domains: LHR I; VHIID motif; LHR II; PFYRE motif; and SAW motif.

Proteins located in different parts of eukaryotic cells possess different functions, and protein function is closely related to subcellular localization. Therefore, subcellular localization of proteins is essential for studying protein function. In the plant GRAS gene family, fusion of green fluorescent proteins with GRAS proteins, such as MtSymSCL1 (Kim and Nam, 2013), NtGRAS1 (Czikkel and Maxwell, 2007), AtRGA (Silverstone et al., 1998), AtRGL1 (Wen and Chang, 2002), AtSCL14 (Fode et al., 2008), OsCIGR1 and OsCIGR2

(Day et al., 2003), OsGAI (SLR1) (Ogawa et al., 2000), and LiSCL (Morohashi, et al., 2003) generated green fluorescence in the nucleus, consistent with the functional characteristics of transcription factors. The predicted subcellular location of BkGRAS1, 2, 5, 7, 8, 9, 13, and 17 proteins was the nucleus, which is consistent with the known characteristics of transcription factors. In addition, BkGRAS2 was predicted to be localized in the cytoplasm. In *Arabidopsis*, the GRAS proteins AtSCL13, AtSCL21, and AtPAT1 were also localized in the nucleus and cytoplasm (Torres-Galea et al., 2006, 2013). In the present study, the BkGRAS3, 4, 11, 16, and 19 proteins were localized to the cytoplasm; therefore, these may be GRAS transcription factors. However, there is no literature to support the localization of GRAS transcription factors in the chloroplast, and further study is needed to determine whether the BkGRAS6, 10, 12, 14, 15, and 18 proteins are located in the chloroplast and if they are true members of the GRAS gene family.

The preferred codons, including AGG, AGA, UCU, GCU, GGG, CCA, GUU, GUG, AUU, GAU, and AAG, were used by most BkGRAS genes. However, only the coding region of BkGRAS17 had a strong usage bias for the UUA codon. Values for A3s, T3s, G3s, C3s, and GC3s indicated that the BkGRAS6, 7, 8, 11, 13, and 18 genes preferred to use codons ending in G or C, whereas codons ending in A or T were favored by the rest of the BkGRAS genes. ENC values of all BkGRAS genes were analyzed, and the results showed that the majority of BkGRAS genes had weak codon bias. In conclusion, a series of comprehensive analyses of synonymous codon usage patterns provided a basic understanding of the mechanisms underlying codon usage in *B. kirghisorum*. This information could be helpful in the future selection of appropriate expression systems or for the improved expression of foreign genes via codon reconstruction.

GRAS proteins of *Populus* can be divided into 13 subfamilies: LISCL, AtSHR, AtPAT1, AtSCR, AtSCL4/7, AtLAS, Os19, HAM, Os4, Pt20, DLT, AtSCL3, and DELLA. Of these, Os4, Os19, and Pt20 subfamilies were newly identified. Of note, Pt20 represented a *Populus*-specific subfamily (Liu and Widmer, 2014). Although, *Populus* and *B. kirghisorum* are woody plants, no members in *B. kirghisorum* were found in the Pt20 subfamily. As shown in Figure 3, 19 BkGRAS proteins were divided into six subfamilies: SCR, LISCL, SCL3, DELLA, HAM, and PAT1.

Taken together, these results for *B. kirghisorum* GRAS genes provide an understanding for exploring the relationship between different members of this gene family. In addition, gene cloning, expression analysis, and transgenic experiments are needed to further validate the biological function of these genes. To date, we have successfully cloned the *BkGRAS2* gene, which was found to respond to 0.6% $NaHCO_3$ stress (Li et al., 2015). Assays overexpressing the *B. kirghisorum BkGRAS2* gene in *Arabidopsis* are underway, which will help to further understand the biological functions of the *BkGRAS2* gene.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the Fundamental Research Funds for the Central Universities (#DL12CA13).

REFERENCES

Audic S and Claverie JM (1997). The significance of digital gene expression profiles. *Genome Res.* 7: 986-995.

Benfey PN, Linstead PJ, Roberts K, Schiefelbein JW, et al. (1993). Root development in *Arabidopsis*: four mutants with dramatically altered root morphogenesis. *Development* 119: 57-70.

Benjamini Y and Yekutieli D (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29: 1165-1188.

Chandler PM, Marion-Poll A, Ellis M and Gubler F (2002). Mutants at the Slender1 locus of barley cv Himalaya. Molecular and physiological characterization. *Plant Physiol.* 129: 181-190. http://dx.doi.org/10.1104/pp.010917

Chen YQ, Tai SS, Wang DW, Ding AM, et al. (2015). Homology-based analysis of the GRAS gene family in tobacco. *Genet. Mol. Res.* 14: 15188-15200. http://dx.doi.org/10.4238/2015.November.25.7

Czikkel BE and Maxwell DP (2007). NtGRAS1, a novel stress-induced member of the GRAS family in tobacco, localizes to the nucleus. *J. Plant Physiol.* 164: 1220-1230. http://dx.doi.org/10.1016/j.jplph.2006.07.010

Day RB, Shibuya N and Minami E (2003). Identification and characterization of two new members of the GRAS gene family in rice responsive to N-acetylchitooligosaccharide elicitor. *Biochim. Biophys. Acta* 1625: 261-268. http://dx.doi.org/10.1016/S0167-4781(02)00626-7

Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pysh L, et al. (1996). The *SCARECROW* gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. *Cell* 86: 423-433. http://dx.doi.org/10.1016/S0092-8674(00)80115-4

Fode B, Siemsen T, Thurow C, Weigel R, et al. (2008). The *Arabidopsis* GRAS protein SCL14 interacts with class II TGA transcription factors and is essential for the activation of stress-inducible promoters. *Plant Cell* 20: 3122-3135. http://dx.doi.org/10.1105/tpc.108.058974

Hirsch S and Oldroyd GED (2009). GRAS-domain transcription factors that regulate plant development. *Plant Signal. Behav.* 4: 698-700. http://dx.doi.org/10.4161/psb.4.8.9176

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. http://dx.doi.org/10.1007/s00705-008-0260-1

Kim GB and Nam YW (2013). A novel GRAS protein gene MtSymSCL1 plays a role in regulating nodule number in Medicago truncatula. *Plant Growth Regul.* 71: 77-92. http://dx.doi.org/10.1007/s10725-013-9814-7

Li GY, Yang CJ and Liu GJ (2015). Cloning and expression pattern analysis of *BkGRAS2* from *Betula kirghisorum. Genet. Mol. Res.* 14: 11335-11347. http://dx.doi.org/10.4238/2015.September.22.27

Li J, Huang Y, Liu Z, Zhu Q, et al. (2014). Codon bias analysis of Kunitz protease inhibitors. *Biotechnology* 24: 39-44.

Li X, Qian Q, Fu Z, Wang Y, et al. (2003). Control of tillering in rice. *Nature* 422: 618-621. http://dx.doi.org/10.1038/nature01518

Lim J, Helariutta Y, Specht CD, Jung J, et al. (2000). Molecular analysis of the *SCARECROW* gene in maize reveals a common basis for radial patterning in diverse meristems. *Plant Cell* 12: 1307-1318. http://dx.doi.org/10.1105/tpc.12.8.1307

Liu X and Widmer A (2014). Genome-wide comparative analysis of the GRAS gene family in *Populus*, *Arabidopsis* and Rice. *Plant Mol. Biol. Report.* 32: 1129-1145. http://dx.doi.org/10.1007/s11105-014-0721-5

Ma H, Xia X and Yin W (2011). Cloning and analysis of *SCL7* gene from *Populus euphratica*. *J. Beijing Forest. Univ* 33: 1-10.

Morohashi K, Minami M, Takase H, Hotta Y, et al. (2003). Isolation and characterization of a novel GRAS gene that regulates meiosis-associated gene expression. *J. Biol. Chem.* 278: 20865-20873. http://dx.doi.org/10.1074/jbc.M301712200

Mortazavi A, Williams BA, McCue K, Schaeffer L, et al. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5: 621-628. http://dx.doi.org/10.1038/nmeth.1226

Ogawa M, Kusano T, Katsumi M and Sano H (2000). Rice gibberellin-insensitive gene homolog, OsGAI, encodes a nuclear-localized protein capable of gene activation at transcriptional level. *Gene* 245: 21-29. http://dx.doi.org/10.1016/S0378-1119(00)00018-4

Peng J, Carol P, Richards DE, King KE, et al. (1997). The *Arabidopsis GAI* gene defines a signaling pathway that negatively regulates gibberellin responses. *Genes Dev.* 11: 3194-3205. http://dx.doi.org/10.1101/gad.11.23.3194

Pysh LD, Wysocka-Diller JW, Camilleri C, Bouchez D, et al. (1999). The GRAS gene family in *Arabidopsis*: sequence characterization and basic expression analysis of the SCARECROW-LIKE genes. *Plant J.* 18: 111-119. http://dx.doi.org/10.1046/j.1365-313X.1999.00431.x

Sharp PM and Li WH (1986). An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* 24: 28-38. http://dx.doi.org/10.1007/BF02099948

Silverstone AL, Ciampaglio CN and Sun T (1998). The *Arabidopsis RGA* gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway. *Plant Cell* 10: 155-169. http://dx.doi.org/10.1105/tpc.10.2.155

Solé A, Sánchez C, Vielba JM, Valladares S, et al. (2008). Characterization and expression of a *Pinus radiata* putative ortholog to the *Arabidopsis SHORT-ROOT* gene. *Tree Physiol.* 28: 1629-1639. http://dx.doi.org/10.1093/treephys/28.11.1629

Song XM, Liu TK, Duan WK, Ma QH, et al. (2014). Genome-wide analysis of the GRAS gene family in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). *Genomics* 103: 135-146. http://dx.doi.org/10.1016/j.ygeno.2013.12.004

Stuurman J, Jäggi F and Kuhlemeier C (2002). Shoot meristem maintenance is controlled by a GRAS-gene mediated signal from differentiating cells. *Genes Dev.* 16: 2213-2218. http://dx.doi.org/10.1101/gad.230702

Sun X, Wang C, Fang J, Mu Q, et al. (2011). Bioinformatics analysis of GRAS gene family in Grapevine. *Acta Agric. Jiangxi* 23: 1-8.

Tian C, Wan P, Sun S, Li J, et al. (2004). Genome-wide analysis of the GRAS gene family in rice and *Arabidopsis. Plant Mol. Biol.* 54: 519-532. http://dx.doi.org/10.1023/B:PLAN.0000038256.89809.57

Torres-Galea P, Huang LF, Chua NH and Bolle C (2006). The GRAS protein SCL13 is a positive regulator of phytochrome-dependent red light signaling, but can also modulate phytochrome A responses. *Mol. Genet. Genomics* 276: 13-30. http://dx.doi.org/10.1007/s00438-006-0123-y

Torres-Galea P, Hirtreiter B and Bolle C (2013). Two GRAS proteins, SCARECROW-LIKE21 and PHYTOCHROME A SIGNAL TRANSDUCTION1, function cooperatively in phytochrome A signal transduction. *Plant Physiol.* 161: 291-304. http://dx.doi.org/10.1104/pp.112.206607

Wang L, Feng Z, Wang X, Wang X, et al. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26: 136-138. http://dx.doi.org/10.1093/bioinformatics/btp612

Wen CK and Chang C (2002). *Arabidopsis* RGL1 encodes a negative regulator of gibberellin responses. *Plant Cell* 14: 87-100. http://dx.doi.org/10.1105/tpc.010325

Wis'niewska A, Pietraszewska-Bogiel A, Zuzga S, Tagashira N, et al. (2013). Molecular characterization of SCARECROW(CsSCR) gene expressed during somatic embryo development and in root of cucumber (*Cucumis sativus* L.). *Acta Physiol. Plant.* 35: 1483-1495. http://dx.doi.org/10.1007/s11738012-1189-2

Wright F (1990). The 'effective number of codons' used in a gene. *Gene* 87: 23-29. http://dx.doi.org/10.1016/0378-1119(90)90491-9

Wu ZY, Wu PZ, Chen YP, Li MR, et al. (2015). Genome-wide analysis of the GRAS gene family in physic nut (*Jatropha curcas* L.). *Genet. Mol. Res.* 14: 19211-19224. http://dx.doi.org/10.4238/2015.December.29.31

Yang C, Liu G, Zhang X, Chen J, et al. (2013). Screening of salt tolerant birch family under salt stress. *Jiangsu Agric. Sci.* 41: 139-142.

Supplementary material

Table S1. Significant enrichment of different unigenes.