DOI: 10.1007/s10535-016-0693-4 BIOLOGIA PLANTARUM 61 (3): 473-482, 2017

Soybean *NAC* gene family: sequence analysis and expression under low nitrogen supply

X. WANG[1,2]*, D. LI[1,2], J. JIANG[1,2], Z. DONG[1], and Y. MA[1]

Department of Agronomy Sciences, Anhui Agricultural University, Hefei, 230036, P.R. China[1]
Key Laboratory of Wheat Biology and Genetic Improvement on South Yellow & Huai River Valley, Ministry of Agriculture, Hefei, 230036, P.R. China[2]

Abstract

NAM, ATAF1/2, and CUC2 (NAC) proteins are plant-specific transcription factors playing essential roles in plant development and various abiotic stress responses. In the present study, we identified 173 full-length *NAC* genes in soybean, which were phylogenetically clustered into 15 groups (*NACa* - *NACo*). The soybean *NAC* genes (*GmNACs*) were non-randomly located across the 20 chromosomes, and 128 genes (86.5 %) were preferentially located in duplicated regions of chromosome arms, which implied long segmental duplication and contributed to evolution of the *GmNAC* gene family. Most *GmNACs* genes showed a distinct tissue-specific expression pattern and the redundant expression patterns of active duplicate genes suggested that *GmNACs* have been retained by substantial subfunctionalization during soybean evolution. Furthermore, active *GmNACs* genes that had undergone strong artificial selection during soybean domestication were identified based on selection analysis. After low nitrogen treatment, enhanced expression of some selected *GmNAC* genes were noticed in soybean shoot and root, which implied that *GmNACs* might play an important role in nitrogen metabolism. Here, we summarize the sequence and expression analysis of the *NAC* gene family in the soybean.

Additional key words: chromosome location, evolution, *Glycine max*, nitrogen metabolism, transcription factors.

Introduction

Transcription factors (TFs) refer to a group of proteins controlling the rate of transcription of the target genes by binding to specific regulatory promoter elements in the genome (Tucker *et al*. 2002, Chuck *et al*. 2014). Identification and functional characterization of TFs are essential for understanding the regulation of gene expression and the reconstruction of intricate transcriptional regulatory networks (Udvardi *et al*. 2007). The plant-specific NAC [no apical meristem (NAM), *Arabidopsis* transcription activation factor (ATAF1/2), and cup-shaped cotyledons (CUC2)] proteins, one of the largest plant-specific TFs, play crucial roles in several biological processes, such as plant growth, development, and abiotic stress-responsive networks (Fang *et al*. 2008, Jensen *et al*. 2008, Gao *et al*. 2010, Han *et al*. 2012, Huang *et al*. 2012, Fan *et al*. 2015). These proteins commonly have a conserved NAC domain at the N-terminus, which has been implicated in nuclear localization, DNA binding, and the formation of dimers (Duval *et al*. 2002, Ernst *et al*. 2004, Olsen *et al*. 2005). NAC domain contained a unique TF fold consisting of a twisted β-sheet bounded by a few helical elements rather than a classical helix-turn-helix motif (Ernst *et al*. 2004). Moreover, some studies indicated that most of the NAC

Submitted 21 December 2015, *last revision* 11 August 2016, *accepted* 12 August 2016.

Abbreviations: AREB - ABA responsive element binding protein; ATAF - *Arabidopsis* transcription activation factor; CUC - cup-shaped cotyledon; Mya - milliard year ago; NAC - NAM, ATAF1/2 and CUC2; NAM - no apical meristem; SNP - single nucleotide polymorphism; TF - transcription factor; WRKY - WRKY DNA-binding protein.

Acknowledgments: This study was supported by grants from the Natural Science for Key Project of Anhui Education Committee (No. KJ2016A843), the Cultivating Academic Backbone Fundation of Anhui Agricultural University (No. 2014XKPY-04), the Introduced Leading Talent Research Team for Universities in Anhui Province and the Special Fund for Agro-scientific Research in the Public Interest (No. 201503121-02). We wish to thank the anonymous reviewers for their helpful comments on this manuscript.

* Corresponding author; e-mail: wxbphd@163.com

proteins are dependent on the C-terminal region for transactivational activity (Xie *et al.* 2000, Duval *et al.* 2002, Taoka *et al.* 2004, Olsen *et al.* 2005).

The extensive studies have been performed on the identification analysis of the *NAC* gene family in plant species including *Arabidopsis* (Ooka *et al.*, 2003), rice (Nuruzzaman *et al.* 2010), poplar (Hu *et al.* 2010), soybean (Le *et al.* 2011), grapevine (Wang *et al.* 2013), pigeon pea (Satheesh *et al.* 2014), foxtail millet (Puranik *et al.* 2013), Chinese cabbage (Liu *et al.* 2014), physic nut (Wu *et al.* 2015), and banana (Cenci *et al.* 2014). NAC proteins have been implicated to participate in diverse processes in plants, including root development (Hao *et al.* 2011), leaf and fruit senescence (Guo and Gan 2006, Kou *et al.* 2012), flower morphogenesis (Hendelman *et al.* 2013), hormone signalling (Bu *et al.* 2008, Dong *et al.* 2014), secondary wall formation (Mitsuda and Ohme-Takagi 2008, Dong *et al.* 2014), and biotic/abiotic stress responses (Chen *et al.* 2013, Li *et al.* 2014, Movahedi *et al.* 2015). According to the previous studies (Sakuraba *et al.* 2015, Wang *et al.* 2015), the *NAC* gene is considered as a negative factor in plant stress response and regulation. For example, in *Arabidopsis*, *NAC016-overexpressing* (*NAC016-OX*) plants show low drought tolerance, while the *NAC016* mutants show a very high drought tolerance (Sakuraba *et al.* 2015). For the mechanism, *NAC016* could repress ABRE binding protein 1 (*AREB1*) transcription by binding directly to the *NAC16BM* in the *AREB1* promoter that encoded a central TF in the stress-responsive abscisic acid signalling pathway (Sakuraba *et al.* 2015). The soybean *NAC29* is also a negative factor of stress tolerance as indicated by the performance of transgenic soybean hairy roots under drought stress (Wang *et al.* 2015). Herein, the interaction between *GmWRKY27* and *GmMYB174* could suppress *GmNAC29* expression and enhance drought stress tolerance. The authors explained that *GmWRKY27* and *GmMYB174* may have evolved to bind to neighbouring *cis* elements in the *GmNAC29* promoter to co-reduce promoter activity and gene expression.

Nowadays, there are only limited studies on the functional characterization of *NAC* TFs in soybean. Pinheiro *et al* (2009) reported three *GmNAC* genes (*i.e.* *GmNAC1*, *GmNAC5*, and *GmNAC6*) involved in cell death program, based on the transient expression in tobacco leaves (Pinheiro *et al.* 2009). Besides, Le *et al.* (2011) found the expression of some soybean *NACs* genes was closely related to the response to drought stress in roots and shoots. Recently, a *NAC* gene (*SHAT1-5*), that underwent a strong artificial selection during soybean domestication, was proved to activate secondary wall biosynthesis and contribute to the significant thickening of secondary walls in fiber cap cells (Dong *et al.* 2014). *SHAT1-5* is functionally similar to two *AtNAC* TFs (NST1 and NST3) that are involved in regulating pod shattering in *Arabidopsis* (Mitsuda and Ohme-Takagi 2008). Completion of the soybean genome, RNA-seq and resequencing data greatly contributed to the identification of gene families and tracing gene domestication history at the whole-genome level (Lam *et al.* 2010, Schmutz *et al.* 2010, Li *et al.* 2013). However, no detailed analysis including genome organization, evolutionary analysis of *GmNAC* TFs has been conducted up to now.

The cultivars of soybean containing abundant proteins beneficial to human health are highly dependent on the nitrogen nutrition. They could obtain N through N fixation process with rhizobia, but also take up residual and mineralized N from the soil. Traditionally, soybean has been grown without addition of N fertilizer, and N is the limiting factor for soybean yield in today's agricultural production systems especially in developing countries (Salvagiotti *et al.* 2008). In this study, a genome-wide identification of soybean *NAC* domain was carried out to gain a comprehensive understanding of soybean *NAC* gene family and its role in soybean growth, development, and evolution. In addition, we analyzed the *GmNAC* gene sequence phylogeny, chromosomal location, tissue-specific expression patterns, and the expression compendium under low nitrogen stress. Moreover, the evolutionary and selective effects of *GmNAC* genes during soybean domestication and some of the active *GmNAC* genes undergone strong artificial selection were also investigated. Our genome-wide systematic analysis of *GmNAC* TFs contributes to the further evolutionary and nitrogen use efficiency analysis of the *NAC* gene family in soybean.

Materials and methods

Database search, sequence retrieval, and phylogenetic analysis: The *NAC* gene was retrieved from soybean genome annotation database at *Phytozome* (http://www.phytozome.net/) (Goodstein *et al.* 2012) and *SoyKB* (http://soykb.org/) (Joshi *et al.* 2014). For the incorrectly predicted genes, manual reannotation was then performed using *FGENESH* (http://linux1.softberry.com/) (Salamov and Solovyev 2000). To confirm the full-length of a NAM/NAC domain, the reannotated sequences were further manually analyzed using *NCBI* (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and *InterProScan* program (http://www.ebi.ac.uk/Tools/InterProScan/) (Quevillon *et al.* 2005). Multiple sequence alignments of the full-length *Gm*NAC protein sequences were performed using *Clustal X* software (*v. 1.83*) (Thompson *et al.* 1997). The unrooted phylogenetic trees were constructed with *MEGA 5.0* using the neighbour-joining (NJ) method and the bootstrap test carried out with 1 000 iterations (Tamura *et al.* 2011). The evolutionary distances were computed using the *p*-distance method.

Chromosome location of *GmNAC* genes: The chromosome visualization tool (*CViT*) genome search and *Synteny* viewer (http://comparative-legumes.org/) (Cannon and Cannon 2011) were used to locate the *Gm*NACs on soybean genome and identify the homologous chromosome segments resulting from whole-genome duplication events (Schmutz *et al*. 2010). Blocks of the same colour represented the homologous chromosome segments, while *GmNAC* genes separated in a range of less than 100 kb distance were considered to be tandem duplicates (Hu *et al*. 2010).

Expression analysis of *GmNAC* genes: Genome-wide transcriptome data were obtained from the *Soybase* database (http://soybase.org/), which includes vegetative tissues (*e.g*., young leaf, root, and nodule), reproductive tissues (*e.g*., flower, 1-cm pod, pod shell at 10 and 14 d after flowering), and seed at seven developmental stages (*i.e*., 10, 14, 21, 25, 28, 35, and 42 d after flowering). All transcriptome data were analyzed with *Heatmapper Plus* (http://bar.utoronto.ca/welcome.htm), *Cluster 3.0* (De Hoon *et al*. 2004), and the heat map was viewed in *Java Treeview* (1996).

Calculation of Ka/Ks values and evolutionary analysis of *GmNAC* genes: Pairwise alignments of the paralogous nucleotide sequences were performed using *Clustal X* (*v. 2.0*) (Larkin *et al*. 2007) and *EMBOSS Needle* (http://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html). Ka (non-synonymous substitution per non-synonymous site) and Ks (synonymous /silent substitution rate substitutions per synonymous site) were estimated using the *DnaSp v.5* program (Librado and Rozas 2009). The Ka/Ks ratio was used to measure the selective pressure on protein-coding genes (Hurst 2002). Divergence time (T) was calculated using the formula $T = Ks/2\lambda$, where the synonymous mutation rate λ is 6.1×10^{-9} for soybean (Lavin *et al*. 2005, Lynch and Conery 2000, Schmutz *et al*. 2010).

Single nucleotide polymorphisms (SNPs) of the *GmNAC* genes were downloaded from the *SoyKB* database (http://soykb.org/) based on the resequencing of seventeen wild and fourteen cultivated soybean genomes (Lam *et al*. 2010). The ratio of each SNP in wild and cultivated soybean populations was analyzed. The reverse distribution of SNPs in a different evolutionary type of soybeans was defined as selected sites, and the ratio of selected sites with total SNPs was used to evaluate the extent of artificial selection that *GmNAC* genes had undergone during soybean domestication (Wang *et al*. 2014a).

***Cis*-regulatory element analysis:** For the promoter analysis, 1000-bp sequences upstream from the start codon of the twelve flower specific expression *GmNAC* genes and *NAC-o* subfamily genes were retrieved. These sequences were then subjected to search in the *PlantCARE* database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot *et al*. 2002) to predict *cis*-regulatory elements.

Stress treatments and real-time quantitative PCR: Seeds of soybean (*Glycine max* L.) cv. WS01-15 were germinated in the *Vermiculite* moistened with Hoagland nutrient solution in a chamber under a 16-h photoperiod, an irradiance of 200 μmol m^{-2} s^{-1}, day/night temperatures of 25/22 °C, and a relative humidity of 60 - 70 %. The seedlings were cultivated for about one week until the first trifoliate leaves were fully developed. For low nitrogen treatment, the plants were transferred to a solution containing nitrogen decreased to 15 % of the original amount and cultivated for 3 d. The root and shoot were harvested separately, frozen in liquid nitrogen, and stored at -80 °C for further analysis. The data were presented as mean ± standard deviation (SD) and were obtained from at least three independent replicates.

Total RNA was isolated from the root and shoot using *RNAiso Plus* (*TaKaRa*, Toyoto, Japan) according to the manufacture's handbook. The cDNA synthesis and real-time qPCR were performed with *SYBR Premix ExTaq Taq™II* (*TaKaRa*) on an *ABI Prism 7000* according to the manufacturer's protocols (*Applied Biosystem*, Foster City, CA, USA). The *GmNAC* gene primers for real time qPCR were designed using the *Primer Premier 5.0* software (Table 1 Suppl.). The amplification efficiency of a qPCR reaction was analyzed based on the slope of the standard curve, and the amplification efficiency of more than 90 % was chosen for final analysis. The mRNA level was normalized by *ACT11* (Table 1 Suppl.). Reactions were performed in a total volume of 20 mm^3 containing: 10 mm^3 of 2× *SYBRPremix*, 0.4 mm^3 of each *GmNAC* gene specific primer to a final concentration of 200 nM, and 1 mm^3 of cDNA template. The PCR conditions consisted of denaturation at 95 °C for 3 min, followed by 40 cycles of denaturation at 95 °C for 30s, annealing at 58 °C for 30 s, and extension at 72 °C for 30 s. For each sample, three technical replicates were conducted to calculate the averaged Ct values. Relative expression was calculated by the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). The expression patterns were analyzed with an *ABI Prism 7000* sequence detection system (*Applied Biosystems*). Statistical analyses were performed using *SPSS 17.0* software. Student's *t*-test was used for the comparison and $P < 0.05$ was considered as statistically significant.

Results and discussion

A total of 208 putative *NAC* gene sequences were downloaded from soybean annotation databases (*Phytozome* and *SoyKB*). Among these genes, 173 full-length genes (83.2 %) with a complete *NAM/NAC* domain were used for further analysis (Table 2 Suppl.). Most genes (162/173) had a NAM domain. These *GmNAC* genes were numbered from *GmNAC01.1* to *GmNAC* 20.5 according to their localization on chromosomes. Genome-wide analyses identified a large number of *NAC* family members in other species, including 117 *NAC* genes in the *Arabidopsis thaliana* (Ooka *et al*. 2003), 151 *NAC* genes in *Oryza sativa* (Nuruzzaman *et al*. 2010), 124 *NAC* genes in *Zea mays* (Fan *et al*. 2014), 163 *NAC* genes in *Populus trichocarpa* (Hu *et al*. 2010), 147 putative *NAC* genes in *Setaria italica* (Puranik *et al*. 2013), and 145 *NAC* genes in *Gossypium raimondii* (Shang *et al*. 2013). To test the phylogenetic relationships among soybean *GmNAC* genes, an unrooted tree was constructed using alignments of the full-length amino acid sequences in all GmNAC proteins (Fig. 1 Suppl.). The *GmNAC* gene family was classified into fifteen subgroups (*NAC*a - *NAC*o) with 2 - 30 members in each subgroup, which was similar to other plants with 13 distinct subfamilies in *Zea mays* (Fan *et al.* 2014), 16 distinct subfamilies in rice (Nuruzzaman *et al.* 2010), 11 distinct subfamilies in *Setaria italica* (Puranik *et al.* 2013), and 18 distinct subfamilies in *Populus trichocarpa* (Hu *et al*. 2010). Phylogenetic tree topology revealed that 74 *GmNAC* pairs located at the terminal nodes shared high similarities. Thus, they were assigned as paralogous pairs (homologous genes that diverged by gene duplication) (Xu *et al*. 2012).

In silico mapping revealed that 173 *GmNAC* genes were distributed across all 20 soybean chromosomes. The distribution of these *GmNAC* genes appeared to be non-random, with 10 *GmNAC* genes (from 10 - 16) on 10 chromosomes *(i.e.*, 2, 4, 5, 6, 7, 8, 12, 13, 16, and 19) and 4 - 7 *GmNAC* genes on the other 10 chromosomes (*i.e,*. 1, 3, 9, 10, 11, 14, 15, 17, 18, and 20). Substantial clustering of *GmNAC* genes was evident on several chromosomes (Fig. 2 Suppl.). Interestingly, all *GmNACs* except two genes on chromosome 12 (*NAC12.7* and *NAC12.8*) were tended to distribute far from pericentromeric regions which comprised 57 % of the soybean genome accounting for only 6.9 % of recombination (Du *et al.* 2015). This indicated that the *GmNAC* genes were distributed in a region that might be apt to develop recombination.

Table 1. Selected sites of *NAC* genes during soybean domestication.

Name	Chromosome	Position	Wild	Cultivar
GmNAC19.4	Gm19	36057148	5T/10C	13T/0C
GmNAC16.5	Gm16	3992171	9T/8C	14T/0C
	Gm16	3992342	9G/7C	14G/0C
GmNAC20.4	Gm20	41983148	7T/8C	10T/4C
	Gm20	41984019	1G/16C	8G/6C
	Gm20	41984098	4A/13G	8A/6G
	Gm20	41984535	1A/16C	7A/6C
	Gm20	41985716	8G/8C	3G/10C
	Gm20	41985780	8T/7G	4T/9G
GmNAC08.12	Gm08	46359664	6A/10C	14A/0C
	Gm08	46360262	6T/8C	14T/0C
GmNAC02.7	Gm02	27361899	9T/3G	2T/6G
GmNAC13.7	Gm13	36983661	9A/6G	0A/14G
	Gm13	36984750	9T/6G	0T/14G
GmNAC12.15	Gm12	38104886	4A/13C	12A/3C
	Gm12	38104958	10T/6G	3T/12C
	Gm12	38104991	11T/6C	3T/12C
	Gm12	38105103	9A/6G	2A/11G
	Gm12	38105342	6T/11C	12T/3C
	Gm12	38106325	3T/9C	12T/3C
GmNAC17.5	Gm17	23912265	12A/1G	3A/11G
	Gm17	23912266	12A/1T	3A/11T
	Gm17	23912924	14T/2C	3T/11C
	Gm17	23913055	12A/2G	3A/9G
	Gm17	23913065	11A/5G	3A/10G
	Gm17	23913208	2G/7C	10G/2C
	Gm17	23913924	2A/15T	11A/3T
GmNAC17.5	Gm17	23913975	14T/2C	3T/11C
	Gm17	23914036	11A/2G	3A/9G
	Gm17	23914159	2A/15G	9A/2G
	Gm17	23914181	2T/15C	9T/3C
GmNAC11.5	Gm11	8695502	7A/8G	0A/13G
	Gm11	8695967	8A/8C	0A/14C
	Gm11	8696471	9T/6C	14T/0C
	Gm11	8696481	8A/7T	14A/0T
	Gm11	8696536	8A/9T	0A/14T
GmNAC11.4	Gm11	8688340	8A/6G	0A/14G
	Gm11	8688667	10A/6G	0A/14G
	Gm11	8688927	6A/10T	14A/0T
	Gm11	8689219	7T/9C	0T/14C
	Gm11	8689726	9A/6G	0A/13G
	Gm11	8690317	6T/8G	14T/0G
	Gm11	8690516	6A/9G	14A/0G
	Gm11	8690724	6A/7T	0A/13T
	Gm11	8690768	8T/6G	0T/14G
	Gm11	8690912	10G/6C	0G/14C
	Gm11	8691076	10T/6G	0T/14G
GmNAC04.10	Gm04	46565321	6A/8C	8A/5C
GmNAC03.1	Gm03	35304422	7T/10G	14T/0G
	Gm03	35306021	7A/7G	14A/0G
GmNAC03.2	Gm03	35309231	12A/4T	6A/7T
GmNAC19.5	Gm19	38159634	14A/2G	5A/9G
	Gm19	38161439	1T/15C	6T/5C

It has been well documented that the soybean genome has undergone two rounds of genome-wide duplication followed by multiple segmental duplication, tandem duplication, and transposition events, which are considered as the major forces for soybean gene-family expansion (Schlueter *et al*. 2007). In the sequence of the 1.1 Gb soybean genome, 15 632 gene pairs (almost 67.34 % of total 46 430 genes) exist as "recent" paralogs, which are believed to have been duplicated and retained after the 13-Mya tetraploidy event (Schmutz *et al*. 2010). Segmental duplications have been shown to contribute to the expansion of other gene families in soybean (Guo and Qiu 2013, Wang *et al*. 2014b). To determine the possible relationship between the paralogous pairs of *GmNAC* genes and potential segmental duplications, we mapped *GmNACs* to the duplicated blocks (Fig. 2 Suppl.). Within the identified duplicated blocks associated with a duplication event, about 86.5 % (128/148) of *GmNACs* were preferentially retained duplicates that located in duplicated regions, with 52 putative paralogous pairs located in a segmental duplication of a long fragment (> 1 Mb) and 12 located in a segmental duplication of a short fragment (< 1 Mb, Table 3 Suppl.). Meanwhile, other two putative paralogous pairs (*GmNAC19.5*/ *GmNAC19.6* and *GmNAC19.9*/*GmNAC19.10*) were formed, which were supposed to be possibly due to tandem duplication in the same orientation. In contrast, only 16 *GmNAC* genes were located outside of duplicated blocks. Based primarily on the genomic organization of *GmNAC* genes, we implied that long segmental duplication exclusively contributed to the evolution of *GmNAC* gene family, which may be associated with the short segment and tandem duplication.

A central problem in evolutionary and molecular biology is to identify factors that determine the rate of protein evolution. Proteins with rigorous function are subject to strong purifying selective pressure, which tend to evolve more slowly than proteins with weaker constraints (Liu *et al*. 2008). A classic measure for selective pressure on protein-coding genes is the Ka/Ks ratio (Hurst 2002). Ka/Ks of 66 *GmNAC* paralogous pairs are calculated and Ks was used to calculate the approximate dates of duplication events. There, two tandem duplication *GmNAC* gene pairs, both of which were distributed on chromosome 19, with the dating duplication event at 11.62 Mya (Ks = 0.1418) for *GmNAC19.5* and *GmNAC19.6* and 3.04 Mya (Ks = 0.0371) for *GmNAC19.9* and *GmNAC19.10* respectively. Meanwhile, the segmental duplications of the *GmNAC* genes in soybean was supposed to originate from 4.76 Mya (Ks = 0.0581) to 19.48 Mya (Ks = 0.2377), with a mean value of 10.79 Mya (Ks = 0.1317, Table 3 Suppl.). As the soybean genome has undergone two polyploidy events at 13 and 58 Mya, all the tandem and segmental duplications of the *GmNAC* genes occurred around 13 Mya when *Glycine*-specific duplication occurred in the soybean genome (Schmutz *et al*. 2010). The Ka/Ks ratios of 33 segmental duplication pairs and one tandem duplication pair (*GmNAC19.5/ GmNAC19.6*) were less than 0.3, while the ratios of the other 31 segmental duplication pairs and another tandem duplication pair (*GmNAC19.9/GmNAC19.10*) were more than 0.3, which demonstrated a possibility of significant functional divergence of some *GmNAC* genes after the duplication events. The Ka/Ks ratios of seven paralogous gene pair were slightly larger than 0.5 (Table 3 Suppl.). This suggests that they experienced relatively rapid evolution following duplication. It could be concluded that *GmNAC* genes have been subjected to strong purifying selection pressure with limited functional divergence after segmental duplications.

The previously publicly-available *RNA-Seq* data was considered as a useful means of highlight the expression profiles of *GmNAC* genes. From the *Soybase* database (http://soybase.org/), most of *GmNAC* genes showed a very low expression (Fig. 3 Suppl.). For example, thirteen *GmNACs* genes had no expression at 14 soybean tissues/developmental stages and 74 *GmNACs* represented by less than one read on average. The low expression of *GmNAC* genes might either transcribe at relatively low abundance to be detected or had special temporal and spatial expression patterns which cannot be easy examined in the libraries. This is consistent with the commonly low transcript abundance feature of transcription factor genes (Wilkins *et al*. 2009). Some of *GmNACs* showed a distinct tissue-specific expression pattern. For example, twelve genes (*GmNAC13.2*, *GmNAC1.3*, *GmNAC2.2*, *GmNAC2.6*, *GmNAC8.12*, *GmNAC16.6*, *GmNAC16.9*, *GmNAC6.12*, *GmNAC6.13*, *GmNAC12.8*, *GmNAC12.15*, and *GmNAC13.7*) were specifically expressed in soybean flower. Two genes (*GmNAC19.4* and *GmNAC16.5*) had a significantly higher transcript accumulation in the nodules. It is noteworthy that compared with the other 13 subfamily, the expression of most genes in subfamily *NAC-h* and *NAC-o* was relatively high and genes in subfamily *NAC-o* were the most active *GmNAC* genes of family (Fig. 3 Suppl.).

In this study, the promoter regions of 12 NAC-flower genes and *NAC-o* subfamily genes were analyzed using the *PlantCARE* database to establish the relationship between the expression pattern of gene families with the putative *cis*-elements distribution in the gene promoters. According to the *PlantCARE* results, 21 high ratio *cis*-acting regulatory elements were identified in the promoter regions of this two groups (Fig. 1). The basal regulatory elements, CAAT-box and TATA-box elements were the most abundant motifs in both gene groups. Half of these *cis*-elements (10/21) were development related, such as Box4, Box1, circadian, G-Box, GA-motif, GCN4_motif, I-box, Skn-1_motif, Sp1, and TCT-motif (Lescot *et al*. 2002). Five were abiotic stress related, such as ARE, TCA-element, HSE, MBS, and TC-rich repeats (Lescot *et al*. 2002). Four were hormone response related, such as ABRE, CGTCA-motif, TGACG-motif, and TGA-element. As expected, most of the closely-related members in the phylogenetic tree had common motif compositions. For example, three *cis*-elements (ABRE, GCN4_motif and TGA-element) were present in most *NAC*-flower genes but not in *NAC-o* subfamily, while the ratio of GA-motif in *NAC-o* subfamily was significantly higher than that in *NAC*-flower genes (Fig. 1).

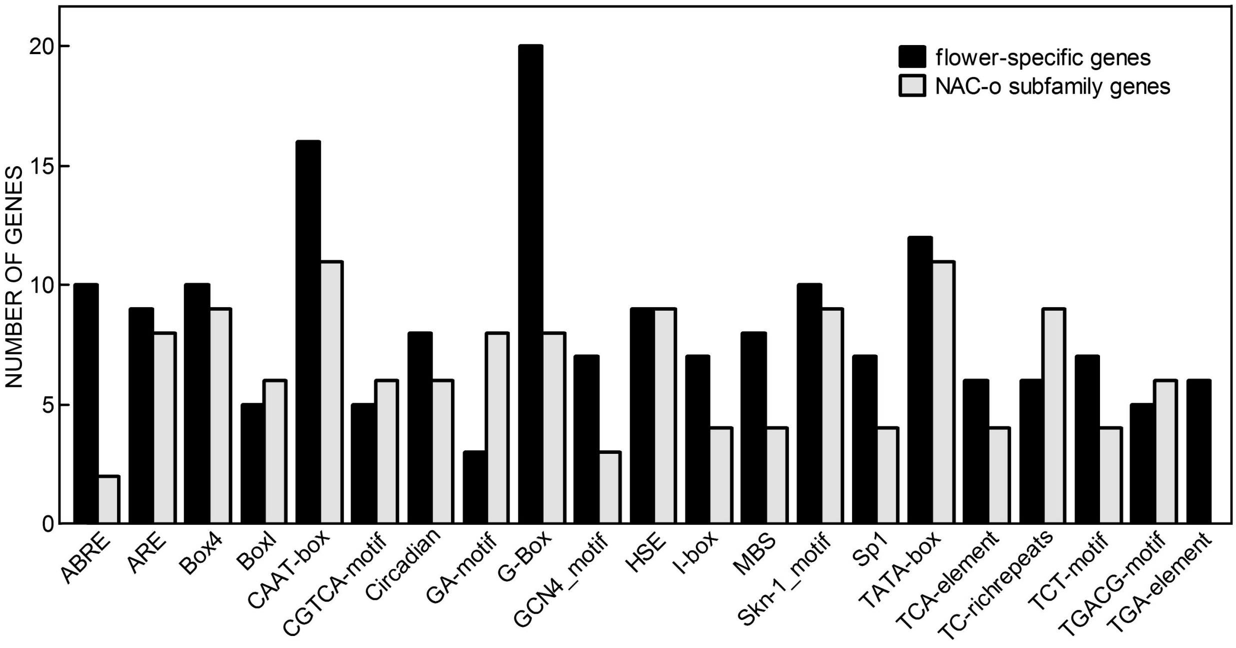

Fig. 1. The distribution of motifs in the promoter region of *NAC-o* subfamily and *NAC*-flower genes

We also investigated the functional redundancy of *GmNAC* genes with high proportion of segmental/tandem duplications with genes in subfamilies *NAC-h* and *NAC-o*. Four paralogous pairs (*GmNAC04.11/06.3*, *GmNAC04.5/06.9*, *GmNAC12.15/13.7* and *GmNAC06.12/12.8*) in subfamily *NAC-h* from segmental duplications, three paralogous pairs (*GmNAC11.5/12.4*, *GmNAC04.9/06.5* and *GmNAC04.10/06.4*) derived from segmental duplications and one paralogous pair (*GmNAC19.5/19.6*) derived from tandem duplications in subfamily *NAC-o* shared almost identical expression patterns. A paralogous pair (*GmNAC16.5/19.4*) belonging to subfamily *NAC-a*, which was specifically expressed in soybean nodule, had the same expression in roots and nodules. Therefore, *GmNAC* genes might have been retained by functional redundancy during soybean evolutionary processes.

Many agronomically important traits in soybean are associated with the domestication (Tian *et al.* 2010, Li *et al.* 2013, Dong *et al.* 2014, Zhou *et al.* 2015). Based on the expression profile, 28 genes with relatively high expression (more than 100 reads in the expression profile) were defined as active *GmNAC* genes. The variations of these genes in 17 wild soybeans and 14 cultivated soybeans were analyzed to identify the selective effects during the domestication of soybean cultivars. Half of these active *GmNAC* genes (14/28) had selected site(s), among which more than 50 % SNP sites were selected in *GmNAC11.5* (5/8) and *GmNAC17.5* (11/21). Soybean had undergone several genetic bottlenecks and most of rare sequence variants were lost as numerous allele frequency changes during its domestication (Hyten *et al.* 2006). The genetic diversity of most *GmNAC* genes in cultivars was declined notably compared with that of wild soybeans. We found that *GmNAC11.4* and *GmNAC11.5* belonging to subfamily *o* and *GmNAC19.4* belonging to subfamily *a* had only one haplotype in cultivar soybean, and the selected haplotypes are now the main genotype of cultivated soybean (Table 1). All these facts suggest that these *GmNAC* genes may have undergone a strong artificial selection during soybean domestication, which may contributed to the cultivation of soybeans.

Nitrogen is an essential nutrient for plant growth, development, and reproduction. Soybean has a higher demand for N than other crops such as rice, wheat, and maize. Although soybean could fix most of the N from atmosphere, a half of the total N is also taken from the soil (Salvagiotti *et al.* 2008). To understand the role of *GmNAC* genes in soybean development, evolution, and the improvement of the soybean nitrogen use efficiency, the expression of active *GmNAC* genes that had undergone artificial selection were analyzed under low nitrogen supply by real time qPCR. In soybean shoots, a significant up-regulation was noticed in three *GmNAC* genes, such as *GmNAC12.15*, *GmNAC13.7*, and *GmNAC17.5*, while a remarkable down-regulation was noticed in *GmNAC11.4* (Fig. 2). Compared with the baseline level, a 3-fold increase was noticed in the expression of *GmNAC12.15* and *GmNAC17.5*. In soybean root, *GmNAC16.5* showed the highest transcription, which was consistent with the tissue specific analysis (Fig. 3 Suppl.), and a remarkable down-regulation was noticed in *GmNAC08.12*, while the expression of *GmNAC12.15* was increased 2.0-fold compared with the baseline levels. *GmNAC12.15* gene was significantly induced in soybean shoots and roots under low nitrogen stress. Expression of several *NAC* genes induced by abiotic stresses has been reported to play crucial roles in the regulation of tolerance to abiotic stresses. For instance, three *Arabidopsis NAC* genes (*ANAC019*, *ANAC055*, and *ANAC072*) were up-regulated at transcription levels after drought, salinity, and abscisic acid treatment, which resulted in increased tolerance to drought (Tran *et al.* 2004). With the extensive fertilization, N content in fields increased significantly. Wild plants growing in nature are considered to be superior than cultivated species in resistance to low nitrogen stress. In future, further studies are needed to discover whether *GmNAC* genes (*GmNAC12.15*, *GmNAC13.7* and *GmNAC17.5*) contribute to improvement of soybean resistance to low nitrogen stress during soybean domestication.

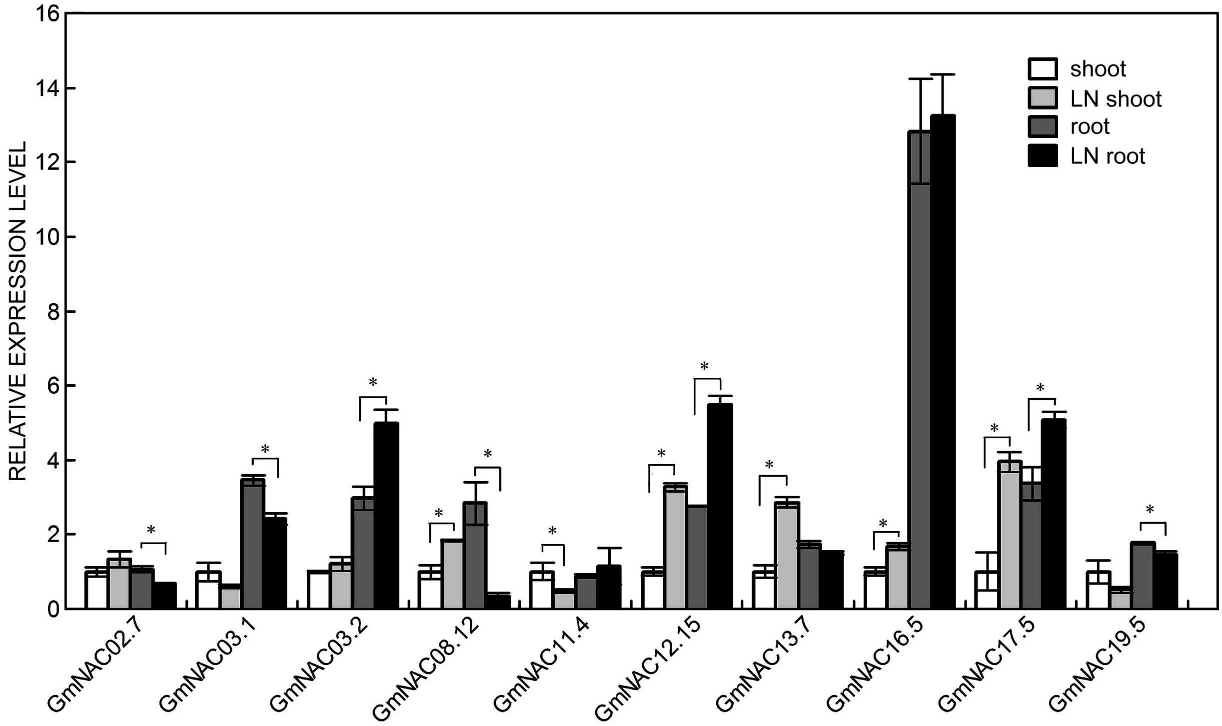

Fig. 2. Expression analysis of *GmNAC* genes in soybean shoots and roots under low nitrogen (LN) supply. Real time qPCR data were normalized using soybean *ACT11* gene. X-axis shows *GmNAC* genes that underwent strong artificial selection during soybean domestication and y-axis shows relative expression (means ± SDs, $n = 3$, * - indicates significant difference at $P < 0.05$ between the LN and control plants).

References

Bu, Q.Y., Jiang, H.L., Li, C.B., Zhai, Q.Z., Zhang, J.Y., Wu, X.Q., Sun, J.Q., Xie, Q., Li, C.Y.: Role of the *Arabidopsis thaliana* NAC transcription factors ANAC019 and ANAC055 in regulating jasmonic acid-signaled defense responses. - Cell Res. **18**: 756-767, 2008.

Cannon, E.K., Cannon, S.B.: Chromosome visualization tool: a whole genome viewer. - Int. J. Plant Genomics **2011**: 373875, 2011.

Cenci, A., Guignon, V., Roux, N., Rouard, M.: Genomic analysis of NAC transcription factors in banana (*Musa acuminata*) and definition of NAC orthologous groups for monocots and dicots. - Plant. mol. Biol. **85**: 63-80, 2014.

Chen, Y.J., Perera, V., Christiansen, M.W., Holme, I.B., Gregersen, P.L., Grant, M.R., Collinge, D.B., Lyngkjaer M.F.: The barley HvNAC6 transcription factor affects ABA accumulation and promotes basal resistance against powdery mildew. - Plant mol. Biol. **83**: 577-590, 2013.

Chuck, G.S., Brown, P.J., Meeley, R., Hake, S.: Maize SBP-box transcription factors unbranched2 and unbranched3 affect yield traits by regulating the rate of lateral primordia initiation. - Proc. nat. Acad. Sci. USA **111**: 18775-18780, 2014.

De Hoon, M.J.L., Imoto, S., Nolan, J., Miyano, S.: Open source clustering software. - Bioinformatics **20**: 1453-1454, 2004.

Dong, Y., Yang, X., Liu, J., Wang, B.H., Liu, B.L., Wang, Y.Z.: Pod shattering resistance associated with domestication is mediated by a *NAC* gene in soybean. - Nat. Commun. **5**: 3352, 2014.

Dong, Y.P., Fan, G.Q., Zhao, Z.L., Deng, M.J.: Compatible solute, transporter protein, transcription factor, and hormone-related gene expression provides an indicator of drought stress in *Paulownia fortunei*. - Funct. integr. Genomics **14**: 479-491, 2014.

Du, J.C., Tian, Z.X., Sui, Y., Zhao, M.X., Song, Q.J., Cannon, S.B., Cregan, P., Ma, J.X.: Pericentromeric effects shape the patterns of divergence, retention, and expression of duplicated genes in the paleopolyploid soybean. - Plant Cell **24**: 21-32, 2015.

Duval, M., Hsieh, T.F., Kim, S.Y., Thomas, T.L.: Molecular characterization of AtNAM: a member of the *Arabidopsis* NAC domain superfamily. - Plant mol. Biol. **50**: 237-248, 2002.

Ernst, H.A., Olsen, A.N., Larsen, S., Lo, Leggio L.: Structure of the conserved domain of ANAC, a member of the NAC family of transcription factors. - EMBO Rep. **5**: 297-303, 2004.

Fan, K., Bibi, N., Gan, S., Li, F., Yuan, S., Ni, M., Wang, M., Shen, H., Wang, X.: A novel NAP member GhNAP is involved in leaf senescence in *Gossypium hirsutum*. - J. exp. Bot. **15**: 4669-4682, 2015.

Fan, K., Wang, M., Miao, Y., Ni, M., Bibi, N., Yuan, S., Li, F., Wang, X.D.: Molecular evolution and expansion analysis of the NAC transcription factor in *Zea mays*. - PLoS ONE **9**: e111837, 2014.

Fang, Y.J., You, J., Xie, K.B., Xie, W.B., Xiong, L.Z.: Systematic sequence analysis and identification of tissue-specific or stress-responsive genes of NAC transcription factor family in rice. - Mol. Genet. Genomics **280**: 547-563, 2008.

Gao, F., Xiong, A.S., Peng, R.H., Jin, X.F., Xu, J., Zhu, B., Chen, J.M., Yao, Q.H.: OsNAC52, a rice NAC transcription factor, potentially responds to ABA and confers drought tolerance in transgenic plants. - Plant Cell Tissue Organ Cult. **100**: 255-262, 2010.

Goodstein, D.M., Shu, S.Q., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., Rokhsar, D.S.: Phytozome: a comparative platform for green plant genomics. -Nucl. Acids Res. **40**: D1178-D1186, 2012.

Guo, Y., Qiu, L.J.: Genome-wide analysis of the Dof transcription factor gene family reveals soybean-specific duplicable and functional characteristics. - PloS ONE **8**: e76809, 2013.

Guo, Y.F., Gan, S.S.: AtNAP, a NAC family transcription factor, has an important role in leaf senescence. - Plant J. **46**: 601-612, 2006.

Han, Q.Q., Zhang, J.H., Li, H.X., Luo, Z.D., Ziaf, K., Ouyang, B., Wang, T.T., Ye Z.B.: Identification and expression pattern of one stress-responsive *NAC* gene from *Solanum lycopersicum*. - Mol. Biol. Rep. **39**: 1713-1720, 2012.

Hao, Y.J., Wei, W., Song, Q.X., Chen, H.W., Zhang, Y.Q., Wang, F., Zou, H.F., Lei, G., Tian, A.G., Zhang, W.K., Ma, B., Zhang, J.S., Chen, S.Y.: Soybean NAC transcription factors promote abiotic stress tolerance and lateral root formation in transgenic plants. - Plant J. **68**: 302-313, 2011.

Hendelman, A., Stav, R., Zemach, H., Arazi, T.: The tomato NAC transcription factor SlNAM2 is involved in flower-boundary morphogenesis. - J. exp. Bot. **64**: 5497-5507, 2013.

Hu, R.B., Qi, G.A., Kong, Y.Z., Kong, D.J., Gao, Q.A., Zhou, G.K.: Comprehensive analysis of NAC domain transcription factor gene family in *Populus trichocarpa*. - BMC Plant Biol. **10**: 145, 2010.

Huang, H., Wang, Y., Wang, S.L., Wu, X., Yang, K., Niu, Y.J., Dai, S.L.: Transcriptome-wide survey and expression analysis of stress-responsive *NAC* genes in *Chrysanthemum lavandulifolium*. -Plant Sci. **193**: 18-27, 2012.

Hurst, L.D.: The Ka/Ks ratio: diagnosing the form of sequence evolution. - Trends Genet. **18**: 486-487, 2002.

Hyten, D.L., Song, Q.J., Zhu, Y.L., Choi, I.Y., Nelson, R.L., Costa, J.M., Specht, J.E., Shoemaker, R.C., Cregan, P.B.: Impacts of genetic bottlenecks on soybean genome diversity. - Proc. nat. Acad. Sci. USA **103**: 16666-16671, 2006.

Jensen, M.K., Hagedorn, P.H., De Torres-Zabala, M., Grant, M.R., Rung, J.H., Collinge, D.B., Lyngkjaer, M.F.: Transcriptional regulation by an NAC (NAM-ATAF1,2-CUC2) transcription factor attenuates ABA signalling for efficient basal defence towards *Blumeria graminis* f. sp. *hordei* in *Arabidopsis*. - Plant J. **56**: 867-880, 2008.

Joshi, T., Fitzpatrick, M.R., Chen, S.Y., Liu, Y., Zhang, H.X., Endacott, R.Z., Gaudiello, E.C., Stacey, G., Nguyen, H.T., Xu, D.: Soybean knowledge base (SoyKB): a web resource for integration of soybean translational genomics and molecular breeding. - Nucl. Acids Res. **42**: D1245-D1252, 2014.

Kou, X.H., Watkins, C.B., Gan, S.S.: *Arabidopsis* AtNAP regulates fruit senescence. - J. exp. Bot. **63**: 6139-6147, 2012.

Lam, H.M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F.L., Li, M.W., He, W., Qin, N., Wang, B., Li, J., Jian, M., Wang, J., Shao, G., Sun, S.S., Zhang, G.: Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. - Natur. Genet. **42**: 1053-1059, 2010.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G.: Clustal W and Clustal X version 2.0. - Bioinformatics **23**: 2947-2948, 2007.

Lavin, M., Herendeen, P.S., Wojciechowski, M.F.: Evolutionary rates analysis of *Leguminosae* implicates a rapid diversification of lineages during the tertiary. - Syst. Biol. **54**: 575-594, 2005.

Le, DT, Nishiyama, R., Watanabe, Y., Mochida, K., Yamaguchi-Shinozaki, K., Shinozaki, K., Tran, L.S.P.: Genome-wide survey and expression analysis of the plant-specific NAC transcription factor family in soybean during development and dehydration stress. - DNA Res. **18**: 263-276, 2011.

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., De Peer, Y.V., Rouzé, P., Rombauts, S.: PlantCARE: a database of plant *cis*-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. - Nucl. Acids Res. **30**: 325-327, 2002.

Li, X.L., Yang, X., Hu, Y.X., Yu, X.D., Li, Q.L.: A novel NAC transcription factor from *Suaeda liaotungensis* K. enhanced transgenic *Arabidopsis* drought, salt, and cold stress tolerance. - Plant Cell Rep. **33**: 767-778, 2014.

Li, Y.H., Zhao, S.C., Ma, J.X., Li, D., Yan, L., Li, J., Qi, X.T., Guo, X.S., Zhang, L., He, W.M., Chang, R.Z., Liang, Q.S., Guo, Y., Ye, C., Wang, X.B., Tao, Y., Guan, R.X., Wang, J.Y., Liu, Y.L., Jin, L.G., Zhang, X.Q., Liu, Z.X., Zhang, L.J., Chen, J., Wang, K.J., Nielsen, R., Li, R.Q., Chen, P.Y., Li, W.B., Reif, J.C., Purugganan, M., Wang, J., Zhang, M.C., Wang, J., Qiu, L.J.: Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. - BMC Genomics **14**: 579, 2013.

Librado, P., Rozas, J.: DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. - Bioinformatics **25**: 1451-1452, 2009.

Liu, J.F., Zhang, Y., Lei, X.Y., Zhang, Z.M.: Natural selection of protein structural and functional properties: a single nucleotide polymorphism perspective. - Genome Biol. **4**: R69, 2008.

Liu, T.K., Song, X.M., Duan, W.K., Huang, Z.N., Liu, G.F., Li, Y., Hou, X.L.: Genome-wide analysis and expression patterns of NAC transcription factor family under different developmental stages and abiotic stresses in chinese cabbage. - Plant mol. Biol. Rep. **32**: 1041-1056, 2014.

Livak, K.J., Schmittgen, T.D.: Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. - Methods **25**: 402-408. 2001.

Lopez, S., Stuhl, L., Fichelson, S., Dubart-Kupperschmitt, A., St Arnaud R., Galindo J.R., Murati A., Berda N., Dubreuil P., Gomez S.: NACA is a positive regulator of human erythroid-cell differentiation. - J. cell. Sci. **118**: 1595-1605, 2005.

Lynch, M., Conery, J.S.: The evolutionary fate and consequences of duplicate genes. - Science **290**: 1151-1155, 2000.

Mitsuda, N., Ohme-Takagi, M.: NAC transcription factors NST1 and NST3 regulate pod shattering in a partially redundant manner by promoting secondary wall formation after the establishment of tissue identity. - Plant J. **56**: 768-778, 2008.

Movahedi, A., Zhang, J.X., Gao, P.H., Yang, Y., Wang, L.K., Yin, T.M., Kadkhodaei, S., Ebrahimi, M., Qiang, Z.G.: Expression of the chickpea *CarNAC3* gene enhances salinity and drought tolerance in transgenic poplars. - Plant Cell Tissue Organ Cult. **120**: 141-154, 2015.

Nuruzzaman, M., Manimekalai, R., Sharoni, A.M., Satoh, K., Kondoh, H., Ooka, H., Kikuchi, S. Genome-wide analysis of NAC transcription factor family in rice. - Gene **465**: 30-44, 2010.

Olsen, A.N., Ernst, H.A., Leggio, L.L., Skriver, K.: NAC transcription factors: structurally distinct, functionally diverse. - Trends Plant Sci. **10**: 79-87, 2005.

Ooka, H., Satoh, K., Doi, K., Nagata, T., Otomo, Y., Murakami, K., Matsubara, K., Osato, N., Kawai, J., Carninci, P., Hayashizaki, Y., Suzuki, K., Kojima, K., Takahara, Y., Yamamoto, K., Kikuchi, S.: Comprehensive analysis of NAC family genes in *Oryza sativa* and *Arabidopsis thaliana*. - DNA Res. **10**: 239-247, 2003.

Pinheiro, G.L., Marques, C.S., Costa, M.D.B.L., Reis, P.A.B., Alves, M.S., Carvalho, C.M., Fietto, L.G., Fontes, E.P.B.: Complete inventory of soybean NAC transcription factors: sequence conservation and expression analysis uncover their distinct roles in stress response. - Gene **444**: 10-23, 2009.

Puranik, S., Sahu, P.P., Mandal, S.N., Venkata Suresh, B.**,** Parida, S.K., Prasad, M.: Comprehensive genome-wide survey, genomic constitution and expression profiling of the NAC transcription factor family in foxtail millet (*Setaria italica* L.). - PloS ONE **8**: e64594, 2013.

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., Lopez, R.: InterProScan: protein domains identifier. - Nucl. Acids Res **33**: W116-W120, 2005.

Sakuraba, Y., Park, S. Y., Paek, N. C.: The divergent roles of STAYGREEN (SGR) homologs in chlorophyll degradation. - Mol. Cells **38**: 390-395, 2015.

Salamov, A.A., Solovyev V.V.: *Ab initio* gene finding in *Drosophila* genomic DNA. - Genome Res. **10**: 516-522, 2000.

Salvagiotti, F., Cassman, K.G., Specht, J.E., Walters, D.T., Weiss, A., Doberman, A.: Nitrogen uptake, fixation, and response to fertilizer N in soybeans: a review. - Field Crops Res. **108**: 1-13, 2008.

Satheesh, V., Jagannadham, P.T., Chidambaranathan, P., Jain, P.K., Srinivasan, R.: NAC transcription factor genes: genome-wide identification, phylogenetic, motif and *cis*-regulatory element analysis in pigeonpea (*Cajanus cajan* (L.) Millsp.). - Mol. Biol. Rep. **41**: 7763-7773, 2014.

Schlueter, J.A., Lin, J.Y., Schlueter, S.D., Vasylenko-Sanders, I.F., Deshpande, S., Yi, J., O'Bleness, M., Roe, B.A., Nelson, R.T., Scheffler, B.E., Jackson, S.A., Shoemaker, R.C.: Gene duplication and paleopolyploidy in soybean and the implications for whole genome sequencing. - BMC Genomics **8**: 330, 2007.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., Xu, D., Hellsten, U., May, G.D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M.K., Sandhu, D., Valliyodan, B., Lindquist, E., Peto, M., Grant, D., Shu, S., Goodstein, D., Barry, K., Futrell-Griggs, M., Abernathy, B., Du, J., Tian, Z., Zhu, L., Gill, N., Joshi, T., Libault, M., Sethuraman, A., Zhang, X.C., Shinozaki, K., Nguyen, H.T., Wing, R.A., Cregan, P., Specht, J., Grimwood, J., Rokhsar, D., Stacey, G., Shoemaker, R.C., Jackson, S.A.: Genome sequence of the palaeopolyploid soybean. - Nature **463**: 178-183, 2010.

Shang, H., Li, W., Zou, C., Yuan, Y.: Analyses of the NAC transcription factor gene family in *Gossypium raimondii* Ulbr.: chromosomal location, structure, phylogeny, and expression patterns. - J. integr. Plant Biol. **55**: 663-676, 2013.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S.: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. - Mol. Biol. Evol. **28**: 2731-2739, 2011.

Taoka, K., Yanagimoto, Y., Daimon, Y., Hibara, K., Aida, M., Tasaka, M.: The NAC domain mediates functional specificity of Cup-Shaped Cotyledon proteins. - Plant J. **40**: 462-473, 2004.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G.: The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. - Nucl. Acids Res. **25**: 4876-4882, 1997.

Tian, Z.X., Wang, X.B., Lee, R., Li, Y.H., Specht, J.E., Nelson, R.L., McClean, P.E., Qiu, L.J., Ma, J.X.: Artificial selection for determinate growth habit in soybean. - Proc. nat. Acad. Sci. USA **107**: 8563-8568, 2010.

Tran, L.S., Nakashima, K., Sakuma, Y., Simpson, S.D., Fujita, Y., Maruyama K., Yamaguchi-Shinozaki K.: Isolation and functional analysis of *Arabidopsis* stress inducible NAC transcription factors that bind to a drought responsive cis-element in the early responsive to dehydration stress 1 promoter. - Plant Cell **16**: 2481-2498, 2004.

Tucker, M.L., Whitelaw, C.A., Lyssenko, N.N., Nath, P.: Functional analysis of regulatory elements in the gene promoter for an abscission-specific cellulase from bean and isolation, expression, and binding affinity of three TGA-type basic leucine zipper transcription factors. - Plant Physiol. **130**: 1487-1496, 2002.

Udvardi, M.K., Kakar, K., Wandrey, M., Montanari, O., Murray, J., Andriankaja, A., Zhang, J.Y., Benedito, V., Hofer, J.M., Chueng, F., Town, C.D.: Legume transcription factors: global regulators of plant development and response to the environment. - Plant Physiol. **144**: 538-549, 2007.

Wang, N., Zheng, Y., Xin, H., Fang, L., Li, S.: Comprehensive analysis of NAC domain transcription factor gene family in *Vitis vinifera*. - Plant Cell. Rep. **32**: 61-75, 2013.

Wang, X.B., Zhang, H.W., Sun, G.L., Jin, Y., Qiu, L.J.: Identification of active VQ motif-containing genes and the expression patterns under low nitrogen treatment in soybean. - Gene **543**: 237-243, 2014a.

Wang, X.B., Zhang, H.W., Gao, Y.L., Sun, G.L., Zhang, W.M., Qiu, L.J.: A comprehensive analysis of the *Cupin* gene family in soybean (*Glycine max*). - PloS ONE **9**: e110092, 2014b.

Wang, F., Chen, H.W., Li, Q.T., Wei, W., Li, W., Zhang, W.K., Ma, B., Bi, Y.D., Lai, Y.C., Liu, X.L., Man, W.Q., Zhang, J.S., Chen, S.Y.: *GmWRKY27* interacts with *GmMYB174* to reduce expression of *GmNAC29* for stress tolerance in soybean plants. - Plant J. **83**: 224-236, 2015.

Wilkins, O., Nahal, H., Foong, J., Provart, N.J., Campbell, M.M.: Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors. - Plant Physiol. **149**: 981-993, 2009.

Wu, Z., Xu, X., Xiong, W., Wu, P., Chen, Y., Li, M., Wu, G., Jiang, H.: Genome-wide analysis of the *NAC* gene family in physic nut (*Jatropha curcas* L.). - PloS ONE **10**: e0131890, 2015.

Xie, Q., Frugis, G., Colgan, D., Chua, N.H.: *Arabidopsis* NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. - Genes Dev. **14**: 3024-3036, 2000.

Xu, G.X., Guo, C.C., Shan, H.Y., Kong, H.Z.: Divergence of duplicate genes in exon-intron structure. - Proc. nat. Acad. Sci. USA **109**: 1187-1192, 2012.

Zhou, Z.K., Jiang, Y., Wang, Z., Gou, Z.H., Lyu, J., Li, W.Y., Yu, Y.J., Shu, L.P., Zhao, Y.J., Ma, Y.M., Fang, C., Shen, Y.T., Liu, T.F., Li, C.C., Li, Q., Wu, M., Wang, M., Wu, Y.S., Dong, Y., Wan, W.T., Wang, X., Ding, Z.L., Gao, Y.D., Xiang, H., Zhu, B.G., Lee, S.H., Wang, W., Tian, Z.X.: Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. - Nat. Biotechnol. **33**: 408-414, 2015.