Genome-wide characterization of the ankyrin repeats gene family under salt stresses in soybean

Dayong Zhang^{a†}, Qun Wan^{a*}, Xiaolan He^a, Lihua Ning^a, Yihong Huang^a, Zhaolong Xu^a, Jia Liu^a, Hongbo Shao^{a,b†}

^aProvincial Key Laboratory of Agrobiolology, Institute of Biotechnology, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China.

^bKey Laboratory of Coastal Biology & Bioresources Utilization, Yantai Institute of Coastal Zone Research (YIC), Chinese Academy of Sciences (CAS), Yantai 264003, China

*These authors contributed equally to this work.

† To whom correspondence should be addressed. E-mail:shaohongbuchu@126.com
cotton.z@126.com

Graphical abstract

Abstract:

Ankyrin repeats (ANK) gene family are common in diverse organisms and play important roles in cell growth, development and response to environmental stresses. Recently, genome-wide identification and evolutionary analyses of the ANK gene family have been carried out in *Arabidopsis*, rice and maize. However, little is known about the ANK genes in the whole soybean genome. In this study, we described the identification and structural characterization of 162 ANK genes in soybean (GmANK). Then, comprehensive bioinformatics analyses of GmANK genes family were performed including gene locus, phylogenetic, domain composition analysis, chromosomal localization and expression profiling. Domain composition analyses showed that GmANK proteins formed eleven subfamilies in soybean. In silico expression analysis of these GmANK genes demonstrated that GmANK genes show a diverse/various expression pattern, suggesting that functional diversification of GmANK genes family. Based on digital gene expression profile (DGEP) data between cultivated soybean and wild type under salt treatment, some GmANKs related to salt/drought response were investigated. Moreover, the expression pattern and subcellular localization of GmANK6 were performed. The results will provide important clues to explore ANK genes expression and function in future studies in soybean.

Key words: GmANKs, bioinformatics analyses, GmANK39-like; salt stresses; expression, subcellular localization

1. Introduction

Soil salinization, which affects an estimated 1 to 3 million hectares in the enlarged EU and more than 3.5 million hectares in China, is one of the serious problems restricting agricultural development in the world (<http://www.fao.org/soils-portal/soil-management/management-of-some-problemsoils/salt-affected-soils/more-information-on-salt-affected-soils/en/>). Most crop plants are sensitive to excess salt concentration in soil (Xiong et al., 2002), it is of great importance to insight into the mechanisms underlying plant response to salinity stress, especially for plant-salt soil

interaction. Cultivating the novel varieties with high tolerance to salt is one of the most effective methods, therefore, mining the key genes controlling crop tolerance to salt becomes more and more important for modern agriculture, especially via high throughput technologies.

Ankyrin-repeat domains contain a 33-amino acid motif that is widely present in prokaryotes, eukaryotes and some viruses, and they play an important role in mediating protein–protein interactions (Sedgwick et al., 1999). In plants, proteins containing ankyrin repeats are involved in diverse cellular functions, including cell cycle regulation, cytoskeleton interactions, signal transduction, ion transport (Sedgwick et al., 1999) and organ development, including leaf morphogenesis (Ha et al., 2004) and chloroplast biogenesis (Bae et al., 2008). Additionally, they play important roles in plant responses to biotic and abiotic stresses (Yan et al., 2002; Lu et al., 2005; Seong et al., 2007). In recent years, genome-wide analysis of ankyrin-repeat proteins in maize (Jiang et al., 2013) has been reported. Sakamoto et al (2008) identified *itn1*, a novel gene with an ankyrin-repeat motif in *Arabidopsis thaliana* that affects the ABA-mediated production of reactive oxygen species and is involved in salt-stress tolerance, suggesting that *itn1* is involved in the ABA-dependent salinity stress pathway. Ito et al (1999) characterized a protein containing ankyrin repeats named *Ankhzn*, which localized to the endosomal membrane and was involved in vesicle or protein transport, and it bound to a zinc finger motif. In general, ankyrin-repeat domains mediate protein–protein interactions and act as molecular scaffolds for protein recognition (Sedgwick et al., 1999).

Soybean (*Glycine max* .L) is one of the important crops providing plant protein and oil. Here, we identified 162 GmANK genes in soybean by database searches, which genes were classified by protein domains. We analyzed the gene locus and Chromosome(Chr) localization the protein length, number of ANK domain, molecular weight (MW) ,isoelectric points(pI) ,expression patterns and phylogenetic relationship of the ANK genes in soybean. We also surveyed the expression patterns of *GmANK39-like(GmANK6)* gene in the developmental stages of soybean and its responses to abiotic stress and subcellular localization.

2.Materials and methods

2.1.Obtaining of protein sequences containing ANK domain in soybean genome

To identify soybean proteins containing Ankyrins domain at the genome level, all protein-coding genes were downloaded from the latest version of the soybean genome (Wm82.a2.v1) in Phytozome v.10.0. Hidden Markov models (HMMs) of the 'typical' Ankyrin clan (PF00023) was to search for putative PKs using HMMER v.3.0 with an E-value cut-off of <1.0. 275 sequences containing the Ankyrin domain were identified. Only the longest variant of each gene was retained and all other redundant sequences were deleted. Domain composition was searched by SMART and PFAM database, Molecular Weight and isoelectric points was calculated with DNASTar software.

To further reveal the conserved motif of the GmANKs protein, the motif investigation software Multiple EM for Motif Elicitation (MEME) (Bailey et al., 2006) (version 4.0; <http://meme.sdsc.edu/meme4/cgi-bin/meme.cgi>) was carried out. The parameters of this analysis were as follows: optimum motif width ranged from 6 to 50; the maximum number of motifs for identification was 20, 3 motifs should find and all other parameters were defaulted.

2.1.In silico subcellular predicted localization and expression analyses for GmANKs

The protein subcellular localization was predicted using the online tool WoLF PSORT available at <http://wolfpsort.org/>, and the expression data for GmANKs in different tissues was downloaded from the <http://soybase.org>, analyzed in silico and displayed using Java-Treeview software.

2.3.Plant material

The *Glycine max* var. Willimas 82 variety was used to grow seedlings and extract total RNA for GmANK-39 gene cloning, tissue expression and induced expression analysis experiments.

2.4, Gene cloning and sequence analysis

The *GmANK39-like* gene primers (forward primer: ATGCTTGCCATCGGGAAGGCCACC, reverse

primer:TCAATGAAAGAGCAAAAAACATA) were designed according to the full-length coding sequence and used to clone the genes from soybean root tissue cDNA using RT-PCR (Reverse transcriptase-polymerase chain reaction).

2.5,Phylogenetic analyses

The Neighbour Joining (NJ) tree of GmANKs from soybean and other plants was performed using MEGA5.0 software using neighbour-joining (NJ) and maximum-likelihood (ML) methods (Tamura et al., 2011).

2.6.qPCR analysis

The soybean root, stem, leaf at young seedling stage, and flower, pod were harvested and frozen in liquid nitrogen for RNA extraction. The soybean roots were collected from plants treated with PEG6000 for 0 h, 2 h, 4 h and 8 h or with 200mM NaCl for 0h, 1h, 2h, 4h, 8h, 12h, 24h, 48h and 72h. Total RNA was extracted from soybean samples according to the TRIZOL Kit (Invitrogen) manual. Single-stranded cDNA was synthesized using 2 µg total RNA and Oligod(T)18 primer with the Takara RT-PCR system in a total volume of 25 µl according to the manufacturer's protocol (TaKaRa Bio Inc.) and used for qRT-PR analysis. qPCR analysis was performed as described previously (Zhang et al., 2013). The gene primers used were forward primer:ATGCTTGCCATCGGGAAGGCCACC,reverse

primer:TCAATGAAAGAGCAAAAAACATA, the internal control gene was *GmActin2* forward primer:CGGTGGTTCTATCTTGGCATC,reverse primer:GTCTTTCGCTTCAATAACCCTA. The assays were repeated three times.

2.7.Subcellular localization

PCR-generated *Hind III*-*BamHI* fragment containing the open reading frame of *GmANK39-like* without stop codon (Forward primer: cgcGTTCGACATGCTTGCCATCGGG,reverse primer:cgcCCATGGATGAAAGAGCAAAAAACAT) was subcloned in-frame upstream of the GFP gene in the pJIT166GFP plasmid. The construct was validated by sequencing.

Arabidopsis leaf protoplasts were isolated according to Yoo et al (2007). The resulting fusion construct (p35S::GmANK39-like-GFP) or empty control vector (p35S::GFP)

were introduced into *Arabidopsis* protoplasts by the PEG4000-mediated method (Abel and Theologis, 1994). After 24 h at room temperature, GFP signal was detected by confocal fluorescence microscopy (Zeiss, LSM510 Meta, Carl Zeiss AG).

2.8. Statistical analysis

Data were analyzed using an ANOVA test in Microsoft Excel. Significant differences among means were determined by * at $P < 0.05$ and ** at $P < 0.01$.

3. Results

3.1. Genome-wide identification of proteins containing ANK domain in soybean genome

The HMM profile (PF00023) was used to search for conserved ANK domain in soybean genome data, totally, 162 putative ANK-containing proteins were identified and validated by on-line SMART software and pfam and BLAST2GO analysis. The gene locus and Chromosome (Chr) localization were provided and gene names (GmANK1-162) were designated according to the order of Chr1-20, indicating that the GmANKs distributed in each Chr, and the Chr8 (14) contained the most, while the Chr20 included 5 GmANKs. Furthermore, the peptide length, number of ANK domain, molecular weight (MW) and isoelectric points (pI) were deduced from the protein sequences respectively. In details, the largest gene (*Glyma11G157400*, *GmANK90*) encoded 1,638 amino acids with the MW of 179.85, while the shortest one (*Glyma02G033700*, *GmANK6*) contained only 143 with a MW of 15.86. The number of ANK domain ranged from 1-11 and the average is 4.5. The protein subcellular localization of GmANKs mainly distributed nucl(nucleus, 35.19%), cyto(cytoplasm, 22.22%), plas(plasma membrane, 22.22%) and chlo(chloroplast, 12.35%), the remaining localized other organelles, such as mitochondrion, ER(endoplasmic reticulum) (Table 1)

Based on the results of SMART analysis for every GmANKs, we found that ANK domain mainly combined with the BAR(Endocytosis and intracellular transport), PH(Pleckstrin homology domain), ArfGap(Putative GTP-ase activating proteins for the small GTPase), BTB(Broad-Complex, Tramtrack and Bric a

brac),ZnFC3H1(Zinc fingerC3H1),TM(Transmembrane region),cNMP(Cyclic nucleotide-monophosphate binding domain),TPR(Tetratricopeptide repeats),RING(Ring finger),CG1(CalModulin -binding Transcription Activator),IQ(Calmodulin-binding motif),AAA(ATPases associated with a variety of cellular activities) and CHROMO(CHRromatin Organization Modifier) domains, indicating the functional diversity of ANK(Fig.1).

To reveal the sequence conservation of soybean ANK repeats motifs, we performed MEME analysis of GmANK1-80 to generate sequence logos , which showed that there were several highly conserved amino acid residues within the ANK repeats motifs, including the T at position 3, the A at position 8, the L at position 20, LH at position 39-40 and G at position 46 respectively(Fig.2). Each logo of the ANK motifs was similar to the Pfam sequence (PF00023: DGFTPLHLAALRGNLEVVKLLLSQGADLNAQDD)

3.2.In silico expression patterns analysis of *GmANKs* genes in soybean different tissues

To insight into the expression patterns of *GmANKs* genes in soybean different tissues, the expression data in different tissues was downloaded from soybase (<http://soybase.org>) and analyzed in silico and displayed using Java-Treeview software. The heatmap showed that some *GmANKs* had no expression in any tissues, such as *GmANK18,27,45,46,61,107,118,150,151*, some *GmANKs* were organ-specific, including *GmANK9,17* in nodule, *GmANK19,152* in flower and *GmANK48,49,70,102,142* in root respectively. Some *GmANKs* were constitutively expressed, including *GmANK25,90,100,108*. Seven *GmANKs*(32,44,79,93,124,161,162) had no expression data in soybean database(Fig.3).

3.3.Phylogenetic analysis of *GmANKs*

To gain the phylogenetic relationship among the *GmANKs* proteins and acquire the evolutionary history of this gene family, a phylogenetic tree was constructed except for the *GmANK2,44,55,79* with the low similarities, 158 *GmANKs* were clustered using MEGA5.0 software. The results indicated that *GmANK* proteins were divided into eight distinct major groups(A-H), and some major groups could be divided into

different subgroups. In general, most of the closely related members in the phylogenetic tree had the same or very similar domain composition, such as Zinc finger CCCH domain-containing protein (Group G) and Ankyrin-kinase family (Group H).(Fig.4)

3.4.The cloning and expression patterns of *GmANK39-like*(*GmANK6*)

GmANK6 was a shortest one in soybean ANK family and was isolated by screening a salt-treated yeast two hybrid library in our lab previously.BLASTx search showed that *GmANK6* had a higher identities with *GmANK-39-Like*, and had a 75% ,68% and 47%identities with *Cajanus cajan* (CcAnk2 accession no:KYP71517),*Cicer arietinum* (CaAnk50,accession no:XP_004496053)and *Prunus mume* (ankyrin-3-like,accession no:XP_008234958) respectively. SMART on-line software analysis indicated that the *GmANK-39-Like* protein sequence only contains two Ankyrin repeats domains from amino acids 47-76 and 80-108. However, the others above mentioned possessed five Ankyrin repeats domains and had a relatively conserved sequence characterization at Ankyrin repeats domain sites for these proteins(Supplemental Fig. 1).

Phylogenetic analysis was used to find out the relationship among the *GmANK-39-Like* and other ANKs proteins from *Arabidopsis thaliana* (AtAnk2), *Cicer arietinum* (CaAnk50),*Cajanus cajan* (CcAnk2),*Gossypium raimondii* (GrAnk50),*Medicago truncatula*(MtAnk), *Ricinus communis* (RcAnk1) , *Theobroma cacao* (TcAnk) ,*Vigna radiata*(VrAnk1) and *Vitis vinifera* (VvAnk65).*GmANK-39-Like* was found to cluster together with other ANKs and had a higher homology with VrAnk1and CcAnk2, however, had the lower homology with AtAnk2, indicating that *GmANK-39-Like* really originates from the legume(Supplemental Fig.2).

To examine the tissue expression pattern of *GmANK-39-Like*,the root,stem,leaf,flower and pod were investigated. The result showed that *GmANK-39-Like* mainly expressed the root,leaf and pod, weak in stem and flower(Fig.5 a), which was consistent with the data of in silico expression.To determine the expression pattern of *GmANK-39-Like* under abiotic stresses, the soybean seedlings of three-leaf stage were treated with 20%PEG and 200mM NaCl solutions for different time points, then QRT-PCR was carried out using treated roots tissue as template. The results showed that the expression

of *GmANK-39-Like* was increased at 2h, then significantly declined at 4h and went back to previous expression level at 8h(Fig.5 b). However, the expression under salinity stress was decreased all the time points(Fig.5c).

3.5.Subcellular localization of GmANK39-like(GmANK6)

To examine the functional localization of GmANK-39-Like, the PEG-mediated protoplast transformation method was performed. Enhanced green fluorescent protein(EGFP) in-frame fused C terminally to GmANK-39-Like and only EGFP (empty vector) were transiently expressed in *Arabidopsis* mesophyll protoplast driven by cauliflower mosaic virus 35S promoter respectively. The results of subcellular localization demonstrated that GFP was detected all the whole cells for the 35S::GFP, but mainly distributed cytoplasm for 35S::GmANK-39-Like:GFP(Fig.6), indicating GmANK-39-Like is not a membrane protein and functions at the cytoplasm .

4.Discussion

Soil salinity becomes more and more one of the major obstacles in the world. In China, more than 600 hectares soil field were increased every year in JiangSu province because of sea movement, and in Australia, large amounts of salt accumulated from cyclical deposition via rainfall, weathering of saline material and connate salts laid down as marine sediments in earlier geological time (Spies and Woodgate.,2004). The activities of man, including agronomy and hydrology, can also cause soil salinization via mobilisation of salts(Huang et al.,2016). Plant directly touched with the saline soil, therefore, understanding the principles of plant-soil interaction and revealing the mechanisms of plant response to salt stress will provide the important theoretical basis for modern salt-soil agriculture.

The ankyrin repeats are about 33-amino-acid in length repeats found in numerous proteins, and serve as domains for protein–protein interactions (Michaely and Bennett., 1992) .The ankyrin repeat was first identified in the yeast cell cycle regulator Swi6/Cdc10 and the Drosophila signaling protein Notch (Breedon and Nasmyth 1987).The number of ankyrin repeats varies for different proteins. Analysis of the SMART and PFAM databases show that the number of repeats per protein ranges from

1 to 33. To date, the protein with the most ankyrin repeats is the ORF EAA39756 from *Giardia lamblia* containing 34 repeats when PFAM and SMART predictions are combined (Mosavi et al., 2004). In the present study, we found the most ankyrin repeats in soybean is Glyma16G060400, Glyma18G008500 and Glyma19G086700, which contain 11 repeats. However, the least ones (Glyma01G052600, Glyma05G117000, Glyma08G008000, Glyma09G038300, Glyma15G143400) only include one repeat. The average of the ankyrin repeats is 4.5, which is identical with the rice ANK proteins (Becerra et al., 2004). These may indicate that the structure of ankyrin repeats is of importance for ANK proteins, especially for protein-protein interactions. In recent years, the number of ANK proteins has been found to be 175 in rice, 105 in the *Arabidopsis thaliana*, 130 in tomato and 71 in maize (Becerra et al., 2004; Huang et al., 2009; Yuan et al., 2013; Jiang et al., 2013). In this study, we identified 162 ANK proteins in soybean, which can be classified into 11 subfamilies according to domain composition and they are distributed on all 20 chromosomes. Moreover, some ANK proteins showed gene duplications on some chromosomes, such as Glyma06G294100, Glyma06G294200 and Glyma06G294400. This is the first report about the ANK family in soybean.

In addition, ANK domain mainly combined with the other domains (Fig. 1). Huang et al. (2009) found that rice ANK proteins contained Motile_Sperm MSP (major sperm protein) domain, ACBP (acyl CoA binding Protein), PPR (pentatricopeptide repeats); Ion_trans (ion transport Protein), RCC1 (regulator of chromosome condensation); rve (integrase core domain); RVT (reverse transcriptase RNA-dependent DNA polymerase); RRM (RNA recognition motif); ubiquitin (ubiquitin family) instead of domains mentioned in this study. This inferred that rice ANKs possessed more diverse functions than maize and soybean, which was consistent with the statement that ANK repeats mediate protein-protein interactions also in plants (Becerra et al., 2004).

In plants, one of the first characterized ANK proteins was AKR (*Arabidopsis* Ankyrin Repeat). The transcripts levels of AKR were found to be light dependent and was associated with the regulation of chloroplast differentiation (Zhang et al., 1992). To insight into the expression profile of *GmANK* genes which will provide clues to the

function of each gene.,the in silico expression patterns for each gene were performed. The data showed that the diverse or different expression types, including organ-specific, constitutive and developmental stage-specific, indicating that they play important roles in plant development and response to stresses. In this study, the GmANK6 showed an increase expression after drought treatment at early stage, however, a decreasing after salt treatment ,indicating ANK proteins have the response to abiotic stresses. The cytoplasm localization of GmANK6 complied with the putative result by WoLF PSORT tool. The results will provide important clues to explore ANK genes expression and function in future studies in soybean

5.Conclusions

To mining the key salt tolerance related genes in soybean for planting in salt soil, 162 *GmANK* genes were identified by database searches and classified by protein domains, the gene locus and Chromosome(Chr) localization the protein length, number of ANK domain, molecular weight,isoelectric points,expression patterns and phylogenetic relationship of the ANK genes in soybean were analyzed ,and the expression patterns of *GmANK39-like(GmANK6)* gene in the developmental stages of soybean and its responses to abiotic stress and subcellular localization were also surveyed.

Acknowledgments

This study was sponsored by the National Science Foundation of China (31101166); the Jiangsu Natural Science Foundation, China (BK20151364) ,Shuangchuang Talent Plan of Jiangsu Province, China and Jiangsu Agriculture Science and Technology Innovation Fund (JASTIF, CX(15)1005).

Author contributions

DY. Zhang, and Q. Wan conceived, designed and conducted the experiments. ZL. Xu, XL. He, and HY. Huang analyzed the data and results. DY. Zhang, wrote the manuscript. HB. Shao and DY. Zhang monitored the experiments and critically commented on the manuscript. All authors read and approved the final manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

References

- Abel, S., and Theologis, A.,1994. Transient transformation of Arabidopsis leaf protoplasts: a versatile experimental system to study gene expression. *Plant. J.* 5, 421-427.<http://dx.doi:10.1111/j.1365-313X.1994.00421.x>.
- Bae, W., Lee, Y. J., Kim, D. H., Lee, J., Kim, S., Sohn, E. J.,Hwang, I., 2008.AKR2A-mediated import of chloroplast outer membrane proteins is essential for chloroplast biogenesis. *Nat .Cell. Biol.* 10, 220-227.<http://dx.doi:10.1038/ncb1683>.
- Bailey, T.L., Williams, N., Misleh, C., Li, W.W., 2006. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic. acids. Res.* 34, 369-373.<http://dx.doi:10.1093/nar/gkl198>.
- Becerra, C., Jahrmann, T., Puigdomenech, P., Vicient, C. M. ,2004. Ankyrin repeat-containing proteins in Arabidopsis: characterization of a novel and abundant group of genes coding ankyrin-transmembrane proteins. *Gene.* 340, 111–121.[http://dx. doi:10.1016/j.gene.2004.06.006](http://dx.doi:10.1016/j.gene.2004.06.006).
- Breeden, L., Nasmyth, K., 1987. Similarity between cell-cycle genes of budding yeast and fission yeast and the Notch gene of Drosophila. *Nature* .329, 651-654. [http://dx. doi:10.1038/329651a0](http://dx.doi:10.1038/329651a0).
- Ha, C.M., Jun, J.H., Nam, H.G., Fletcher, J.C., 2004. BLADE-ON-PETIOLE1 encodes a BTB/POZ domain protein required for leaf morphogenesis in Arabidopsis thaliana. *Plant .Cell .Physiol*, 45, 1361-1370.[http://dx. doi:10.1093/pcp/pch201](http://dx.doi:10.1093/pcp/pch201).
- Huang, J., Prochazka, M.J., Triantafilis, J.,2016.Irrigation salinity hazard assessment and risk mapping in the lower Macintyre Valley,Australia.*Sci .Total .Environ.*551-552,460-73. [http://dx. doi:10.1016/j.scitotenv.2016.01.200](http://dx.doi:10.1016/j.scitotenv.2016.01.200).
- Huang, J., Zhao, X., Yu, H., Ouyang, Y., Wang, L., Zhang, Q. ,2009. The ankyrin repeat gene family in rice: genome-wide identification, classification and expression profiling. *Plant. Mol. Biol.* 71, 207–226.[http://dx. doi:10.1016/j.pmb.2009.05.006](http://dx.doi:10.1016/j.pmb.2009.05.006).

10.1007/s11103-009-9518-6.

- Ito, K., Ishii, N., Miyashita, A., Tominaga, K., Kuriyama, H., Maruyama, H., Shirai, M., Naito, M., Arakawa, M., Kuwano, R., 1999. Molecular cloning of a novel 130-kDa cytoplasmic protein, Ankhzn, containing Ankyrin repeats hooked to a zinc finger motif. *Biochem. Biophys. Res. Commun.* 257, 206-213.<http://dx.doi:10.1006/bbrc.1999.0430>.
- Jiang, H., Wu, Q., Jin, J., Sheng, L., Yan, H., Cheng, B., Zhu, S., 2013. Genome-wide identification and expression profiling of ankyrin-repeat gene family in maize. *Dev. Genes. Evol.* 223, 303-318. <http://dx.doi:10.1007/s00427-013-0447-7>.
- Lu, H., Liu, Y., Greenberg, J. T., 2005. Structure-function analysis of the plasma membrane- localized Arabidopsis defense component ACD6. *Plant. J.* 44, 798-809.<http://dx.doi:10.1111/j.1365-313X.2005.02567.x>.
- Michaely, P., Bennett, V., 1992. The ANK repeat: a ubiquitous motif involved in macromolecular recognition. *Trends. Cell. Biol.* 2, 127-129.[http://dx.doi:10.1016/0962-8924\(92\)90084-Z](http://dx.doi:10.1016/0962-8924(92)90084-Z).
- Mosavi, L.K., Cammett, T.J., Desrosiers, D.C., Peng, Z.Y., 2004. The ankyrin repeat as molecular architecture for protein recognition. *Protein. science.* 13, 1435-1448.<http://dx.doi:10.1110/ps.03554604>.
- Sakamoto, H., Matsuda, O., Iba, K., 2008. ITN1, a novel gene encoding an ankyrin-repeat protein that affects the ABA-mediated production of reactive oxygen species and is involved in salt-stress tolerance in *Arabidopsis thaliana*. *Plant. J.* 56, 411-422.<http://dx.doi:10.1111/j.1365-313X.2008.03614.x>.
- Sedgwick, S.G., Smerdon, S.J., 1999. The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends. Biochem. Sci.* 24, 311-316.[http://dx.doi:10.1016/S0968-0004\(99\)01426-7](http://dx.doi:10.1016/S0968-0004(99)01426-7).
- Seong, E. S., Choi, D., Cho, H. S., Lim, C. K., Cho, H. J., Wang, M. H. ,2007. Characterization of a stress-responsive ankyrin repeat-containing zinc finger protein of *Capsicum annuum* (CaKR1). *J. Biochem. Mol. Biol.* 40, 952-958.<http://dx.doi:10.5483/BMBRep.2007.40.6.952>.
- Sharma, M., Pandey, G.K., 2015. Expansion and Function of Repeat Domain Proteins

- During Stress and Development in Plants. *Front. Plant Sci.* 6, 1218.
[http://dx.doi: 10.3389/fpls.2015.01218](http://dx.doi.org/10.3389/fpls.2015.01218).
- Spies, B., Woodgate, P., 2004. Review of methods for mapping dry land salinity. *Aseg. Extended Abstracts.* 2004, 1-4.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011.
 MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731-2739. [http://dx.doi: 10.1093/molbev/msr121](http://dx.doi.org/10.1093/molbev/msr121).
- Xiong, L., Lee, H., Ishitani, M., Zhu, J. K., 2002. Regulation of osmotic stress-responsive gene expression by the LOS6/ABA1 locus in *Arabidopsis*. *J. Biol. Chem.* 277, 8588-96. [http://dx.doi: 10.1074/jbc.M109275200](http://dx.doi.org/10.1074/jbc.M109275200).
- Yan, J., Wang, J., Zhang, H., 2002. An ankyrin repeat-containing protein plays a role in both disease resistance and antioxidation metabolism. *Plant. J.* 29, 193-202. [http://dx.doi: 10.1046/j.0960-7412.2001.01205.x](http://dx.doi.org/10.1046/j.0960-7412.2001.01205.x)
- Yoo, S. D., Cho, Y. H., and Sheen, J., 2007. *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.* 2, 1565-72. [http://dx.doi: 10.1038/nprot.2007.199](http://dx.doi.org/10.1038/nprot.2007.199).
- Yuan, X., Zhang, S., Qing, X., Sun, M., Liu, S., Su, H., Shu, H., Li, X., 2013.
 Superfamily of ankyrin repeat proteins in tomato. *Gene.* 523, 126-136. [http://dx.doi: 10.1016/j.gene.2013.03.122](http://dx.doi.org/10.1016/j.gene.2013.03.122).
- Zhang, D. Y., Ali, Z., Wang, C. B., Xu, L., Yi, J. X., Xu, Z. L., Liu, X. Q., He, X. L., Huang, Y. H., Khan, I. A., Ma, H. X., 2013. Genome-wide sequence characterization and expression analysis of major intrinsic proteins in soybean (*Glycine max* L.). *PloS. one.* 8, e56312. [http://dx.doi: 10.1371/annotation/e3307d0c-bb59-4f75-89b4-8e0d5af087d5](http://dx.doi.org/10.1371/annotation/e3307d0c-bb59-4f75-89b4-8e0d5af087d5).
- Zhang, H., Scheirer, D. C., Fowle, W. H., and Goodman, H. M., 1992. Expression of antisense or sense RNA of an ankyrin repeat-containing gene blocks chloroplast differentiation in *Arabidopsis*. *Plant. Cell.* 4, 1575-1588. [http://dx.doi: http://dx.doi.org/10.1105/tpc.4.12.1575](http://dx.doi.org/10.1105/tpc.4.12.1575).

Table1

The ANK gene family in soybean

Gene locus	Chr	Gene name	No. AN K	Length (aa)	MW (KD)	PI	BLAST2GO Results	SL
Glyma01G052600	1	GmANK 1	1	428	48.02	7.16	Mitogen-activated protein kinase kinase Raf17.1	cysk
Glyma01G055500	1	GmANK 2	5	276	29.99	5.27	Ankyrin repeat domain-containing protein	cyto
Glyma01G149900	1	GmANK 3	3	252	26.87	6.18	Ankyrin repeat domain-containing protein	nucl
Glyma01G150400	1	GmANK 4	3	173	19.22	7.81	Ankyrin repeat domain-containing protein	cyto
Glyma01G161900	1	GmANK 5	5	443	50.61	7.58	Ankyrin repeat domain-containing protein	chlo
Glyma02G033700	2	GmANK 6	2	143	15.86	6.9	Ankyrin repeat domain-containing protein	cyto
Glyma02G113900	2	GmANK 7	5	244	26.03	4.96	Ankyrin repeat family protein	cyto
Glyma02G151200	2	GmANK 8	5	456	49.13	7.49	Ankyrin repeat domain-containing protein	nucl
Glyma02G188700	2	GmANK 9	2	477	51.97	6.15	Ankyrin repeat family protein	ER
Glyma02G227000	2	GmANK 10	2	696	75.83	5.9	Zinc finger CCCH domain-containing protein	nucl
Glyma02G243400	2	GmANK 11	5	850	97.4	6.28	Potassium channel SKOR	plas
Glyma02G263600	2	GmANK 12	4	352	38.4	6.66	Ankyrin repeat domain-containing protein	cyto

Glyma02G28830	2	GmANK	3	455	51.49	7.15	Protein kinase family protein	cyto
0		13						
Glyma02G29660	2	GmANK	2	658	71.84	7.19	Zinc finger CCCH domain-containing protein	nucl
0		14						
Glyma03G12860	3	GmANK	2	488	53.34	6.44	BTB/POZ ankyrin repeat protein	cyto
0		15						
Glyma03G13860	3	GmANK	3	681	74.91	5.89	Zinc finger CCCH domain-containing protein	nucl
0		16						
Glyma03G17040	3	GmANK	5	628	69.54	5.44	Ankyrin repeat family protein	plas
0		17						
Glyma03G17050	3	GmANK	4	574	63.81	6.57	Ankyrin repeat family protein	plas
0		18						
Glyma03G17440	3	GmANK	7	579	63.24	9.6	Ankyrin repeat-containing protein	plas
0		19						
Glyma03G17450	3	GmANK	8	522	56.97	8.8	Ankyrin repeat-containing protein	plas
0		20						
Glyma03G24760	3	GmANK	5	465	49.69	5.1	Ankyrin repeat family protein	chlo
0		21						
Glyma03G26450	3	GmANK	8	567	62.44	9.4	Ankyrin repeats-containing protein	plas
0		22						
Glyma04G04450	4	GmANK	3	372	41.28	5.45	Arginine N-methyltransferase	cyto
0		23						
Glyma04G06940	4	GmANK	5	879	99.05	6.43	Potassium channel	plas
0		24						
Glyma04G11650	4	GmANK	3	351	37.633	4.39	Ankyrin-repeat protein	mito
0		25						
Glyma04G13470	4	GmANK	9	963	104.15	5.86	E3 ubiquitin-protein ligase KEG	chlo
0		26						
Glyma04G20080	4	GmANK	2	786	87.9	3.69	ADP-ribosylation factor GTPase-activating protein	nucl
0		27						
Glyma04G23360	4	GmANK	3	347	37.88	4.32	ACYL-COA binding family protein	ER
0		28						
Glyma05G01060	5	GmANK	4	894	100.16	6.91	Potassium channel	plas
0		29						
Glyma05G08840	5	GmANK	9	539	58.1	8.6	NA	chlo
0		30						
Glyma05G11700	5	GmANK	1	1089	121.94	5.67	Calmodulin-binding transcription activator	nucl
0		31						
Glyma05G12470	5	GmANK	9	1032	113.99	7.45	NA	plas
0		32						

Glyma05G14520	5	GmANK	9	675	74.51	7.79	Serine/threonine protein phosphatase 6 regulatory ankyrin repeat subunit C	chlo
Glyma05G14830	5	GmANK	2	984	109.31	5.47	Calmodulin-binding family protein	nucl
Glyma05G16760	5	GmANK	2	476	53.66	9.05	Ankyrin-kinase	mito
Glyma05G17820	5	GmANK	2	1123	126.99	5.99	Ethylene-responsive calmodulin-binding family protein	nucl
Glyma05G20070	5	GmANK	2	801	90.32	7.05	ADP-ribosylation factor GTPase-activating protein	nucl
Glyma05G23860	5	GmANK	5	531	58.73	7.74	Ankyrin repeat-containing protein	chlo
Glyma05G24750	5	GmANK	6	780	89.15	6.65	Outward rectifying potassium channel	plas
Glyma06G04500	6	GmANK	3	329	36.33	4.64	Arginine N-methyltransferase	nucl
Glyma06G07110	6	GmANK	4	880	99.49	6.46	Inward rectifying shaker-like K ⁺ channel	plas
Glyma06G13120	6	GmANK	3	355	38.8	4.36	ACYL-COA binding family protein	ER
Glyma06G16470	6	GmANK	2	739	83.58	7.61	ADP-ribosylation factor GTPase-activating protein	cyto
Glyma06G19860	6	GmANK	2	156	17.43	6.32	Ankyrin repeat domain-containing protein	cyto
Glyma06G24130	6	GmANK	5	418	46.98	6.57	Ankyrin repeat protein	cyto
Glyma06G24200	6	GmANK	5	434	48.62	7.22	Ankyrin repeat protein	plas
Glyma06G29410	6	GmANK	9	599	66.44	7.46	Ankyrin repeats-containing protein	plas
Glyma06G29420	6	GmANK	9	665	74.99	8.64	Ankyrin repeats-containing protein	plas
Glyma06G29440	6	GmANK	9	668	75.05	8.95	Ankyrin repeats-containing protein	plas
Glyma06G32020	6	GmANK	3	353	37.88	4.41	Ankyrin-repeat protein	nucl
Glyma07G08580	7	GmANK	3	481	53.87	6.3	Ankyrin repeat domain-containing protein	cyto
Glyma07G16950	7	GmANK	6	526	57.73	8.41	Ankyrin repeat-containing protein	plas

Glyma07G18500 0	7	GmANK 53	6	541	59.59	7.24	S-acyltransferase	plas
Glyma07G22800 0	7	GmANK 54	2	422	47.78	7.95	Mitogen-activated protein kinase kinase kinase Raf17.1	cyto
Glyma07G23470 0	7	GmANK 55	3	232	26.08	6.34	NA	cyto
Glyma07G24200 0	7	GmANK 56	2	922	104.29	7.28	Calmodulin-binding transcription activator	nucl
Glyma07G26120 0	7	GmANK 57	2	569	64.11	9.49	Ankyrin repeat family protein	chlo
Glyma08G00800 0	8	GmANK 58	1	690	77.59	8.18	Signal recognition particle subunit SRP68	nucl
Glyma08G06430 0	8	GmANK 59	6	542	59.68	7.26	S-acyltransferase	cyto
Glyma08G07210 0	8	GmANK 60	2	1080	121.26	5.89	Calmodulin-binding transcription activator	nucl
Glyma08G07980 0	8	GmANK 61	7	618	68.39	6.65	NA	plas
Glyma08G10180 0	8	GmANK 62	9	677	75.04	8.35	Serine/threonine protein phosphatase 6 regulatory ankyrin repeat subunit C	nucl
Glyma08G10520 0	8	GmANK 63	2	966	107.77	6	Calmodulin-binding family protein	nucl
Glyma08G12600 0	8	GmANK 64	2	476	53.71	8.98	Ankyrin-kinase	mito
Glyma08G13520 0	8	GmANK 65	2	1103	124.43	5.81	Ethylene-responsive calmodulin- binding family protein	nucl
Glyma08G15020 0	8	GmANK 66	3	158	16.84	4.41	Ankyrin repeat domain-containing protein	cyto
Glyma08G17890 0	8	GmANK 67	3	1082	121.61	5.72	Calmodulin-binding transcription activator	nucl
Glyma08G18760 0	8	GmANK 68	4	835	95.46	7.2	Potassium channel	cyto
Glyma08G25760 0	8	GmANK 69	2	824	92.1	7.32	ADP-ribosylation factor GTPase- activating protein AGD3	nucl
Glyma08G31550 0	8	GmANK 70	6	347	38.05	8.24	Hop-interacting protein THI015	cyto
Glyma08G35830 0	8	GmANK 71	6	439	47.54	7.05	Hop-interacting protein THI012	cyto
Glyma09G03830 0	9	GmANK 72	1	912	102.97	6.84	Calmodulin-binding transcription activator	nucl

Glyma09G10290 0	9	GmANK 73	3	729	78.39	6.12	Zinc finger CCCH domain- containing protein	nucl
Glyma09G14360 0	9	GmANK 74	5	512	54.91	6.98	Ankyrin repeat-containing protein	nucl
Glyma09G19110 0	9	GmANK 75	3	485	54.32	7.2	Ankyrin repeat domain-containing protein	cyto
Glyma09G21400 0	9	GmANK 76	4	250	26.54	6.1	Ankyrin repeat domain-containing protein	nucl
Glyma09G21540 0	9	GmANK 77	3	173	19.32	7.99	Ankyrin repeat domain-containing protein	nucl
Glyma10G05980 0	10	GmANK 78	3	205	21.81	4.25	Ankyrin repeat family protein	cyto
Glyma10G09720 0	10	GmANK 79	2	233	26.19	6.25	NA	cyto
Glyma10G10040 0	10	GmANK 80	3	711	77.37	5.88	Zinc finger CCCH domain- containing protein	nucl
Glyma10G12450 0	10	GmANK 81	3	558	62.93	9.38	Ankyrin repeat family protein	cysk
Glyma10G18110 0	10	GmANK 82	2	457	48.88	6.68	Zinc finger family protein	cyto
Glyma10G23800 0	10	GmANK 83	5	513	54.97	6.66	Ankyrin repeat-containing protein	nucl
Glyma10G29190 0	10	GmANK 84	7	593	64.11	8.41	NA	plas
Glyma11G08180 0	11	GmANK 85	5	445	50.57	7.24	Ankyrin repeat domain-containing protein	chlo
Glyma11G08190 0	11	GmANK 86	4	422	47.83	9.84	Ankyrin repeat domain-containing protein	cyto
Glyma11G09720 0	11	GmANK 87	3	383	41.8	4.82	Chaos family protein	chlo
Glyma11G10450 0	11	GmANK 88	2	777	87.81	6.28	Arf GTPase-activating domain- containing protein	nucl
Glyma11G14000 0	11	GmANK 89	6	448	48.31	7.35	Ankyrin repeat family protein	nucl
Glyma11G15740 0	11	GmANK 90	9	1638	179.85	5.95	E3 ubiquitin-protein ligase KEG	chlo
Glyma11G17440 0	11	GmANK 91	2	606	65.79	6.44	Zinc finger CCCH domain- containing protein	nucl
Glyma11G24770 0	11	GmANK 92	8	653	71.9	7.25	Serine/threonine protein phosphatase 6 regulatory ankyrin repeat subunit C	chlo

Glyma11G24880 0	11	GmANK 93	3	359	39.27	9.03	Ankyrin repeat domain-containing protein	chlo
Glyma11G25190 0	11	GmANK 94	2	911	10.24	5.62	Calmodulin-binding transcription activator	nucl
Glyma12G02950 0	12	GmANK 95	2	777	87.63	6.28	Arf GTPase-activating domain-containing protein	nucl
Glyma12G06360 0	12	GmANK 96	6	448	48.44	7.49	Ankyrin repeat family protein	nucl
Glyma12G07490 0	12	GmANK 97	8	549	59.13	9.48	Ankyrin repeat-containing protein	plas
Glyma12G11140 0	12	GmANK 98	9	658	73.75	9.12	Ankyrin repeats-containing protein	plas
Glyma12G17050 0	12	GmANK 99	4	832	95.07	7.68	Potassium channel	cyto
Glyma12G23730 0	12	GmANK 100	3	705	75.82	6.49	Zinc finger CCCH domain-containing protein	nucl
Glyma13G09190 0	13	GmANK 101	5	509	54.97	7.28	Ankyrin repeat-rich protein	chlo
Glyma13G13030 0	13	GmANK 102	6	470	52.46	8.89	Ankyrin repeat family protein	plas
Glyma13G14650 0	13	GmANK 103	3	205	21.92	4.26	Ankyrin repeat family protein	mito
Glyma13G15240 0	13	GmANK 104	5	331	36.96	4.87	Ankyrin repeat family protein	chlo
Glyma13G16340 0	13	GmANK 105	5	633	69.06	6.71	S-acyltransferase	plas
Glyma13G19520 0	13	GmANK 106	9	1422	157.54	5.72	E3 ubiquitin-protein ligase KEG	cyto
Glyma13G20200 0	13	GmANK 107	5	332	37.23	8.42	Ankyrin repeat family protein	cyto
Glyma13G20280 0	13	GmANK 108	3	702	75.66	6.75	Zinc finger CCCH domain-containing protein	nucl
Glyma13G22340 0	13	GmANK 109	3	456	51.45	7.44	Ankyrin protein kinase	cyto
Glyma13G33150 0	13	GmANK 110	8	546	59.38	9.04	Ankyrin repeat-containing protein	plas
Glyma13G33530 0	13	GmANK 111	7	452	48.55	7.35	Ankyrin repeat family protein	nucl
Glyma14G01630 0	14	GmANK 112	2	681	74.6	6.88	Zinc finger CCCH domain-containing protein	nucl
Glyma14G02670 0	14	GmANK 113	3	454	51.34	7.49	Protein kinase family protein	mito

Glyma14G09960 0	14	GmANK 114	2	318	34.92	10.12	Calmodulin-binding transcription factor SR2L	nucl
Glyma14G12650 0	14	GmANK 115	6	876	98.94	6.61	Shaker-like potassium channel 1	plas
Glyma14G19410 0	14	GmANK 116	2	691	75.13	6.01	Zinc finger CCCH domain- containing protein	nucl
Glyma14G21250 0	14	GmANK 117	5	851	97.56	6.97	Potassium channel SKOR	plas
Glyma15G01860 0	15	GmANK 118	8	648	72.03	8.4	Ankyrin repeat-containing protein	plas
Glyma15G03910 0	15	GmANK 119	7	445	47.87	7.05	Ankyrin repeat family protein	nucl
Glyma15G04260 0	15	GmANK 120	8	546	59.22	8.85	Ankyrin repeat-containing protein	mito
Glyma15G05360 0	15	GmANK 121	3	1089	122.48	5.87	Calmodulin-binding transcription activator	nucl
Glyma15G08850 0	15	GmANK 122	3	457	51.4	7.92	Ankyrin protein kinase	pero
Glyma15G14340 0	15	GmANK 123	1	912	103.08	7.68	Calmodulin-binding transcription activator	nucl
Glyma15G23300 0	15	GmANK 124	2	234	25.96	8.87	A subunit of NADH dehydrogenase	cyto
Glyma15G27480 0	15	GmANK 125	2	723	77.74	5.95	Zinc finger CCCH domain- containing protein	nucl
Glyma16G03820 0	16	GmANK 126	7	517	56.21	7.29	NA	cyto
Glyma16G03890 0	16	GmANK 127	2	1081	117.59	9.09	Ankyrin repeat domain-containing protein	nucl
Glyma16G06040 0	16	GmANK 128	11	754	81.32	6.58	Ankyrin repeat domain-containing protein	cyto
Glyma16G06270 0	16	GmANK 129	8	669	73.63	9.27	Ankyrin-like protein	nucl
Glyma16G19640 0	16	GmANK 130	5	512	54.88	6.82	Ankyrin repeat-containing protein	nucl
Glyma17G01280 0	17	GmANK 131	2	563	63.25	9.21	Ankyrin repeat family protein	chlo
Glyma17G03190 0	17	GmANK 132	2	923	103.87	7.51	Calmodulin-binding transcription activator	nucl
Glyma17G06840 0	17	GmANK 133	5	511	55.3	7.53	Ankyrin repeat-rich protein	chlo
Glyma17G10750 0	17	GmANK 134	5	634	69.58	6.73	S-acyltransferase	plas

Glyma17G118300	17	GmANK135	5	876	98.11	7.3	Potassium channel	plas
Glyma17G207900	17	GmANK136	5	867	97.56	6.23	Shaker-like potassium channel 1	plas
Glyma18G005100	18	GmANK137	2	963	108.04	5.4	Calmodulin-binding family protein	nucl
Glyma18G008500	18	GmANK138	11	762	83.11	7.5	Ankyrin repeat domain-containing protein	cyto
Glyma18G009500	18	GmANK139	8	652	71.55	6.87	Serine/threonine protein phosphatase 6 regulatory ankyrin repeat subunit C	nucl
Glyma18G044300	18	GmANK140	5	332	36.81	5.11	Ankyrin repeat family protein	chlo
Glyma18G062200	18	GmANK141	2	609	66.47	5.96	Zinc finger CCCH domain-containing protein	nucl
Glyma18G100300	18	GmANK142	5	727	81.37	8.86	Nuclear factor NF-kappa-B p105 subunit	plas
Glyma18G176600	18	GmANK143	6	442	47.78	6.96	Hop-interacting protein THI012	nucl
Glyma18G282000	18	GmANK144	2	840	93.81	7	ADP-ribosylation factor GTPase-activating protein AGD3	nucl
Glyma19G086700	19	GmANK145	11	755	81.24	6.51	Ankyrin repeat domain-containing protein	cyto
Glyma19G113700	19	GmANK146	2	1078	117.28	8.99	Ankyrin repeat domain-containing protein	nucl
Glyma19G114600	19	GmANK147	8	528	57.52	7.53	NA	chlo
Glyma19G131000	19	GmANK148	2	414	45.48	6.54	BTB/POZ ankyrin repeat protein	cyto
Glyma19G141700	19	GmANK149	2	668	73.47	6.23	Zinc finger CCCH domain-containing protein	nucl
Glyma19G171500	19	GmANK150	5	311	34.7	8.28	Ankyrin repeat family protein	nucl
Glyma19G171600	19	GmANK151	3	270	30.61	7.84	Ankyrin repeat family protein	nucl
Glyma19G175300	19	GmANK152	7	582	63.94	9.55	Ankyrin repeat-containing protein	plas
Glyma19G175400	19	GmANK153	8	531	58.2	9.08	Ankyrin repeat-containing protein	plas
Glyma19G245700	19	GmANK154	7	461	49.04	5.04	Ankyrin repeat family protein	chlo

Glyma19G26370	19	GmANK 155	8	559	61.46	9.51	Ankyrin repeats-containing protein	plas
Glyma20G03130	20	GmANK 156	2	424	48	7.68	Mitogen-activated protein kinase kinase kinase Raf17.1	cyto
Glyma20G07660	20	GmANK 157	2	549	61.65	9.44	Ankyrin repeat family protein	nucl
Glyma20G15660	20	GmANK 158	5	513	54.93	6.87	Ankyrin repeat-containing protein	chlo
Glyma20G20940	20	GmANK 159	2	458	49.03	5.96	Zinc finger family protein	cyto
Glyma20G24120	20	GmANK 160	7	586	63.58	8.83	NA	plas
GlymaU025300	X	GmANK 161	6	453	50.14	8.9	Ankyrin repeat protein	ER
GlymaU033200	X	GmANK 162	8	550	59.2	9.42	Ankyrin repeat-containing protein	plas

Note: Chromosome:Chr, Number of ANK: No.ANK ,Length(amino acids):Length(aa),molecular weight(kiloDalton):MW(KD), isoelectric points:PI, X presents the unknown,SL: Subcellular localization

Figure legends

Fig.1. Structure and domain composition of representative GmANK proteins from each subfamily.

Domain abbreviations are:ANK ankyrin repeat domain; TM transmembrane; TPR tetratricopeptide repeat domain; RING ring finger domain; ZnF zinc finger; BTB broad-complex, tramtrack and bric a brac domains; IQ calmodulin-binding domain; BAR BAR domain; PH pleckstrin homology domain;ArfGap putative GTP-ase activating proteins for the small GTPase,ARF; AAA ATPase family associated with various cellular activities; CHROMO CHRromatin organisation MODifier domain; cNMP cyclic nucleotide-binding domain;CG1 CAIModulin -binding Transcription Activator. The length and order of domains represent actual situation in each protein

Fig. 2.Sequence logos for the GmANK repeats motifs of soybean.

The number of X-axis represent the sequence positions in the ANK repeats motifs. The Y-axis represents the information content measured in bits

Fig.3.Phylogenetic relationships of 158 soybean ANK repeats proteins. The tree was constructed using the MEGA5.0 software.

Fig.4.Expression analysis of *GmANK* genes in the different developmental stages of soybean.

Hierarchical cluster displaying expression profiles for 162*GmANK* genes in 14 tissues or developmental stages.The tissues and color are indicated on the top of expression heatmap. Different color lines representing different expression patterns for each gene.

Fig.5. Normalized expression of differentially expressed *GmANK* genes in *Glycine*

soja and Glycine max

(a).Heatmap of differentially expressed *GmANK* genes based on their relative gene expression values (Z-score) computed using DGEP data under 0 and 200 mM NaCl stress.

(b). Heatmap of differentially expressed *GmANK* genes based on their relative gene expression values (Z-score) computed using DGEP data under 0 and 20%PEG stress. Red to green color indicates high to low expression levels.

Fig.6.The expression pattern of *GmANK-39-Like* in different tissues and under different abiotic stresses in soybean

(a).The expression pattern of *GmANK-39-Like* in root, stem, leaf, flower and pod at seedling, flowering and podding stages, respectively using quantitative RT-PCR.

(b).The expression pattern of *GmANK-39-Like* in soybean roots under PEG6000 for 0 h, 2 h, 4 h and 8 h.

(c).The expression pattern of *GmANK-39-Like* in soybean roots under 200mM NaCl for 0h, 1h, 2h, 4h, 8h, 12h,24h,48h and 72h.

* p <0.05 level; **, p<0.01.

Fig.7.Subcellular localization of *GmANK-39-Like* protein in *Arabidopsis* protoplasts.

The *GmANK-39-Like*-green fluorescent protein (p35S::*GmANK-39-Like*-GFP) fusion construct was generated by inserting *GmANK-39-Like* into pJIT166-GFP without a termination codon to creating an in-frame fusion between the CDS and GFP. p35S::*GmANK-39-Like*-GFP and the GFP control plasmid (p35S::*GFP*) were transformed into *Arabidopsis* protoplasts by the PEG4000-mediated method .

subcellular localization demonstrated that GFP was detected all the whole cells for the 35s::*GFP*, but mainly distributed cytoplasm for 35S::*GmANK-39-Like*:GFP Scale bars = 20 μ m.

Fig 1

Fig2

Fig 3

Fig 4

Fig5

Fig6

Fig 7