Genome-wide identification and analysis of the *CNGC* gene gamily in maize

Lidong Hao[1], Xiuli Qiao[2]*

[1]College of Agriculture and Hydraulic Engineering, Suihua University, 18 Huanghe Road, Suihua, Heilongjiang Province, 152000, PR China

[2]College of Food and Pharmaceutical Engineering, Suihua University, 18 Huanghe Road, Suihua, Heilongjiang Province, 152000, PR China

*Corresponding author:

Xiuli Qiao

Email: shxynxhld@163.com

12 **Abstract**

13 As one of the non-selective cation channel gene families, the cyclic nucleotide-gated channel
14 (CNGC) gene plays a vital role in plant physiological processes that are related to signal pathways,
15 plant development, and environmental stresses. However, genome-wide identification and analysis
16 of the *CNGC* gene family in maize has not yet been conducted. In the present study, twelve
17 *ZmCNGC* genes were identified from the maize genome, which were unevenly distributed on
18 chromosomes 1, 2, 4, 5, 6, 7, and 8. They were classified into five major groups as Group I, II, III,
19 IVa, and IVb. Phylogenetic analysis showed that gramineous plant CNGC genes expand unequally
20 during evolution. And Group IV CNGC genes emerged the earliest while Group I and II CNGCs
21 appeared later. Prediction analysis of cis-acting regulatory elements showed that 137 putative cis-
22 elements were related to hormone-response, abiotic stress, and organ development. Furthermore,
23 120 protein pairs were predicted to interact among the 12 ZmCNGC proteins and other maize
24 proteins. The expression profiles of ZmCNGC genes were shown to express in tissue-specific
25 patterns. Our results provide valuable information to increase our understanding of the *CNGC* gene
26 family in maize and other plants.

27

INTRODUCTION

During the process of organism evolution, the formation of complex nutrient absorption and transport systems including ion channels, ion pumps, and carriers. It has been previously shown that these systems respond to endogenous and abiotic stimuli (Saand et al. 2015b). The cyclic nucleotide-gated channel (CNGC) is a Ca^{2+}-permeable cation transport channel, and has been suggested as being one of the fundamental mechanisms in organism systems (Yuen and Christopher 2013; Nawaz et al. 2014). Acting as molecular switches, secondary messengers such as cyclic nucleotide monophosphates (3',5'-cAMP and 3',5'-cGMP) and Ca^{2+}/calmodulin (CaM) can regulate CNGCs. CNGCs are activated by direct binding of cyclic nucleotides as well as by being inhibited by the binding of CaM to the CaM binding domain (Saand et al. 2015b; Borsics et al. 2007; Defalco et al. 2016; Kaplan et al. 2007).

In plants, CNGCs are composed of six transmembrane (TM) domains and one pore region between the fifth and sixth TM domains. The cyclic nucleotide-binding domain (CNBD) is a highly conserved region and carries a phosphate-binding cassette (PBC) and hinge region (Saand et al. 2015b). However, the existence of these domains is a necessary but not significant condition with which to judge a CNGC protein because other ion transporters such as potassium AKT/KAT channels (Shaker type) also contain both a CNBD and a TM domain (Su et al. 2001; Chérel 2004). To address this issue, previous studies have proposed that a plant CNGC-specific motif, [LIMV0]-X(2)-[GSANCR]-X-[FVIYASCL]-X-G-X(0,1)-X(0,1)-[EDAQGH]-L-[LIVFA]-X-[WRCMLS0]-X-[LMSIQAFT0]-X(7,37)-[SAC]-X(9)-[VTIALMS]-X(0,1)-[EQDN]-[AGSVT]-[FYL]-X-[LIVF], in the PBC and hinge region within the CNBD of CNGC proteins only exists in plant CNGCs and not in other ion transporters (Zelman et al. 2013; Saand et al. 2015b).

The first plant CNGC was identified from *Hordeum vulgare* and named *HvCBT1* (Schuurink et al. 1998). With the application of bioinformatics tools, the identification of the *CNGC* gene family in *Arabidopsis (20)*, rice (16) and other plants (18 in tomato and 26 in *Brassica oleracea*) (Bridges et al. 2005; Nawaz et al. 2014; Ward et al. 2009; Zelman et al. 2012; Saand et al. 2015a; Chen et al. 2015; Zelman et al. 2013; Guo et al. 2017; Kakar et al. 2017).

CNGCs are key components of plant development, with many previous studies conducted. Most

*CNGC*s have been characterized by genetic methods and found to be related to plant physiological and molecular functions, including playing a vital role in multiple physiological processes involved in signal pathways, plant development, and environmental stresses. For example, *Arabidopsis CNGC7* and *CNGC8* are essential for male reproductive fertility (Tunc-Ozdemir et al. 2013a); *CNGC16* and *CNGC18* participate in pollen development (Tunc-Ozdemir et al. 2013b; Frietsch et al. 2007; Gao et al. 2016); *AtCNGC2* is involved in jasmonic acid induced apoplastic Ca(2+) influx in epidermal cells (Lu et al. 2015; Wang et al. 2017); and *Arabidopsis CNGC6*, *CNGC19*, and *CNGC20* are involved in abiotic stress (Kugler et al. 2009; Gao et al. 2012). *Arabidopsis* CNGCs structures have six TM domains with a pore domain, and they also possess a cyclic nucleotide-binding domain and CaM-binding domain in the C-terminus. These various domains have diverse functions (Talke et al. 2003; Chin et al. 2009a; Hua et al. 2003; Köhler and Neuhaus 2000). For example, *AtCNGC2* plays a key role in stress signaling pathways, including changes to the cytosolic free Ca^{2+} in *Arabidopsis*, and *CNGC4* is permeable to both K+ and Na+, and is activated by both cGMP and cAMP (Balague 2003; Ali et al. 2007).

In recent years, efforts had been made to study the CNGC gene family in plants, yet even though it is as one of the most important food crops and sources of industrial materials worldwide, the maize *CNGC* gene family has rarely been reported. In the present study, utilizing maize genome-wide sequence information and research information on *Arabidopsis* and rice CNGC families, we conducted genome-wide identification of CNGCs in maize via comprehensive bioinformatics analyses. To the best of our knowledge, this is the first systematic study of *CNGC* genes in maize and provides the basis for future research on the *ZmCNGC* gene family.

MATERIALS AND METHODS

Identification of *CNGC* genes in maize genome

To identify the *CNGC* genes in maize, 20 *Arabidopsis* and 16 rice *CNGC* protein sequences were retrieved from the *Arabidopsis* Information Resource (TAIR10) database (http://www.arabidopsis.org/index.jsp) and the Rice Genome Annotation Project (RGAP)

database (http://rice.plantbiology.msu.edu/), respectively. Two methods were applied to search the maize protein sequences, one was built using a Hidden Markov Model (HMM) to search against maize protein sequences and the other used the local BLASTP method with a threshold of e-value < 1e-5. After the searches were conducted, a manual correction was performed to remove any redundancy and proteins without PBC and hinge regions within the CNBD of CNGC proteins were removed. To further confirm whether the ZmCNGC proteins contained the CNBD domain, those putative ZmCNGC protein sequences were submitted to SMART (Letunic and Bork 2018) and NCBI-CDD (Marchler-Bauer et al. 2017), where proteins without CNBD domains or with amino acids (aa) below 200 were removed and the *ZmCNGC*s confirmed. The identification another 11 gramineae plant CNGCs was also performed applying the same method as that described above. The PI (theoretical isoelectric point), MW (molecular weight), and GRAVY of ZmCNGCs were predicted by ExPASy (Artimo et al. 2012). CELLO v.2.5 software was used to predict the subcellular location of the ZmCNGCs (Yu et al. 2010). The information on chromosome distribution of ZmCNGCs and the sequences including DNA sequences, CDS, cDNA, and up-stream 1,500 base pairs (bps) of the ZmCNGCs were obtained from the BLASTN search results in the Ensembl Plant database (Bolser et al. 2016).

Multiple alignments, phylogenetic analysis and gene duplication analysis

Multiple sequence alignments were performed using the T-COFFEE method (Di Tommaso et al. 2011) and visualized by ESPript using the default program setting (Robert and Gouet 2014). A maximum likelihood (ML) phylogenetic tree was constructed using the MEGA 7 software program with 1,000 bootstrap replications and the Jones-Taylor-Thornton model (Kumar et al. 2016). To further validate the accuracy of the ML tree, an un-root phylogenetic tree was constructed with 1,000 bootstrap replications using the MEGA 7 software program based on full-length protein sequence alignments. ML analysis of 12 gramineae plants was performed using the IQTree program with LG+G4 and state frequencies determined from the amino acid matrix and other default parameters (Lam-Tung et al. 2015) and visualized by Evolview online (He et al.

2016). Segmental duplication between maize genes as well as the synteny block between maize and *Sorghum*, rice and *Brachypodium* were obtained from the Plant Genome Duplication database (Lee et al. 2013). The substitution rates (Ka/Ks) of duplication events were calculated by using the DnaSP v5 software program (Librado and Rozas 2009), and the divergence times (Mya) were calculated as a formula $Mya = Ks/2\lambda \times 10^{-6}$, therein $\lambda = 6.5 \times 10^{-9}$ (Lynch and Conery 2000).

Analyses of gene structure and conserved motifs

The gene structure (exon-intron) of *ZmCNGC* genes were determined by the Gene Structure Display Server (Hu et al. 2015) using the CDS and DNA sequences of the *ZmCNGC* genes. The conserved motifs of ZmCNGC proteins were used in the MEME Suite web server (Bailey et al. 2015) with the maximum number of motifs set at 10 and the optimum width of motifs from 6 to 200 aa.

Cis-acting regulatory elements and the prediction of protein-protein interaction in ZmCNGCs

The up-stream 1,500bp DNA sequences of the *ZmCNGC* genes were used to locate cis-acting regulatory elements by 'Signal Scan Search' programs in the NEW PLACE database (Higo et al. 1999). The interactions between ZmCNGCs and other maize proteins were predicted with the STRING 10 online program (Szklarczyk et al. 2017) and visualized by the Cytoscape v3.4.0 software program (Shannon et al. 2003).

Expression profiles of *ZmCNGC* genes and network interaction analysis

To understand the expression of *ZmCNGC* genes in different tissues, two high throughput datasets of maize were obtained from the Expression Atlas datasets (https://www.ebi.ac.uk/gxa/home/) under accession numbers E-MTAB-3826 and E-MTAB-439. These data were used to analyze the expression of *ZmCNGC* genes in six different tissues (ear, embryo, endosperm, pollen, root, and tassel) and different developmental stages (embryo, endosperm, and seed). FPKM values were

used to calculate each *ZmCNGC* genes and were visualized by OmicShare tools, which is a free online platform for data analysis (http://www.omicshare.com/tools).

RESULTS

Identification of *CNGC* genes in maize

To identify the complete overview of *CNGC* genes in maize, we first used 20 *Arabidopsis* and 16 rice CNGC protein sequences that were BLAST aligned against maize protein sequences. After BLAST alignment, a total of 18 putative *ZmCNGC* genes were identified in the maize genome. Then, to confirm these 18 putative *ZmCNGC* genes, we used the SMART and NCBI CDD to determine whether they contained the CNGC-specific domains (CNBD and TM). After removing redundancy genes, a total of 12 *ZmCNGC* genes were detected, which was lower than the number of rice and *Arabidopsis* ~~CNGC gene~~ (Paterson et al. 2004; Yu et al. 2005). To better understand the following analysis, the predicted *ZmCNGC* genes were designated as ZmCNGC1 to ZmCNGC12 based on family classification (Table 1). To further access the existence of the ZmCNGCs, we identified all the expressed sequence tags (EST) that had aligned to the *ZmCNGC* genes using the BLASTN program from NCBI. The results demonstrated that only ZmCNGC3 showed no EST hits, whereas the other ZmCNGCs had more than 13 representative matches to ESTs. Five of them were located on chromosome 5, whereas the others were unevenly located on chromosomes 1, 2, 4, 6, 7, and 8. The characteristic features of these 12 *ZmCNGC* genes are listed in Table 1. The ZmCNGCs lengths ranged from 326 to 745 aa with an average of 612 aa. The molecular weight of these proteins ranged from 38.63 kDa (ZmCNGC2) to 85.52 kDa (ZmCNGC8) and the pI values ranged from 8.92 (ZmCNGC4) to 9.75 (ZmCNGC12). Subcellular localization prediction analysis showed that all the ZmCNGCs were localized in the plasma membrane except for ZmCNGC3 that was localized in the nuclear fraction. The result is consistent with that of *Arabidopsis* (Lemtiri-Chlieh and Berkowitz 2004).

Multiple alignments of maize CNGCs and potassium AKT/KAT channel genes

Many ion transporters other than CNGCs also have CNBD in the C-terminus and a hexa- TM in the N-terminus, for example, potassium AKT/KAT channels (Shaker type) also contain a CNBD and a TM domain. All AKT/ KAT-type channels consist of six TM regions with one P region (Su et al. 2001). Therefore, we aligned the 11 AKT/KAT proteins of maize from NCBI and 12 ZmCNGC protein sequences (Supplemental File 1). Results showed that they were highly conserved and all of them were composed of six TM domains (S1-S6) and a pore region, which were not highly conserved in the PBC and hinge domain. A ML phylogenetic tree indicated that maize CNGC and AKT/KAT-type channel genes were clustered into two separate sections (Supplemental File1).

The CNBD is a gene structural feature element in plant CNGCs that contains the PBC and the hinge region (Diller et al. 2001). As shown in Figure 1, within the maize PBCs, a conserved phenylalanine (F), a stabilizing glycine (G) and an acidic residue (D or E), as well as two aliphatic leucines (L) were 100% conserved inside the PBCs. Additionally, aromatic phenylalanine (F) and leucine (L) were 100% conserved within the hinge region. These two conserved regions occurred between the CNBD and CaMBD regions. Based on the corresponding alignment, a stringent motif (L-X(2)-G-[ED]-ELL-[TSG]-W-[ACY]-L-X(10,20)-[SA]-X-T-X(7)-[EQ]-[AG]-F-X-L) which was recognized in the 12 maize CNGCs that included the PBC and hinge domain, was consistent with other plant species (Saand et al. 2015b; Nawaz et al. 2014). The alignment of maize, rice, and *Arabidopsis* CNGCs was also performed. Results showed that there were no positions that were specific to the maize CNGC consensus, which suggested that the PBCs and hinge domain were highly conserved among plants (Supplemental File 2).

Phylogenetic and duplication analyses of ZmCNGCs

Large phylogenetic trees with minimal homologous characters increase the likelihood of confounding relationship between different species; therefore, we constructed a ML phylogenetic tree based on the alignment of 12 gramineae plants CNGC proteins, which included 24 in *Aegilops tauschii*, 16 in *Brachypodium distachyon*, 20 in *Hordeum vulgare*, 23 in *Leersia perrieri*, 21 in

192 *Nicotiana attenuata*, 16 in *Oryza sativa,* 28 in *Setaria italic*, 13 in *Sorghum bicolor*, 79 in *Triticum*
193 *aestivum*, 21 in *Triticum urartu* and 12 in maize CNGCs in the present study (Supplemental File
194 5: Table 1). Based on the phylogenetic tree (Supplemental File 4), we clustered these plant CNGC
195 proteins into six groups, named Group I, II, III, IV, Iva, and IVb with significant bootstrap values.
196 Results showed that CNGC proteins of *B. distachyon*, *O. sativa*, *S. bicolor* and *maize* did not
197 cluster in Group IV, and *N. attenuata* CNGC did not cluster in Group I. Additionally, the number
198 in each group was unevenly distributed. Group IV was the largest with 86 genes, followed by 69
199 in Group III, 44 in Group II, 33 in Group IVb, 25 in Group I and 16 in Group IVa. These data
200 demonstrate that gramineae plant CNGC gene expansion occurred unequally during evolution.
201 To better understand the evolutionary relationship among CNGC proteins, a ML phylogenetic tree
202 was created based on the full-length protein alignments of 12 *ZmCNGCs*, 19 *AtCNGCs* and 16
203 *OsCNGCs* (Nawaz et al. 2014; Maser et al. 2001). As shown in the phylogenetic tree, the 47 CNGC
204 proteins were classified into five groups with significant values (Figure 2), named Group I, II, III,
205 IVa, and IVb. This is consistent with what has been previously reported for flowering plant CNGCs
206 (Saand et al. 2015b). Group I contained three maize *CNGC* genes (*ZmCNGC1*, *ZmCNGC2*, and
207 *ZmCNGC3*), five in *Arabidopsis*, and two in rice. Group II contained two maize CNGC genes
208 (*ZmCNGC4* and *ZmCNGC5*), five in *Arabidopsis* and three in rice. Group III contained three maize
209 *CNGC* genes (*ZmCNGC6*, *ZmCNGC7*, and *ZmCNGC8*). Group IV included five CNGC genes in
210 rice and four in *Arabidopsis*, which were separated into one in Group IVa (*ZmCNGC9*) and three
211 in Group IVb (*ZmCNGC10*, *ZmCNGC11*, and *ZmCNGC12*).
212 Based on the phylogenetic tree, maize *CNGC* genes were also grouped into five groups (Figure
213 3A). Consistent with other plant *CNGC* genes, Group IVa contained only one or two gene members
214 and formed the smallest group (Saand et al. 2015b). Results suggested that during evolution of
215 CNGCs, Group IV *CNGC* genes emerged the earliest whereas Group I and II CNGC genes
216 appeared later. Furthermore, tree topology resulting from neighbor joining analyses was the same
217 as the ML tree in Figure 2, with all the groups retrieved (Supplemental File 3).
218 Additionally, only one segmental duplication gene pair, ZmCNGC10-ZmCNGC11, was formed

in the maize genome in the present study (Table 2). To further investigate the evolutionary process between maize CNGCs and other gramineae plants, the genome synteny among *Sorghum*, rice, and *Brachypodium* were explored. Results showed that there were two, two and one *ZmCNGC* genes that showed syntenic bias towards particular chromosomes of *Sorghum*, rice, and *Brachypodium*, respectively (Table 2). In addition, Ka/Ks was used to evaluate the specific positions under positive selection pressure after duplication (Mayrose et al. 2007). Ka/Ks =1, < 1 or > 1 indicates neutral, purifying, and positive selection (Lynch and Conery 2000). The Ka/Ks of each gene pair was calculated and the Ka/Ks of all gene pairs was less than 1, suggesting that the selection pressure after duplication was strongly purifying. Moreover, results indicated that the divergence time between maize CNGCs and other gramineae plants was unevenly divergence.

Gene structure and conserved motif of ZmCNGCs

Gene structure analysis could improve our understanding of gene function and evolution. The number of introns ranged from 0 to 7 (Figure 3B), which was different from rice and *Arabidopsis* CNGCs. In rice, OsCNGC ranged from 1 to 11 introns, whereas *Arabidopsis* CNGCs ranged from 4 to 10 introns (Nawaz et al. 2014). The Group IVa and IVb ZmCNGCs had distinct gene structures compared to those of the other groups, with more introns at different phases and lengths, which is consistent with most flowering plant species Group IV CNGC genes (Saand et al. 2015b). Furthermore, motif-based recognition of proteins provides understanding of the evolution history (Seoighe and Gehring 2004). Ten putative motifs were characterized and named as motif1 to motif10 in the ZmCNGCs. The relative positions of the motifs in the five groups were found to have various patterns (Figure 3C and 3D). All the ZmCNGCs harbored motif1, motif2, motif3 and motif4, which represent they were the typical ZmCNGC domain, motif3 was the combination of the calmodulin binding (CaMB) domain and IQ domain (QWRTWAA[CV]FIQ[AL]AW[RH]RY), and motif4 was the cyclic nucleotide-binding (CNB) domain which located in C-terminal. In addition, motif10, 9, 5, 8(6), 2, and 1 were the motif logo of the transmembrane domains, they represented the S1, S2, S3, S4, S5, and S6 of the

transmembrane domain in N-terminal. And motif7 represent the ion transport protein (ITP) domain. In addition, 10 ZmCNGCs except *ZmCNGC2* and *ZmCNGC3* possessed motif 5 and 6 that are associated with ion transport (Nawaz et al. 2014). Although other motifs have not been reported in plants or animals, they hold an important position in the function of the organism.

Prediction of cis-acting regulatory elements and protein-protein interaction of ZmCNGC proteins

Cis-acting regulatory elements are important molecular switches that are associated with the transcriptional regulation of genes when environmental stresses encountered (Nakashima et al. 2009). To better understand the possible biological processes of the ZmCNGCs involved, 1.5 kb upstream of *ZmCNGC* genes genomic sequences were used to identify cis-regulatory elements. A total of 137 different putative cis-elements were found to be associated with the identified *ZmCNGC* genes and only 12, including CACTFTPPCA1, EBOXBNNAPA, DOFCOREZM, MYCCONSENSUSAT, CAATBOX1, GTGANTG10, WRKY71OS, GT1CONSENSUS, ROOTMOTIFTAPOX1, POLLEN1LELAT52, MYBCORE, and OSE2ROOTNODULE were in the promoter region of all *ZmCNGC* genes (Supplemental File 5: Table S2) and which was highly consistent with rice CNGCs. Additionally, five cis-elements were gene-specific, with ACGTCBOX, TATABOX3, CTRMCAMV35S and HDZIP2ATATHB2 ~~were~~ unique to *ZmCNGC6*, *ZmCNGC7*, *ZmCNGC8*, *ZmCNGC11*, respectively. Additionally, some cis-elements were involved in different abiotic/biotic stimuli, including hormone-response (e.g. abscisic acid, auxin, ethylene), stress-related (e.g. drought, temperatures, disease) and development-related (e.g. mesophyll specific, tissue specific), indicating that these *ZmCNGC* genes might be involved in regulating diverse stress responses.

To better understand the role of ZmCNGC proteins, the prediction of ZmCNGC protein-protein interaction was performed. A total of 120 protein pairs were predicted to interact between the 12 ZmCNGC proteins and the other 11 maize proteins, and ZmCNGC8 was found to interact with *ZmCNGC10* by prediction analysis (Figure 4 and Supplemental File 5: Table S3). We also used

the homologous genes of ZmCNGCs in *Arabidopsis* to predict the protein-protein interaction. Results showed that three were validated by experimental (Supplemental File 5: Table S3).

Expression profiles of *ZmCNGC* genes in different tissues

To investigate the physiological function of *ZmCNGC* genes, we performed transcriptome sequencing to evaluate tissue-specific expression levels of *ZmCNGC* genes in different tissues based on previous methodology (see Materials and Methods section). As shown in Figure 5A and Supplemental File 5: Table S4, the expression levels among *ZmCNGC* genes were tissue-specific in their expression. For example, *ZmCNGC5* was specifically expressed in pollen compared to other tissues, implying it play particular role in the physiological process of pollen development. *ZmCNGC2*, *ZmCNGC4*, *ZmCNGC6*, and *ZmCNGC8* had higher expression in the roots than in the other tissues, suggesting their important roles in root growth and development. All Group IVb *ZmCNGC* genes, including *ZmCNGC10*, *ZmCNGC11*, and *ZmCNGC12*, were relatively high in the embryo, implying that these genes play crucial roles in growth and development of maize embryo.

We also evaluated several *ZmCNGC* genes in the embryo, endosperm, and seed expression in several days after pollination. *ZmCNGC3*, *ZmCNGC5* and *ZmCNGC7* were not detected or had no expression in any of the tissues; however, *ZmCNGC5* did show expression in the pollen (Fig. 5A, B, C, and D). Cis-acting regulatory elements analysis showed that only *ZmCNGC5* and *ZmCNGC7* did not contain CANBNNAPA, which is the element required for embryo and endosperm-specific transcription (Ellerström et al. 1996). This might be the reason why these did not show any expression in these tissues. The embryo specific-expression gene *ZmCNGC10* gradually increased over time in the embryo (Figure 5B). *ZmCNGC8* was highly expressed in the embryo, endosperm, and seeds after pollination, and the Group IVb gene showed a similar expression pattern.

DISCUSSION

Features and evolution of plant CNGC family genes

Plant CNGC family genes are characterized by the presence of a CNBD at the C-terminal and a hexa-TM at the N-terminals (Saand et al. 2015b). After blast against the maize genome protein sequences, a total of 12 *ZmCNGC* genes were identified. Among them, ZmCNGC3 showed no EST and did not have expression in all tissue in the present study, suggested that it was a non-expressed pseudogene. In *Arabidopsis*, AtCNGC16 (AT3G48010) showed no EST and might be a pseudogene (Mäser et al. 2001). Many ion transporters other than CNGCs also possess these domains; for example, potassium AKT/KAT channels (Shaker type) contain both a CNBD domain and a TM domain. All AKT/ KAT-type channels consist of six transmembrane (TM) regions with one P region (Su et al. 2001). A ML phylogenetic tree showed that the maize CNGC and AKT/KAT-type channel genes were clustered into two separate sections (Supplemental File 1). Previous studies have shown that the CNGC-specific motif with PBC and hinge domain (L-X(2)-G-[ED]-ELL-[TSG]-W-[ACY]-L-X(10,20)-[SA]-X-T-X(7)-[EQ]-[AG]-F-X-L) only exists in plant CNGCs rather than in other ion transporters, for example, in rice, *Arabidopsis*, and tomato (Saand et al. 2015b; Nawaz et al. 2014; Zelman et al. 2013). In the present study, 12 ZmCNGCs were identified and the PBCs and hinge domains were highly conserved after aligning, which further confirmed the previous hypothesis (Saand et al. 2015b). Conserved motif analysis showed that the motif3 (QWRTWAA[CV]FIQ[AL]AW[RH]RY) pattern was the IQ domain among these 12 ZmCNGCs. IQ domain is conserved among plant CNGCs and enhances the changeable Ca^{2+}-dependent channel control mechanisms in plants (Fischer et al. 2013). In the present study, all CNGC proteins contained the IQ motif, suggesting that those proteins bind CaM in a Ca^{2+}-dependent manner. Notably, motif4 is the sequence logo of the CNBD domain, which is conserved the most plants and animals (Jackson et al. 2007).

We observed that the CNGC family in 12 gramineae plants at various evolutional nodes provides a good platform for analyzing the phylogeny and evolution of the CNGC gene family in gramineae plants. As showen in Supplemental File 5: Table 1, 4 out of 12 plants under analysis contained < 20 CNGC genes, including 16 in *B. distachyon*, 16 *O. sativa*, 13 *S. bicolor* and 12 in maize. The ML tree for the 273 CNGCs clearly showed that gramineae CNGCs clustered into six groups (I,

II, III, IV, IVa, and IVb) with significant bootstrap values, Group IV, IVa and IVb were with high support value, but group IV CNGCs did not cluster *B. distachyon*, *O. sativa*, *S. bicolor* and maize CNGCs. This reveal that CNGCs are likely missing or there was duplication during evolution. Davidson et al. (2013) showed that the gene duplication was a significant origin for generating novel genes, thus, these results imply that the duplication events play a principal role in gene evolution. Based on the phylogenetic, gene structure, and conserved motifs analyses, we classified 12 *ZmCNGC*s into five groups with significant values. Among these five groups, Group IV *ZmCNGC*s emerged the earliest and had more introns than the genes in the other groups, followed by Groups I and II CNGCs genes during plant evolution.

CNGC genes play an important role in plant development

CNGC genes are involved in the regulation of plant growth and development (Chin et al. 2009b). In the present study, we focused on the role that maize CNGCs play in different tissues, including in the plant embryo, endosperm and seed expression. We found that some *ZmCNGC* genes were expression specific, e.g. *ZmCNGC2*, *ZmCNGC4*, *ZmCNGC6* and *ZmCNGC8* had high expression in roots, *ZmCNGC5* was specifically expressed in the pollen, and all Group IVb genes were expressed in the embryo. As the ancestor of CNGC genes, Group IVb genes were mainly expressed in the embryo and seeds after pollination, which suggested that Group IVb genes play a significant role in embryo development. In addition, all Group IVb genes were related to gene duplication and they had a similar expression pattern in different tissue of this study, which indicated that they obtained functions to enhance adaptability during gene evolution.

The majority of previous studies have shown that *CNGC* genes are related to pollen development and response to environmental stimuli. For example, *Arabidopsis CNGC16* is critical for pollen fertility under conditions of heat stress and drought stress (Tunc-Ozdemir et al. 2013b), and *CNGC18* has been shown to function in pollen tube tip growth (Frietsch et al. 2007). In rice, *OsCNGC13* promotes the seed-setting rate by facilitating pollen tube growth in stylar tissues (Xu et al. 2017). *ZmCNGC1* and *ZmCNGC5*, two homologous genes of *CNGC16* and *CNGC18* in the

present study are mainly expressed in pollen, indicating that they are mainly involved in pollen development. Previous studies have shown that the *AtCNGC3* is mainly expressed in the embryo, leaves, and roots. The expression level of *ZmCNGC4* is consistent with that of *AtCNGC3* which was highly expressed in plant development except for in pollen (Kaplan et al. 2007). These results imply that these genes play crucial roles during the growth and development of maize.

CONCLUSION

In the present study, a total of 12 CNGC genes were identified in maize using the bioinformatics method based on the presence of plant CNGC-specific motif spanning the PBCs and hinge domain with the CNBD of CNGC proteins. Phylogenetic analyses showed that Group IV ZmCNGCs emerged the earliest and had more introns than that of the other group ZmCNGCs, whereas Groups I and II evolved later. Phylogenetic analysis of 12 gramineae plants revealed that CNGCs are likely missing or duplicating during evolution. Significantly, *ZmCNGC* genes were diverse in gene structure, protein length and size. We modified a maize stringent motif (L-X(2)-G-[ED]-ELL-[TSG]-W-[ACY]-L-X(10,20)-[SA]-X-T-X(7)-[EQ]-[AG]-F-X-L) that included the PBC and hinge domain. Expression profiles of *ZmCNGC* genes were tissue-specific expressed and related to pollen development. Our results provide a reference for plant CNGCs during gene evolution.

ACKNOWLEDGEMENTS

We extend our gratitude to Dr. Lihe Yu for providing valuable suggestions.

ADDITIONAL INFORMATION AND DECLARATIONS

Supplemental Information

Supplemental File 1

A ML phylogenetic tree, conserved regions and multiple sequences alignments between maize AKT/KAT channels genes and ZmCNGCs.

Supplemental File 2

Multiple sequences alignments among maize, *Arabidopsis* and rice CNGCs by using T-COFFEE method.

Supplemental File 3

The un-root neighbor joining (NJ) tree based on the maize, rice and *Arabidopsis* CNGCs protein sequences by using MEGA7.

Supplemental File 4

The Maximum likelihood (ML) tree based on the maize and other 11 gramineous plants CNGCs protein sequences by using IQTREE.

Supplemental File 5

Table S1 The GNGC information of 12 gramineae plants

Table S2 Numbers of known stress-related elements in the promoter regions of ZmCNGCs

Table S3 The protein-protein interaction information of maize CNGC protein with other maize proteins

Table S4 The FPKM data of ZmCNGC genes in different tissues.

Supplemental File 6

The logo of ten motifs in study.

Supplemental File 7

The gene sequences used in this research.

Data Availability

The following information was supplied regarding data availability:

The raw data has been supplied as a Supplemental File 7.

REFERENCES

Ali R, Ma W, Lemtiri-Chlieh F, Tsaltas D, Leng Q, Bodman SV, Berkowitz GA (2007) Death Don't Have No Mercy and Neither Does Calcium: Arabidopsis CYCLIC NUCLEOTIDE GATED CHANNEL2 and Innate Immunity. The Plant cell 19 (3):1081-1095

Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, Stockinger H (2012) ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res 40 (Web Server issue):W597-603. doi:10.1093/nar/gks400