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

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

[1] College of Agriculture and Hydraulic Engineering, Suihua University, Suihua, HeiLongjiang province, China
[2] College of Food and Pharmaceutical Engineering, Suihua University, Suihua, HeiLongjiang province, China

Corresponding Authors: Lidong Hao, Xiuli Qiao
Email address: haolidong1987@163.com, shxynxhld@163.com

As one of the non-selective cation channel gene family, the cyclic nucleotide-gated channel (CNGC) gene plays vital roles in plant physiological processes which are related to signal pathways, plant development, and environment stresses. However, genome-wide identification and analysis of the *CNGC* gene family in maize has not yet been conducted. In this study, 12 *ZmCNGC* genes are detected in the maize genome, which are unevenly distributed on chromosomes 1, 2, 4, 5, 6, 7 and 8. They are classified into four major groups, including group I, II, III, and IV. Prediction of cis-acting regulatory elements show that 137 putative cis-elements which are related to hormones-response, abiotic stress and organ development. Synteny analysis showed that 2, 2 and 1 ZmCNGCs had homologous genes in *Sorghum*, rice and *Brachypodium*, respectively. Gene ontology (GO) analysis demonstrated that most of the ZmCNGCs are involved in various biological processes including cellular processes, establishment of localization, and transmembrane transport. Furthermore, the co-expression network analysis of *ZmCNGC* genes may establish the importance of better understanding *ZmCNGC* transduction pathways in maize. Additionally, expression profiles of *ZmCNGC* genes are shown to express in a tissue-specific pattern. Our results provide valuable information to increase our understanding of the *CNGC* gene family in maize.

Genome-wide Identification and Analysis of the *CNGC* Gene Family 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

Abstract

As one of the non-selective cation channel gene family, the cyclic nucleotide-gated channel (CNGC) gene plays vital roles in plant physiological processes which are related to signal pathways, plant development, and environment stresses. However, genome-wide identification and analysis of the *CNGC* gene family in maize has not yet been conducted. In this study, 12 *ZmCNGC* genes are detected in the maize genome, which are unevenly distributed on chromosomes 1, 2, 4, 5, 6, 7 and 8. They are classified into four major groups, including group I, II, III, and IV. Prediction of cis-acting regulatory elements show that 137 putative cis-elements which are related to hormones-response, abiotic stress and organ development. Synteny analysis showed that 2, 2 and 1 ZmCNGCs had homologous genes in *Sorghum*, rice and *Brachypodium*, respectively. Gene ontology (GO) analysis demonstrated that most of the ZmCNGCs are involved in various biological processes including cellular processes, establishment of localization, and transmembrane transport. Furthermore, the co-expression network analysis of *ZmCNGC* genes may establish the importance of better understanding *ZmCNGC* transduction pathways in maize. Additionally, expression profiles of *ZmCNGC* genes are shown to express in a tissue-specific pattern. Our results provide valuable information to increase our understanding of the *CNGC* gene family in maize.

INTRODUCTION

In the process of organism evolution, the formation of complex nutrient absorption and transport system includes ion channels, ion pumps and carriers, and it has been previously shown that those 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, which is suggested to have one of the fundamental mechanisms for organismal systems (Nawaz et al. 2014; Yuen & Christopher 2013). As a molecular switch, secondary messengers such as cyclic nucleotide monophosphates (cNMPs; 3',5'-cAMP and 3',5'-cGMP) and Ca^{2+}/calmodulin (CaM) can regulate CNGCs, those messengers are activated by directing binding of cyclic nucleotides as well as are inhibited by binding of CaM to the CaM binding domain (Borsics et al. 2007; Defalco et al. 2016; Kaplan et al. 2007; Saand et al. 2015b).

As CNGCs are key components for plant development, many previous research studies have been conducted (Harada et al. 2003; Hua et al. 2003; Wang et al. 2013). With the application of bioinformatics tools, identification of *CNGC* gene family in *Arabidopsis*, rice and other plants were carried out. So far, *CNGC* genes have been identified in many plants, with 20 for *Arabidopsis*, 16 for rice, 18 for tomato and 26 in *B. oleracea*. (Bridges et al. 2005; Chen et al. 2015; Guo et al. 2017; Kakar et al. 2017; Nawaz et al. 2014; Saand et al. 2015a; Ward et al. 2009; Zelman et al. 2013; Zelman et al. 2012). Most *CNGC*s have been characterized by genetic methods and found to be related to plant physiological and molecular functions, including playing vital roles in multiple physiological processes which are involved in signal pathways, plant development, and environmental stresses. For example, *AtCNGC18* is expressed primarily in *Arabidopsis* pollen (Frietsch et al. 2007); *phot1* and *phot2* increases cytosolic Ca^{2+} in *Arabidopsis* leaves (Harada et al. 2003); *AtCNGC2* play key roles in stress signaling pathways, including changes the cytosolic free Ca^{2+} in *Arabidopsis* and *CNGC4* is permeable to both K+ and Na+ and activated by both cGMP and cAMP (Balague 2003; Reddy et al. 2011; Tracy et al. 2008). Meanwhile, the structure of *Arabidopsis* CNGCs have six transmembrane domains with a pore domain, also have cyclic nucleotide-binding domain and CaM-binding domains in the C-terminus, and these domains have

diverse functions (Chin et al. 2009; Hua et al. 2003; Köhler & Neuhaus 2000; Talke et al. 2003). For example, *Arabidopsis CNGC6*, *CNGC7*, *CNGC8*, *CNGC9*, *CNGC10* and *CNGC16* participate in the pollen development (Gao et al. 2012; KW et al. 2006; Tunc-Ozdemir et al. 2013; Wang et al. 2013); *CNGC2*, *CNGC4*, *CNGC11* and *CNGC12* are activated in response to pathogenic microorganisms (Dodd et al. 2010); and *CNGC6*, *CNGC10*, *CNGC19* and *CNGC20* are involved in abiotic stress (Kugler et al. 2009; Mosher et al. 2010).

In recent years, efforts had been made in studying the CNGC gene family in plants, but as one of the most important food crops and source of industrial materials in the world, the maize *CNGC* gene family was rarely reported. In this study, with the benefit from genome-wide sequence information in maize and research information on *Arabidopsis* and rice CNGC families, we conducted genome-wide identification of CNGCs in maize through comprehensive bioinformatics analyses. Furthermore, comprehensive analyses were conducted including multiple alignments, gene structure, conserved motifs and gene duplication of ZmCNGCs, and prediction of cis-acting regulatory elements, GO analysis, expression profiles of *ZmCNGC* genes and a co-expression network between ZmCNGC and other maize genes. This is the first systematic study of CNGC genes in maize and will provide the basis for further 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 obtained from the TAIR10 database (http://www.arabidopsis.org/index.jsp) and the RGAP database (http://rice.plantbiology.msu.edu/), respectively. After that, two method were used to search against the maize protein sequences, one is build a Hidden Markov Model (HMM) to search against maize protein sequences, another one is use local BLASTP method with a threshold of e-value < 1e-5. After that, the putative non-redundant protein sequence of maize *CNGC* genes were retrieved. To further confirm the ZmCNGC proteins whether have the CNBD domain, those putative ZmCNGC protein sequences were summited to SMART (http://smart.embl-

heidelberg.de/) (Letunic & Bork 2018) and NCBI-CDD (https://www.ncbi.nlm.nih.gov/cdd/) (Marchler-Bauer et al. 2017), genes without the CNBD domains or the amino acids size of below < 200 were removed and the *ZmCNGC* genes were confirmed.

The PI (theoretical isoelectric point), MW (molecular weight), and GRAVY of ZmCNGCs were predicted by ExPASy (http://web.expasy.org/protparam/) (Artimo et al. 2012). And the prediction subcellular location of ZmCNGCs were using the CELLO v.2.5 (http://cello.life.nctu.edu.tw/). The information of chromosome distribution of ZmCNGCs and the sequences including DNA sequences, CDS, cDNA, up-stream 1500bps of *ZmCNGC* genes were obtained from result of BLASTN search in the Ensembl Plant database (http://plants.ensembl.org/index.html) (Bolser et al. 2016).

Multiple alignments, phylogenetic analysis and gene duplication analysis

Multiple sequences alignments were performed using T-COFFEE web tool (http://tcoffee.crg.cat/apps/tcoffee/index.html) (Di Tommaso et al. 2011) and visualized by ESPript with default program (Robert & Gouet 2014). An un-root phylogenetic tree was **c**onstructed with 1000 bootstrap replication using MEGA 7 (Kumar et al. 2016) based on the full-length protein sequences alignment. 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 (Librado & Rozas 2009), and the divergence times (Mya) were calculated as a formula $Mya= Ks/2\lambda \times 10^{-6}$, therein $\lambda=6.5\times10^{-9}$ (Lynch & Conery 2000).

Analyses of genes structure and conserved motifs

The gene structure (exon-intron) of *ZmCNGC* genes were performed by the Gene Structure Display Server (GSDS, http://gsds.cbi.pku.edu.cn/) (Hu et al. 2015) using CDS and genome sequences of *ZmCNGC* genes. The conserved motif of ZmCNGC proteins were using the MEME

Suite web server (http://meme-suite.org/index.html) (Bailey et al. 2015) with the maximum number of motif sets at 9 and optimum width of motifs from 6 to 200 amino acid.

Cis-acting regulatory elements and GO annotation of ZmCNGCs

The up-stream 1500bps of *ZmCNGC* DNA sequences were used to find cis-acting regulatory elements by 'Signal Scan Search' programs in the NEW PLACE database (https://sogo.dna.affrc.go.jp/cgi-bin/sogo.cgi?lang=en) (Higo et al. 1999). The gene ontology (GO) annotation of ZmCNGC proteins were submitted to Monocots PLAZA 4.0 (Van Bel et al. 2017) to predict the function of ZmCNGC proteins, then the annotation was visualized and plotted by BGI WEGO (Ye et al. 2006).

Expression profiles of *ZmCNGC* genes and network interaction analysis

For understanding 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 number E-MTAB-3826 and E-MTAB-439. These data were used to analyze the expression of *ZmCNGC* in six different tissues (i.e. ear, embryo, endosperm, pollen, root and tassel) and different development stages in embryo, endosperm and seed, respectively. The FPKM values were used to calculate for each *ZmCNGC* genes. The interaction network was constructed on the base on the orthologs between maize and *Arabidopsis* using the AraNet v2 (Lee & Lee 2017) and visualized by the Cytoscape v3.4.0 (Shannon et al. 2003).

RESULTS AND DISSCUSSIONS

Genome-wide identification of *CNGC* genes in maize

To identify a complete overview of CNGC genes in maize, we firstly used 20 *Arabidopsis* and 16 rice CNGC protein sequences align blast against maize protein sequences. After BLAST alignment, a total of 18 putative *ZmCNGC* genes were identified in the maize genome. Then, to

confirm those 18 putative *ZmCNGC* genes, we used the SMART and NCBI CDD to find whether they contained the CNGC-specific domains (CNBD and transmembrane). After removing redundancy genes, a total of 12 *ZmCNGC* genes were detected, less than rice and *Arabidopsis* CNGC genes, the reason for it may be the gene duplication occur in rice and *Arabidopsis* (Paterson et al. 2004; Yu et al. 2005). As shown in Table 1, five of them were located in chromosome 5, others were unevenly located in chromosomes 1, 2, 4, 6, 7, and 8. The physiological and biochemical properties of these 12 *ZmCNGC* genes are listed in Table 1. The protein lengths ranged from 326 to 745 aa with average of 612 aa. The molecular weight of these proteins ranged from 38.63 kDa (GRMZM2G129375) to 85.52 kDa (GRMZM2G135651) and the pI value ranged from 8.92 (GRMZM2G023037) to 9.75 (GRMZM2G090528). Subcellular localization analysis indicated that all of ZmCNGCs localized in the plasma membrane except for GRMZM2G066269 which was localized in the nuclear fraction, this result is consistent with *Arabidopsis*, for example, previous studies have revealed that CNGCs are majorly localized in the plasma membrane (Lemtiri-Chlieh & Berkowitz 2004). In addition, some are distributed in vacuole membrane and nuclear envelope (Borsics et al. 2007; Christopher et al. 2007; Yuen & Christopher 2013).

To further access the existence of ZmCNGCs we identified, all the expressed sequence tags (EST) which aligned to *ZmCNGC* genes using the BLASTN program form NCBI; results found only GRMZM2G066269 showed no EST hits, other ZmCNGCs had more than 13 representative matches to ESTs.

Phylogenetic, classification and duplication analyze of CNGC genes

To further understanding the evolutionary relationship of *CNGCs*, an unrooted Neighbor-Joining (NJ) phylogenetic tree was generated based on the full-length protein alignments of *ZmCNGCs*, *AtCNGCs* and *OsCNGCs* (Maser et al. 2001; Nawaz et al. 2014). Thus, 20 from *Arabidopsis*, 20 from rice, and 12 from maize were used for constructing un-rooted phylogenetic tree. As shown in Figure 1, the phylogenetic tree clustered the *CNGCs* into four major groups, including group I, II, III, and IV. Of those four groups, group I, II and III are monophyletic, group IV was divided

into subgroups IVa and IVb. Among them, group I contains three maize *CNGC* genes (GRMZM2G066269, GRMZM2G148118, and GRMZM2G129375), six in *Arabidopsis* and two in rice. Group II contained two maize CNGC genes (GRMZM2G077828, GRMZM2G023037), five in *Arabidopsis* and three in rice. Similarly, group III contains three maize *CNGC* genes (GRMZM2G090528, GRMZM2G074317, and GRMZM2G858887). Group IV embraced seven in rice and seven in *Arabidopsis*, forming the largest CNGC group with four members in maize which included three in IVa (GRMZM2G068904, GRMZM2G005791 and GRMZM2G135651) and one in IVb (GRMZM2G141642). Based on the phylogenetic tree among *ZmCNGC*s, *AtCNGC*s and *OsCNGC*s, we also grouped maize *CNGC* genes into four sub-groups (Figure 2A). In the process of gene evolution, gene duplication play a significant role in generating new members and creating novel gene function**s** (Magadum et al. 2013). In this study, results found that two segmental duplications gene pairs were formed in maize genome (Table 2), including GRMZM2G005791-GRMZM2G436583 and GRMZM5G858887-GRMZM2G074317. To further investigated the evolutionary process between maize CNGCs and other gramineae plants, the genome synteny among *Sorghum*, rice and *Brachypodium* were also explored. Results showed that 2, 2 and 1 ZmCNGCs had homologous genes in *Sorghum*, rice and *Brachypodium*, respectively (Table 2). The substitution rate (Ka/Ks) is use 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 & Conery 2000). As result shown in Table 2, the Ka/Ks of each gene pair was calculated and Ka/Ks of all gene pairs were less than 1, suggested that the selection pressure after duplication was strongly purifying selection. Moreover, the divergence time were also calculated, results indicated that the divergence time between maize CNGCs and other gramineae plants were unevenly divergence, it may be the reason why maize CNGCs were a small gene family. Previous research showed that the gene duplication is a significant origin to generate novel genes (Davidson et al. 2013), thus, these results implied that the duplication events gave principal role in gene evolution.

Multiple alignments, gene structure and conserved motif of ZmCNGCs

The CNBD domain is a gene structural feature element in plant CNGCs which contain the PBC and the hinge region (Diller et al. 2001). As shown in Figure 3, after aligning the CNBD region of maize CNGCs, the putative PBC (phosphate binding cassette, from site 217 to 232 in Figure 3) and hinge domain (from site 260 to 266 in Figure 3) were also identified, which were consistent with rice CNGCs (Nawaz et al. 2014). Results showed that glycine (G), acidic residue glutamate (E), leucine (L) and aromatic tryptophan (W) inside the PBCs were 100% conserved, consistent with rice CNGCs (Nawaz et al. 2014).

As well the aliphatic alanine (A), aromatic phenylalanine (F) and leucine (L) were the most conserved within the hinge region. Compared to rice and *Arabidopsis* (Supplemental File 1), we found that glycine (G) and leucine (L) were found to be conserved 100% in CNGCs PBC domains, while aromatic phenylalanine (F) and leucine (L) were 100% conserved in the hinge domain. The diversity conserved motifs in maize, rice and *Arabidopsis* suggested that the function conserved among them.

Furthermore, gene structure analysis could add better understanding to the gene function and evolution. As a whole, the number of introns ranged from 1 to 7 while GRMZM2G077828 was intronless (Figure 3B), different from rice and Arabidopsis CNGCs. Previously showed that rice CNGC ranged from 1 to 11 introns, while *Arabidopsis* CNGCs ranged from 4 to 10 introns (Nawaz et al. 2014). Motif-based recognition of proteins give understanding the evolution history (Seoighe & Gehring 2004). Eight putative motifs were characterized and named as motif 1 to motif 8 in ZmCNGCs. The relative positions of motif in the four groups were found to have various patterns (Figure 3C). The conserved domain of most ZmCNGCs harbor motif 1, representing it is the typical ZmCNGC domain also as sequence logo of CNBD domain (Supplemental file 2). Fischer et al. (2013) have showed that IQ (motif 2 in this study) as a functional motif within CaMBD and downstream of the CNBD domain, also conserved among plant CNGCs. Study has showed that IQ motif enhances the changeable of Ca^{2+}-dependent channel control mechanisms in plant. In this study, 66.7% CNGC proteins (including group I, II and IVa) contained IQ (motif 2), suggesting

that those proteins bind CaM in a Ca^{2+}-dependent manner. In addition, 9 ZmCNGCs except GRMZM2G129375, GRMZM2G066269, and GRMZM2G090528 possess motif 4 which associated with associated with ion transport (Nawaz et al. 2014). Other motifs have not been identified in other plants or animals, suggested that these motif is maize specific.

Prediction of cis-acting regulatory elements and GO analysis of ZmCNGC proteins

Cis-acting regulatory elements are important molecular witches which associated with the transcriptional regulatory of genes when encounter environment stresses (Nakashima et al. 2009). To better understand the possible biological processes of these ZmCNGCs involved, 1.5 kb upstream of *ZmCNGC* genes genomic sequences were used to identify cis-regulatory elements and these were submitted to the NEW PLACE web tool. 137 different putative cis-elements were found to be associated with identified *ZmCNGC* genes and only 12, including CACTFTPPCA1, EBOXBNNAPA, DOFCOREZM, MYCCONSENSUSAT, CAATBOX1, GTGANTG10, WRKY71OS, GT1CONSENSUS, ROOTMOTIFTAPOX1, POLLEN1LELAT52, MYBCORE, and OSE2ROOTNODULE, were apparently in the promoter region of all *ZmCNGC* genes (Supplemental File 3: Table S1) and highly consistent with rice CNGCs, maybe these elements in the upstream region were conserved (Nawaz et al. 2014). Additionally, five cis-elements were gene-specific, such as ACGTCBOX, TATABOX3, HDZIP2ATATHB2, CTRMCAMV35S, -300CORE, ABREMOTIFAOSOSEM and MRNA3ENDTAH3 were unique to GRMZM2G005791, GRMZM2G068904, GRMZM2G074317, GRMZM2G135651, GRMZM2G135651, GRMZM2G005791 and GRMZM2G135651, respectively. Also, some cis-elements were involved in different abiotic/biotic stimuli, including those such as hormone-response (i.e. abscisic acid, auxin, ethylene, etc.), stress-related (i.e. drought, temperatures, disease, etc.) and development-related (i.e. mesophyll specific, tissue specific, etc.), indicating that these *ZmCNGC* genes may be involved in regulating diverse stress responses. For example, study showed that the CACTFTPPCA1 motif is for mesophyll-specific gene expression in the C4 plant (Gowik et al. 2004); MYCCONSENSUSAT is MYC recognition site which related to abiotic

stress signaling (Liu et al. 2015); and WRKY71OS is reported as binding site of WRKY71 which involved in gibberellin signaling pathway (Zhang et al. 2004). Different cis-elements presenting in *ZmCNGC* genes indicated that they may relate to different regulatory networks.

Furthermore, gene ontology (GO) terms were used to predict the functions of ZmCNGCs by classifying them into categories with three independent ontologies including those for biological process (BP), molecular function (MF), and cellular components (CC) (Consortium 2017). As shown in Figure 4, the biological process of ZmCNGCs were involved in cellular processes for establishment of localization and transmembrane transport. The molecular function ZmCNGCs participated in substrate-specific and transmembrane transport. Further, cellular component analysis revealed the localization of ZmCNGCs in the cell and membrane, may be the reason why the subcellular localization of most ZmCNGCs localized in the plasma membrane.

Co-expression network between ZmCNGC and other maize genes

To get the detailed information about the interaction relationship between *ZmCNGC* genes and other maize genes, the co-expression network based on the orthology-based predictions following the network in *Arabidopsis* were constructed. As shown in Figure 5, a total of 5 *ZmCNGC* genes including GRMZM2G068904, GRMZM2G077828, GRMZM2G005791, GRMZM2G141642 and GRMZM2G858887 with 76 gene pairs of network interactions were identified. GO annotations of interacting genes were also performed (Supplemental File 3: Table S2). Some symbols such as SOS1, BPM2, SKOR and BPM4, which play an essential role in regulation of organ development and osmotic stress response were shown. The co-expression network analysis of *ZmCNGC* genes may provide comprehensive information for understanding ZmCNGC genes transduction pathways in maize.

Expression profiles of *ZmCNGC* genes in different tissues

We performed transcriptome sequencing to evaluate tissue-specific expression levels of *ZmCNGC* genes in different tissues based on previous study. As shown in Supplemental File 4, the expression

levels among *ZmCNGC* genes were tissue-specific in their expression. For example, *GRMZM2G077828* and *GRMZM2G148118* were highly expressed in pollen and all group III genes (*GRMZM5G858887*, *GRMZM2G074317* and *GRMZM2G090528*) were mainly expressed in embryo, while *GRMZM2G129375* and *GRMZM2G005791* were expressed at low levels in pollen and embryo, indicted that III sub-group genes contribute to maize embryo development. We also evaluated some *ZmCNGC* genes in embryo, endosperm and seed expression in some days after pollination. As shown in Supplemental File 4 B, C and D, the embryo specific-expression gene *GRMZM5G858887* is gradually went up with time, and *GRMZM2G090528* is highly expression in embryo, endosperm and seed. Interestingly, we found that all group III genes including *GRMZM5G858887*, *GRMZM2G074317* and *GRMZM2G090528* had participated in gene duplication. Previous research have demonstrated that duplication genes often obtain new functions to enhance plant adapt to the environment (Dias et al. 2003). Compared to other *ZmCNGC* genes, the gene expression of these three duplication genes were more intense, suggested that these three *ZmCNGC* genes obtained gene functions from other plants after duplication. Further illustrated that group III gene play important role in maize gene duplication, evolution and expression.

Most researches showed that cyclic nucleotide-gated channels (CNGCs) genes have been related to pollen development and in response to environmental stimulus. For example, *Arabidopsis CNGC16* is critical for pollen fertility under conditions of heat stress and drought stress (Tunc-Ozdemir et al. 2013), and *CNGC18* has been shown to function in pollen tube tip growth . In rice, *OsCNGC13* promotes seed-setting rate by facilitating pollen tube growth in stylar tissues (Xu et al. 2017). *GRMZM2G148118* and *GRMZM2G077828*, two homologous gene of *CNGC16* and *CNGC18* in our study are mainly expressed in pollen, indicated they mainly involved in pollen development. Previous study showed that the AtCNGC3 promoter::GUS construct in transgenic plants revealed expression throughout plant development mainly in the embryo, leaves and roots, the expression level of *GRMZM2G023037* is consistent with *AtCNGC3* which highly expressed in plant development except pollen (Kaplan et al. 2007).

CONCLUSION

We identified 12 *ZmCNGC* genes from the maize genome, and classed them into four major groups. We comprehensively analyzed the gene structure and conserved motif in ZmCNGCs. Aligning the maize CNGCs and other plants showed that PBC and hinge domain is the most conserved in CNBD domain. Also, a total of 137 putative cis-elements were identified and related to hormones response, abiotic stress and organ development. GO analysis indicated that most of them are involved in various biological processes, including cellular process, establishment of localization and transmembrane transport. Furthermore, the co-expression network analysis of *ZmCNGC* genes may provide important information for the better understanding ZmCNGC transduction pathways. Expression profiles of *ZmCNGC* genes were tissue-specific expressed and related to pollen development. In addition, gene duplication analysis indicated that *ZmCNGC* genes obtained gene functions from other plants after duplication. In summary, our results provide a solid foundation for further evolutionary and functional investigations on ZmCNGCs.

ACKNOWLEDGEMENTS

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

ADDITIONAL INFORMATION AND DECLARATIONS

Grant Disclosures

The following grant information was disclosed by the authors:

Natural Science Foundation of Heilongjiang Province, China: B2016010.

Supplemental Information

Supplemental File 1

Multiple sequences alignments among maize, *Arabidopsis* and rice.

Supplemental File 2

The logo of eight motifs in study. Eight motifs were identified in this study.

Supplemental File 3

Table S1 Numbers of known stress-related elements in the promoter regions of ZmCNGCs

Table S2 Detail information of network of *ZmCNGC* genes with other maize genes

Supplemental File 4

Expression profiles of *ZmCNGC* genes. (A) the expression of *ZmCNGC* in 6 different tissues which including ear, embryo, endosperm, pollen, root and tassel, and different stage in (B) embryo, (C)endosperm and (D) seed, respectively.

Supplemental File 5

The gene sequences were used in this research.

Data Availability

The following information was supplied regarding data availability:

The raw data has been supplied as a Supplemental File 5.

REFERENCES

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, and Stockinger H. 2012. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res* 40:W597-603. 10.1093/nar/gks400

Bailey TL, Johnson J, Grant CE, and Noble WS. 2015. The MEME Suite. *Nucleic Acids Res* 43:W39-49. 10.1093/nar/gkv416

Balague C. 2003. HLM1, an Essential Signaling Component in the Hypersensitive Response, Is a Member of the Cyclic Nucleotide-Gated Channel Ion Channel Family. *The Plant Cell Online* 15:365-379. 10.1105/tpc.006999

Bolser D, Staines DM, Pritchard E, and Kersey P. 2016. Ensembl Plants: Integrating Tools for Visualizing, Mining, and Analyzing Plant Genomics Data. *Methods Mol Biol* 1374:115-140. 10.1007/978-1-4939-3167-5_6

Borsics T, Webb D, Andemeondzighi C, Staehelin LA, and Christopher DA. 2007. The cyclic nucleotide-gated calmodulin-binding channel AtCNGC10 localizes to the plasma membrane and influences numerous growth responses and starch accumulation in Arabidopsis thaliana. *Planta* 225:563-573. 10.1007/s00425-006-0372-3

Bridges D, Fraser ME, and Moorhead GB. 2005. Cyclic nucleotide binding proteins in the Arabidopsis thaliana and Oryza sativa genomes. *Bmc Bioinformatics* 6:1-12. 10.1186/1471-2105-6-6

Chen J, Yin H, Gu J, Li L, Liu Z, Jiang X, Zhou H, Wei S, Zhang S, and Wu J. 2015. Genomic characterization, phylogenetic comparison and differential expression of the cyclic nucleotide-gated channels gene family in pear (Pyrus bretchneideri Rehd.). *Genomics* 105:39. 10.1016/j.ygeno.2014.11.006

Chin K, Moeder W, and Yoshioka K. 2009. Biological roles of cyclic-nucleotide-gated ion channels in plants: what we know and don't know about this 20 member ion channel family. *Botany-botanique* 87:668-677. 10.1139/B08-147

Christopher DA, Borsics T, Yuen CY, Ullmer W, Andèmeondzighi C, Andres MA, Kang BH, and Staehelin LA. 2007. The cyclic nucleotide gated cation channel AtCNGC10 traffics from the ER via Golgi vesicles to the plasma membrane of Arabidopsis root and leaf cells. *Bmc Plant Biology* 7:48. 10.1186/1471-2229-7-48

Consortium TGO. 2017. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res* 45:D331-d338. 10.1093/nar/gkw1108

Davidson CJ, Guthrie EE, and Lipsick JS. 2013. Duplication and maintenance of the Myb genes of vertebrate animals. *Biology Open* 2:101-110.

Defalco TA, Marshall CB, Munro K, Kang HG, Moeder W, Ikura M, Snedden WA, and Yoshioka K. 2016. Multiple Calmodulin-Binding Sites Positively and Negatively Regulate Arabidopsis CYCLIC NUCLEOTIDE-GATED CHANNEL12. *Plant Cell* 28:1738. 10.1105/tpc.15.00870

Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang JM, Taly JF, and Notredame C. 2011. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res* 39:W13-17. 10.1093/nar/gkr245

Dias AP, Braun EL, McMullen MD, and Grotewold E. 2003. Recently duplicated maize R2R3 Myb genes provide evidence for distinct mechanisms of evolutionary divergence after duplication. *Plant Physiol* 131:610-620. 10.1104/pp.012047

Diller TC, Madhusudan, Xuong NH, and Taylor SS. 2001. Molecular Basis for Regulatory Subunit Diversity in cAMP-Dependent Protein Kinase - Crystal Structure of the Type IIbeta Regulatory Subunit. *Structure* 9:73-82. 10.1016/S0969-2126(00)00556-6

Dodd AN, Kudla J, and Sanders D. 2010. The Language of Calcium Signaling. *Annual Review of Plant Biology* 61:593. 10.1146/annurev-arplant-070109-104628

Fischer C, Kugler A, Hoth S, and Dietrich P. 2013. An IQ Domain Mediates the Interaction with Calmodulin in a Plant Cyclic Nucleotide-Gated Channel. *Plant Cell Physiol* 54:573-584. 10.1093/pcp/pct021

Frietsch S, Wang YF, Sladek C, Poulsen LR, Romanowsky SM, Schroeder JI, and Harper JF. 2007. A cyclic nucleotide-gated channel is essential for polarized tip growth of pollen. *Proc Natl Acad Sci U S A* 104:14531-14536. 10.1073/pnas.0701781104

Gao F, Han X, Wu J, Zheng S, Shang Z, Sun D, Zhou R, and Li B. 2012. A heat-activated calcium-permeable channel--Arabidopsis cyclic nucleotide-gated ion channel 6--is involved in heat shock responses. *Plant Journal for Cell & Molecular Biology* 70:1056-1069. 10.1111/j.1365-313X.2012.04969.x

Gowik U, Burscheidt J, Akyildiz M, Schlue U, Koczor M, Streubel M, and Westhoff P. 2004. cis-Regulatory elements for mesophyll-specific gene expression in the C4 plant Flaveria trinervia, the promoter of the C4 phosphoenolpyruvate carboxylase gene. *Plant Cell* 16:1077-1090. 10.1105/tpc.019729

Guo S, Zuo Y, Zhang Y, Wu C, Su W, Jin W, Yu H, An Y, and Li Q. 2017. Large-scale transcriptome comparison of sunflower genes responsive to Verticillium dahliae. *Bmc Genomics* 18:42. 10.1186/s12864-016-3386-7

Harada A, Sakai T, and Okada K. 2003. Phot1 and phot2 mediate blue light-induced transient increases in cytosolic Ca2+ differently in Arabidopsis leaves. *Proceedings of the National Academy of Sciences of the United States of America* 100:8583-8588.

Higo K, Ugawa Y, Iwamoto M, and Korenaga T. 1999. Plant cis-acting regulatory DNA elements (PLACE) database:

1999. *Nucleic Acids Res* 27:297-300.

Hu B, Jin J, Guo AY, Zhang H, Luo J, and Gao G. 2015. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31:1296-1297. 10.1093/bioinformatics/btu817

Hua BG, Mercier RW, Zielinski RE, and Berkowitz GA. 2003. Functional interaction of calmodulin with a plant cyclic nucleotide gated cation channel. *Plant Physiology & Biochemistry* 41:945-954.

Köhler C, and Neuhaus G. 2000. Characterisation of calmodulin binding to cyclic nucleotide-gated ion channels from Arabidopsis thaliana. *FEBS Lett* 471:133-136. 10.1016/S0014-5793(00)01383-1

Kakar KU, Nawaz Z, Kakar K, Ali E, Almoneafy AA, Ullah R, Ren XL, and Shu QY. 2017. Comprehensive genomic analysis of theCNGCgene family inBrassica oleracea: novel insights into synteny, structures, and transcript profiles. *Bmc Genomics* 18:869. 10.1186/s12864-017-4244-y

Kaplan B, Sherman T, and Fromm H. 2007. Cyclic nucleotide-gated channels in plants. *FEBS Lett* 581:2237-2246. 10.1016/j.febslet.2007.02.017

Kugler A, Köhler B, Palme K, Wolff P, and Dietrich P. 2009. Salt-dependent regulation of a CNG channel subfamily in Arabidopsis. *Bmc Plant Biology* 9:140. 10.1186/1471-2229-9-140

Kumar S, Stecher G, and Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology & Evolution* 33:1870. 10.1093/molbev/msw054

KW B, D H, JM W, S P, EP N, KD H, D T, and H S. 2006. Integrating Membrane Transport with Male Gametophyte Development and Function through Transcriptomics. *Plant Physiol* 140:1151. 10.1104/pp.105.074708

Lee T, and Lee I. 2017. AraNet: A Network Biology Server for Arabidopsis thaliana and Other Non-Model Plant Species. *Methods in Molecular Biology* 1629:225. 10.1007/978-1-4939-7125-1_15

Lee TH, Tang H, Wang X, and Paterson AH. 2013. PGDD: a database of gene and genome duplication in plants. *Nucleic Acids Research* 41:1152-1158.

Lemtiri-Chlieh F, and Berkowitz GA. 2004. Cyclic adenosine monophosphate regulates calcium channels in the plasma membrane of Arabidopsis leaf guard and mesophyll cells. *J Biol Chem* 279:35306-35312. 10.1074/jbc.M400311200

Letunic I, and Bork P. 2018. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res* 46:D493-d496. 10.1093/nar/gkx922

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

Liu J, Wang F, Yu G, Zhang X, Jia C, Qin J, and Pan H. 2015. Functional Analysis of the Maize C-Repeat/DRE Motif-Binding Transcription Factor CBF3 Promoter in Response to Abiotic Stress. *Int J Mol Sci* 16:12131-12146. 10.3390/ijms160612131

Lynch M, and Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151.

Magadum S, Banerjee U, Murugan P, Gangapur D, and Ravikesavan R. 2013. Gene duplication as a major force in evolution. *J Genet* 92:155-161.

Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, and Bryant SH. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* 45:D200-d203. 10.1093/nar/gkw1129

Maser P, Thomine S, Schroeder JI, Ward JM, Hirschi K, Sze H, Talke IN, Amtmann A, Maathuis FJ, Sanders D, Harper JF, Tchieu J, Gribskov M, Persans MW, Salt DE, Kim SA, and Guerinot ML. 2001. Phylogenetic

relationships within cation transporter families of Arabidopsis. *Plant Physiol* 126:1646-1667.

Mayrose I, Doron-Faigenboim A, Bacharach E, and Pupko T. 2007. Towards realistic codon models: among site variability and dependency of synonymous and non-synonymous rates. *Bioinformatics* 23:i319-327. 10.1093/bioinformatics/btm176

Mosher S, Moeder W, Nishimura N, Jikumaru Y, Joo SH, Urquhart W, Klessig DF, Kim SK, Nambara E, and Yoshioka K. 2010. The Lesion-Mimic Mutant cpr22 Shows Alterations in Abscisic Acid Signaling and Abscisic Acid Insensitivity in a Salicylic Acid-Dependent Manner. *Plant Physiol* 152:1901. 10.1104/pp.109.152603

Nakashima K, Ito Y, and Yamaguchi-Shinozaki K. 2009. Transcriptional regulatory networks in response to abiotic stresses in Arabidopsis and grasses. *Plant Physiol* 149:88-95. 10.1104/pp.108.129791

Nawaz Z, Kakar KU, Saand MA, and Shu QY. 2014. Cyclic nucleotide-gated ion channel gene family in rice, identification, characterization and experimental analysis of expression response to plant hormones, biotic and abiotic stresses. *BMC Genomics* 15:1-18.

Paterson AH, Bowers JE, and Chapman BA. 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc Natl Acad Sci U S A* 101:9903-9908. 10.1073/pnas.0307901101

Reddy AS, Ali GS, Celesnik H, and Day IS. 2011. Coping with stresses: roles of calcium- and calcium/calmodulin-regulated gene expression. *Plant Cell* 23:2010-2032. 10.1105/tpc.111.084988

Robert X, and Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* 42:W320. 10.1093/nar/gku316

Saand MA, Xu YP, Li W, Wang JP, and Cai XZ. 2015a. Cyclic nucleotide gated channel gene family in tomato: genome-wide identification and functional analyses in disease resistance. *Frontiers in Plant Science* 6:303. 10.3389/fpls.2015.00303

Saand MA, Xu YP, Munyampundu JP, Li W, Zhang XR, and Cai XZ. 2015b. Phylogeny and evolution of plant cyclic nucleotide-gated ion channel (CNGC) gene family and functional analyses of tomato CNGCs. *DNA research : an international journal for rapid publication of reports on genes and genomes* 22:471-483. 10.1093/dnares/dsv029

Seoighe C, and Gehring C. 2004. Genome duplication led to highly selective expansion of the Arabidopsis thaliana proteome. *Trends Genet* 20:461-464. 10.1016/j.tig.2004.07.008

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and Ideker T. 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* 13:2498-2504. 10.1101/gr.1239303

Talke, Ina N, Blaudez, Damien, Maathuis, Frans JM, Sanders, and Dale. 2003. CNGCs: prime targets of plant cyclic nucleotide signalling? *Trends in Plant Science* 8:286. 10.1016/S1360-1385(03)00099-2

Tracy FE, Gilliham M, Dodd AN, Webb AAR, and Tester M. 2008. NaCl‐induced changes in cytosolic free Ca2+ in Arabidopsis thaliana are heterogeneous and modified by external ionic composition. *Plant Cell & Environment* 31:1063. 10.1111/j.1365-3040.2008.01817.x

Tunc-Ozdemir M, Tang C, Ishka MR, Brown E, Groves NR, Myers CT, Rato C, Poulsen LR, Mcdowell S, and Miller G. 2013. A cyclic nucleotide-gated channel (CNGC16) in pollen is critical for stress tolerance in pollen reproductive development. *Plant Physiol* 161:1010-1020. 10.1104/pp.112.206888

Van Bel M, Diels T, Vancaester E, Kreft L, Botzki A, Van de Peer Y, Coppens F, and Vandepoele K. 2017. PLAZA

4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res*. 10.1093/nar/gkx1002

Wang YF, Munemasa S, Nishimura N, Ren HM, Robert N, Han M, Puzõrjova I, Kollist H, Lee S, and Mori I. 2013. Identification of cyclic GMP-activated nonselective Ca2+-permeable cation channels and associated CNGC5 and CNGC6 genes in Arabidopsis guard cells. *Plant Physiol* 163:578.

Ward JM, Mäser P, and Schroeder JI. 2009. Plant Ion Channels: Gene Families, Physiology, and Functional Genomics Analyses. *Annual Review of Physiology* 71:59. 10.1146/annurev.physiol.010908.163204

Xu Y, Yang J, Wang Y, Wang J, Yu Y, Long Y, Wang Y, Zhang H, Ren Y, Chen J, Wang Y, Zhang X, Guo X, Wu F, Zhu S, Lin Q, Jiang L, Wu C, Wang H, and Wan J. 2017. OsCNGC13 promotes seed-setting rate by facilitating pollen tube growth in stylar tissues. *PLoS Genet* 13:e1006906. 10.1371/journal.pgen.1006906

Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L, and Wang J. 2006. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res* 34:W293-297. 10.1093/nar/gkl031

Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W, Hu S, Zeng C, Zhang J, Zhang Y, Li R, Xu Z, Li S, Li X, Zheng H, Cong L, Lin L, Yin J, Geng J, Li G, Shi J, Liu J, Lv H, Li J, Wang J, Deng Y, Ran L, Shi X, Wang X, Wu Q, Li C, Ren X, Wang J, Wang X, Li D, Liu D, Zhang X, Ji Z, Zhao W, Sun Y, Zhang Z, Bao J, Han Y, Dong L, Ji J, Chen P, Wu S, Liu J, Xiao Y, Bu D, Tan J, Yang L, Ye C, Zhang J, Xu J, Zhou Y, Yu Y, Zhang B, Zhuang S, Wei H, Liu B, Lei M, Yu H, Li Y, Xu H, Wei S, He X, Fang L, Zhang Z, Zhang Y, Huang X, Su Z, Tong W, Li J, Tong Z, Li S, Ye J, Wang L, Fang L, Lei T, Chen C, Chen H, Xu Z, Li H, Huang H, Zhang F, Xu H, Li N, Zhao C, Li S, Dong L, Huang Y, Li L, Xi Y, Qi Q, Li W, Zhang B, Hu W, Zhang Y, Tian X, Jiao Y, Liang X, Jin J, Gao L, Zheng W, Hao B, Liu S, Wang W, Yuan L, Cao M, McDermott J, Samudrala R, Wang J, Wong GK, and Yang H. 2005. The Genomes of Oryza sativa: a history of duplications. *PLoS Biol* 3:e38. 10.1371/journal.pbio.0030038

Yuen CCY, and Christopher DA. 2013. The group IV-A cyclic nucleotide-gated channels, CNGC19 and CNGC20, localize to the vacuole membrane in Arabidopsis thaliana. *AoB Plants* 5:plt012-plt012. 10.1093/aobpla/plt012

Zelman AK, Dawe A, and Berkowitz GA. 2013. Identification of cyclic nucleotide gated channels using regular expressions. *Methods in Molecular Biology* 1016:207. 10.1007/978-1-62703-441-8_14

Zelman AK, Dawe A, Gehring C, and Berkowitz GA. 2012. Evolutionary and Structural Perspectives of Plant Cyclic Nucleotide-Gated Cation Channels. *Frontiers in Plant Science* 3:95. 10.3389/fpls.2012.00095

Zhang ZL, Xie Z, Zou X, Casaretto J, Ho TH, and Shen QJ. 2004. A rice WRKY gene encodes a transcriptional repressor of the gibberellin signaling pathway in aleurone cells. *Plant Physiol* 134:1500-1513. 10.1104/pp.103.034967

Table 1(on next page)

Characteristic features of ZmCNGC gene family identified in *Zea mays*

Table 1 Characteristic features of ZmCNGC gene family identified in *Zea mays*

Group	Gene name	Chr	start	end	Length (aa)	MW (Da)	pI	GRAVY	Localization	EST
I	GRMZM2G148118	4	23019490	23020538	701	79914.66	9.18	-0.046	PlasmaMembrane	17
	GRMZM2G129375	6	10982435	10982592	326	38405.22	9.17	-0.484	PlasmaMembrane	23
	GRMZM2G066269	4	23038271	23038459	329	38632.18	9.59	-0.52	Nuclear	0
II	GRMZM2G023037	2	5966501	5978989	723	82856.52	8.92	-0.114	PlasmaMembrane	29
	GRMZM2G077828	5	17453069	17455739	699	80110.81	9.42	-0.091	PlasmaMembrane	15
III	GRMZM2G005791	5	21883443	21884091	700	80057.39	8.97	-0.078	PlasmaMembrane	31
	GRMZM2G068904	5	19160904	19161178	689	80038.95	9.87	-0.117	PlasmaMembrane	15
	GRMZM2G135651	7	15065251	15065706	739	85523.26	9.33	-0.148	PlasmaMembrane	61
IVa	GRMZM2G141642	5	21711922	21712546	463	53303.64	9.48	-0.059	PlasmaMembrane	13
IVb	GRMZM5G858887	5	6938133	6943695	745	83672.62	9.46	0.081	PlasmaMembrane	84
	GRMZM2G074317	1	28340150	28340803	730	81440.82	9.43	0.067	PlasmaMembrane	74
	GRMZM2G090528	8	17724486	17724728	505	57042.47	9.75	0.014	PlasmaMembrane	48

Table 2(on next page)

The Ka/Ks ratios and estimated divergence time for orthologous CNGC proteins between maize and other Gramineae plants

Segmental duplication between maize genes as well as the synteny block between maize and other gramineae plants were obtained from the Plant Genome Duplication database. The substitution rates (Ka/Ks) of duplication events were calculated by using the DnaSP v5, and the divergence times (Mya) were calculated as a formula $Mya = Ks/2\lambda \times 10^{-6}$, therein $\lambda = 6.5 \times 10^{-9}$.

Table 2

The Ka/Ks ratios and estimated divergence time for orthologous CNGC proteins between maize and other Gramineae plants

Gene ID	Gene ID	Ka	Ks	Ka/Ks	Mya
GRMZM2G005791	GRMZM2G436583	0.1202	1.3677	0.087885	105.2077
GRMZM2G074317	GRMZM5G858887	0.0258	0.1829	0.141061	14.06923
GRMZM2G077828	Sobic.001G155100	0.5298	0.8917	0.594146	68.59231
GRMZM2G090528	Sobic.009G188800	0.095	0.4878	0.194752	37.52308
GRMZM2G090528	Sobic.003G317700	0.0306	0.0993	0.308157	7.638462
GRMZM2G077828	LOC_Os03g44440	0.7415	1.2253	0.605158	94.25385
GRMZM2G090528	LOC_Os05g42250	0.1187	0.4278	0.277466	32.90769
GRMZM2G090528	LOC_Os01g57370	0.073	0.3017	0.241962	23.20769
GRMZM2G077828	Bradi1g13740	0.7912	1.1721	0.675028	90.16154

Figure 1(on next page)

Phylogenetic analysis of CNGC proteins among *Zea mays*, *Arabidopsis* and rice

The CNGC genes of maize, Arabidopsis thaliana, rice and maize were clustered into four major groups, including Group I, II, III, and IV (a and b).

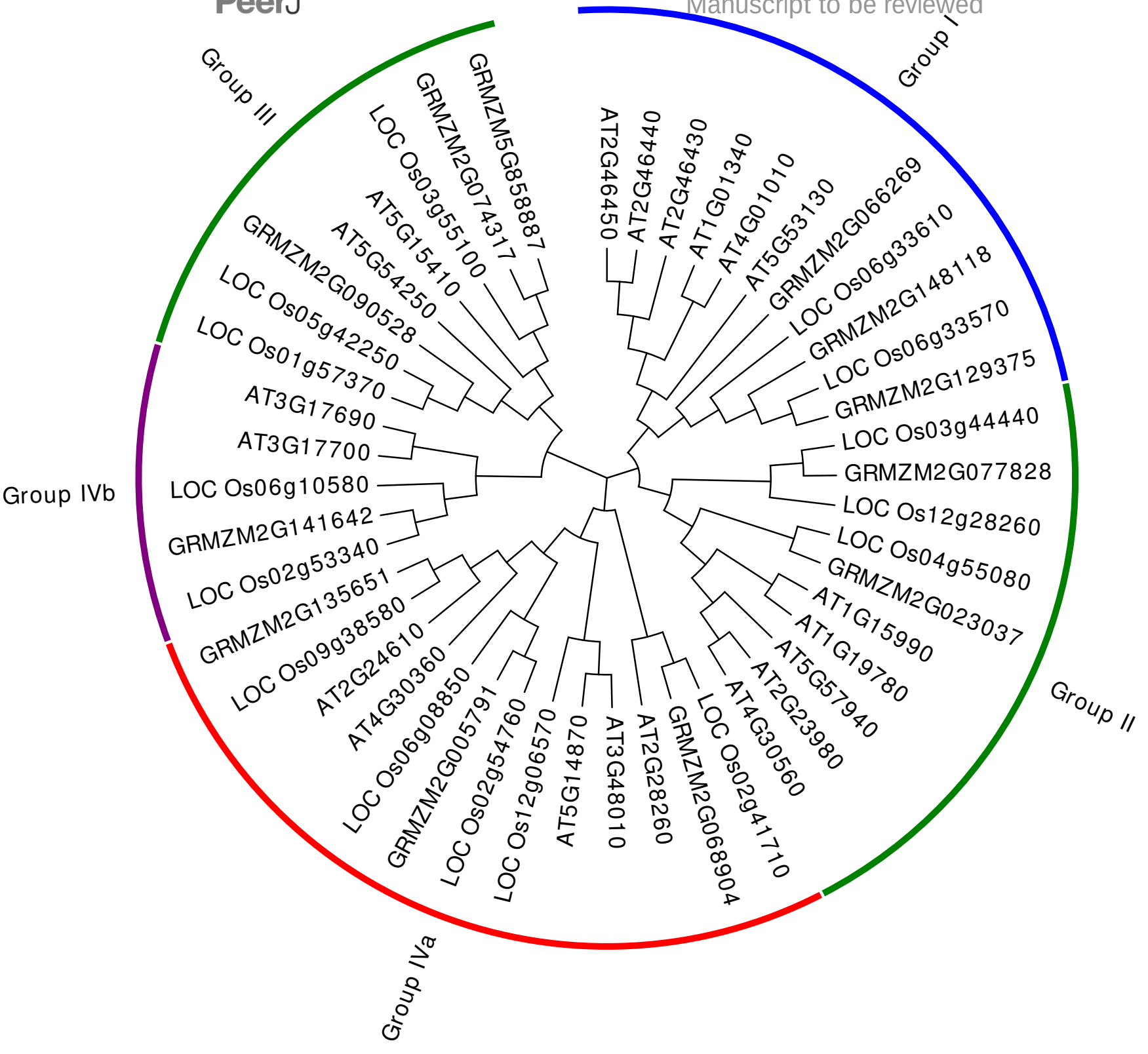

Group I
Group II
Group IVa
Group IVb
Group III
AT2G46450
AT2G46440
AT2G46430
AT1G01340
AT4G01010
AT5G53130
GRMZM2G066269
LOC Os06g33610
GRMZM2G148118
LOC Os06g33570
GRMZM2G129375
LOC Os03g44440
GRMZM2G077828
LOC Os12g28260
LOC Os04g55080
GRMZM2G023037
AT1G15990
AT1G19780
AT5G57940
AT2G23980
AT4G30560
LOC Os02g41710
GRMZM2G068904
AT2G28260
AT3G48010
AT5G14870
LOC Os12g06570
LOC Os02g54760
GRMZM2G005791
LOC Os06g08850
AT4G30360
AT2G24610
LOC Os09g38580
GRMZM2G135651
LOC Os02g53340
GRMZM2G141642
LOC Os06g10580
AT3G17700
AT3G17690
LOC Os01g57370
LOC Os05g42250
GRMZM2G090528
AT5G54250
AT5G15410
LOC Os03g55100
GRMZM2G074317
GRMZM5G858887

Figure 2(on next page)

The structure of ZmCNGCs in Zea mays

Phylogenetic relationships (A), gene structure (B) and motif compositions (C) of ZmCNGCs.

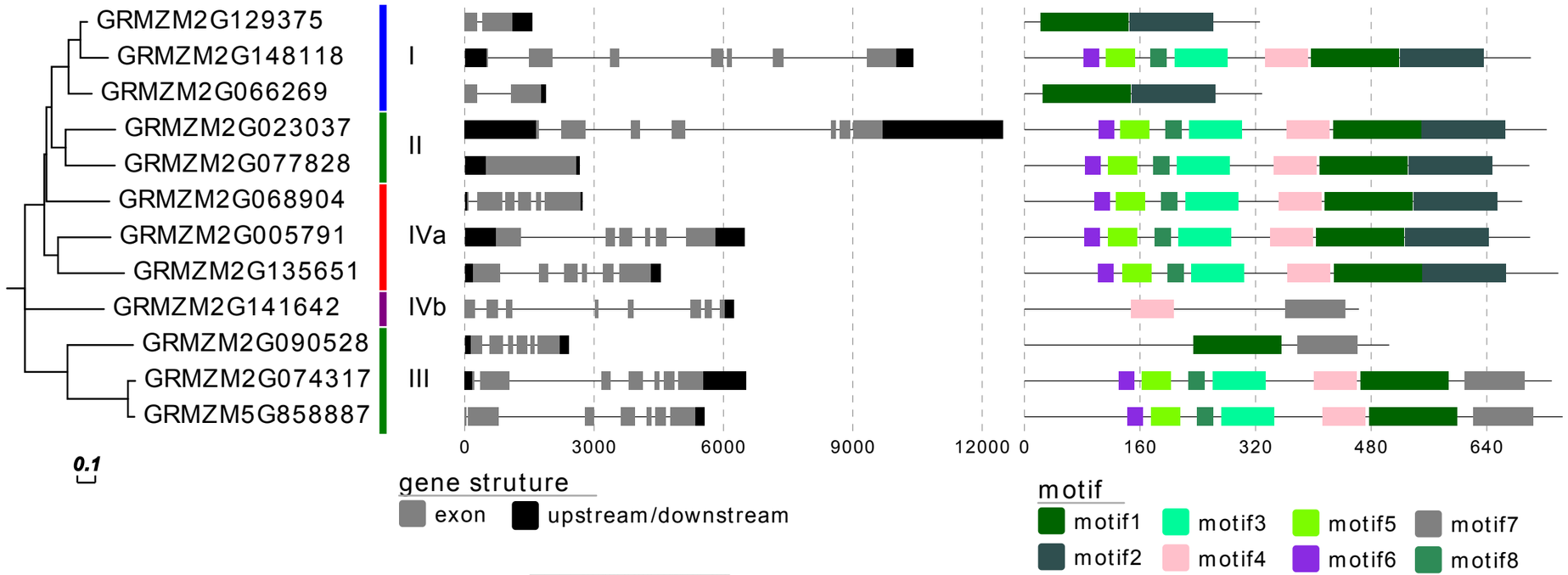

GRMZM2G129375
GRMZM2G148118
GRMZM2G066269
GRMZM2G023037
GRMZM2G077828
GRMZM2G068904
GRMZM2G005791
GRMZM2G135651
GRMZM2G141642
GRMZM2G090528
GRMZM2G074317
GRMZM5G858887
I
II
IVa
IVb
III
0.1
0
3000
6000
9000
12000
0
160
320
480
640
gene struture
exon
upstream/downstream
motif
motif1
motif2
motif3
motif4
motif5
motif6
motif7
motif8

Figure 3(on next page)

Multiple alignments of ZmCNGC proteins.

The putative PBC (phosphate binding cassette, from site 217 to 232 in Figure 3) and hinge domain (from site 260 to 266 in Figure 3) were identified in maize CNGC proteins.

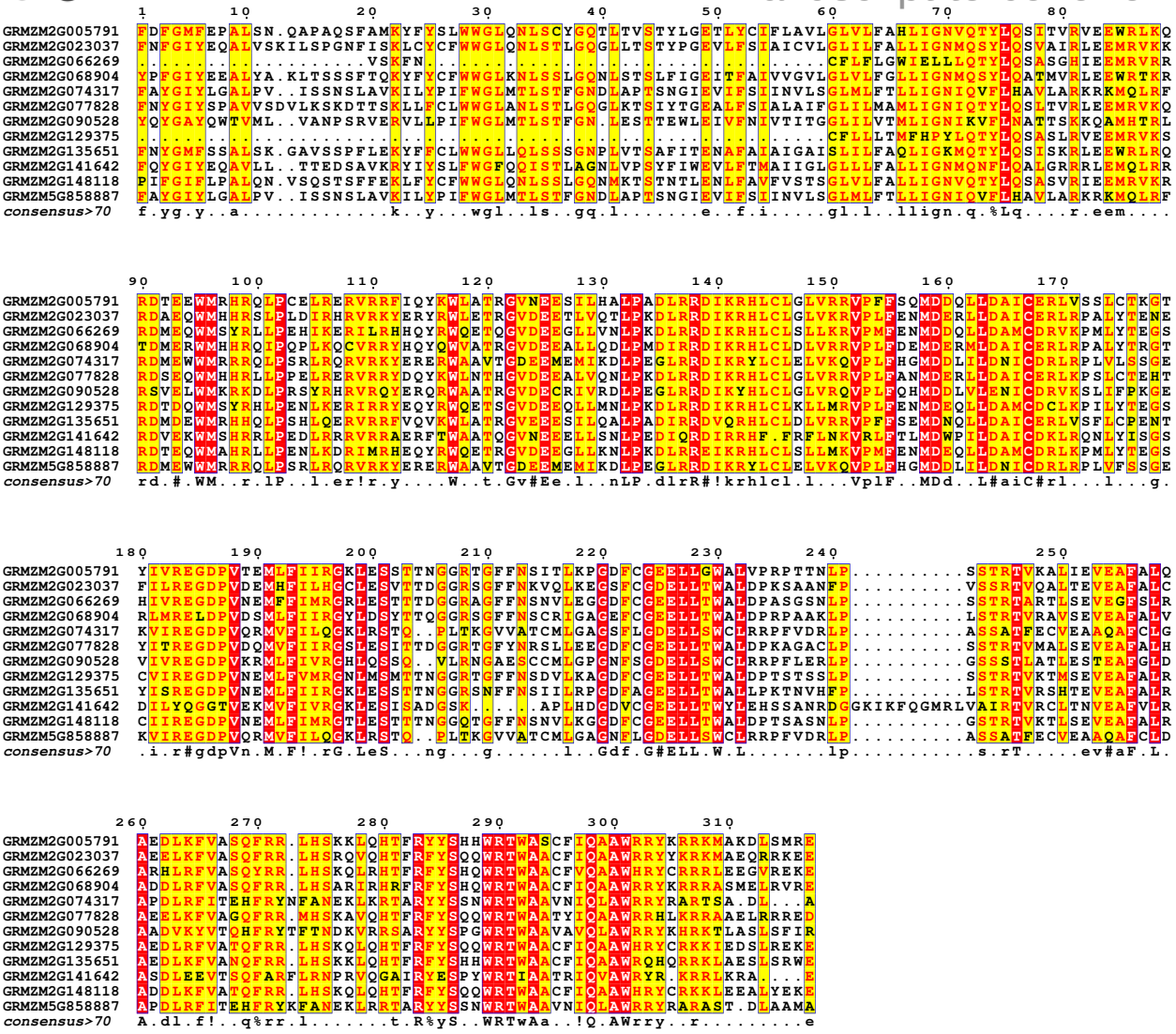

GRMZM2G005791
GRMZM2G023037
GRMZM2G066269
GRMZM2G068904
GRMZM2G074317
GRMZM2G077828
GRMZM2G090528
GRMZM2G129375
GRMZM2G135651
GRMZM2G141642
GRMZM2G148118
GRMZM5G858887
consensus>70

Figure 4(on next page)

Gene Ontology annotation of *ZmCNGC*s

The CC, MF and BP represent cellular components, molecular function and biological process in GO annotation, respectively.

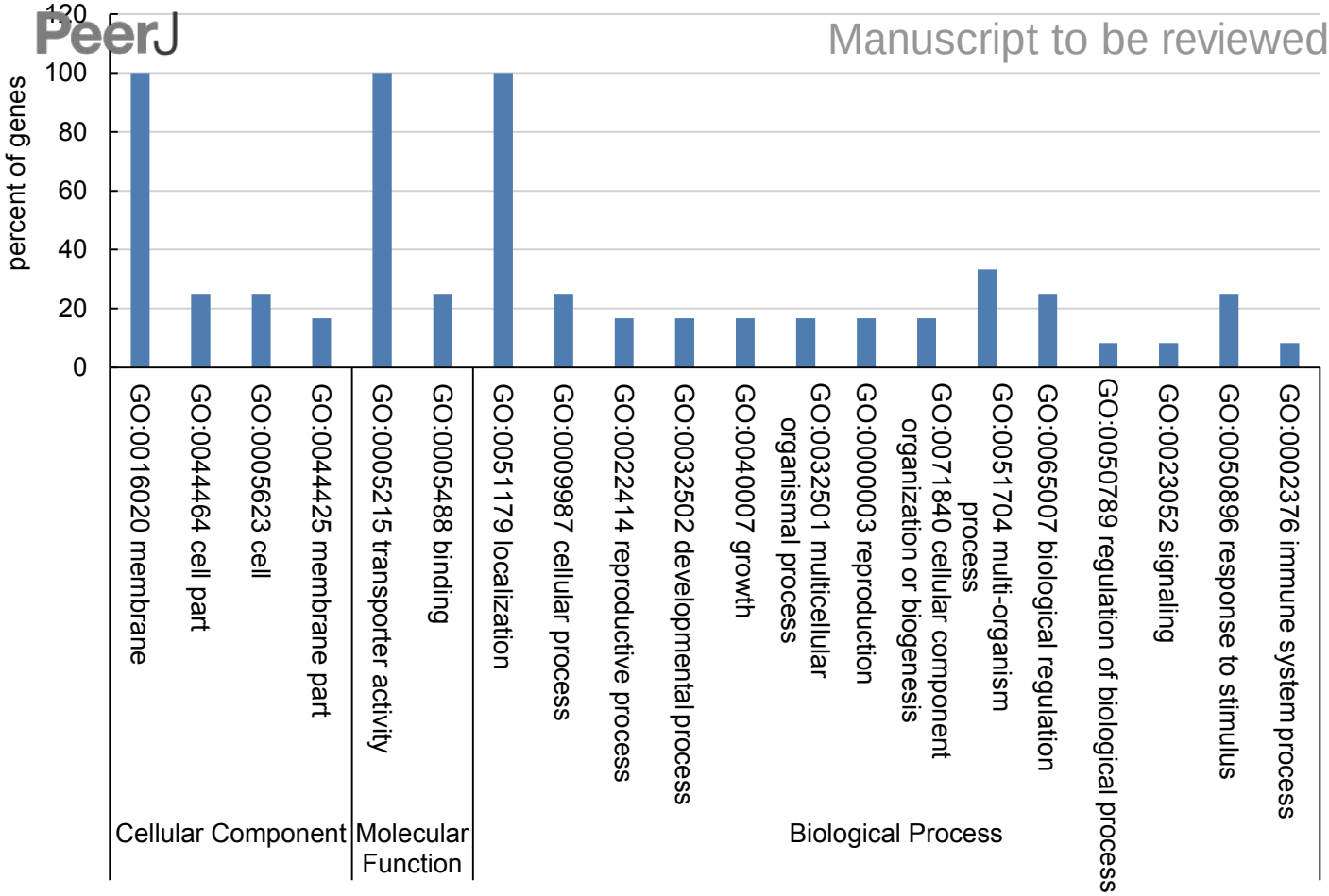
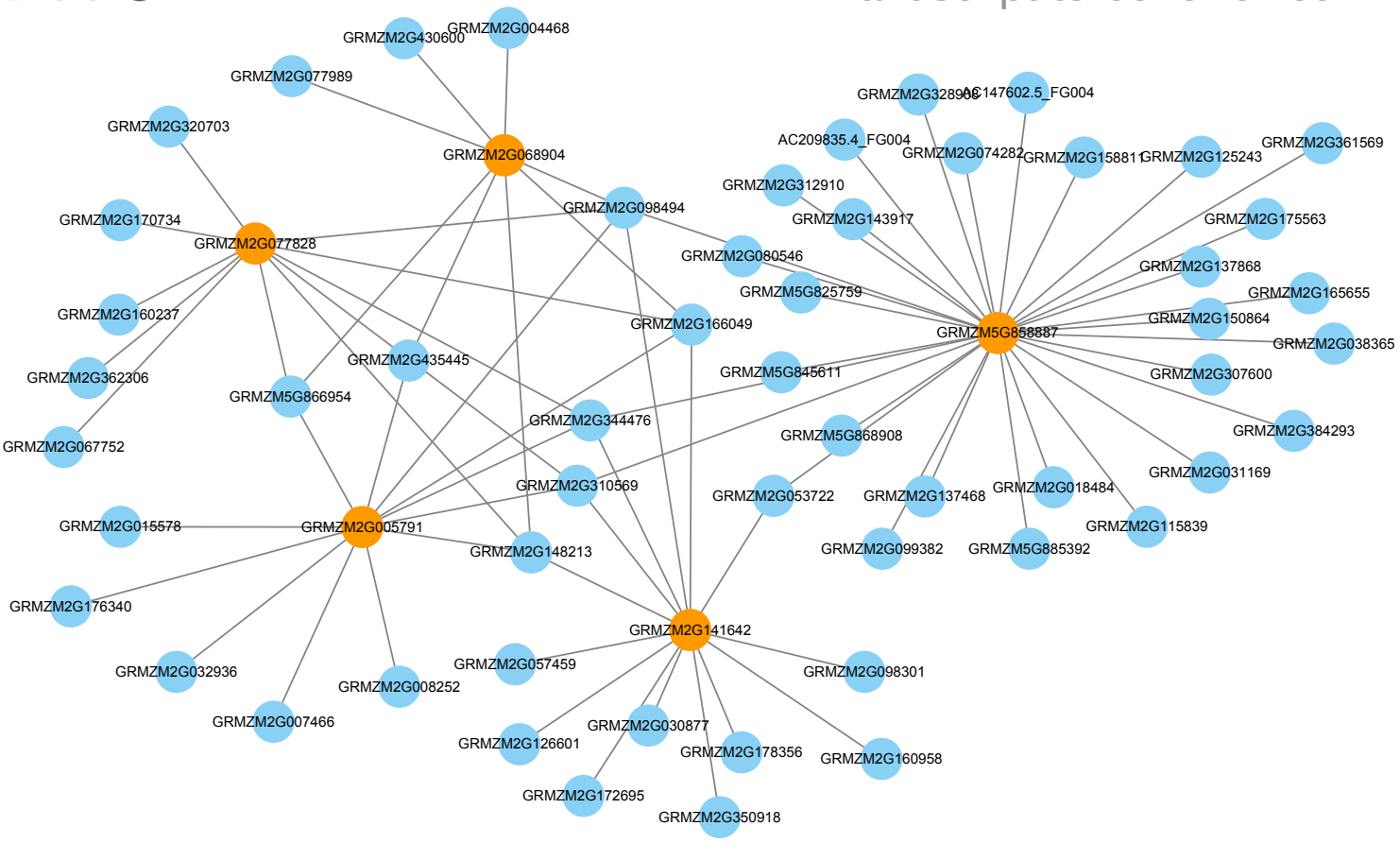

Go annotation
percent of genes
120
100
80
60
40
20
0
GO:0016020 membrane
GO:0044464 cell part
GO:0005623 cell
GO:0044425 membrane part
GO:0005215 transporter activity
GO:0005488 binding
GO:0051179 localization
GO:0009987 cellular process
GO:0022414 reproductive process
GO:0032502 developmental process
GO:0040007 growth
GO:0032501 multicellular organismal process
GO:0000003 reproduction
GO:0071840 cellular component organization or biogenesis
GO:0051704 multi-organism process
GO:0065007 biological regulation
GO:0050789 regulation of biological process
GO:0023052 signaling
GO:0050896 response to stimulus
GO:0002376 immune system process
Cellular Component
Molecular Function
Biological Process

Figure 5(on next page)

The interaction network of *ZmCNGC*s in *Zea mays* according to the orthologs in *Arabidopsis*

Yellow circus represent the *ZmCNGC*s in this study, others represent the genes in *Zea mays* genome.


GRMZM2G004468
GRMZM2G430600
GRMZM2G077989
GRMZM2G320703
GRMZM2G068904
GRMZM2G170734
GRMZM2G077828
GRMZM2G098494
GRMZM2G160237
GRMZM2G362306
GRMZM2G067752
GRMZM2G435445
GRMZM5G866954
GRMZM2G166049
GRMZM2G344476
GRMZM2G310569
GRMZM2G015578
GRMZM2G005791
GRMZM2G148213
GRMZM2G176340
GRMZM2G032936
GRMZM2G007466
GRMZM2G008252
GRMZM2G141642
GRMZM2G057459
GRMZM2G126601
GRMZM2G030877
GRMZM2G178356
GRMZM2G172695
GRMZM2G350918
GRMZM2G098301
GRMZM2G160958
GRMZM2G080546
GRMZM5G825759
GRMZM5G845611
GRMZM5G858887
GRMZM2G328908
AC147602.5_FG004
AC209835.4_FG004
GRMZM2G074282
GRMZM2G158811
GRMZM2G125243
GRMZM2G361569
GRMZM2G312910
GRMZM2G143917
GRMZM2G175563
GRMZM2G137868
GRMZM2G165655
GRMZM2G150864
GRMZM2G038365
GRMZM2G307600
GRMZM2G384293
GRMZM2G031169
GRMZM2G115839
GRMZM5G868908
GRMZM2G053722
GRMZM2G137468
GRMZM2G018484
GRMZM2G099382
GRMZM5G885392