Genome-wide Identification and Analysis of the CNGC Gene Family in Zea mays

Lidong Hao¹, Xiuli Qiao²*

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

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

*Corresponding author:

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 have 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-related and organ development-related. Gene ontology (GO) analysis demonstrate 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 expressed pattern. Our results provide valuable information to increase our understanding of the CNGC gene family in maize.

Subjects: Bioinformatics, Genomics, Plant Science

Keywords: CNGC, Zea mays, Gene family, Expression profiles

**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. 2015). The cyclic nucleotide-gated channel (CNGC) is a Ca²⁺-permeable cation transport channel, which is suggested to behave one of the fundamental mechanism for organismsal 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 Ca²⁺/calmodulin (CaM) can regulate CNGCs; those messengers are activated by directing binding of cyclic nucleotides as well as being inhibited by binding of CaM to the CaM binding domain (Borsics et al. 2007; Defalco et al. 2016; Kaplan et
As CNGCs are key components for plant development, many previous researches have been conducted. With the application of bioinformatics tools, the identification of CNGC gene family members 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. Most CNGCs 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; phot1 and phot2 increases cytosolic Ca\textsuperscript{2+} in Arabidopsis leaves; AtCNGC2 play key roles in stress signaling pathways, including changes the cytosolic free Ca\textsuperscript{2+} in Arabidopsis and CNGC4 is permeable to both K+ and Na+ and activated by both cGMP and cAMP. Meanwhile, the structure of Arabidopsis CNGCs have six transmembrane domains with a pore domain, also have a cyclic nucleotide-binding domain and CaM-binding domains in the C-terminus, and these domains have diverse functions. For example, Arabidopsis CNGC6, CNGC7, CNGC8, CNGC9, CNGC10 and CNGC16 participate in the pollen development; CNGC2, CNGC4, CNGC11 and CNGC12 are activated in response to pathogenic microorganisms; and, CNGC6, CNGC10, CNGC19 and CNGC20 are involved in abiotic stress.

In recent years, efforts had been made in studying the CNGC 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 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 systematically study of CNGC genes in maize and will provide the basis for further research on the ZmCNGC gene family.

5MATERIALS AND METHODS

Identification of CNGC genes in maize genome
To identify the maize CNGC genes, 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. Then, two methods were used to search against the maize protein sequences, one is built using a Hidden Markov Model (HMM) to search against maize protein sequences, and another is used the local BLASTP method with the e-value set to 1e-5. After that, the putative non-redundant protein sequence of maize CNGC genes were retrieved. To further confirm whether the ZmCNGC proteins have contained the CNBD domain, those putative ZmCNGC protein sequences were submitted 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 of a size of below 200 were removed and the ZmCNGC genes were confirmed. The information of chromosome distribution of ZmCNGCs and the sequences including DNA sequences, CDS, cDNA, up-stream 1500bps of ZmCNGC genes were obtained from results of the BLASTN search in the Ensembl Plant database (http://plants.ensembl.org/index.html) (Bolser et al. 2016).

Analyses of genes structure, conserved motifs, and cis-acting regulatory elements

The gene structure 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 DNA sequences. The conserved motif domins were identified using the MEME software algorithm (http://meme-suite.org/index.html) (Bailey et al. 2015) with the maximum number of motifs obtained set at 9 and the optimum width of motifs ranging from 6 to 200 amino acids. The up-stream 1500bps of ZmCNGC genes were used to find cis-acting regulatory elements by using ‘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 PI (theoretical isoelectric point), MW (molecular weight), and GRAVY of ZmCNGCs were predicted by ExPASy (http://web.expasy.org/protparam/) (Artimo et al. 2012). The prediction subcellular location of ZmCNGCs were identified using the CELLO v.2.5 (http://cello.life.nctu.edu.tw/) package. The gene ontology (GO) annotation of ZmCNGC genes were submitted to the Monocots PLAZA 4.0 (Van Bel et al. 2017) database, then were visualized and plotted by BGI WEGO (Ye et al. 2006).

Multiple alignments and phylogenetic analysis

Multiple sequences alignments were performed using T-COFFEE (http://tcoffee.crg.cat/apps/tcoffee/index.html) (Di Tommaso et al. 2011) and visualized by ESPript (Robert & Gouet 2014). An un-root phylogenetic tree was constructed with 1000 bootstrap replication using MEGA 7 (Kumar et al. 2016) based on the full-length protein sequences alignment.

Expression profiles of ZmCNGC genes by RNA-seq datasets and network interaction analysis

For understanding the expression of ZmCNGC genes in different tissues, two RNA-seq datasets of Zea mays obtained from the Expression Atlas datasets database were obtained (https://www.ebi.ac.uk/gxa/home/) with the accession numbers E-MTAB-353826 and E-MTAB-4395. Those data were used to analyze the expression of ZmCNGC in 6 different tissues (including ear, embryo, endosperm, pollen, root and tassel) and different development stages in embryo, endosperm, and seed, respectively. The FPKM values were used to calculated for each of the ZmCNGC genes. The interaction network was constructed on the based on the orthologs between maize and Arabidopsis using the AraNet v2 (Lee & Lee 2017) annotations and visualized by the Cytoscape v3.4.0 (Shannon et al. 2003).

RESULTS AND DISCUSSIONS

Identification of CNGC genes in Zea mays

To identify a complete overview of CNGC genes in maize, we firstly used 20 Arabidopsis and 16 rice CNGC protein sequences to blast align tBLAST n against maize protein sequences. After blastBLAST 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. 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 were listed in Table 1. The protein lengths ranged from 326 to 745 aa with average of 612.42 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, 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) were blasted to ZmCNGC genes using the BLASTN program from NCBI; results found only GRMZM2G066269 showed no EST hits, other ZmCNGCs had more than 13 representative matches to ESTs.

Phylogenetic analysis of CNGC genes in maize, Arabidopsis and rice

To further understanding the evolutionary relationship of CNGCs, an unrooted neighbor-joining 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 contains two maize CNGC genes (GRMZM2G077828, GRMZM2G023037), five in Arabidopsis, and three in rice. Similarly, group III contains three maize CNGC genes (GRMZM2G090528, 11 12...
And Group IV embraced seven in rice and seven in Arabidopsis, forming the largest GNGC group with four members in maize—CNGC genes—which included three in IVa (GRMZM2G068904, GRMZM2G005791 and GRMZM2G135651) and one in IVb (GRMZM2G141642). Based on the phylogenetic tree among ZmCNGCs, AtCNGCs and OsCNGCs, we also grouped maize CNGC genes into four sub-groups (Figure 3A).

**Multiple alignments, gene structure and conserved motif of ZmCNGCs**

Due to the CNBD domain is the gene's a structural feature element in plant CNGCs which contain the PBC and the hinge region (Diller et al. 2001). As shown in Figure 2, after aligning the CNBD region of maize CNGCs, the putative PBC and hinge domain were also identified, which were consistent with rice CNGCs (Nawaz et al. 2014). Results showed that glycine (G), acidic residue glutamate (E) and leucine (L) and aromatic tryptophan (W) inside the PBCs were 100% conserved. As well as, the aliphatic alanine (A), aromatic phenylalanine (F) and leucines (L) were the most conserved within the hinge region. Compared to rice and Arabidopsis (Supplemental File 2), we found that glycine (G) and leucine (L) were found to be conserved at 100% in CNGCs PBC domains, while aromatic phenylalanine (F) and leucine (L) were 100% conserved in the hinge domain. Furthermore, gene structure analysis could add better understanding to the gene functions and evolution. As a whole, the number of exons ranged from 1 to 8 while GRMZM2G077828 was intronless (Figure. 3B). In addition, 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 the most ZmCNGCs harbor motif 1, representing it is the typical ZmCNGC domain.

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

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. There are 137 different putative cis-elements were found to be presented in at least one associated with identified ZmCNGC genes and only 12, including CACTFTPPCA1, EBOXBNNAPA, DOFCOREZM, MYCCONSENSUSAT, 7 13 14
CAATBOX1, GTGANTG10, WRKY71OS, GT1CONSENSUS, ROOTMOTIFTAPOX1, POLLEN1LELAT52, MYBCORE, and OSE2ROOTNODULE, out of them apparently appeared in the promoter region of all ZmCNGC genes (Supplemental File 1: Table S1). Additionally, five cis-elements were gene-specific, such as ACGTCBOX, TATABOX3, HDZIP2ATATHB2, ABREMOTIFAOSOSEM and MRNA3ENDTAH3 and were unique to GRMZM2G005791, GRMZM2G068904, GRMZM2G074317 and GRMZM2G135651, respectively [5 compared to 4? one missing].

Also, some cis-elements were involved in different abiotic/biotic stimuli, including those such as hormone-response (abscisic acid, auxin, ethylene, etc.), stress-related (drought, temperatures, disease, etc.) and development-related (mesophyll specific, tissue specific, etc.), indicating that these ZmCNGC genes might be involved in regulating diversity stress responses. 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 transporter. Further, cellular component analysis revealed the localization of ZmCNGCs in the cell and 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 1: 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 at different tissues based on previous studies. As shown in Figure 6, the expression levels among ZmCNGC genes were tissue-specific in their expression. For example, GRMZM2G077828 and GRMZM2G148118 were highly expressed in pollen and all IVb sub-group genes (GRMZM5G858887, 258GRMZM2G074317 and GRMZM2G090528) were mainly expressed in embryo, while 259GRMZM2G129375 and GRMZM2G005791 were lowly expressed at low levels in pollen and embryo, indicated that IVb 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 showed in Figure 6 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.

Most researches showed that cyclic nucleotide-gated channels (CNGCs) have been related to pollen development and in response to environment 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**

To study the CNGC gene family in the maize, we identified 12 CNGC genes distributed in 7 chromosomes which were classified into four major groups. 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 found.
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. Taken together, 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

Funding

This work was supported by the Natural Science Foundation of Heilongjiang Province, China (B2016010). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

Natural Science Foundation of Heilongjiang Province, China: B2016010.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

Lidong Hao performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, and wrote the paper, prepared figures and/or tables. Xiuli Qiao conceived and designed the experiments, reviewed drafts of the paper.

Supplemental Information

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

Multiple sequences alignments among maize, Arabidopsis and rice.

Supplemental File 1

Supplemental File 2

Supplemental File 3
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 3.

REFERENCES


Balagué 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


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


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 Ib beta Regulatory Subunit. Structure 9:73-82. 10.1016/S0969-2126(00)00556-6


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


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


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
