Genome-Wide Identification, Characterization, and Expression Analysis of Cyclic Nucleotide-Gated Channel (CNGC) Gene Family in Daucus carota Under Drought Stress and Arbuscular Mycorrhizal Symbiosis

Hafiz Muhammad Shahzaib Siraj
Government College University Faisalabad
Sahab Sahel Alotaibi
Princess Nourah bint Abdulrahman University
Hidayah Alotaibi
hefalotaibi@pnu.edu.sa
Princess Nourah bint Abdulrahman University

Research Article

Keywords: Daucus carota, CNGC, cyclic nucleotide-gated channels, genome-wide analysis, drought stress, arbuscular mycorrhizal symbiosis, calcium signaling, ion transport

Posted Date: May 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9663686/v1

Additional Declarations: No competing interests reported.

Abstract

Background

Cyclic nucleotide-gated channels (CNGCs) are non-selective cation channels that are involved in key processes in plant signal transduction, ion homeostasis and stress adaptation. Although the CNGC gene family has well-characterized roles in model plants like *Arabidopsis thaliana*, the roles of these genes in carrot (*Daucus carota*), a highly valuable root crop, have not been well studied.

Results

A global genome-wide overview revealed 24 *DcCNGC* genes that are spread out on all eight *D. carota* chromosomes. These genes were clustered into six groups (I-VI) with phylogenetic analysis which showed conserved orthologous relationships with *AtCNGC* members and evidence of duplication events. Physicochemical profiling revealed that *DcCNGC* proteins have a size between 610–2027 amino acids (70.0-226.3 kDa), most of them with a basic isoelectric point (pI 7.17–9.66). The structural analysis showed the presence of conserved domains of cyclic nucleotide-binding (cNMP_binding), ion transport (Ion_trans), and CAP effector domains. The promoter analysis revealed various cis-regulatory factors in response to drought, abscisic acid, jasmonate, salicylate, and light. Ka/Ks ratio analysis revealed that all paralogous pairs of *DcCNGC* genes diverged under purifying selection (Ka/Ks < 1) that represents a high functional constraint. The AlphaFold3 protein structure prediction showed three-dimensional structure and was used to predict the protein structure of *DcCNGC* members, which revealed the presence of conserved transmembrane topology and cyclic nucleotide-binding pocket architecture. GO and KEGG enrichment analyses verified functional relevance in ion transport, voltage-gated channel activity, plant-pathogen interaction pathways.The expression analysis of the RNA-seq of drought stress and arbuscular mycorrhizal (AM) symbiosis identified differential expression of several *DcCNGC* genes, as *DcCNGC2a*, *DcCNGC5*, and *DcCNGC8* were upregulated during drought and AM inoculation.

Conclusions

This research presents the first systematic characterization of the CNGC family of genes in *D. carota* and a baseline resource to study calcium-mediated responses to abiotic stress. The identified candidate genes, especially, those with drought- and AM-responsive expression patterns, are promising targets in genetic enhancement of carrot under the conditions of water scarcity.

Introduction

Drought stress is one of the most severe environmental constraints affecting plant growth, productivity, and survival, particularly in arid and semi-arid regions. Water deficit disrupts cellular homeostasis, reduces photosynthetic efficiency, and induces oxidative damage, ultimately leading to major losses in crop yield [1]. In plants, adaptation to drought depends on a complex network of signaling pathways that regulate water balance, ion transport, osmotic adjustment, and stress-responsive gene expression [1]. In this context, calcium-mediated signaling is considered a central component of plant responses to abiotic stress, especially drought.

Cyclic nucleotide-gated channels (CNGCs) are a class of non-selective cation transporters that facilitate the movement of Ca^{2+}, K^{+}, and Na^{+} across cellular membranes [2]. In plants, CNGCs belong to the ancient P-loop ion channel superfamily, which arose early in evolution and diversified to regulate ion homeostasis and a wide range of cellular processes. Activated by cyclic nucleotides, these channels play a central role in calcium-mediated signaling by controlling cytosolic Ca^{2+} dynamics [3]. This regulation enables plants to perceive and respond to diverse developmental and environmental stimuli, including hormonal signals, pathogen attack, light, circadian rhythms, and abiotic stresses [4, 5]. Increasing evidence indicates that CNGCs are involved in drought stress responses across plant species. In Arabidopsis, *CNGC16* is critical for pollen fertility under heat and drought stress, with knockout mutants showing a strong reduction in pollen fitness [6]. In tomato, silencing of *SlCNGC15* resulted in increased water loss and reduced relative water content under drought conditions, indicating a positive role in maintaining water balance [7]. Similarly, in kenaf, CNGC genes were shown to enhance drought tolerance by improving antioxidant enzyme activity, reducing ROS accumulation, and maintaining membrane stability and photosynthetic performance [8]. These results suggest that CNGCs play important roles in plant adaptation to water deficit.

In addition, *CNGC15* plays a central role in root endosymbiosis by generating symbiont-induced nuclear Ca2 oscillations required for arbuscular mycorrhizal symbiosis and nitrogen-fixing bacterial symbiosis [9]. This calcium-dependent signaling mechanism connects ion channel activity with the establishment of beneficial plant–microbe interactions. Because *Arbuscular mycorrhizal* (AM) symbiosis improves nutrient uptake and enhances plant tolerance to drought, CNGC-mediated calcium signaling may represent an important molecular bridge between symbiosis and stress adaptation [10]. Previous studies have also shown that AM symbiosis alters host transcriptional responses under drought, highlighting the importance of symbiotic signaling during water limitation [11].

Carrot is highly sensitive to water deficit, and drought stress can significantly affect root development, biomass accumulation, and overall yield [12]. At the same time, carrot can establish beneficial associations with AM fungi, which improve nutrient uptake, enhance water acquisition, and increase tolerance to adverse environmental conditions [13]. In carrot, drought has been reported to influence gene expression differently in mycorrhizal and non-mycorrhizal systems, indicating that symbiotic regulation may contribute to drought adaptation [13]. As a diploid species with 2n = 2x = 18 chromosomes and a relatively compact genome, carrot provides an excellent model for genome-wide analysis of gene families [14]. Nevertheless, little information is available regarding the CNGC gene family in *D. carota*, including their evolutionary relationships, conserved features, and expression patterns under drought stress and AM symbiosis.

Therefore, in this study, we performed a genome-wide identification and characterization of the CNGC gene family in Daucus carota. We analyzed their physicochemical properties, phylogenetic relationships, gene structures, conserved motifs, chromosomal distribution, and cis-regulatory elements. In addition, we investigated their expression profiles under drought stress and arbuscular mycorrhizal symbiosis to identify candidate genes potentially involved in stress adaptation and symbiotic interaction. This study provides a theoretical foundation for future functional analyses of CNGC genes and may contribute to improving drought tolerance in carrot through molecular breeding strategies.

Material and Methods

2.1 Sequence identification of CNGC genes in *Daucus Carota*

To determine the CNGC family genes of *D. carota* in a genome-wide bioinformatics analysis, twenty CNGC protein sequences from *A. thaliana* were used as query sequences. These protein sequences were retrieved from the TAIR10 database (https://www.arabidopsis.org/) [15]. These 20 *AtCNGC* protein sequences served as queries to expectant the *D. carota* genome with the Ensembl Plants BLASTp (https://plants.ensembl.org/Multi/Tools/Blast) tool with default settings [16, 17]. The three most popular hits of each query were taken and the redundant sequences were eliminated manually. Following elimination of redundancies, non redundant potential *DcCNGC* gene sequences were retrieved to undergo further analysis. To verify the completeness and authenticity of the identified sequences, domain analysis was performed to exclude truncated sequences and those lacking conserved domains. SMART (http://smart.embl-heidelberg.de/) and CCD with default cut off parameters were also used for domain analysis (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi/) [18, 19].

2.2 Phylogenetic analysis and Multiple Sequence Alignment

To evaluate the sequence conservation and homology of the proteins, multiple sequence alignment (MSA) of *D. carota* CNGC (*DcCNGC*) protein sequences alongside A. thaliana CNGC (*AtCNGC*) protein sequences were compared using MEGA12 with the default parameters [20]. IQ-TREE (http://iqtree.cibiv.univie.ac.at) was then used to create a phylogenetic tree, basing on the Maximum Likelihood (ML) method, using the aligned sequences [21]. The software automatically selected the best fit substitution model based on the Bayesian Information Criterion (BIC). A total of 1000 bootstrap replicates were carried out to determine the strength of the phylogenetic relationships. The phylogenetic tree thus generated was visualized and annotated with the help of the interactive Tree of Life (iTOL) (https://itol.embl.de/) [22]. The tree layout was refined using iTOL, major clades were color-coded, and proteins in the *DcCNGC* and *AtCNGC* groups were labeled as per their evolutionary groupings.

2.3 Protein characterization and amino acid properties

ExPASy proteomics Server was used for the analysis of protein sequence of putative *DcCNGC* members. Online available tool ProtParam (https://web.expasy.org/protparam/) of ExPASy (https://www.expasy.org/) was used for the determination of amino acid properties [23], such asmolecular weight, protein length, instability index, aliphatic index and isoelectric points of *DcCNGC* proteins. To predict the subcellular localization of the identified DcCNGC proteins, WoLf PSORT (https://wolfpsort.hgc.jp/) was used [24]. The analysis was done with default parameters to predict which cellular compartments each protein probably belongs to based on their amino acid composition and sorting signals.

2.4 Structural Analysis

The online available tool Gene Structure Display Server (GSDS) with DNA and CDS as the input information with default settings (http://gsds.gao-lab.org/Gsds_about.php) was used for analysis of the structural and exon-intron distribution among the *DcCNGC* genes [25]. MEME tool was used for the detection of conserved protein motif of encoded CNGC proteins (http://meme.sdsc.edu/meme/) [26]. It included the following parameters, 10 maximum number of different motifs and each motif with any number of non-overlapping occurrences.

2.5 Location of *DcCNGC* genes on chromosomes

The position of the *DcCNGC* genes was identified and analyzed in the NCBI Genome Database (https://www.ncbi.nlm.nih.gov/genome/), where the number and position of the start and end of each gene were obtained [27]. The genetic linkage map showing the distribution of the *DcCNGC* genes over the chromosomes was then created using the MG2C v2.1 tool (http://mg2c.iask.in/mg2c_v2.1/) [28].

2.6 Cis-acting regulatory elements

To determine how the *DcCNGC* genes cis-acting regulatory elements in their promoter regions could have a regulatory role, their presence was determined. The 2,000 bp upstream sequences of the translation start position of every *DcCNGC* gene were obtained in NCBI Genome Database (https://www.ncbi.nlm.nih.gov/) [27]. They were subsequently examined with the PlantCARE database (https://bioinformatics.psb.ugent.be/webtools/Plantcare/html) to forecast putative cis-regulatory factors related to growth, development and stress responses in plants [29]. The identified factors were divided as per their functional roles, including light responsiveness, hormone regulation, abiotic or biotic stress response.

2.7 Analysis of Gene Expression pattern

The patterns of expression of the *DcCNGC* genes in the drought stress and *arbuscular mycorrhizal* (AM) symbiosis were determined using publicly available RNA-seq data in the NCBI Sequence Read Archive (SRA) with accession number PRJNA784577 [13]. The data consisted of four treatments namely, non-mycorrhizal control (C_Clean), mycorrhizal control (C_Infected), non-mycorrhizal drought (D_Clean) and mycorrhizal drought (D_Infected) of four replicates each. Aligning HISAT2 to clean reads reference to the Daucus carota reference genome and quantification of gene expression levels were performed using StringTie and giving the value as FPKM. The gene patterns of DcCNGC were set to trace the expression patterns with TBtools [30].

2.8 Synonymous and Non-synonymous Substitution Rate Analysis

To measure the evolutionary selection pressures on the *DcCNGC* gene family, the synonymous (Ks) and non-synonymous (Ka) substitution rates were computed across all known paralogous *DcCNGC* gene pairs. The coding sequences of every pair of genes were aligned with MUSCLE and Ka/Ks ratios were calculated with a Ka/Ks calculation option deployed in the TBtools software [31], using Yang and Nielsen (YN) method. Ka/Ks < 1 represents purifying (negative) selection, Ka/Ks = 1 represents neutral evolution, and Ka/Ks > 1 represents positive selection. This was calculated in all 32 paralogous gene pairs that were found in the duplication analysis.

2.9 Three-Dimensional Protein Structure Prediction.

AlphaFold3 (https://alphafoldserver.com), a deep learning-based protein structure prediction framework that uses amino acid sequence to predict protein structure with high precision, was used to predict three-dimensional (3D) protein structures of all *DcCNGC* members [32]. UCSF ChimeraX (version 1.7) was used to visualize and analyze the predicted structures, and to render high-quality structural images, annotate transmembrane helices and identify the location of the cyclic nucleotide-binding domain (CNBD) and ion transport domain in each model [33]. Structural conservation across *DcCNGC* members was assessed by superimposition of predicted models within ChimeraX.

2.10 GO and KEGG Enrichment Analysis.

The functional annotations of the 24 *DcCNGC* genes were characterized by performing Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses. A web based tool shinyGO v0.76 (https://bioinformatics.sdstate.edu/go76/) was used to identify these GO terms and KEGG pathways [34]. Enriched GO terms in the three ontology classes, Biological Process (BP), Cellular Component (CC) and Molecular Function (MF) and enriched KEGG pathways were displayed as bubble/lollipop plots plotted with ggplot2.

Results

3.1 Identification of DcCNGC Genes

Genome-wide searches of CNGC genes in *D. carota* found over 50 putative sequences first by querying with the *A. thaliana* CNGC proteins. Having eliminated all the duplicated and redundant sequences, 32 non-redundant candidates that had over 60% sequence similarity with *A. thaliana* CNGCs remained to be further validated. Pfam, SMART and NCBI CDD databases were then used to conduct a conserved domain analysis that would validate the presence of characteristic domains. It was found during the analysis that the majority of the sequences had conserved domains, which included: the cyclic nucleotide-binding domain (cNMP_binding; pfam00027), ion transport protein domain (Ion_trans; pfam00520), cAMP-binding domain (Crp; COG0664), voltage-dependent potassium channel domain (PLN03192), and effector domain of the CAP family (CAP_ED; cd00038). According to the availability of these domain features, 24 sequences were verified as putative *DcCNGC* genes which were then subjected to further structural and functional analysis.

3.2 Phylogenetic Analysis

The protein sequences of *DcCNGC* and *AtCNGC* were used to build a phylogenetic tree to analyze their evolutionary connections. The 24 *DcCNGC* genes were clustered into six separate groups (I-VI) based on sequence similarity and clustering patterns each group is represented by a specific color in the phylogeny tree, as shown in Fig. 1. Evolutionary relationships were found several orthologous pairs including *DcCNGC1a-b/AtCNGC1* in group 2, *DcCNGC15a-d/AtCNGC15* in group 3, *DcCNGC14/AtCNGC14, DcCNGC16/AtCNGC16, DcCNGC17/AtCNGC17*, and *DcCNGC18/AtCNGC18* in group 4, *DcCNGC19/AtCNGC19* and *DcCNGC20/AtCNGC20 in group 5, DcCNGC4a-d/AtCNGC4* and *DcCNGC2a-e/AtCNGC2* in group 6. Several paralogous pairs such as *DcCNGC2a-e, DcCNGC4a-d* and *DcCNGC15a-d* indicate that duplication of the genes contributed significantly to the development of the CNGC gene family in carrot. The grouping pattern indicates conservation and species specific diversification of *DcCNGCs*.

Figure 1: **Maximum Likelihood (ML) phylogenetic tree of D. carota (DcCNGC) and A. thaliana (AtCNGC) proteins generated with IQ-TREE and visualized with iTOL. The tree is subdivided into six major groups (I–VI) represented in distinct colors. Bootstrap values ≥ 70 are shown at branch nodes.**

3.3 Protein Characterization and Properties

The identified *DcCNGC* proteins had varying physicochemical properties (Table 1). The size of the protein was 610 (*DcCNGC1a*) to 2027 amino acids with a molecular weight of about 70.0 kDa up to 226.3 kDa. Theoretical isoelectric points (pI) of all *DcCNGC* proteins were basic in the range of 7.17 to 9.66, which implied that all *DcCNGC* proteins were positively charged in physiological conditions. The values of the instability index showed that the majority of *DcCNGCs* are unstable in vitro with high thermostability of their aliphatic index values (between 83.5 and 98.8). Most of the GRAVY (Grand Average of Hydropathy) scores were found to be negative which suggests that most of the *DcCNGC* proteins are hydrophilic.

Subcellular localization analysis with WoLF PSORT suggested that the majority of *DcCNGC* proteins are localized in the plasma membrane, which is in line with their role as ion channels. It is worth noting that *DcCNGC16* was expected to be located to the endoplasmic reticulum, whereas *DcCNGC2e* was located in the chloroplast, implying the potential functional diversification of the gene family.

Table 1
Physicochemical properties and predicted subcellular localization of DcCNGC proteins.

Name	ID (NCBI)	AA	MW (Da)	pI	Instability Index	Aliphatic Index	GRAVY	Chr	Start	End	Strand	Subcellular Localization
DcCNGC13	KZN04546	777	89141.9	9.04	43.71	90.64	-0.17	2	12453109	12462955	1	Plasma membrane
DcCNGC14	KZN08016	719	82541.7	9.32	50.63	85.84	-0.23	1	1457321	1461226	-1	Plasma membrane
DcCNGC15a	KZM97113	2027	226395.4	8.20	53.43	83.53	-0.25	4	10940676	10961402	-1	Plasma membrane / ER
DcCNGC15b	KZM95553	678	77142.1	8.93	46.23	92.45	-0.11	5	33272045	33275667	1	Plasma membrane
DcCNGC15c	KZM95555	1082	120996.8	8.06	40.98	86.53	-0.07	5	33299544	33306560	1	Plasma membrane
DcCNGC15d	KZN00544	703	80169.1	8.60	49.68	97.11	0.05	3	6248017	6250877	-1	Plasma membrane
DcCNGC16	KZN05228	695	80192.5	8.69	44.70	87.73	-0.21	2	21443047	21445764	1	Endoplasmic reticulum
DcCNGC17	KZM92527	728	83742.3	9.11	48.00	92.31	-0.20	6	34106471	34115581	-1	Plasma membrane
DcCNGC18	KZM90958	695	79497.9	9.14	48.63	90.52	-0.07	6	20838864	20841497	-1	Plasma membrane
DcCNGC19	KZM97901	703	81356.8	9.27	40.07	86.69	-0.29	4	19244824	19266177	-1	Plasma membrane
DcCNGC1a	KZN01854	610	70003.8	9.58	47.29	98.80	-0.08	3	26130095	26133972	1	Plasma membrane
DcCNGC1b	KZM92325	709	82106.6	9.24	51.02	88.98	-0.23	6	32528756	32538268	-1	Plasma membrane
DcCNGC20	KZM89376	778	89336.1	9.32	39.15	88.96	-0.19	6	212197	221244	1	Plasma membrane
DcCNGC2a	KZN02306	697	79735.2	9.57	51.34	96.61	0.02	3	31873610	31877526	1	Plasma membrane
DcCNGC2b	KZM88330	713	81348.8	9.48	51.39	93.11	-0.01	7	26866034	26869195	-1	Plasma membrane
DcCNGC2c	KZN02307	686	80160.4	9.32	50.27	91.94	-0.12	3	31879242	31883555	1	Plasma membrane
DcCNGC2d	KZM85917	621	71428.7	9.50	48.52	91.48	0.01	8	29904952	29908808	-1	Plasma membrane
DcCNGC2e	KZM85919	616	71531.3	9.66	43.51	94.77	0.01	8	29920330	29923876	1	Chloroplast
DcCNGC4a	KZM93714	667	77729.5	7.77	51.96	95.83	-0.03	5	9602435	9605796	1	Plasma membrane
DcCNGC4b	KZM86622	659	76090.4	7.17	58.56	91.71	-0.04	7	4987560	4990639	1	Plasma membrane
DcCNGC4c	KZM90146	670	77758.3	8.59	48.12	93.55	-0.12	6	11692043	11695455	-1	Plasma membrane
DcCNGC4d	KZN01233	663	76994.7	9.02	47.98	96.27	-0.07	3	13379789	13384344	1	Plasma membrane
DcCNGC5	KZM97830	728	82998.6	8.93	48.82	90.98	-0.12	4	18186340	18192824	-1	Plasma membrane
DcCNGC8	KZM92547	699	80530.9	9.23	50.43	93.78	-0.15	6	34254392	34258925	1	Plasma membrane

AA: amino acids; MW: molecular weight; pI: isoelectric point; GRAVY: grand average of hydropathy; Chr: chromosome; ER: endoplasmic reticulum.

3.4 Gene Structure, Conserved Motif, and Domain Analysis

The structures, conserved motifs, and domain architecture of *DcCNGC* genes were analyzed to comprehend their evolutionary connections and their possible functions. The *DcCNGC* genes had various exons and introns most of which varied in the numbers and length across the family members as indicated by Fig. 2. This inconsistency implies genetic family deviation. The *DcCNGC15a* gene had an unusually large genomic region and had multiple introns which suggested the possibility of gene expansion or alternative splicing.

Figure 2. **Exon–intron organization of DcCNGC genes generated by GSDS. Red boxes represent coding sequences (CDS), blue boxes indicate upstream/downstream regions, and black lines represent introns. Gene structures are displayed from 5' to 3'.**

Most of the *DcCNGC* proteins had a similar pattern of motif organization, specifically motifs 1, 2, 3, 5, and 7 as in Fig. 3. These conserved motifs represent intense structural and functional conservation throughout the family, and minor differences among some of the members can be a sign of functional diversification.

Figure 3. **Conserved motifs of DcCNGC proteins identified by MEME analysis. Each colored box represents a distinct conserved motif (Motifs 1–10) as indicated in the legend. Motif positions are displayed from 5' to 3'.**

Analysis of domains showed that a number of important conserved domains existed including those of CAP_ED, Voltage-dependent potassium channel, Ion_trans and cyclic nucleotide-binding domains, which include Crp, cNMP and cNMP binding as indicated in Fig. 4. The similarity in Ion trans and cNMP binding domains in almost all members prove their classification as cyclic nucleotide-gated channels, which are possibly considered to have a role in the ion transport and signal transduction in carrot.

Figure 4. **Conserved domain architecture of DcCNGC proteins predicted by Pfam, SMART, and NCBI-CDD databases. The color segments depict separate functional domains as indicated in the legends.**

3.5 Chromosomal Distribution of *DcCNGC* Genes

The genomic coordinate data available in NCBI Genome Database was used to map all 24 *DcCNGC* genes to the eight chromosomes of *D. carota* and visualize the data with PhenoGram (Fig. 3). The highest number of *DcCNGC* genes was found in chromosome 6 (n = 6: *DcCNGC17, DcCNGC18, DcCNGC1b, DcCNGC20, DcCNGC4c*, and *DcCNGC8*), which indicated that the region had undergone gene duplication. Chromosome 3 had 5 genes (*DcCNGC15d, DcCNGC1a, DcCNGC2a, DcCNGC2c* and *DcCNGC4d*), whereas chromosome 4 and chromosome 5 possess four genes each. Each chromosome (1, 2, 7, and 8) had one or two members. The physical clustering of paralogous genes, including *DcCNGC2a* and *DcCNGC2c* on chromosome 3, and *DcCNGC15b* and *DcCNGC15c* on chromosome 5, is in line with tandem or segmental duplication events that led to the expansion of the CNGC family in *D. carota*.

Figure 5: **Chromosomal distribution of DcCNGC genes across the eight D. carota chromosomes, showing their physical positions. Gene names are labeled adjacent to their chromosomal locations.**

3.6 *Cis*-acting Regulatory Elements

The 2 kb upstream promoter regions of each of the 24 *DcCNGC* genes were analyzed with PlantCARE, which indicated a highly diverse set of cis-acting regulatory elements (Fig. 4). The presence of universal core promoter elements such as TATA-box and CAAT-box was observed in almost all *DcCNGC* promoters. A set of stress- and hormone-sensitive factors was identified as well: ABRE (abscisic acid responsiveness), MBS (MYB binding site involved in drought induction), W-box (WRKY transcription factor binding), ARE (anaerobic induction), as-1 (salicylic acid responsiveness), CGTCA-motif and TGACG-motif (methyl jasmonate Several light-reactive motifs: G-box, Box-4, and TCT-motif, were widely diffused in the promoters. Some of the most common were MYB-, MYC-, and WRKY-binding motifs, indicating that *DcCNGC* genes were extensively regulated by defense-related and drought-related transcriptional regulatory networks.

DcCNGC14, DcCNGC1b, DcCNGC2a, DcCNGC5, and DcCNGC8 had the largest number of *cis*-regulatory elements (as shown in supplementary Table S1), indicating that these genes could be integrators of several stress and hormone signals. The presence of ABRE, MBS and W-box motifs in the promoters of these genes suggest that they are actively involved in ABA-mediated drought signaling and WRKY-mediated pathogen defense mechanisms.

Figure 6: **Cis-regulatory elements in the 2 kb promoter regions of DcCNGC genes. Elements are grouped by functional category: hormone responsiveness, abiotic stress, light responsiveness, and core promoter elements.**

3.7 Expression Analysis of *DcCNGC* Genes Under Drought and *Mycorrhizal Association*

To determine the transcriptional regulation of the *DcCNGC* genes in response to stress and symbiosis, the FPKM expression values of the *DcCNGC* genes in the four treatment groups were compared and visualized as a heatmap in TBtools. Hierarchical clustering resulted in three different expression clusters with each having a distinct transcriptional signature across the non-mycorrhizal control (C_Clean), mycorrhizal control (C_Infected), non-mycorrhizal drought (D_Clean) and mycorrhizal drought (D_Infected) treatments.

Cluster 1, including *DcCNGC2a, DcCNGC2b, DcCNGC4d, DcCNGC2e, DcCNGC15d*, and *DcCNGC5*, showed a complicated, condition-dependent expression pattern. *DcCNGC15d, DcCNGC5* were significantly repressed under C_Clean, with *DcCNGC2b*, and *DcCNGC4d* showing moderate transcriptional induction, indicating divergent basal regulation in this cluster. AM colonization under non-stressed conditions (C_Infected) strongly repressed *DcCNGC2a, DcCNGC2b* and *DcCNGC4d*, suggesting that the signaling of mycorrhiza actively down-regulates a sub-set of these genes in the non-stressed situation. This trend was reversed by drought stress alone (D_Clean) in both *DcCNGC15d* and *DcCNGC5*, both of which were transcriptionally induced, which is consistent with a role in drought-responsive calcium signaling. Remarkably, under the combined mycorrhizal drought treatment (D_Infected), all members of Cluster 1 were upregulated in concert, suggesting a convergent transcriptional response to the combined effect of the drought stress and the mycorrhizal colonization. This evidence indicates that although under isolation conditions, individual Cluster 1 genes do not respond in a similar manner to either symbiosis or drought, in isolation, the collective induction of these genes by combined conditions might represent a concerted Ca 2 + signaling program that enhances stress tolerance in mycorrhizal roots.

Those clusters included in the analysis (Cluster 2) which consists of *DcCNGC4a, DcCNGC4c, DcCNGC14, DcCNGC1b, DcCNGC17, DcCNGC15a, DcCNGC16, DcCNGC4b, DcCNGC18*, and *DcCNGC2c* exhibited a significantly homogenous expression pattern that is characterized by consistent upregulation in C_Clean conditions only. Throughout C_Infected, D_Clean and D_Infected treatments, all members of Cluster 2, showed minor down-regulation or slight transcriptional changes, indicating an impressive insensitivity to mycorrhizal colonization and drought stress, either individually or in their combination. This C_Clean-restricted expression pattern indicates that Cluster 2 genes mainly act in basal calcium homeostasis under optimum growth conditions, and that they are transcriptionally suppressed generally in response to either symbiotic or abiotic stress signals. The upstream regulators of this coordinated repression, along with their promoter architectures, should be studied further.

Cluster 3 (*DcCNGC8, DcCNGC15b, DcCNGC19, DcCNGC20, DcCNGC2d, DcCNGC15c, DcCNGC13*, and *DcCNGC1a*) had the most dynamic and treatment-responsive expression profile of the three clusters. *DcCNGC15c* was highly expressed at C_Clean with moderate upregulation of *DcCNGC13, DcCNGC1a, DcCNGC19*, and *DcCNGC8*. The expression of most of the Cluster 3 members was significantly increased by AM colonization (C_Infected), with the highest overall level of expression of the genes in this treatment being the expression of *DcCNGC13*. It is in agreement with the established role of cyclic nucleotide-gated channels in mediating the calcium oscillations that are required to mediate mycorrhizal signal transduction and symbiotic organ development. *DcCNGC8, DcCNGC15b, DcCNGC19* were up regulated and *DcCNGC20, DcCNGC2d, DcCNGC15c*, and *DcCNGC13* were moderately repressed, indicating a functional divergence in response to water deficit within Cluster 3. Notably, the combined D_Infected treatment causes strong and widespread transcriptional repression across *DcCNGC8, DcCNGC15b, DcCNGC19, DcCNGC2d, DcCNGC15c, DcCNGC13* and *DcCNGC1a* indicating that the combined imposition of mycorrhizal colonization and drought stress fundamentally reconfigures the transcriptional landscape of this cluster. This general down-regulation can be indicative of a possible strategic reallocation of calcium signaling resources, in which symbiosis-related CNGC signaling is suppressed under extreme water deficit conditions to favor the drought adaptation signaling.

Figure 7: **Heatmap showing expression profiles (FPKM) of DcCNGC genes across four treatment conditions: non-mycorrhizal control (C_Clean), mycorrhizal control (C_Infected), non-mycorrhizal drought (D_Clean), and mycorrhizal drought (D_Infected). Color scale represents log2-transformed FPKM values; red indicates high expression and blue indicates low expression.**

3.8 Ka/Ks Ratio Analysis of DcCNGC Paralogous Gene Pairs

The rates of synonymous (Ks) and non-synonymous (Ka) replacements were then compared to assess the evolutionary selection forces that have operated on the *DcCNGC* gene family since duplication of the gene families (Table 2). A total of 32 paralogous *DcCNGC* gene pairs were identified by chromosome and phylogenetic analyses and the synonymous (Ks) and non-syn Ka/Ks ratios were between 0.0463 (*DcCNGC1b/DcCNGC13*) and 0.3556 (*DcCNGC15b/DcCNGC15c*), all of which were less than 1.0. The universal nature of Ka/Ks < 1 in all gene pairs gives clear support that the whole *DcCNGC* family of genes has undergone purifying (negative) selection meaning that the protein functions have remained highly conserved following instances of gene duplication.

The Ks values among the pairs of genes varied significantly, with 0.2623 (*DcCNGC2d/DcCNGC2e*) as the lowest and 5.1950 (*DcCNGC1b/DcCNGC13*) as the highest, which is a large range of divergence times. Pairs with lowest Ks values (e.g., *DcCNGC2d/DcCNGC2e* (Ks = 0.2623)) are probably products of relatively recent duplications, and pairs with high Ks values (e.g., *DcCNGC1b/DcCNGC13* (Ks = 5.1950)) are more ancient Gene pairs with the lowest Ka/Ks ratios, such as *DcCNGC1b/DcCNGC13* (0.0463), *DcCNGC4c/DcCNGC4d* (0.0595) and *DcCNGC15a/DcCNGC15b* (0.0994), imply the strictest functional constraint, possibly critical conserved functions in ion channel .Although *DcCNGC* gene copy number has expanded through duplication, functional divergence has been constrained by purifying selection, preserving the essential biophysical properties required for cyclic nucleotide-gated channel activity.

Table 2
Ka/Ks ratio analysis of paralogous *DcCNGC* gene pairs.

Gene Pair	Ka	Ks	Ka/Ks	S-sites	N-sites
DcCNGC2d / DcCNGC2e	0.0758	0.2623	0.2890	429.92	1412.08
DcCNGC15a / DcCNGC15b	0.0533	0.5363	0.0994	478.75	1546.25
DcCNGC4a / DcCNGC4b	0.1747	0.5547	0.3149	438.75	1508.25
DcCNGC19 / DcCNGC20	0.1299	0.5560	0.2336	485.17	1605.83
DcCNGC15b / DcCNGC15c	0.1978	0.5562	0.3556	449.00	1459.00
DcCNGC2c / DcCNGC2e	0.1863	0.8453	0.2204	422.75	1404.25
DcCNGC2c / DcCNGC2d	0.1814	0.8719	0.2081	425.50	1437.50
DcCNGC4c / DcCNGC4d	0.0585	0.9827	0.0595	448.25	1537.75
DcCNGC15a / DcCNGC15c	0.3771	1.0627	0.3549	699.33	2252.67
DcCNGC1a / DcCNGC1b	0.0936	1.1097	0.0843	430.92	1369.08
DcCNGC2a / DcCNGC2b	0.1276	1.3181	0.0968	497.33	1587.67
DcCNGC16 / DcCNGC18	0.3114	1.4044	0.2217	471.75	1556.25
DcCNGC5 / DcCNGC8	0.2480	1.6188	0.1532	480.50	1568.50
DcCNGC4a / DcCNGC4d	0.3018	1.6732	0.1804	442.92	1507.08
DcCNGC4a / DcCNGC4c	0.3036	1.6889	0.1797	438.83	1520.17
DcCNGC2b / DcCNGC2c	0.3495	1.8593	0.1880	469.58	1558.42
DcCNGC15b / DcCNGC15d	0.2311	1.9925	0.1160	467.83	1512.17
DcCNGC15a / DcCNGC15d	0.2334	2.0425	0.1143	495.42	1589.58
DcCNGC4b / DcCNGC4d	0.3005	2.1104	0.1424	433.08	1480.92
DcCNGC14 / DcCNGC17	0.2216	2.1230	0.1044	476.67	1608.33
DcCNGC2a / DcCNGC2e	0.3430	2.1795	0.1574	435.25	1403.75
DcCNGC4b / DcCNGC4c	0.2927	2.2369	0.1309	427.67	1483.33
DcCNGC2a / DcCNGC2d	0.3497	2.5490	0.1372	428.08	1404.92
DcCNGC15c / DcCNGC15d	0.3641	3.0803	0.1182	467.58	1515.42
DcCNGC15d / DcCNGC18	0.4442	3.1060	0.1430	469.92	1513.08
DcCNGC2e / DcCNGC4d	0.5251	3.3314	0.1576	424.67	1396.33
DcCNGC1a / DcCNGC5	0.3334	3.4981	0.0953	427.25	1339.75
DcCNGC5 / DcCNGC13	0.3756	3.5798	0.1049	492.83	1616.17
DcCNGC1b / DcCNGC8	0.4010	3.6339	0.1104	473.17	1554.83
DcCNGC2a / DcCNGC4b	0.5790	4.0165	0.1442	447.00	1488.00
DcCNGC8 / DcCNGC20	0.5998	4.3103	0.1392	417.67	1394.33
DcCNGC1b / DcCNGC13	0.2406	5.1950	0.0463	481.25	1600.75

Ka: non-synonymous substitution rate; Ks: synonymous substitution rate; S-sites: synonymous sites; N-sites: non-synonymous sites. All gene pairs show Ka/Ks < 1, indicating purifying (negative) selection.

3.9 Three-Dimensional Protein Structure Prediction

The AlphaFold3 was used to predict three-dimensional protein structures of representative *DcCNGC* members of each phylogenetic clade, and visualized in UCSF ChimeraX (Fig. 6). The canonical CNGC structural architecture was present in all the predicted models; six transmembrane helices (S1-S6), a pore-forming loop between the S5 and S6, and a C-terminal cyclic nucleotide-binding domain (CNBD). The general topology of the structure of *DcCNGC* proteins is very conserved among the members, which explains the similarity in domain composition as determined by sequence analysis.

All the modeled *DcCNGC* proteins had well-defined alpha-helical structures with the S4 helix comprising positively charged residues which are typical of the voltage-sensing domains of similar potassium channels. The pore-forming loop, determining ion selectivity, was structurally conserved among all representatives, but small differences in loop geometry were found between the phylogenetic groups, and could be due to different ion selectivity and channel gating behaviour between *DcCNGC* clades.

The C-terminal CNBD was well defined in all models, assuming a beta-roll scaffold with alpha-helices on either side which is typical of the catabolite activator protein (CAP) fold. All models were identifiable based on the cyclic nucleotide-binding pocket that is lined by conserved residues that mediate cAMP and cGMP coordination in AtCNGC homologs. The structural superimposition of *DcCNGC* members into ChimeraX indicated that the CNBD structure is strongly conserved with root mean square deviation (RMSD) values less than 2.0 A between pairwise comparisons of phylogenetic groups showing that structural conservation is in line with the purifying selection found in Ka/Ks analysis.

Figure 8: **Three-dimensional protein structures of representative DcCNGC members predicted by AlphaFold3 and visualized using UCSF ChimeraX. Transmembrane helices (S1–S6) are shown in ribbon representation.**

3.10 GO and KEGG Enrichment Analysis

To characterize the functional roles of the *DcCNGC* gene family at the systems biology level, GO and KEGG enrichment analyses were performed and visualized as lollipop plots separated by ontology category (Fig. 7). The Biological Process (BP) category showed the largest fold enrichment value (> 100-fold), with potassium ion transmembrane transport and potassium ion transport as the most significantly enriched within the BP category. There was also strong enrichment on ion transmembrane transport as well as ion transport with medium-high gene counts and significant FDR values. Other enriched BP terms were metal ion transport, inorganic cation transmembrane transport, transmembrane transport, inorganic ion transmembrane transport, cation transmembrane transport, and cation transport.

The most significantly enriched terms in the Cellular Component (CC) category were apical plasma membrane and apical part of cell, both with very high fold enrichment values (> 300-fold) with very low FDR, though with lower numbers of genes. This enrichment pattern is coherent with the highly plasma-membrane-targeted localization that *DcCNGC* proteins are predicted to exhibit and indicates their specific targeting to the apical region of the polarized plant cells, which are applicable to root hair growth, pollen tube development, and directional ion uptake.

The Molecular Function (MF) category provided the most diverse list of enriched terms. The activity of voltage-gated potassium channels was most strongly enriched (> 300-fold) and then potassium channel activity (> 200-fold) and ion channel activity, which was enriched with the most genes (≥ 20 genes) of all MF terms and had the highest statistical significance (lowest FDR). Passive transmembrane transporter activity, channel activity, cation channel activity, inorganic molecular entity transmembrane transporter activity, ion transmembrane transporter activity, transmembrane transporter activity and transporter activity were also significantly enriched, with the range of transport activities proposed to be attributed to CNGC proteins.

The analysis of the KEGG pathways showed that the only known pathway, but the most enriched one, was that of plant-pathogen interaction with moderate fold enrichment (~ 100-fold) and the number of genes being about 10. This enhancement is in line with the already established functions of CNGCs in the innate immunity of plants, especially in the mediation of Ca2 + influx in the process of hypersensitive response and systemic acquired resistance.

Figure 9: **GO and KEGG enrichment analysis of DcCNGC genes. Lollipop plots show enriched terms across Biological Process (BP), Cellular Component (CC), Molecular Function (MF), and KEGG Pathway categories. The x-axis represents fold enrichment; dot size indicates gene count; dot color represents -log10(FDR), with warmer colors (red/yellow) indicating higher statistical significance.**

Discussion

The present work provides the initial genome-wide analysis of the CNGC gene family in *D. carota* and 24 gene members that are structurally as well as functionally conserved to the Arabidopsis orthologs. The 24 *DcCNGCs* identification is in agreement with CNGC family sizes of other dicotyledonous crops including 20 in tomato, 35 in soybean and 20 in *Arabidopsis* and is reflective of the conserved evolutionary pressure to retain this gene family across angiosperm lineages [7, 35].

The Phylogenetic grouping of *DcCNGCs* into six groups reflects the classification system that has been determined on *Arabidopsis* CNGCs, and supports the usefulness of the *AtCNGC* framework in making evolutionary inferences on other plant species. The numerous pairs of paralogs identified, especially *DcCNGC2a-2e, DcCNGC4a-4d*, *DcCNGC15a-15d*, suggest a large amount of gene duplication activity, and tandem and segmental duplications are likely the reasons why the *DcCNGC* family has expanded in carrot. This interpretation is supported by physical clustering of paralogs on chromosomes 3, 5 and 6. CNGCs in wheat and *Brassica napus* have been reported to have similar patterns of chromosomal expansion with duplication [36, 37].

The Ka/Ks ratio analysis offers a critical evolutionary view on *DcCNGC* family evolution after duplication. The universal Ka/Ks < 1 observed in all 32 paralogous gene pairs is a clear demonstration of the widespread purifying selection on the *DcCNGC* gene family. The result suggests that even though the genes copy number has increased, functions encoded by the genes have been highly conserved, perhaps due to the fact that changes in channel biophysics, including ion selectivity, gating kinetics, or ligand-binding affinity, would be counteradaptive to cellular ion homeostasis. The most extreme functional constraint is on the gene pairs with the lowest Ka/Ks values e.g., *DcCNGC1b/DcCNGC13* (0.0463) and *DcCNGC4c/DcCNGC4d* (0.0595), which may be the most essential in the basal calcium channel activity. The broad range of Ks values across gene pairs indicates that *DcCNGC* duplications occurred across multiple evolutionary timepoints rather than as a single burst, suggesting ongoing, stepwise expansion of the gene family in *D. carota*. This is compatible with the Ka/Ks measurements reported of CNGCs in other plant species, in which purifying selection was also predominant within the gene family [38].

The Physicochemical characteristics of *DcCNGC* proteins predict plant membrane-integrated ion channels. The largely low pI values indicate that *DcCNGC* proteins preferentially interact with negatively charged phospholipid membranes and the plasma membrane localization that is predicted of most members is similar to their current uses as membrane ion channels. The localization of *DcCNGC16* to the endoplasmic reticulum and of *DcCNGC2e* to the chloroplast are interesting predictions, which might suggest compartment-specific functions in Ca2 + homeostasis as some *Arabidopsis* CNGCs have been proposed to have specialized roles in regulating intracellular Ca^{2+} fluxes within distinct organelles [39].

The distinctive structural hallmark of the CNGC superfamily is the conserved domain architecture that is found across *DcCNGC* proteins- including Ion-trans, cNMP-binding, and CAP-ED domains. The exon intron structure of *DcCNGC* members, especially the larger than normal and multi intronic *DcCNGC15a*, could be indicative of variation in post-transcriptional regulation through alternative splicing as reported in CNGCs in other species [40]. The structural integrity of the *DcCNGC* proteins is also supported by the conserved core motifs found in the MEME analysis and offers a foundation upon which structure-function studies may be conducted in future.

DcCNGC genes are of particular interest in terms of their promoter landscape. The abundance of ABRE, MBS and W-box elements highlights the involvement in ABA, drought and defense signaling networks with *DcCNGC* expression. The main cis-regulatory elements that mediate ABA-responsive gene expression in response to drought and osmotic stress are the ABA-responsive regulatory elements found in the promoters of *DcCNGC2a, DcCNGC5*, and *DcCNGC8*, which are in line with the upregulation of these genes in response to drought in the RNA-seq data. MYB-binding sites, and WRKY-binding W-box elements, which are involved in drought and cold stress response, further indicate that *DcCNGC* genes are entrenched in complex regulatory networks that regulate stress adaptation.

The three-dimensional structural models predicted by AlphaFold3 and visualized in ChimeraX represent the first structural characterization of *DcCNGC* proteins and significantly support the domain analysis of the sequences [41]. The preserved six-membrane topology, pore-forming loop and CNBD geometry found in all modeled members validate that *DcCNGC* proteins take the canonical CNGC structural fold [42]. A RMSD of less than 2.0 A between *DcCNGC* members suggest that the level of structural conservation is very high as is expected by the high level of purifying selection previously established by the Ka/Ks analysis. Reported differences in ion selectivity between phylogenetic groups of plant CNGC clades could be due to minor structural differences in the pore-forming loop, a hypothesis that should be directly experimentally tested by electrophysiological characterization. The intuitive solution of the cyclic nucleotide-binding pocket in each of the models facilitates a structural foundation of how the binding of cAMP or cGMP controls channel gating in *DcCNGC* proteins.

The results of the GO and KEGG enrichment offer systems level confirmation to the functional annotations deduced by sequence and structure analysis. The high-level enrichment of potassium ion transmembrane flux, ion channel activity, and voltage-gated ion channel potassium channel activity goes to back the notion that *DcCNGC* proteins are essentially cation-conducting ion channels, as found in the electrophysiological characterizations of *AtCNGC* homologs [43, 44]. It is especially interesting to note the apical plasma membrane enrichment of the Cellular Component category, which implies that *DcCNGC* proteins are specifically targeted to the apical domain of polarized cells, a localization also used in directional tip growth processes like root hair elongation and pollen tube guidance, which is analogous to the role of AtCNGC18 in pollen tube growth [45]. This enrichment of the plant-pathogen interaction KEGG pathway is in line with the known functions of CNGCs in defense-related Ca2 + signaling and justifies the inference based on the W-box and as-1 cis-elements observed in *DcCNGC* promoters that multiple *DcCNGC* members play a role in pathogen defense in carrot [46].

Some of the most biologically interesting results of this study are the expression analysis under drought and AM symbiosis. *DcCNGC2a, DcCNGC5*, and *DcCNGC8* are upregulated during drought which is in line with taking part in drought induced calcium signaling. *AtCNGC2* has been demonstrated to mediate hypersensitive response-related Ca2 + influx and its close homologs *DcCNGC2a-e* could play similar roles in carrot stress adaptation [47]. The mycorrhizal drought treatment-induced augmented expression of these genes indicates that AM colonization conditions or sensitizes the Ca 2 + channel-mediated response to water deficit, which may be mediated by an alteration in the properties of root cell membranes or the participation of fungus-derived signaling molecules to activate CNGC transcription.

The upregulation of *DcCNGC4* paralogs and *DcCNGC17, DcCNGC19* induced by AM symbiosis is in line with the need of Ca2 + oscillations (Ca2 + spiking) to establish arbuscular mycorrhizal symbiosis. Symbiosis-related transcription factors like CYCLOPS/IPD3 are reported to be activated by symbiosis-associated Ca2 + signals produced in the nuclear envelope in legumes and non-legume plants [48]. The channels that mediate these Ca2 + signatures in carrot are not known, and the current evidence indicates that *DcCNGC4* and *DcCNGC17* can be considered as possible players. These roles will have to be confirmed by functional validation using mutant analysis or by heterologous expression.

Together with these findings, the particular *DcCNGC* members are candidate genes to be manipulated biotechnologically to enhance carrot drought resistance. *DcCNGC* gene overexpression, especially with multiple stress-associated cis-elements and transcriptional upregulation in the combination of drought and AM stress, could be a potential remedy to carrot adaptation to water-stressed agricultural conditions. Future functional genomics research, such as CRISPR/Cas9-mediated gene editing and virus-induced gene silencing (VIGS), will be needed to unravel the individual *DcCNGC* member contributions to carrot stress physiology.

Conclusions

This research is the first genome-wide identification and characterization of the CNGC gene family in Daucus carota. The 24 *DcCNGC* genes were found and described in terms of their phylogenetic relationships, physicochemical properties, gene structure, conserved motifs, chromosomal distribution, cis-regulatory elements, and stress-responsive expression profiles. Phylogenetic distribution of the *DcCNGC* genes was done into six groups with gene duplication playing a significant role in family expansion. The Ka/Ks analysis showed that all 32 paralogous gene pairs underwent universal purifying selection, which supports that there is significant functional constraint after duplication. Three-dimensional structures predicted by AlphaFold3 confirmed that *DcCNGC* proteins have a conserved six-transmembrane and CNBD structure, and structural conservation was supported by superimposition with ChimeraX. GO and KEGG enrichment analyses validated functional roles in cation transport, ion channel activity, and plant–pathogen interactionPromoter examination showed a multitude of stress- and hormone-sensitive cis-elements, and transcriptomic examination showed that *DcCNGC* genes were differentially regulated during drought and arbuscular mycorrhizal symbiosis, separately and in interaction with each other. *DcCNGC2a, DcCNGC2b, DcCNGC4a, DcCNGC8, DcCNGC13* and *DcCNGC19* were identified as the key candidates and should be given priority to undergo functional validation as they

are consistently drought-responsive. The work forms a basic genomic tool to further research on Ca2+-mediated stress signaling in carrot and to guide molecular breeding efforts to enhance drought tolerance in this valuable crop.

Abbreviations

AM
Arbuscular Mycorrhizal
ABRE
Abscisic acid-responsive element
BIC
Bayesian Information Criterion
BLAST
Basic Local Alignment Search Tool
CaM
Calmodulin
CaMBD
Calmodulin-binding domain
CDD
Conserved Domain Database
CNBC
Cyclic nucleotide-binding domain
CNGC
Cyclic Nucleotide-Gated Channel
FPKM
Fragments Per Kilobase per Million mapped reads
GRAVY
Grand Average of Hydropathy
GSDS
Gene Structure Display Server
HISAT2
Hierarchical Indexing for Spliced Alignment of Transcripts
ML
Maximum Likelihood
MSA
Multiple Sequence Alignment
MBS
MYB binding site
pI
Isoelectric point
RNA-seq
RNA sequencing
SRA
Sequence Read Archive
TAIR
The Arabidopsis Information Resource.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The RNA-seq data analyzed in this study are publicly available in the NCBI Sequence Read Archive under accession number PRJNA784577. The *D. carota* reference genome is available via Ensembl Plants (https://plants.ensembl.org). All other data supporting the findings are presented within the manuscript.

Competing interests

The authors declare no competing interests.

Funding

This research project was funded by the Deanship of Scientific Research and Libraries,Princess Nourah bint Abdulrahman University, through the Pioneer Researcher Funding Initiative, Grant No (PRFI- 2026)

References

1. Mahmood, T., et al., *Insights into drought stress signaling in plants and the molecular genetic basis of cotton drought tolerance.* Cells, 2019. **9**(1): p. 105.
2. Liu, J., et al., *Genome-Wide Identification of the Cyclic Nucleotide-Gated Ion Channel Gene Family and Expression Profiles Under Low-Temperature Stress in Luffa cylindrica L.* International Journal of Molecular Sciences, 2024. **25**(20): p. 11330.
3. Dietrich, P., W. Moeder, and K. Yoshioka, *Plant cyclic nucleotide-gated channels: new insights on their functions and regulation.* Plant Physiology, 2020. **184**(1): p. 27-38.
4. Zhang, Z., et al., *Genome-wide identification, characterization, and expression analysis related to low-temperature stress of the CmGLP gene family in Cucumis melo L.* International Journal of Molecular Sciences, 2022. **23**(15): p. 8190.
5. Duszyn, M., et al., *Cyclic nucleotide gated channels (CNGCs) in plant signalling—Current knowledge and perspectives.* Journal of Plant Physiology, 2019. **241**: p. 153035.
6. Tunc-Ozdemir, M., et al., *A cyclic nucleotide-gated channel (CNGC16) in pollen is critical for stress tolerance in pollen reproductive development.* Plant physiology, 2013. **161**(2): p. 1010-1020.
7. Saand, M.A., et al., *Phylogeny and evolution of plant cyclic nucleotide-gated ion channel (CNGC) gene family and functional analyses of tomato CNGCs.* DNA research, 2015. **22**(6): p. 471-483.
8. Chen, C., et al., *Kenaf cyclic nucleotide-gated channel gene HcCNGC27 confers plant drought stress tolerance and involved in flowering regulation.* Molecular Genetics and Genomics, 2025. **300**(1): p. 65.
9. Cook, N.M., et al., *Autoactive CNGC15 enhances root endosymbiosis in legume and wheat.* Nature, 2025. **638**(8051): p. 752-759.
10. Yuan, P., et al., *Calcium/calmodulin-mediated microbial symbiotic interactions in plants.* Frontiers in Plant Science, 2022. **13**: p. 984909.
11. Recchia, G.H., et al., *Arbuscular mycorrhizal symbiosis leads to differential regulation of drought-responsive genes in tissue-specific root cells of common bean.* Frontiers in Microbiology, 2018. **9**: p. 1339.
12. Geem, K.R., et al., *Role of carrot (Daucus carota L.) storage roots in drought stress adaptation: Hormonal regulation and metabolite accumulation.* Metabolites, 2025. **15**(1): p. 56.
13. Keller-Pearson, M., et al., *A dual transcriptomic approach reveals contrasting patterns of differential gene expression during drought in arbuscular mycorrhizal fungus and carrot.* Molecular Plant-Microbe Interactions, 2023. **36**(12): p. 821-832.
14. Carrot, W., *Carrot: history and iconography.* Chronica, 2011. **51**(2): p. 13.
15. Reiser, L., et al., *Recent major changes to TAIR: updates to the database, website, and Arabidopsis genome.* Genetics, 2026: p. iyaf248.
16. Bolser, D., et al., *Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomics data*, in *Plant bioinformatics: Methods and protocols*. 2016, Springer. p. 115-140.
17. Madden, T., *The BLAST sequence analysis tool.* The NCBI handbook, 2013. **2**(5): p. 425-436.
18. Letunic, I. and P. Bork, *20 years of the SMART protein domain annotation resource.* Nucleic acids research, 2018. **46**(D1): p. D493-D496.
19. Marchler-Bauer, A., et al., *CDD: NCBI's conserved domain database.* Nucleic acids research, 2015. **43**(D1): p. D222-D226.
20. Kumar, S., et al., *MEGA12: molecular evolutionary genetic analysis version 12 for adaptive and green computing.* Molecular biology and evolution, 2024. **41**(12): p. msae263.
21. Minh, B.Q., et al., *IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era.* Molecular biology and evolution, 2020. **37**(5): p. 1530-1534.
22. Letunic, I. and P. Bork, *Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool.* Nucleic acids research, 2024. **52**(W1): p. W78-W82.
23. Azimi, R., et al., *Bioinformatic analysis of small humanin like peptides using AlfaFold-2 and Expasy ProtParam.* Investigative Ophthalmology & Visual Science, 2024. **65**(7): p. 1320-1320.
24. Horton, P., et al., *WoLF PSORT: protein localization predictor.* Nucleic acids research, 2007. **35**(suppl_2): p. W585-W587.
25. Hu, B., et al., *GSDS 2.0: an upgraded gene feature visualization server.* Bioinformatics, 2015. **31**(8): p. 1296-1297.
26. Bailey, T.L., et al., *The MEME suite.* Nucleic acids research, 2015. **43**(W1): p. W39-W49.
27. Kitts, P.A., et al., *Assembly: a resource for assembled genomes at NCBI.* Nucleic acids research, 2016. **44**(D1): p. D73-D80.
28. Chao, J., et al., *MG2C: A user-friendly online tool for drawing genetic maps.* Molecular horticulture, 2021. **1**(1): p. 16.
29. Lescot, M., et al., *PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences.* Nucleic acids research, 2002. **30**(1): p. 325-327.
30. Chen, C., et al., *TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining.* Molecular plant, 2023. **16**(11): p. 1733-1742.
31. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic acids research, 2004. **32**(5): p. 1792-1797.

32. Krokidis, M.G., et al., *AlphaFold3: an overview of applications and performance insights.* International Journal of Molecular Sciences, 2025. **26**(8): p. 3671.
33. Meng, E.C., et al., *UCSF ChimeraX: Tools for structure building and analysis.* Protein Science, 2023. **32**(11): p. e4792.
34. Ge, S.X., D. Jung, and R. Yao, *ShinyGO: a graphical gene-set enrichment tool for animals and plants.* Bioinformatics, 2020. **36**(8): p. 2628-2629.
35. Oranab, S., et al., *Genome-wide characterization and expression analysis of cyclic nucleotide-gated ion channels (CNGCs) gene family in Arabidopsis thaliana L. and Helianthus annuus L. for drought stress.* Brazilian Journal of Botany, 2024. **47**(3): p. 885-900.
36. Guo, J., et al., *Genome-wide identification of cyclic nucleotide-gated ion channel gene family in wheat and functional analyses of TaCNGC14 and TaCNGC16.* Frontiers in Plant Science, 2018. **9**: p. 18.
37. Liu, W., et al., *Identification and expression analysis of BnaCNGC family gene in the response to phytohormones, abiotic and biotic stresses in Brassica napus.* Journal of Plant Interactions, 2021. **16**(1): p. 575-586.
38. Nawaz, Z., et al., *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, 2014. **15**(1): p. 853.
39. Gao, F., et al., *A heat-activated calcium-permeable channel–Arabidopsis cyclic nucleotide-gated ion channel 6–is involved in heat shock responses.* The Plant Journal, 2012. **70**(6): p. 1056-1069.
40. Meier, I., et al., *Going green: plants' alternative way to position the Ran gradient.* Journal of microscopy, 2008. **231**(2): p. 225-233.
41. Shimizu, T., et al., *The three-dimensional structure prediction of human bitter taste receptor using the method of AlphaFold3.* Current Research in Food Science, 2025: p. 101146.
42. Hamid, H.A., *Structural-Functional Analysis of Plant Cyclic Nucleotide Gated Ion Channels*. 2013.
43. Kaplan, B., T. Sherman, and H. Fromm, *Cyclic nucleotide-gated channels in plants.* FEBS letters, 2007. **581**(12): p. 2237-2246.
44. Hua, B.-G., et al., *Plants do it differently. A new basis for potassium/sodium selectivity in the pore of an ion channel.* Plant Physiology, 2003. **132**(3): p. 1353-1361.
45. Frietsch, S., et al., *A cyclic nucleotide-gated channel is essential for polarized tip growth of pollen.* Proceedings of the National Academy of Sciences, 2007. **104**(36): p. 14531-14536.
46. Sun, L., Y. Wang, and J. Zhang, *Calcium signaling in plant defense.* New Plant Protection, 2026. **3**(1): p. e70028.
47. Clough, S.J., et al., *The Arabidopsis dnd1 "defense, no death" gene encodes a mutated cyclic nucleotide-gated ion channel.* Proceedings of the National Academy of Sciences, 2000. **97**(16): p. 9323-9328.
48. Cooke, A., *The conservation of arbuscular mycorrhizal symbiosis between a liverwort and an angiosperm*. 2019, University of East Anglia.

Figures

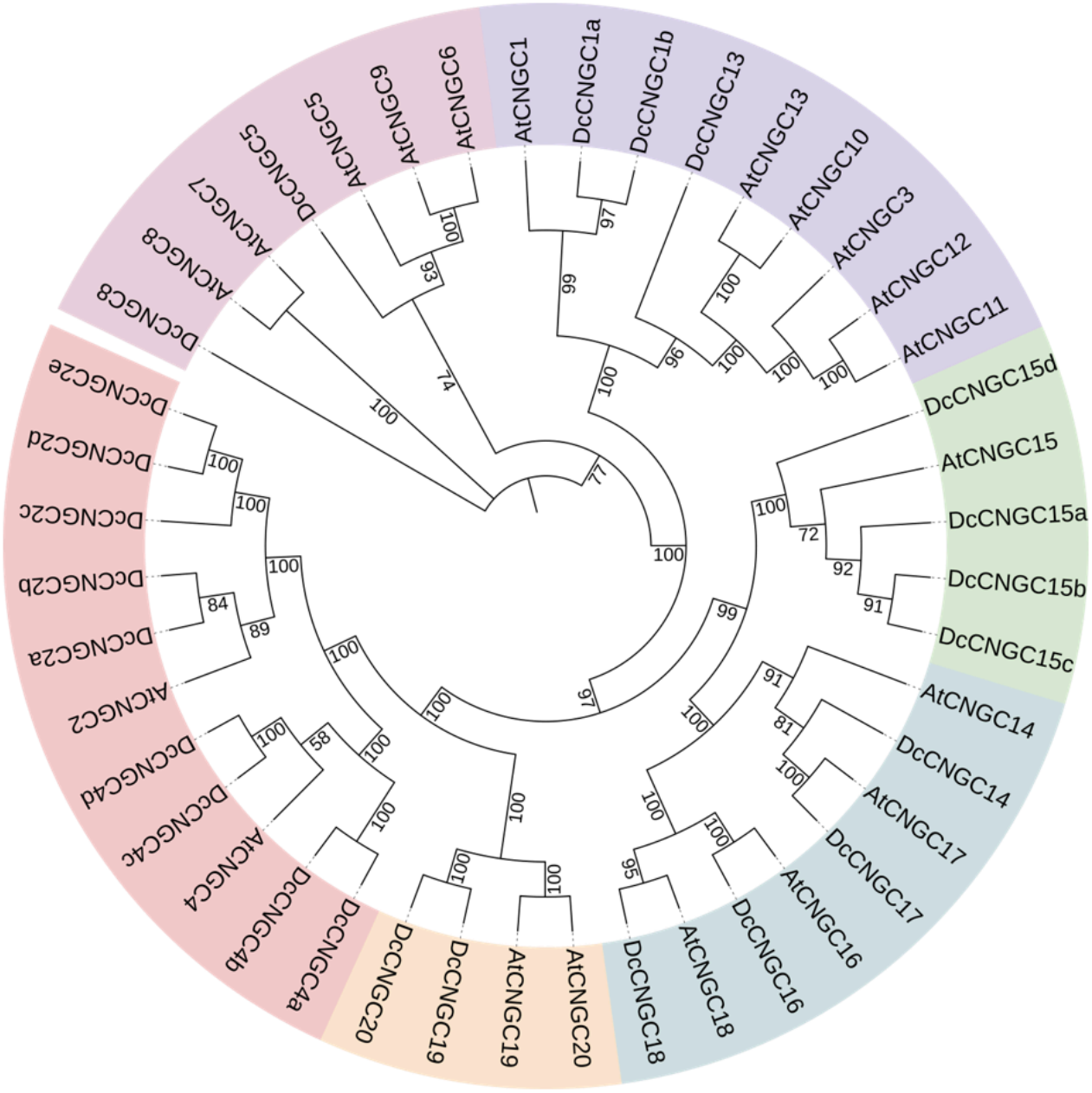

Figure 1

Maximum Likelihood (ML) phylogenetic tree of *D. carota* (*DcCNGC*) and *A. thaliana*(*AtCNGC*) proteins generated with IQ-TREE and visualized with iTOL. The tree is subdivided into six major groups (I–VI) represented in distinct colors. Bootstrap values ≥70 are shown at branch nodes.

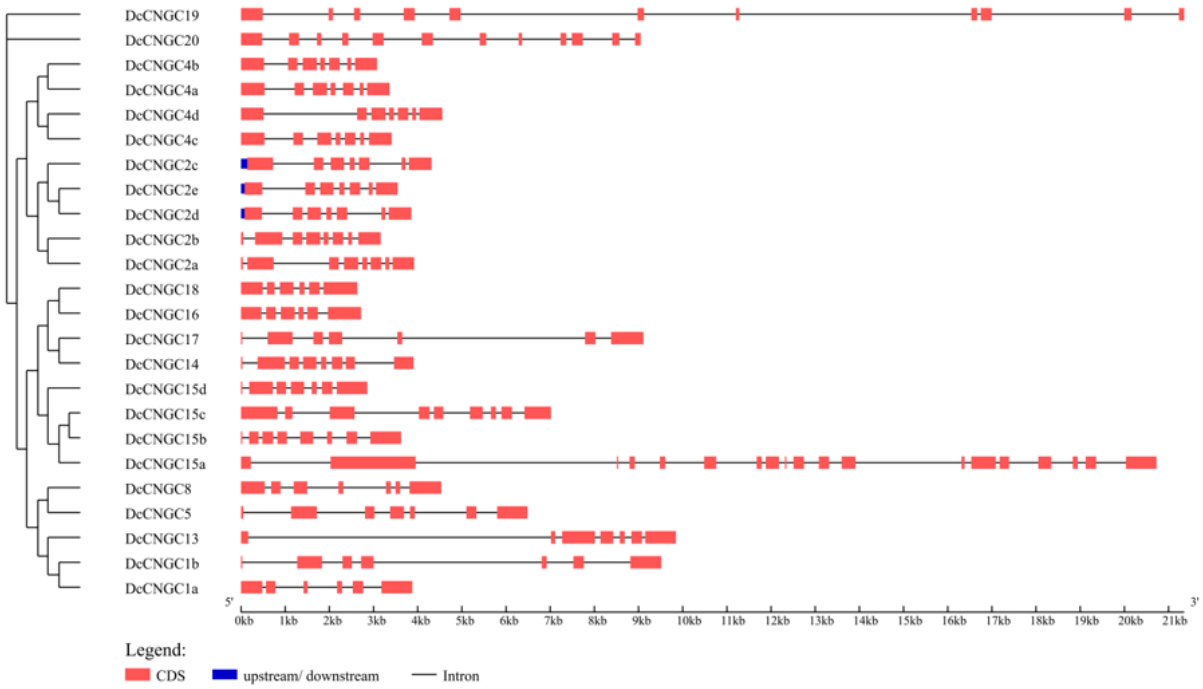

Figure 2

Exon–intron organization of *DcCNGC* genes generated by GSDS. Red boxes represent coding sequences (CDS), blue boxes indicate upstream/downstream regions, and black lines represent introns. Gene structures are displayed from 5' to 3'.

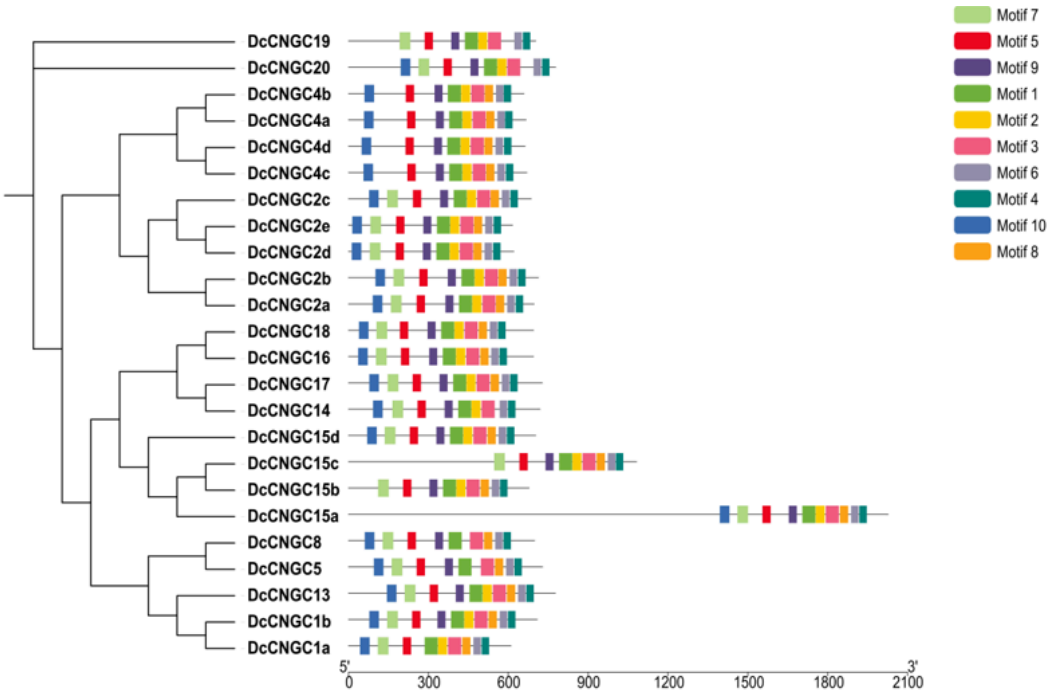

Figure 3

Conserved motifs of *DcCNGC* proteins identified by MEME analysis. Each colored box represents a distinct conserved motif (Motifs 1–10) as indicated in the legend. Motif positions are displayed from 5' to 3'.

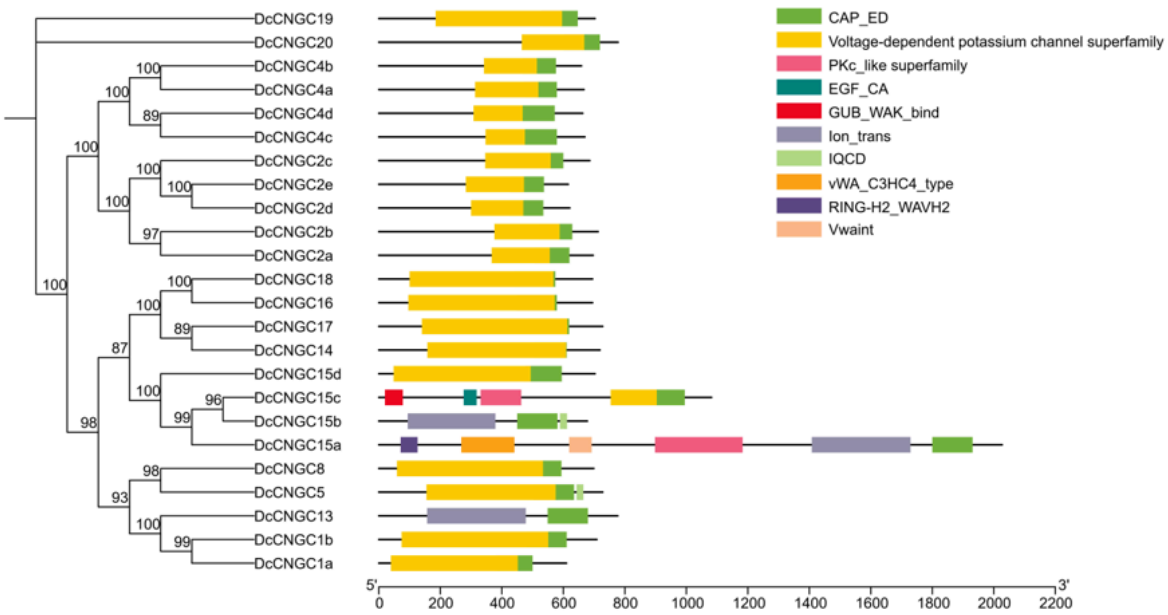

Figure 4

Conserved domain architecture of DcCNGC proteins predicted by Pfam, SMART, and NCBI-CDD databases. The color segments depict separate functional domains as indicated in the legends.

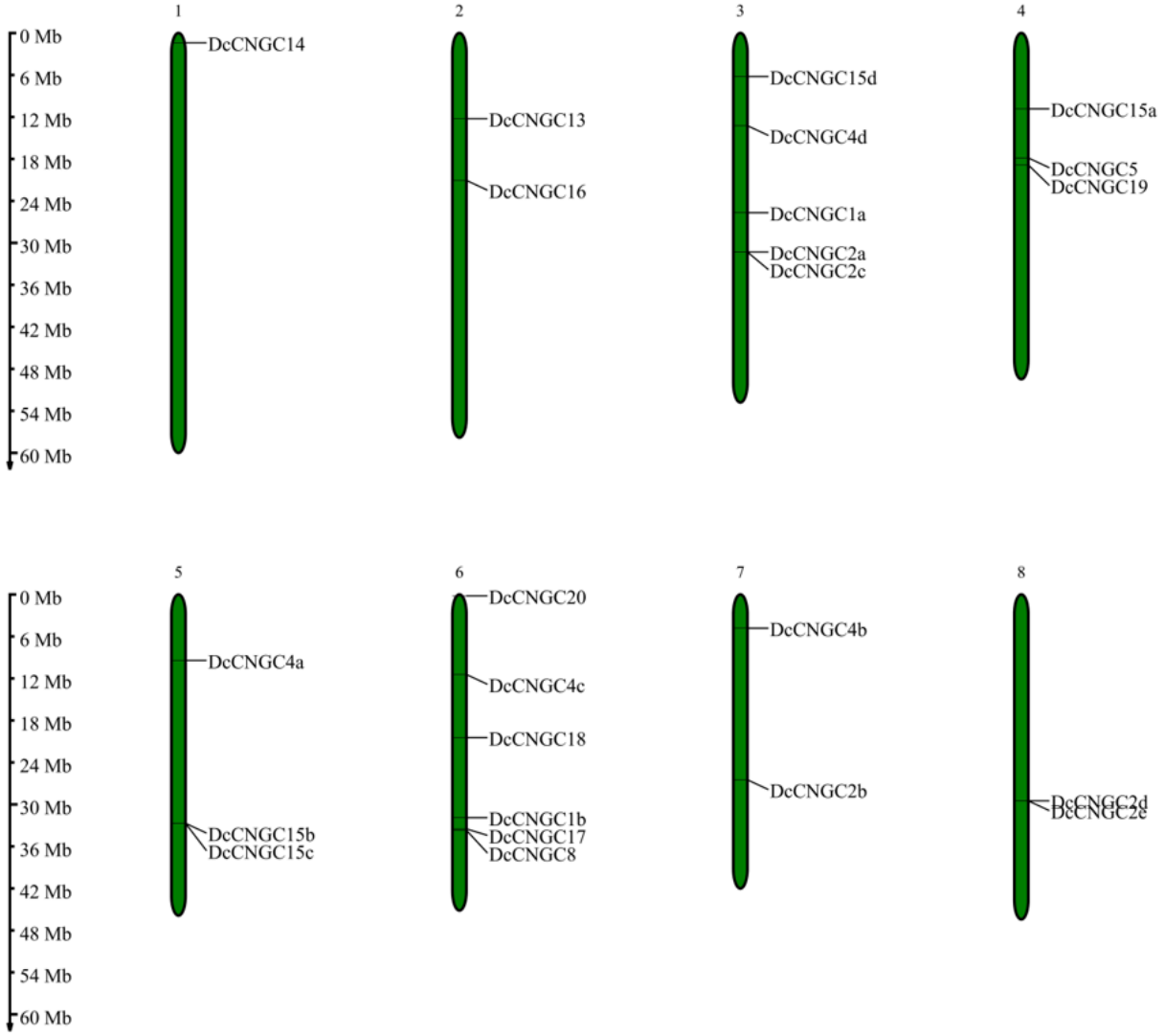

Figure 5

Chromosomal distribution of *DcCNGC* genes across the eight *D. carota*chromosomes, showing their physical positions. Gene names are labeled adjacent to their chromosomal locations.

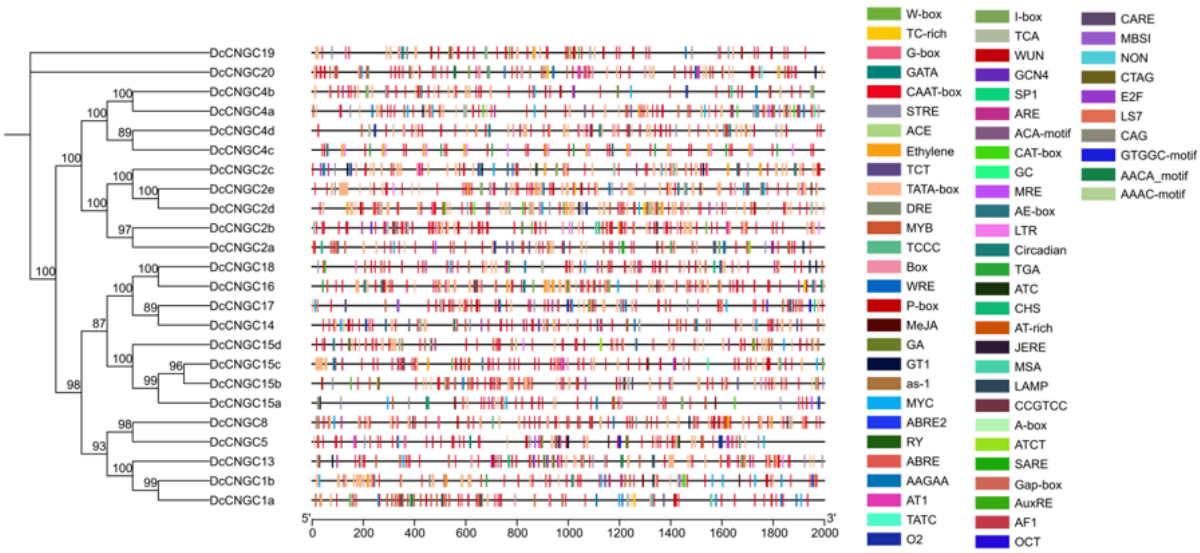

Figure 6

Cis-regulatory elements in the 2 kb promoter regions of *DcCNGC* genes. Elements are grouped by functional category: hormone responsiveness, abiotic stress, light responsiveness, and core promoter elements.

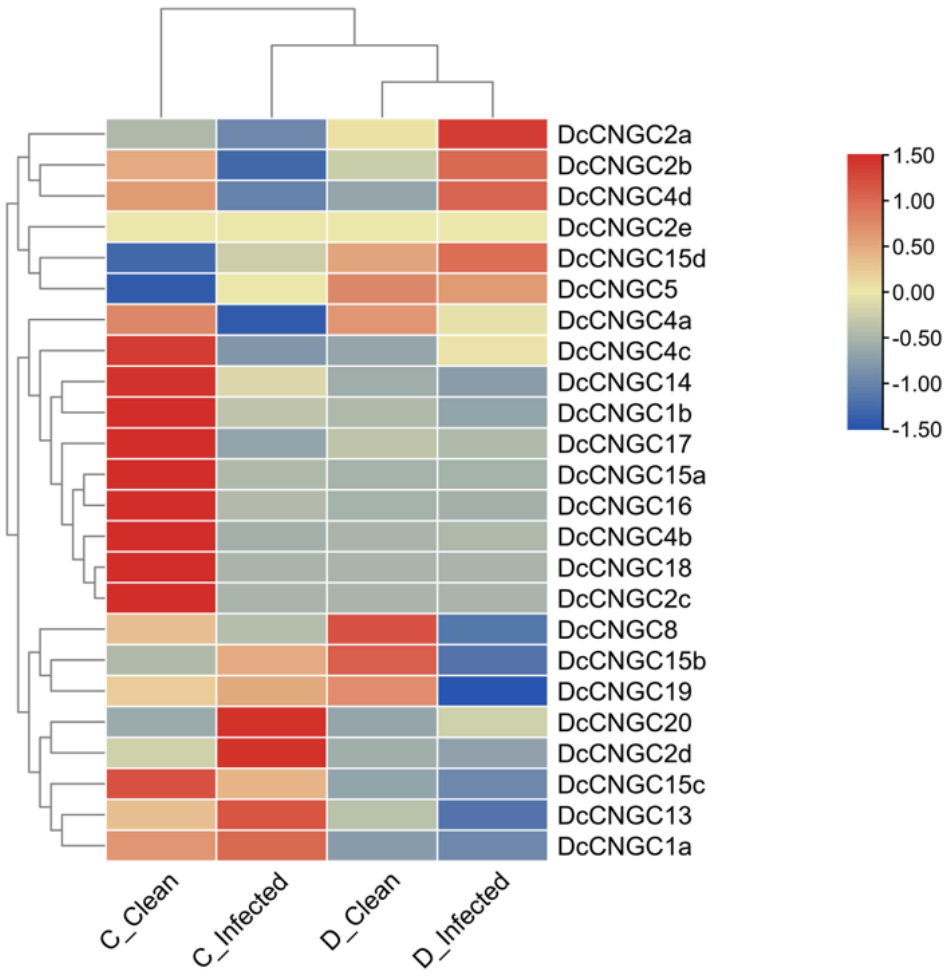

Figure 7

Heatmap showing expression profiles (FPKM) of *DcCNGC* genes across four treatment conditions: non-mycorrhizal control (C_Clean), mycorrhizal control (C_Infected), non-mycorrhizal drought (D_Clean), and mycorrhizal drought (D_Infected). Color scale represents log2-transformed FPKM values; red indicates high expression and blue indicates low expression.

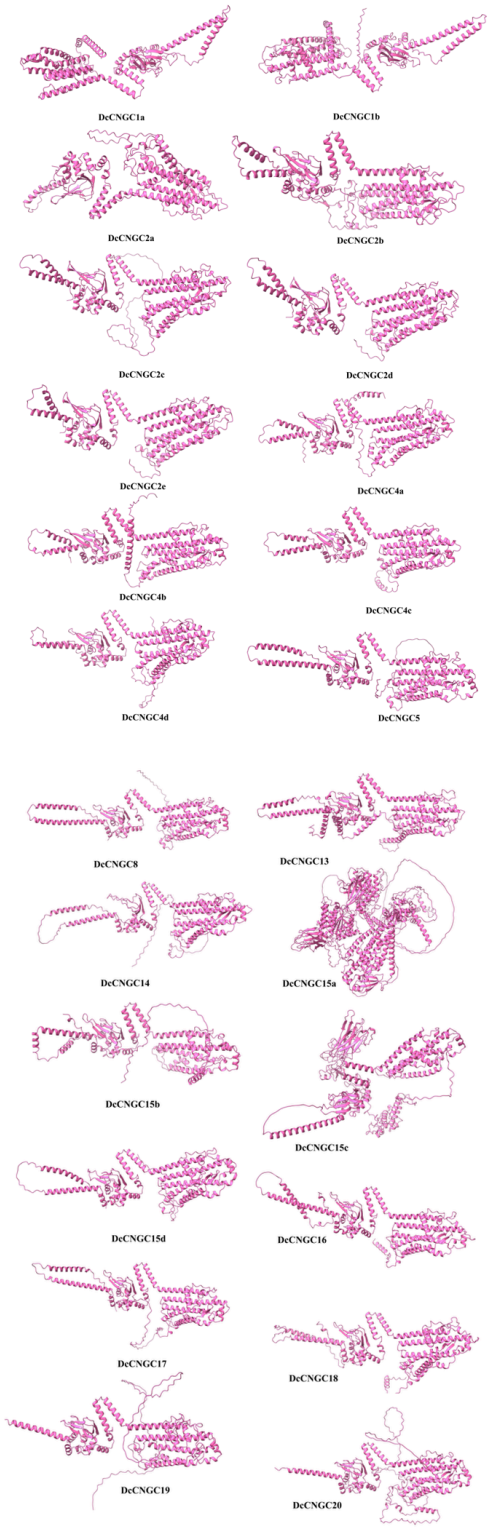

Figure 8

Three-dimensional protein structures of representative DcCNGC proteins predicted using AlphaFold3 and visualized with UCSF ChimeraX. (a) Structural models of DcCNGC1a, DcCNGC1b, DcCNGC2a–e, DcCNGC4a–d, and DcCNGC5. (b) Structural models of DcCNGC8, DcCNGC13–20. Transmembrane helices (S1–S6) are displayed in ribbon representation.

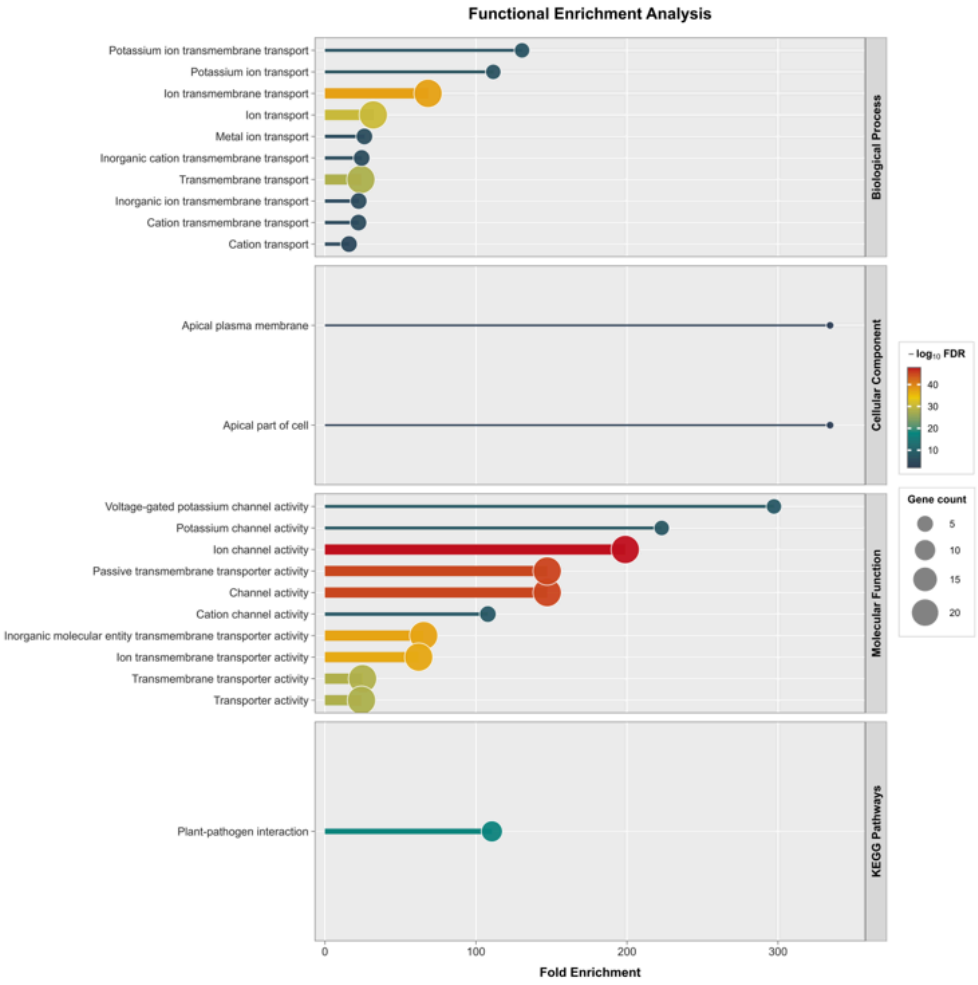

Figure 9

GO and KEGG enrichment analysis of *DcCNGC* genes. Lollipop plots show enriched terms across Biological Process (BP), Cellular Component (CC), Molecular Function (MF), and KEGG Pathway categories. The x-axis represents fold enrichment; dot size indicates gene count; dot color represents -log10(FDR), with warmer colors (red/yellow) indicating higher statistical significance.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.CisregulatoryelementsidentifiedinthepromoterregionsofDcCNGCgenes.xlsx
- CDS.txt
- Genomic.txt
- protein.txt