
Article

Not peer-reviewed version

Genome-Wide Characterization of VDAC Gene Family in Soybean (*Glycine max* L.) and in-silico Expression Profiling in Response to Drought and Salt Stresses

[Muhammad Muneeb Ullah](#) ^{*}, [Muhammad Mudassar Iqbal](#), [Muqadas Aleem](#) ^{*}, Awais Riaz Riaz, [Ainong SHI](#)

Posted Date: 2 June 2025

doi: 10.20944/preprints202506.0068.v1

Keywords: soybean; genome-wide; VDAC; drought; salt stress



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Genome-Wide Characterization of VDAC Gene Family in Soybean (*Glycine max* L.) and *in-silico* Expression Profiling in Response to Drought and Salt Stresses

Muhammad Muneeb Ullah ^{1,*}, Muhammad Mudassar Iqbal ², Awais Riaz ³, Muqadas Aleem ^{2,*} and Ainong Shi ¹

¹ Department of Horticulture, 316 PTSC, University of Arkansas. Fayetteville 72701, AR, USA; mu007@uark.edu (M.M.U.); ashi@uark.edu (A.S.)

² Department of Plant breeding and Genetics, University of Agriculture, Faisalabad, 38000, Pakistan; mudassariq5@gmail.com (M.M.I.); muqadas.aleem@uaf.edu.pk (M.A.)

³ Department of Crop, Soil and Environmental Sciences, 316 PTSC, University of Arkansas. Fayetteville 72701, AR, USA; ariaz@uark.edu (A.R.)

* Correspondence: mu007@uark.edu (M.M.U.); muqadas.aleem@uaf.edu.pk (M.A.)

Abstract: Soybean (*Glycine max* L.) is grown worldwide to obtain edible oil, livestock feed, and biodiesel. However, drought and salt stresses are becoming serious challenges to global soybean cultivation as they retard the growth of soybean plants and cause significant yield losses. Voltage dependent anion-selective channel (VDAC) proteins are well-known for their role in drought and salt tolerance in crop plants. In this study, we identified 111 putative VDAC genes randomly distributed in genomes of 14 plant species, including cultivated soybean (*Glycine max*) and wild soybean (*Glycine soja*). The comparative phylogenetic studies classified these genes into six different clades and found the highest structural similarities among VDAC genes of *G. max* and *G. soja*. Porin-3 was the conserved domain, and seven motifs were also conserved in all VDAC proteins. Furthermore, *GmaVDAC* proteins were found in outer mitochondrial membrane and gene annotation studies revealed their role in voltage gated anion channel activity. These proteins were also found to interact with other proteins especially the mitochondrial receptors. A total of 103 miRNAs were predicted to target fifteen *GmaVDAC* genes. In *G. max*, these genes were found segmentally duplicated and randomly distributed on twelve chromosomes. Transcriptomic analysis revealed that *GmaVDAC18.2* gene showed overexpression in root nodules whereas, *GmaVDAC9.1*, *GmaVDAC18.1*, and *GmaVDAC18.2* genes showed overexpression under drought and salt stress conditions.

Keywords: soybean; genome-wide; VDAC; drought; salt stress

1. Introduction

Soybean (*Glycine max* L.) is a widely grown oilseed crop which accounts for 58% global edible oil production and satisfies approximately 69% of demand for protein meal [1]. Moreover, the rhizobia inside the root nodules of soybean plants can fix atmospheric nitrogen and thus improve the nitrogen content of the soil. However, the drought and salt stresses due to climate change are adversely affecting the growth and development of soybean [2]. These conditions negatively impact various physiological processes and create an ionic imbalance in the cell, thus exacerbating the production of reactive oxygen species [3], which in turn pose oxidative stress. Ultimately, this leads to low soybean yield, poor oil quality, and decreased nodulation. Thus, the cell tends to maintain its internal homeostatic state by regulating ionic balance and scavenging excessive ROS [4].

Voltage dependent anion-selective channel (VDAC) genes have been reported to regulate drought and salt stresses in plants. VDAC proteins are major mitochondrial proteins (also called

mitochondrial porins) found in all eukaryotic organisms, namely plants, animals, and fungi [5]. These cover about thirty percent of the surface area of the outer membrane of mitochondria and give a sieve-like structure to it [6]. These are highly conserved transmembrane β -barrel proteins involved in maintaining cellular homeostasis by regulating metabolite transport, especially chloride ions, between cytoplasm and mitochondria in drought and salt stress conditions. In addition, these proteins also regulate the communication between cytoplasm and mitochondria by interacting with other proteins [3]. VDAC proteins are also involved in regulating the programmed cell death (PCD) that results from mitochondrial-mediated apoptosis [7].

Some studies have reported the evolutionary history, structural and functional characterization, and role in abiotic stress tolerance of VDAC gene family in different crops such as *Arabidopsis* [8], beans [9], *Medicago* and tobacco [10], pearl millet [11], rice [12], and wheat [4] etc. However, to date, the role of VDAC gene family under drought and salt stresses has remained unknown in soybean. Therefore, this study aims at structural characterization and evolutionary relationships of VDAC genes in *Glycine max* L. along with their functional characterization in response to drought and salt stresses based on the RNA-seq data available from previous studies.

2. Results

2.1. Identification, distribution, and renaming of VDAC genes

We identified 111 non-redundant VDAC genes randomly distributed in the genomes of 14 plant species. Fifteen VDAC genes were found in both *G. max* and *G. soja*, whereas five VDAC genes were found in *A. thaliana* (S. Table 1). For convenience, the gene IDs were renamed in such a way that the first three letters indicated the abbreviated botanical name of each crop, followed by the abbreviated name of the gene family. Of the last two numbers, one before the placeholder represented the n^{th} chromosome, whereas the other represented the n^{th} VDAC gene on that chromosome [2]. For example, in our case, GmaVDAC13.2 showed the 2nd VDAC gene lies on the 13th chromosome in the *G. max* genome. We did not consider any sequence of less than 150 bp, encoding a truncated protein, and a scaffold [13].

2.2. Conserved domain, localization, and physico-chemical properties of VDAC proteins

We found PF01459 as the Hidden Markov Model (HMM) indicating a conserved domain (Porin -3) of VDAC proteins. All the peptide sequences devoid of this domain and containing an incomplete domain were not selected for phylogenetic analysis. The sub-cellular localization analysis revealed the presence of all VDAC proteins in mitochondria of the cell.

The physico-chemical properties of GmaVDAC and GsoVDAC include peptide length, molecular weight, isoelectric point, and GRAVY (Grand average of hydropathy) score (S. Table 1). The number of amino acids varied from 245 (GmaVDAC5.1 and GmaVDAC17.2) to 333 (GmaVDAC14.1) in GmaVDAC proteins, whereas this value varied from 245 (GsoVDAC5.1) to 277 (GsoVDAC13.1, GsoVDAC14.1, and GsoVDAC17.1) in GsoVDAC proteins. The molecular weights (kDa) ranged from 26545 (GmaVDAC5.1) to 35938 (GmaVDAC14.1) in GmaVDAC proteins, whereas in GsoVDAC proteins, it ranged from 26545 (GsoVDAC5.1) to 29961 (GsoVDAC17.1). Similarly, the isoelectric point varied from 5.96 (GmaVDAC5.1) to 9.43 (GmaVDAC14.1), and the GRAVY score varied from -0.241 (GmaVDAC13.1) to 0.024 (GmaVDAC17.2) in GmaVDAC proteins. However, in GsoVDAC proteins, the values were found to range from 5.96 (GsoVDAC5.1) to 9.29 (GsoVDAC11.1) for isoelectric point, and -0.24 (GsoVDAC13.1) to 0.024 (GsoVDAC17.2) for GRAVY score. A hundred percent similarity was found between GmaVDAC5.1 and GsoVDAC5.1 proteins for peptide length, molecular weight, pI, and GRAVY score.

2.3. Classification and phylogenetic analysis

The topology of the comparative phylogenetic tree of 111 VDAC genes indicated the distribution of VDAC genes of the concerned species into six different clades (Figure 1). We found 13 genes and one pair of orthologs (*GmaVDAC14.1/GsoVDAC14.1*) in clade I, 16 genes and two pairs of orthologs (*GmaVDAC11.1/GsoVDAC11.1* and *GmaVDAC1.1/GsoVDAC1.1*) in clade II, 16 genes and two pairs of orthologs (*GmaVDAC17.1/GsoVDAC17.1* and *GmaVDAC13.2/GsoVDAC13.2*) in clade III, 23 genes and four pairs of orthologs (*GmaVDAC6.1/GsoVDAC6.1*, *GmaVDAC4.1/GsoVDAC4.1*, *GmaVDAC17.2/GsoVDAC17.2*, and *GmaVDAC5.1/GsoVDAC5.1*) in clade IV, 15 genes and two ortholog pairs (*GmaVDAC18.1/GsoVDAC8.1* and *GmaVDAC8.1/GsoVDAC8.1*) in clade V, and 28 genes and four ortholog pairs (*GmaVDAC19.1/GsoVDAC19.1*, *GmaVDAC13.1/GsoVDAC13.1*, *GmaVDAC18.2/GsoVDAC18.2*, and *GmaVDAC9.1/GsoVDAC9.1*) in clade VI. In all clades, orthologous gene pairs in *G. max* and *G. soja* were found to be distributed on the same chromosome of the respective crop species. A tight clustering between *GmaVDAC* and *GsoVDAC* genes shows that they are more closely related to each other as compared to VDAC genes of any other species.

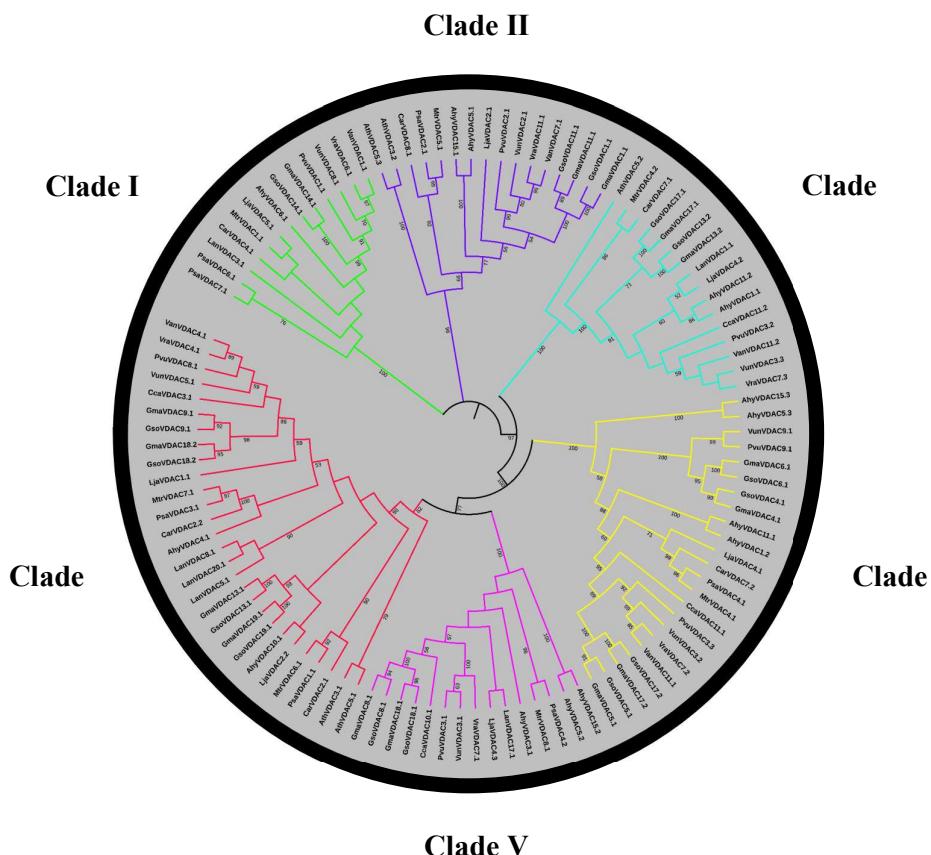


Figure 1. Comparative phylogenetic tree of VDAC proteins constructed with MEGA 11 showing the evolutionary relationships.

2.4. Conserved motif and gene structure analysis

The presence of at least ten conserved motifs was searched out in *GmaVDAC*, *GsoVDAC*, and *MtrVDAC* proteins. Seven (1, 2, 3, 4, 5, 7, and 8) out of ten motifs were found highly conserved as they were present in all VDAC proteins selected for motif analysis (Figure 2).

The diversity in gene structure of *G. max*, *G. soja*, and *M. truncatula* was explored by drawing exon-intron structure along with a phylogenetic tree, which revealed a small variation in number of

exons (5-6) of all VDAC genes (Figure 3). The genes of clade I possessed five exons whereas genes of other five clades consisted of six exons in each. However, it was found that the length and position of introns and exons were more similar within a clade as compared to that among the clades. This shows the conservation of gene structures in each clade.

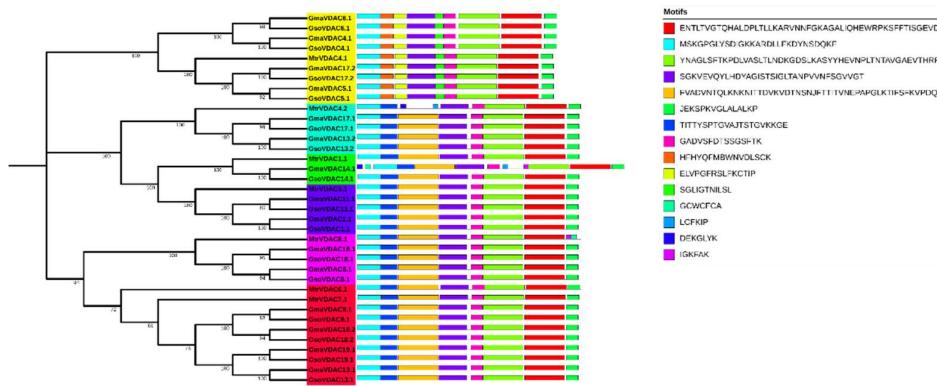


Figure 2. Phylogenetic tree and conserved motifs of VDAC proteins in *G. max*, *G. soja* and *M. truncatula*. (A) There are six clades of phylogenetic tree that are shown with six different colors (Clade I = Yellow, Clade II = Arctic blue, Clade III = Green, Clade IV = Dark blue, Clade V = Pink and Clade VI = Red). (B) Conserved motifs of VDAC proteins shown in different colors. (C) The amino acid sequence of each motif.

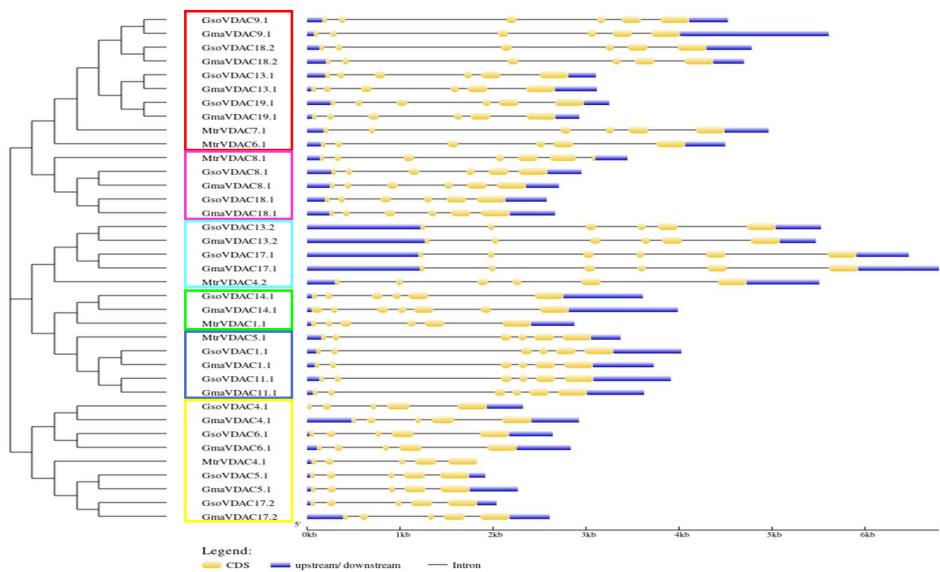


Figure 3. VDAC genes length, no. of exons and introns and their order (A) in *G. max*, *G. soja* and *M. truncatula* species along with phylogenetic tree (B) classified into six clades each is shown with a specific color (Clade I = Yellow, Clade II = Arctic blue, Clade III = Green, Clade IV = Dark blue, Clade V = Pink and Clade VI = Red).

2.5. Identification of *cis*-acting regulatory elements (CREs) in promoters of *Gma*VDAC genes

CREs are specific motifs involved in regulating the transcription of genes by binding to certain transcription factors [5]. The sequence of 2 kb upstream from promoter region of each *Gma*VDAC gene was analyzed for identification of putative CREs. A total of 74 different *cis*-acting elements were found randomly distributed in promoters of all *Gma*VDAC genes. Based on their functions, these were divided into six categories relating to abiotic stress (17 elements), biotic stress (seven

elements), growth and development (10 elements), phytohormones (13 elements), light responsiveness (20 elements), transcription (six elements). A single element, i.e., CTAG-motif, could not be placed in any category due to its unknown function.

The most commonly found CREs in our studies included MYB (water responsive), MYC (drought responsive), ARE (anoxic responsive), WUN-motif (wound and pathogen responsive), ERE (ethylene responsive), ABRE (Abscisic acid responsive), Box 4 (light responsive), and G-box (light responsive). Similarly, AT~TATA-box element (involved in transcription regulation) was found in all genes except *GmaVDAC13.2*. Two *cis*-elements viz. CAAT-box (promoter or enhancer region) and TATA-box (promoter or enhancer region) were the most abundantly present in all *GmaVDAC* genes. Keeping in view the major functions of *GmaVDAC* *cis*-acting elements, it can be predicted that the VDAC gene family plays a critical role in regulating abiotic stress in *G. max*. For better understanding, a graph (Figure 4) was drawn among 17 elements that were more than ten in number in all *GmaVDAC* genes.

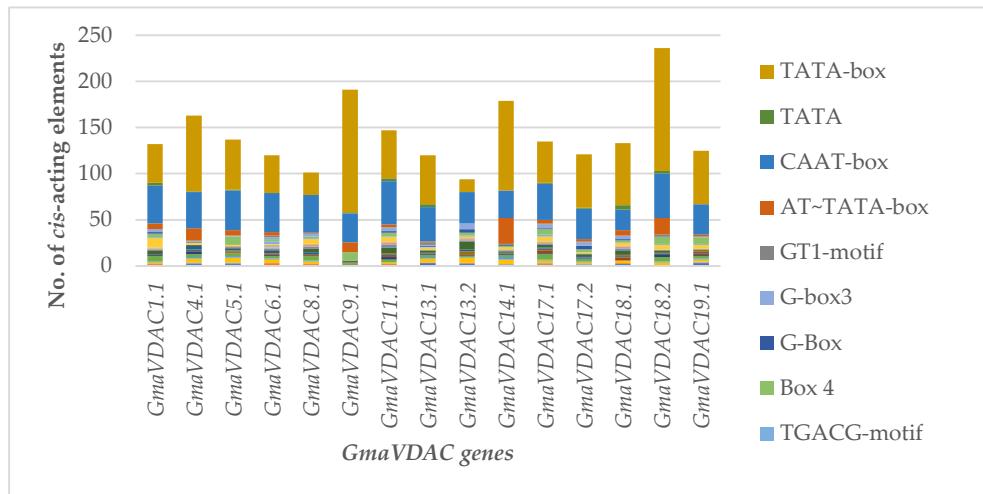


Figure 4. Cis-regulatory elements identified in promoters of 15 *GmaVDAC* genes. Transcription-related elements (TATA-box, TATA, CAAT-box, and AT~TATA-box), light responsive elements (GT1-motif, G-box 3, G-box, Box 4), hormone responsive elements (TGACG-motif, ERE, CGTCA-motif, ABRE 4, ABRE 3a, and ABRE), and growth and development-related elements (MYB-like sequence, Myb-binding site, and AAGAA-motif)

2.6. Gene duplication events and divergence rate

The *G. max* genome has experienced two events of whole-genome duplication (WGD) and one event of whole-genome triplication (WGT), which has led to the presence of numerous duplicated gene copies. Approximately 75% of soybean genome possesses multiple copies of genes [14]. Therefore, the genome duplication has resulted in the accumulation of triplet, tetrad or paired homologous regions, leading to higher sequence similarity. The gene duplication analysis of 15 *GmaVDAC* genes was conducted to explore the diversity among gene sequences. The results of analysis shows that expansion of VDAC gene family in soybean has been primarily driven by segmental duplications. In this study, seven segmental duplications were identified: *GmaVDAC1.1/GmaVDAC11.1*, *GmaVDAC4.1/GmaVDAC6.1*, *GmaVDAC5.1/GmaVDAC17.2*, *GmaVDAC8.1/GmaVDAC18.1*, *GmaVDAC9.1/GmaVDAC18.2*, *GmaVDAC13.1/GmaVDAC19.1*, and *GmaVDAC13.2/GmaVDAC17.1*.

The parameters such as Ka, Ks, and Ka/Ks values are used to assess the type of selection pressure that has influenced the codons. Generally, a Ka/Ks ratio greater than 1 signifies positive or Darwinian selection, while a Ka/Ks ratio less than 1 indicates negative, purifying, or stabilizing selection. However, a Ka/Ks ratio equal to 1 suggests neutral selection. With Ka/Ks ratio ranging from 0.06 to 0.23, purifying selection was observed in all seven duplication events. Furthermore, the approximate

divergence time of gene duplication events was calculated, revealing that these events likely occurred between 5.16 and 12.51 million years ago (MYA). The complete list of *GmaVDAC* duplicates along with Ka, Ks, Ka/Ks values, estimated time of divergence, duplicate and selection type are provided in table 1.

Table 1. List of *GmaVDAC* duplicates including their type, Ka, Ks and Ka/Ks values, time of divergence and selection type.

Duplicated pair	Ka	Ks	Ka/Ks	Duplicate type	Selection type	T (Mya)
<i>GmaVDAC1.1/GmaVDAC11.1</i>	0.01	0.07	0.18	Segmental	Purifying	6.10
<i>GmaVDAC4.1/GmaVDAC6.1</i>	0.03	0.14	0.21	Segmental	Purifying	11.39
<i>GmaVDAC5.1/GmaVDAC17.2</i>	0.02	0.09	0.23	Segmental	Purifying	7.02
<i>GmaVDAC8.1/GmaVDAC18.1</i>	0.01	0.14	0.06	Segmental	Purifying	11.31
<i>GmaVDAC9.1/GmaVDAC18.2</i>	0.005	0.06	0.08	Segmental	Purifying	5.16
<i>GmaVDAC13.1/GmaVDAC19.1</i>	0.02	0.15	0.12	Segmental	Purifying	12.51
<i>GmaVDAC13.2/GmaVDAC17.1</i>	0.02	0.11	0.18	Segmental	Purifying	8.75

2.7. Gene positions and chromosomal chart

The localization pattern of VDAC genes in *Glycine max* was examined and it was found that *GmaVDAC* genes were unevenly distributed among chromosomes. These genes were found on 12 of the 20 pairs of chromosomes (Figure 5). The chromosomes 1, 4, 5, 6, 8, 9, 11, 14, and 19 each carried one VDAC gene, while chromosomes 13, 17, and 18 contained two VDAC genes. The remaining chromosomes did not possess any VDAC gene.

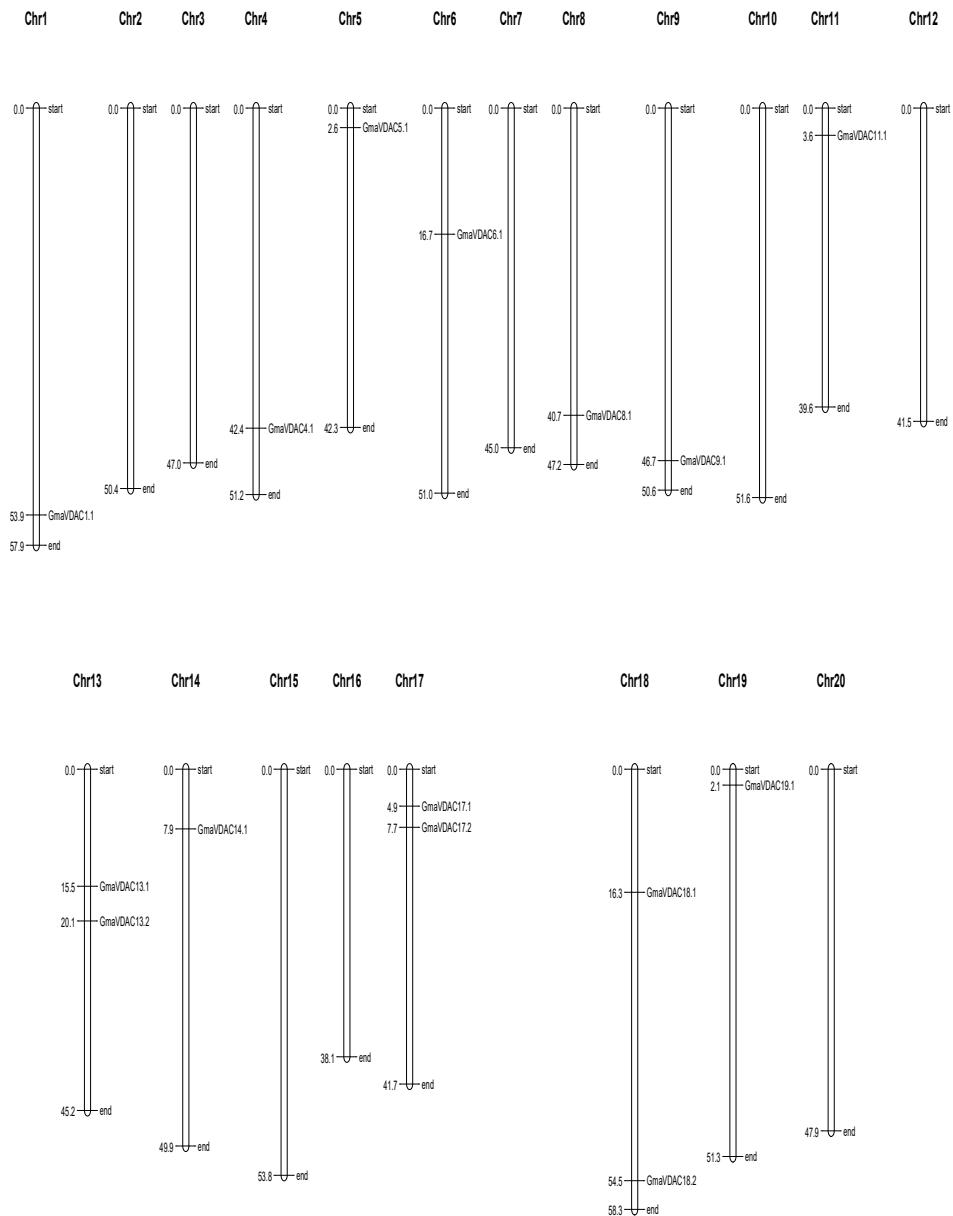


Figure 5. Locations of *GmaVDAC* genes on chromosomes. The positions are given in megabases on left side while the corresponding gene names are given on right side.

2.8. Synteny analysis

To identify orthologous genes, a dual synteny plot was constructed between cultivated soybean (*G. max*) and wild soybean (*G. soja*). We found 14 homologous VDAC gene pairs for 15 *GmaVDAC* genes and there was a single VDAC gene on chromosome 13 of *G. max* for which no homolog was identified in *G. soja*. In addition, it was found that all homologs in two species were present on corresponding chromosomes (Figure 6). This shows that they remain highly conserved during the evolutionary process.

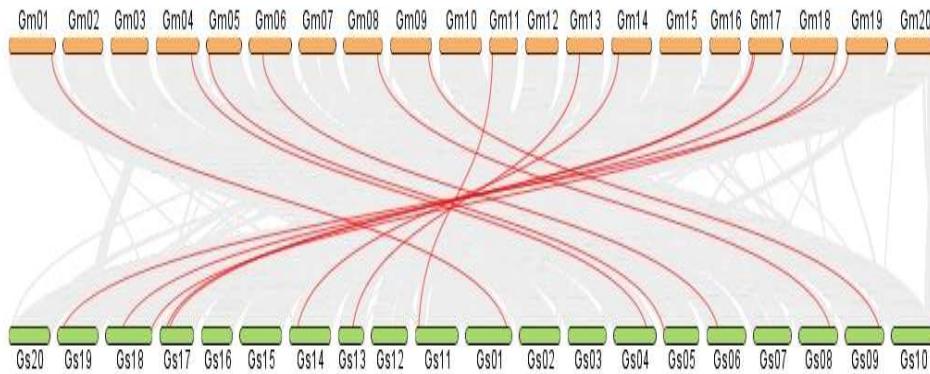


Figure 6. Syntenic plot made between chromosomes of *G. max* (brown color) and *G. soja* (green color) showing orthologous VDAC genes in red lines.

2.9. Expression pattern analysis of *GmaVDAC* genes

The normalized RNA expression data of *GmaVDAC* genes in fourteen plant parts was retrieved from SoyBase and \log_2 transformed. The expression pattern of *GmaVDAC* genes highly varied among fourteen parts (Figure 7a). Four genes viz. *GmaVDAC4.1*, *GmaVDAC5.1*, *GmaVDAC6.1*, and *GmaVDAC17.2* showed the lowest expression in all studied plant parts. However, the genes *GmaVDAC8.1*, *GmaVDAC18.1*, *GmaVDAC9.1*, *GmaVDAC18.2*, and *GmaVDAC19.1* expressed relatively higher in all tissues. Besides, it was found that *GmaVDAC18.2* gene showed the highest expression in nodules of plant.

Similarly, the expression profiles of *GmaVDAC* genes under salt and drought stress conditions were also analyzed (Figure 7b and 7c) respectively. After one hour of salt and drought stress exposure, the genes *GmaVDAC1.1*, *GmaVDAC8.1*, *GmaVDAC9.1*, *GmaVDAC11.1*, *GmaVDAC13.1*, *GmaVDAC13.2*, *GmaVDAC14.1*, *GmaVDAC17.1*, *GmaVDAC18.2*, and *GmaVDAC19.1* were found upregulated. The expression of four genes viz. *GmaVDAC4.1*, *GmaVDAC5.1*, *GmaVDAC6.1*, and *GmaVDAC17.2* was found to be negligible as compared to other VDAC genes under both conditions. After 6 and 12 hours of salt and drought stresses, it was noticed that most of genes downregulated except *GmaVDAC18.1*, *GmaVDAC18.2*, and *GmaVDAC9.1* that expressed upregulation.

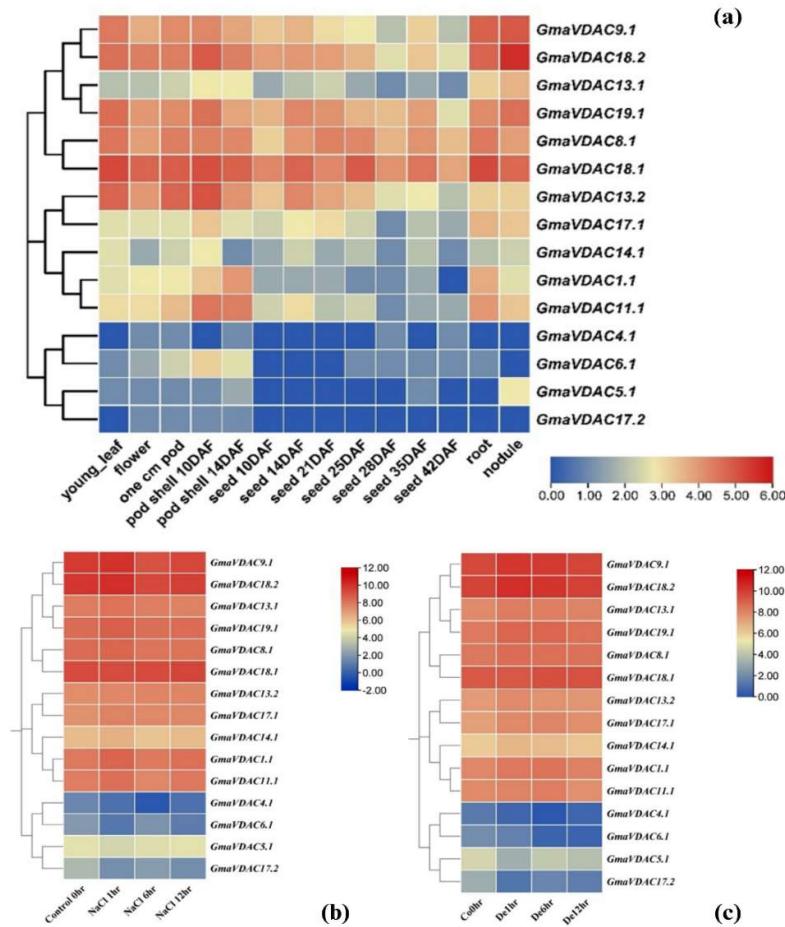


Figure 7: Expression pattern of *GmaVDAC* genes in different tissues (a), under salt stress (b), and drought stress (c)

2.10. Gene ontology analysis

Research in functional genomics relies on the annotation of transcriptomic sequences, and one of the key methods for this annotation is gene ontology (GO) analysis. In our particular study, we performed GO analysis of VDAC gene family in *G. max*, which provided valuable insights into the functional characteristics and roles of *GmaVDAC* genes. Four GO terms obtained were as follows: (i) Biological processes including, inorganic anion transport (GO:0015698) and mono-atomic anion transmembrane transport (GO:0098656); (ii) molecular function of voltage gated anion channel activity (GO:0008308) however, a closely similar molecular function annotation porin activity (GO:0015288) was also obtained from STRING database; and (iii) cellular localization in mitochondrial outer membrane (GO:0005741). Using Blast2GO, a chart showing molecular functions (Figure 8).

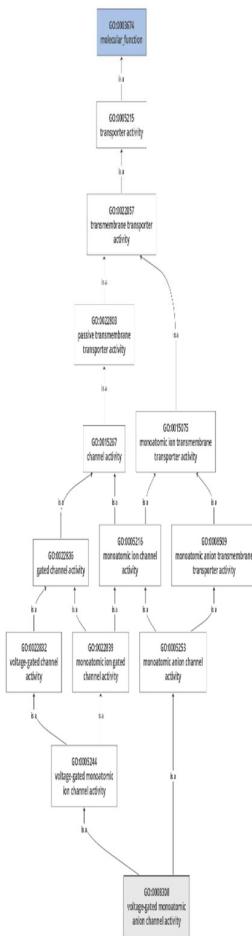


Figure 8. Gene annotations showing molecular function of *GmaVDAC* genes.

2.11. Protein interactions network and microRNA targets prediction

The gene ontology predicted the role of VDAC proteins in porin activity (GO:0015288). The string database was explored to determine interactions among proteins involved in porin activity. The search was performed with the query 'GO:0015288'. The analysis was conducted by setting medium confidence level (0.400) and none of the interactors involved in 1st and 2nd subshell. The analysis revealed 32 interacting proteins along with 52 edges or interactions among them (Figure 9). Interestingly, no interaction was seen among members of *GmaVDAC* protein family however, we found an interaction between each *GmaVDAC* protein and two mitochondrial import receptors (RSVD1 like and Glyma08G05280) that may indirectly activate VDAC proteins during imbalance of ions. Besides, three interacting proteins i.e., Glyma9G18920.1, Glyma6G01780.1, and Glyma14G08240.1 were found outside the network indicating no interaction with *GmaVDAC* proteins. But these are also involved in regulating porin activity. These interactions suggest that VDAC proteins play an important role in regulating the traffic of substances especially anions across mitochondrial membrane.

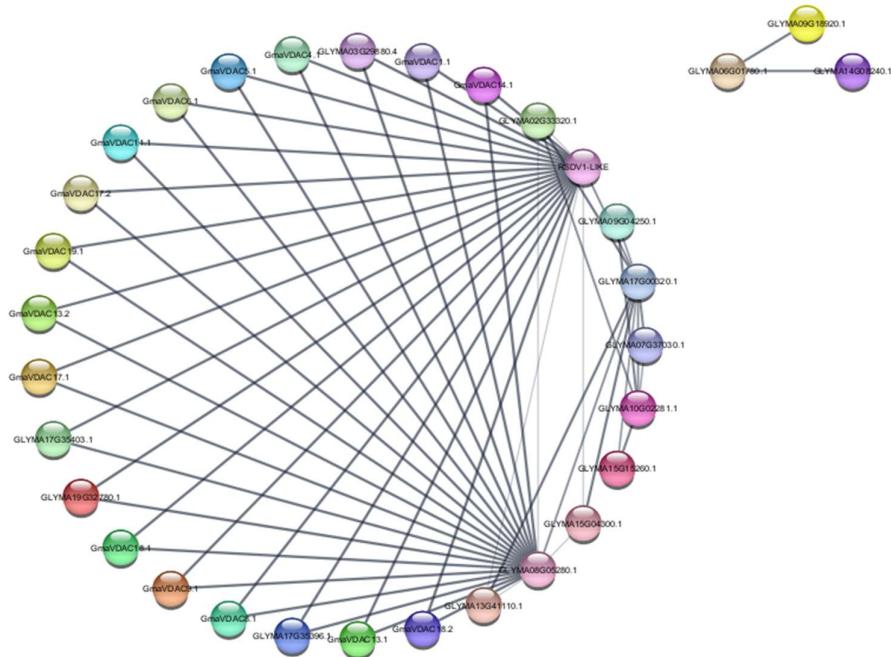


Figure 9. Interaction network of proteins involved in porin activity in *G. max* (The thickness of lines indicates the strength of interaction among them).

MicroRNAs (miRNAs) play a crucial role in regulating gene expression by either degrading mRNA or inhibiting protein translation [15]. However, the transcriptional regulation of miRNA genes remains largely unknown. Therefore, we conducted a systematic analysis to explore potential miRNAs targeting the soybean VDAC gene family. This analysis aimed to provide insights into the regulatory mechanism of *GmaVDAC* genes. Our findings revealed 103 miRNAs targeting 15 *GmaVDAC* genes. Each gene was associated with 2 to 16 miRNAs, and a relationship network was constructed (Figure 10). Noticeably, *GmaVDAC13.1* and *GmaVDAC19.1* were the most targeted genes in soybean, with 16 and 14 predicted miRNA targets, respectively. Additionally, *GmaVDAC1.1* was targeted by a two miRNA (gma-miR5672 and gma-miR9760). Furthermore, we observed that two genes (*GmaVDAC4.1* and *GmaVDAC18.2*) were targeted by three miRNAs, three genes (*GmaVDAC6.1*, *GmaVDAC9.1* and *GmaVDAC11.1*) were targeted by four miRNAs, two genes (*GmaVDAC13.2* and *GmaVDAC17.2*) were targeted by six miRNAs, two genes (*GmaVDAC5.1* and *GmaVDAC14.1*), and two other genes (*GmaVDAC8.1* and *GmaVDAC18.1*) were targeted by 11 miRNAs. Similarly, five miRNAs were found in association with *GmaVDAC17.1* gene.

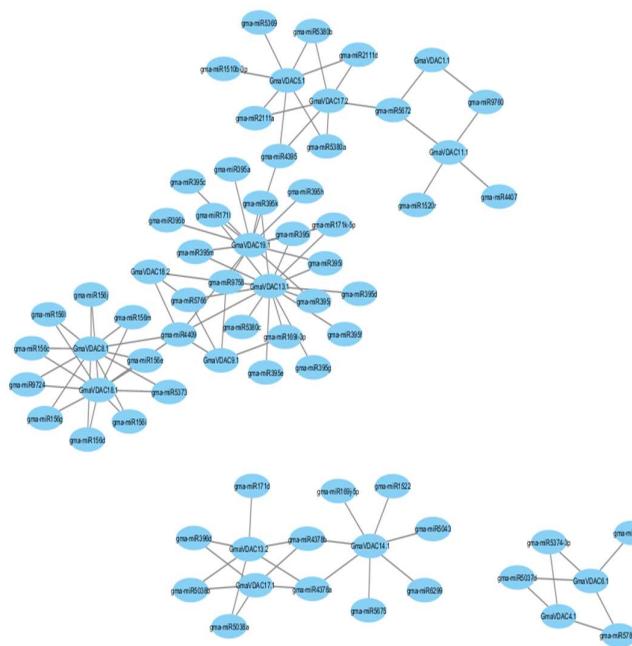


Figure 10. The schematic representation of the regulatory network relationships between putative miRNAs and their targeted soybean VDAC genes.

3. Discussion

The voltage-dependent anion-selective channel (VDAC) is a highly abundant protein in the outer membrane of mitochondria. It serves as a crucial pathway for transporting various substances, ranging from ions to large biomolecules. The expression of VDAC genes is influenced by environmental stresses. In present study, an extensive analysis was conducted on VDAC gene family in 13 legume species (including cultivated and wild soybeans) along with *Arabidopsis thaliana*.

A total of 111 VDAC genes were identified across all studied plant species including 15 genes in each of two soybean species i.e., *G. max* and *G. soja*. This shows a high structural and functional similarity among VDAC genes in soybeans. However, their varying number from four (*C. cajan*) to seventeen (*A. hypogaea*) across other legumes indicates their diversity in sequence structure and functional roles. In a previous study [16], five VDAC genes were reported in *G. max* that contradicted our results. The present study identified five VDAC genes in *Arabidopsis thaliana* however, four genes were previously reported in this species [17]. Similarly, five genes were identified in *Medicago truncatula* and the same number of genes were also reported [10].

The comparative phylogenetic tree exhibited a tight clustering between *GmaVDAC* and *GsoVDAC* genes. This shows the highest sequence similarity among VDAC genes of these two species as compared to those of any other studied species. This close similarity is possibly due to a shared ancestral origin. *G. max* and *G. soja* shared an equal number of VDAC genes in each clade, suggesting that these gene sequences remained conserved during the domestication of cultivated soybean from wild species. Moreover, six clades were formed of the tree in present study. Similarly, VDAC gene family was characterized in wheat and reported the same number of clades in comparative phylogenetic tree [4].

The sub-cellular localization in mitochondrion of all VDAC protein members indicates their significance for proper functioning of this organelle [18]. Their localization in mitochondria and small unknown vesicles was also reported in soybean [16]. A small variation in number of introns and exons of VDAC genes of both cultivated and wild soybeans suggests that the genes remain conserved throughout the process of evolution due to their great functional importance. Seven highly conserved motifs in both *G. max* and *G. soja* species show a very close sequence similarity of VDAC proteins and the preservation of VDAC gene structures in both soybean species. Genes that share similar structures

in both species also exhibit a similar composition of motifs. Most of the conserved *GmaVDAC* protein motifs (i.e., 1, 3, 4, 5, 7, 8, and 9) reported in this study were also described as conserved in wheat [4].

Gene duplications play a significant role in genome evolution and enhances the capacity of genome to adapt to changing environments [19]. The variation in the number of VDAC genes among different legume species can be attributed to this force [20]. In *G. max* and *G. soja*, the higher number of VDAC genes as compared to model legume *M. truncatula* (which has 7 VDACS) is primarily due to segmental duplication. Specifically, seven segmental duplicates were identified in *GmaVDAC* genes. Both genes of each paralogous pair identified in *G. max* exhibited not less than 90 % sequence similarity with each other. The divergence time of duplication events also varied in *GmaVDAC* genes. A higher value of Ka/Ks indicates a greater divergence in the duplication event [21]. In *G. max*, the segmental duplication events of VDAC genes occurred between 5.16 and 12.51 million years ago (Mya). These findings suggest that gene duplications in *G. max* occurred recently for this gene family. These results align closely with the findings of [22], who proposed that duplication events in the Glycine lineage occurred around 13 Mya. Though, the distribution of *GmaVDAC* and *GsoVDAC* genes on chromosomes was uneven their corresponding genomic positions in both species had more than 90% sequence similarities.

VDAC proteins act as gate-keepers in mitochondrial outer membrane and thus contribute to defense against abiotic stresses especially drought and salt stresses by regulating the movement of ions between cytoplasm and mitochondria. Previous studies have demonstrated the roles of VDAC genes in conferring tolerance to drought and salt stress [4]. Various *cis*-regulatory elements were identified in promoter region of *GmaVDAC* genes, which were mainly related to growth and development, abiotic and biotic stress tolerance, transcription regulation and light responsiveness. The varied number of different *cis*-elements is an indication of diverse roles of *GmaVDAC* genes. The presence of certain elements like ARE, as-1, Box-4, Myb, MYC, GT1-motif, WUN-motif, and STRE indicates that these genes respond to drought, salinity, cold, and hormonal conditions as was previously described by [4,10–12].

To validate the expression patterns of VDAC genes at various developmental stages, we examined their expression using RNA-seq data of *G. max*. These findings highlighted the tissue-specific expression of VDAC genes in response to developmental cues. The contiguous expression of VDAC genes was reported in all plant parts including root nodules in *L. japonicus* [16]. Similarly, the overexpression of *GmaVDAC9.1*, *GmaVDAC18.1*, and *GmaVDAC18.2* genes also showed their potential role in drought and salt stress tolerance as evidenced in wheat [4].

The gene ontology results confirmed the role of *GmaVDAC* proteins in voltage-gated anion channel activity or porin activity. These molecular functions were also described by many researchers [3,7]. The protein interactions network revealed that there was no interaction among *GmaVDAC* protein family members. This shows that the expression of one protein does not affect the expression of any other *GmaVDAC* protein. Nonetheless, their interactions with two mitochondrial import receptors suggest that activation of VDAC proteins may be regulated by these receptor proteins. In addition, 103 miRNA targets were predicted in 15 *GmaVDAC* genes. *GmaVDAC13.1* and *GmaVDAC19.1* genes were found to be targeted by 16 and 14 miRNAs respectively.

4. Materials and Methods

4.1. Identification of VDAC genes and retrieval of sequences

In this study, we searched out and retrieved the genomic, transcriptomic, and proteomic sequences of VDAC genes from the genomes of 13 legume species and a model plant species – *Arabidopsis thaliana* (Table 2) [2].

Table 2. List of crop species for phylogenetic studies and database links for retrieval of VDAC genes sequences.

Sr. No.	Crop	Botanical name	Database link
1.	Adzuki bean	<i>Vigna angularis</i>	https://legacy.legumeinfo.org/
2.	Arabidopsis	<i>Arabidopsis thaliana</i>	https://www.arabidopsis.org/
3.	Barrel clover	<i>Medicago truncatula</i>	https://phytozome-next.jgi.doe.gov/
4.	Birdsfoot trefoil	<i>Lotus japonicus</i>	https://phytozome-next.jgi.doe.gov/
5.	Chickpea	<i>Cicer arietinum</i>	https://phytozome-next.jgi.doe.gov/
6.	Common bean	<i>Phaseolus vulgaris</i>	https://phytozome-next.jgi.doe.gov/
7.	Cowpea	<i>Vigna unguiculata</i>	https://phytozome-next.jgi.doe.gov/
8.	Cultivated soybean	<i>Glycine max</i>	https://www.soybase.org/
9.	Mungbean	<i>Vigna radiata</i>	https://legacy.legumeinfo.org/
10.	Narrow leaf lupin	<i>Lupinus angustifolius</i>	https://legacy.legumeinfo.org/
11.	Pea	<i>Pisum sativum</i>	https://legacy.legumeinfo.org/
12.	Peanut	<i>Arachis hypogaea</i>	https://phytozome-next.jgi.doe.gov/
13.	Pigeon pea	<i>Cajanus cajan</i>	https://legacy.legumeinfo.org/
14.	Wild soybean	<i>Glycine soja</i>	https://www.soybase.org/

4.2. Conserved domain, subcellular localization and physico-chemical properties

We found the hidden Markov model (HMM) of VDAC proteins from the Pfam database (<https://pfam.xfam.org/>) and confirmed the presence of PF01459 domain in each VDAC protein from the conserved domain database (CDD) of NCBI (<https://www.ncbi.nlm.nih.gov/cdd>) [23].

For subcellular location, we used WegoLoc (<https://www.btool.org/WegoLoc>) [24]. Moreover, the physicochemical properties of VDAC proteins were determined by using ProtParam (<https://web.expasy.org/protparam/>) tool of Expasy website [25].

4.3. Multiple sequence alignment (MSA) and phylogenetic relationship

We conducted multiple sequence alignment (MSA) of all VDAC proteomic sequences of selected plant species by using ClustalW algorithm of MEGA11 software. Further, we performed the phylogenetic analysis and constructed a comparative phylogenetic tree in MEGA11 [26]. Thereafter, the phylogenetic tree was edited and modified by using iTOL website to make it easier to understand (<https://itol.embl.de/>).

4.4. Gene structure prediction and motif analysis

Gene Structure Display Server (GSDS) (<https://gsds.gao-lab.org/>) was employed to predict structure of VDAC genes in three species *viz.* *G. max*, *G. soja*, and *M. truncatula* [27]. For conserved motifs, MEME tool (<https://meme-suite.org/meme/tools/meme>) was used by considering at least ten motifs for each VDAC protein with motif width ranging from 6-50 [28].

4.5. Cis-regulatory elements (CREs) and Chromosomal distribution

In *G. max*, we predicted the *cis*-regulatory elements (CREs) in 2000 bp upstream promoter region of gene transcript by using CARE searching tool of PlantCARE database (<https://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) [29]. To show gene positions, we used VDAC genomic sequences of *G. max* to draw a physical map in MapChart software [30].

4.6. Gene duplications and syntenic analysis

Genes with 90% or more similarity in sequences were considered as duplicated genes. Generally, there are two types of duplications i.e., tandem duplications (if genes are separated by fewer or up to five genes) and segmental duplications (if genes are scattered on different chromosomes or are separated by more than five genes) [2].

For this purpose, we employed a software MCScanX (Multiple Collinearity Scan Toolkit) with default parameters [31]. After sequence alignment, DnaSP v5.10.01 software was used to perform synonymous (Ks) analysis to calculate Ks values and non-synonymous (Ka) analysis to calculate Ka values of substitution [32]. The Ka/Ks ratio was calculated from SNAP web-based tool to further analyze the selection mode and selection pressure during evolutionary process. The estimated time of divergence event was calculated by using formula $T = Ks/2\lambda$ where $\lambda = 6.1 \times 10^{-9}$ [22]. Similarly, we constructed syntenic plots to show the syntenic relationship among VDAC paralogs in *G. max* by using TBtools [33].

4.7. Protein interactions network and microRNA targets prediction

The possible protein-protein interaction network and microRNA targets of *GmaVDAC* genes were predicted by using STRING database (<https://string-db.org/>) and psRNATarget (<https://plantgrn.noble.org/psRNATarget/>) respectively [34,35]. Both interaction networks were drawn by using Cytoscape software [36].

4.8. Gene ontology analysis

To describe functional annotation of *GmaVDAC* genes, peptide sequences of *GmaVDAC* proteins were uploaded on Blast2GO (<https://www.blast2go.com/>) and the results were visualized using BiNGO (<https://apps.cytoscape.org/apps/bingo>) [37]. The GO analysis provides insights about cellular components, molecular functions, and biological processes of concerned genes [15].

4.8. Expression pattern of *GmaVDAC* genes

The expression analysis of *GmaVDAC* genes in 14 different tissues of soybean plant was studied using transcriptomic data previously submitted to SoyBase website (<https://www.soybase.org/soyseq/>) by [38]. The normalized RNA-seq data of different growth stages was downloaded, \log_2 transformed, and the results were displayed in the form of heatmap using TBtools [33].

Moreover, we conducted an in-depth examination of the *GmaVDAC* gene expression analysis about drought and salt stresses. We achieved this by analyzing publicly accessible RNA-seq data available on Gene Expression Omnibus (GEO) database, specifically under the accession number GSE57252 of *G. max* [39].

5. Conclusions

To date, no comprehensive study of VDAC gene family has been conducted in soybeans. Therefore, this research provides the first systematic and extensive investigation of VDAC genes in soybean. A total of 111 VDAC genes were identified and analyzed for phylogenetic studies in fourteen crop species including *G. max* and *G. soja*. The conserved gene structures and protein domains indicate their importance for proper functioning of a cell. The expansion of VDAC genes in *G. max* occurred through segmental duplication and purifying selection. Their expression pattern varied among plant tissues with the highest expression in root nodules. Similarly, the transcriptomic profile revealed the upregulation of *GmaVDAC* genes under drought and salt stress conditions. The gene annotations and protein interactions suggested their role in voltage-gated anion channel activity. Overall, this study provides valuable insights into the functions of VDAC proteins that can be further explored to harness their potential for drought and salt tolerance in soybeans.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Author Contributions: Conceptualization, M.A. and M.M.U.; methodology, M.M.U., M.A. and M.M.I.; software, M.M.U.; validation, M.M.U.; formal analysis, M.M.U.; investigation, M.M.U.; resources, M.A.; data curation, M.M.U.; writing—original draft preparation, M.M.U.; writing—review and editing, M.M.U., M.A., A.R. and A.S.; visualization, M.M.U.; supervision, M.A.; project administration, M.A.; funding acquisition, A.S.

Acknowledgments: The authors are grateful to the scientists who have contributed to this project, and to the reviewers and editors for their constructive review.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Xu, H.; Zhang, L.; Zhang, K.; Ran, Y. Progresses, Challenges, and Prospects of Genome Editing in Soybean (Glycine Max). *Front Plant Sci* **2020**, *11*, 1–19, doi:10.3389/fpls.2020.571138.
2. Aleem, M.; Aleem, S.; Sharif, I.; Wu, Z.; Aleem, M.; Tahir, A.; Atif, R.M.; Cheema, H.M.N.; Shakeel, A.; Lei, S.; et al. Characterization of SOD and GPX Gene Families in the Soybeans in Response to Drought and Salinity Stresses. *Antioxidants* **2022**, *11*, 460, doi:10.3390/antiox11030460.
3. Rostovtseva, T.K.; Bezrukova, S.M.; Hoogerheide, D.P.; Kmita, H.; Messina, A.A. Regulation of Mitochondrial Respiration by VDAC Is Enhanced by Membrane-Bound Inhibitors with Disordered Polyanionic C-Terminal Domains. *Int J Mol Sci* **2021**, *22*, 7358, doi:10.3390/ijms22147358.
4. Yu, M.; Yu, Y.; Song, T.; Zhang, Y.; Wei, F.; Cheng, J.; Zhang, B.; Zhang, X. Characterization of the Voltage-Dependent Anion Channel (VDAC) Gene Family in Wheat (Triticum Aestivum L.) and Its Potential Mechanism in Response to Drought and Salinity Stresses. *Gene* **2022**, *809*, 146031, doi:10.1016/j.gene.2021.146031.
5. Ling, H.Q.; Wang, J.; Zhao, S.; Liu, D.; Wang, J.; Sun, H.; Zhang, C.; Fan, H.; Li, D.; Dong, L.; et al. Draft Genome of the Wheat A-Genome Progenitor Triticum Urartu. *Nature* **2013**, *496*, 87–90, doi:10.1038/NATURE11997.
6. Gonçalves, R.P.; Buzhynskyy, N.; Prima, V.; Sturgis, J.N.; Scheuring, S. Supramolecular Assembly of VDAC in Native Mitochondrial Outer Membranes. *J Mol Biol* **2007**, *369*, 413–418, doi:10.1016/j.jmb.2007.03.063.
7. Shoshan-Barmatz, V.; Shteiher-Kuzmine, A.; Verma, A. VDAC1 at the Intersection of Cell Metabolism, Apoptosis, and Diseases. *Biomolecules* **2020**, *10*, 1485.
8. Kusano, T.; Tateda, C.; Takahashi, Y.; Berberich, T. Voltage-Dependent Anion Channels: Their Roles in Plant Defense and Cell Death. *Plant Cell Rep* **2009**, *28*, 1301–1308, doi:10.1007/S00299-009-0741-Z.
9. Saidani, H.; Grobys, D.; Léonetti, M.; Kmita, H.; Homblé, F. Towards Understanding of Plant Mitochondrial VDAC Proteins: An Overview of Bean (Phaseolus Vulgaris) VDAC Proteins. *2017*, *4*, 43–62, doi:10.3934/biophy.2017.1.43.
10. Mei Yang, Xinhang Duan, Zhaoyu Wang, Hang Yin, Junrui Zang, Kai Zhu, Y.W. and P.Z. Overexpression of a Voltage-Dependent Anion-Selective Channel (VDAC) Protein-Encoding Gene, MsVDAC, from *Medicago Sativa* Confers Cold and Drought Tolerance to Transgenic Tobacco. **2021**, *12*, 1706.
11. Desai, M.K.; Mishra, R.N.; Verma, D.; Nair, S.; Sopory, S.K.; Reddy, M.K. Structural and Functional Analysis of a Salt Stress Inducible Gene Encoding Voltage Dependent Anion Channel (VDAC) from Pearl Millet (*Pennisetum Glaucum*). *Plant Physiology and Biochemistry* **2006**, *44*, 483–493, doi:10.1016/j.plaphy.2006.08.008.
12. Xu, X.; Tan, Y.P.; Cheng, G.; Liu, X.Q.; Xia, C.J.; Luo, F.Y.; Wang, C.T. Genomic Survey and Gene Expression Analysis of the VDAC Gene Family in Rice. *Genetics and Molecular Research* **2015**, *14*, 15683–15696, doi:10.4238/2015.December.1.20.
13. Ma, T.; Wang, J.; Zhou, G.; Yue, Z.; Hu, Q.; Chen, Y.; ... Liu, J. Genomic Insights into Salt Adaptation in a Desert Poplar. *Nat Commun* **2013**, *4*, 2797.
14. Severin, A.J.; Cannon, S.B.; Graham, M.M.; Grant, D.; Shoemaker, R.C. Changes in Twelve Homoeologous Genomic Regions in Soybean Following Three Rounds of Polyploidy. *Plant Cell* **2011**, *23*, 3129–3136, doi:10.1105/TPC.111.089573.

15. Zhu, Y.; Yin, J.; Liang, Y.; Liu, J.; Jia, J.; Huo, H.; Wu, Z.; Yang, R.; Gong, H. Transcriptomic Dynamics Provide an Insight into the Mechanism for Silicon-Mediated Alleviation of Salt Stress in Cucumber Plants. *Ecotoxicol Environ Saf* **2019**, *174*, 245–254.
16. Wandrey, M.; Trevaskis, B.; Brewin, N.; M.K., U. Molecular and Cell Biology of a Family of Voltage-Dependent Anion Channel Porins in Lotus Japonicus. *Plant Physiol* **2004**, *134*, 182–193.
17. Tateda, C.; Watanabe, K.; Kusano, T. Molecular and Genetic Characterization of the Gene Family Encoding the Voltage-Dependent Anion Channel in Arabidopsis. *J Exp Bot* **2011**, *62*, 4773–4785–4773–4785.
18. Mok, D.W.S.; Peloquin, S.J. Breeding Value of 2n Pollen (Diplandroids) in Tetraploid x Diploid Crosses in Potatoes. *Theoretical and Applied Genetics* **1975**, *46*, 307–314, doi:10.1007/BF00281153.
19. Singh, N.; Bhatt, V.; Rana, N.; Shivaraj, S.M. Advances of Next-Generation Sequencing (NGS) Technologies to Enhance the Biofortifications in Crops. *Advances in Agri-Food Biotechnology* **2020**, *427–450*, doi:10.1007/978-981-15-2874-3_17.
20. Magadum, S.; Banerjee, U.; Murugan, P.; Gangapur, D.; Ravikesavan, R. Gene Duplication as a Major Force in Evolution. *J Genet* **2013**, *92*, 155–161, doi:10.1007/S12041-013-0212-8.
21. Wang, Y.; Wang, Q.; Zhao, Y.; Han, G.; Gene, S.Z.-; 2015, U. Systematic Analysis of Maize Class III Peroxidase Gene Family Reveals a Conserved Subfamily Involved in Abiotic Stress Response. *Gene* **2015**, *566*, 95–108.
22. Lynch, M.; Conery, J.S. The Evolutionary Fate and Consequences of Duplicate Genes. *Science (1979)* **2000**, *290*, 1151–1155, doi:10.1126/SCIENCE.290.5494.1151.
23. Marchler-Bauer, A.; Anderson, J.B.; Derbyshire, M.K.; DeWeese-Scott, C.; ..., Bryant, S.H. CDD: A Conserved Domain Database for Interactive Domain Family Analysis. *Nucleic Acids Res* **2006**, *35*, D237–D240.
24. Chi, S.M.; Nam, D. WegoLoc: Accurate Prediction of Protein Subcellular Localization Using Weighted Gene Ontology Terms. *Bioinformatics* **2012**, *28*, 1028–1030.
25. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook* **2005**, *1*, 571–607, doi:10.1385/1-59259-890-0:571.
26. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol* **2021**, *38*, 3022–3027.
27. Hu, B.; Jin, J.; Guo, A.; Zhang, H.; Luo, J.; Gao, G. GSDS 2.0: An Upgraded Gene Feature Visualization Server. *Bioinformatics* **2015**, *31*, 1296–1297.
28. Bailey, T.L.; Williams, N.; Misleh, C.; Li, W.W. MEME: Discovering and Analyzing DNA and Protein Sequence Motifs. *Nucleic Acids Res* **2006**, *34*, W369–W373.
29. Lescot, M.; Déhais, P.; Thijs, G.; Marchal, K.; Moreau, Y.; Van de Peer, Y.; Rouzé, P.; Rombauts, S. PlantCARE: A Database of Plant Cis-Acting Regulatory Elements and a Portal to Tools for in Silico Analysis of Promoter Sequences. *Nucleic Acids Res* **2002**, *30*, 325–327.
30. Voorrips, R.E. MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. *Journal of heredity* **2002**, *93*, 77–78.
31. Wang, Y.; Tang, H.; Debarry, J.; Tan, X.; Li, J.; Wang, X.; Lee, T.H.; Jin, H.; Marler, B.; Guo, H.; et al. MCScanX: A Toolkit for Detection and Evolutionary Analysis of Gene Synteny and Collinearity. *Nucleic Acids Res* **2012**, *40*, e49.
32. Korber, B. HIV Signature and Sequence Variation Analysis. *Computational analysis of HIV molecular sequences* **2000**, *4*, 55–72.
33. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol Plant* **2020**, *13*, 1194–1202, doi:10.1016/j.molp.2020.06.009.
34. Dai, X.; Zhuang, Z.; Zhao, P.X. PsRNATarget: A Plant Small RNA Target Analysis Server (2017 Release). *Nucleic Acids Res* **2018**, *46*, W49–W54.
35. Szklarczyk, D.; Gable, A.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; et al. STRING V11: Protein–Protein Association Networks with Increased Coverage,

Supporting Functional Discovery in Genome-Wide Experimental Datasets. *Nucleic Acids Res* **2019**, *47*, D607–D613.

- 36. Su, G.; Morris, J.H.; Demchak, B.; Bader, G.D. Biological Network Exploration with Cytoscape 3. *Curr Protoc Bioinformatics* **2014**, *8*, 1–24, doi:10.1002/0471250953.BI0813S47.
- 37. Conesa, A.; Götz, S. Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. *Int J Plant Genomics* **2008**, *12*, 12 pages, doi:10.1155/2008/619832.
- 38. Shen, Y.; Du, H.; Liu, Y.; Ni, L.; Wang, Z.; Liang, C.; Tian, Z. Update Soybean Zhonghuang 13 Genome to a Golden Reference. *Sci China Life Sci* **2019**, *62*, 1257–1260, doi:10.1007/s11427-019-9822-2.
- 39. Belamkar, V.; Weeks, N.T.; Bharti, A.K.; Farmer, A.D.; Graham, M.A.; Cannon, S.B. Comprehensive Characterization and RNA-Seq Profiling of the HD-Zip Transcription Factor Family in Soybean (Glycine Max) during Dehydration and Salt Stress. *BMC Genomics* **2014**, *15*, 1–25, doi:10.1186/1471-2164-15-950.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.