

Unveiling Integrative Computational Analysis and Functional Profiling of GhSOS1 gene family in response to dual stresses Drought and Salinity in *Gossypium hirsutum* L.

Laraib Iqra

Laraibiqra841@gmail.com

University of Education

Muhammad Naveed Shahid

University of Education

Umar Akram

Usman Arif

Research Article

Keywords: Insilico characterization of SOS1, Cis-regulatory elements, phylogenetic analysis, qPCR validation of GhSOS1

Posted Date: October 22nd, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-7788692/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Unveiling Integrative Computational Analysis and Functional Profiling of *GhSOS1* gene family in response to dual stresses Drought and Salinity in *Gossypium hirsutum* L.

LARAIB IQRA^{1*}, Muhammad Naveed Shahid¹, Umar Akram², Usman Arif³

¹ Department of Botany, Division of science and technology, University of Education, Lahore

² MNS- University of Agriculture, Multan, Pakistan

³ Centre of excellence in molecular biology, University of the Punjab, Lahore

Corresponding Author email: laraibiqra841@gmail.com

Abstract

Background

The global production of upland cotton (*Gossypium hirsutum*) has declined in recent years due to increasing drought and salinity stress. *SOS1* gene family plays a key role in maintaining ion homeostasis and enhancing stress resilience.

Methods

A genome-wide study was conducted to identify *SOS1* genes, including analyses of gene structure, motifs, chromosomal distribution, phylogenetic relationships, Ka/Ks ratios, cis-regulatory elements, protein–protein interactions, 3D structures, and phosphorylation sites. Fifteen paralogs of *G. hirsutum* were identified and functionally characterized. For the validation of in-silico findings, expression of 15 genes was profiled during salt and drought stress by RNA-seq data that was available publicly and finally it led to selection of five highly responsive species. After exposing cotton seedlings to 400mM NaCl and 20% PEG, quantitative Realtime PCR was performed with different time intervals (0h, 3h, 6h, and 12 h).

Results

The presence of cis-regulatory elements, including ABRE, DRE/CRT, MYB sites, and other transcription factor motifs, further confirmed their stress response role. Ka/Ks analysis indicated strong conservation among gene paralogs, with most under purifying selection. Protein–protein interaction analysis revealed *SOS1* association with antiporters, facilitating Na⁺ efflux across the plasma membrane to reduce drought and salinity effects. Prominent transcriptional changes

induced by stress was confirmed by qPCR results. The results showed consistency with in-silico predictions by up-regulation of *GhSOS1-5* and *GhSOS1-11*.

Conclusions

Collectively, this is a comprehensive study of the *SOS1* family in *G. hirsutum* which highlights its roles in conferring abiotic stress tolerance. These findings will serve as a valuable resource for molecular breeding.

Keywords: Insilico characterization of *SOS1*, Cis-regulatory elements, phylogenetic analysis, qPCR validation of *GhSOS1*

1. Introduction

Cotton is one of the most economically significant crops, serving as the primary source of natural fiber for the global textile industry. However, its productivity is increasingly threatened by abiotic stressors such as drought and salinity, which cause substantial yield losses. Disruption of ion homeostasis in plants is a key physiological indicator of salinity stress and drought [1].

A major limiting factor in productivity and growth of upland cotton (*Gossypium hirsutum*) is drought. [2]. Under drought conditions, cotton plants undergo a cascade of physiological changes, including reduced stomatal conductance and a decline in photosynthetic efficiency [3]. Another hallmark of drought stress is the excessive accumulation of reactive oxygen species (ROS), which can cause severe cellular damage [4]. In addition, drought alters carbohydrate concentrations in leaves and triggers osmotic adjustment mechanisms to maintain cellular homeostasis in *G. hirsutum* [5].

Salt stress is another major challenge to sustainable agriculture, directly reducing yield along with the growth of upland cotton (*Gossypium hirsutum*) [6]. Globally, nearly 20% of the agricultural land is affected by salinity [7]. Upland cotton, along with staple crops such as wheat and rice, is particularly vulnerable to saline irrigation water [8], necessitating the development of tolerance mechanisms and mitigation strategies [9]. Under salt stress, cotton plants typically show an increase in root length, whereas other physiological parameters, including stem diameter, leaf area, and chlorophyll a and b contents, tend to decline [10].

Traditional practices to mitigate soil salinity, such as soil drainage, soil leaching and strip cropping have shown only limited success [11]. Salinity tolerance, on the other hand, is an intricate phenomenon regulated by diverse biochemical, physiological and molecular responses [12]. Plants employ different mechanisms to tolerate salinity stress such as ion compartmentalization, maintenance of ion homeostasis and reactive oxygen species (ROS) detoxification [13]. Among these mechanism salt overly sensitive (SOS) signaling pathway plays a vital role in maintaining ion equilibrium, thereby mitigating the adverse effects of both salinity and drought stress [14]. This pathway was first characterized in *Arabidopsis thaliana* and comprises three core components: SOS1, SOS2, and SOS3. SOS1 encodes vacuole-localized Na⁺/H⁺ exchanger responsible for pumping excessive sodium ions out of the cell. Additionally, SOS1 may contribute indirectly to high-temperature stress tolerance by maintaining ionic balance.

A crucial pathway of ion homeostasis in cotton under salinity and drought stresses is SOS (salt overly sensitive) pathway [15]. The SOS gene family is an evolutionarily conserved group of genes involved in maintaining ion homeostasis, particularly in response to salt stress [16]. Although SOS1 gene family has been widely studied in many other plants like *Arabidopsis thaliana*, *Zea mays* and *Oryza sativa* [17], genome-wide investigations in *Gossypium hirsutum* remain limited.

In recent years, genome-wide studies have emerged as powerful approaches for analyzing gene families and elucidating their functional roles in abiotic stress responses [18]. Therefore, the present study was undertaken to investigate the *GhSOS1* gene family through comprehensive genome-wide characterization, including analyses of gene structure, conserved domains, motifs, protein homology and phylogeny. To validate insilico results expression profiling of *GhSOS1* gene family qPCR was performed on candidate genes and expression level was noticed. Considering the economic importance of cotton, this study provides valuable insights into the functional characterization of the Salt Overly Sensitive 1 (SOS1) gene family, thereby contributing to the development of stress-resilient cultivars.

2. Materials and Methods

2.1 Identification of *SOS1* Genes in *G. hirsutum*

Genome of *Gossypium hirsutum* was downloaded from Phytozome v.13 (<https://phytozome-next.jgi.doe.gov/>), a publicly available database. Candidate genes were shortlisted by BLASTP

(<https://phytozome-next.jgi.doe.gov/blast-search>) search against queries with standard BLAST parameters e-value $1e-5$ and identity more than 90% . Query sequences of SOS1 were retrieved from model plant *Arabidopsis thaliana* protein; sequences downloaded from its databases TAIR 10 (<https://www.arabidopsis.org/>). The identified protein sequences were subjected to NCBI CDD-batch search for identifying the conserved domains present in *GhSOS1* using standard parameters (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>) and verified by Pfam database (<http://pfam.xfam.org/>) and SMART (<http://smart.embl-heidelberg.de/>) using default setting. All the protein sequences that do not contain Na⁺/H⁺ domain were discarded.

2.2 Construction of Phylogenetic Tree and selection pressure Analysis

To infer the evolutionary relationship of identified candidate *GhSOS1* proteins with other plant species having same proteins , a phylogenetic tree was constructed by neighbor-joining method with 1000 bootstrap values using MEGA V11.0.13 software (<https://www.megasoftware.net/>), and results were saved as newick format and exported to online tool itol (<https://itol.embl.de/upload.cgi>) for constructing a graphic representation. To infer the selection pressure between candidate genes paralogs were determined by MEGA V11.0.13. Ka/Ks (non-synonymous/synonymous) ratio for the duplicated genes were determined to figure out the diversity in function and duplication events in gene pairs.

2.3 Gene Structure and Conserved Domain and motif Analysis

Gene structure was graphically determined by using TBtools v2.313 with the help of *Gossypium hirsutum* Gff annotation file method . Conserved domains were analyzed via Pfam and then motifs were determined by MEME Suite (<https://meme-suite.org/>) in candidate genes .

2.4 Physicochemical Properties and Subcellular Localization of GhSOS1 protein

In-silico characterization like molecular weight (KDa), CDS length (bp) and isoelectric point (pI) and GAVY of identified 15 proteins was measured by using ProtParam expasy (<https://web.expasy.org/protparam/>) . To predict the sub-cellular localization of the candidate proteins CELLO V2.5 (<http://cello.life.nctu.edu.tw/>) and WOLPSORT (<https://wolfsort.hgc.jp/>) were used .

2.5 Chromosomal Distribution

Chromosomal distribution and gene locus of identified 15 *SOS1* genes was manually displayed by using online tool Mapchart (<https://www.wur.nl/en/show/Mapchart.htm>) .

2.6 Cis-Regulatory Element Analysis

PlantCARE database was used (<https://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) in promoter sequences (2kb upstream) of these 15 proteins for determination of cis-regulatory elements and stress-responsive transcriptions factors.

2.7 Protein-Protein Interaction (PPI) Network Prediction

To see the regulatory Protein-protein interaction of 15 shortlisted *hirsutum* proteins online software STRING (<https://string-db.org/>) was used .

2.8 Prediction of Phosphorylation Sites and Transmembrane Domain

Phosphorylation sites were identified using online software Net Phos 3.1a (<https://services.healthtech.dtu.dk/services/NetPhos-3.1/>) in *GhSOS1* and a bioinformatics tool TMHMM - 2.0 (<https://services.healthtech.dtu.dk/services/TMHMM-2.0/>) was used for determining the transmembrane domain.

2.9 Protein 3D structure

Protein tertiary structures of the identified *GhSOS1* proteins was determined by SWIS-MODEL (<https://swissmodel.expasy.org/>) Suitable template structures were selected automatically based on highest GMQE.

2.10 In-Silico Expression Profiling and Candidate Gene Selection

Transcript abundance of the *SOS1* gene family was analyzed under abiotic stress by using RNA-seq datasets of *Gossypium hirsutum* (upland cotton). CottonGVD database (https://db.cngb.org/cottonGVD/tool/pop_expression) was used for retrieval of publicly available RNA-seq datasets. Comparison between stress treatments and tissues was enabled by using normalized expression values (FPKM) of all 15 identified *GhSOS1* genes. TB tools were used for compiling and visualizing these values as heatmap. Five *GhSOS1* genes which showed the

substantial induction and consistent differential expression patterns in salt and drought conditions, were selected for experimental validation using quantitative real-time PCR (qPCR).

2.11 RNA Extraction and cDNA Synthesis

Specific controlled conditions (28 ± 2 °C Day/night temperature and a 16-h photoperiod) were used for growing upland cotton (*Gossypium hirsutum* L., till the tree-leaf stage in a greenhouse. For implementation of abiotic stress like drought stress, seedlings were exposed to heavy salt stress 400 mM NaCl and other set was exposed to 20 % (w/v) polyethylene glycol-6000 (PEG). After treatment at time intervals of 0 h (control), 3 h, 6 h, and 12 h, leaves were harvested having three biological replicants for each time point. Total RNA was extracted from 100mg of fresh cotton leaves using the TRIzol reagent method (Invitrogen, USA) followed by a manual protocol [19]. RNA was precipitated by using 500 μ L of isopropanol after the aqueous phase, and the pellet was air dried in RNase-free water [20]. Nanodrop spectrophotometer was used to assess the purity and concentration of RNA, and integrity was confirmed by 1% agarose gel electrophoresis [21]. Thermoscientific RevertAid First Strand cDNA Synthesis Kit was used to synthesize cDNA and to ensure its reverse transcription; oligo(dT) primers and random hexamers were used as a mix. Diluted cDNA was stored at -80°C for qPCR analysis.

2.12 Gene Expression Analysis (qPCR) of *GhSOS1*

Gene expression of selective *GhSOS1* genes was studied by using SYBR Green ReadyMix - Fast SYBR Green qPCR system. Amplification protocol was followed stepwise denaturation at 95°C for 3 minutes, thermocycle 40 cycles of 95°C followed by annealing at 55°C for 30 sec, and extension [22]. Gene expression levels were calculated by $2^{-\Delta\Delta\text{CT}}$ method [23]. Ubiquitin (*GhUBQ*) gene was used as housekeeping gene as stabilizer. Primers used are given in table [\(S1\)](#). Three biological replicates which were harvested at different time intervals were used for reaction along-with three technical replicates each individually. One-way analysis of variance (ANOVA) was used for assessing statistical significance of expression differences among treatments and time points. After ANOVA Tukey's post-hoc test ($p < 0.05$) was also performed and finally results were presented in bar graphs as mean $\pm$ standard error by using R software.

3. Results

3.1 Selection Pressure Analysis and Phylogenetic Tree

To observe the evolutionary relationship of *Gossypium hirsutum* SOS1 gene with other plant species a phylogenetic tree was constructed by neighbor-joining method in MEGA V11.0.13. The results showed different clades and sub-groups. In clade-I marked as red includes majority of *GhSOS1* members clustered with other SOS1 members of other plants i.e. rice *OsNHX5*. Clade-II colored as blue also elucidated ancestral relationship of several *GhSOS1* members with *Arabidopsis thaliana*. The highest bootstrap value present in Clade-III depicted strong evolutionary relationship of *GhSOS1* gene family members with *Arabidopsis AtNHX2*, *AtNHX4* and *Vitis vinifera VvNHX2-1*. Orange colored clade-IV showed cross species conservation between *Gossypium hirsutum* and other crops like *Glycine max*, *Brassica napus* and few members of *Arabidopsis*. *GhSOS1-9* and *GhSOS1-10* were observed in distinct clade-V showed these genes functionally divergent (Figure 1). Ka/Ks ratio indicated most of the *GhSOS1* gene pairs had strong purifying selection pressure, but few had diversifying type of evolution. Most of the *GhSOS1* gene members had functional conservation but few had shown diversity under different evolutionary selection pressure. Segmental duplication event were observed in most of *GhSOS1* genes while few conserved tandem duplication events (Table 1).

3.2 Analysis of conserved domain, motif and gene structure of *GhSOS1* Proteins

There was substantial variation in structures of these 15 *GhSOS1* proteins. All genes contained Na⁺/H⁺ and few contained K⁺ flux domain in addition. All genes had different length and number of exons and introns present among entire sequences (Figure 2). Conserved domain is predicted via NCBI CDD search domain and results revealed that all 15 *GhSOS1* proteins possessed Na⁺/H⁺ exchanger domains (Figure 3). Sequence conservation and similar functionalities among identified genes were determined by analyzing conserved motif in SOS1 gene family of *G. hirsutum* genes. Among 10 identified conserved motifs of all *GhSOS1* genes, Motif 1, 2 and 9 were the most prevalent. The presence and a specific alignment of these motifs across *GhSOS1* genes showcased functional conservation (Figure 4).

3.3 In-silico Physicochemical Properties of *GhSOS1* Proteins

Molecular weight of *GhSOS1* proteins ranged between 56.8-90.5 k Da and isoelectric point varied between 5.35-8.99 that induces change in solubility of proteins. Amino acid counts ranges from 520-827 AA. The CDS length ranged from 774-1800 bps (Table 2). Significant variations among CDS length, molecular weight and isoelectric point elucidated diversity among protein orthologs.

3.4 Subcellular Localization of *GhSOS1* Proteins

Most of the *GhSOS1* protein members were localized in plasma membrane (Figure 5).

3.5 Chromosomal Distribution and Gene Duplication

Chromosomal mapping via Mapchart tool showed non-random distribution of *GhSOS1* genes across the A and D subgenomes, indicating the occurrence of segmental duplication events that contributed to the expansion of this gene family (Figure 6).

3.6 Promoter Analysis and Cis-Regulatory Elements

Multiple Cis-regulatory elements such as Auxin responsive elements, MYB-binding site transcription factors and drought resistant elements in promotor region were present in promotor region of *GhSOS1* proteins actively work in drought and salinity stress tolerance in cotton (Figure 7).

3.7 Protein-Protein Interaction and Functional Implications

PPI analysis revealed relationship of *GhSOS1* show high confidence bond with sodium hydrogen exchanger proteins, HKT1 and few kinases (Figure 8).

3.8 Prediction of Phosphorylation Sites and Transmembrane Domain Prediction of *GhSOS1*

Phosphorylation sites were identified using Net Phos 3.1a which revealed phosphorylation of multiple potential serine (S), threonine (T), and tyrosine (Y) residues. Multiple phosphorylation sites indicate a versatile regulatory mechanism which potentially govern the function of *GhSOS1* in presence of external signals especially drought and salinity stress. Phosphorylation sites were not uniformly distributed across protein chain, while threonine and tyrosine showing less higher frequency as compared to serine residue which exhibits highest one (Figure 9 a, b, c). Transmembrane topology of *GhSOS1* was predicted using the TMHMM tool, which revealed multiple transmembrane helices distributed across the protein sequence. The analysis confirmed that *GhSOS1* functions as an integral membrane protein, with alternating regions predicted to be

embedded within the membrane (transmembrane domains), cytoplasmic (inside) or outside (Figure 10 a, b).

3.9 Protein 3D Structure Prediction of *GhSOS1*

To dive into structural conformation and functional domains of *GhSOS1*, SWISS-MODEL was used to predict tertiary structure of this protein. The model revealed its role in ion transport by gaining insights into its structural conformation like defined transmembrane topology, Na⁺/H⁺ antiporters along with multiple α -helices spanning the membrane (Figure 11). Preserved domains were identified, critical for cation exchange and their alignment with already known SOS1 homologs from other species of plants.

3.10 In-Silico Expression Profiling and Candidate Gene Selection

Under salt and drought stress expression of all 15 *GhSOS1* genes were profiled by using the RNA-seq data which was retrieved from the cotton transcriptome datasets available publicly. After treatment expression values at 0, 3, 6, 12, and 24 h were normalized and visualized as heatmap. As a result, different members from gene family were highlighted which showed different expression as compared to control. Genes which showed consistent and stress response expression patterns under salt and salinity conditions are *GhSOS1-4*, *GhSOS1-5*, *GhSOS1-10*, *GhSOS1-11*, and *GhSOS1-13*. Because of their expression consistency these genes were chosen for experimental validation (Figure 12).

3.11 Gene Expression Analysis (qPCR) of *GhSOS1* Genes

Distinct temporal responses were observed to both 400 mM NaCl and 20 % PEG treatments after performing quantitative real-time PCR. Overall, up-regulation at early (3 h) and mid (6 h) time points was observed for *GhSOS1-11* and *GhSOS1-5* under salt stress. Similarly, a well-maintained increased expression was observed for *GhSOS1-11* under drought. However, *GhSOS1-13* showcased an average down-regulation at different time intervals and similarly a significant lowered transcript abundance under salinity was observed in *GhSOS1-4*. Meanwhile, *GhSOS1-10* showed no remarkable changes and remained stable throughout treatments and time points. All of the reactions were performed in triplicate and finally the data of gene expression was analyzed by using ANOVA and Tukey's post-hoc test ($p < 0.05$). Bar graphs show mean $\pm$ standard error and asterisks (* $p < 0.05$; ** $p < 0.01$) above each bar display significance and shows treatment- and time-specific differences (Figure 13 a, b, c, d, e).

4. Discussion

Close relationship of *GhSOS1* members with rice *OsNHX5* suggested that both plants have functional conservation [24]. This diversification of SOS1 genes under different phylogenetic groups reflect the adaptive evolution under abiotic stress tolerance [25]. Phylogenetic relationship of SOS1 protein among different plants depicted possible functional evolutionary convergence of *GhSOS1* [26]. The SOS1 motifs were found in association with the Na⁺_H⁺ exchanger domain which plays a pivotal role in transport of sodium ion under drought and salt stress tolerance [27]. Such conserved motif have been reported in other plants for SOS1 genes and a bit of variation may highlight gene-specific roles in response to abiotic stress factors [28]. Na⁺/H⁺ exchanger which plays a key role in ion-homeostasis[29]. Presence of sodium hydrogen and potassium hydrogen domains confirmed the functional role of SOS1 in *G.hirsutum* under drought and salinity stress [30]. This superfamily is known for its role to unload sodium ion from xylem and maintain ion balance under drought and salinity. Presence of these domains in *GhSOS1* proteins clearly predicted that cotton is structurally equipped with sodium ion regulator protein to enhance the salt tolerance at genetic level [31]. By physiochemical properties hydrophilic nature and intracellular stability of protein is ensured. Low GAVY values in Areca palm determined low hydrophilicity among orthologs [33]. It is consistent with the evidence that SOS1 acts as exchanger protein across membrane to maintain ion balance under drought and salt stress [34]. SOS1 is protein showed involvement in stress-receptor signaling functions [35]. Distribution of same gene family over different set of genome depicts the functional divergence among their functions [36]. Segmental duplication, therefore, plays a key role in driving diversification of gene functions across chromosomes [37]. Presence of stress responsive cis-regulatory elements in promotor region illustrated the importance of insect resistance gene in meta-analysis study of barley [38]. Transcription factors present in promotor region regulates drought responsive genes [39]. Interaction between antiporters increases the significance of these proteins and their role in ion transport [40]. All known proteins in the network interacting with *GhSOS1* have active role in stress signaling [41]. Phosphorylation plays a vital role under abiotic conditions in regulation of protein function and signaling pathways [42]. Several phosphorylation sites indicated crucial role

in regulation of SOS1 activity and significance in post-translational modifications [43]. *GhSOS1* as membrane bounded transporter which plays its presumed role in ion homeostasis and stress tolerance in *G.hirsutum* is strongly justified by presence of various transmembrane helices [44]. Transmembrane domains are connected because of intracellular and extracellular loops and may play a vital role in regulatory interactions [45]. These results go parallel with the previous findings about SOS1 proteins from other plants illustrating a typical structural and functional role [46]. 3D structure provided a base for molecular docking and functional characterization [47]. This report was aligned with the previously genome-wide surveys which shows that SOS1 transporters are crucial Na^+/H^+ antiporters in regulating the cytosolic ion balance [48]. These results align with the in-silico predictions by indicating *GhSOS1-5* and *GhSOS1-11* as the most salt and drought responsive members among *GhSOS1* family. Quantitative real time PCR confirmed that under salt and drought stress, *GhSOS1-5* and *GhSOS1-11* show up-regulation particularly at early (3 h) and mid (6 h) stage, which suggest a signaling role for Na^+ extrusion and osmotic adjustment. Similarly, in *Arabidopsis thaliana* and *Oryza sativa* an early induction of SOS1 orthologs has been observed, where increased SOS1 activity is associated with high salt tolerance by maintaining K^+/Na^+ ratio[49,50]. The stable expression of *GhSOS1-10* throughout treatment shows tissue specific and housekeeping functions which are not related to extreme level of stress and a consistency with cotton SOS1 paralogs showing constitutive expression. Interestingly, a remarkable down-regulation was observed in *GhSOS1-4* which indicates negative regulation or different response towards osmotic stress. Certain SOS1 members show down-regulation under high salinity like in tomato and barley and it indicates feedback mechanism to tune ion flux for preventing overconsumption of energy[51]. However, *GhSOS1-13* displayed mixed regulation, which suggest context dependent expression and influenced by post-translational controls or intervention of other signaling pathways like ABA or calcium signaling [52].

5.Conclusion

This study presents the first genome-wide identification and functional characterization of the *GhSOS1* gene family in *Gossypium hirsutum*. The findings highlight the pivotal role of *GhSOS1* genes in drought and salt stress tolerance, while also uncovering their structural diversity and evolutionary conservation. Comprehensive in-silico analyses provided strong evidence of their function, including the presence of stress-responsive cis-regulatory elements, conserved Na^+/H^+

exchanger domains, and evolutionary relationships with SOS1 genes from other plant species, confirming functional conservation. Motif analysis and protein homology modeling further predicted strong functional similarity of *GhSOS1* proteins with sodium/hydrogen exchanger porters. Most *GhSOS1* proteins were localized to the plasma membrane, suggesting their direct involvement in ion transport and regulation of the SOS signaling pathway under abiotic stress. Substantially, five candidate genes were shortlisted based on FPKM values retrieved from publicly available RNA-seq data and tested for qPCR different intervals under 400Mm Salt and 20% drought. Subsequently *GhSOS1-5* and *GhSOS1-11* showed upregulation but *GhSOS1-4* was downregulated under both stresses meanwhile *GhSOS1-10* showed stability. These expression levels were supported by statistical analysis corroborating with in-silico predictions emphasized *GhSOS1-5* and *GhSOS1-11* as key regulators for ionic homeostasis. Collectively, these results demonstrate the critical role of *GhSOS1* genes in enabling cotton to adapt to drought and salinity stress, and provide a valuable foundation for their targeted utilization in future molecular breeding programs aimed at developing stress-resilient cotton cultivars.

References

1. Saud, S., & Wang, L. (2022). Mechanism of cotton resistance to abiotic stress, and recent research advances in the osmoregulation related genes. *Frontiers in Plant Science*, 13, 972635.
2. Ul-Allah, S., Rehman, A., Hussain, M., & Farooq, M. (2021). Fiber yield and quality in cotton under drought: Effects and management. *Agricultural Water Management*, 255, 106994.
3. Farooq, M., Wahid, A., Zahra, N., Hafeez, M. B., & Siddique, K. H. (2024). Recent advances in plant drought tolerance. *Journal of Plant Growth Regulation*, 1-33.
4. Zhang, D., Zhang, Y., Sun, L., Dai, J., & Dong, H. (2023). Mitigating salinity stress and improving cotton productivity with agronomic practices. *Agronomy*, 13(10), 2486.

5. Soni, S., Kumar, A., Sehrawat, N., Kumar, A., Kumar, N., Lata, C., & Mann, A. (2021). Effect of saline irrigation on plant water traits, photosynthesis and ionic balance in durum wheat genotypes. *Saudi Journal of Biological Sciences*, 28(4), 2510-2517.
6. Sharif, I., Aleem, S., Farooq, J., Rizwan, M., Younas, A., Sarwar, G., & Chohan, S. M. (2019). Salinity stress in cotton: effects, mechanism of tolerance and its management strategies. *Physiology and Molecular Biology of Plants*, 25, 807-820.
7. Hasanuzzaman, M., & Fujita, M. (2022). Plant responses and tolerance to salt stress: physiological and molecular interventions. *International Journal of Molecular Sciences*, 23(9), 4810.
8. Guo, M., Wang, X. S., Guo, H. D., Bai, S. Y., Khan, A., Wang, X. M., ... & Li, J. S. (2022). Tomato salt tolerance mechanisms and their potential applications for fighting salinity: A review. *Frontiers in Plant Science*, 13, 949541.
9. Maryum, Z., Luqman, T., Nadeem, S., Khan, S. M. U. D., Wang, B., Ditta, A., & Khan, M. K. R. (2022). An overview of salinity stress, mechanism of salinity tolerance and strategies for its management in cotton. *Frontiers in Plant Science*, 13, 907937.
10. Wang, Q., Lu, X., Chen, X., Malik, W. A., Wang, D., Zhao, L., ... & Ye, W. (2021). Transcriptome analysis of upland cotton revealed novel pathways to scavenge reactive oxygen species (ROS) responding to Na₂SO₄ tolerance. *Scientific Reports*, 11(1), 8670.
11. Cha, J. Y., Kim, J., Jeong, S. Y., Shin, G. I., Ji, M. G., Hwang, J. W., ... & Kim, W. Y. (2022). The Na⁺/H⁺ antiporter SALT OVERLY SENSITIVE 1 regulates salt compensation of circadian rhythms by stabilizing GIGANTEA in Arabidopsis. *Proceedings of the National Academy of Sciences*, 119(33), e2207275119.
12. Chaudhary, M. T., Majeed, S., Rana, I. A., Ali, Z., Jia, Y., Du, X., ... & Azhar, M. T. (2024). Impact of salinity stress on cotton and opportunities for improvement through conventional and biotechnological approaches. *BMC Plant Biology*, 24(1), 20.
13. Bui, H. H. T., Nguyen, D. H., Dinh, L. T. T., Trinh, H. T. T., Vu, T. K., & Bui, V. N. (2025). Comprehensive In Silico Analysis of the NHX (Na⁺/H⁺ Antiporter) Gene in Rice (*Oryza sativa* L.). *International Journal of Plant Biology*, 16(1), 6.

14. Jiang, W., Pan, R., Buitrago, S., Wu, C., Abou-Elwafa, S. F., Xu, Y., & Zhang, W. (2021). Conservation and divergence of the TaSOS1 gene family in salt stress response in wheat (*Triticum aestivum* L.). *Physiology and Molecular Biology of Plants*, 27(6), 1245-1260.
15. Liu, H., Wang, K., Mei, Q., Wang, X., Yang, J., Ma, F., & Mao, K. (2023). Genome-wide analysis of the *Actinidia chinensis* NHX family and characterization of the roles of AcNHX3 and AcNHX7 in regulating salt tolerance in *Arabidopsis*. *Environmental and Experimental Botany*, 214, 105477.
16. Abedi, A., Pourkarimi Daryakenari, M., Zare, F., & Hajiahmadi, Z. Evolutionary Dynamics of the Calcium/Cation Antiporter Superfamily in Brassicaceae: Codon Usage, Selection Pressure, and BnCaCAs Role in Abiotic Stress Response. *Frontiers in Plant Science*, 16, 1506461.
17. El Mahi, H., Pérez-Hormaeche, J., De Luca, A., Villalta, I., Espartero, J., Gámez-Arjona, F., ... & Quintero, F. J. (2019). A critical role of sodium flux via the plasma membrane Na⁺/H⁺ exchanger SOS1 in the salt tolerance of rice. *Plant physiology*, 180(2), 1046-1065.
18. Bhattacharjee, S., Paul, A., Jana, A., Unnati, G. M., Li, W., Chatterjee, A., ... & Mishra, N. (2025). Genome Wide Identification, Expression Profiles and Regulatory Network Analysis of SOS Group of Genes in *Camellia Sinensis*. *Tropical Plant Biology*, 18(1), 1-27.
19. Rio DC, Ares M, Jr., Hannon GJ, Nilsen TW. Purification of RNA using TRIzol (TRI reagent). *Cold Spring Harb Protoc*. 2010;2010(6):pdb.prot5439.
20. Chomczynski, P., & Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Analytical Biochemistry*, 162(1), 156–159.
21. Desjardins, P., & Conklin, D. (2010). NanoDrop microvolume quantitation of nucleic acids. *Journal of Visualized Experiments*, (45), e2565.
22. Bustin, S. A., et al. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*, 55(4), 611–622.

23. Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods*, 25(4), 402–408.
24. Cui, J. Q., Hua, Y. P., Zhou, T., Liu, Y., Huang, J. Y., & Yue, C. P. (2020). Global landscapes of the Na⁺/H⁺ antiporter (NHX) family members uncover their potential roles in regulating the rapeseed resistance to salt stress. *International Journal of Molecular Sciences*, 21(10), 3429.
25. Feng, L., Chen, Y., Ma, T., Zhou, C., Sang, S., Li, J., & Ji, S. (2024). Integrative physiology and transcriptome sequencing reveal differences between *G. hirsutum* and *G. barbadense* in response to salt stress and the identification of key salt tolerance genes. *BMC Plant Biology*, 24(1), 787.
26. Guo, H., Li, S., Min, W., Ye, J., & Hou, Z. (2019). Ion omic and transcriptomic analyses of two cotton cultivars (*Gossypium hirsutum* L.) provide insights into the ion balance mechanism of cotton under salt stress. *PLoS One*, 14(12), e0226776.
27. Chen, X., Lu, X., Shu, N., Wang, D., Wang, S., Wang, J., ... & Ye, W. (2017). GhSOS1, a plasma membrane Na⁺/H⁺ antiporter gene from upland cotton, enhances salt tolerance in transgenic *Arabidopsis thaliana*. *PLoS One*, 12(7), e0181450.
28. Wang, J., Hu, Y., Guo, D., Gao, T., Liu, T., Jin, J., ... & Song, C. (2025). Evolution and functional divergence of glycosyltransferase genes shaped the quality and cold tolerance of tea plants. *The Plant Cell*, 37(1), koae268.
29. Zhang, F., Shi, X., Xu, J., Yuan, W., & Li, Z. (2024). Tandem gene duplication selected by activation of horizontally transferred gene in bacteria. *Applied Microbiology and Biotechnology*, 108(1), 340.
30. Sun, J., Li, S., Guo, H., & Hou, Z. (2021). Ion homeostasis and Na⁺ transport-related gene expression in two cotton (*Gossypium hirsutum* L.) varieties under saline, alkaline and saline-alkaline stresses. *PLoS One*, 16(8), e0256000.
31. Cushman, K. R., Pabuayon, I. C., Hinze, L. L., Sweeney, M. E., & de Los Reyes, B. G. (2020). Networks of physiological adjustments and defenses, and their synergy with

sodium (Na⁺) homeostasis explain the hidden variation for salinity tolerance across the cultivated *Gossypium hirsutum* germplasm. *Frontiers in plant science*, 11, 588854.

32. Mohanta, T. K., Khan, A., Hashem, A., Abd_Allah, E. F., & Al-Harrasi, A. (2019). The molecular mass and isoelectric point of plant proteomes. *BMC genomics*, 20(1), 631.
33. Alam, O. (2023). In Silico Analysis of SOS Family in Areca palm (*Areca catechu* L.).
34. Heilmann, M., & Heilmann, I. (2022). Regulators regulated: Different layers of control for plasma membrane phosphoinositides in plants. *Current opinion in plant biology*, 67, 102218.
35. Murugan, N., Palanisamy, V., Channappa, M., Ramanathan, V., Ramaswamy, M., Govindakurup, H., & Chinnaswamy, A. (2022). Genome-wide in silico identification, structural analysis, promoter analysis, and expression profiling of PHT gene family in sugarcane root under salinity stress. *Sustainability*, 14(23), 15893.
36. Joshi, S., Patil, S., Hossain, M. A., & Kumar, V. (2024). In-silico identification and characterization of sucrose transporter genes and their differential expression pattern under salinity stress in *Glycine max*. *Plant Stress*, 11, 100314.
37. Wang, Y., Pan, C., Chen, Q., Xie, Q., Gao, Y., He, L., ... & Zhao, Y. (2023). Architecture and autoinhibitory mechanism of the plasma membrane Na⁺/H⁺ antiporter SOS1 in *Arabidopsis*. *Nature Communications*, 14(1), 4487.
38. Shariatipour, N., & Heidari, B. (2020). Meta-analysis of expression of the stress tolerance associated genes and uncover their cis-regulatory elements in rice (*Oryza sativa* L.). *The Open Bioinformatics Journal*, 13(1).
39. Baldoni, E., Genga, A., & Cominelli, E. (2015). Plant MYB transcription factors: their role in drought response mechanisms. *International Journal of Molecular Sciences*, 16(7), 15811-15851.
40. Liu, L., Luo, S., Ma, L., Zhang, Y., Wang, T., Wang, J., ... & Xue, S. (2024). Analysis of ion transport properties of *Glycine max* HKT transporters and identifying a regulation of GmHKT1; 1 by the non-functional GmHKT1; 4. *Plant and Cell Physiology*, 65(9), 1399-1413.

41. Damaris, R. N., & Yang, P. (2021). Protein phosphorylation response to abiotic stress in plants. *Plant phosphoproteomics: methods and protocols*, 17-43.
42. Zhang, X., Jiang, L., & Liu, H. (2021). Forkhead box protein O1: functional diversity and post-translational modification, a new therapeutic target. *Drug Design, Development and Therapy*, 1851-1860.
43. Zhang, X. Y., Tang, L. H., Nie, J. W., Zhang, C. R., Han, X., Li, Q. Y., ... & Chen, Y. H. (2023). Structure and activation mechanism of the rice Salt Overly Sensitive 1 (SOS1) Na⁺/H⁺ antiporter. *Nature Plants*, 9(11), 1924-1936.
44. Aamir, M., Kashyap, S. P., Krishna, R., Mishra, P., Singh, P., Ansari, W. A., ... & Singh, A. K. (2025). Advances in salinity tolerance breeding in tomato: from conventional breeding tools to genome editing. *Environmental Sustainability*, 1-19.
45. Khan, N. (2025). Molecular Insights into ABA-Mediated Regulation of Stress Tolerance and Development in Plants. *International Journal of Molecular Sciences*, 26(16), 7872.
46. Zhao, Z., Wang, L., Liu, Y., Sun, J., Xiao, J., Dong, Q., ... & Wu, J. (2024). Effects of drought on non-structural carbohydrates and C, N, and P stoichiometric characteristics of *Pinus yunnanensis* seedlings. *Journal of Forestry Research*, 35(1), 12.
47. Xie, X., Gan, L., Wang, C., & He, T. (2024). Salt-tolerant plant growth-promoting bacteria as a versatile tool for combating salt stress in crop plants. *Archives of Microbiology*, 206(8), 341.
48. Borena, F. R., & Hassen, J. M. (2022). Impacts of soil salinity on irrigation potential: In the case of Middle Awash, Ethiopian Review. *Open Access Library Journal*, 9(4), 1-18.
49. Ayars, J. E., Zaccaria, D., & Bali, K. M. (Eds.). (2023). *Microirrigation for crop production: design, operation, and management*. Elsevier.
50. Haasdijk, E., & Heinerman, J. (2018). Quantifying selection pressure. *Evolutionary computation*, 26(2), 213-235.

51. Babar, M., Khalid, M. N., Haq, M. W. U., Hanif, M., Ali, Z., Awais, M., ... & Amjad, I. (2023). 12. A comprehensive review on drought stress response in cotton at physiological, biochemical and molecular level. *Pure and Applied Biology (PAB)*, 12(1), 610-622.
52. IQRA, L., PREM, A., SANA, K., AFZAL, K., SAHER, M., AKHTAR, T., ... & GILL, S. A. (2021). Effect of salt on different morphological traits in wheat varieties. *Plant Cell Biotechnol. Mol. Biol.*, 16, 25-32.

Figure legends

Figure 1. Phylogenetic tree of SOS1 proteins in *Gossypium hirsutum* and other plant species. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The evolutionary distances were computed using the p-distance method and are in the units of the number of amino acid differences per site. This analysis involved 31 amino acid sequences. All positions containing gaps and missing data were eliminated (complete deletion option). There were a total of 493 positions in the final dataset. Evolutionary analyses were conducted in MEGA11.

Figure 2. Exon-intron structure of *GhSOS1* genes. Number of exons and introns are shown by using Tbttools-II v.2.019 showing different numbers of axons introns across length of every protein. All containing conserved domains.

Figure 3. Distribution of conserved motifs in *GhSOS1* proteins. Conserved motifs were identified using MEME Suite. Motifs common to all proteins are linked to Na⁺/H⁺ exchanger domains, crucial for salinity response.

Figure 4. Conserve motif analysis among 15 *GhSOS1* proteins. Conserved motif architecture of *GhSOS1* proteins in *Gossypium hirsutum* as identified by MEME suite. A total of 10 conserved motifs (Motif 1–Motif 10) were detected across all *GhSOS1* protein sequences, each represented by a distinct colored box. The motifs were distributed in varying patterns, with Motif 1, 2, and 9 being the most prevalent across all family members. These motifs were

primarily associated with the Na⁺/H⁺ exchanger domain, indicating their potential involvement in ion transport and salt stress response.

Figure 5. Predicted subcellular localization of *GhSOS1* proteins. Heatmap showing predicted subcellular localization of *GhSOS1* proteins across different cellular compartments. The majority of proteins are predicted to localize to the plasma membrane, with few showing possible dual localization to vacuoles and cytoplasm. Color intensity represents prediction confidence.

Figure 6. Chromosomal location of *GhSOS1* genes in the *G.hirsutum* genome. Chromosomal distribution across nine different chromosome is determined by using Mapchart tool.

Figure 7. Cis-regulatory elements detected in the promoter regions of *GhSOS1* genes revealed multiple stress- and hormone-responsive motifs. These elements suggest transcriptional regulation of *GhSOS1* genes under drought and salinity stresses, supporting their functional role in abiotic stress tolerance.

Figure8. Protein-protein interaction network of *GhSOS1* proteins. Predicted using STRING and visualized in Cytoscope, SOS1 interacts with HKT1, NHX1, and CIPK family members.

Figure 9. Predicted phosphorylation sites in *GhSOS1* proteins predicted using Net Phos 3.1. Most of the proteins have high level of serine threonine content indicating their potential role in phosphorylation.

Figure 10 Predicted transmembrane domains in *GhSOS1-1* protein. TMHMM prediction shows multiple helices indicating *GhSOS1*- nature as a membrane transporter.

Figure11. 3D structure prediction of *GhSOS1-1* protein. Built via SWISS-MODEL, the structure reveals transmembrane helices and functionally important regions aligned with Na⁺/H⁺ antiporter activity.

Figure 12. Heatmap of 15 *GhSOS1* genes showing log₂-normalized RNA-seq expression under salt and drought stress (0–24 h). Genes *GhSOS1*-4, *GhSOS1*-5, *GhSOS1*-10, *GhSOS1*-

11, and *GhSOS1*-13 displayed consistent stress-responsive patterns selected for qPCR validation.

Figure 13. Relative expression of selected *GhSOS1* genes under drought (20% PEG) and salt (400 mM NaCl) at 0, 3, 6 and 12 h. Expression was normalized to GhUBQ and calculated by the $2^{-\Delta\Delta C_t}$ method. Bars represent mean $\pm$ SE of three biological replicates (each with three technical replicates). Different letters above bars indicate significant differences at $p < 0.05$ (one-way ANOVA, Tukey's HSD).

Table legends

Table 1: Ka/Ks ratio selection pressure analysis among *GhSOS1* genes

Table 2: Physiochemical properties of *GhSOS1* proteins

Supplementary File

Table S1: The list of primers used for qPCR analysis

Declarations

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Ethics approval

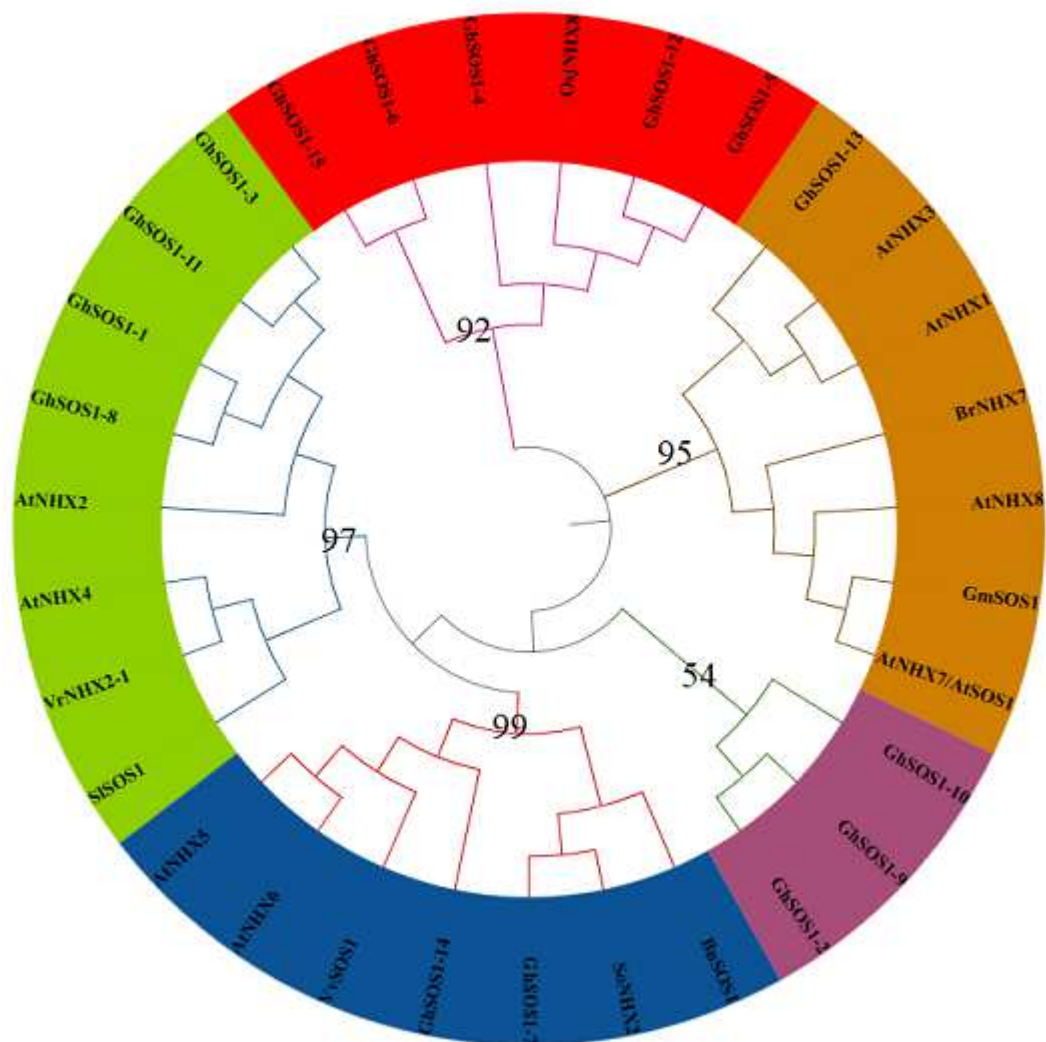
Not applicable.

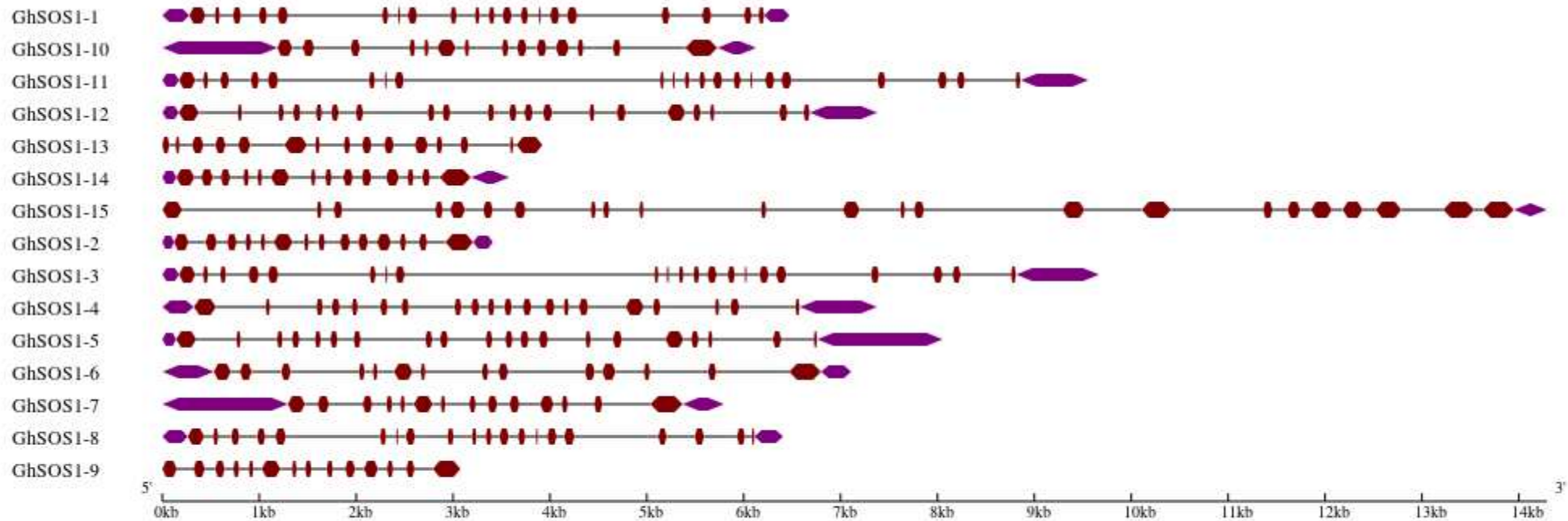
Data availability

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Author contributions

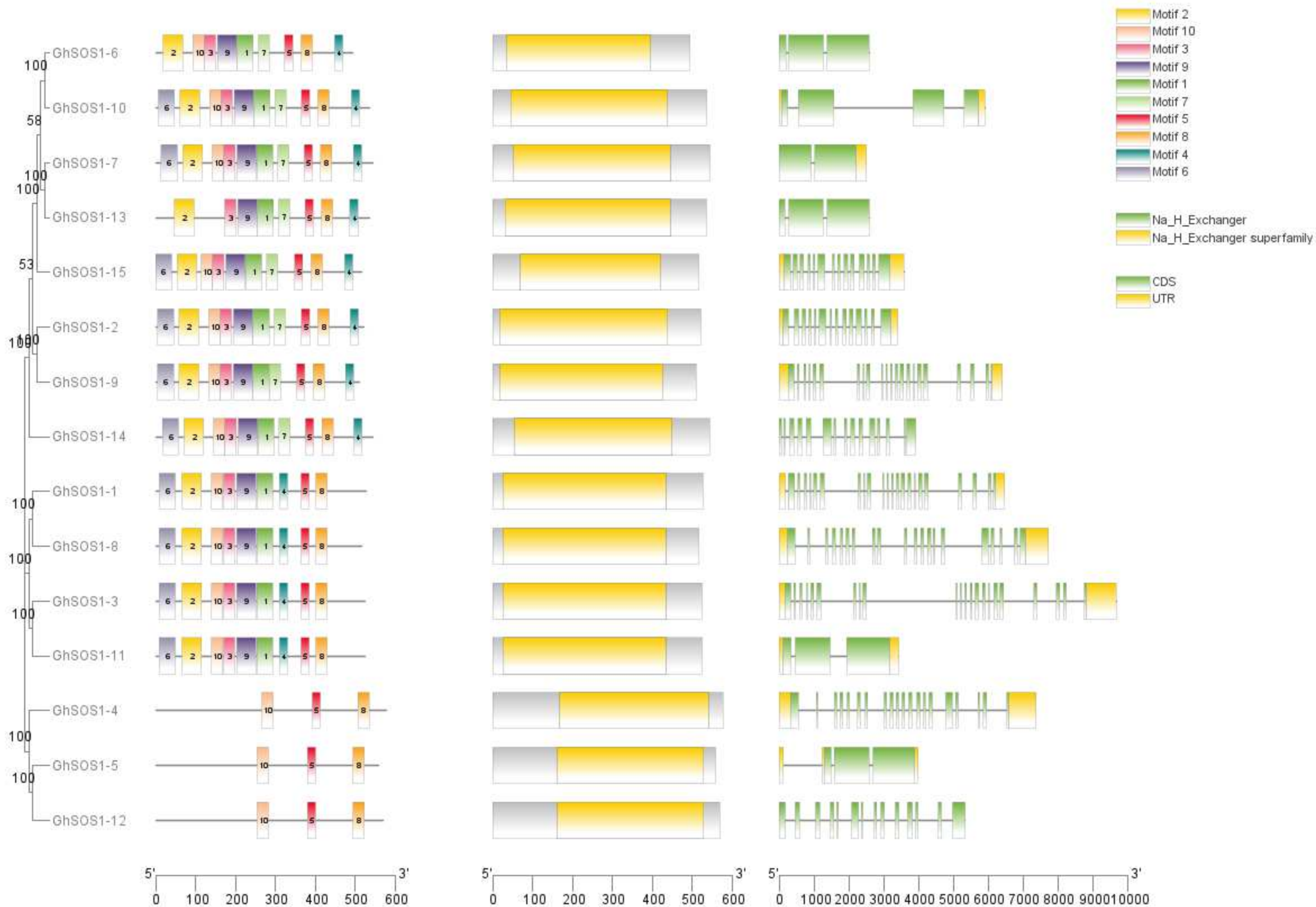
Laraib Iqra: Conceptualization, Methodology, Investigation, Writing - original draft. Muhammad Naveed Shahid: Supervision, Resources, Validation. Umar Akram: Data analysis, Software. Usman Arif: Formal analysis, Review & editing. All authors read and approved the final manuscript.

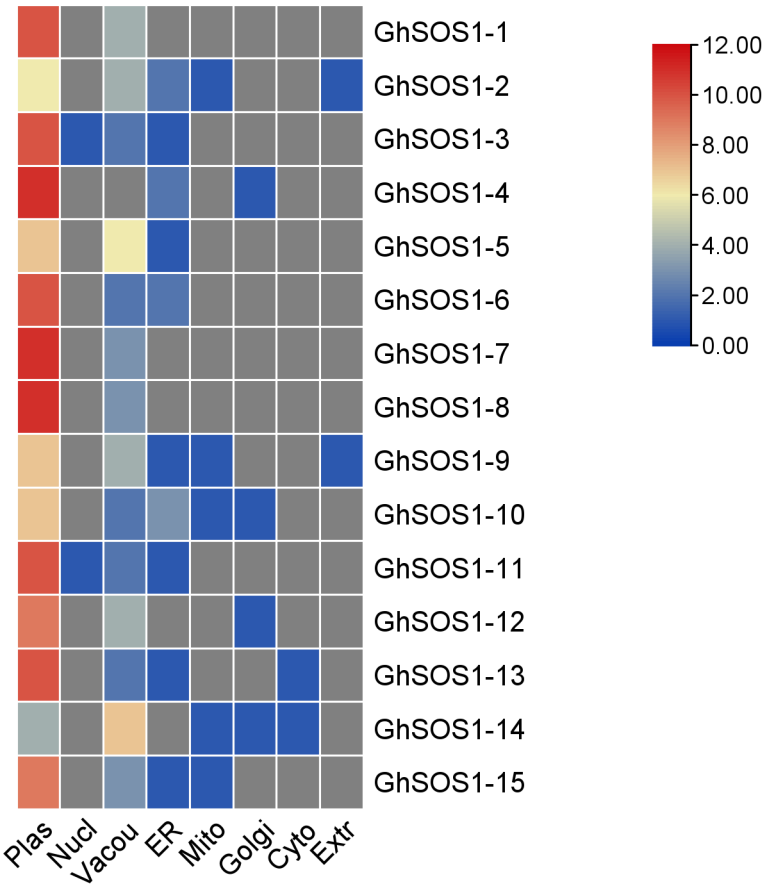




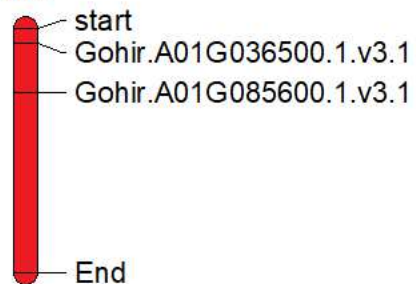
Legend:

Red diamond: CDS Purple arrow: upstream/ downstream Black line: Intron

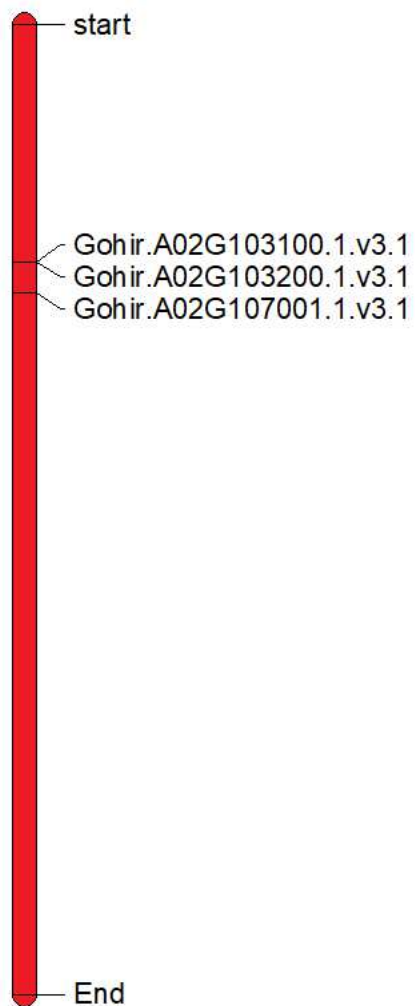




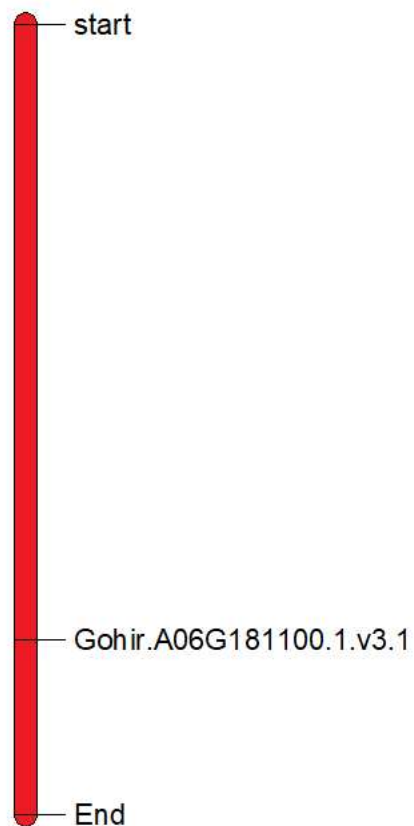
A01



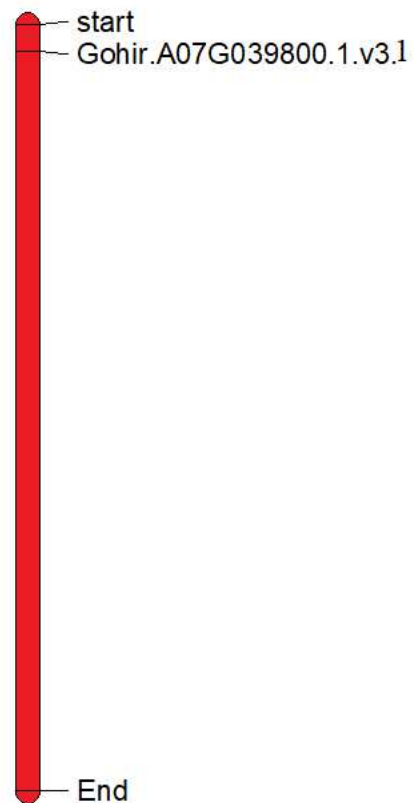
A02



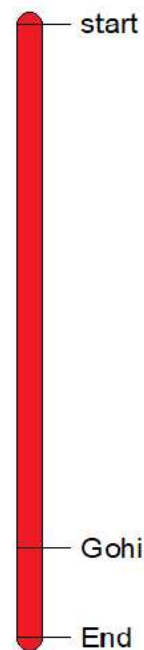
A06



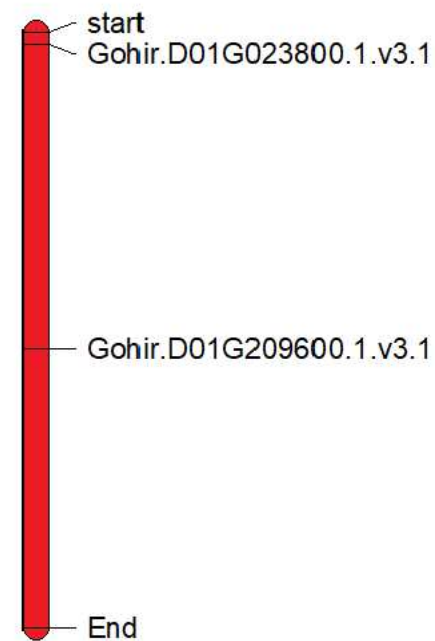
A07



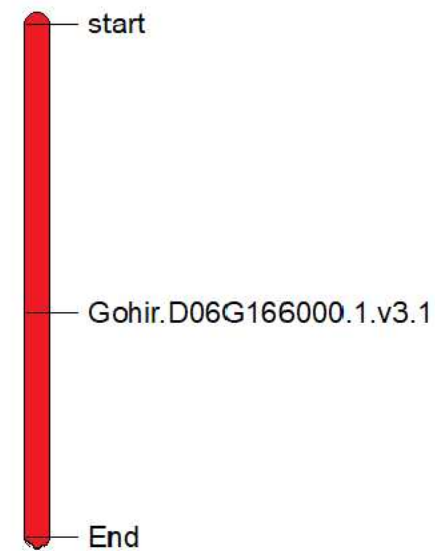
A13



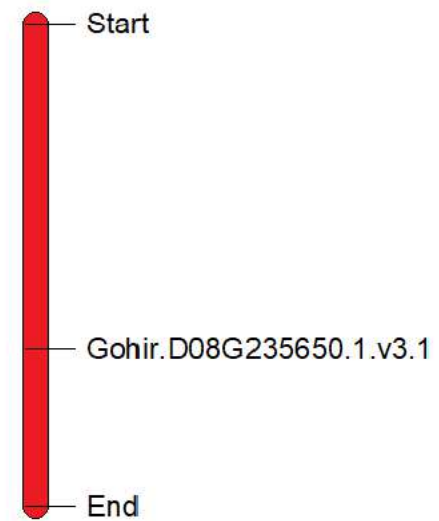
D01



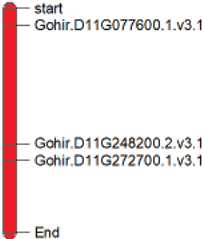
D06

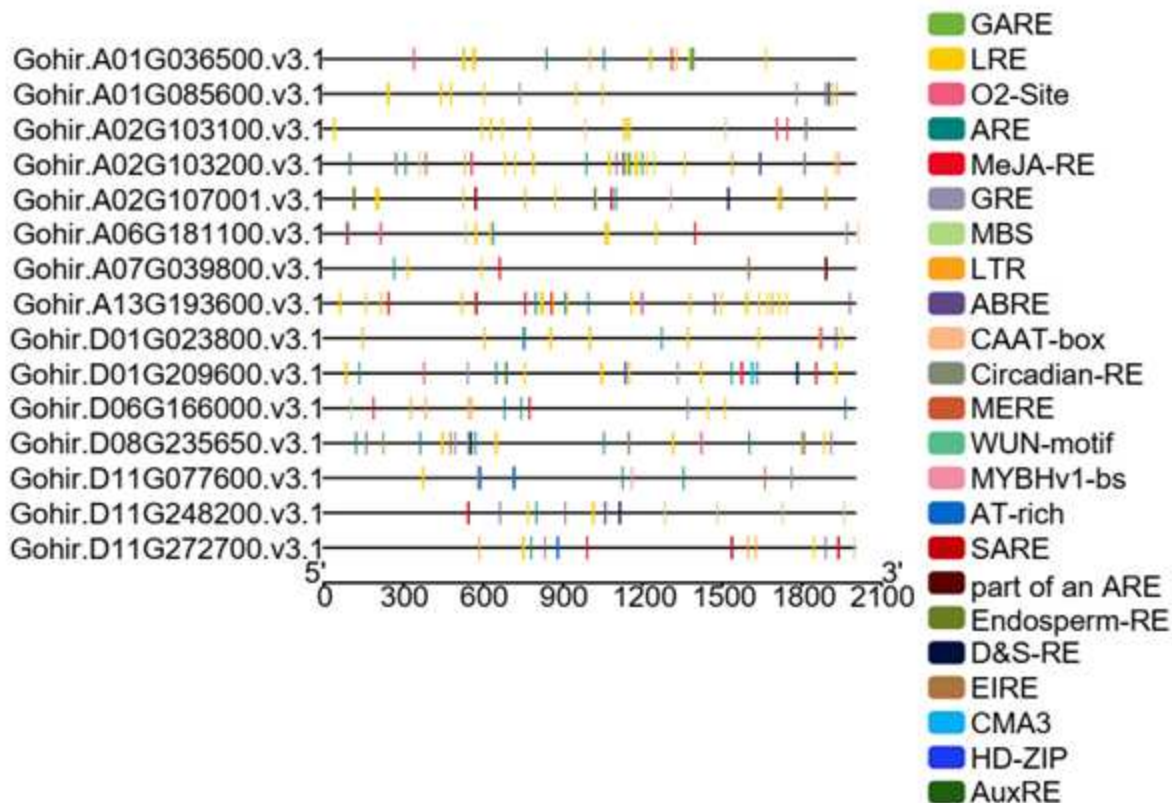


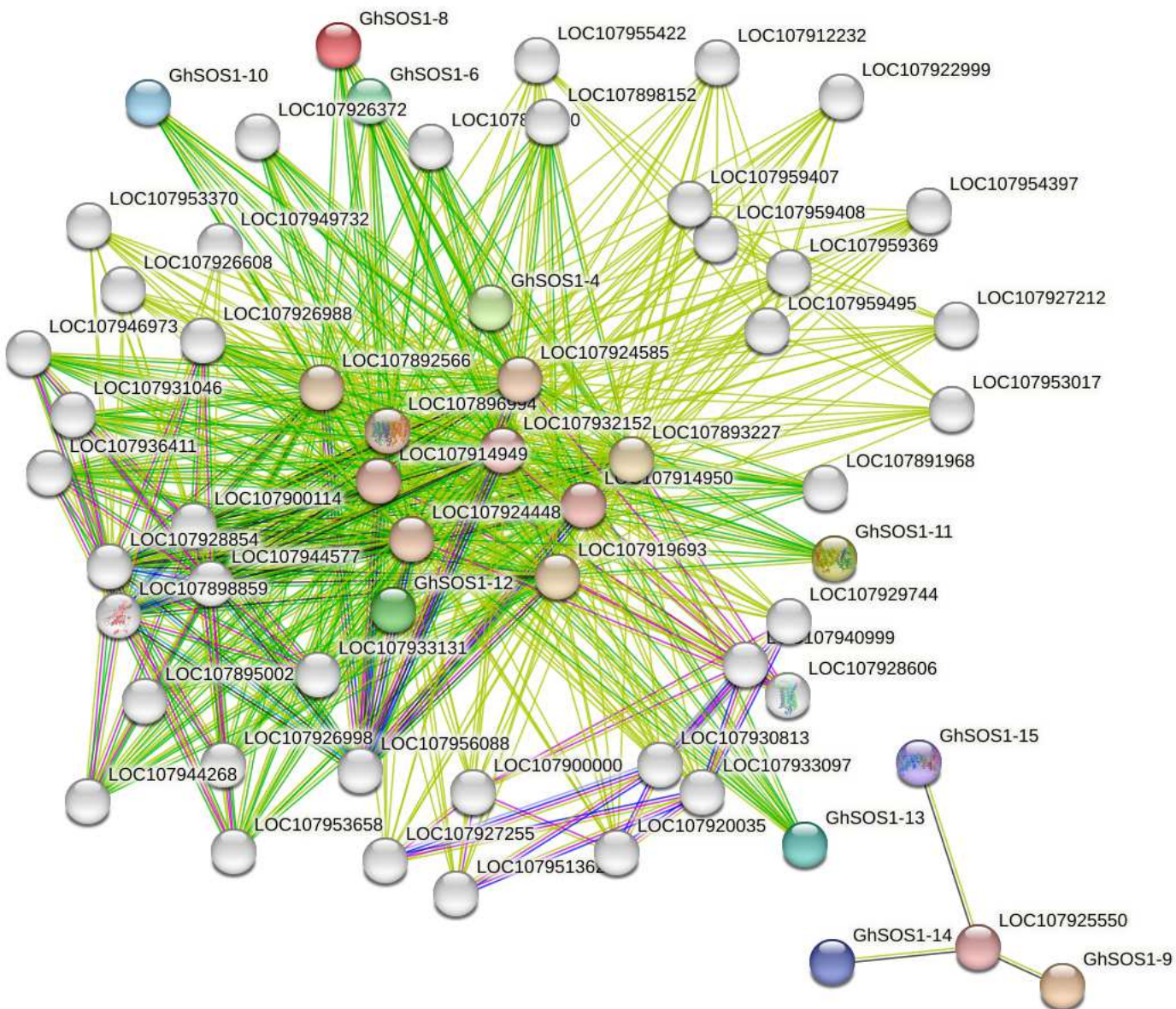
D08



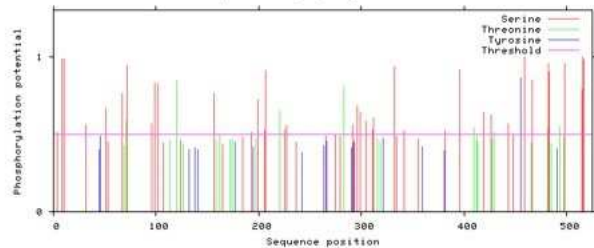
D11



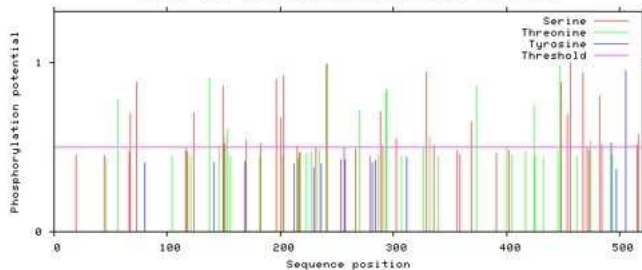




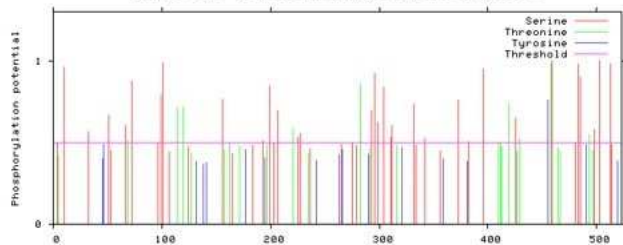
NetPhos 3.1a: predicted phosphorylation sites in GhSOS1-1



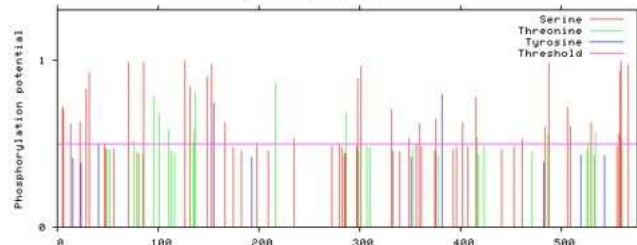
NetPhos 3.1a: predicted phosphorylation sites in GhSOS1-2



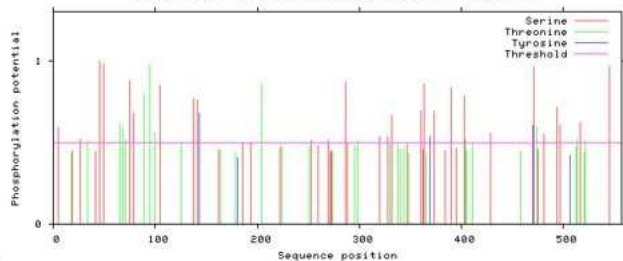
NetPhos 3.1a: predicted phosphorylation sites in GhSOS1-3



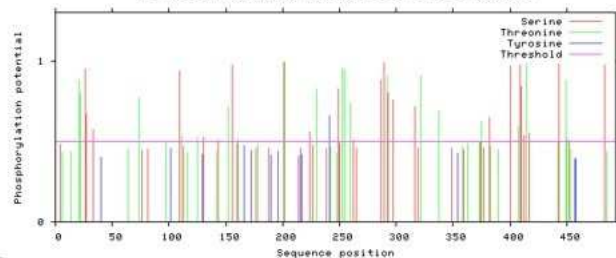
NetPhos 3.1a: predicted phosphorylation sites in GhSOS1-4



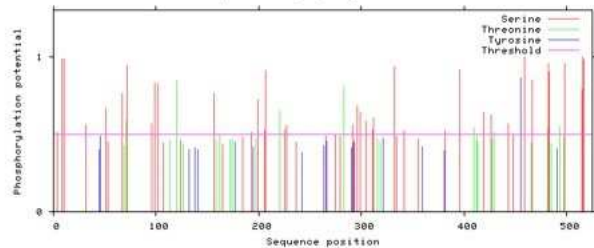
NetPhos 3.1a: predicted phosphorylation sites in GhSOS1-5



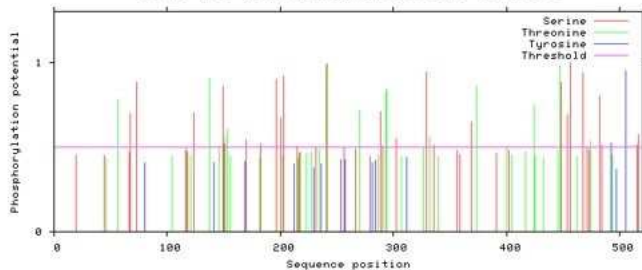
NetPhos 3.1a: predicted phosphorylation sites in GhSOS1-6



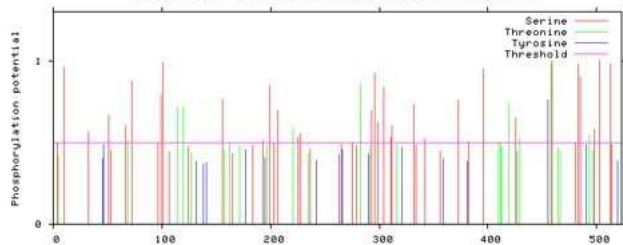
NetPhos 3.1a: predicted phosphorylation sites in GhSOS1-1



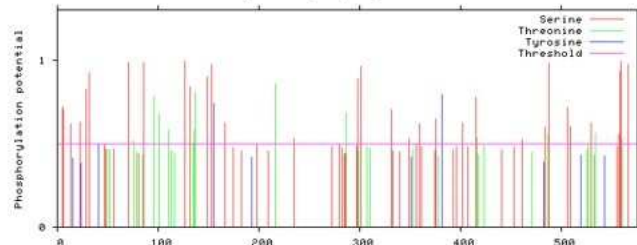
NetPhos 3.1a: predicted phosphorylation sites in GhSOS1-2



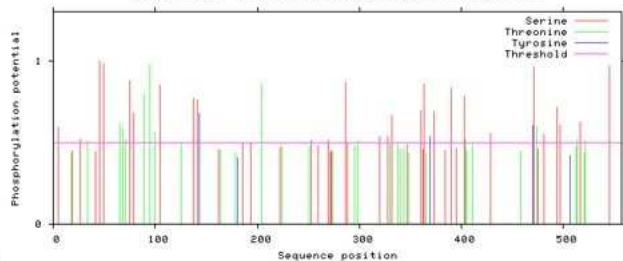
NetPhos 3.1a: predicted phosphorylation sites in GhSOS1-3



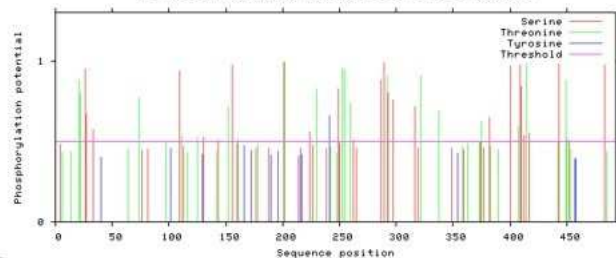
NetPhos 3.1a: predicted phosphorylation sites in GhSOS1-4



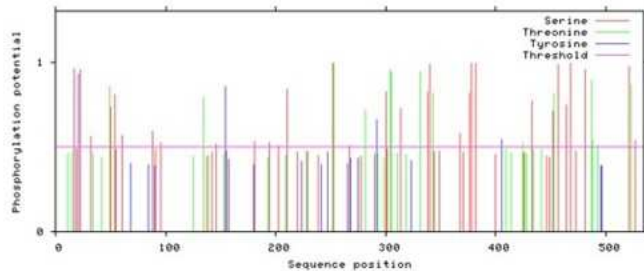
NetPhos 3.1a: predicted phosphorylation sites in GhSOS1-5



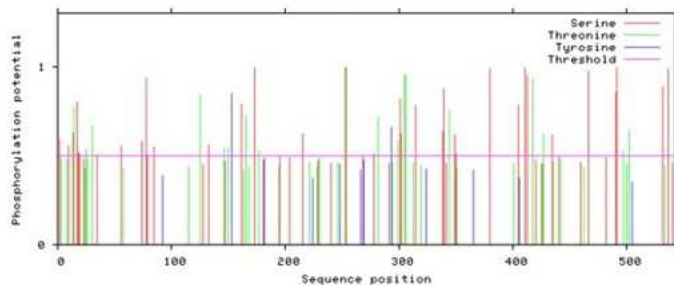
NetPhos 3.1a: predicted phosphorylation sites in GhSOS1-6



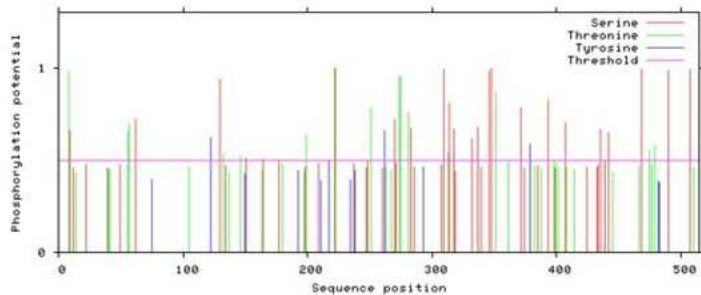
NetPhos 3.1a: predicted phosphorylation sites in GhSOS1-13



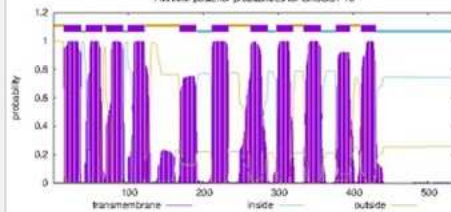
NetPhos 3.1a: predicted phosphorylation sites in GhSOS1-14



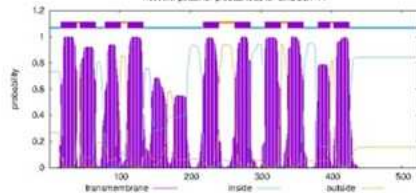
NetPhos 3.1a: predicted phosphorylation sites in GhSOS1-15



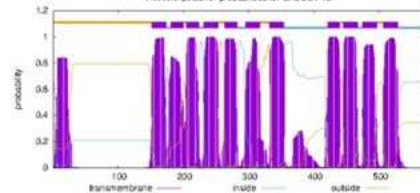
TMHMM posterior probabilities for GhSOS1-10



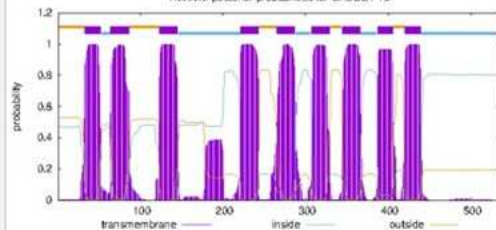
TMHMM posterior probabilities for GhSOS1-11



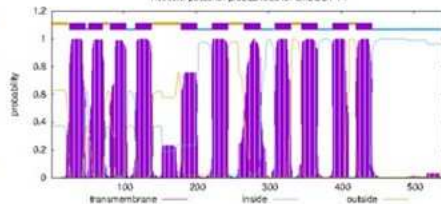
TMHMM posterior probabilities for GhSOS1-12



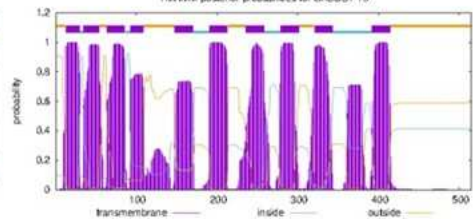
TMHMM posterior probabilities for GhSOS1-13

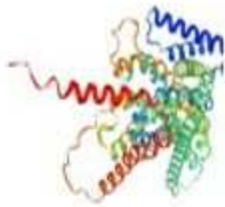


TMHMM posterior probabilities for GhSOS1-14

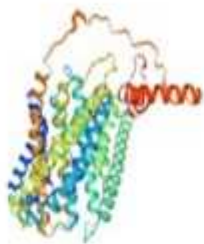


TMHMM posterior probabilities for GhSOS1-15

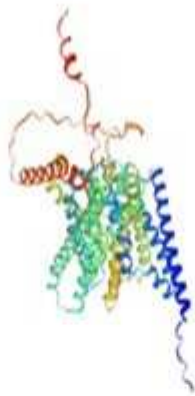




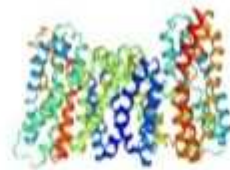
GhSOS1-1



GhSOS1-2



GhSOS1-3



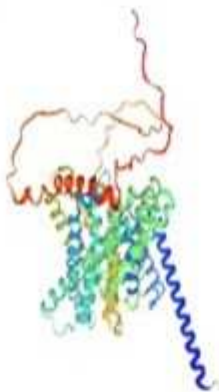
GhSOS1-4



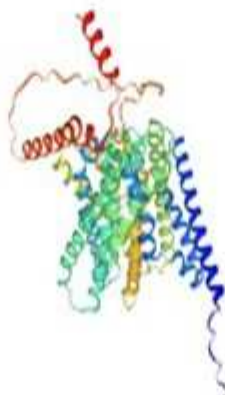
GhSOS1-5



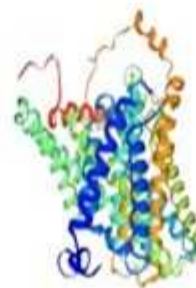
GhSOS1-6



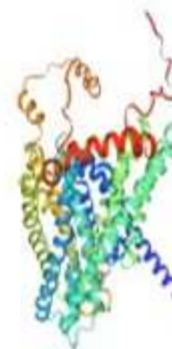
GhSOS1-7



GhSOS1-8



GhSOS1-9



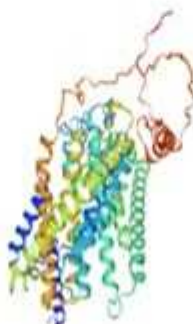
GhSOS1-10



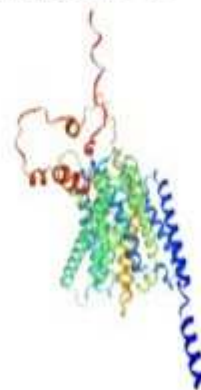
GhSOS1-11



GhSOS1-12



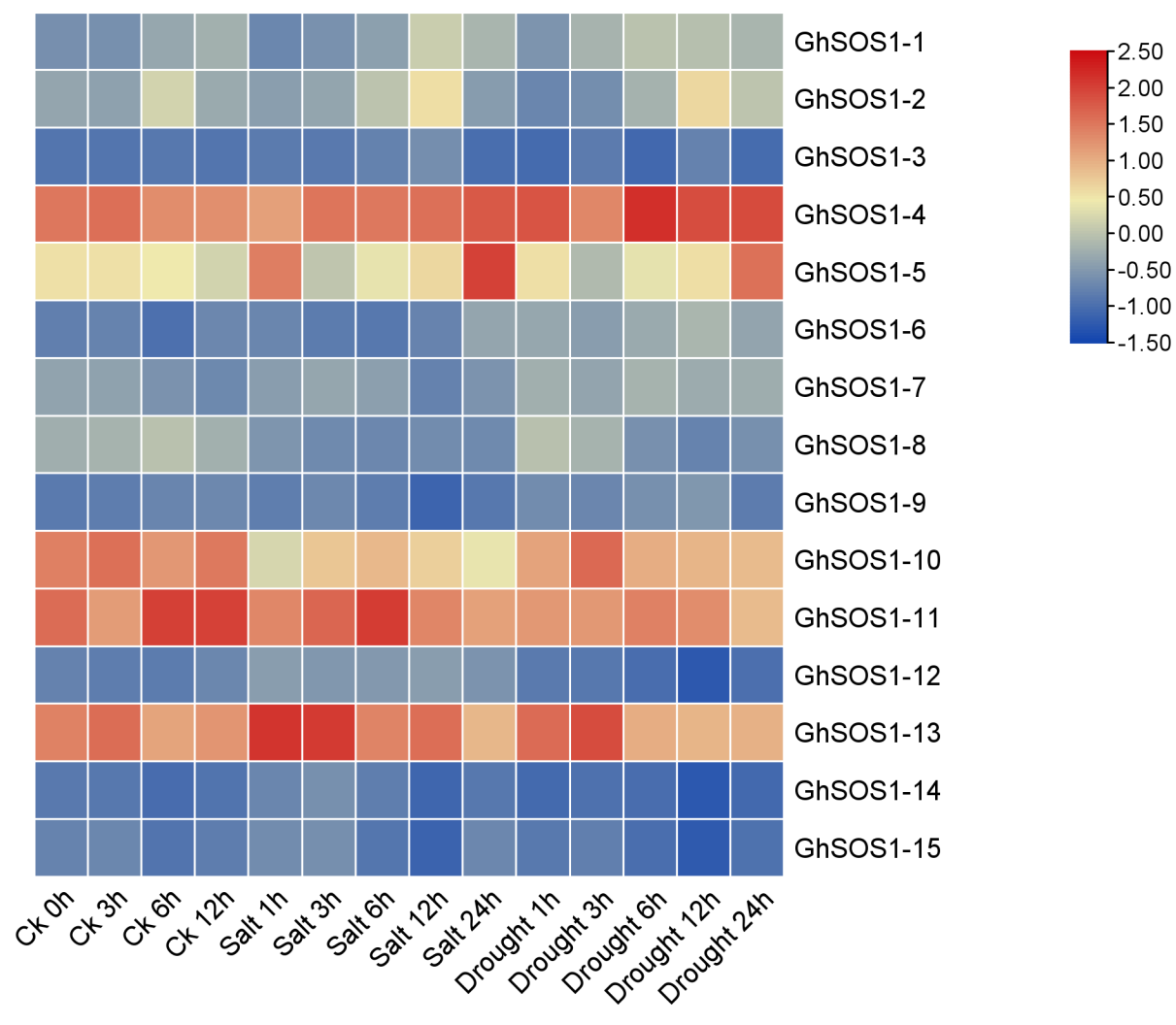
GhSOS1-13



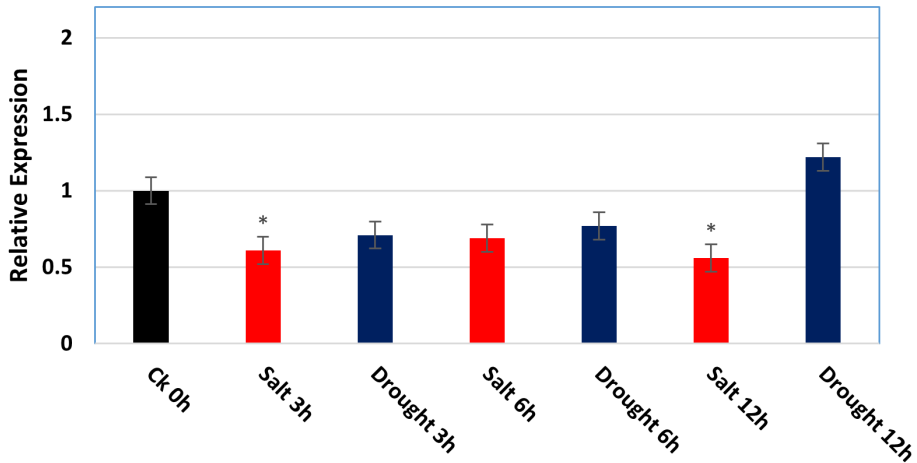
GhSOS1-14



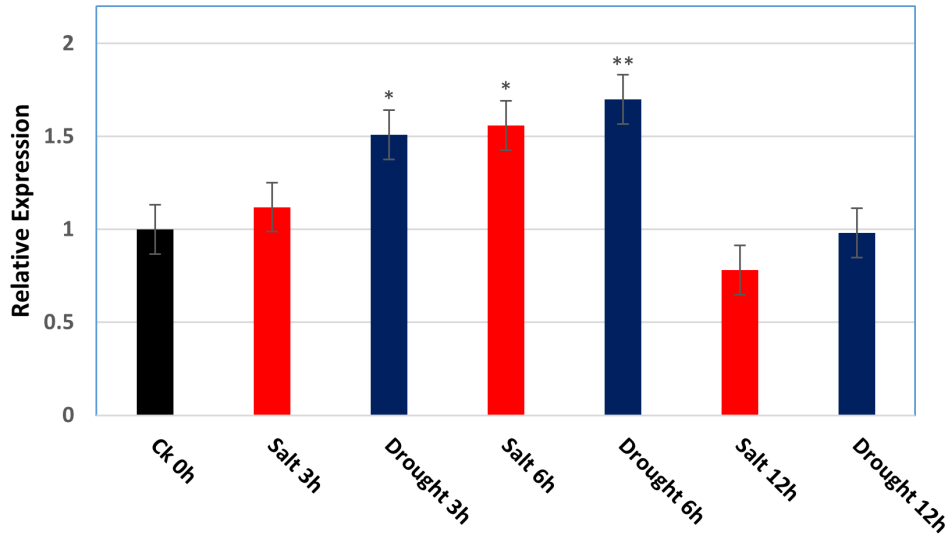
GhSOS1-15



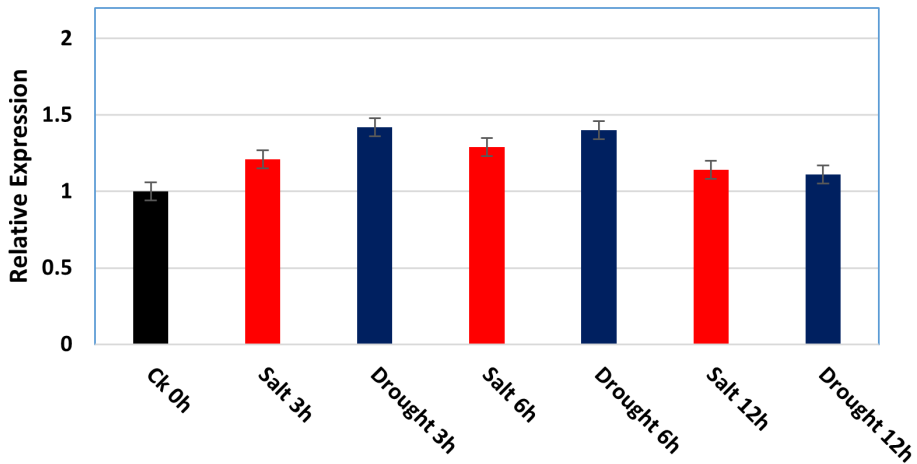
GhSOS1-4



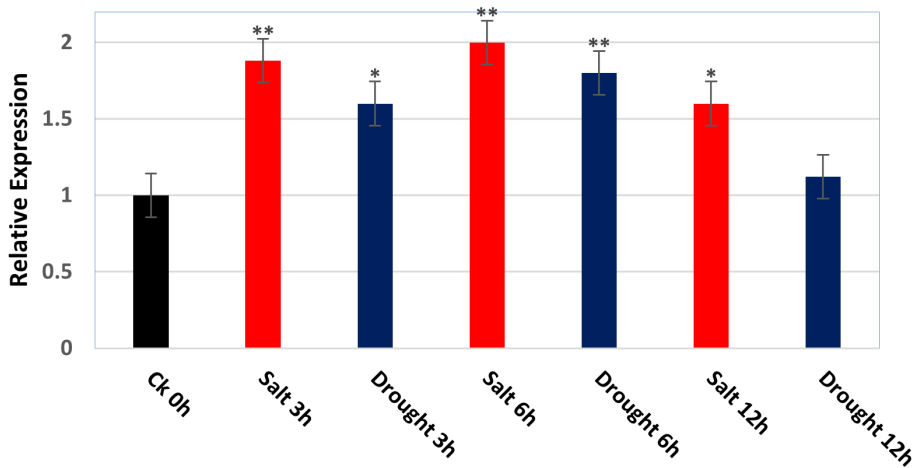
GhSOS1-5



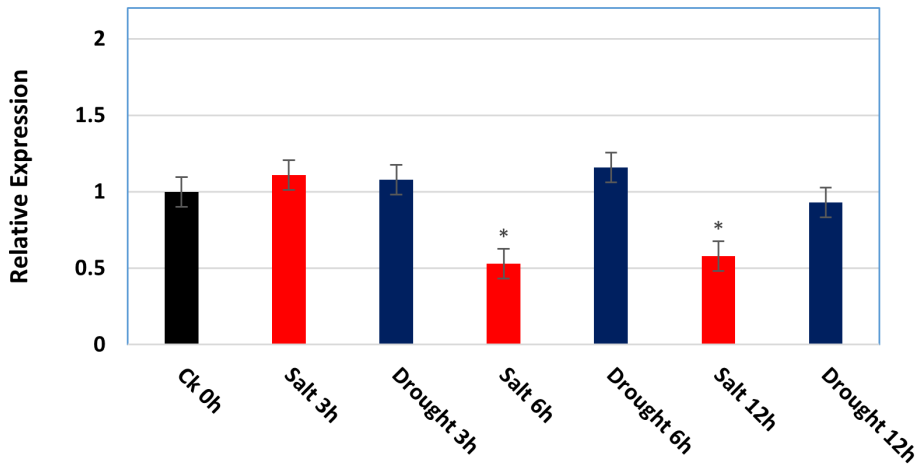
GhSOS1-10



GhSOS1-11



GhSOS1-13



Supplementary Files

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

- [Table2.xlsx](#)
- [Table1.xlsx](#)
- [TableS1.docx](#)