In Silico Analysis of SOS Family in Areca palm (Areca catechu. L)

Osama Alam (osamaalam9988@gmail.com)
University of Science & Technology Bannu

Research Article

Keywords: SOS gene, Areca catechu, Salinity

Posted Date: November 6th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3156169/v2

Additional Declarations: No competing interests reported.

Abstract

The salt-overly sensitive (*SOS*) gene is a signaling gene that helps plants to respond to high salinity stress. It is activated when cells are exposed to high levels of sodium ions, and it leads to a number of changes in gene expression that help the plant cope with high salinity stress. However, The *SOS* gene in Arecha palm has not been studied, despite being present in various plant species. We have identified 10 members of the *AcatSOS* family of genes, which are dispersed randomly over 5 chromosomes in the genome of Areca catechu. Using the phylogenetic analysis of 10 *AcatSOS* genes and two additional species rice and Arabidopsis—the tree is divided into 5 subfamilies: Group 1, Group 2, Group 3, Group 4 and Group 5. Each gene in the *AcatSOS* family was identified on its physiochemical properties. The plastid and vacuole contain most of the localization signals. Furthermore, there were numerous cis-acting areas and light-responsive elements in the promoter sequences of the *AcatSOS* genes. After the Areca palm genome was sequenced, segmental and tendom duplications were discovered, providing evidence of a major driving force in the evolution of the *AcatSOS* gene family. Areca palms share numerous genes, in terms of orthologous relationships, with other monocotyledonous plants because of their evolutionary affinity and related heritage. The development of the leaf, endocarp, and epicarp is regulated by the *AcatSOS* genes, according to circular heat. *AcatSOS*-4, *AcatSOS*-6 and *AcatSOS*-8 are highly regulated in plants. Genome-wide characterization of the *SOS* gene expression patterns has revealed that the expression of these genes is regulated in a complex manner. It is influenced by several factors, including the level of salinity stress, the developmental stage of the plant, and the tissue type.

1. Introduction

To survive, all organisms must maintain cellular homeostasis, a stable internal environment that is balanced about many factors, including temperature, pH, and ions. Exposure of an organism to abiotic stresses, such as high salinity, can lead to the excessive accumulation of certain molecules, which can disrupt homeostasis or cause it to fail [1]. Excessive salt can inhibit enzyme activity, damage metabolism, and impair cell growth and development [2]. Previous studies have found that salt stress delays the differentiation and growth of plant organs and tissues via three main aspects, namely, ion toxicity, osmotic stress and nutritional imbalance, which further inhibit the growth and development of whole plant [3]. Plant salt tolerance mainly involves three aspects: cell homeostasis maintenance (including ion and osmosis), detoxification (such as removal of reactive oxygen species) and growth regulation (such as cell division and stretching) [3, 4]. When plants are exposed to salt stress, the salt overly sensitive (SOS) pathway is activated. This pathway helps to regulate ion homeostasis and improve the sodium tolerance of plants [4]. The salt overly sensitive (SOS) signaling pathway was first discovered in the model plant *Arabidopsis thaliana* [5], SOS pathway has been shown to be active in a variety of plants, including rice and tomato [6, 7]. The SOS pathway is a signaling pathway that helps plants to respond to high salinity stress. It comprises of three genes: SOS1, SOS2, and SOS3. SOS1 encodes a Na+/H + antiporter, which is a protein that helps to pump sodium ions out of the cells. SOS2 and SOS3 do not directly transport Na+, but instead regulate the ion-transporting activity of SOS1 [8]. The SOS protein family in Arabidopsis includes SOS2 and SOS3, which are located in the cytoplasm. They regulate the homeostasis of K + and Na + inside and outside the cell by regulating SOS1 on the plasma membrane. This can increase the salt tolerance of plants. The SOS1 gene is a salt-tolerance effector gene that encodes a plasma membrane Na+/H + antiporter. The activity and expression of SOS1 are controlled by the Ca2+-responsive SOS3-SOS2 protein kinase complex [9]. Some research suggests that the SOS3-SOS2 kinase complex may directly increase the activity of the Na+/H + exchanger, which could then regulate the flow of sodium through the cell membrane [3]. Moreover, AtHKT1, an Arabidopsis Na+-preferential transporter, primarily controls Na + delivery to the roots and shoots [10]. The hypersensitivity of SOS3 and hkt double mutants to Na + was significantly lower than that of SOS3 single mutants, suggesting that the SOS3-SOS2 complex may negatively regulate Na + delivery. In addition, the activities of H+-ATPase, H + PPase, and Na+/H + reverse transporters on the vacuolar membrane may also be regulated by the SOS3-SOS2 complex [11].

Areca catechu is commonly known as areca palm, is one of the most significant tropical industrial crops [12]. Areca catechu (2n = 32), also called areca nut or betel nut, is a member of the genus Areca in the family Arecaceae. It is a tropical plant that

is commonly grown in the tropics and has many uses in both industry and agriculture. The genome of A. catechu was sequenced in 2018 and was found to be 2.59 Gb in size. The genome contained 31,571 protein-coding genes, of which 92.92% were functionally annotated. The genome also contained a large number of transposable elements, which are mobile genetic elements that can insert themselves into the genome. The areca genome is enriched in genes involved in the biosynthesis of flavonoids, anthocyanins, and monoterpenoids. These compounds are responsible for the characteristic taste and aroma of areca nuts. The genome also contains genes involved in the defense against pathogens and pests. The sequencing of the areca genome has provided a valuable resource for researchers studying the biology and genetics of this important plant. The genome can be used to identify genes that are involved in the development, growth, and reproduction of areca nuts. It can also be used to develop new varieties of areca nuts with improved traits, such as resistance to diseases and pests [13].

In this study, we performed a comprehensive genome-wide analysis of the SOS family in the Areca palm genome. We identified 10 *AcatSOS* members, which were mapped to 6 chromosomes. The SOS1 gene was identified, and phylogenetic relationships were established with the previously reported *AcatSOS* proteins of rice (Oryza sativa L.), Arabidopsis thaliana, and Areca catechu. The Areca palm genome has recently been sequenced with high chromosomal resolution. This allows for the empirical inquiry of selected gene families with significant functions based on the entire genome. The purpose of the current study was to provide the groundwork for future research on the function of *AcatSOS* genes by thoroughly analyzing the traits, evolution, and expression of the SOS1 gene in Areca palm. We found that the SOS1 gene is expressed in a variety of tissues and at different developmental stages. The expression of SOS1 was also induced by salt stress. These findings suggest that SOS1 plays an important role in the response of Areca palm to salt stress. Our study provides a foundation for future research on the function of *AcatSOS* genes in Areca palm. By understanding the function of these genes, we can develop new strategies to improve the salt tolerance of this important crop.

2. Materials and Methods

2.1. Identification of SOS Genes in Areca palm genome

The genome of the Areca palm was downloaded from the Areca palm genome database, http://arecaceae-gdb.com/#/ accessed on 22 June, 2023 [14]. The NCBI website was used to find all recently available research data on *SOS* proteins, https://www.ncbi.nlm.nih.gov/protein (accessed on 22 June, 2023), and Phytozome, https://phytozome-next.jgi.doe.gov/ (accessed on 22 June 2023) [15]. The InterPro tool, https://www.ebi.ac.uk/interpro/ (accessed on 22, June 2023), was required to determine the *SOS* proteins' domains. Previous research revealed the described sequence data for tomatoes (Solanum lycopersicum), and corn (Zea Junes L.) Arabidopsis thaliana. For finding the *SOS* physiological and chemical characteristics Expasy ProtParam Tool, https://web.expasy.org/protparam/ (accessed on 23 June, 2023) was used to compute the protein's molecular weight, Pi (isoelectric point), and GRAVY (grand average of hydropathicity).

2.2. Identification of AcatSOS Protein Domains

The CDD NCBI tool https://www.ncbi.nlm.nih.gov/Structure/cdd/ wraps.cgi (accessed on 23 June 2023) was used to detect conserved domains in all 10 *AcatSOS* protein sequences, Using TBTools [16]. the collected information was exploited to illustrate the *SOS* domain's structure. https://www.genome.jp/tools/motif/ (accessed on 24 June, 2023) is a motif finder that was further utilized to calculate the relevant themes.

2.3. Phylogenetic Analysis of the AcatSOS Family Using Multiple Alignments

From Phytozome-13, all *SOS* protein sequences for rice (Oryza sativa—v7.0) and Arabidopsis (A. thaliana thair 7) were retrieved (https://phytozome-next.jgi.doe.gov (accessed on 24 June, 2023) [15], the "Areca palm -genome-website" was used to extract Areca palm protein sequences" [15]. For the molecular evolutionary genetic study(MEGA-11), the cluster was used

to align all 30 protein sequences [17]. Phylogenetic analysis using the standard statistical tree test and 1000 bootstrap repetitions was performed on the aligned protein sequences. iTOL software [18]was used to design and visualize the tree.

2.4. Chromosomal distribution and circus analysis of AcatSOS70 Genes on the Areca palm

The Areca palm genomic library was used to collect information on all 10 *SOS* genes from the areca database http://arecaceae-gdb.com/#/ (accessed on 25 June, 2022) [16] With TBTools, the chromosome length was determined "FASTA stats" [14]. Then, by using PhenoGram software chromosome ideogram was generated, http://visualization.ritchielab.org/phenograms/plot (accessed on 25 June, 2023). The output image was used to mention all *AcatSOS* genes on the chromosome. A multiple collinearity scan toolbox was used to examine the duplication analysis (MCScanX) [19] having an e-value of 1 × 10 − 5. *AcatSOS* gene-related duplication events were chosen and shown using CIRCOS, and a connecting line connected the potential duplication genes.

2.5. Characterization of Gene Structure, Conserved Motif, and Sub-cellular Localization

The gene structure of *AcatSOS* was determined by downloading CDS and Genomic data fromAreca palm database http://arecaceae-gdb.com/#/ (accessed on 25 June, 2023). The MEME suite program was used to analyze the 10 *AcatSOS* genes and identify conserved motifs. The default parameters were used to identify the maximum 10 motifs at https://meme-suite.org/meme/tools/meme (accessed on 25 June, 2023). Subcellular localization of *AcatSOS* was predicted by Wolf PSORT (https://www.genscript.com/wolf-psort.html, accessed on (25 June, 2023) and CELLO v.2.5 (http://cello.life.nctu.edu.tw/ (accessed on 25 June, 2023).

2.6. Genes Expression of AcatSOS genes

The Areca palm genomic library was used to determine the *AcatSOS* gene family's expression profile http://arecaceae-gdb.com/#/tools/efp-gene (accessed on 25 June, 2023) [14]. Different parts of plant tissues have been used to assess the expression levels of all 10 *AcatSOS* genes in different plant parts: root, stem, leaf, fruit, endocarp, mesocarp and epicarp. These all data wwereobtained from **eFP**-Gene tool that is present in the Areca palm database.

2.7. Gene Correlation networks of AcatSOS genes

The data of all 10 *AcatSOS* genes' interactions with other genes were obtained from Gene correlation network http://arecaceae-gdb.com/#/tools/cor-network (accessed on 25 June, 2023) [14]. All of the discovered and interacting genes' data were utilized to construct and visualize the network using the Cytoscape software [20].

2.8. Analysis of Promoter Regions in AcatSOS Genes

All the promoter sequences of the *AcatSOS* genes file (3000 bp upstream of the start codon) were obtained from the Areca palm database http://arecaceae-gdb.com/#/tools/sequence-fetch (accessed on 26 June, 2023). TBTool was used to draw various elements in the promoter region from the Areca palm genomic [16]. The cis-acting regulatory elements were retrieved from the PlantCARE library [21].

2.9. KEGG Enrichment and Go Enrichment Analysis of AcatSOS

All the 10 *AcatSOS* genes are introduced to the Areca palm database of KEGG(Kyoto Encyclopedia of Genes and Genomes) and GO(Gene Ontology) analysis http://arecaceae-gdb.com/#/tools/kegg-enrichment (accessed on 26 June, 2023).

3. Result

3.1. Characterization of the SOS Family in Areca Palm by Genome-Wide Analysis

From the genome of the Areca palm, 10 *AcatSOS* genes were identified through genome-wide analysis, and these genes have been classified as *AcatSOS*-1 to *AcatSOS*-10. The table is organized by chromosomal number, with each gene listed in its corresponding location on the chromosome. The CDS length is the number of nucleotides in the gene's coding sequence, and the protein length is the number of amino acids in the protein that the gene encodes. The protein molecular weight is the calculated weight of the protein, and the pI is the isoelectric point, which is the pH at which the protein has no net charge. The GRAVY score is a measure of the gene's predicted stability, with higher scores indicating more stable proteins. The range of **Amino acid length** and molecular weight, from 195 (*AcatSOS_8g000210*) to 254 AA (*AcatSOS_10g003600*), and **molecular weight** ranged from 21060 kDa (*AcatSOS_8g000210*)) to 71482.68 kDa (*AcatSOS_10g003600*). Due to the 10 Areca palm *AcatSOS* proteins' representative **GRAVY** values being less than 0, hydrophilicity was anticipated for all of them ranging from 0.403 (*AcatSOS_8g000210*) to -0.244 (*AcatSOS_1g019830*). All *AcatSOS* proteins comprised varying degrees of **pI** values, ranging from 4.77 (*AcatSOS_1g009860*) to 9.16 (*AcatSOS_10g003600*), with an average value of 6.466

Table 1. Physiochemical properties of *SOS* genes in the Areca palm (Areca catechu. L)

Transcript ID	*RenameID*	Chr No	Start Position	End Position	A.A Length	M w (kda)	Pi	GRAVY	CDS
Acat_1g009860	*AcatSOS*-1	1	90846470	90857128	212	23929.8	4.77	0.565	2953
Acat_1g019830	*AcatSOS*-2	1	163147651	163000000	203	22114.79	5.62	0.801	2465
Acat_1g017970	*AcatSOS*-3	1	125883653	126000000	557	61807.54	8.65	0.559	2093
Acat_6g002750	*AcatSOS*-4	6	65473952	65522354	516	56962.24	6.76	0.52	1934
Acat_8g000210	*AcatSOS*-5	8	2484041	2521105	195	21060.2	4.97	0.403	1839
Acat_8g002990	*AcatSOS*-6	8	68924847	68949552	458	50069.78	8.19	0.548	1781
Acat_10g013220	*AcatSOS*-7	10	45521591	45530482	288	32120.47	5.58	0.648	996
Acat_10g003600	*AcatSOS*-8	10	12033769	12048297	654	71482.68	9.16	0.463	808
Acat_10g009410	*AcatSOS*-9	10	32014283	32040464	538	58608.1	5.79	0.477	639
Acat_13g008520	*AcatSOS*-10	13	26536273	26561123	305	33406.2	5.17	0.676	588

pI; isoelectric point, **GRAVY**; grand average of hydropathy, **A.A;** Amino acid length, **M w;** Molecular weight

Additionally, domain-based analyses of all 10 *AcatSOS* proteins were performed using an NCBI domain search. The domain structure was demonstrated using the obtained information and TBTools. The *SOS* superfamily was present on 10 chosen sequence data chosen, according to this domain-based study (Fig. 1).

Figure 1. Domains of *AcatSOS* proteins ranged from *(AcatSOS-1 to AcatSOS-10).***Green** color represents Na-H Exchanger, **Yellow** shows the Na-H Exchanger superfamily and the **Blackline** shows protein sequences that are not conserved. It was designed inTB tool [16].

3.2. Analysis of Phylogenetic Tree of SOS Gene

A phylogenetic tree was constructed using protein transcripts from known species that were gathered from earlier studies on things like Arabidopsis, Areca catechu, and rice. Using the Phytozome database and the published *SOS* protein sequences from three distinct species, all *SOS* proteins from each species' genomes were obtained. The phylogenetic tree was constructed of 27 protein sequences altogether, including sequence data from the other three species and 10 *AcatSOS*

protein sequences from the Areca palm genomes. On the bases of previous studies, we divide the comparative phylogenetic tree based on Conserved motif into 5 subfamilies: Group 1, Group 2, Group 3, Group 4, and Group 5. The data is available in (Supplementary File S1).

Figure 2. Comparative phylogenetic analysis of *AcatSOS* genes along with other three species, including rice, areca catechu, and Arabidopsis. It is classified based on the conserved motif [17]. Each group is differentiated through various color: 1; Green, 2; Yellow, 3; Orange, 4; Blue, and 5; Purple. It was illustrated through ITool software and the boundaries were designed through Adobe Illustrator.

3.3. Analysis of AcatSOS Genes Structure

The *AcatSOS* gene structures were identified by comparing the intron and exon structures of each gene (Fig. 3). The majority of gene sequences showed introns and exons that show the phylogeny and evolution of their gene family .

Figure 3. Gene structure of *AcatSOS*-1 to *AcatSOS*-10. **The red** color shows exon structure. While, **Blue** color line shows Intron in the gene structure. It was designed by taking CDS and Genomic data of *AcatSOS genes* from Areca database and run in Gene Structure Display server https://mhttp://gsds.gao-lab.org/ (accessed on 25 June, 2023)..

The Motif of *AcatSOS*-1 to *AcatSOS*-10 was constructed by using the MEME software, and each *SOS* protein was checked for homologous motifs. Some Motifs were found in almost all of the *AcatSOS* proteins. However, certain genes do not consistently include motifs (Fig. 4). Most of the SOS proteins in the group had nearly identical motifs. These findings also helped us to better *SOS* proteins and the close evolutionary connections among the same protein family. We make a classification of *AcatSOS* genes based on conserved motif and divide them into 5 subfamilies.

Figure 4. Phylogenetic and Motif analysis of *AcatSOS* genes [17]. Five subfamilies according to previous literature. Group 1 and Group 4 contain 1 gene, Group 2 and Group 5 contain 3 genes, and Group 3 contain 2 genes.

Figure 5. Show the motif color sequences ranging from 1 to 10 https://meme-suite.org/meme/tools/meme (accessed on 25 June, 2023). Each color has own specific sequence that is conserved among the motifs. It was obtained from the meme suit website. Motif logo data is available in (Supplementary File S3).

3.4. Chromosomal Distribution and Circus Analysis of the AcatSOS Genes

In the Areca palm genome, the *AcatSOS* genes were structurally placed on 5 chromosomes(chr). Using the data from the Areca palm genome database, each Areca palm species was given a specific *SOS* gene http://arecaceae-gdb.com/#/ (accessed on 1 June 2023). The chr length and position of each gene in the Areca palm genome are displayed in 5 chrs. The data is available in (Supplementary File S2).

Figure 6. Chromosomal distribution of *SOS* genes on 5 Areca palm chromosomes. It is visualized in http://visualization.ritchielab.org/phenograms/plot (accessed on 25 June, 2023). **White** color shapes represent Chromosome and the names of the genes are in **Black**. The distribution of the *AcatSOS* genes was variable. Chr 1 and 10 contained the majority of the 3 *AcatSOS* genes. There is 1 *AcatSOS* gene in Chr 13 and 6. Chr 8 contains only two genes.

On Areca palm chromosomes, there were 10 *AcatSOS* genes spread irregularly, with four, three, two, and one genes, respectively. We examined whether gene duplication events occurred in the genome of the Areca palm

Figure 7. Graph showing the circus analysis of *AcatSOS* proteins. Events of segmental and tandem duplication are shown by colored lines. There is 1pair of tendon duplications: *AcatSOS*-1/ *AcatSOS*-2 and 1 pair of segmental duplication: *AcatSOS-9/ AcatSOS-10* [19].

3.5. Gene Localization of AcatSOS genes

In gene localization, we predict the function of *AcatSOS* genes in cell organelles; Nucleus, Chloroplast, Peroxisome, Mitochondria, etc. From the result, we concluded that most of the genes are involved in chloroplast and cytoplasm.

Figure 8. *AcatSOS* heat map in quantitative form. It was illustrated in TB tool[16]. The color bar with **Red** signifies a high concentration of proteins, **Blue** represents a small value and **Yellow** shows an intermediate intensity. Most of the genes are localized in Plastid and Vacule. The gene localization data is available in (Supplementary File S4).

3.6. Synteny Assessment of AcatSOS genes

Numerous conserved regions/genes on the Areca plant's chromosome number were discovered using synteny analysis. We have analyzed the synteny assessment of Oryza sativa and Arabidopsis thaliana with Areca catechu.

Figure 9. Synteny analysis of **(A)***Arabidopsis thaliana* and *Areca catechu***(B)***Oryza sativa* and *Areca catechu*, **Orange** color Oryza sativa, Arabidopsis thaliana and **Green** color show Areca catechu. It was illustrated through the tb tool using the **MCScanX** tool [19].

3.7. Protein-Protein Interaction among AcatSOS genes

We have analyzed 10 *AcatSOS* genes through Gene Correlation Network from the Areca palm database. It was illustrated through Cytoscape that demonstrate the interaction among genes. There was strong interaction among *AcatSOS* genes that enhance the stability of genes.

Figure 10. The protein-protein interaction (PPI) network was built through Cytoscape [20] and the data was obtained from the areca palm database [14]. **Bluish color** represents Genes, the **Black line** show interaction among *AcatSOS* genes, **Black** represents *AcatSOS* genes and the **Black** arrow head and **Yellow** tail represent the start and end of gene interaction.

3.8. Characterization of cisacting elements in AcatSOS Promoter Sequence

All 10 *AcatSOS* gene sequences were analyzed to determine the biological roles (release of stress hormones, development, and growth) of the *AcatSOS* genes (3000 bp upstream of start codon) were picked using the PlantCARE search resource for the cis-element analysis displays 10 *AcatSOS* proteins' cis-regulatory elements. The *AcatSOS* gene family of the areca plant had nine genes with increased expression that displayed a variety of cis-acting regulatory factors as shown in (Fig. 7). These cis-acting elements were categorized into nine groups: AURE; auxin responsive elements, LIRE; light-responsive elements. AARE; abscisic acid-responsive elements, MejARE; meja responsive elements, DSRE; defense and stress response, GIRE; giberline responsive elements, LTRE; low-temperature elements, SARE; salicylic acid responsive elements and MYB binding elements. Promoter region data is available in (Supplementary File S5).

Figure 11. The cis-acting components are found in the promoter regions of 10 *AcatSOS* genes, which are located 3000 bp upstream of the start codon. Each element have been represented with different colors [21]. Most of the genes responsible for promoter regions are light-responsive elements. It was designed through TB Tool.

Figure 12. Pichart of promoter regions in *AcatSOS* genes.

A) Growth regulators include Low-temperature RE, MYB for drought, Stress RE, and MYB binding RE

B) Hormone-related elements such as Gibrelline RE, Auxin RE, Salic acid RE and MeJA RE

C) Zein metabolism, Meristem Expression, and Circadian control

D) LIRE, MYB in RE, DNA in RE, and part of RE. it was designed through EXCEL.

3.9. KEGG Enrichment and Go Enrichment Analysis of AcatSOS genes

KEGG (Kyoto Encyclopedia of Genes and Genomes) is used to identify the enriched pathways in a given set of genes or proteins. From result, it is proved that *AcatSOS* are mostly involved in photosynthesis proteins. X-axis represents Gene ratio and Y-axis represents count number of *AcatSOS* genes.

Figure 13. Graphical representation of KEGG enrichment of *AcatSOS* genes in which the highest count is 1 and responsible for Photosynthesis protein with > 0.0125 gene ratio [16].

All 10 *AcatSOS* genes' functions were determined by using of Areca database. X-axis shows the gene ratio while Y-axis represents the count number of genes. From the result it is concluded that these genes are mostly involved in cell development and response to heat(most significant).

Figure 14. Graphical representation of GO enrichment of *AcatSOS* genes in which the highest count is 3 that is responsible for protein transmembrane transport with 0.2 gene ratio [16]. **Red** shows high significance and **Yellow** shows low significance.

3.10. Expression pattern of AcatSOS

From the Areca palm database we extracted all the eFP-Gene data of *AcatSOS* genes. All the *AcatSOS* genes were compiled and constructed the circular heat map through the TB tool. Among all, 3 genes are highly expressed; *AcatSOS*-4, *AcatSOS*-6, and *AcatSOS*-8 are involved in leaf, endocarp, and epicarp development.

Table 2. Gene expression of *AcatSOS* genes in leaf, root, endocarp, stem, epicarp, and mesocarp[16]. Highly expressed genes are represented in **Brown** Highlighted form.

RenameID	Mesocarp	Endocarp	Fruit	Leaf	Root	Epicarp	Stem
AcatSOS-1	NaN	0	NaN	0	NaN	0	NaN
AcatSOS-2	NaN	1.276024333	NaN	0	NaN	1.982273667	NaN
AcatSOS-3	NaN	9.991280667	NaN	5.351949	NaN	104.1729407	NaN
AcatSOS-4	NaN	16.040033	NaN	3.893313	NaN	6.999934	NaN
AcatSOS-5	NaN	0	NaN	0	NaN		NaN
AcatSOS-6	NaN	48.922114	NaN	36.798729	NaN	44.344672	NaN
AcatSOS-7	NaN	0.981822333	NaN	0.033856	NaN	1.460505667	NaN
AcatSOS-8	NaN	27.095471	NaN	14.250592	NaN	28.44155833	NaN
AcatSOS-9	NaN	15.30590267	NaN	16.263288	NaN	27.19187367	NaN
AcatSOS-10	NaN	4.339673	NaN	2.731218333	NaN	2.176102333	NaN

We have revealed the expression pattern in the leaf, endocarp, fruit, stem, and epicarp. From the circular heat map, we concluded that most of the *AcatSOS* genes are involved in leaf, endocarp, and epicarp.

Figure 15. The circular heat map of *AcatSOS* expression of genes in several Areca palm tissues. In this circular heat map, 10 columns indicate the genes, and 7 rows that reflect the responses of various tissues: Root, Stem, Leaf, Fruit, Endocarp, Mesocarp, and Epicarp. From the heat map it is concluded that three genes; *AcatSOS*-4, *AcatSOS*-6, and *AcatSOS*-8 are involved in the growth and development of leaf, endocarp and epicarp.

Figure 16. Homology 3D structure of highly expressed genes. They are mostly similar in shape and help to maintain salt concentration.

Figure 17. Expression of highly expressed genes in Leaf, Endocarp and Epicarp. **(A)***AcatSOS*-4 gene expression, **(B)***AcatSOS*-6 expression and **(C)***AcatSOS*-8 expression

4. Discussion

One of the most commercially significant tropical palm plants is the areca palm (Areca catechu L.). It is a Southeast Asian native known as the areca palm or betel nut palm. The areca catechu nut is a common component in the development of traditional medicines as well as a stimulant used recreationally in various cultures. The nut contains several bioactive substances, such as alkaloids, tannins, and flavonoids, which have been associated with several medicinal and health benefits. [22]. The genetic diversity of 19 Areca catechu accessions from throughout the world was analyzed using Areca palm genome sequencing. The study found that various genomic areas were linked to variance in nut and leaf attributes [23]. The current study identified 10 *AcatSOS*genes, numbered *AcatSOS*-1 through *AcatSOS*-10, which are widely scattered over 5 chromosomes in the areca palm genome (Fig. 4). The identification of *AcatSOS* genes, along with the physicochemical properties of each gene's CDS sequence, protein length, and molecular weight, start and end points on each chromosome, PI, GRAVY, and predicted sub-cellular localization of each *AcatSOS* family member protein, have all been significant findings of the current study. In a phylogenetic study using sequences from rice, Arabidopsis, and areca catechu, we matched the 10 *AcatSOS* gene sequences with the 27 *SOS* protein sequences. *AcatSOS* might be divided into 5 subfamilies, according to the previous studies; 1, 2, 3, 4, and 5 (Fig. 2). All genes were assigned to the appropriate *SOS* family according to the conserved motif and gene structure.Exon-intron structure study showed that the gene patterns of several subgroups were very distinct from one another. However, the structure was largely similar (Fig. 3). Mostly *AcatSOS* have between 0 and 2 introns, which is a comparatively low amount, and most of its genes had just one or no introns. The low number has been linked to stress response in earlier research [24]. The results suggest that the function of *AcatSOS* varies significantly across various subgroups. All *AcatSOS* proteins between the Groups had about the same amount, type, and order of motifs in the motif research. All motif sequences had similarities, according to motif-I. In addition, nearly every grouping had its own distinctive and essentially conservative themes. Motif 3 is a highly conserved [25]. Gene duplication, which often includes tandem and segmental duplication, is a frequent occurrence in the evolution of angiosperms. This can increase the result of the gene family [26]. Tandem and segmental duplication events are also present in *AcatSOS* at the same time. In the current study there are 1 pairs of segmental duplication events and 1 pair of segmental duplication were found in the current study. As we could observe, the *AcatSOS* promoters had a large number of cis-acting elements, such as gibberellin, auxin, ABA, jasmonic acid and salicylic-acid-responsive elements. These genes can be utilized to investigate the mechanisms that control particular plant features. From gene expression, it was concluded that most of the genes are involved in leaf, epicarp, and mesocarp development such as *AcatSOS*-4, *AcatSOS*-6, and *AcatSOS*-8.

Conclusion

We conclude that the SOS1 gene is expressed in a variety of tissues and at different developmental stages. The expression of SOS1 was also induced by salt stress. These findings suggest that SOS1 plays an important role in the response of Areca palm to salt stress. The study provides a foundation for future research on the function of *AcatSOS* genes in Areca palm. By understanding the function of these genes, we can develop new strategies to improve the salt tolerance of this important crop. This study is the first to describe the *SOS* gene family in tropical fruit. It provides a foundation for further research into the functional traits of the areca palm *SOS* gene family, including the role of this gene family in stress tolerance, development, and growth. It will also aid future developments of Areca palm variants.

Declarations

Ethics approval and consent to participate

My work is based on computational method and there is no experimental work involved in this analysis.

Consent for publication

I give permission for the results of a research study to be published.

Authors contributions

Osama Alam wrote the initial draft of the manuscript, prepared the results and interpret it.

Sobia Asmat made revisions to the manuscript, including adjusting the citations and providing feedback on the overall content.

All authors read and approved the final manuscript.

Data Availability Statement:

The data can be freely used, shared, and adapted, as long as the original authors are credited. The dataset includes information on the identification and phylogenetic analysis of the *AcatSOS* gene family in Areca catechu, as well as the characterization of their physiochemical properties, localization signals, promoter sequences, and expression patterns. The dataset also encompasses the identification of segmental and tandem duplications in the Areca palm genome, as well as the involvement of *AcatSOS* genes in the regulation of leaf, endocarp, and epicarp development.

Declaration of interests

I declare that I have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

I believe that my Declaration of Interest statement accurately reflects my interests and relationships, and I have no conflicts of interest that could have influenced the results of my research.

I have no conflicts of interest to disclose.

Conflict of interest

The authors declare that they have no conflict of interest.

Funding Source

There is no funding source involved in this research.

References

1. Kaur, R., K. Singh, and J. Singh, *A root-specific wall-associated kinase gene, HvWAK1, regulates root growth and is highly divergent in barley and other cereals.* Functional & integrative genomics, 2013. **13**: p. 167-177.
2. Flowers, T.J. and T.D. Colmer, *Salinity tolerance in halophytes.* New phytologist, 2008: p. 945-963.
3. Zhu, J.-K., *Salt and drought stress signal transduction in plants.* Annual review of plant biology, 2002. **53**(1): p. 247-273.
4. Zhu, J.-K., *Plant salt tolerance.* Trends in plant science, 2001. **6**(2): p. 66-71.
5. Zhu, J.-K., *Genetic analysis of plant salt tolerance using Arabidopsis.* Plant physiology, 2000. **124**(3): p. 941-948.
6. Martínez-Atienza, J., et al., *Conservation of the salt overly sensitive pathway in rice.* Plant physiology, 2007. **143**(2): p. 1001-1012.
7. Olias, R., et al., *The plasma membrane Na+/H+ antiporter SOS1 is essential for salt tolerance in tomato and affects the partitioning of Na+ between plant organs.* Plant, cell & environment, 2009. **32**(7): p. 904-916.

8. Ji, H., et al., *The salt overly sensitive (SOS) pathway: established and emerging roles.* Molecular plant, 2013. **6**(2): p. 275-286.
9. Shi, H., et al., *The Arabidopsis thaliana salt tolerance gene SOS1 encodes a putative Na+/H+ antiporter.* Proceedings of the national academy of sciences, 2000. **97**(12): p. 6896-6901.
10. Rus, A., et al., *AtHKT1 is a salt tolerance determinant that controls Na+ entry into plant roots.* Proceedings of the national academy of sciences, 2001. **98**(24): p. 14150-14155.
11. Shi, H., et al., *The Arabidopsis SOS5 locus encodes a putative cell surface adhesion protein and is required for normal cell expansion.* The Plant Cell, 2003. **15**(1): p. 19-32.
12. Henderson, A., *Palms of southern Asia*. Vol. 50. 2009: Princeton University Press.
13. Cai, X., et al., *Genome-wide identification of GRF gene family and their contribution to abiotic stress response in pitaya (Hylocereus polyrhizus).* International Journal of Biological Macromolecules, 2022. **223**: p. 618-635.
14. Yang, Y.; Huang, L.; Xu, C.; Qi, L.; Wu, Z.; Li, J.; Chen, H.; Wu, Y.; Fu, T.; Zhu, H.; et al. Chromosome-scale genome assembly of areca palm (Areca catechu). Mol. Ecol. Resour. 2021, 21, 2504–2519
15. Goodstein, D.M., et al., *Phytozome: a comparative platform for green plant genomics.* Nucleic acids research, 2012. **40**(D1): p. D1178-D1186.
16. Chen, C., et al., *TBtools: an integrative toolkit developed for interactive analyses of big biological data.* Molecular plant, 2020. **13**(8): p. 1194-1202.
17. Tamura, K., G. Stecher, and S. Kumar, *MEGA11: molecular evolutionary genetics analysis version 11.* Molecular biology and evolution, 2021. **38**(7): p. 3022-3027.
18. Letunic, I. and P. Bork, *Interactive Tree Of Life (iTOL) v4: recent updates and new developments.* Nucleic acids research, 2019. **47**(W1): p. W256-W259.
19. Hirsch, S. and G.E. Oldroyd, *GRAS-domain transcription factors that regulate plant development.* Plant signaling & behavior, 2009. **4**(8): p. 698-700.
20. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome research, 2003. **13**(11): p. 2498-2504.
21. Lescot, M., et al., *PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences.* Nucleic acids research, 2002. **30**(1): p. 325-327.
22. Ding, H., et al., *Genome-Wide Association Analysis of Fruit Shape-Related Traits in Areca catechu.* International Journal of Molecular Sciences, 2023. **24**(5): p. 4686.
23. Chia-Yi Ko, H.-C.S., Wen-Tzu Liu, Hui-Ling Huang, Fang-l Hsieh, and Yen-Tzu Wu, *Genetic diversity and population structure of Areca catechu L. revealed by restriction site-associated DNA sequencing.* BMC Genomics, 2017. **18**: p. 668.
24. Sidhu, N.S., et al., *Genome-wide identification and analysis of GRAS transcription factors in the bottle gourd genome.* Scientific reports, 2020. **10**(1): p. 1-9.
25. Zhang, H., et al., *Genome-wide identification, characterization, interaction network and expression profile of GRAS gene family in sweet orange (Citrus sinensis).* Scientific Reports, 2019. **9**(1): p. 2156.
26. Liu, B., et al., *Genome-wide characterization and expression analysis of GRAS gene family in pepper (Capsicum annuum L.).* PeerJ, 2018. **6**: p. e4796.

Figures

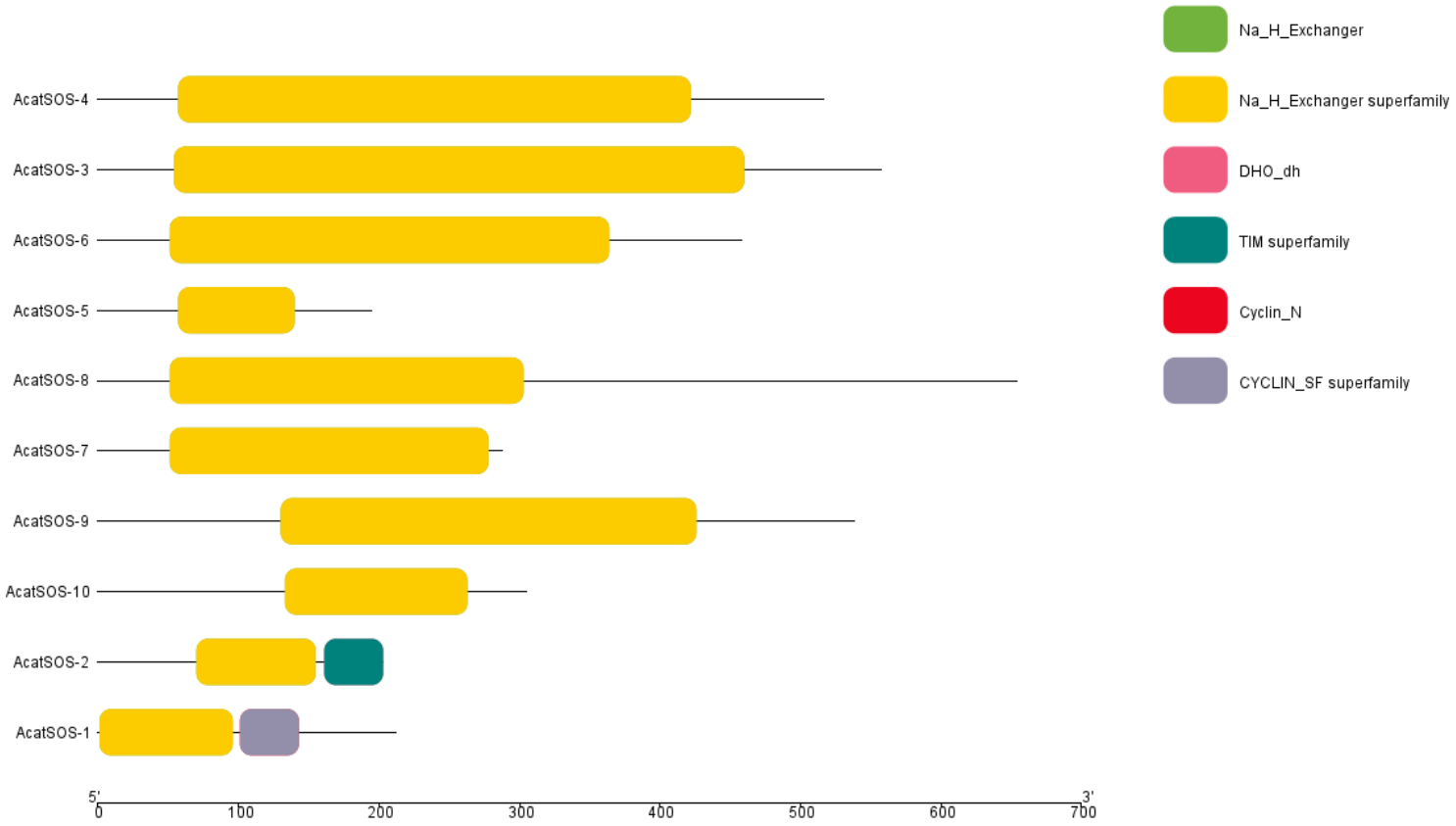

Figure 1

Domains of *AcatSOS* proteins ranged from *(AcatSOS-1 to AcatSOS-10).* **Green** color represents Na-H Exchanger, **Yellow** shows the Na-H Exchanger superfamily and the **Blackline** shows protein sequences that are not conserved. It was designed inTB tool [16].

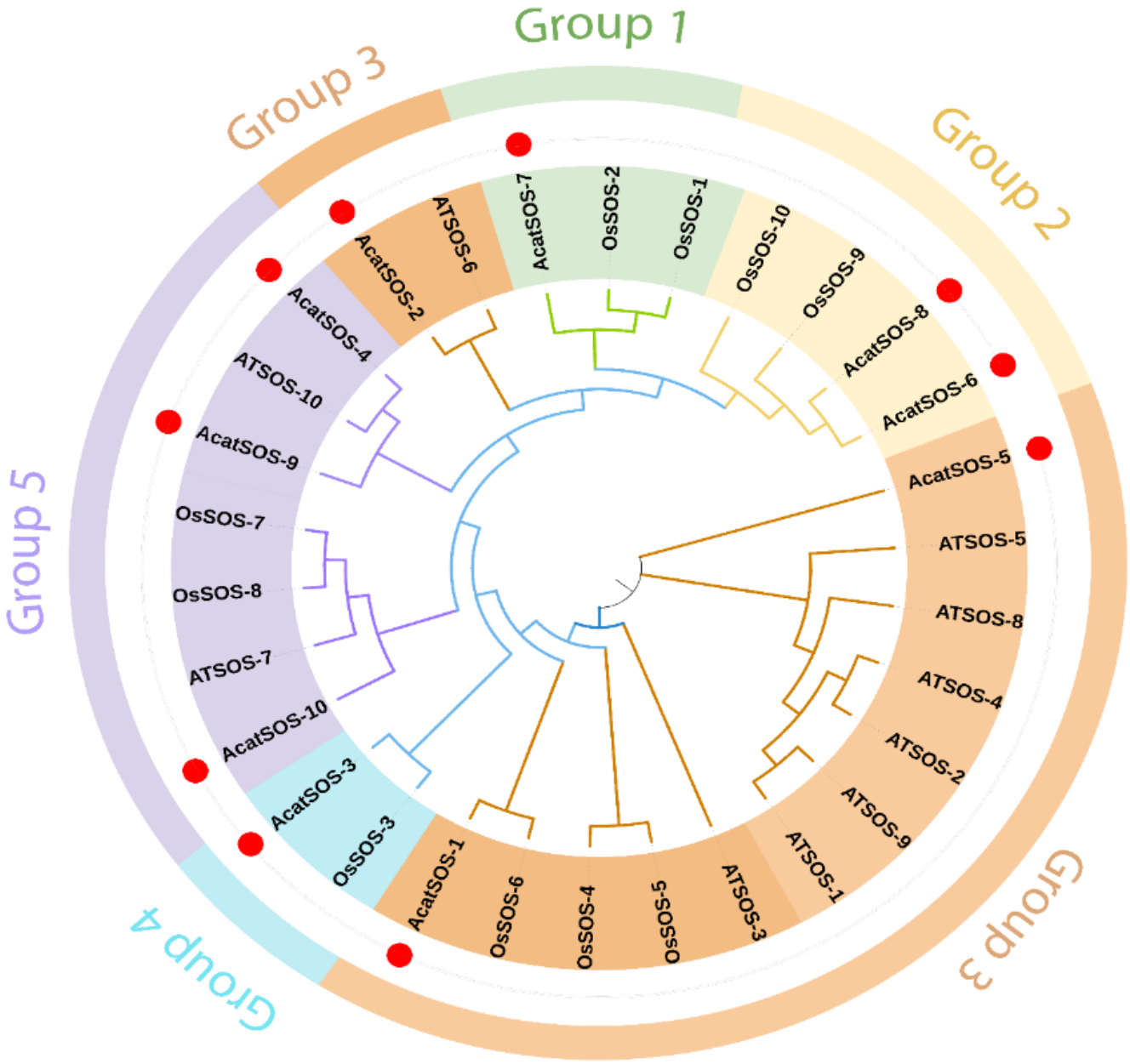

Figure 2

Comparative phylogenetic analysis of *AcatSOS* genes along with other three species, including rice, areca catechu, and Arabidopsis. It is classified based on the conserved motif [17]. Each group is differentiated through various color: 1; Green, 2; Yellow, 3; Orange, 4; Blue, and 5; Purple. It was illustrated through ITool software and the boundaries were designed through Adobe Illustrator.

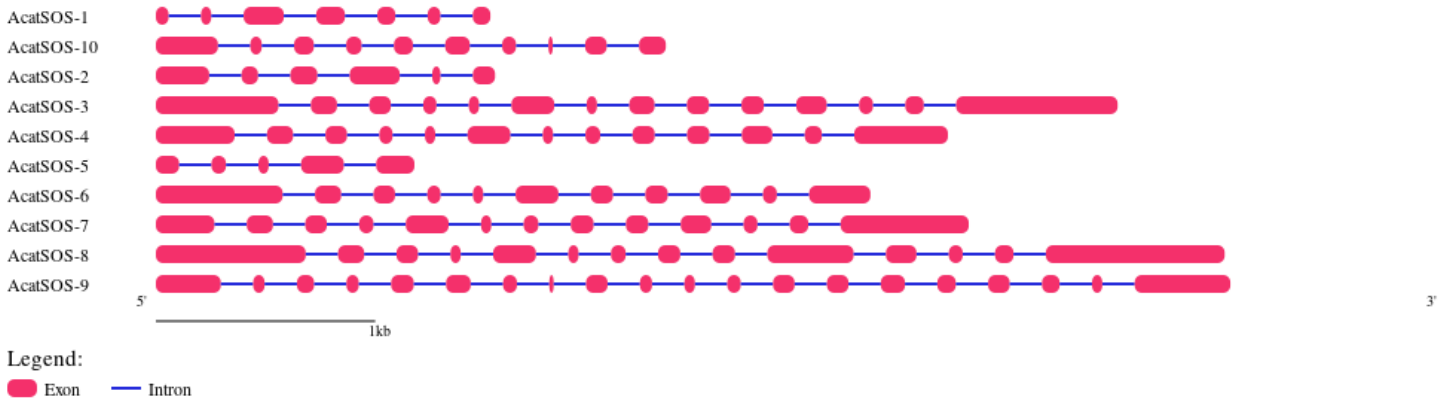

Figure 3

Gene structure of *AcatSOS*-1 to *AcatSOS*-10. **The red** color shows exon structure. While, **Blue** color line shows Intron in the gene structure. It was designed by taking CDS and Genomic data of *AcatSOS genes* from Areca database and run in Gene Structure Display server https://mhttp://gsds.gao-lab.org/ (accessed on 25 June, 2023).

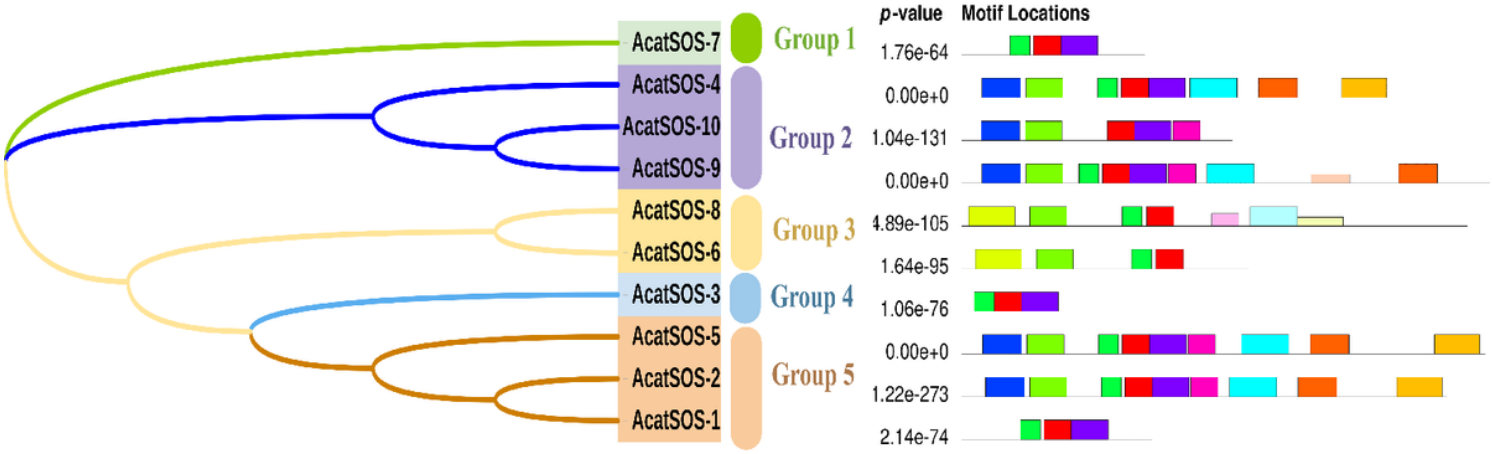

Figure 4

Phylogenetic and Motif analysis of *AcatSOS* genes [17]. Five subfamilies according to previous literature. Group 1 and Group 4 contain 1 gene, Group 2 and Group 5 contain 3 genes, and Group 3 contain 2 genes.

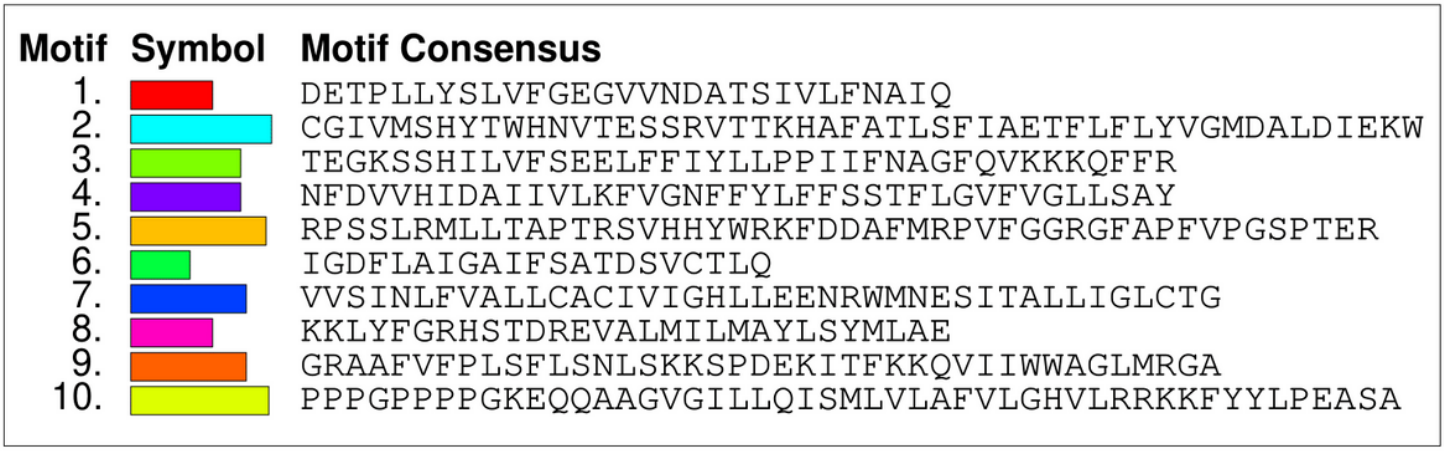

Figure 5

Show the motif color sequences ranging from 1 to 10 https://meme-suite.org/meme/tools/meme(accessed on 25 June, 2023). Each color has own specific sequence that is conserved among the motifs. It was obtained from the meme suit website. Motif logo data is available in (Supplementary File S3).

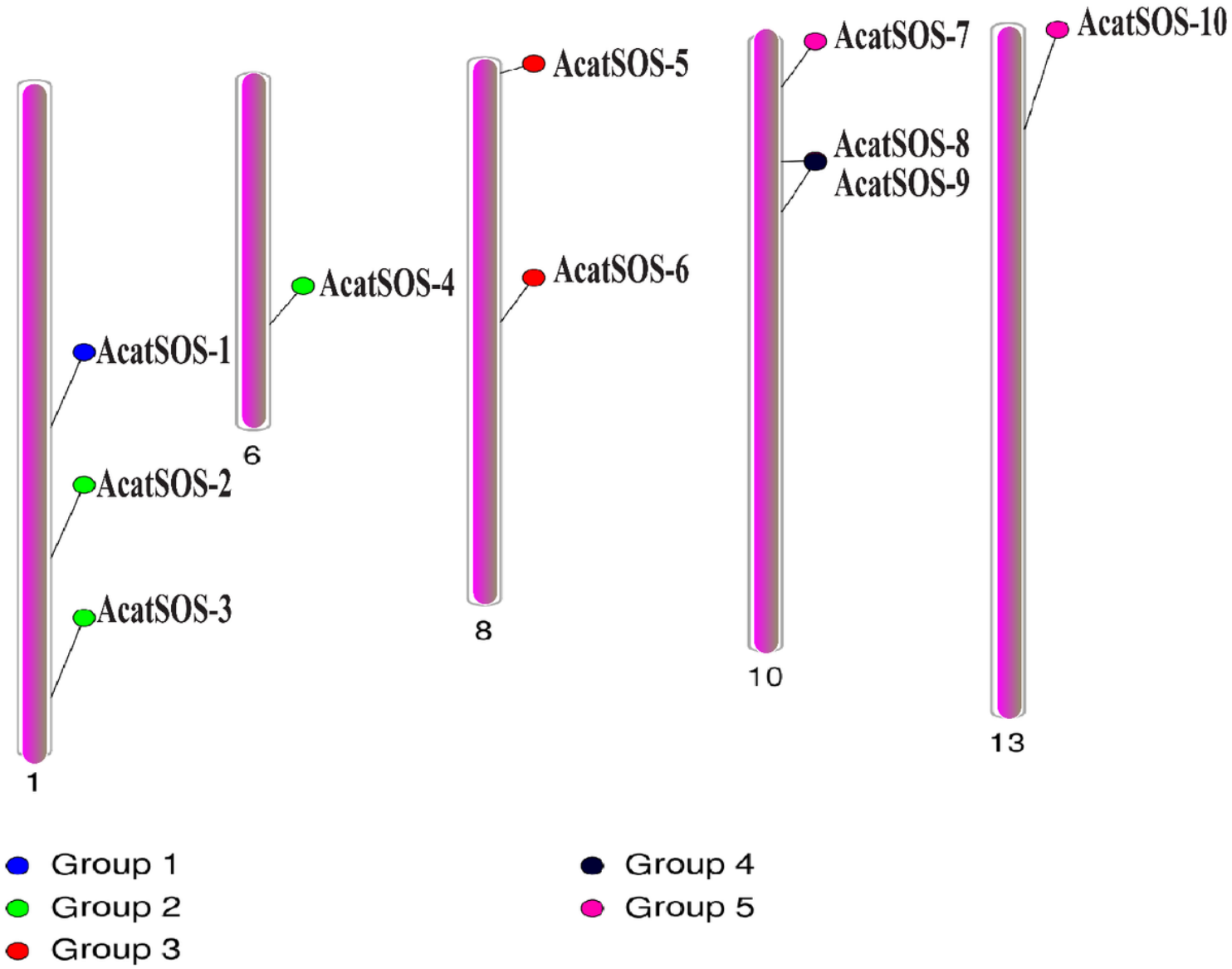

Figure 6

Chromosomal distribution of *SOS* genes on 5 Areca palm chromosomes. It is visualized in http://visualization.ritchielab.org/phenograms/plot (accessed on 25 June, 2023). **White** color shapes represent Chromosome and the names of the genes are in **Black**. The distribution of the *AcatSOS* genes was variable. Chr 1 and 10 contained the majority of the 3 *AcatSOS* genes. There is 1 *AcatSOS* gene in Chr 13 and 6. Chr 8 contains only two genes.

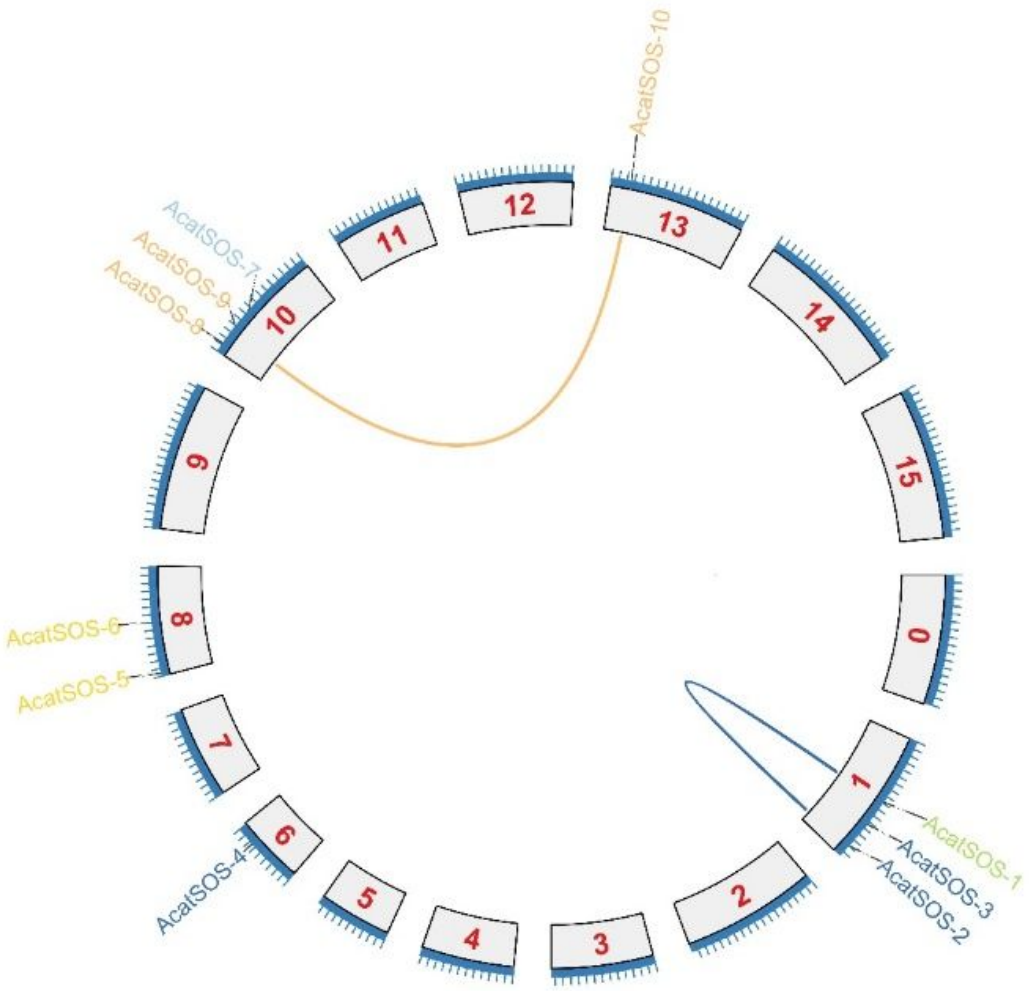

Figure 7

Graph showing the circus analysis of *AcatSOS* proteins. Events of segmental and tandem duplication are shown by colored lines. There is 1 pair of tendon duplications: *AcatSOS*-1/*AcatSOS*-2 and 1 pair of segmental duplication: *AcatSOS-9/ AcatSOS-10* [19].

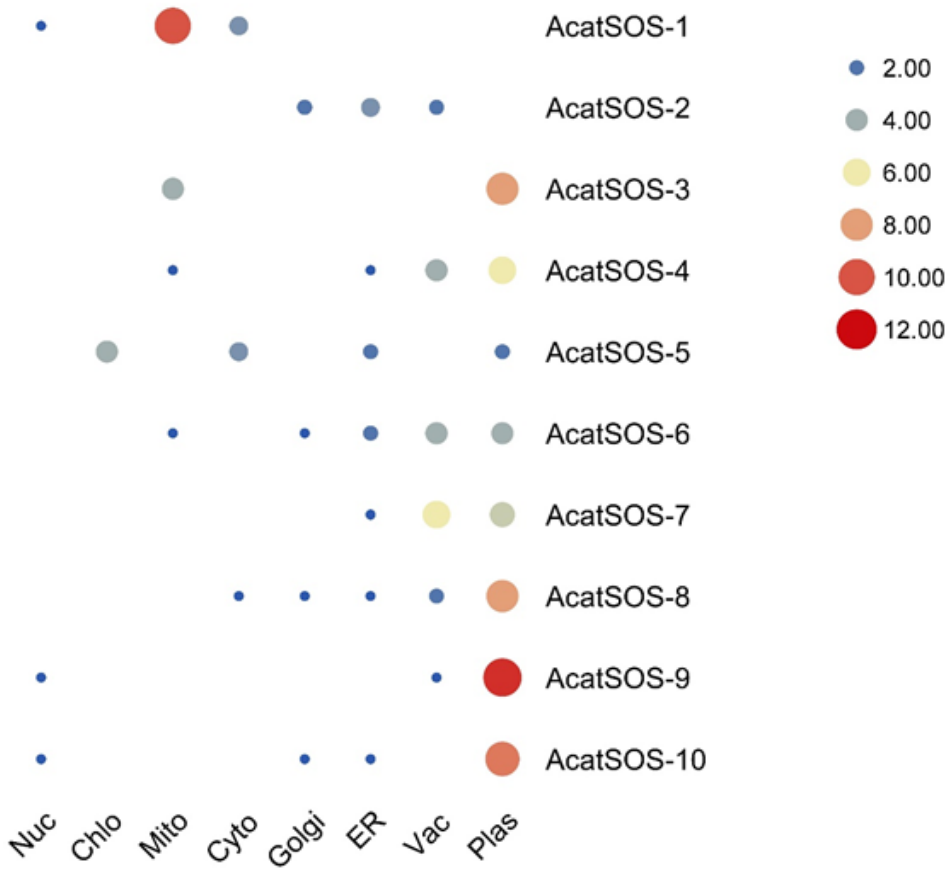

Figure 8

AcatSOS heat map in quantitative form. It was illustrated in TB tool[16]. The color bar with **Red** signifies a high concentration of proteins, **Blue** represents a small value and **Yellow** shows an intermediate intensity. Most of the genes are localized in Plastid and Vacule**.**The gene localization data is available in (Supplementary File S4).

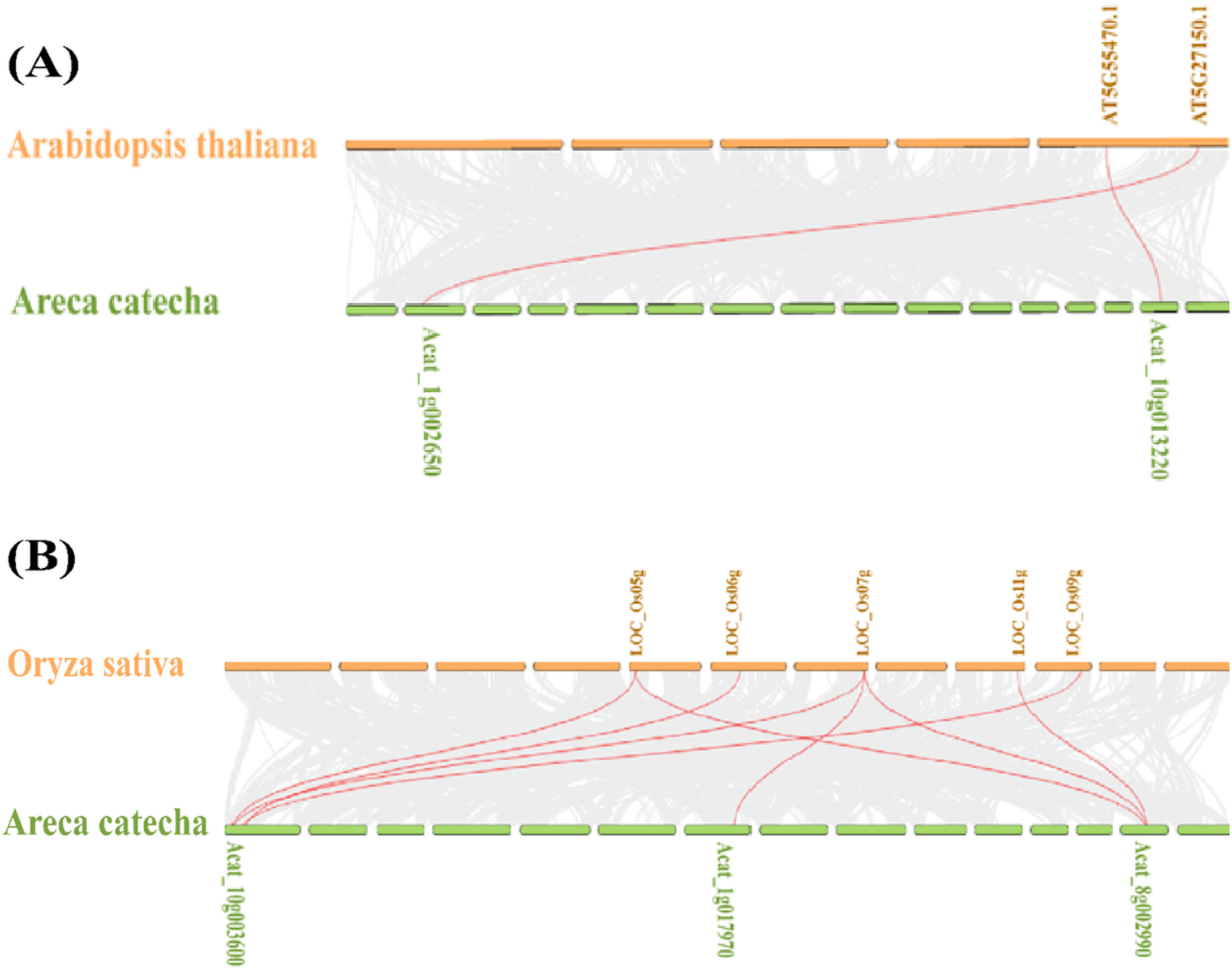

Figure 9

Synteny analysis of **(A)***Arabidopsis thaliana* and *Areca catechu* **(B)***Oryza sativa* and *Areca catechu,* **Orange** color Oryza sativa, Arabidopsis thaliana and **Green** color show Areca catechu. It was illustrated through the tb tool using the **MCScanX** tool [19].

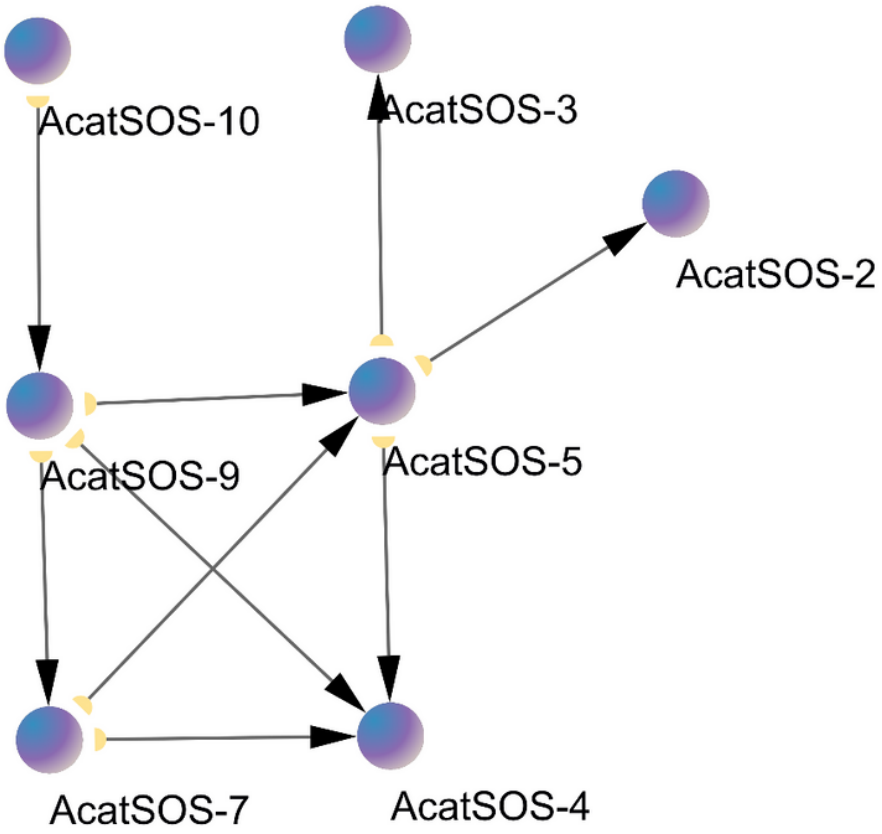

Figure 10

The protein-protein interaction (PPI) network was built through Cytoscape [20] and the data was obtained from the areca palm database [14]. **Bluish color** represents Genes, the **Black line** show interaction among *AcatSOS* genes, **Black** represents *AcatSOS* genes and the **Black** arrow head and **Yellow** tail represent the start and end of gene interaction.

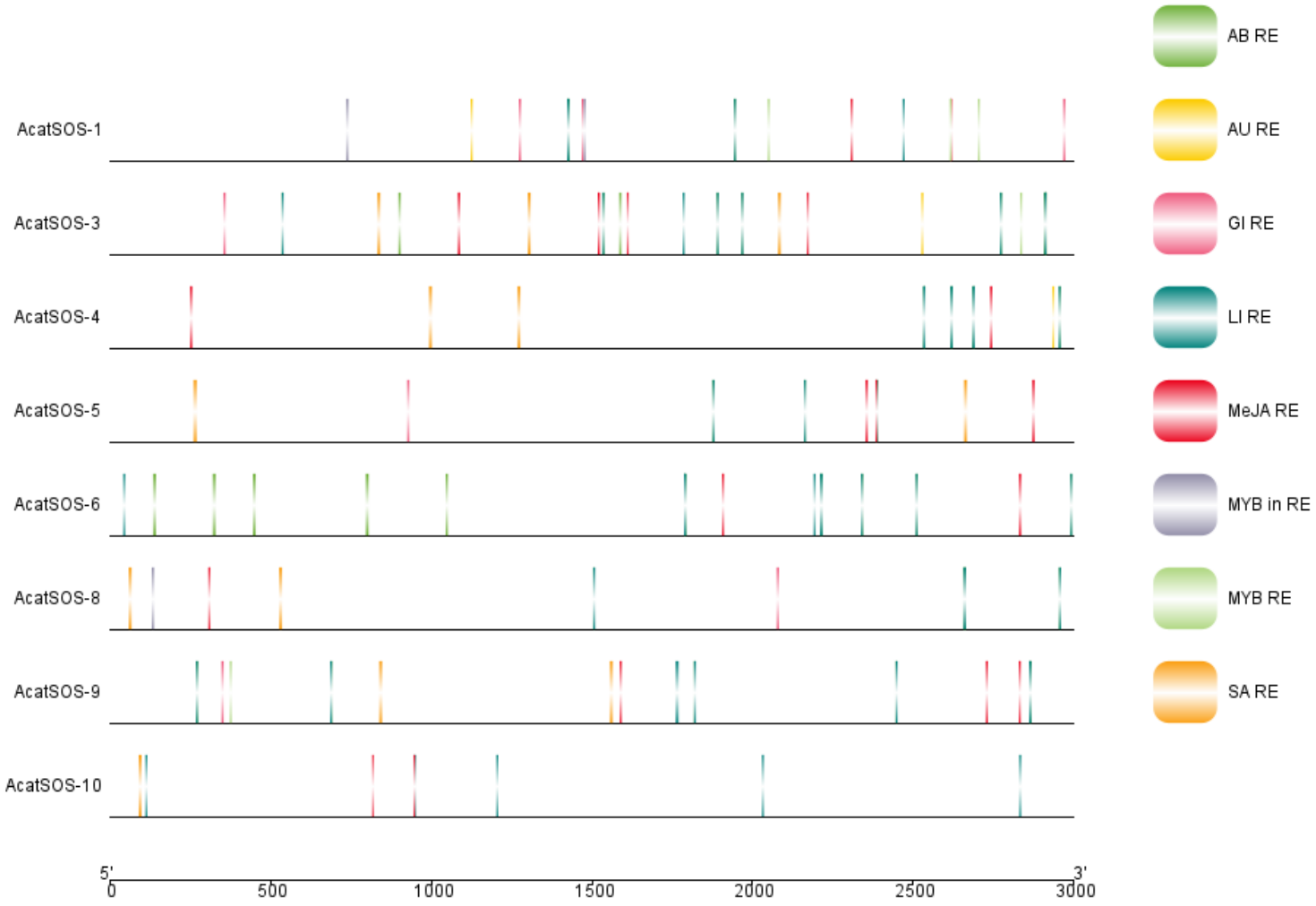

Figure 11

The cis-acting components are found in the promoter regions of 10 *AcatSOS* genes, which are located 3000 bp upstream of the start codon. Each element have been represented with different colors [21]. Most of the genes responsible for promoter regions are light-responsive elements. It was designed through TB Tool.

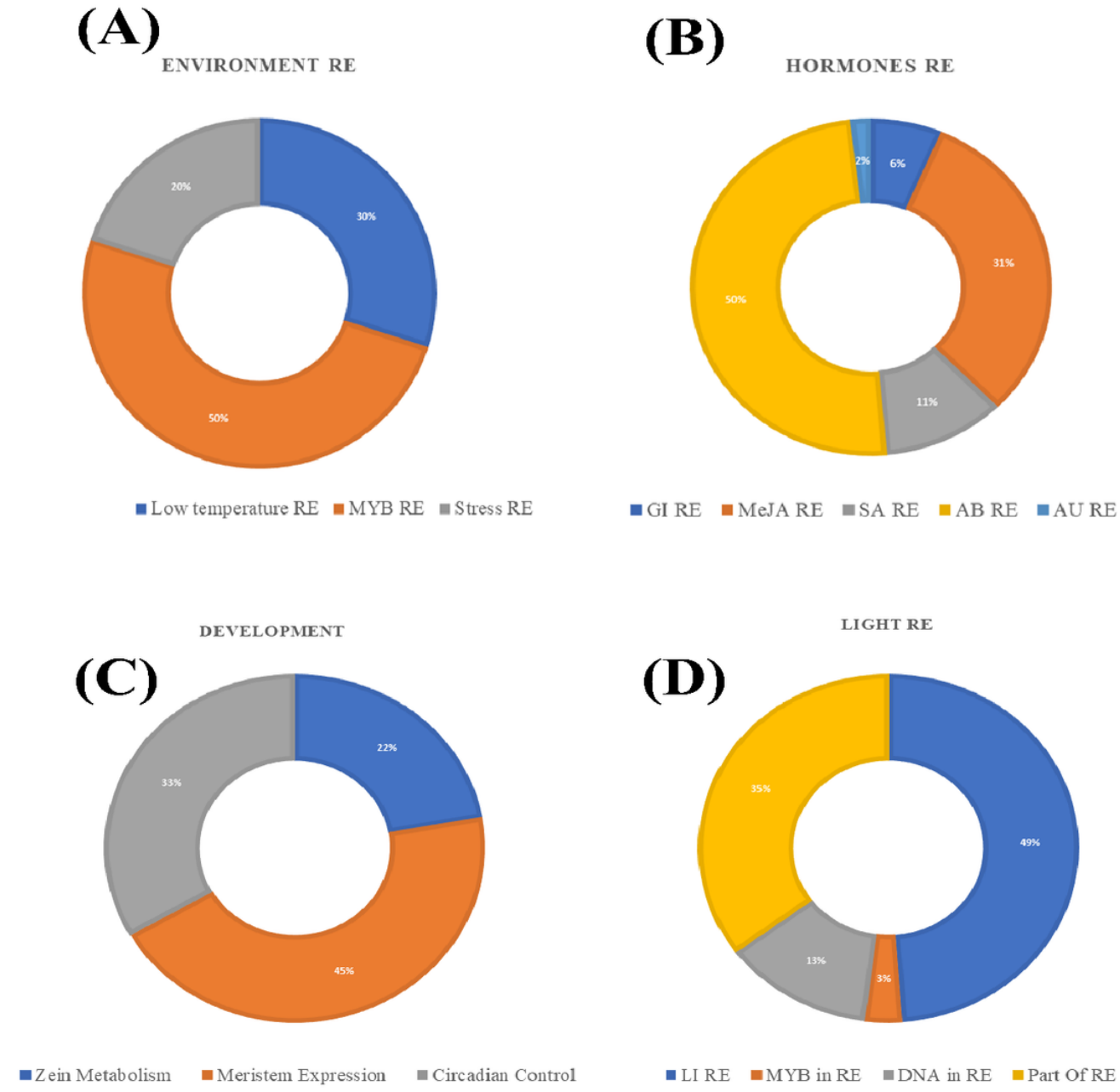

Figure 12

Pichart of promoter regions in *AcatSOS* genes.

(A) Growth regulators include Low-temperature RE, MYB for drought, Stress RE, and MYB binding RE

(B) Hormone-related elements such as Gibrelline RE, Auxin RE, Salic acid RE and MeJA RE

(C) Zein metabolism, Meristem Expression, and Circadian control

(D) LIRE, MYB in RE, DNA in RE, and part of RE. it was designed through EXCEL.

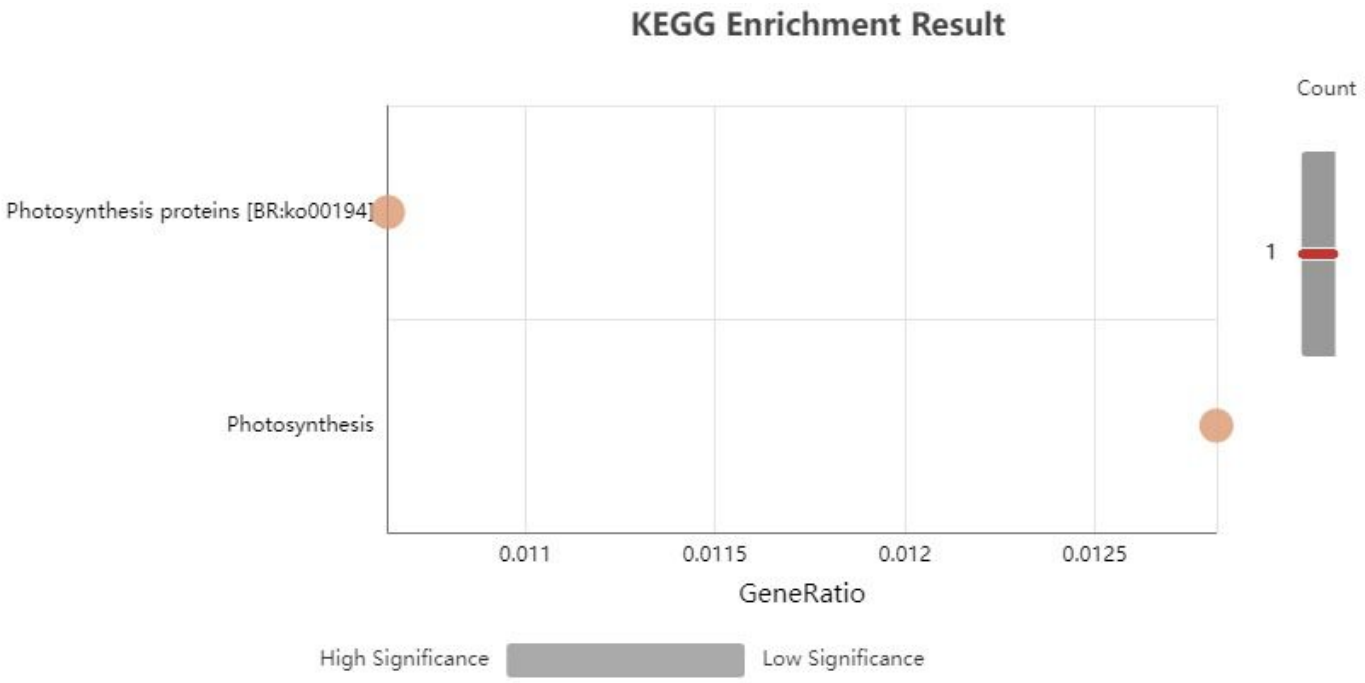

Figure 13

Graphical representation of KEGG enrichment of *AcatSOS* genes in which the highest count is 1 and responsible for Photosynthesis protein with >0.0125 gene ratio [16].

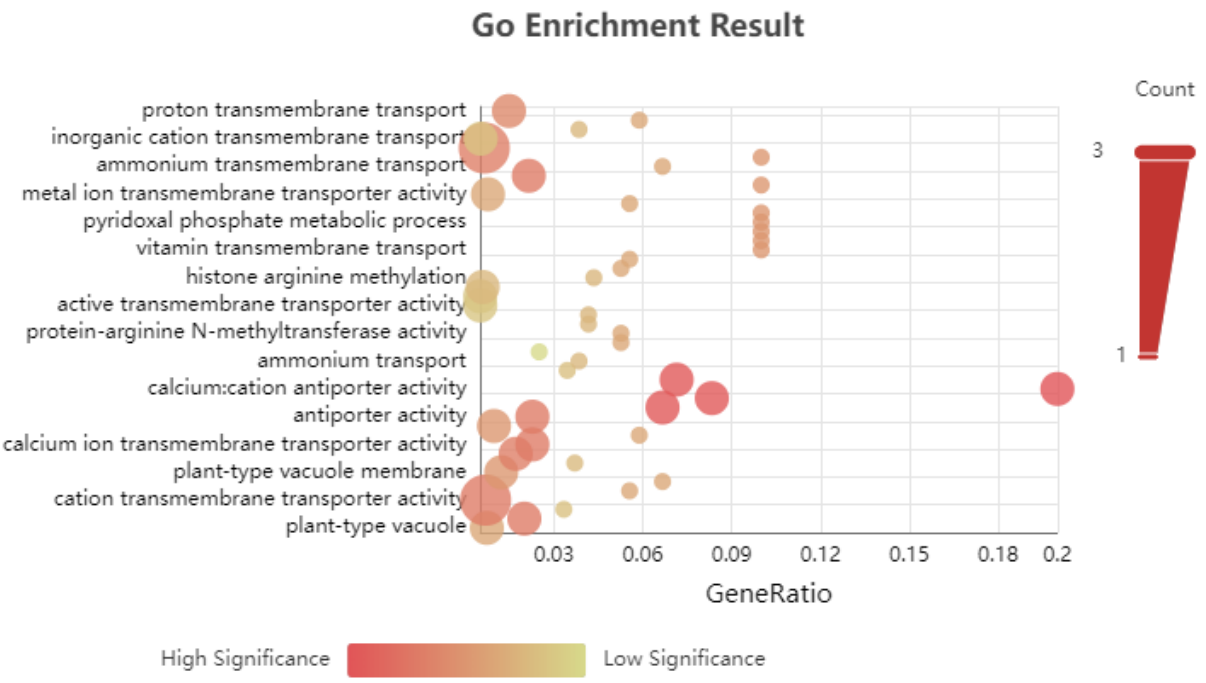

Figure 14

Graphical representation of GO enrichment of *AcatSOS* genes in which the highest count is 3 that is responsible for protein transmembrane transport with 0.2 gene ratio [16]. **Red**shows high significance and **Yellow** shows low significance.

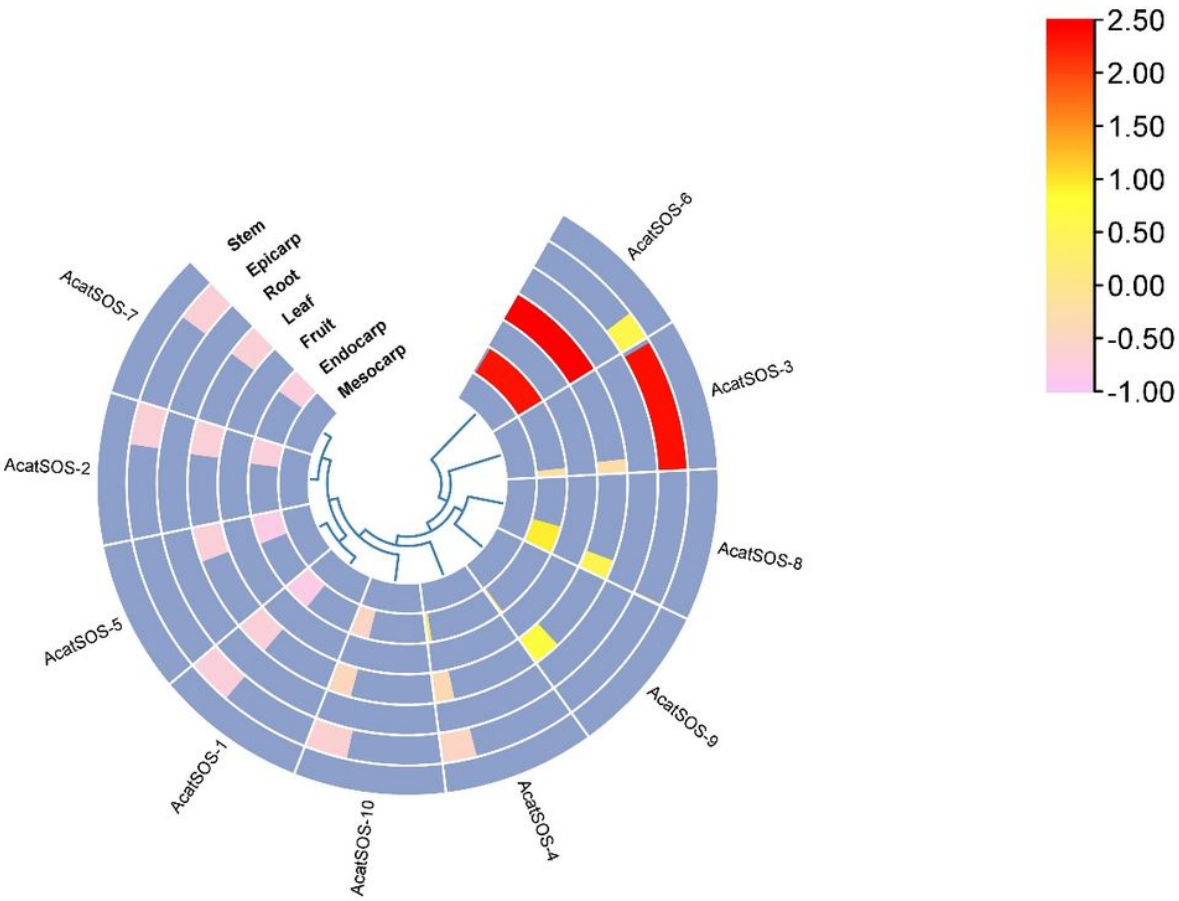

Figure 15

The circular heat map of *AcatSOS* expression of genes in several Areca palm tissues. In this circular heat map, 10 columns indicate the genes, and 7 rows that reflect the responses of various tissues: Root, Stem, Leaf, Fruit, Endocarp, Mesocarp, and Epicarp. From the heat map it is concluded that three genes; *AcatSOS*-4, *AcatSOS*-6, and *AcatSOS*-8 are involved in the growth and development of leaf, endocarp and epicarp.

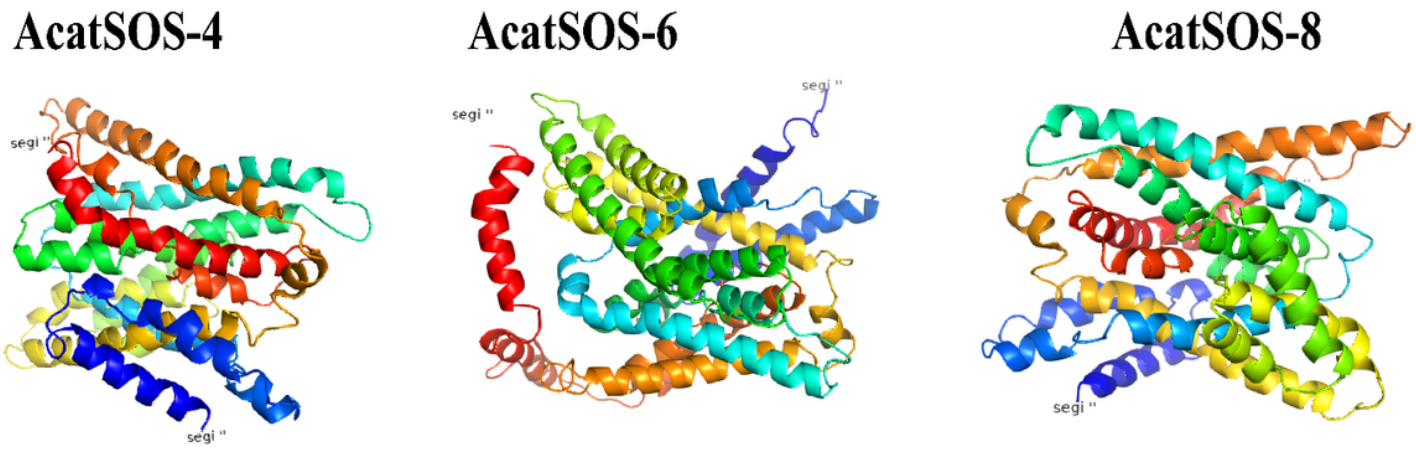

Figure 16

Homology 3D structure of highly expressed genes. They are mostly similar in shape and help to maintain salt concentration.

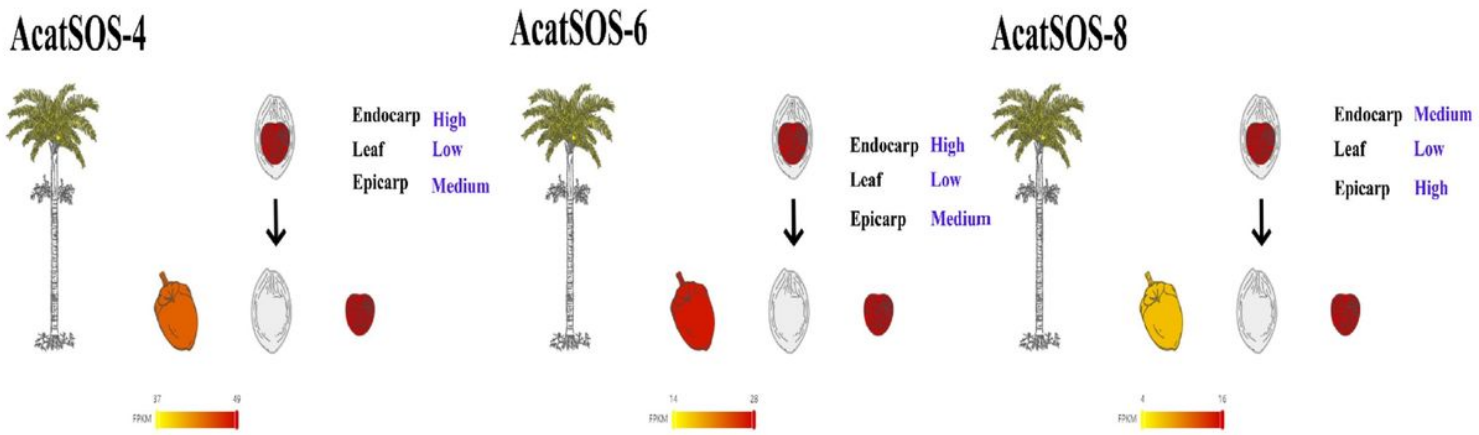

Figure 17

Expression of highly expressed genes in Leaf, Endocarp and Epicarp.**(A)** *AcatSOS*-4 gene expression, **(B)** *AcatSOS*-6 expression and **(C)** *AcatSOS*-8 expression

Supplementary Files

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

- SupplementaryFileS1.xlsx
- SupplementaryFileS1A.pdf
- SupplementaryFileS2.xlsx
- SupplementaryFileS3.png
- SupplementaryFileS4.xlsx
- plantCAREoutputPlantCARE8687.tab