Galactinol synthases (GolSs) are the key enzymes that participate in raffinose family oligosaccharides (RFO) biosynthesis, which perform a big role in modulating plant growth and response to biotic or abiotic stresses. To date, no systematic study of this gene family has been conducted in cassava (*Manihot esculenta* Crantz). Here, eight *MeGolS* genes are isolated from the cassava genome. Based on phylogenetic background, the *MeGolS*s are clustered into four groups. Through predicting the cis-elements in their promoters, it was discovered that all *MeGolS* members act as hormone-, stress-, and tissue-specific related elements to different degrees. *MeGolS* genes exhibit incongruous expression patterns in various tissues, indicating that different *MeGolS* proteins might have diverse functions. *MeGolS1* and *MeGolS3–6* are highly expressed in leaves and midveins. *MeGolS3–6* are highly expressed in fibrous roots. Quantitative real-time Polymerase Chain Reaction (qRT-PCR) analysis indicates that several *MeGolS*s, including *MeGolS1*, 2, 5, 6, and 7, are induced by abiotic stresses. microRNA prediction analysis indicates that several abiotic stress-related miRNAs target the *MeGolS* genes, such as mes-miR156, 159, and 169, which also respond to abiotic stresses. The current study is the first systematic research of GolS genes in cassava, and the results of this study provide a basis for further exploration the functional mechanism of GolS genes in cassava.

**Keywords:** galactinol synthase; gene expression pattern; abiotic stress; *Manihot esculenta*; cassava

1. Introduction

Galactinol synthases (GolS, EC 2.4.1.123) are the enzymes that catalyze the reaction of myo-inositol and UDP-galactose to generate galactinol. Galactinol is one of the organic osmolytes in plants [1], and also is considered to be one of plant defense molecules [2]. It is known as the substrate to provide galactosyl in raffinose family oligosaccharides (RFOs) biosynthesis pathway for generation of the raffinose, stachyose, verbascose, and other larger soluble oligosaccharides [3]. Galactinol and RFOs participate in a variety of physiological and developmental functions in plant life [4]. The RFOs
accumulate in seeds to protect the embryo from dehydration damage, and are associated with seed longevity [5,6], and act as galactose stores for rapid germinating requirements [7]. They accumulate in vegetative tissues and act as signaling molecules in answer to a series of biotic [2,8] and abiotic stresses [1,9,10] suffered by plants.

GolS participates in the initial stage of RFO biosynthesis. Since it plays important roles in plants, more and more GolS genes have been investigated from various plants, for instance, Arabidopsis (Arabidopsis thaliana L.) [11], rapeseed (Brassica napus L.) [12], tobacco (Nicotiana tabacum L.) [12], maize (Zea mays L.) [13], Tea plant (Camellia sinensis (L.) O. Ktze) [14], wheat (Triticum aestivum L.) [15], tomato (Solanum lycopersicum Mill.) [16], chickpea (Cicer arietinum L.) [17], and so on. Previous studies indicate that the expression of the GolS genes went hand in hand with abiotic stress. Seven GolS genes have been characterized in Arabidopsis, in which the expressions of AtGolS1 and AtGolS2 responded to drought and salinity while AtGolS3 responded to cold [11]. Some studies have also revealed that AtGolS1 also responded to heat stress [18–20]. Over-expression of AtGolS2 in rice significantly improved the capability of transgenic plants to tolerant to drought, and increased the grain yield under drought conditions [21]. Regarding chickpea, CaGolS1 and CaGolS2 responded to abiotic stresses and CaGolS1 was induced prior to CaGolS2 under both heat and oxidative stressors [17]. Considering wheat, TaGolS1 and TaGolS2 responded to cold stress, but not drought, heat stressors, or abscisic acid (ABA). Over-expression of TaGolS1 or TaGolS2 in rice found that the transgenic rice could improve cold tolerance and accumulate significantly higher standards of galactinol and raffinose [22]. In wheat, the TaGolS3 was upregulated by heavy metal, cold, and salinity stressors. By over-expression of TaGolS3 in transgenic Arabidopsis and transgenic rice, it was revealed that TaGolS3 can improve their zinc stress tolerance by regulating the reactive oxygen species (ROS) production [15]. Found in Ammopiptanthus nanus (M. Pop.) Cheng f., AnGolS1 responded to cold, salinity, and drought pressures [23].

Cassava (Manihot esculenta Crantz), which belongs to Euphorbiaceae, is a starch-rich root crop found in tropical and subtropical areas. It has outstanding soil nutrient- and water-use efficiency and can grow in barren and arid soil where other crops cannot grow and develop normally. It has a strong tolerance to drought and can withstand 4–6 months of continuous drought stress, quickly restoring growth and development when the rainy season arrives [24]. However, it is sensitive to low temperatures and salinity [25], and the storage root deteriorates easily after harvest. The molecular processes of cassava in answering to abiotic stresses are still unclear. Although several papers have reported the functions of GolS genes in other species, their features and functions in cassava are still unclear. Considering the significant roles of galactinol synthase in plant growth and response to environmental stressors, eight MeGolS genes in cassava are amplified according to the latest cassava genome database shared online (http://www.phytozome.net/cassava). The phylogenetic relationship, gene structures, conserved motifs, chromosome localization, and mode of expression of MeGolSs in plant organs or tissue and their response to abiotic stressors are systematically analyzed. The results will lay a basis for further investigation of the regulating functions of MeGolS genes in cassava.

2. Materials and Methods

2.1. Plant Materials and Treatments

One-month-old cassava seedlings (Manihot esculenta Crantz cultivar, SC8) were treated with 200 mM NaCl for salt stress, 20% PEG6000 for drought stress, and 4 °C for cold stress for inducible expression analysis of MeGolS genes under abiotic stresses. The one-month-old cassava seedlings without NaCl, PEG6000, and 4 °C treatment were used as control. Young leaves (the first to third fully expanded leaves from top position) were harvested at 0 h, 3 h, 6 h, 9 h, 12 h, and 24 h after the three abiotic stress treatments, respectively. Three biological replicates were set in this experiment. The harvested leaves were immediately frozen in liquid nitrogen and analyzed by qRT-PCR.
2.2. Identification, Isolation, and Analysis of MeGolS Family Genes

To identify the MeGolS members in cassava, the previously identified Galactinol synthase amino acid sequences from the *Arabidopsis thaliana* genome database were downloaded as well as the BLASTP of the *Manihot esculenta* V6.1 database. Reverse transcription PCR was performed to obtain the full-length cDNAs of GolS from cassava, and the primers utilized are listed in Table S1. The PCR products were confirmed by sequencing and uploaded to GenBank. The predicted amino acids of MeGolSs were analyzed by Jellyfish software (LabVelocity Inc., Los Angeles, CA, USA). Each amino acids sequence of MeGolSs was further verified by the Pfam [26] and SMART [27] online tools. The physicochemical properties of each MeGolS protein was calculated by the online tool ProtParam, including molecular weight (MW), theoretical isoelectric point (pI), and hydrophilic mean (GRAVY) [28]. The possible subcellular location of GolS proteins from cassava and Arabidopsis were predicted using the CELLO2GO tool [29].

2.3. Chromosomal Localization of MeGolS Family Genes

The information of chromosomal localization of the MeGolS family genes from the cassava genomics resource (Bioproject: PRJNA234389) were searched. A chromosomal localization map of these genes was drawn by Mapinspect software (Mike Lischke, Berlin, Germany) based on the relative location of each gene on each chromosome. Then the Figure was enhanced by Photoshop software CS (San Jose, CA, USA).

2.4. Prediction of MeGolS Gene Structures and the Conserved Motifs

The GolS amino acid sequences from Arabidopsis and cassava were aligned by ClustalW to produce the phylogenetic tree based on the NJ-method (neighbor-joining) using molecular evolutionary genetics analysis (MEGA7.0, Institute of Molecular Evolutionary Genetics The Pennsylvania State University, University Park, PA 16802, USA) software [30]. The gene structures of AtGolSs and McGolSs were drawn by the online software GSDS (Gene Structure Display Server, Center for Bioinformatics (CBI), Peking University, Beijing, China) [31]. The conserved motifs in GolSs were analyzed by using the online Motif alignment and search tool-MEME Suite 4.12.0 [32].

2.5. Cis-Element Analysis within the MeGolS Family Gene Promoters

To analyze the cis-elements of MeGolS promoters, the 1500 base pair (bp) sequences upstream of the initiation codons (ATG) of each MeGolS genes were obtained from the cassava genome database. The website of PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was used to predict the cis-elements within promoters of MeGolS genes. The data was processed by Microsoft Office Excel software to obtain table and figure.

2.6. The Expression Profiles of MeGolS Genes in Cassava

To further understand the diverse gene expression profiles of the MeGolS family genes, the authors downloaded the publicly available *M. esculenta* RNA sequencing (RNA-seq) data from the NCBI website and constructed a heatmap using the OmicShare tools (http://www.omicshare.com/tools). Expression data of the 8 MeGolS genes were drawn from the Transcriptome sequencing datasets (Bioproject ID PRJNA324539). The 11 different cassava tissues were the leaf blade, leaf mid-veins, petioles, stems, lateral buds, shoot apical meristems (SAM), storage roots, fibrous roots, root apical meristems (RAM), somatic organized embryogenic structures (OES), and friable embryogenic callus (FEC).

2.7. MeGolS Gene Expression Profiles under Drought, Cold, and Salt Stresses

The relative expression profiles of MeGolSs were obtained by qRT-PCR using the ABI7900HT FastReal-Time PCR System (Applied Biosystems, Foster City, CA, USA). To evaluate the relative
expression level of MeGolS genes, the $2^{-\Delta\Delta Ct}$ method was used. The relative expression levels were equal to the mean values and transform log2 values to draw heatmap by the OmicShare tools (http://www.omicshare.com/tools). The MeGolS gene primers, and internal reference gene primers used for qRT-PCR, are listed in Table S1. Each sample was analyzed with three technical replicates.

2.8. Prediction of microRNAs Targeting to MeGolS Genes

Genome sequences of 8 MeGolS genes were used in an online tool (psRNATarget Server, http://plantgrn.noble.org/psRNATarget/) to search against cassava microRNAs to predict the potential miRNAs targeting MeGolS genes. The interaction network between the MeGolS genes and predicted miRNAs was built by Cytoscape software (http://www.cytoscape.org/).

3. Results

3.1. Identification and Cloning of MeGolS Family Genes in Cassava

There were eight GolS genes in the M. esculenta genome were identified based on the BLASTP program constructed on the known AtGolS amino acid sequences as references. The full-length CDS sequences of MeGolS1–8 were isolated by reverse transcription PCR. Detailed information about the CDS sequences of MeGolS1–8 and the corresponding amino acids were deposited in GenBank, the accession numbers can be found in Table 1. The protein length and molecular weight of these MeGolS proteins ran from 320–338 aa, and 37.14–38.60, respectively. All the MeGolS proteins were acidic since their isoelectric points were smaller than seven. All the MeGolS proteins were hydrophilic proteins since their GRAVY scores were negative. Subcellular localization predication indicated that MeGolS3 was located in two organelles, the cytoplasm and chloroplast. The other seven MeGolSs were located in one organelle; MeGolS1, 2, 4, and 8 were located in the cytoplasm, MeGolS5 and MeGolS6 were located in the plasma membrane, and MeGolS7 was located in ER (Table 1).

3.2. Chromosomal Location and Distribution of MeGolS Family Genes

The physical chromosomal locations of MeGolS genes were searched on phytozome (Manihot esculenta V6.1 database). Eight members of MeGolSs were distributed in seven chromosomes of M. esculenta, respectively (Figure 1). Two genes (MeGolS5 and MeGolS7) were located on chromosome 08. The remaining six genes, MeGolS1, MeGolS2, MeGolS3, MeGolS4, MeGolS6, and MeGolS8, were distributed separately on six different chromosomes, Chr03, Chr15 Chr13, Chr05, Chr09, and Chr01, respectively. The MeGolS1, 2, 3, 6, and 8 had the same orientation, and MeGolS4, 5, and 7 had the opposite orientation.

3.3. Analysis of Phylogenesis, Conserved Motifs, and Gene Structures of GolS Family Genes in Cassava and Arabidopsis

Multiple sequence alignment was used to manage the GolS protein sequences from cassava and Arabidopsis. After that, a neighbor-joining tree was established by MEGA7.0 (Figure 2A). All the GolSs were divided into four (groups I–IV) groups. Group I contained MeGolS5–7, AtGolS2–3. Group II contained MeGolS4, MeGolS8, and AtGolS1. Group III contained MeGolS1–2, AtGolS5–6. Group IV contained MeGolS3, AtGolS4, and AtGolS7. To investigate the GolS gene structures, the GSDS online software was used to analyze the MeGolS and AtGolS gene exon-intron structures. Except for MeGolS8 having 5 exons and 4 introns, all the other GolS genes had 4 exons and 3 introns. Both MeGolS8 and AtGolS7 did not have upstream and downstream factors. MeGolS7, AtGolS5, and AtGolS6 contained only downstream and no upstream features. The other genes had both upstream and downstream aspects. Obviously, the first intron length of MeGolS8 was significantly longer than the other genes, at more than 8 Kb (Figure 2B). To predict the function of MeGolS proteins, the conserved motifs of GolSs from cassava and Arabidopsis were compared. Ten conserved motifs of GolS proteins were detected by the MEME online tool (Figure 2C, Table S2). These conserved motifs ranged in length
from 11 to 50 aa. All the GolS proteins from cassava and Arabidopsis contained the motifs of 1, 2, 3, 4, 5, 7, 8, and 9. However, motif 6 was found in all the GolS proteins except MeGolS8; Motif 10 was only distributed on MeGolS4, 5, 6, 7, 8, not on any AtGolSs. Thus, it was deduced that the same conserved motifs in the homologous GolSs might have similar functions; and motif 10 in MeGolSs might have a special function that is different from Arabidopsis.

3.4. The Cis-Element Analysis within MeGolS Family Gene Promoters

To better uncover the potential function of the MeGolS genes, the cis-elements within these promoters were identified by PlantCARE. A total of 74 types of cis-acting elements were detected in the promoters of MeGolSs (Table S3), among which were 10 types of hormone-related, 7 tissue specific-related, and 7 stress-related elements (Figure 3). These hormone-related elements included abscisic acid responsive elements (ABRE), MeJA-responsive elements (CGTCA-motif and TGACG-motif), ethylene-responsive elements (ERE), salicylic acid responsive elements (TCA-element), gibberellin-responsive (GARE-motif, P-box and TATC-box), and auxin-responsive elements (TGA-box and TGA-element). ABRE elements were found to be the most abundant hormone-related elements in six of eight MeGolSs. MeGolS4 had nine of the 10 hormone-related elements. MeGolS1 and MeGolS2 had one hormone-related element ABRE or TCA-element, respectively. These tissue specific-related elements included a shoot-specific expressive element (as-2-box), meristem expressive elements (CAT-box and CCGTCC-box), endosperm expressive elements (GCN4-motif and Skn-1-motif), a seed-specific expressive element (RY-element), and a nodule-specific expressive element (Nodule-site2). The Skn-1 motif was the most abundant tissue specific-related element in MeGolS promoters. The numbers of tissue specific-related elements in each MeGolS promoter ranged from one to three. These stress-responsive elements included heat stress (HSE), low-temperature stress (LTR), drought stress (MBS), wound stress (WUN-motif), anaerobic stress (ARE), anoxic stress (GC-motif), and defense and stress responsive element (TC-rich repeats). At least three of these stress-responsive elements were found to be included in each MeGolS promoter. HSE was the most abundant stress responsive related element and was found in all the 8 MeGolS promoters.

Figure 1. Chromosomal locations of 8 predicted MeGolS genes. The chromosome number is marked upon each chromosome. The red arrows indicate gene orientation on chromosomes. The number below each chromosome indicates chromosome size. The number next to each gene represents the position of the gene on the chromosome.
Figure 2. Comparative analysis of the phylogenetics, structure, and conserved motifs of GolSs from cassava and *Arabidopsis*. (A) The phylogenetic tree of MeGolSs and AtGolSs was constructed by using MEGA 7.0 and all the GolSs were classed into four groups. (B) GSDS software was employed to generate the gene structures of MeGolSs and AtGolSs. The yellow boxes are CDS, the blue boxes are 5′ UTR or 3′ UTR, and the black lines are introns. (C) Conserved motifs of MeGolSs and AtGolSs proteins from cassava and *Arabidopsis*. The gene order in B and C is similar to A.
Table 1. Information of 8 MeGolSs.

<table>
<thead>
<tr>
<th>GenBank ID</th>
<th>Gene Name</th>
<th>Locus Name</th>
<th>Chromosomal Location</th>
<th>Genomic Position (5′-3′)</th>
<th>gDNA (bp)</th>
<th>ORF (bp)</th>
<th>Protein Length (aa)</th>
<th>pI</th>
<th>MW (KD)</th>
<th>Exon</th>
<th>GRAVY</th>
<th>Subcellular Location Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>KF793045</td>
<td>MeGolS1</td>
<td>Manes.03G208800</td>
<td>Chr03</td>
<td>28896616–28898137</td>
<td>1522</td>
<td>984</td>
<td>327</td>
<td>5.04</td>
<td>37.70</td>
<td>4</td>
<td>–0.378</td>
<td>Cytoplasmic</td>
</tr>
<tr>
<td>KF793046</td>
<td>MeGolS2</td>
<td>Manes.15G003200</td>
<td>Chr15</td>
<td>272870–274578</td>
<td>1709</td>
<td>963</td>
<td>320</td>
<td>5.2</td>
<td>37.14</td>
<td>4</td>
<td>–0.339</td>
<td>Cytoplasmic</td>
</tr>
<tr>
<td>KJ722606</td>
<td>MeGolS3</td>
<td>Manes.13G029000</td>
<td>Chr13</td>
<td>2755754–2759184</td>
<td>3431</td>
<td>1017</td>
<td>338</td>
<td>5.27</td>
<td>38.60</td>
<td>5</td>
<td>–0.193</td>
<td>Cytoplasmic, Chloroplast</td>
</tr>
<tr>
<td>KJ722605</td>
<td>MeGolS4</td>
<td>Manes.05G012000</td>
<td>Chr05</td>
<td>822143–824202</td>
<td>2060</td>
<td>1008</td>
<td>335</td>
<td>5.39</td>
<td>38.30</td>
<td>4</td>
<td>–0.213</td>
<td>Cytoplasmic</td>
</tr>
<tr>
<td>KJ722604</td>
<td>MeGolS5</td>
<td>Manes.08G109300</td>
<td>Chr08</td>
<td>27446317–27447941</td>
<td>1625</td>
<td>1008</td>
<td>335</td>
<td>5.08</td>
<td>38.10</td>
<td>4</td>
<td>–0.162</td>
<td>Plasma Membrane</td>
</tr>
<tr>
<td>KJ722603</td>
<td>MeGolS6</td>
<td>Manes.09G179100</td>
<td>Chr09</td>
<td>28784797–28786402</td>
<td>1606</td>
<td>1014</td>
<td>337</td>
<td>5.47</td>
<td>38.19</td>
<td>4</td>
<td>–0.19</td>
<td>Plasma Membrane</td>
</tr>
<tr>
<td>KJ816317</td>
<td>MeGolS7</td>
<td>Manes.08G109400</td>
<td>Chr08</td>
<td>27451435–27453176</td>
<td>1742</td>
<td>1005</td>
<td>334</td>
<td>5.33</td>
<td>38.10</td>
<td>4</td>
<td>–0.167</td>
<td>ER</td>
</tr>
<tr>
<td>MH535931</td>
<td>MeGolS8</td>
<td>Manes.01G231500</td>
<td>Chr01</td>
<td>31476735–31486544</td>
<td>9810</td>
<td>984</td>
<td>327</td>
<td>5.27</td>
<td>37.35</td>
<td>5</td>
<td>–0.182</td>
<td>Cytoplasmic</td>
</tr>
</tbody>
</table>

gDNA, genomic DNA; ORF, open reading frame; aa, amino acid; MW, molecular weight; pI, isoelectric point; GRAVY, Grand average of hydropathicity.
Figure 3. Cis-acting elements in cassava GolS gene promoters. (A) Number of every selected cis-acting element in the MeGolS gene promoters. (B) Statistics for total number of MeGolS genes containing the corresponding cis-acting elements (red dots) and the total number of cis-acting elements in every MeGolS (blue boxes).

3.5. Tissue-Specific Expression of MeGolS in Cassava

To understand the expression abundance of the MeGolS gene family in different tissues or organs, 11 tissues or organs of cassava were analyzed based on the RNA-seq datasets from cassava. The heatmap results showed that the eight MeGolS genes had different expression patterns in tissues and organs (Figure 4). Seven of eight MeGolS genes were expressed in mid-vein (except MeGolS7), fibrous roots (FR) (except MeGolS1), stem (except MeGolS7), and petiole (except MeGolS2); six of eight MeGolS genes were expressed in root apical meristem (RAM) (except MeGolS1, and MeGolS7); five of eight MeGolS genes were expressed in somatic organized embryogenic structures (OES) (except MeGolS1, 2, and 7), shoot apical meristem (SAM) (except MeGolS1, 2, and 8); only 4 MeGolS genes (MeGolS3, 4, 6, and 8) were expressed in storage roots (SR). The genes with the highest expression levels were, MeGolS3 in leaf, mid-vein, stem, and petiole; MeGolS4 in the lateral bud, OES, FEC, SR, and SAM, MeGolS5 in RAM; and MeGolS6 in FR. At gene expression levels, three MeGolS genes (MeGolS3, 4 and 6) were expressed in all 11 tissues and organs; MeGolS5 and 8 were slightly expressed in 10 tissues and organs, respectively, except MeGolS5 in SR and MeGolS8 in SAM; MeGolS2 is slightly expressed in seven tissues and organs except OES, SR, petiole, and SAM; MeGolS1 was only expressed in leaf, mid-vein, lateral bud, FEC, stem, and petiole; while MeGolS7 was only slightly expressed in leaf, lateral bud, FEC, FR, petiole, and SAM.
3.6. The Inducible Expressions of MeGolSs under Abiotic Stress

To establish the MeGolS gene expressions under abiotic stresses, quantitative real-time RT-PCR data from different abiotic stressors (drought, cold, and salinity) and different treatment times (3 h, 6 h, 9 h, 12 h, and 24 h) were gathered to construct a heatmap. The expression abundance of each MeGolSs was obviously different among different stresses and treatment times (Figure 5). The expression patterns of MeGolS5 and MeGolS6 were similar in response to drought stress, during which both were upregulated and peaked at 3 h; MeGolS1 and MeGolS2 had similar expression patterns and peaked at 12 h; MeGolS3, 4, and 8 were downregulated. Regarding response to cold stress, six of eight tested MeGolSs except MeGolS3 and 8 were upregulated, during which MeGolS5 and MeGolS6 had similar expression patterns, with higher expressions from 3 h to 24 h except at 12 h; MeGolS1, MeGolS2, and MeGolS7 peaked their expressions at 12 h; MeGolS4 was slightly upregulated, and peaked at 24 h; however, MeGolS3 and 8 were downregulated at all time points. MeGolS1 was highly induced at all
time points in response to salinity and peaked at 12 h; MeGolS5 and MeGolS7 had similar expression patterns, which were downregulated at 3 h, then upregulated from 6 h to 24 h, with a peak expression at 12 h; MeGolS2 and 6 were downregulated from 3 h to 9 h, and upregulated from 12 h to 24 h; MeGolS2 was highly expressed at 24 h; and MeGolS4 was upregulated at 6 h and 24 h. MeGolS3 and 8 were downregulated in all stages.

Figure 5. Expression abundance of MeGolSs under drought, cold, and salinity abiotic stresses in cassava. The bars indicated the relative gene expression levels, calculated based on the $2^{-\Delta\Delta C_{t}}$ method. The expression levels were equal to the mean values and transform log2 values.

3.7. Prediction of MicroRNA Targeting MeGolS Genes

To predict the microRNA (miRNA) targeting MeGolS genes, the genomic sequences of all MeGolS genes were used to BLAST against the cassava microRNA database. A total of 70 putative miRNAs were predicted to target eight MeGolS genes (Figure 6, Table S4). The numbers of miRNAs predicted to target MeGolS1–8 were 13, 2, 6, 3, 30, 3, 8, and 19, respectively. Nine mes-miR156 families and two mes-miR403 families targeted MeGolS1. Twenty-eight members of the mes-miR169 family and two members of the mes-miR2111 family target MeGolS5. MeGolS8 is targeted by six miRNA families, including mes-miR319, mes-miR171, mes-miR1446, mes-miR2275, mes-miR403, and mes-miR827. Two mes-miR403 families target MeGolS2. Four mes-miR164 families and two mes-miR403 families target MeGolS3. mes-miR2275 and two mes-miR159 families target MegolS4. mes-miR397 and two
mes-miR172 families target MeGolS6. Four mes-miR169 families, two mes-miR159 and two mes-miR828 families target MeGolS7. Furthermore, mes-miR403a and mes-miR403b were identified to target three MeGolS genes (MeGolS1, MeGolS5 and MeGolS8), the most targeted MeGolS genes in this study. Both mes-miR403a and mes-miR403b were found to have two target sites on MeGolS8 as well.

![Figure 6. Prediction of microRNA targeting MeGolS genes.](image)

The red dot indicates the MeGolS genes, the blue dot indicates the miRNAs that targeting MeGolSs.

4. Discussion

Cassava has a strong tolerance to drought and infertility but is sensitive to low temperature and salinity. However, there still insufficient information to reveal the mechanisms by which cassava treats various abiotic stress. Generally, raffinose family oligosaccharides (RFOs) have been reported to be kinds of important metabolites in response to various abiotic stresses including drought, cold, and salinity. Galactinol synthase, as a key enzyme in RFO synthases, plays vital roles in various aspects of the plant growth, including seed maturation and abiotic and biotic stress tolerances. However, the systematic analysis of the GolS gene family in cassava is still lacking. GolS family genes have been identified in seven GolS genes from Arabidopsis [11], nine GolS genes from poplar (Populus trichocarpa Torr. & Gray) [33], four GolS genes in tomato [16], nine GolS genes from tobacco [12], three GolS genes from tea plant [14], and others. Eight GolS genes were isolated from cassava in this study. The diversity of gene number might be associated with the diversity in evolution and genome duplication or genome sizes in plants.

Phylogenetic analysis revealed that GolS proteins in cassava and Arabidopsis are divided into four groups. This classification is consistent with that previous reported in other species, such as poplar [33], tomato [16], rapeseed [12], tobacco [12], and sesame (Sesamum indicum L.) [34]. The proteins clustered together might have similar functions. MeGolS4 and MeGolS8 are clustered with AtGolS1, MeGolS5–7 are clustered with AtGolS2–3, MeGolS1–2 are clustered with AtGolS5 and AtGolS6, and MeGolS3 is clustered with AtGolS4 and AtGolS7 (Figure 2A) in this study. Previous studies indicated that
AtGolS1, AtGolS2, and AtGolS3 are related to response to abiotic stresses, such as drought, salinity [11], and temperature stresses [19]. Thus, MeGolS4–8 might have the function to respond to abiotic stresses.

Genomic comparison is a rapid means of obtaining knowledge about less-studied taxon. Eight MeGolS protein sequences have been identified and isolated in this study, which have a very close amino acid length (320–338 amino acids) and molecular weight (37.1–38.6 kDa) with FI value in the range of 5.04–5.47. Similar results can be found in other plant species, including coffee (Coffeea canephora Pierre ex Froehn) [35], cotton (Gossypium hirsutum L.) [36], tomato [16,37], and purple false-brome (Brachypodium distachyon) [16]. All the eight MeGolS genes in this study have four exons and three introns. According to previous research, the exon numbers of GolS genes in different species vary between two and four, such as GolS genes in cotton have three exons [36], in pea (Pisum sativum L.) with at least three exons [38], in tomato with three to four exons, and in B. distachyon with four exons [16]. This suggests that there is a slightly different gene structure diversity of GolS genes in different species.

Expression abundance of MeGolS genes in different tissues or organs of cassava revealed that MeGolS genes exhibit incongruous expression patterns in various tissues (Figure 4), indicating that different MeGolS proteins might have diverse functions. The tissue specific expression levels of GolS genes have been described in other species, including poplar [33], tomato [16,37], rapeseed [12], tobacco [12], and sesame [34]. Previous research indicates that GolS expression in leaves is associated with phloem export, carbon storage, and plant protection against abiotic stress and oxidative damage [39]. GolS expression in roots is associated with the osmotic stress [40]. In this study, MeGolS1 and MeGolS3–6 are highly expressed in leaves and midveins. MeGolS3–6 are highly expressed in fibrous roots. It is reported that Pa3gGolSII homologue and BnGolS3 subfamily members were highly expressed in leaves [3,12]. However, not all the plants have high GolS gene expression in leaves. For example, there were no highly expressed GolS genes observed in tobacco leaves [12]. Thus, these tissue specific expression MeGolS genes might have important functions in phloem export, carbon storage, and plant protection against abiotic stress and oxidative damage.

Plants frequently face abiotic stressors such as drought, cold, and high salinity. In order to help plants avoid the damage of adverse conditions, researchers are constantly exploring genetic resources that are beneficial to plants against cold [41–44], drought [45–47], salt [48–51], and so on. In this paper, qRT-PCR was employed to calculate the transcript levels of each MeGolS under different abiotic stressors. Differential expressions of GolS linking to abiotic stresses have been reported in several plants [11,12,16,33,34,52]. The stresses of salt, low temperature, and drought could be strongly represented by MeGolS genes in the current study. The response to the three treatments varied among the respective MeGolS genes (Figure 4), which also was found in other plants, for instance in Verbascum phoeniceum L., where VpGolSs showed diverse expression under salt, cold, and drought stresses [53]. In poplar, the expression of Pa × gGolSII isoform were highly temperature related, while the Pa3gGolS isoform persistent expression throughout the year [3], which suggests that each member of the gene family may play a different physiological role. MeGolS5 and MeGolS6 were induced by drought stress and could rapidly peak at three hours after drought treatment. MeGolS1, 2 and 7 were also responsive to drought stress (Figure 4). As described in the phylogenetic analysis, MeGolS5–7 are clustered with AtGolS2 and 3 (which were proven to be significant for drought stress [11]), thus there was no surprise that they were induced by drought stress. MeGolS5 was highly responsive to cold, followed by MeGolS6 (Figure 4). These results indicated that the GolS genes expression mechanisms under drought and cold stress might be similar. MeGolS1 could rapidly respond to salt stress and reached peak expression at 12 h, thus it is a very salt sensitive MeGolS gene. Additionally, MeGolS2 and 7 are involved in salt stress, but their response is comparatively weaker than MeGolS1. The upregulated GolS genes to salt stress are also found in other plants, such as AtGolS1 and 2 from Arabidopsis [11], PtrGolS3, 4, and 6 from P. Trichocarpa [52]. However, in cassava, the most salt sensitive gene, MeGolS1, is not clustered with AtGolS1 and AtGolS2 in phylogenetic analysis. Thus, the current results indicate that cassava might have a specific mechanism to respond to salt stress. Generally, several MeGolS genes are induced by
salt, drought, and cold treatments, indicating that these genes might be crucial for cassava response to abiotic stresses.

miRNAs were crucial for regulating gene expression to adapt to biotic and abiotic stresses [54–57]. There were 14 miRNAs in wild eggplant roots (*Solanum linnaeanum* L.) [58], 71 miRNAs in radish (*Raphanus sativus*, L.) [59], 123 miRNAs in tomato [60], and 37 miRNAs in maize (*Zea mays* L.) [61] that showed significant changes in expression under salt stress. There were 18 miRNAs in rice (*Oryza sativa* L.) [62] and 30 miRNAs in *Populus tomentosa* Carr [63] that were identified to be cold-responsive. There were also 35 miRNAs in tef (*Eragrostis tef*, (Zucc.) Trotter) [64], 13 miRNAs in wild emmer wheat (*Triticum dicoccoides* ssp. *dicoccoides* (Korn.) Thell.) [65], and 34 miRNAs in citrus (*Citrus junos* Siebold) [66] responses to drought stress. Regarding cassava, 881 miRNAs have been reported to identified by high-throughput sequencing [67], among which another 38 new miRNAs have been identified from cassava [68]. Here, 70 mes-miRNAs have been predicted to target eight *MeGolS* genes in cassava. These miRNAs belong to 14 miRNA families, mes-miR156, 159, 169, 171, 172, 319, 397, 403, 827, 828, 2111, 1446, and 2275. Previous studies indicate that miR156, 159, 169 are related to temperature, drought, and salt stresses [69–76]. Thus, the cleavage or translation activities of mes-miR156, 159, and 169 targeted to *MeGolS1*, *MeGolS5*, *MeGolS6*, or *MeGolS7* might account for the qRT-PCR results of these genes response to abiotic stresses. The prediction of miRNA targeting to *MeGolS* may be useful to understanding the regulatory network of *MeGolS* involved in responding to abiotic stress in cassava.

5. Conclusions

In the cassava genome, 8 *MeGolSs* were identified and their physicochemical properties, gene structure, protein motifs, and classification were investigated. By combining promoter cis-element prediction, expression patterns in tissue and in response to abiotic stress, and with the prediction of miRNA targeting to the *MeGolSs*, we have increased the functional knowledge of MeGolS genes. Finally, these results would be useful in further study of the biological functions of *MeGolSs* in cassava.

**Supplementary Materials:** The following are available online at [http://www.mdpi.com/2073-4395/8/11/250/s1](http://www.mdpi.com/2073-4395/8/11/250/s1), Table S1: The primers were used for gene cloning and qRT-PCR analysis of *MeGolS* genes, Table S2: List of the putative motifs of GolS proteins from cassava and Arabidopsis, Table S3: Predicted cis-acting elements in the promoter regions of *MeGolSs*, Table S4: Predicted miRNAs targeted to *MeGolS* genes.

**Author Contributions:** R.L. and Y.H. were responsible for all aspects of the research, including experimental design, data acquisition and analysis, and manuscript preparation. S.Y., J.F., Y.Z., T.Q. and X.L. worked on the preparation of the studied materials, gene cloning and qRT-PCR. Y.Y., J.L. and S.F. worked on primer design, and technical and informatics analyses of these genes. X.H. and J.G. were responsible for the programs and all experiments, critically revised the manuscript, and provided the final approval of the article.

**Funding:** This research was funded by the National Natural Science Foundation of China (No.31601359; 31600196 and 31671767), the key research and development projects of Hainan Province (No. ZDYF2017073), the Earmarked Fund for China Agriculture Research System (No.CARS-11-HNGJC), and Central Public-interest Scientific Institution Basal Research Fund for Chinese Academy of Tropical Agricultural Sciences (No. 1630052016004).

**Acknowledgments:** We fully appreciate the editors and all anonymous reviewers for their constructive comments on this manuscript.

**Conflicts of Interest:** The authors declare that they have no competing interests.

**References**


40. Lahuta, L.B.; Góręcki, R.J. Raffinose in seedlings of winter vetch (Vicia villosa, Roth.) under osmotic stress and followed by recovery. Acta Physiol. Plant. 2011, 33, 725–733. [CrossRef]


69. Cui, N.; Sun, X.; Sun, M.; Jia, B.; Duanmu, H.; Lv, D.; Duan, X.; Zhu, Y. Overexpression of OsmiR156k, leads to reduced tolerance to cold stress in rice (Oryza sativa). Mol. Breed. 2015, 35, 214. [CrossRef]


