http://journal.unoeste.br/index.php/ca/index
DOI: 10.5747/ca.2024.v20.h525
ISSN on-line 1809-8215

Colloquium Agrariae

Submetido: 07/10/2023 Revisado: 23/02/2024 Aceito: 06/03/2024

CHARACTERIZATION OF *GALACTINOL SYNTHASE* (*GOLS*) GENES IN *IPOMOEA TRILOBA* AND *I. TRIFIDA*: AN APPROACH USING BIOINFORMATICS

Wuesly Rufino[1], Luís Gustavo Gomes Lobo[1], Diliane Harumi Yaguinuma[1], Edinara Maria Barbosa[2], Tiago Benedito dos Santos[1], Silvia Graciele Hülse de Souza[2]

[1]Universidade do Oeste Paulista – UNOESTE, Presidente Prudente, SP. [2]Universidade Paranaense – UNIPAR, PR. E-mail: silviahulse@prof.unipar.br

Abstract

Galactinol synthase (GolS - EC 2.4.1.123) is classified as a key enzyme that catalyzes the first step in the synthesis pathway of the raffinose family (RFOs). Although *GolS* genes have been characterized in several important species, their characterization in sweet potatoes has yet to be explored. *Ipomoea trifida* (Kunth) G. Don (2n = 2x = 30) is currently described as one of the closest ancestors of the sweet potato and considered an excellent crossbreeding species, allowing the introgression of important genes such as *GolS*. This study aimed to identify and characterize *in silico* the *GolS* genes in *I. trifida* and compare with *I. triloba*. We identified nine *GolS* genes, five in *I. triloba* and four in *I. trifida*. Our study encompassed various aspects, including gene structure analysis, motif identification, chromosomal distribution, synteny analysis, and gene expression. The presence of gene duplications and purifying selection were highlighted, suggesting the evolutionary significance of *GolS* genes in these species. Phylogenetic analysis categorized *GolS* proteins into three groups, potentially reflecting distinct functional roles. Furthermore, synteny analysis revealed orthologous relationships between *GolS* genes in the studied species and related plants, contributing to our understanding of their evolutionary history. *In silico* expression analysis across diverse tissues unveiled tissue-specific expression patterns, hinting at specialized roles for *GolS* genes in different plant organs. These findings contribute to the broader field of plant genetics, carbohydrate metabolism, and agriculture, offering opportunities for crop improvement and sustainable food production.

Keywords: bioinformatics, genetic improvement, sweet potato, raffinose family (RFO).

Caracterização dos genes *galactinol sintase* (*GolS*) em *Ipomoea triloba* e *I. trifida*: uma abordagem usando bioinformática

Resumo

A galactinol sintase (GolS - EC 2.4.1.123) é classificada como uma enzima chave que catalisa a primeira etapa na via de síntese da família da rafinose (RFOs). Embora os genes *GolS* tenham sido caracterizados em diversas espécies importantes, a sua caracterização na batata-doce ainda não foi explorada. *Ipomoea trifida* (Kunth) G. Don (2n = 2x = 30) é atualmente descrita como um dos ancestrais mais próximos da batata-doce e considerada uma excelente espécie de cruzamento, permitindo a introgressão de genes importantes como o *GolS*. Este estudo teve como objetivo identificar e caracterizar *in silico* os genes *GolS* em *I. trifida* e compará-los com *I. triloba*. Identificamos nove genes *GolS*, cinco em *I. triloba* e quatro em *I. trifida*. Nosso estudo abrangeu vários aspectos, incluindo análise da estrutura genética, identificação de motivos, distribuição cromossômica, análise de sintenia e expressão gênica. A presença de duplicações gênicas e seleção purificadora foram destacadas, sugerindo o significado evolutivo dos genes *GolS* nessas espécies. A análise filogenética categorizou as proteínas GolS em três grupos, refletindo potencialmente papéis funcionais distintos. Além disso, a análise de sintenia revelou relações ortólogas entre os genes *GolS* nas espécies estudadas e plantas relacionadas, contribuindo para a nossa compreensão da sua história evolutiva. A análise de expressão *in silico* revelou padrões de expressão específicos em diferentes tecidos, sugerindo papéis especializados para genes *GolS* em diferentes órgãos vegetais. Estas descobertas contribuem para o campo mais amplo da genética das plantas, do metabolismo de carboidrato e da agricultura, oferecendo oportunidades para o melhoramento das culturas e a produção sustentável de alimentos.

Palavras-chave: bioinformática, melhoramento genético, batata-doce, família da rafinose (RFO).

Introduction

Galactinol synthase (GolS, EC 2.4.1.123) is classified and named as an enzyme that catalyzes the first reaction in the biosynthetic pathway of the raffinose family of oligosaccharides (RFOs) in plants (Salvi *et al.*, 2020). We can say that galactinol is mentioned as a precursor and is also involved in plant defense. It is a galactosyl donor that generates an essential molecule of oligosaccharides (RFOs) of the raffinose, stachyose, and verbascose family, among others (Sengupta *et al.*, 2012; Meyer *et al.*, 2018). The biosynthetic pathway of galactinol and RFOs catalyzed by GolS is very detailed in some plant species (Sengupta *et al.*, 2012; Santos; Vieira, 2020). This GolS belongs, according to some studies, to family 8 (GT8) of glycosyltransferases, a

family of enzymes responsible for the synthesis of various sugars, and is related to the storage, energy, and signaling of important molecules (Jing *et al.*, 2023). The GT8 glycosyltransferase family (GTs – EC 2.4.x.y.) can be confirmed through confirmed in the CAZy tool (Carbohydrate Active EnZymes; CAZypedia Consortium 2018).

GolS gene has been identified and characterized in different plant species, both in monocot and dicot: *Arabidopsis thaliana* (Nishizawa *et al.*, 2008), *Coffea arabica* L. and *C. canephora* (Santos *et al.*, 2011; 2015), *Pisum sativum* (Lahuta *et al.*, 2014), *Solanum lycopersicum* and *Brachypodium distachyon* (Filiz *et al.*, 2015), *Manihot esculenta* Crantz (Li *et al.*, 2018), *Musa acuminata* (Dolcimasculo *et al.*, 2018), *Panicum virgatum L.* and *P. hallii* (Góis *et al.*, 2020), *Phaseolus vulgaris* L. (Koning *et al.*, 2023). Plants have more than one isoform of *GolS* and can act on specific tissues (Sengupta *et al.*, 2012; revised by Santos; Vieira, 2020). In the recent past, we also found information in the literature that the expression of *GolS* is in the involvement, for example, of environmental adversities (Santos *et al.*, 2011; 2015; Kim *et al.*, 2011; Santos; Vieira, 2020). In *C. arabica* L., Santos *et al.* (2011) identified three *CaGolS* isoforms (*CaGolS1*, *2*, and *3*) and evaluated their expression under abiotic stress conditions. Mukherjee *et al.* (2019) studied two isoforms of *GolS* and reported that these genes play an essential role in transcriptional and post-transcriptional regulation mechanisms under stress conditions.

Sweet potato (*Ipomoea batatas* (L.) Lam.) (2n = 6x= 90) is a root crop important to the family Convolvulacea. It belongs to the genus *Ipomoea* and includes approximately 500-600 species (Austin, 1988). Austin (1988) suggested that the wild ancestor of sweet potato originated from the natural hybridization between *Ipomoea triloba* and *I. trifida* (diploid ancestors). The recent complete genome sequences these two diploid species, named as ancestors, *I. triloba* and *I. trifida*, are important resources to investigate genes in the hexaploid sweet potato, (*I. batatas* (L.) Lam.), respectively (Wu *et al.*, 2018). For example, there are indications of interrelationships between *I. triloba* and *I. trifida*, among other species, was recognized by Van Ooststroom in 1953, and these species were included in the Potatoes section, respectively.

Despite its crucial role in agriculture, this crop remains underexplored, with particularly limited research focused on genetic enhancement for this specific species. Furthermore, there is a need for more extensive information regarding genotypes well-suited to the conditions faced by producers. In light of these gaps, based on genome sequencing of this species, our study involved a detailed investigation of *GolS* genes with the use of bioinformatics tools. This investigation thoroughly examined gene structure, motif analysis, chromosomal distribution, syntenic analysis, and gene expression patterns.

Material and Methods

First, the galactinol synthase's predicted amino acid sequences, genomics, and coding sequence (CDS) were downloaded from the Sweet Potato Database (https://sweetpotato.plantbiology.msu.edu/). To confirm, all sequences were selected and submitted to the National Center for Biotechnology Information database (NCBI) (Altschul *et al.*, 1990) using the BLASTP tool. We used the PROTPARAM tool (Expert Protein Analysis System – ExPASy - http://expasy.org/tools/) to determine the molecular weight (MW) and isoelectric points (pIs) of the ItlbGolSs and ItfGolSs proteins. Also, the subcellular location of these genes was passed through the Plant-mPloc software (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/) (Chou; Shen, 2010). We determined the exon–intron structure of the *ItlbGolS* and *ItfGolS* genes by the Gene Structure Display Server program (GSDS; http://gsds.cbi.pku.edu.cn/) (HU *et al.*, 2015). In addition, we did conserved motif analysis using the MEME algorithm (Multiple Em for Motif Elicitation - https://meme-suite.org/meme/tools/meme), according to Martins *et al.* (2022). Individually, the physical locations of *GolS* genes were obtained from the Sweet Potato Database, and posteriorly, the map of the chromosome location of genes was constructed through the Mapgene2chrom v. 2.1. software (Chao *et al.*, 2015). We then further investigated the synonymous (*Ks*) and non-synonymous (*Ka*) substitution rates of the paralogs genes by using the Ka_Kscalculator 2.0 (Zhang *et al.*, 2006). Multiple sequence alignment was performed using ClustalW (Thompson *et al.*, 1994), and we constructed a phylogenetic tree using MEGA7.0 according to the Neighbor-Joining (NJ) method and performed bootstrap testing with 1000 iterations in all analyses (Kumar *et al.*, 2016). To analyze the *GolS* genes and establish orthologous relationships among *I. trifida* and *A. thaliana*; *I. trifida* and *O. sativa*; and *I. trifida* and *I. triloba,* we carried out reciprocal BlastP. The hit threshold values were set as E-value <1e-50, score >200, minimum 80% coverage, and 70% identity (Huang *et al.*, 2016). Then, the ClicO FS tool (http://103.47.253.210:3000/home) was used to represent the syntenic relationships. Expression patterns *in silico* of *GolS* were verified in different tissues named (flower, flower bud, leaf, root1, root2, stem, in *I. triloba* genome), and (callus_flower, callus_stem, flower, flower bud, leaf, root1, root2, stem, in *I. trifida* genome) that were obtained from RNA-seq data available at the Sweet Potato Database.

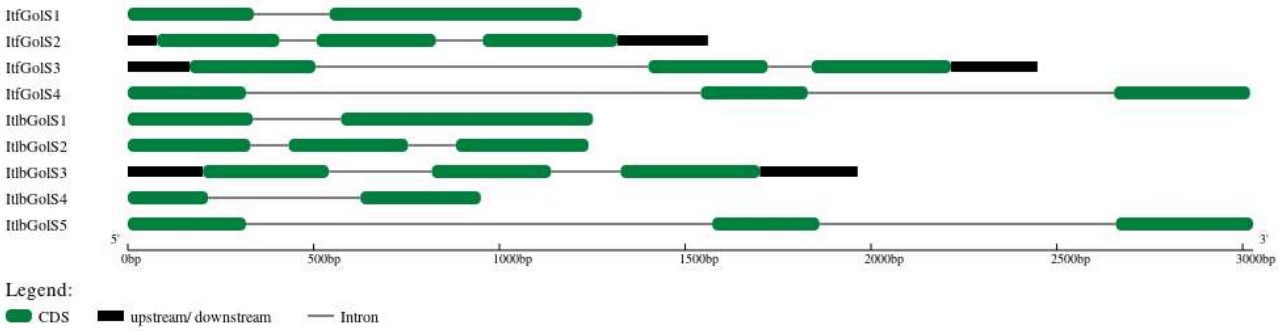
Results and Discussion

Especially in this study, we identified a total of nine *GolS* genes, which were named according to their position on the chromosomes plus the species name, *I. triloba* (five "Itlb" GolS) and *I. trifida* (four "Itf" GolS), respectively (see Table Supplementary 1). All sequences are available at the Sweet Potato Database (https://sweetpotato.plantbiology.msu.edu/). We retrieved all the sequences and checked them one by one, and they were confirmed as having the conserved GT8

domain (PF01501) characteristic of this gene using the InterPro tool (www.ebi.ac.uk/interpro). In plants, especially in the GT8 group, these enzymes are derived from a single ancestral sequence, and it is believed that a part of the sequence was lost and evolved differently in different existing plant species (Sengupta *et al.*, 2012). We observed that the *GolS* genes of *I. triloba* encode proteins ranging from 179 (ItlbGolS4 - 20.45 kDa) to 344 (ItlbGolS3 - 38.96 kDa) amino acids in length and pI values ranging from 5.40 (ItlbGolS3) to 9.22 (ItlbGolS4) (Table Supplementary 1). In *I. trifida*, the *GolS* genes encode proteins ranging from 323 (ItfGolS4 - 37.10 kDa) to 344 (ItfGolS3 - 38.97 kDa) amino acids in length and pI values ranging from 5.40 (ItfGolS3) to 8.12 (ItfGolS4) (Table Supplementary 1). Martins *et al.* (2022) observed that all our proteins have a negative GRAVY value, indicating that all proteins are hydrophilic (Table Supplementary 1). The number of *GolS* genes identified in the genomes of *I. triloba* and *I. trifida* varied somewhat from that found in other plant species, such as eight in *A. thaliana* (Sprenger; Keller, 2000) and *Malus* × *domestica* (Falavigna *et al.*, 2018) and nine in *Populus trichocarpa* (Zhou *et al.*, 2014) and *N. tabacum* (Fan *et al.*, 2017) eight in *Citrus sinensis* (Martins *et al.*, 2022). Subcellular localization predictions for ItlbGolS and ItfGolS suggest that diversity is found in the cell wall, chloroplast, cytoplasm, mitochondrion, peroxisome, cell membrane, and nucleus (Table Supplementary 1). As previously reported, some subcellular locations are in the same membrane tissue (Santos *et al.*, 2011; Fan *et al.*, 2017; Dolcimasculo *et al.*, 2018).

Over the years, genes can show diversity and a specific signature species, and such information can provide researchers with the behavior of that gene (Cao; Shi, 2012). In the present study, we analyzed the structure of all putative genes mentioned above. We used CDS and genomic DNA sequences. The *ItlbGolS* and *ItfGolS* genes exhibited one or two introns (Figure 1). Based on this valuable information, it can be suggested that in sweet potato, this gene presents less diversity in its structure. In *Musa acuminata*, the exon/intron of the *GolS* genes shows three or four exons (Dolcimasculo *et al.*, 2018). According to Figures 1 and 2, there may be a variation in the number of exons and introns in the *ItlbGolS* and *ItfGolS* genes, and we can also suggest that some genes are duplicated. Previously, Filiz *et al.* (2015) observed variations in the structure and number of genes.

Figure 1. Illustration of the structure of *GolS* genes. Green bars represent the exon, gray lines represent the intron, and black color bars indicate the untranslated region (UTR).



In *I. triloba* and *I. trifida*, the distribution of *GolS* genes on chromosomes was similar. Chromosomal location showed that *GolS* genes were randomly distributed on three chromosomes in *I. triloba* and *I. trifida* (Figure 2). In *I. triloba*, two *ItbGolS genes* were on Chr3 and Chr12 and one on Chr15. In *I. trifida*, two *ItfGolS genes* were on Chr3, and one was on each chromosome: Chr12 and Chr15. Expansion analysis of the *GolS* genes in *I. triloba* and *I. trifida* genomes were examined. Based on their chromosomal distribution and the high rate of sequence similarity, we determined three duplication pairs (Figure 2, Table 1). In *I. trifida* we found one tandem duplication among *ItfGolS1* and *ItfGolS2*, and two duplication pairs arose from segmental events, one between *ItfGolS1*and *ItfGolS3* and another between *ItfGolS2* and *ItfGolS3* genes. The same pattern of duplication events was observed in *I. triloba*. Black lines in Figure 2 show the connections among these paralogs. Our results indicated that *GolS* genes were possibly generated by gene duplication, and the segmental duplication played a major driving force for expansion in the studied species. Duplication enhances functional divergence, a critical factor for adapting to dynamic environmental changes (Conant; Wolfe, 2008).

We have calculated the divergence and evolutionary relationship by *Ka*, *Ks*, and *Ka*/*Ks* ratios between two paralogs *GolS* genes. *Ka*/*Ks* ratio ranged from 0.38 to 0.49 for *GolS* genes as detailed in Table 1. Generally, a *Ka*/*Ks* ratio of less than 1 signifies negative selection, equal to 1 indicates neutral selection, and greater than 1 signifies positive selection (as described by Ali *et al.*, 2017). The observed trends in *Ka*/*Ks* ratios within the coding sequences of the duplicate gene pairs strongly suggest that *GolS* genes in *I. triloba* and *I. trifida* have undergone a process of purifying selection (*Ka*/*Ks* < 1).

Figure 2. Chromosomal distribution and duplication events of the *GolS* genes in sweet potato. **A.** Chromosome of *I. triloba.* **B.** Chromosome of *I. trifida*. The black lines represent duplicated genes (see Table 1). The number of chromosomes and their size in Mb are indicated at the top of each bar, and the vertical scale represents the size of the chromosome.

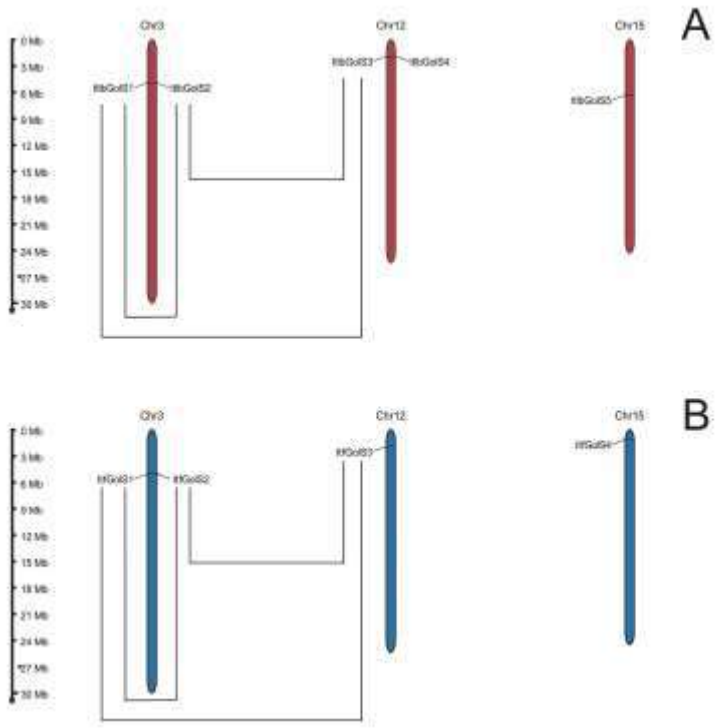


Table 1. Duplication date of paralogous genes pairs among *I. triloba* and *I. trifida GolS* genes. *Ka* represents the non-synonymous substitution number per non-synonymous site, *Ks* is the number of the synonymous substitution site. *Ka*/*Ks* represents the ratio of non-synonymous (*Ka*) to synonymous (*Ks*) substitutions.

Paralogous Pairs	Chromosomal location	Duplication event	*Ka*	*Ks*	*Ka*/*Ks*	Purifyng Selectioin
ItlbGolS1/ ItlbGolS2	Chr03/Chr03	Tandem	0.27	0.55	0.49	Yes
ItlbGolS1/ ItlbGolS3	Chr03/Chr012	Segmental	0.25	0.51	0.49	Yes
ItlbGolS2/ ItlbGolS3	Chr03/Chr12	Segmental	0.16	0.42	0.38	Yes
ItfGolS1/ ItfGolS2	Chr03/Chr03	Tandem	0.27	0.60	0.45	Yes
ItfGolS1/ ItfGolS3	Chr03/Chr12	Segmental	0.26	0.55	0.47	Yes
ItfGolS2/ ItfGolS3	Chr03/Chr12	Segmental	0.16	0.39	0.41	Yes

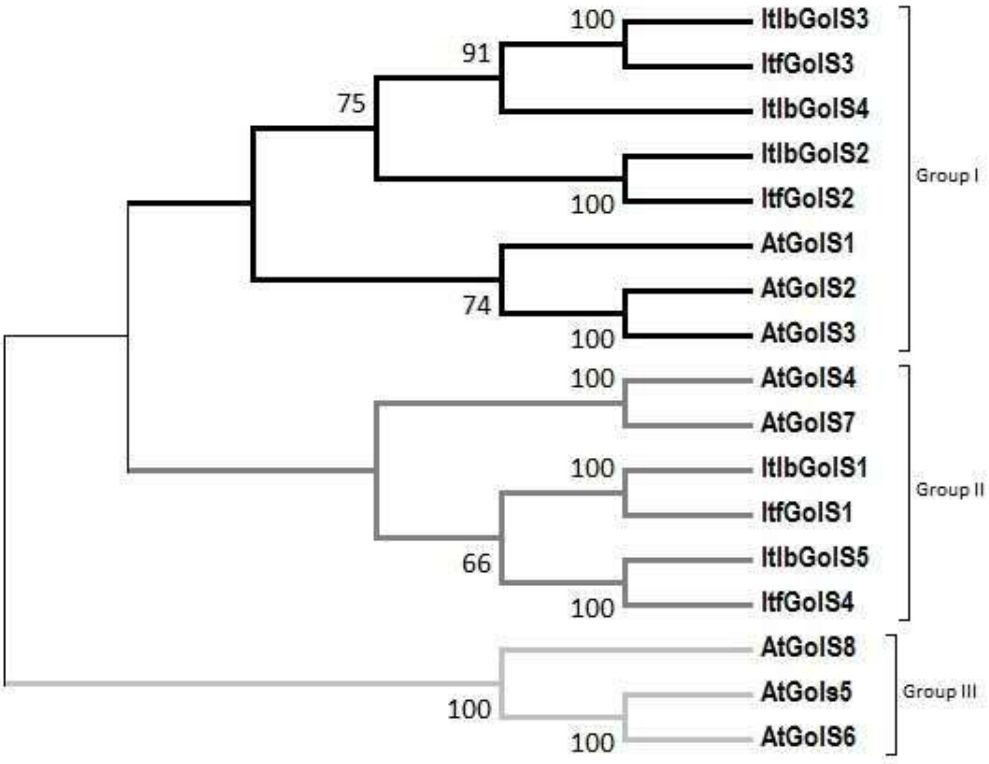
All ItlbGolS and ItfGolS amino acid sequences contained the C-terminal hydrophobic pentapeptide APSAA, a signature of this gene (Sprenger; Keller, 2000) (Figure Supplementary 2).

The conserved catalytic residues of a manganese-ligation motif denominated (DxD) were also observed (Santos *et al.*, 2011; 2015; Zhou *et al.*, 2017- see Figure Supplementary 2).

We performed the multiple sequence alignment to verify the GolS protein sequences from *I. triloba*, *I. trifida*, and *A. thaliana.* In addition, a Neighbor-Joining (NJ) tree was established by MEGA7.0 (Figure 3). We observed that all the GolSs were divided into three (groups I–III), respectively. Group I contained ItlbGolS3-4-2, ItfGolS2-3, and AtGolS1-2-3. Group II contained AtGolS4-7, ItlbGolS1-5, and ItfGolS1-4. Group III is formed only by Arabidopsis sequences (AtGolS8-5-6) (Figure 3).

Based on the phylogenetic background in *M. esculenta* Crantz, they were revealed that the GolS are clustered into four groups (Li *et al.*, 2018). This classification is a little inconsistent with that previously reported in other plant species: *Solanum lycopersicum* (Filiz *et al.*, 2015), *N. tabacum* (Fan *et al.*, 2017), and *S. indicum* L. (You *et al.*, 2018). Although sequences have been grouped into slightly different clades (ItlbGolSs and ItfGolSs), these sequences may have the same predicted function in *A. thaliana*. Previous studies *in A. thaliana* indicated that the allelic variants *AtGolS1*, *AtGolS2*, and *AtGolS3* are related to the response to abiotic stresses (Taji *et al.*, 2002; Panikulangara *et al.*, 2004).

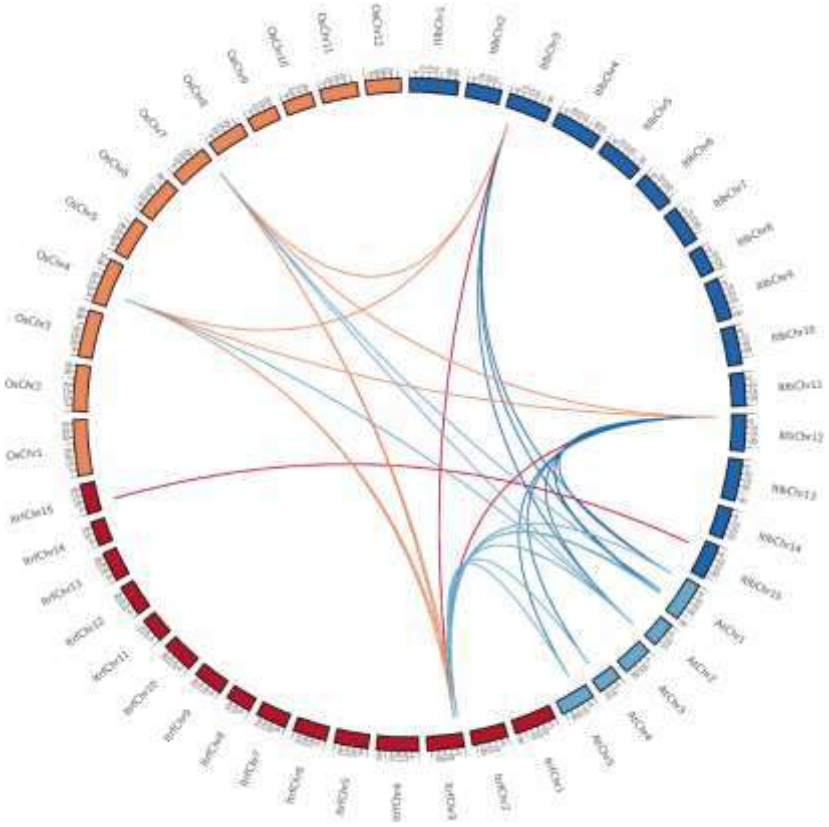
Figure 3. Phylogenetic tree of GolSs proteins sequence from *I. triloba*, *I. trifida*, and *A. thaliana*. A Neighbor-Joining phylogenetic tree was constructed using MEGA7.0 program with 1.000 bootstrap replicates. Groups I, II, and III are grouped as indicated by square brackets, respectively.



Synteny analysis and evolutionary history among *I. trifida* with other species (*A. thaliana*, *O. sativa*, and *I. triloba*) were performed to identify the orthologous pairs of *GolS* genes (Figure 4; Table Supplementary 2). We identified 38 pairs of orthologous genes. Four pairs of *GolS* genes were found with synteny relationships between *I. triloba* and *O. sativa* and I. *trifida* and *O. sativa*. Among *I. trifida* and *A. thaliana* we found 12 pairs and 11 pairs between *I. trifida* and *A. thaliana*. Many genes in *I. trifida* and *I. triloba* (one *I. trifida* or *I. triloba* gene associated with various *A. thaliana* genes) had multiple orthologous gene pairs in *A. thaliana* (Table Supplementary 2). This indicates that *GolS* genes might be derived from the same ancestor and arose before the divergence of eudicots and monocots. Indeed, we suggest this gene may have played an essential adaptative role during evolution. Besides, we also found that other *GolS* genes were present in *I. triloba* (*ItlbGolS4* and *ItlbGolS5*) and *I. trifida* (*ItfGolS4)* and were not found in *Arabidopsis* (Figure 4; Table Supplementary 2), which may indicate that these genes were formed after the divergence of eudicot and monocot plants.

We identified four *GolSs* gene pairs, consisting of three *ItfGolS* and *ItlbGolS* genes, respectively, between *I. trifida* and *I. triloba* (Figure 4; Table Supplementary 2). The high sequence identity and synteny between the *GolSs* genes of these two species can be attributed to the recent divergence between *I. trifida* and *I. triloba* genomes. A previous report has proposed that an ancient ancestor of the Ipomoea lineage underwent whole-genome triplication (WGT) approximately 46.1 million years ago (Mya). This event predates the divergence of *I. nil* from the lineage encompassing *I. trifida* and *I. triloba* (~3.6 Mya), as well as the subsequent divergence between *I. trifida* and *I. triloba* (~2.2 Mya) (Wu *et al.*, 2018).

Figure 4. Representation of synteny analysis of *GolS* genes. The lines indicate the relationship between genes in that region (see Supplementary Table 2). The prefix ItlbChrs is referring to *I. triloba* chromosomes; ItfChrs is referring to *I. trifida*; OsChrs: *O. sativa* chromosomes, and AtChrs: *A. thaliana* chromosomes.

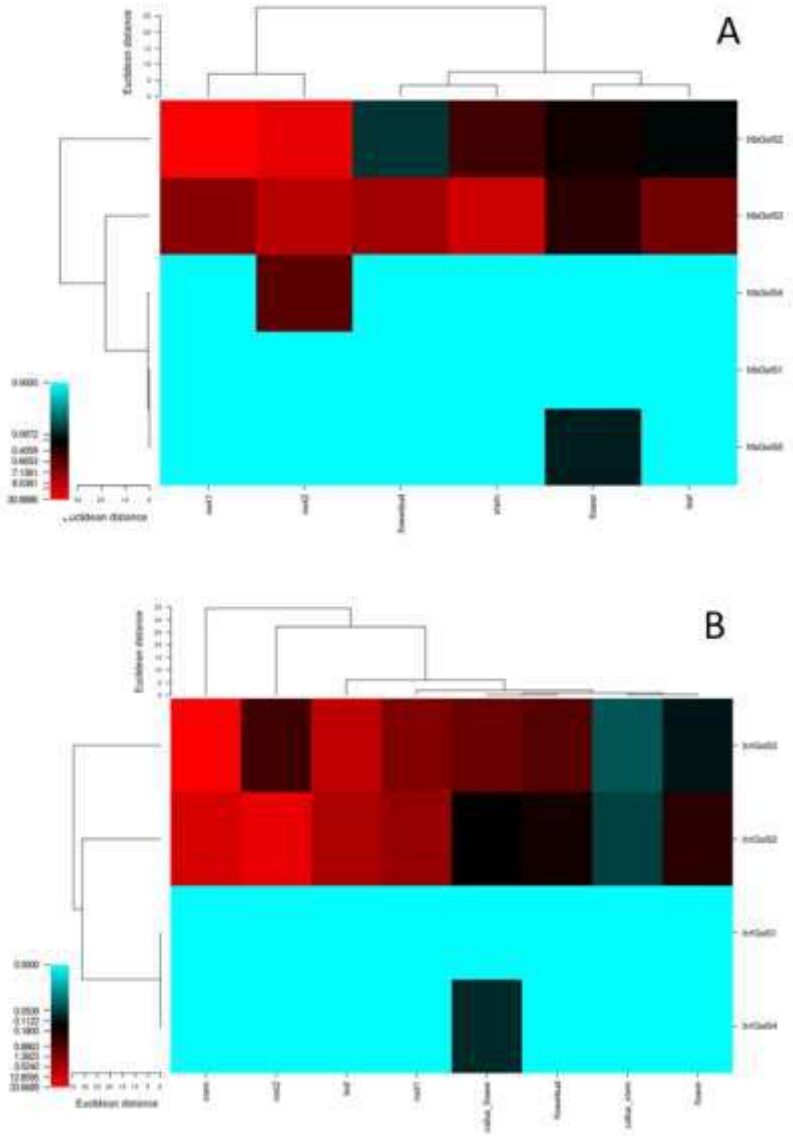


In this study, we also verified and analyzed *in silico* the *GolS* genes' mRNA abundance using publicly available data (https://sweetpotato.plantbiology.msu.edu/). We collected data for a total of six tissues of *I. triloba*: flower, flower bud, leaf, root1, root2, stem, and a total of six tissues of *I. trifida*: callus_flower, callus_stem, flower, flower bud, leaf, root1, root2, and stem, respectively. Both sequences are derived from FPKM (fragments per kilobase of exon per million mapped). In figure 5A shows expression profiles *in silico* of *ItlbGolS* genes. *ItlbGolS2* and *ItlbGolS3* genes are expressed in almost all tissues studied. Interestingly, the *ItlbGolS4* and *ItlbGolS5* genes show slight expression in root2, and flower (Figure 5A). The remainder of the genes exhibited low mRNA levels (Figure 5A).

The heatmap in *I. trifida* also shows the expression of *ItfGolS2* genes, and *ItfGolS3* genes in callus_flower, callus_stem, flower, flower bud, leaf, root1, root2, and stem, respectively. The *ItfGolS4* gene was also expressed in callus_flower (Figure 5B). The remaining genes also showed low mRNA levels (Figure 5B). Our data suggest that *GolS* genes in *I. triloba* and *I. trifida* may be expressed differently in the studied tissues. When we observe that specific genes are

expressed in certain tissues, we can suggest that these genes show some functional preservation (Li *et al.*, 2015). In a previous study, for example, they related *GolS* expression in leaves to phloem export and carbon storage, among other characteristics (Nishizawa *et al.*, 2008). According to the study by Fan *et al.* (2017), no expression of *GolS* genes was observed in *N. tabacum* leaves. In contrast, the *StGolS4* gene showed high expression levels in tissues related to transport, suggesting that this gene may play an important role in material transport in *Solanum tuberosum* L. (Jiang *et al.*, 2023).

Figure 5. *In silico e*xpression profiles. **A.** *Ipomoea triloba* **B.** *Ipomoea trifida*. The color bar represents the FPKM (fragments per kilobase of exon per million reads mapped) value obtained in the genome.



Conclusion

Five *GolS* genes were identified in *I. triloba* and four in *I. trifida*. Both species are mentioned as ancestors of the sweet potato. A conserved motifs demonstrated group-specificity in all *ItlbGolS* and *ItfGolS* proteins. Phylogenetic analysis of the ItlbGolS and ItfGolS proteins revealed four groups. The synteny *GolS* genes are highly conserved among the studied species. *In silico* analysis of gene expression suggests a certain functional divergence of *GolS* genes in sweet potato. These findings will provide new insights into the *GolS* genes in the sweet potato and its wild ancestors (*I. triloba* and *I. trifida*) and contribute to the molecular breeding of sweet potatoes, mainly in the producing regions of this species.

References

ALI, E.; KAKAR, K.U.; SHAH, J.M.; SAAND, M.A.; HUSSAIN, N.; XUE, D.; JIANG, L. Bioinformatics study of Tocopherol biosynthesis pathway genes in *Brassica rapa*. **International Journal of Current Microbiology and Applied Sciences**, v.4, n.3, p.721-732, 2015.

ALI, H.; LIU, Y.; AZAM, S.M.; PRIYADARSHANI, S.V.G.N.; LI, W.; HUANG, X.; HU, B.; XIONG, J.; ALI, U.; QIN, Y. Genomic survey, characterization, and expression profile analysis of the SBP genes in pineapple (*Ananas comosus* L.). **International Journal of Genomics**, v.2017, 2017. https://doi.org/10.1155/2017/1032846

ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.; W.; LIPMAN, D.J. Basic local alignment search tool. **Journal of molecular biology**, v.215, n.3, p.403-410, 1990. https://doi.org/10.1016/S0022-2836(05)80360-2

AUSTIN, D.F. The taxonomy, evolution and genetic diversity of sweet potatoes and related wild species. *In*: **Exploration, maintenance, and utilization of sweetpotato genetic resources.** p.27-60, 1988.

CAO, J.; SHI, F. Dynamics of arginase gene evolution in metazoans. **Journal of Biomolecular Structure and Dynamics**, v.30, n.4, p.407-418, 2012. https://doi.org/10.1080/07391102.2012.682207

CHOU, K.C.; SHEN, H.B. Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization. **PloS one**, v.5, n.6, p.e11335, 2010. https://doi.org/10.1371/journal.pone.0011335

CONANT, G.C.; WOLFE, K.H. Probabilistic cross-species inference of orthologous genomic regions created by whole-genome duplication in yeast. **Genetics**, v.179, n.3, p.1681-1692, 2008. https://doi.org/10.1534/genetics.107.074450

DOLCIMASCULO, F.; RIBAS, A. F.; VIEIRA, L.G.E.; DOS SANTOS, T.B. A genome-wide analysis of the galactinol synthase gene family in banana (*Musa acuminata*). **Colloquium Agrariae,** p.01-11, 2018.

FALAVIGNA, V.D.S.; PORTO, D. D.; MIOTTO, Y.E.; SANTOS, H. P. D.; OLIVEIRA, P.R.D.D.; MARGIS-PINHEIRO, M.; PASQUALI, G.; REVERS, L.F. Evolutionary diversification of galactinol synthases in Rosaceae: adaptive roles of galactinol and raffinose during apple bud

dormancy. **Journal of Experimental Botany**, v.69, n.5, p.1247-1259, 2018. https://doi.org/10.1093/jxb/erx451

FAN, Y.; YU, M.; LIU, M.; ZHANG, R.; SUN, W.; QIAN, M.; DUAN, H.; CHANG, W.; MA, J.; QU, C., ZHANG, K.; LEI, B.; LU, K. Genome-wide identification, evolutionary and expression analyses of the *GALACTINOL SYNTHASE* gene family in rapeseed and tobacco. **International Journal of Molecular Sciences**, v.18, n.12, p.2768, 2017. https://doi.org/10.3390/ijms18122768

FILIZ, E.; OZYIGIT, I.I.; VATANSEVER, R. Genome-wide identification of galactinol synthase (*GolS*) genes in *Solanum lycopersicum* and *Brachypodium distachyon*. **Computational Biology and Chemistry**, v.58, p.149-157, 2015. https://doi.org/10.1016/j.compbiolchem.2015.07.006

GÓIS, E.H.B.; MENEGAZZO, R.F.; DOS SANTOS, T.B.; DE SOUZA, S.G.H. Identification, evolutionary and expression analysis of the *galactinol synthase* (*GolS*) genes in *Panicum virgatum* L. and *Panicum hallii*: An in silico approach. **Plant Gene**, v.24, p.100262, 2020. https://doi.org/10.1016/j.plgene.2020.100262

HU, B.; JIN, J.; GUO, A.Y.; ZHANG, H.; LUO, J.; GAO, G. GSDS 2.0: an upgraded gene feature visualization server. **Bioinformatics**, v.31, n.8, p.1296-1297, 2015. https://doi.org/10.1093/bioinformatics/btu817

HUANG, Z.; ZHONG, X.J.; HE, J.; JIN, S.H.; GUO, H.D.; YU, X.F.; ZHOU, Y.J.; LI, X.; MA, M.D.; CHEN, Q.B.; LONG, H. Genome-wide identification, characterization, and stress-responsive expression profiling of genes encoding LEA (late embryogenesis abundant) proteins in Moso bamboo (*Phyllostachys edulis*). **PloS one**, v.11, n.11, p.e0165953, 2016. https://doi.org/10.1371/journal.pone.0165953

JARRET, R. L.; GAWEL, N.; WHITTEMORE, A. Phylogenetic relationships of the sweetpotato [*Ipomoea batatas* (L.) Lam.]. **Journal of the American Society for Horticultural Science**, v.117, n.4, p.633-637, 1992. https://doi.org/10.21273/JASHS.117.4.633

JIANGTAO, C.; YINGZHEN, K.; QIAN, W.; YUHE, S.; DAPING, G.; JING, L.V.; GUANSHAN, L. Mapgene2chrom, a tool to draw gene physical map based on perl and svg languages. **Hereditas**, v.37, n.1, p.91-97, 2015. https://doi.org/10.16288/j.yczz.2015.01.013

JIANG, Q.; CHEN, A.; LV, Z.; DONG, Z.; WANG, L.; MENG, X.; FENG, Y.; WAN, Y.; SU, C.; CUI, Y.; XU, W.; HOU, H.; ZHU, X. Systematic analysis of galactinol synthase and raffinose synthase gene families in potato and their expression patterns in development and abiotic stress responses. **Genes**, v.14, n.7, p.1344, 2023. https://doi.org/10.3390/genes14071344

KIM, J.H.; HOSSAIN, A.M.; KIM, N.H.; LEE, D.H.; LEE, H.J. Identification and functional characterization of the GALACTINOL SYNTHASE (MoGolS1) gene in Melissa officinalis plants. **Journal of Applied Biological Chemistry,** v. 54, n.4, p. 244-251, 2011. http://dx.doi.org/10.3839/jabc.2011.040

KONING, R.; WILS, G.E.; KIEKENS, R.; DE VUYST, L.; ANGENON, G. Impact of drought and salt stress on galactinol and raffinose family oligosaccharides in common bean (*Phaseolus vulgaris*). **AoB Plants**, v.15, n.4, p.plad038, 2023. https://doi.org/10.1093/aobpla/plad038

KUMAR, S.; STECHER, G.; TAMURA, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. **Molecular biology and evolution**, v.33, n.7, p.1870-1874, 2016. https://doi.org/10.1093/molbev/msw054

LAHUTA, L.B.; PLUSKOTA, W.E.; STELMASZEWSKA, J.; SZABLIŃSKA, J. Dehydration induces expression of *GALACTINOL SYNTHASE* and *RAFFINOSE SYNTHASE* in seedlings of pea (*Pisum sativum* L.). **Journal of plant physiology**, v.171, n.14, p.1306-1314, 2014. https://doi.org/10.1016/j.jplph.2014.04.012

LI, R.; YUAN, S.; HE, Y.; FAN, J.; ZHOU, Y.; QIU, T.; LIN, X.; YAO, Y.; LIU, J.; FU, S.; HU, X.; GUO, J.; Genome-wide identification and expression profiling analysis of the galactinol synthase gene family in cassava (*Manihot esculenta* Crantz). **Agronomy**, v.8, n.11, p.250, 2018. https://doi.org/10.3390/agronomy8110250

MARTINS, C.P.; FERNANDES, D.; GUIMARÃES, V.M.; DU, D.; SILVA, D.C.; ALMEIDA, A.A.F.; GMITTER JR.,F.G.; OTONI, W.C.; COSTA, M.G. Comprehensive analysis of the *GALACTINOL SYNTHASE* (*GolS*) gene family in citrus and the function of *CsGolS6* in stress tolerance. **Plos one**, v.17, n.9, p.e0274791, 2022. https://doi.org/10.1371/journal.pone.0274791

MEYER, T.; VIGOUROUX, A.; AUMONT-NICAISE, M.; COMTE, G.; VIAL, L.; LAVIRE, C.; MORÉRA, S. The plant defense signal galactinol is specifically used as a nutrient by the bacterial pathogen *Agrobacterium fabrum*. **Journal of Biological Chemistry**, v. 293, n. 21, p. 7930-7941, 2018. https://doi.org/10.1074/jbc.RA118.001856

MUKHERJEE, S.; SENGUPTA, S.; MUKHERJEE, A.; BASAK, P.; MAJUMDER, A. L. Abiotic stress regulates expression of galactinol synthase genes post-transcriptionally through intron retention in rice. **Planta**, v.249, p.891-912, 2019. https://doi.org/10.1007/s00425-018-3046-z

NISHIZAWA, A.; YABUTA, Y.; SHIGEOKA, S. Galactinol and raffinose constitute a novel function to protect plants from oxidative damage. **Plant physiology**, v.147, n.3, p.1251-1263, 2008. https://doi.org/10.1104/pp.108.122465

PANIKULANGARA, T.J.; EGGERS-SCHUMACHER, G.; WUNDERLICH, M.; STRANSKY, H.; SCHÖFFL, F. *Galactinol synthase1*. A novel heat shock factor target gene responsible for heat-induced synthesis of raffinose family oligosaccharides in Arabidopsis. **Plant physiology**, v.136, n.2, p.3148-3158, 2004. https://doi.org/10.1104/pp.104.042606

SALVI, P.; KAMBLE, N.U.; MAJEE, M. Ectopic over-expression of ABA-responsive Chickpea galactinol synthase (*CaGolS*) gene results in improved tolerance to dehydration stress by modulating ROS scavenging. **Environmental and Experimental Botany**, v.171, p.103957, 2020. https://doi.org/10.1016/j.envexpbot.2019.103957

SANTOS, T.B.; BUDZINSKI, I.G.; MARUR, C.J.; PETKOWICZ, C.L.; PEREIRA, L.F.; VIEIRA, L.G. Expression of three *galactinol synthase* isoforms in *Coffea arabica* L. and accumulation of raffinose and stachyose in response to abiotic stresses. **Plant Physiology and Biochemistry**, v.49, n.4, p.441-448, 2011. https://doi.org/10.1016/j.plaphy.2011.01.023

SANTOS, T.B.D.; LIMA, R.B.D.; NAGASHIMA, G.T.; PETKOWICZ, C.L.D.O.; CARPENTIERI-PÍPOLO, V.; PEREIRA, L.F.P.; DOMINGUES, D.S.; VIEIRA, L.G.E. *Galactinol synthase* transcriptional profile in two genotypes of *Coffea canephora* with contrasting tolerance to drought. **Genetics and Molecular Biology**, v.38, p.182-190, 2015. https://doi.org/10.1590/S1415-475738220140171

SANTOS, T.B.; VIEIRA, L.G.E. Involvement of the *galactinol synthase* gene in abiotic and biotic stress responses: A review on current knowledge. **Plant Gene**, v.24, p.100258, 2020. https://doi.org/10.1016/j.plgene.2020.100258

SENGUPTA, S.; MUKHERJEE, S.; PARWEEN, S.; MAJUMDER, A.L. Galactinol synthase across evolutionary diverse taxa: functional preference for higher plants? **FEBS letters**, v.586, n.10, p.1488-1496, 2012. https://doi.org/10.1016/j.febslet.2012.04.003

SPRENGER, N.; KELLER, F. Allocation of raffinose family oligosaccharides to transport and storage pools in Ajuga reptans: the roles of two distinct galactinol synthases. **The Plant Journal**, v.21, n.3, p.249-258, 2000. https://doi.org/10.1046/j.1365-313x.2000.00671.x

TAJI, T.; SEKI, M.; SATOU, M.; SAKURAI, T.; KOBAYASHI, M.; ISHIYAMA, K.; NARUSAKA, Y.; NARUSAKA, M.; ZHU, J.K. SHINOZAKI, K. Comparative genomics in salt tolerance between Arabidopsis and Arabidopsis-related halophyte salt cress using Arabidopsis microarray. **Plant Physiology**, v.135, n.3, p.1697-1709, 2004. https://doi.org/10.1104/pp.104.039909

THOMPSON, J.D.; HIGGINS, D.G.; GIBSON, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic acids research**, v.22, n.22, p.4673-4680, 1994. https://doi.org/10.1093/nar/22.22.4673

WU, S.; LAU, K.H.; CAO, Q.; HAMILTON, J.P.; SUN, H.; ZHOU, C.; ESERMAN, L.; GEMENET, D.C.; OLUKOLU, B.A.; WANG, H.Y.; CRISOVAN, E.; GODDEN, G.T.; JIAO, J.; WANG, X.; MERCY, K.; MANRIQUE-CARPINTERO, N.; VAILLANCOURT, B.; WIEGERT-RININGER, K.; YANG, X.S.; BAO, K.; SCHAFF, J.; KREUZE, J.; GRUNEBERG, W.; KHAN, A.; GHISLAIN, M.; MA, D.F.; JIANG, J.M.; MWANGA, R.; LEEBENS-MACK, J.; COIN, L.; YENCHO, C.; ROBIN, B.R.; FEI, Z. Genome sequences of two diploid wild relatives of cultivated sweetpotato reveal targets for genetic improvement. **Nature Communications**, v.9, n.1, p.4580, 2018. https://doi.org/10.1038/s41467-018-06983-8

YOU, J.; WANG, Y.; ZHANG, Y.; DOSSA, K.; LI, D.; RONG, Z.; WANG, L.; ZHANG, X. Genome-wide identification and expression analyses of genes involved in raffinose accumulation in sesame. **Scientific Reports**, v.8, n.1, p.4331, 2018. https://doi.org/10.1038/s41598-018-22585-2

ZHANG, Z.; LI, J.; ZHAO, X.Q.; WANG, J.; WONG, G.K.S.; YU, J. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. **Genomics, proteomics & bioinformatics**, v.*4*, n.4, p. 259-263, 2006. https://doi.org/10.1016/S1672-0229(07)60007-2

ZHOU, L.; BAWA, R.; HOLLIDAY, J. A. Exome resequencing reveals signatures of demographic and adaptive processes across the genome and range of black cottonwood (*Populus trichocarpa*). **Molecular Ecology**, v.23, n.10, p.2486-2499, 2014. https://doi.org/10.1111/mec.12752

ZHOU, Y.; LIU, Y.; WANG, S.; SHI, C.; ZHANG, R.; RAO, J.; Wang, X.; GU, X.; WANG, Y.; LI, D.; WEI, C. Molecular cloning and characterization of galactinol synthases in *Camellia sinensis* with different responses to biotic and abiotic stressors. **Journal of Agricultural and Food Chemistry**, v.65, n.13, p.2751-2759, 2017. https://doi.org/10.1021/acs.jafc.7b00377