Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Phenotypic and RNA-seq profiles identified key pathways and genes involved in gleditsioside biosynthesis in Gleditsia sinensis Lam

Jing Wang
Beijing Forestry University
Yuzhang Yang
Beijing Forestry University
Dandan Xiao
Beijing Forestry University
Fuli Chang
Taihang Mountains National Nature Reserve, Jiaozuo Administration
Hui Chen
Beijing Forestry University
Chun Wang
Beijing Forestry University
Xiaoqian Yang
Beijing Forestry University
Tiantian Fu
Beijing Forestry University
Yanping Liu
Henan Academy of Forestry
Dingchen Fan
Henan Academy of Forestry
Yanwei Wang
ywwang@bjfu.edu.cn

Beijing Forestry University

Research Article

Keywords: Gleditsia sinensis, gleditsioside biosynthesis, transcriptome analysis, pod development

Posted Date: June 16th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3021898/v1

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

Additional Declarations: No competing interests reported.

Abstract

Background

Gleditsia sinensis Lam (*G. sinensis*) is one of the important medicinal herb plant species, and its major bioactive compounds are gleditsioside in pods. The essential economic and medicinal value of gleditsioside has been increasingly recognized. Nonetheless, the molecular mechanisms underlying the changes in the content of gleditsioside during the development of *G. sinensis* are largely unknown.

Results

In this resrarch, there is a positive relationship between the accumulation of gleditsioside content and the variation of pod length, and we performed a transcriptome-wide analysis involving gleditsioside biosynthesis in the pods of *G. sinensis* using RNA-seq. 9000 differentially expressed unigenes (DEGs) were found among six development stages of *G. sinensis* pods. 703 and 162 DEGs participated in terpenoid backbone and triterpenoid biosynthesis pathways, respectively. 99 unigenes were identified, which can encode 17 enzymes, including key enzyme families, such as ENIN, cytochrome P450 (CYP93E1), and UDP-glucosyltransferase in the gleditsioside biosynthesis pathway. Moreover, DEGs encoding crucial enzymes (HMGCR and AGBH) can determine the gleditsioside synthesis during the development of *G. sinensis* pods. According to the generation of different hormones, 10 pathways have been expanded outward in the gleditsioside synthesis pathway, forming a relationship network together. They shared the same precursor substances (IPP and DMAPP), and the 11 pathways should be inhibitory with the gleditsioside synthesis pathway. In addition, WGCNA analysis was further conducted combined with the phenotype of pods and gleditsioside content. As a result, it was found that Unigene32740 (HMGCS) and CL4789.Contig4 (COL) were involved in the gleditsioside biosynthesis and the pod development, respectively.

Conclusions

Overall, this study shows an important gene resource for the future functional researches and provides new insights into the fundamental mechanisms of the gleditsioside biosynthesic process in *G. sinensis*.

Background

Gleditsia sinensis Lam. (*Leguminosae*) (*G. sinensis*) is a crucial and native tree species in China, distributed around the world [1]. It can be utilized as an economic and medical tree with long circadian. The pods are the main medicinal materials of *G. sinensis* [2], with high medicinal values, including antimicrobial [3], anticancer [4], anti-inflammatory [5], and anticoagulant activities [6]. The major bioactive components are gleditsiosides, belonging to triterpene saponins. Therefore, it is of great economic and medical value to study the changes of gleditsioside content and the regulation mechanisms of

gleditsiosides pathway during the development of *G. sinensis*, which is helpful to figure out key genes regulating gleditsioside synthesis, verify their functions and breed high-yield varieties of saponin by molecular design.

Triterpenoid saponin is a triterpene compound containing one or more glycosyl groups connected to triterpenoid sapogenin. According to its distinct chemical skeleton structures, it includes oleanane-type and dammarane-type, etc [7]. The precursors of these active compounds are terpenoids, which are the largest class of secondary metabolites produced by plants, including monoterpenes, sequiterpenes, diterpenes, treipenes, and tetraterpenes, which play significant roles in plant growth and development [8]. Based on previous studies, the triterpene saponins biosynthesis process consists of three stages: terpene precursor biosynthesis, triterpene backbone biosynthesis and various triterpene saponins generation. The first stage is to generate two types of terpene precursors, isopentenyl pyrophosphate (IPP) and its isomer, dimethylallyl diphosphate (DMAPP) [9], through two isoprenoid biosynthetic pathways, mevalonate (MVA) pathway and 2-C-methyl-d-erythritol 4-phosphate/1-deoxy-d-xylulose 5-phosphate (MEP/DOXP) pathway [10, 11]. Interestingly, most plants including *G. sinensis* commonly have both cytosolic MVA pathway and chloroplastic MEP/DOXP pathway for IPP and DMAPP, and algae plants mainly contain MEP/DOXP pathway [12, 13]. Moreover, five-carbon units form via the cytosolic MVA pathway served as precursors mainly for the biosynthesis of sesquiterpenoids, phytosterols, brassinosteroids, triterpenoids, and terpenoid biosynthesis in mitochondria (e.g., ubiquinones). IPP and DMAPP derived from the plastidial MEP/DOXP pathway are preferably used in the formation of hemi-, mono-, di-, sester-, tetraterpenoids (carotenoids and their breakdown products), phytohormones such as cytokinins and gibberellins, chlorophyll, tocopherols, and plastoquinones [8].

Over the past few years, next-generation sequencing has been extensively used to explore the novel genes underlying saponin biosynthesis pathways which has led to a better understanding of pathways and regulatory mechanisms involved in the biosynthesis of specialized terpenoids. Studies have revealed the presence of noncanonical routes and gene clusters involved in the formation of precursors and downstream terpenoids in a few plant species, which added to the complexities of terpenoids biosynthesis. Some include the excavation of genes encoding enzymes (P450 and UGTs) that catalyze distinct steps related to the biosynthetic pathway of saponin in *Panax ginseng* and *Bacopa monnieri* [14, 15]. Comparative transcriptome analysis including root and leaf tissues to excavate transcripts related to saponin biosynthesis has been reported for a few plants, such as *Hedera helix*, *Panax notoginseng* and *Asparagus racemosus* [16, 17, 18]. In addition, the maximum triterpenoid saponin content was identified in pod of *G. sinensis* [2]. Elucidation of regulatory factors and gene clusters involved in the biosynthesis of specialized terpenoids in different plant species has provided insights into enhancing the biosynthesis of specialized terpenoids.

Further understanding of the regulation of terpenoid pathways, and identification of novel pathway genes or gene clusters for terpenoid pathways in *G. sinensis* could pave the way to realize enhanced production of terpenoids through planta or heterologous engineering. In our previous research, we have discovered several small RNAs that regulate gleditsioside synthesis [19]. However, the transcriptional level of the

gleditsioside biosynthesis pathway during different developmental stages of pods has not been reported so far. Thus, we performed transcriptome analysis based on six stages of pods in *G. sinensis* to deeply explore the regulatory mechanisms of the gleditsioside synthesis pathway. Cooperating with WGCNA analysis, we found key genes and enzymes that participated in the gleditsioside synthesis and its parallel pathway. Additionally, some genes and transcription factors were revealed to be related to pod development. This study should provide essential candidate genes for the molecular improvement of gleditsioside in pods of *G. sinensis* and help to uncover the molecular mechanisms of gleditsioside synthesis and pod development, which could accelerate breeding cultivated *G. sinensis* varieties with high-yield gleditsioside.

Results

Phenotype and gleditsioside content variation during pod development

During the six developmental periods, we conducted morphological observation on the development of the pods in *G. sinensis*, the developmental changes of pods were mainly embodied in length and color (Fig. 1A, B). The length of pods changed from 20.98 cm to 31.22 cm significantly from May to July, while the change in length was not significant in the later period. In addition, a noticeable change in color from green to brown suggested that *G. sinensis* pods started to develop in May and matured in October (Fig. 1A). From the variation data, the pod length was mainly the phenotypic change (Fig. 1C).

In the six development periods, the content of gleditsioside was measured by spectrophotometric colorimetry with echinocystic acid (EA) [20]. We found that only maximum absorption wavelength of methanol extraction from the *G. sinensis* pods similar to EA in methanol solution was much close to 538 nm (Figure S1), suggesting that we can regard EA as the standard to quantify gleditsioside content. Contents of gleditsioside gradually increased from 9.3 to 10.6 in the early stage, then dropped in October and reached a maximum in October (Fig. 1C). It shows that gleditsioside was accumulated during the development of pods. Interestingly, we found that there was a closely positive relationship between the accumulation of gleditsioside content and the variation of pod length (Figure S2).

Annotation and classification of unigenes of G. sinensis with pods of six development stages

12 transcriptome samples produced 163319 unigenes, with the longest size above 3000 nt (Fig. 2A). By comparing and annotating the clean reads, 133646 unigenes (81.83%) were obtained, of which GC over 43.84% per sample, and a percentage of Q20 bases above 95.83%, suggesting the high quality of RNA sequencing (Table S1). All unigenes were functionally annotated in NR, NT, Swiss Prot, KEGG, COG, and GO databases, the most and least unigenes were obtained in the NR and COG databases, which were 125434 (76.55%) and 54258 (33.22%), respectively. Unigenes e-value less than 1e-60 was 57.74%, and

unigenes similarity distribution most reached 79.8% (Figure S3), which suggested that these unigenes had a higher degree of compatibility with the NR database. In addition, 91645 (56.11%) and 78334 (47.96%) unigenes were annotated in the GO and KEGG databases, respectively (Fig. 2B). Intriguingly, 31739, 15931, and 11559 unigenes separately were annotated with high similarity to genes of the Fabaceae family members (*Glycine max*, *Glycine soja*, and *Cicer arietinum*) (Fig. 2C). This result confirmed *G. sinensis* belongs to legume family from a molecular perspective. Moreover, we searched unigene sequences against the COG database. A total of 54,258 unigenes were mapped to the COG database, among gene functions mainly annotated in “Secondary metabolites biosynthesis, transport, and catabolism” (Q = 2352, 2.23%), indicating a connection with the secondary metabolite synthesis (Fig. 2D).

Comparison of the differentially expressed gene from different development stages

By differential comparison with SOAP software, we detected 9000 differentially expressed unigene (DEGs) among the six development stages of *G. sinensis* pods. The number of up-regulated DEGs was lower than that of the down-regulated DEGs in all comparison groups, except for GS5-vs-GS6 and GS4-vs-GS6, which supported that many unigenes were differentially and intermittently expressed during the pod development. In the six comparison groups (GS1-vs-GS2, GS2-vs-GS3, GS3-vs-GS4, GS4-vs-GS5, and GS5-vs-GS6), the number of DEGs showed a gradually increasing trend (Fig. 3). The largest number of DEGs were identified in the GS1-vs-GS6 comparison, with 9190 up-regulated and 5543 down-regulated genes. In addition, compared to last month, numerous pseudogenes that do not have function in many plants have been obtained through unigenes, and the DEGs were far less than that of GS2 ~ GS6 compared with GS1 (Figure S4).

In most comparison groups of the six development stages, the down-regulated genes were greater than the up-regulated genes, on the contrary, in GS4-vs-GS5 and GS5-vs-GS6, the number of up-regulated DEGs increased the largest, which might be related to the aging of the pods (Figure S5). Moreover, the number of up-regulated DEGs increased firstly and then decreased from GS1 to GS5 periods and increased again in GS6, which was consistent with changing trend of gleditsioside content in different periods. Hence, it was the process from tender to mature of pods from GS1 to GS3, and the gradual accumulation of gleditsioside in the later stage. Since then, parts of unigenes were not expressed or have no differential change in the GS4 to avoid the excessive accumulation of secondary metabolites. Ultimately, in the development of pods, gleditsioside gradually accumulated, reaching maximum levels when the pods were mature and the gene encoding the gleditsioside might increase then inhibiting gleditsioside synthesis.

GO annotation and KEGG enrichment of differentially expressed genes in G. sinensis

Comparing two adjacent periods (GS1-vs-GS2, GS2-vs-GS3, GS3-vs-GS4, GS4-vs-GS5, and GS5-vs-GS6), the number of differential genes was far less than the DEGs of GS2 ~ GS6 compared with GS1 (Fig. 3, Figure S4). Consequently, to excavate DEGs related to the gleditsioside synthesis pathway, DEGs obtained from the comparisons between the GS2 ~ GS6, and GS1 stages were used for functional GO annotation and KEGG enrichment.

23, 20, and 15 classes of unigenes are involved in biological processes, cellular components, and molecular functions, respectively (Table S2). In the biological processes, there were three categories of genes, among which “metabolic process” (n = 62,940, 10.13%) and “growth” (n = 1,850, 0.30%) were dominant. In molecular function, 418 unigenes were annotated with “antioxidant activity” (n = 418, 0.07%). Interestingly, we found that “nucleic acid binding transcription factor activity” (n = 2,228, 0.36%), and “transcription factor activity, protein binding” (n = 585, 0.1%) may be related to gleditsioside production or pod development (Fig. 4A, Table S3). Most of them were terms with enzymatic catalytic activity (oxidoreductase, transferase, etc.), binding function and transport function. UGTs (GO:0008194), an important enzyme in the gleditsioside synthesis pathway, and ENIN (enzyme inhibitor activity, GO:0004857), a critical factor inhibiting the derivative branch of the gleditsioside synthesis pathway, were significantly enriched. Moreover, Hydrolase activity, acting on glycosyl bonds (AGBH, GO:0016798) was highly significantly enriched, which may be the main factor for the easy degradation of gleditsioside (Table 1).

78,334 genes were annotated to 53 KEGG pathways and significantly enriched ($q \leq 0.05$). Six differently comparable groups all were enriched for metabolic pathways in the top 20 pathways (Fig. 4B-F). A total of 162 unigenes and 703 unigenes were significantly enriched in sesquiterpenoid and triterpenoid biosynthesis (STTB) and terpenoid backbone biosynthesis (TBB), respectively. These pathways are related to triterpene saponin unit synthesis, suggesting that *G. sinensis* pods show a marked response to gleditsioside synthesis. Likewise, ten branch pathways such as carotenoid biosynthesis (CTB), zeatin biosynthesis (ZTB), and ubiquinone and another terpenoid quinone biosynthesis (UTB) associated were also significantly enriched with terpenoid synthesis pathways. These pathways were involved in glesitsioside synthesis pathways, and diterpenoid biosynthesis (DTB), CTB, and ZTB were highly enriched per stage. TBB was highly enriched in GS2/GS1 stage, which was significantly higher than the other stages. Additionally, seven secondary metabolic pathways were significantly enriched, including anthocyanin biosynthesis; flavone and flavonol biosynthesis; flavonoid biosynthesis; etc. This proved *G. sinensis* pods contained alkaloids, phenolic components, organic acids, and other chemicals. Ultimately, a monoterpene degradation pathway (Limonene and pinene degradation) was significantly enriched in each stage. Thus, the main secondary metabolite of pods was triterpene saponins, which might be due to the gradual degradation or utilization of other secondary metabolites during the development of pods (Table 1).

Table 1
Differentially expressed unigene numbers of KEGG enrichment pathways in *G. sinensis*

Pathway	Gene Number				
	GS2/GS1	GS3/GS1	GS4/GS1	GS5/GS1	GS6/GS1
Alanine, aspartate and glutamate metabolism	41**	54*	62**	56**	74**
alpha-Linolenic acid metabolism	27*	52*	56**	46*	67**
Amino sugar and nucleotide sugar metabolism	51	125*	126**	108*	150**
Anthocyanin biosynthesis	9**	4	5*	7**	7**
Ascorbate and aldarate metabolism	63**	91**	93**	93**	96**
Biosynthesis of secondary metabolites	742**	1129**	1113**	1054**	1347**
Biosynthesis of unsaturated fatty acids	40	18	21	14	40*
Brassinosteroid biosynthesis	26	45	46	8	12*
Butanoate metabolism	19**	21	20	14	16
Carbon fixation in photosynthetic organisms	55**	87	91*	86*	110**
Carotenoid biosynthesis	60**	85**	72**	83**	99**
Cutin, suberine and wax biosynthesis	36**	49**	49**	45**	66**
Cyanoamino acid metabolism	35*	48	52	55	84**
Cysteine and methionine metabolism	30	75	93**	76*	94**
Diterpenoid biosynthesis	39**	45**	49**	39**	47**
DNA replication	37*	75**	67*	70**	67
Endocytosis	80	216**	228**	212**	244**
Ether lipid metabolism	61	134*	151**	138**	166**
Fatty acid metabolism	9	63	63	45	63*
Flavone and flavonol biosynthesis	82**	76**	64**	63**	89**
Flavonoid biosynthesis	132**	152**	148**	145**	172**
Fructose and mannose metabolism	3	48	41	50	80*
Glucosinolate biosynthesis	6	10	15**	12*	14*

Pathway	Gene Number				
	GS2/GS1	GS3/GS1	GS4/GS1	GS5/GS1	GS6/GS1
Glycosylphosphatidylinositol -anchor biosynthesis	6	18**	20**	19**	17*
Glycosphingolipid biosynthesis - globo series	9*	7	6	4	5
Glyoxylate and dicarboxylate metabolism	44**	16	13	57**	80**
Indole alkaloid biosynthesis	9**	14**	11*	10*	16**
Limonene and pinene degradation	54**	74**	73**	60**	61**
Metabolic pathways	1117**	1967**	2018**	1871**	2293**
Monoterpenoid biosynthesis	13**	77	80	24**	27**
N-Glycan biosynthesis	28*	47	47	44	62**
Other glycan degradation	45**	73**	77**	73**	81**
Pentose and glucuronate interconversions	56**	130**	136**	129**	146**
Phagosome	29	7	8	68	108*
Phenylalanine metabolism	61**	90**	90**	82**	88**
Phenylalanine, tyrosine and tryptophan biosynthesis	33**	60**	59**	46**	57**
Phenylpropanoid biosynthesis	153**	189**	187**	191**	235**
Photosynthesis	82**	65**	94**	72**	72**
Photosynthesis - antenna proteins	60**	40**	68**	62**	45**
Plant hormone signal transduction	242**	648**	683**	602**	664**
Plant-pathogen interaction	214	843**	904**	734**	700**
Porphyrin and chlorophyll metabolism	49**	50	68**	63**	65*
Propanoate metabolism	28*	8	9	43	45
Pyruvate metabolism	42	101**	73	88**	117**
Sesquiterpenoid and triterpenoid biosynthesis	23**	38	35	20	28*
Starch and sucrose metabolism	114**	244**	252**	223**	301**

Pathway	Gene Number				
	GS2/GS1	GS3/GS1	GS4/GS1	GS5/GS1	GS6/GS1
Steroid biosynthesis	90**	100**	94**	82**	106**
Synthesis and degradation of ketone bodies	9*	20	24	4	2
Terpenoid backbone biosynthesis	69**	87	81	86**	77
Tropane, piperidine and pyridine alkaloid biosynthesis	21**	31**	30**	25*	42**
Tryptophan metabolism	22**	32	37*	46**	41**
Ubiquinone and other terpenoid-quinone biosynthesis	26**	44*	41*	39*	42
Zeatin biosynthesis	72**	78**	64*	76**	93**

** represents significant difference (q-value ≤ 0.05); * represents extremely significant difference (q-value ≤ 0.01); GS1. May, GS2. June, GS3. July, GS4. August, GS5. September, GS6. October.

Gleditsioside biosynthesis-related unigenes were differentially expressed in Gleditsia sinensis

The gleditsioside synthesis pathway is mainly composed of the terpenoid skeleton pathway (TBB) and sesquiterpene triterpene synthesis pathway (STTB) in *G sinensis*. To further indicate the comprehensive gene expression characteristics of the whole gleditsioside biosynthesis route in pods, we drew the gene expression profiles based on $\log_2$Ratio values of the transcribed genes from the perspective of all metabolic modules (Fig. 5). A total of 372 unigenes encoding 23 enzymes were found in the six stages, among them, only 99 unigenes encoding 17 enzymes were differently expressed (Table S3). The main pathways of gleditsioside biosynthesis include mevalonate (MVA), methylerythritol phosphate (MEP/DOXP) and triterpene cyclization process (TCP). Both the MVA and MEP/DOXP pathways were essential biosynthetic processes for the formation of the triterpene backbone [2]. The expression profiles of putative genes associated with the MVA and MEP/DOXP pathways participated in the gleditsioside biosynthesis (Fig. 5). The number of DEGs in the MVA was eight more than that in the MEP/DOXP, thus, MVA plays a more significant role in IPP biosynthesis than in the MEP/DOXP.

Among the MVA and MEP/DOXP pathways, most enzymes and genes showed a trend of high expression in the G1 ~ G4, especially HMGCS, HMGCR, MVK and DXPS. In addition, CL5845.Contig1 (HMGCR), CL1651.Contig1 (IDI), CL13160.Contig1 (DXR), CL4067.Contig3 (ISPG), and CL4067.Contig4 (ISPG) were the highly expressed DEGs in all stages of MEP/DOXP. Additionally, through the enrichment of the GO function, CL4067.Contig3 and CL4067.Contig4 not only participated in the isoprene process of gleditsioside synthesis but participated in other processes of pod growth, such as the jasmonic acid biosynthetic process (GO: 0009697; salicylic acid biosynthetic process, GO: 0019760). And their

differential expression pattern might affect gleditsioside synthesis in the MVA and MEP/DOXP pathways (Fig. 5, Table S4).

2,3-oxidosqualene needs to be cyclized by oxidation and oxygen exchange enzymes (OSCs) in a plant, it was the watershed for steroid and triterpenoid biosynthesis and the initial step for oxidosqualene cyclases (OSCs) [21]. During the Triterpene cyclization process (TCP), different OSCs catalyzed 2,3-oxidosqualene to cycloartenol or triterpenoid to generate different types of pentacyclic triterpenoid saponin skeletons [22]. We annotated 79 unigenes encoding four OSCs that were differentially expressed, including LUP1, LUP2, LUP4, and CAMS1. Among them, the first three OSCs could generate pentacyclic triterpenes of oleanolane, ursolidane, and lupane types. In addition, triterpenoids were modified with hydroxylation by cytochrome P450 monooxygenases (P450s) and glycosidation by UDP-glycosyltransferases (UGTs), which could increase the structural diversity of triterpenoids. We annotated 22 beta-amyrin 24-hydroxylase and eight UDP-glucosyltransferase (UGT), respectively, P450s (CYP93E1) were differentially expressed corresponding to the oleanolane type. Highly expressed Unigene6088 (LUP4), Unigene29200 (CYP93E1) and CL11248.Contig4 (CYP93E1) were gradually up-regulated in the early stage, and then down-regulated in the late stage (Fig. 5). Moreover, the CL8823.Contig2 (LUP4) and Unigene37260 (UGT) with high expression indicated two cycles of declined and then increased expression. In total, Unigene6088, Unigene29200, CL11248.Contig4, CL8823.Contig2 and Unigene37260 might affect gleditsioside synthesis in the TCP synthesis pathway. These findings also provide implications for regulating the production of saponin by synthetic biological methods in the future.

Differentially expressed unigenes in the parallel pathways of gleditsioside biosynthesis

We further investigated the main pathways of gleditsioside synthesis and their parallel pathways. KEGG results showed ten synthesis pathways, such as carotenoid biosynthesis (CTB), zeatin biosynthesis (ZTB), ubiquinone and another terpenoid-quinone biosynthesis (UTB), diterpenoid biosynthesis (DTB), and monoterpenoid biosynthesis (MTB), etc, which were significantly enriched (q-value $\leq$ 0.05). In the ten enrichment pathways, DTB, CTB, and ZTB were rich in all periods, while TBB was only rich in the period of GS2/GS1 (Table 1). According to the TBB pathway, the intermediates DMAPP and GPP were catabolized as enzymatic substrates for ZTB, MTB, and IAB by five enzymes IPT, TRIT1, TPS14, TPS10, and TPS-Cin, respectively (Table S5, Table S6). The intermediate product of FPP was also a key precursor for the enzymatic action of DHDDS and GGPS in the synthesis of other secondary metabolites such as CTB, DTB, UTB, BRB, and NGB. FPP was produced by the DHDDS enzyme, which could derive the N-glycan pathway (NGB). MTB and NGB could derive the indole alkaloid pathway (IAB) and glycosyl phosphatidylinositol pathway (GPIB), respectively. Then, FPP was producing intermediate GGPP by GGPS enzyme, and GGPP was directly reacted by three enzymes (PSY, GA1, CAS) to derive the carotenoid pathway (CTB) and diterpene pathway (DTB); Finally, GGPP also enters the ubiquinone and terpenoid quinone pathway (UTB) together with the intermediate product (PHPP) produced by the action of chIP enzyme and the intermediate product (Solanesyl pyrophosphate, SP) produced by the action of SPS

enzyme under the action of four enzymes (menA, COQ2, HPT, HSP). In addition, during the oxidative cyclization of squalene, the 2,3-oxsqualene produced was cyclized by another kind of cyclizing enzyme CAS1 to generate sterol, which will enter the sterol pathway (STB), and then the STB derived the brassinolide pathway (BRB). These ten branch pathways and gleditsioside synthesis pathways, sharing the same precursor substance IPP, DMAPP, GPP, and FPP, might compete with each other (Fig. 6B).

Through the difference analysis of ten parallel pathways, a total of 700 unigenes were counted, among which ZTB (n = 161) and CTB (n = 141) had more than 100 unigenes. Interestingly, compared with GS1 in the last five periods, we found that the amount of gene change in the parallel pathways was similar to that in the gleditsioside synthesis pathway, The down-regulation genes in all periods showed a gradual upward trend, and the up-regulation genes showed a gradual downward trend, among which up-regulation genes decreased to the lowest in the GS5 vs GS1 period. In addition, the amount of changes in up-regulation genes was similar to the amount of changes in gleditsioside content (Fig. 3). This result further verified that the trend of gene changes in gleditsioside biosynthesis pathway is worthy of reference, which also indicates that the increased gene change might better explain the gene change in gleditsioside biosynthesis pathway and parallel pathway. Additionally, Unigene5728 (SPS/IPT), Unigene6524 (SPS/IPT), Unigene10127 (TPT), and CL333.Contig3 (TPS14) were highly expressed in the parallel pathway, and these genes were gradually down-regulated over time (Fig. 6A, C). They might act as a negative regulator in gleditsioside synthesis.

Transcription factors related to gleditsioside synthesis

According to the report, transcription factors (TFs) regulate triterpenoid biosynthesis through generally interacting with structural genes during plant development and response to stress [19, 23]. We annotated 1196 TFs belonging to 92 TF families in the *G. sinensis* transcriptome dataset. They contained the most members such as bHLH (n = 129 genes), MYB (n = 122 genes), and WRKY (n = 120 genes) families, and played vital roles in the regulation of triterpenoid and terpene biosynthesis in many plant species, such as *Salvia miltiorrhiza*, etc [24, 25, 26, 27] (Table S7). By Further strict screening criteria [FPKM $\geq$ 1 and FDR $\leq$ 0.05], a total of 22 TFs were obtained (Table S8, Table S9).

In addition, TFs EIN3, AGBH, ARF, and MADS also have critical regulatory roles in plant fruit growth and development [18, 28, 29]. The molecular function of GO enrichment (GO-MF) showed ENINs was the last important enzyme for gleditsioside biosynthesis. 18 ENIN genes had a mostly downward trend throughout pod formation, whereas the five AGBHs genes with high expression were all up-regulated (Fig. 7A). Interestingly, AGL and SEP1 were two subfamilies of MADS, in which SEP1 promotes fruit ripening in a similar regulatory pathway to ethylene synthesis [29]. AGL During *G. sinensis* pod development, EIN3, ERF, AGL, and two ARFs (CL4767.Contig2, Unigene28867) were up-regulated. Nevertheless, Dof, GRAS, and one ARF (Unigene28867) showed down-regulation expression trend, especially GRAS was significantly down-regulated and the differential ploidy increased gradually with pod development (Fig. 7B). The Dof, GRAS and ARF (Unigene28867) might have detrimental regulatory

effects on pod development, whereas NAC (CL3360.Contig6), EIN3, ERF (CL2043.Contig2), AGL, SEP1 and two ARFs (CL4767.Contig2, Unigene28867) have positive regulatory effects on pod development.

Identification of a WGCNA Module Related to gleditsioside biosynthesis and pod development during G. sinensis

To expore the correlation between phenotype and gene, we conducted a co-expression network analysis (WGCNA) using the increasing variables of the two shapes of *G. sinensis* compared with the last month, respectively (Table S10). 8000 DEGs were divided into 21 different hierarchical clustering modules (Fig. 8A). Among them, the turquoise module has the largest number of genes (n = 2240), and the gray module with the least number of genes (n = 40)(Table S11). MElightcyan module was significantly negatively correlated with the change in gleditsioside content with a correlation coefficient of 0.79 (P = 0.0021). And the MEpink module was significantly positively associated with the change in *G. sinensis* pod length with a correlation coefficient of 0.78 (P = 0.0025) (Fig. 8B). CL1983.Contig15, CL4789.Contig4, CL4870.Contig1, CL616.Contig7, Unigene40853, and Unigene41074 were the gleditsioside-related to hub genes in the MEpink module network (Fig. 8C). Among them, the differential expression of Unigene32740 (HMGCS) was highest in gleditsioside biosynthesis (Figure S6A). On the other hand, CL13675.Contig7, CL8461.Contig2l, Unigene25298, Unigene29764, Unigene32740, and Unigene35613 were the length-related hub genes in the MEpink module network (Fig. 8F). CL4789.Contig4 (COL) has the highest differential expression in pod development (Figure S6B) Overall, Unigene32740 (HMGCS) and CL4789.Contig4 (COL) were separately considered to the key genes related to gleditsioside biosynthesis and pod development.

To further characterize the functional differences and relationships between modules MElightcyan and Mepink, GO enrichment and KEGG pathway enrichment analyses were performed. The lightcyan module genes were prominent in circadian rhythm, alanine, aspartate, and glutamate metabolism, and ubiquinone and other terpenoid quinone biosynthesis (UTB) pathways in the GO and KEGG enrichment analysis (Fig. 8D-E). During the development of pods, the UTB pathway was involved in and impacted gleditsioside production, and it was inhabited by the biological gleditsioside synthesis pathway and other parallel pathways. Thiamin was frequently found in the active form of TPP, which functions as a coenzyme in glycolysis, the TCA cycle, and the PPP route. Thiamine deficiency slowed photosynthesis and respiration, preventing plants from producing photosynthetic pigments [30]. Besides, the GO enrichment analysis of the pink module revealed the biosynthetic pathway of thiamin as well as a large number of GO terms associated with sugar metabolism, especially sucrose metabolism, showing that the formation of *G. sinensis* pods was accompanied by a significant quantity of energy release. The development process was accompanied by a substantial amount of energy discharge. Furthermore, the biosynthetic pathway of thiamin emerged from GO enrichment analysis, and thiamin was found in the active form of TPP, which was capable of participating in metabolism as a coenzyme in glycolysis, the TCA cycle, and the PPP pathway. The glycosylation pattern of gleditsioside was considered to have an important position in their biological activity, it might involve the sequential activity of different enzymes

of the multigene family of UDP glycosyltransferases (UGTs), which catalyzed the transfer of reactive glycosyl residues from uridine sugar nucleotides to a large number of low molecular weight receptor molecules (Fig. 8G). UDP glycosyltransferases were key enzymes in the biosynthesis of plant triterpenoid saponins [31]. KEGG enrichment analysis of the pink module focused on flavonoid biosynthesis, terpene skeleton synthesis, sesquiterpenoids, and other pathways. The biosynthetic pathway of triterpenoid saponin elements in *G. sinensis* pods also consisted precisely of the terpene skeleton biosynthetic pathway and the triterpenoid synthesis pathway (Fig. 8H). These results indicated that there was also a strong correlation between pod development, gleditsioside synthesis, and accumulation of flavonoids, respectively.

Discussion

Key unigenes to code enzymes in gleditsioside biosynthesis

MVA in the cytosol and MEP/DOXP in plastids were the two main processes for saponin biosynthesis, of which IPP was one of the most significant intermediate products [32, 33]. Inhibition of any pathway would affect the synthesis of saponins. When one pathway was blocked, the other would become a supplement to the saponin synthesis pathway [33]. In the gleditsioside synthesis pathway, the HMGCR enzyme was the key rate-limiting enzyme of MVA, and the genes encoded by HMGCR show a downward trend. Studies have proved that HMGCR was the key rate-limiting enzyme of MVA in ginsenoside [34], and the trend was consistent with our research (Fig. 5). A previous study reported that one TSAR1/2 belonging to bHLH can activate *HMGCR* and all precursor genes in MVA to promote saponin biosynthesis in *Medicago truncatula* [35]. In our investigation, the highest expressed unigenes of *CL5845.Contig1* was up-regulated in each period compared with GS1, indicating it might be the key gene to the coding enzyme of HMGCR to affect gleditsioside content in *G. sinensis*. Other eight unigenes coding HMGCR, belonging to a family of *CL1058*, were gradually down-regulated from GS3 to GS6 (Fig. 9, Table S4). The change of DXR expression was similar to HMGCR, in addition, the expression change of DXR (CL13160.Contig1) and ISPG (CL4067.Contig3, CL4067.Contig4) was the opposite trend in MEP/DOXP pathway. DXR enzyme gene was considered to be the most important rate-limiting enzyme in the MEP pathway, and the gene activity decreases significantly with time [36]. Moreover, the key enzyme IDI (CL1651.Contig1) for the mutual transformation of IPP and DMAPP was gradually up-regulated in MVA, and the expression of this enzyme might be the key enzyme for gleditsioside synthesis, in other studies, IDI was an IPP isomerase gene which can promote the synthesis of triterpenes [16].

Gleditsioside had been identified to be oleanane-type triterpenoids [37, 38]. From the annotation of unigenes in KEGG, CYP93E1 belonging to the family cytochrome P450 can be coding the enzyme to synthesize oleanane-type triterpenoid, and the former OSCs were LUP4 and LUP2. However, some LUP2s were coded by the LUP4s (Table S3). Comparing the last two steps in gleditsioside biosynthesis, most of the LUPs, like higher expressed Unigene6088 (LUP4) with the higher expressed CYP93E1 (CL11248.Contig4), were up-regulated in the early stage, and then down-regulated in the late stage (Fig. 9). The results indicate that the inhibition of gleditsioside synthesis in the late stage, particularly in

GS6, may be related to the down-regulation of these genes. Moreover, the UGT (Unigene37260) and LUP4 (CL8823.Contig2) with high expression showed an increasing trend from GS1 to GS5, while the expression changes of unigenes were reduced in GS6, and these changes could also prove that gleditsioside biosynthesis were inhibited in the later stage (GS6) of pod development. The large accumulation of gleditsioside not only leads to toxicity, but shows antibacterial and cytotoxic activities to prevent plants from biotic stresses [39]. The changes of these unigenes expression levels suggested that pods might be protected in the early stage to increase gleditsioside content, and then in the late stage to decrease the rate of gleditsioside biosynthesis from autotoxicity, which might be one reason for the decrease in gleditsioside in October.

Parallel and inhibitors in gleditsioside biosynthesis

Despite their large structural diversity, saponin synthesis pathway and all parallel pathways for synthesis of other terpenoids share terpenoid precursors IPP and DMAPP, which might compete with each other [40]. Therefore, the weakening of the parallel synthesis pathway would promote the gleditsioside synthesis pathway, and the up-regulation of parallel pathway-related genes can inhibit gleditsioside synthesis. In this study, productions of zeatin catalyzed by ZTB and brassinolide catalyzed by BRB were indispensable for gleditsioside synthesis. In other terpenoid synthesis, BRB and ZTB pathways occur in the cytoplasm, and ZTB occurs in plastids and mitochondria [41]. Therefore, the weakening of the parallel synthesis pathway would promote the gleditsioside synthesis pathway, and the up-regulation of parallel pathway-related genes can inhibit gleditsioside synthesis. In this study, productions of zeatin catalyzed by ZTB and brassinolide catalyzed by BRB were indispensable for gleditsioside synthesis. In other terpenoid synthesis, BRB and ZTB pathways occur in the cytoplasm, and ZTB occurs in plastids and mitochondria.

As a precursor in the biosynthesis of C10-monoterpenoids, GPP was synthesized from IPP and DMAPP, which were usually targeted to plastids and mitochondrion. GPP produced in plastids is exported to the cytosol, where it can be used for monoterpene biosynthesis [42]. We found that GPP was acted by different TPS enzymes in MTB to produce multiple monoterpene precursors, and most of the differentially expressed TPSs showed a down-regulation state, such as CL3333. Contig3 (TPS14) (Fig. 9). Moreover, GGPP synthesized by GGPSs was a major branching point for several downstream terpenoid pathways in primary and specialized metabolism. These included the biosynthesis of CTB, DTB, and UTB (all synthesized in plastids). We found that GGPP catalyzed CTB and DTB via the SPS enzyme, as the key catalyst to enter these pathways. There were two kinds of SPS enzymes, one only encodes one enzyme SPS, and the other can encode IPT and SPS two enzymes, among them, Unigene5782 (SPS/IPT) and Unigene6524 (SPS/IPT) showed a trend of high expression and gradual down-regulation (Fig. 9).

Unigenes and transcription factors in pod development

In pod development, TFs interacted to form a regulatory network that together regulates the rate of fruit development. In bananas, ERF could promote fruit ripening in response to ethylene biosynthesis [43], while Dof was used as a transcriptional repressor of fruit ripening [44]. In our investigate, GRAS has a

time-specific expression, which was abundantly expressed during embryo development at the early stage of fruit development, but declines sharply during fruit ripening [45]. In contrast, SCL, a member of the GRAS family, has a similar expression that promotes gibberellin synthesis and thus embryo development after ovary fertilization at flowering, while showing a decreasing trend until fruit ripening [46]. Among the positively regulated TFs, ERF, NAC, and ARF can respond to ethylene and thus promote fruit ripening senescence [47]. In addition, they could promote CTB formation collectively [48], which explained why PSY in the gleditsioside-derived pathway was significantly up-regulated during GS6, especially NAC (CL3360.Contig6) was significantly up-regulated differentially at GS6, which might be the period of senescence in *G. sinensis* pods (Fig. 9). Therefore, at the early stage of *G. sinensis* pod development (before GS1 and GS1), which was embryo development, SCL and GRAS were abundantly expressed to promote ovule development, when the transcriptional repressor Dof acted as a repressor of fruit ripening, however, by the fruit ripening stage (GS2/GS1 ~ GS5/GS1), SCL, GRAS, and Dof were significantly repressed, while the fruit ripening-promoting ERF (CL2043.Contig2), NAC (CL3360.Contig6), and ARF (CL4767.Contig2, Unigene28867) were gradually and significantly up-regulated, increasing gibberellin content and promoting cell division. Based on the synthesis process of gleditsioside, gleditsioside would be significantly accumulated at the end of pod ripening.

Conclusions

Our study analyzed the changes in gleditsioside content and pod length of *G. sinensis* Lam from May to October. 163,319 genes were obtained and annotated, and analyzed by high-throughput sequencing of pods at the six developmental stages. In differential gene expression analysis, 162 and 703 unigenes were annotated to the triterpene and sesquiterpene biosynthesis pathway (STTB) and terpene backbone synthesis pathway (TBB). In addition, CTB, ZTB, NGB, and the other ten parallel pathways were elucidated, as well as their interrelationships. Among these synthesis pathways, we analyzed key DEGs encoding enzymes including ENIN, cytochrome P450 (CYP93E1), and UDP glossyltransfer, and key genes such as CL5845. Contig1 (HMGCR), CL1651. Contig1 (IDI), CL13160. Contig1 (DXR), and also found some TFs were related to development, such as NAC, EIN3, ERF, AGL, SEP1, and two ARFs (CL4767. Contig2 and Unigene28867). Through further WGCNA combing with phenotype analysis, two highly expressed genes Unigene32740 (HMGCS) and CL4789. Contig4 (COL) related to gleditsioside synthesis and pod development were screened. These results revealed the function of key genes in gleditsioside synthesis pathway and provide new insights into understanding the global gene function distribution of this species.

Methods

Plant materials

From May to October, *G. sinensis* pods were collected at Beijing Forestry University (40°0'18"N, 116°20'13' 'E). Five to six pods were collected at identical heights and sites at each month. These pods were divided into two parts, one part was wrapped in tin foil, immediately frozen in liquid nitrogen and stored at -80°C

for total RNA extraction, and the remaining material was dissected out of the seeds, dried, and crushed and the powder was stored in a dark place for the gleditsioside extraction.

Gleditsioside extraction and quantification

Three grams of *G. sinensis* husk powder for the pods from each period with two replicates were dipped in 30 mL absolute methanol for an hour, respectively. Subsequently, they were transferred in an ultrasonic cleaner with 100% (800 w) power to facilitate the accumulation of gleditsioside for an hour. Afterward, each sample was filtrated by applying a cartridge filter with a size of 0.2 μm. The experiment was repeated the three times. And then, 0.5 ml filtrate was employed for the chromogenic reaction with a chromogenic reagent of 5 mL 77% sulfuric acid and 0.5 mL 8% vanillin ethanol. On the condition that mixed evenly, they were heated in a water bath at 60℃. After 30 minutes, all liquid samples were put in an ice bath for 10 minutes. Following closely, each instance was demonstrated with a full wavelength scan by ultraviolet-visible absorption spectrometry (UV-2550) from 400 nm to 600 nm to confirm the maximum absorption wavelength of the gleditsioside after the chromogenic reaction. Consequently, echinocystic acid (Yuanye Biotechnology Company, Shanghai, China, HPLC ≧ 98%) was used as a standard substance to make a standard curve. Accordingly, in combination with the absorbance of each sample at the maximum absorption wavelength, we would acquire the content of gleditsioside each month for the foundation of the following bioinformatics analysis.

$$\text{Gleditsioside } \% = x \times \frac{V1}{V2} \times {}^{N}/_{M}$$

x: absorbancy of methanol solution after color reaction in 538 nm; V1: volume of gleditsioside extract for the first time; V2: volume of the sample for color reaction; N: dilution multiple of gleditsioside extract; M: powder of *G. sinensis* husk for extraction.

Total RNA extraction and RNA-seq

Two individual pods for each month from May to October 2016 were randomly selected for total RNA extraction using Trizol (Invitrogen life technologies, California, USA). Total RNA bands were detected with a 1% gel (180v, 16 min), followed by mass detection and concentration determination of total RNA with NanoDrop 2000 spectrophotometer (Thermo Scientific, Pittsburgh, PA, USA). 12 samples of GS1-GS6 were sequenced to obtain transcriptome sequences in 1 GENE Biotechnology Company (Hangzhou, China). High-quality RNA was adopted to construct cDNA libraries, and the Illumina NovaSeq 6000 platform was used to sequence the resulting 150 bp paired-end reads. High-quality reads were obtained by trimming sequencing adaptors and removing low-quality reads from the raw sequencing data, clean reads were mapped to the assembled transcripts, and gene expression levels were calculated as Fragments Per Kilobase per Million (FPKM). Afterward, sequences that had low-quality reads and contaminants, including arrangements with poly N or 5’ adaptor and without 3’ adaptor insert tag, containing poly A/T, etc, were removed from the raw datum to get the clean reads for the further analysis.

Assembly, annotation and genes expression analysis

After filtration, clean reads of 12 samples were collected by Trinity to obtain all unigenes. Next, the unigenes will be classified into two groups based on gene family clustering. The first is the cluster with the prefix "CL" and the cluster id behind it. There were numerous unigenes in this cluster that were over 70% identical. The other cluster was a singleton with the prefix "unigene." Subsequently, the annotations were obtained by matching all unigenes in the RefSeq non-redundant protein sequence database (Nr; ftp://ftp.ncbi.nih.gov/blast/db/FASTA), nucleotide sequence (NT, https://www.uniprot.org/uniprotkb?query=NT) database, Swiss-Prot database (https://www.uniprot.org/uniprotkb?query=Swiss-Prot), Kyoto encyclopedia of genes and genomes database (KEGG; http://www.genome.jp/kegg/tool/map_pathway1.html) and clusters of orthologous groups (COG, https://www.ncbi.nlm.nih.gov/research/cog-project) database (e-value < 0.00001). Expression levels of all unigenes were calculated using the method of FPKM (Fragments Per kb per Million fragments) with the following formula.

$$FPKM\,(A) = \frac{10^6 C}{NL/10^3}$$

Assigns FPKM (A) to be the expression of gene A, C to be several fragments that were uniquely aligned to gene A, N to be a total number of fragments that were uniquely aligned to all genes, and L to be the number of bases on gene A. The FPKM method could eliminate the influence of distinct gene lengths and sequencing discrepancies on the calculation of gene expression.

Differentially expressed unigenes analysis

The calculated gene expression can be directly used for comparing the difference in gene expression among samples. First, the same unigenes were averaged in the same period. Next, SOAP was used for the differential expression analysis of all unigenes from six periods in *G. sinensis* with default parameters. During the same time, unigenes which agreed with the criterion of FDR ≤ 0.05, FPKM ≥ 1, and |log2Ratio| ≥1 were selected to be expressed differentlly.

For the KEGG (http://www.genome.jp/kegg/kaas/) or gene ontology (GO) (http://www.geneontology.org/) annotation of all unigenes, enrichment analysis can ensure the further function of differentially expressed unigenes which will participate in the main pathway or physiological function. First of all, according to the following formula, the p-value or corrected p-value (q-value) of the hypergeometric test was calculated. And then, a pathway or GO term whose p-value or q-value was not more than 0.05 can be defined as significant enrichment. In the calculation, N represents the total number of unigenes annotated to the KEGG or GO databases; n represents the number of differentially expressed unigenes in N, and M represents the total number of unigenes annotated to a specific pathway or phrase. The number of differentially expressed unigenes in M is represented by m. TBTools created the heat maps (version: 1.007) [49].

$$P = 1 - \sum_{\mathrm{i}=0}^{\mathrm{m}-1} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$

Weighted gene co-expression network analysis

Weighted gene co-expression network analysis (WGCNA) was performed on the transcriptome data to identify highly coordinated gene sets during fruit gleditsioside biosynthesis, as well as associations between gleditsioside biosynthesis and pod development in *G. sinensis*. Genes with FPKM > 1 were used to construct a weighted gene co-expression network and partition module with the WGCNA v1.6.6 package in R v3.4.4 [50], and visualization were performed with Cytoscape (version 1.7) [51](Shannon et al., 2003).

Declarations

Ethics approval and consent to participate

The collection of plant material complies with relevant guidelines and legislation. Our manuscript does not report on or involve the use of any animal or human data or tissue

Consent for publication

Not applicable.

Availability of data and materials

Raw Illumina sequencing data was deposited in the CNGB Sequence Archive (CNSA) database (https://db.cngb.org/cnsa/handbook/). The accession number is sub039366. All data generated or analyzed during this investigation are included in this published article and its supplementary information files.

Competing Interest

The authors declare that they have no competing interests.

Funding

This work was supported by Basic Research Funds of Henan Academy of Forestry (2022JB01008), Science and Technology Innovation Team of Henan Academy of Agricultural (2023TD28) and the National Natural Science Foundation of China (32071504).

Author Contributions

Y.L. D.F. and Y.W. designed the experiment and edited the manuscript. J.W. and Y.Y. performed most of the experiments and analyzed the data. J.W. and X.D. wrote the manuscript; F.C., H.C., C.W., X.Y. and T.F. revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Supplementary Materials

We have provided detailed information about the materials and methods in our manuscript. Supplementary materials can be found in the online version at https://doi.org……

References

1. Hao X, Han L, Wang Z. Research progress and application of *Gleditsia sinensis*. J Anhui Agricultural Sci. 2012;40:5989–91.
2. Kuwahara Y, Nakajima D, Shinpo S, Nakamura M, Kawano N, Kawahara N, Yamazaki M, Saito K, Suzuki H, Hirakawa H. Identification of potential genes involved in triterpenoid saponins biosynthesis in *Gleditsia sinensis* by transcriptome and metabolome analyses. J Nat Med. 2019;73:369–80. 10.1007/s11418-018-1270-2.
3. Jin SK, Yang HS, Choi JS. Effect of *Gleditsia sinensis* Lam. extract on physico-chemical properties of emulsion-type pork sausages. Korean J Food Sci Anim Resour. 2017;37:274–87. 10.5851/kosfa.2017.37.2.274.
4. Kim KH, Han CW, Yoon SH, Kim YS, Kim JI, Joo M, Choi JY. 2016, The Fruit Hull of *Gleditsia sinensis* Enhances the Anti-Tumor Effect of cis-Diammine Dichloridoplatinum II (Cisplatin). Evid Based Complement Alternat Med 2016, 7480971. doi:10.1155/2016/7480971.
5. Kim Y, Koh JH, Ahn YJ, Oh S, Kim SH. The synergic anti-inflammatory impact of *Gleditsia sinensis* Lam. and lactobacillus brevis KY21 on intestinal epithelial Cells in a DSS-induced Colitis Model. Korean J Food Sci Anim Resour. 2015;35:604–10. 10.5851/kosfa.2015.35.5.604.
6. Zhang JP, Tian XH, Yang YX, Liu QX, Wang Q, Chen LP, Li HL, Zhang WD. *Gleditsia* species: An ethnomedical, phytochemical and pharmacological review. J ETHNOPHARMACOL. 2016;178:155–71. 10.1016/j.jep.2015.11.044.
7. Han JY, Kwon YS, Yang DC, Jung YR, Choi YE. Expression and RNA interference-induced silencing of the dammarenediol synthase gene in *Panax ginseng*. PLANT CELL PHYSIOL. 2006;47:1653–62. 10.1093/pcp/pcl032.
8. Tholl D. Biosynthesis and biological functions of terpenoids in plants. Adv Biochem Eng Biotechnol. 2015;148:63–106. 10.1007/10_2014_295.
9. Xue L, He Z, Bi X, Xu W, Wei T, Wu S, Hu S. Transcriptomic profiling reveals MEP pathway contributing to ginsenoside biosynthesis in *Panax ginseng*. BMC Genomics. 2019;20:383. 10.1186/s12864-019-

5718-x.

10. Vranova E, Coman D, Gruissem W. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. ANNU REV PLANT BIOL. 2013;64:665–700. 10.1146/annurev-arplant-050312-120116.
11. Wang X, Ort DR, Yuan JS. Photosynthetic terpene hydrocarbon production for fuels and chemicals. PLANT BIOTECHNOL J. 2015;13:137–46. 10.1111/pbi.12343.
12. Hildebrand M, Abbriano RM, Polle JE, Traller JC, Trentacoste EM, Smith SR, Davis AK. Metabolic and cellular organization in evolutionarily diverse microalgae as related to biofuels production. CURR OPIN CHEM BIOL. 2013;17:506–14. 10.1016/j.cbpa.2013.02.027.
13. Wu S, Schalk M, Clark A, Miles RB, Coates R, Chappell J. Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants. NAT BIOTECHNOL. 2006;24:1441–7. 10.1038/nbt1251.
14. Li C, Zhu Y, Guo X, Sun C, Luo H, Song J, Li Y, Wang L, Qian J, Chen S. Transcriptome analysis reveals ginsenosides biosynthetic genes, microRNAs and simple sequence repeats in *Panax ginseng* C. A. Meyer BMC GENOMICS. 2013;14:245. 10.1186/1471-2164-14-245.
15. Jeena GS, Fatima S, Tripathi P, Upadhyay S, Shukla RK. Comparative transcriptome analysis of shoot and root tissue of *Bacopa monnieri* identifies potential genes related to triterpenoid saponin biosynthesis. BMC Genomics. 2017;18:490. 10.1186/s12864-017-3865-5.
16. Sun MC, Chao EK, Su XY, Zhu M, Su Y, Qian GT, Chen SL, Wang CX, Xue JP. Study of heterologous efficient synthesis of beta-amyrin and high-density fermentation. Zhongguo Zhong Yao Za Zhi. 2019;44:1341–9. 10.19540/j.cnki.cjcmm.20190129.011.
17. Liu MH, Yang BR, Cheung WF, Yang KY, Zhou HF, Kwok JS, Liu GC, Li XF, Zhong S, Lee SM, Tsui SK. Transcriptome analysis of leaves, roots and flowers of *Panax notoginseng* identifies genes involved in ginsenoside and alkaloid biosynthesis. BMC Genomics. 2015;16:265. 10.1186/s12864-015-1477-5.
18. Liu M, Diretto G, Pirrello J, Roustan JP, Li Z, Giuliano G, Regad F, Bouzayen M. The chimeric repressor version of an ethylene response factor (ERF) family member, Sl-ERF.B3, shows contrasting effects on tomato fruit ripening. NEW PHYTOL. 2014;203:206–18. 10.1111/nph.12771.
19. Yang Y, Zhang Y, He J, Wu Q, Li Y, Li W, Zhou G, Kamol R, Yang X. Transcription factor GlbHLH regulates hyphal growth, stress resistance, and polysaccharide biosynthesis in *Ganoderma lucidum*. J BASIC MICROB. 2022a;62:82–91. 10.1002/jobm.202100334.
20. Yang Y, Wang J, Wang C, Chen H, Liu Y, Wang Y, Gao W. Comprehensive Identification and Profiling of miRNAs Involved in Terpenoid Synthesis of *Gleditsia sinensis* Lam. FORESTS. 2022b;13:108. 10.3390/f13010108.
21. Nelson D, Werck-Reichhart D. A P450-centric view of plant evolution. PLANT J. 2011;66:194–211.
22. Thimmappa R, Geisler K, Louveau T, O'Maille P, Osbourn A. Triterpene biosynthesis in plants. ANNU REV PLANT BIOL. 2014;65:225–57. 10.1146/annurev-arplant-050312-120229.
23. He L, Zhang S, Qiu Z, Zhao J, Nie W, Lin H, Zhu Z, Zeng D, Qian Q, Zhu L. 2018, FRUCTOKINASE-LIKE PROTEIN 1 interacts with TRXz to regulate chloroplast development in rice. J INTEGR PLANT BIOL

60, 94–111. doi:10.1111/jipb.12631.

24. Ambawat S, Sharma P, Yadav NR, Yadav RC. MYB transcription factor genes as regulators for plant responses: an overview. Physiol Mol Biol Plants. 2013;19:307–21. 10.1007/s12298-013-0179-1.
25. Tamura K, Yoshida K, Hiraoka Y, Sakaguchi D, Chikugo A, Mochida K, Kojoma M, Mitsuda N, Saito K, Muranaka T, Seki H. The Basic helix-loop-helix transcription factor GubHLH3 positively regulates soyasaponin biosynthetic genes in *Glycyrrhiza uralensis*. PLANT CELL PHYSIOL. 2018;59:778–91. 10.1093/pcp/pcy046.
26. Vom ED, Kijne JW, Memelink J. 2002, Transcription factors controlling plant secondary metabolism: what regulates the regulators? PHYTOCHEMISTRY 61, 107–114. doi:10.1016/s0031-9422(02)00185-1.
27. Zhang W, Xu F, Cheng S, Liao Y. Characterization and functional analysis of a MYB gene (GbMYBFL) related to flavonoid accumulation in *Ginkgo biloba*. Genes Genomics. 2018;40:49–61. 10.1007/s13258-017-0609-5.
28. de Vries S, Kloesges T, Rose LE. Evolutionarily dynamic, but robust, targeting of resistance genes by the miR482/2118 gene family in the *Solanaceae*. GENOME BIOL EVOL. 2015;7:3307–21. 10.1093/gbe/evv225.
29. Ubi BE, Saito T, Bai S, Nishitani C, Ban Y, Ikeda K, Ito A, Moriguchi T. Characterization of 10 MADS-box genes from Pyrus pyrifolia and their differential expression during fruit development and ripening. Gene. 2013;528:183–94. 10.1016/j.gene.2013.07.018.
30. Bocobza SE, Malitsky S, Araujo WL, Nunes-Nesi A, Meir S, Shapira M, Fernie AR, Aharoni A. Orchestration of thiamin biosynthesis and central metabolism by combined action of the thiamin pyrophosphate riboswitch and the circadian clock in *Arabidopsis*. PLANT CELL. 2013;25:288–307. 10.1105/tpc.112.106385.
31. Augustin JM, Kuzina V, Andersen SB, Bak S. Molecular activities, biosynthesis and evolution of triterpenoid saponins. Phytochemistry. 2011;72:435–57. 10.1016/j.phytochem.2011.01.015.
32. Hampel D, Mosandl A, Wust M. Biosynthesis of mono- and sesquiterpenes in carrot roots and leaves (*Daucus carota* L.): metabolic cross talk of cytosolic mevalonate and plastidial methylerythritol phosphate pathways. Phytochemistry. 2005;66:305–11. 10.1016/j.phytochem.2004.12.010.
33. Zhao S, Wang L, Liu L, Liang Y, Sun Y, Wu J. Both the mevalonate and the non-mevalonate pathways are involved in ginsenoside biosynthesis. PLANT CELL REP. 2014;33:393–400. 10.1007/s00299-013-1538-7.
34. Kevei Z, Lougnon G, Mergaert P, Horvath GV, Kereszt A, Jayaraman D, Zaman N, Marcel F, Regulski K, Kiss GB, Kondorosi A, Endre G, Kondorosi E, Ane JM. 3-hydroxy-3-methylglutaryl coenzyme a reductase 1 interacts with NORK and is crucial for nodulation in *Medicago truncatula*. PLANT CELL. 2007;19:3974–89. 10.1105/tpc.107.053975.
35. Mertens J, Pollier J, Vanden BR, Lopez-Vidriero I, Franco-Zorrilla JM, Goossens A. The bHLH transcription factors TSAR1 and TSAR2 regulate triterpene saponin biosynthesis in *Medicago truncatula*. PLANT PHYSIOL. 2016;170:194–210. 10.1104/pp.15.01645.

36. Samad AFA, Rahnamaie-Tajadod R, Sajad M, Jani J, Murad AMA, Noor NM, Ismail I. Regulation of terpenoid biosynthesis by miRNA in *Persicaria minor* induced by Fusarium oxysporum. BMC Genomics. 2019;16(1):586. 10.1186/s12864-019-5954-0.
37. Zhang Z, Koike K, Jia Z, Nikaido T, Guo D, Zheng J. Four new triterpenoidal saponins acylated with one monoterpenic acid from *Gleditsia sinensis*. J NAT PROD. 1999a;62:740–5. 10.1021/np980441k.
38. Zhang Z, Koike K, Jia Z, Nikaido T, Guo D, Zheng J. Triterpenoidal saponins from *Gleditsia sinensis*. Phytochemistry. 1999b;52:715–22. 10.1016/s0031-9422(99)00238-1.
39. Miyase T, Melek FR, Warashina T, Selim MA, El FN, Kassem IA. 2010, Cytotoxic triterpenoid saponins acylated with monoterpenic acids from fruits of *Gleditsia* caspica Desf. PHYTOCHEMISTRY 71, 1908–1916. doi:10.1016/j.phytochem.2010.08.001.
40. Henry LK, Gutensohn M, Thomas ST, Noel JP, Dudareva N. Orthologs of the archaeal isopentenyl phosphate kinase regulate terpenoid production in plants. Proc Natl Acad Sci U S A. 2015;112:10050–5. 10.1073/pnas.1504798112.
41. Nagegowda DA, Gupta P. Advances in biosynthesis, regulation, and metabolic engineering of plant specialized terpenoids. PLANT SCI. 2020;294:110457. 10.1016/j.plantsci.2020.110457.
42. Gutensohn M, Orlova I, Nguyen TT, Davidovich-Rikanati R, Ferruzzi MG, Sitrit Y, Lewinsohn E, Pichersky E, Dudareva N. Cytosolic monoterpene biosynthesis is supported by plastid-generated geranyl diphosphate substrate in transgenic tomato fruits. PLANT J. 2013;75:351–63. 10.1111/tpj.12212.
43. Xiao YY, Chen JY, Kuang JF, Shan W, Xie H, Jiang YM, Lu WJ. Banana ethylene response factors are involved in fruit ripening through their interactions with ethylene biosynthesis genes. J EXP BOT. 2013;64:2499–510. 10.1093/jxb/ert108.
44. Feng BH, Han YC, Xiao YY, Kuang JF, Fan ZQ, Chen JY, Lu WJ. The banana fruit Dof transcription factor MaDof23 acts as a repressor and interacts with MaERF9 in regulating ripening-related genes. J EXP BOT. 2016;67:2263–75. 10.1093/jxb/erw032.
45. Li M, Wang X, Li C, Li H, Zhang J, Ye Z. Silencing GRAS2 reduces fruit weight in tomato. J INTEGR PLANT BIOL. 2018;60:498–513. 10.1111/jipb.12636.
46. Heo JO, Chang KS, Kim IA, Lee MH, Lee SA, Song SK, Lee MM, Lim J. Funneling of gibberellin signaling by the GRAS transcription regulator scarecrow-like 3 in the *Arabidopsis* root. Proc Natl Acad Sci U S A. 2011;108:2166–71. 10.1073/pnas.1012215108.
47. Kou X, Watkins CB, Gan SS. *Arabidopsis* AtNAP regulates fruit senescence. J EXP BOT. 2012;63:6139–47. 10.1093/jxb/ers266.
48. Fu CC, Han YC, Kuang JF, Chen JY, Lu WJ. Papaya CpEIN3a and CpNAC2 co-operatively regulate carotenoid biosynthesis-related genes CpPDS2/4, CpLCY-e and CpCHY-b during fruit ripening. PLANT CELL PHYSIOL. 2017;58:2155–65. 10.1093/pcp/pcx149.
49. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: An integrative toolkit developed for interactive analyses of big biological data. MOL PLANT. 2020;13:1194–202. 10.1016/j.molp.2020.06.009.

50. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559. 10.1186/1471-2105-9-559.
51. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. GENOME RES. 2003;13:2498–504. 10.1111/j.1365-313X.2011.04529.x.

Figures

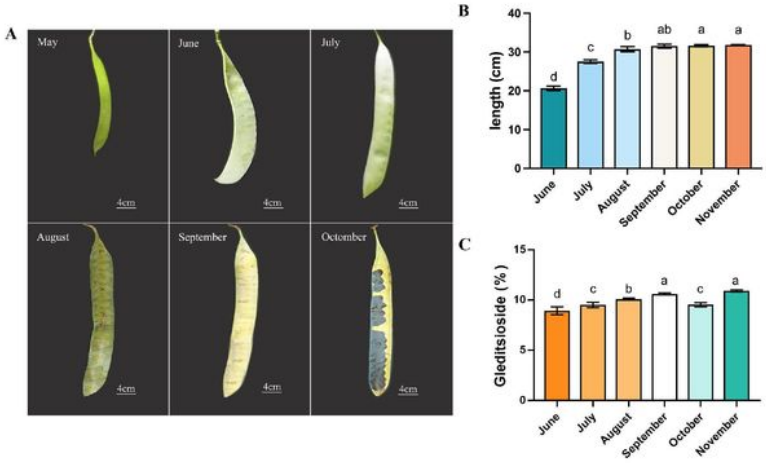


Figure 1. Phenotype, length and gleditsioside content of *G. sinensis* during pod developmental from May to October. (**A**) *G. sinensis* pods developmental phenotype; (**B**) The length of pods; (**C**) The gleditsioside contents of pods. The scale bar = 4 cm. *** represent p-value < 0.001, ** represent p-value < 0.01.

Figure 1

See image above for figure legend.

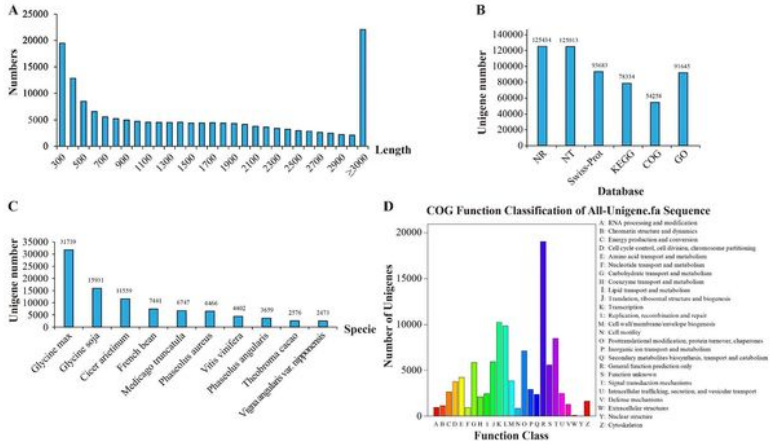


Figure 2. Annotation and classification of unigenes in *G. sinensis*. (**A**) Length distribution of all unigenes in pods; (**B**) Annotation of unigenes of *G. sinensis*; (**C**) Expressed genes in the top ten KEGG pathways; (**D**) Cluster of Orthologous Groups; (COG) function categories of unigenes in *G. sinensis*. COG function classes were demonstrated on the x-axis and the full names of the functions were annotated on the right. The numbers of unigenes in one class were displayed on the y-axis.

Figure 2

See image above for figure legend.

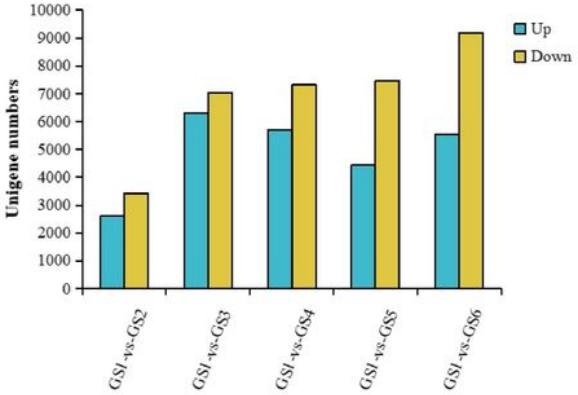


Figure 3. Differentially expressed unigene statistics in comparable groups of *G. sinensis* pods. (GS1. May; GS2. June; GS3. July; GS4. August; GS5. September; GS6. October).

Figure 3

See image above for figure legend.

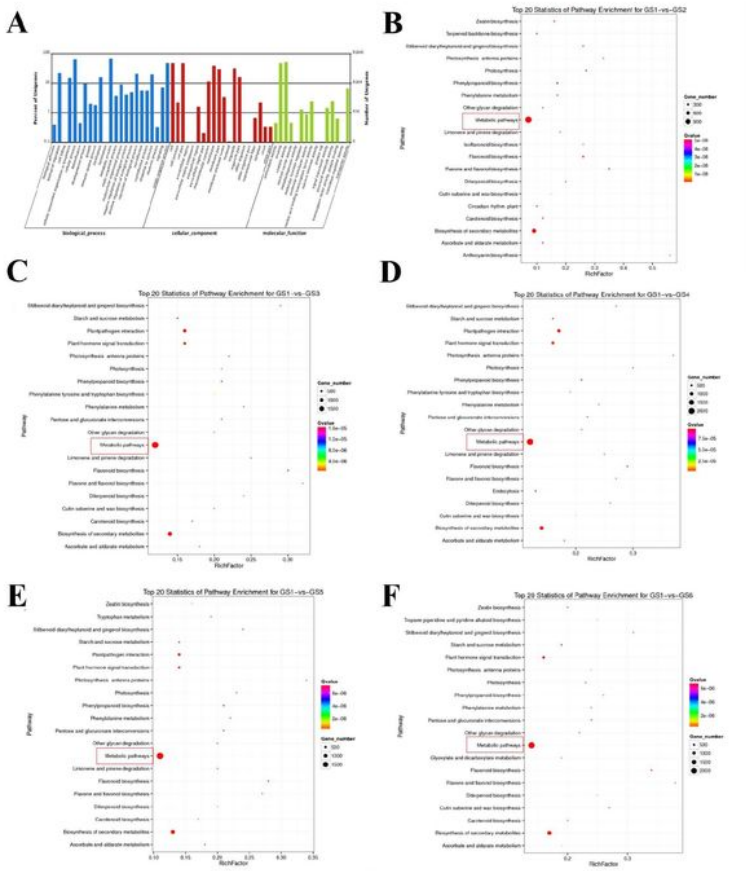


Figure 4. GO annotation and KEGG enrichment of DEGs in *G. sinensis*. (**A**) The GO classification of all blasted unigenes. The percentage of unigenes in one class was shown on the y-axis, while unigenes during six development stages of *G. sinensis* were classified into three classes (biological processes, cellular component, molecular function) on the x-axis; (**B**) Top 20 statistics of pathway enrichment for GS1-vs-GS2; (C) Top 20 statistics of pathway enrichment for GS1-vs-GS3; (**D**) Top 20 statistics of pathway enrichment for GS1-vs-GS4; (**E**) Top 20 statistics of pathway enrichment for GS1-vs-GS5; (**F**) Top 20 statistics of pathway enrichment for GS1-vs-GS6. The red asterisk represents the main pathway of enrichment. The red boxes indicate significant enrichment.

Figure 4

See image above for figure legend.

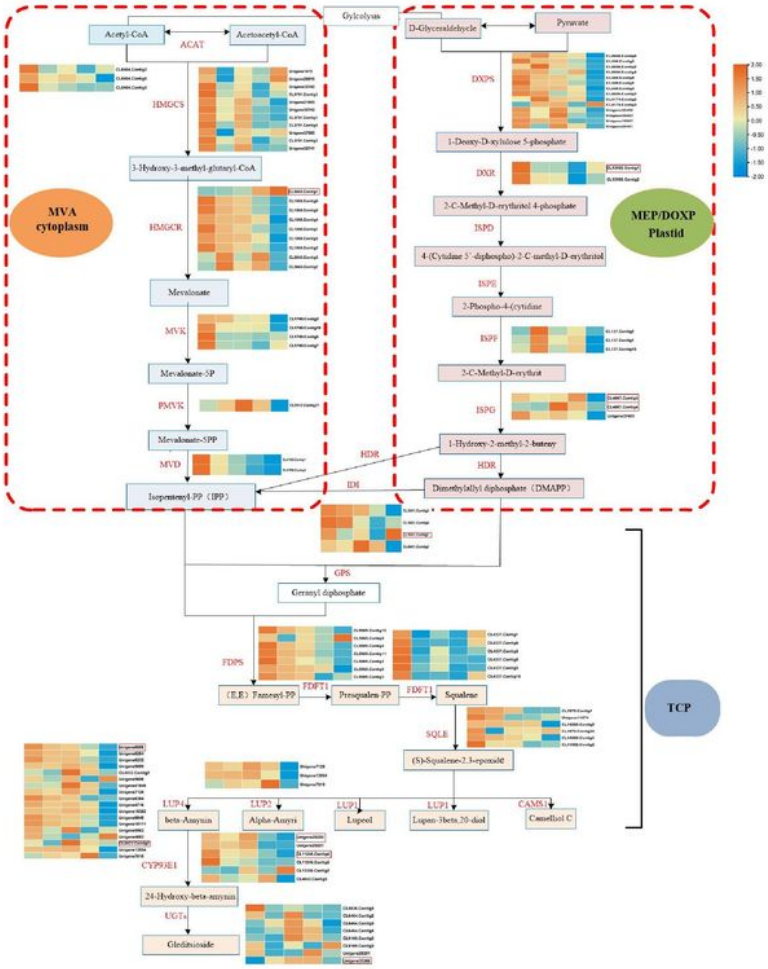


Figure 5. Heatmap and schematic diagram for genes related to terpenoid biosynthesis in *G. sinensis* pods (Reference from KEGG database) (red font: differentially expressed genes; IPP: isopentenyl pyrophosphate; DMAPP: dimethylallyl pyrophosphate; GPP: geranyl diphosphate; FPP: farnesyl pyrophosphate). TCP: Triterpene cyclization process. MVA pathway: mevalonate pathway in the cytoplasm; MEP/DOXP pathway: methylerythritol phosphate pathway in the plastid. The heatmap from left to right shows the DEGs of log2Ratio (GS2/GS1), log2Ratio (GS3/GS1), log2Ratio (GS4/GS1), log2Ratio (GS5/GS1), log2Ratio (GS6/GS1), respectively. The red boxes represent the key differential genes in the gleditsioside synthesis pathway.

Figure 5

See image above for figure legend.

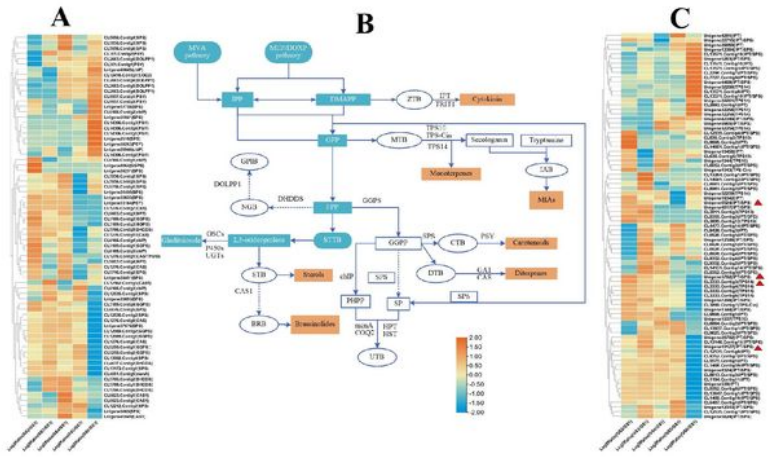


Figure 6. Parallel pathways of gleditsioside biosynthesis and gene expression profile of the key enzymes in *G. sinensis*. (**A**) Gene expression profiles of STB, BRB, NGB, and UTB pathways; (**B**) Linking of gleditsioside biosynthesis with other parallel pathways. Blue Box is the main gleditsioside biosynthesis process; blue oval: parallel pathways; orange square framework: hormones synthesized by parallel pathways; GGPP: geranylgeranyl diphosphate; SP: solanesyl pyrophosphate; PHPP: phytyl diphosphate. Solid arrows indicate direct transformation, and dashed arrows indicate indirect or complicated connections; (**C**) Gene expression profiles of ZTB, MTB, IAB, CTB, and DTB pathways. The red triangle represents the key genes.

Figure 6

See image above for figure legend.

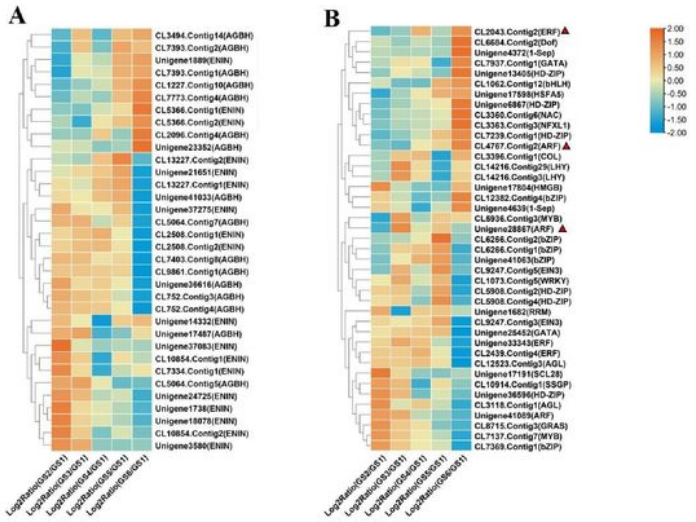


Figure 7. Differentially expressed unigenes in *G. sinensis* pods. (**A**) Differential expressed enzyme genes related to gleditsioside biosynthesis from GO enrichment; (**B**) Highly differentially expressed unigene coding transcription factors. ENIN: term of enzyme inhibitor activity (GO: 0004857), AGBH: term of hydrolase activity acting on glycosyl bonds (GO: 0016798.From left to right in the heatmap, the periods are May, June, July, August, September, and October. The red triangle represents the key differential genes in the gleditsioside synthesis pathway.

Figure 7

See image above for figure legend.

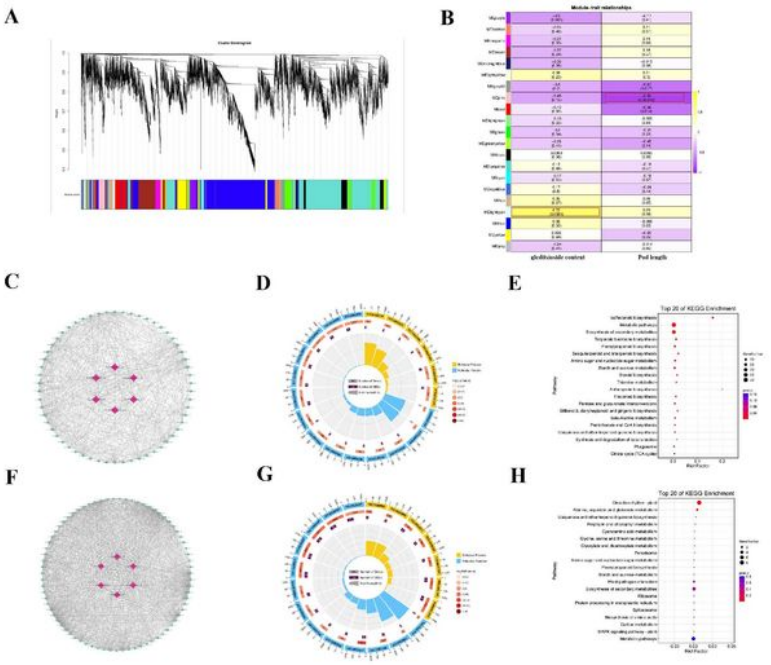


Figure 8. Construction of the gene co-expression network through weighted gene co-expression network analyses (WGCNA). (A) Gene dendrogram obtained by hierarchical clustering with the module color. A total of 21 distinct modules were identified; (B) Heat map of correlation between modules and traits. Red boxes represent the correlation between gene modules and gleditsioside content and pod length, respectively; (C) Hub genes were associated with gleditsioside; (D) GO enrichment analysis results of gleditsioside; (E) KEGG enrichment analysis results of gleditsioside; (F) Hub genes associated with pods development; (G) GO enrichment analysis results of pod development; (H) KEGG enrichment analysis results of pod development.

Figure 8

See image above for figure legend.

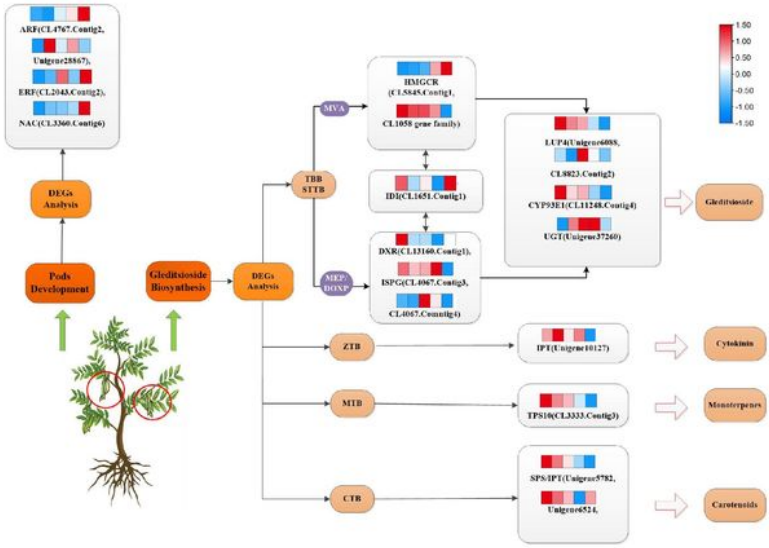


Figure 9. Mechanisms of gleditsioside synthesis and pod developmental associations in *G. sinensis*. Two kinds of orange represent some key genes and TFs obtained by DEGs analysis in gleditsioside synthesis and pod development. Brown represents the final product of the gleditsioside synthesis pathway and other pathway.

Figure 9

See image above for figure legend.

Supplementary Files

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

- SupplementaryFigure1.pdf
- SupplementaryFigure2.pdf
- SupplementaryFigure3.pdf
- SupplementaryFigure4.pdf
- SupplementaryFigure5.pdf
- SupplementaryFigure6.pdf
- SupplementaryTable1.xlsx
- SupplementaryTable2.xlsx
- SupplementaryTable3.docx
- SupplementaryTable4.xlsx
- SupplementaryTable5.docx
- SupplementaryTable6.xlsx
- SupplementaryTable7.xlsx
- SupplementaryTable8.docx
- SupplementaryTable9.xlsx
- SupplementaryTable10.docx
- SupplementaryTable11.docx