scientific reports

Check for updates

OPEN

A candidate gene identified in converting platycoside E to platycodin D from *Platycodon grandiflorus* by transcriptome and main metabolites analysis

Xinglong Su[1,2], Yingying Liu[3], Lu Han[1], Zhaojian Wang[1,2], Mengyang Cao[1,2], Liping Wu[1], Weimin Jiang[4], Fei Meng[1], Xiaohu Guo[1], Nianjun Yu[1], Shuangying Gui[1,5], Shihai Xing[1,2,6]✉ & Daiyin Peng[1,2,7]✉

Platycodin D and platycoside E are two triterpenoid saponins in *Platycodon grandiflorus*, differing only by two glycosyl groups structurally. Studies have shown β-Glucosidase from bacteria can convert platycoside E to platycodin D, indicating the potential existence of similar enzymes in *P. grandiflorus*. An $L_9(3^4)$ orthogonal experiment was performed to establish a protocol for calli induction as follows: the optimal explant is stems with nodes and the optimum medium formula is MS + NAA 1.0 mg/L + 6-BA 0.5 mg/L to obtain callus for experimental use. The platycodin D, platycoside E and total polysaccharides content between callus and plant organs varied wildly. Platycodin D and total polysaccharide content of calli was found higher than that of leaves. While, platycoside E and total polysaccharide content of calli was found lower than that of leaves. Associating platycodin D and platycoside E content with the expression level of genes involved in triterpenoid saponin biosynthesis between calli and leaves, three contigs were screened as putative sequences of β-Glucosidase gene converting platycoside E to platycodin D. Besides, we inferred that some transcription factors can regulate the expression of key enzymes involved in triterpernoid saponins and polysaccharides biosynthesis pathway of *P. grandiflorus*. Totally, a candidate gene encoding enzyme involved in converting platycoside E to platycodin D, and putative genes involved in polysaccharide synthesis in *P. grandiflorus* had been identified. This study will help uncover the molecular mechanism of triterpenoid saponins biosynthesis in *P. grandiflorus*.

Abbreviations

AACT	Acetoacetyl-coenzyme A
6-BA	6-Benzylaminopurine
Amylosucrase	1,4-α-D-glucan-4-α-D-glucosyltransferase-glucan
AXS	UDP-apiose/xylose synthase
B5	Gamborg B5 Medium
CAS	Cycloartenol synthase
CMK	4-Diphosphocytidyl-2-C-methyl-D-erythritol kinase
CMS	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (4-diphosphocytidyl-2C-methyl-D-erythritol synthase)

[1]School of Pharmacy, Anhui University of Chinese Medicine, Hefei 230012, China. [2]Institute of Traditional Chinese Medicine Resources Protection and Development, Anhui Academy of Chinese Medicine, Hefei 230012, China. [3]College of Humanities and International Education Exchange, Anhui University of Chinese Medicine, Hefei 230012, China. [4]College of Life Sciences and Environment, Hengyang Normal University, Hengyang 421008, Hunan, China. [5]Anhui Province Key Laboratory of Pharmaceutical Preparation Technology and Application, Anhui University of Chinese Medicine, Hefei 230012, China. [6]Anhui Province Key Laboratory of Research and Development of Chinese Medicine, Hefei 230012, China. [7]Synergetic Innovation Center of Anhui Authentic Chinese Medicine Quality Improvement, Hefei 230038, China. ✉email: xshshihai@163.com; pengdy@ahtcm.edu.cn

DXR	1-Deoxy-D-xylulose 5-phosphate reductoisomerase
DXS	1-Deoxy-D-xylulose 5-phosphate synthase
FPPS	Farnesyl-diphosphate synthase
GALE	UDP-glucose 4-epimerase
GMDS	GDP-mannose-4,6-dehydratase
GMPP	Mannose-1-phosphate Guanylyltransferase
GPI	Glucose-6-phosphate isomerase
GPPS	Geranylgeranyl pyrophosphate synthase
HDR	4-Hydroxy-3-methylbut-2-enyl diphosphate reductase
HDS	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase
HK	Hexokinase
HMGR	3-Hydroxy-3-methylglutaryl-coenzyme A reductase
HMGS	Hydroxymethyl glutaryl-CoA synthase
IDI	Sopentenyl-diphosphate Delta-isomerase
INV	Isopentenyl-diphosphate Delta-isomerase
LS	Lupeol synthase
MCS	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase
MEP/DOXP	Non-mevalonate pathway
MPI	Mannose-6-phosphate isomerase
MS	Murashige & Skoog Medium
MVA	Mevalonic acid pathway
MVD	Mevalonate diphosphate decarboxylase
MVK	Mevalonate kinase
PD	Platycodin D
PE	Platycoside E
PGM	Phosphoglucomutase
PGPS	*Platycodon grandiflorus* Polysaccharides
PMK	Phosphomevalonate kinase
PMM	Phosphomevalonate kinase
RHM	UDP-glucose-4,6-dehydratase
SacA	β-Fructofuranosidase
ScrA	Sucrose-specific enzyme II
ScrK	Fructokinase
SE	Squalene epoxidase
SS	Squalene synthase
SUS	Sucrose synthase
TSTA3	GDP-L-fucose synthase
UER1	3,5-Epimerase-4-reductase
UGDH	UDP-glucose-6-dehydrogenase
UGE	UDP-glucuronate 4-epimerase
UGE	UDP-glucuronate 4-epimerase
UGP2	UTP-glucose-1-phosphate Uridylyltransferase
USP	UDP-sugar pyrophosphorylase
UXE	UDP-arabinose-4-epimerase
UXS1	UDP-glucuronate decarboxylase
WPM	Lloyd & McCown Woody Plant Basal Medium
α-NAA	α-Naphthylacetic acid
β-A28O	Isolate CYP716A140 beta-amyrin 28-oxidase
β-AS	Beta-amyrin synthase

Platycodon grandiflorus (Jacq.) A. DC., a perennial herb, is the sole species in genus Platycodon within the Campanulaceae family. The flowers of *P. grandiflorus* are blue purple or white, which can be used for ornamental and horticultural purpose. The root (platycodi radix) of *P. grandiflorus* has diverse pharmacological activities and can be used for the treatment of some chronic inflammatory diseases such as asthma, bronchitis and tuberculosis[1]. The dried form of the platycodi radix is officially listed as a traditional herbal medicine in the Chinese, Korean and Japanese Pharmacopoeia[2]. It is also being pickled in northeast China, and made into kimchi in the Korean Peninsula. It has medicinal, edible, ornamental value in one, with immeasurable development prospects.

As a traditional Chinese herb, *P. grandiflorus* is a rich source of natural secondary metabolic products that have various chemical structural types. More than 100 compounds have been isolated from *P. grandiflorus*, including steroidal saponins, flavonoids, phenolic acids, polyacetylenes, and sterols, etc.[3]. Triterpenoid saponins are the main active components in *P. grandiflorus*, including platycodin D (PD), platycoside E (PE), platycodigenin and platyconic acid A, etc. PD is one of the major triterpenoid saponins with higher pharmacological activity than the other platycosides from platycodi radix, and have multiple pharmacological effects, such as immunostimulation[4], anti-inflammation[5], anti-obesity[6], anti-atherosclerosis[1], and anticancer[7]. The structure of platycoside E is similar to that of platycodin D, and both of them are oleanane-type triterpenoid saponin. PE has two additional glucose groups compared to PD[8]. PE could convert to PD through the hydrolysis action of a de-glucosidase. *P. grandiflorus* polysaccharides (PGPs) are another important active component in this medicinal plant, and studies have confirmed that PGPs are involved in antioxidant activity[9], it can activate macrophage

and enhance non-specific immunity function[10]. A research has shown that a selenium polysaccharide from the platycodi radix may be considered as a potential and useful antioxidant agent[9].

Pentacyclic triterpenoid saponins have been well-known as important secondary metabolites in plants. 2, 3-oxidosqualene, a direct precursor of triterpenoid saponins, is synthesized mainly by the mevalonic acid (MVA) pathway[11]. However, the operator of the MVA pathway in regulating the biosynthesis of triterpenoids even phytosterols in *P. grandiflorus* has not been clearly described[12]. Farnesyl pyrophosphate (farnesyl-PP) is synthesized from isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) under the catalysis of farnesyl pyrophosphate synthase (FPP)[11,13]. Under the catalysis of enzymes such as squalene synthase (SS) and squalene epoxidase (SE), 2,3-oxidosqualene is produced thereafter[14]. Subsequently, triterpenoids saponins synthesized from 2,3-oxidosqualene by three kinds of enzymes which are oxidized squalene cyclase, cytochrome P450 monooxygenase and uridine diphosphate-dependent glycosyltransferase. In other words, 2,3-oxidosqualene is converted to polygalacic acid, platycodigenin and platycogenic acid A by successive enzymes such as β-amyrin synthase (β-AS), β-amyrin 28-oxidase (β-A28O) and a series of cytochrome P450 (CYPs)[15]. Subsequently, the conversion of polygalacic acid into polygalacin D, platycodigenin into platycoside E, and platycogenic acid A into platyconic acid A are all catalyzed by certain kinds of GTs (Glycosyltransferases)[16]. Platycodin D, an oleanane-type triterpenoid saponin, is the main bioactive component and has stronger pharmacological activities, but little is clear on its biosynthesis in *P. grandiflorus* at present. It has been found that platycoside E is a precursor of platycodin D, and PE can be converted to PD by enzyme catalysis. Their biological activities can be increased and their bioavailability and cell permeability would be improved due to their reduced size resulted by de-glycosylation of saponins[17].

From the chemical structure of platycodin D and platycoside E, it can be predicted that there are enzymes which can catalyze degradation of glycosyl group from PE to PD. Two extracellular experiments showed that β-D-glucosidase from *Aspergillus usamii* and *Caldicellulosiruptor bescii* can successfully catalyze conversion of PE and platycodin D3 into PD under optimal reactions conditions[18,19]. In this study we attempt to find out appropriate candidate genes encoding enzymes involving in conversion of PE, platycodin D3 into PD in traditional Chinese medicinal plant *P. grandiflorus*. The pathway of triterpenoid saponins in *P. grandiflorus* is predicted referring to the terpenoid backbone and saponin biosynthesis in KEGG[20] (https://www.kegg.jp/dbget-bin/www_bget?map00900), as shown in Fig. 1.

Polysaccharides are extremely important bio-macromolecules, and have a wide range of industrial value and clinical role[21]. There are enormous types of polysaccharides in *P. grandiflorus*, named as *P. grandiflorus* polysaccharides (PGPs), which have important pharmacological activities. However, few relevant studies and analyses on its biosynthetic pathways were reported. Sucrose is firstly produced in plants through photosynthesis in plant chloroplast. Sucrose generates UDP-Glc in the presence of sucrose synthase (SUS), and then UDP-Glc is converted into GDP-Glc under successive catalytic reactions, including UDP-sugar pyrophosphorylase (USP)[22], UDP-glucose phosphorylase (UGP), GDP-glucose pyrophos-phorylase (GGP)[23]. Sucrose-6P is biosynthesized from sucrose under the catalysis of sucrose-specific enzyme II (ScrA)[24,25]. Subsequently, a series of enzymes such as β-Fructofuranosidase (SacA), Mannose-6-phosphate isomerase[11], Mannose-1-phosphate Guanylyl-transferase (GMPP) and other enzymes catalyze sucurose-6P to generate GDP-Man. Then, GDP-Fuc is synthesized from GDP-Man by enzymes such as GDP-Mann 4,6-dehydratase (GMDS) and GDP-L-Fucose (TSTA3)[26,27]. In addition, through the catalysis of nucleotide-diphospho-sugar (NDP-sugar) interconversion enzymes (NSEs), many other NDP sugars were produced from UDP-Glc and GDP-Man[28]. Another way is that sucrose invertase (INV) and hexokinase (HK) catalyze the production of D-Glucose-6P from sucrose[29]. Subsequently, D-Glucose-6P is converted into UDP-Gal, UDP-GalA, UDP-Rha and other polysaccharide precursors under the catalysis of enzymes such as Phosphoglucomutase (PGM), UTP-glucose-1-phosphate Uridylyltransferase (UGP2), and UDP-apiose/xylose synthase (AXS) et al[30]. Finally, through the catalysis of various glycosyltransferases (GTs), the active monosaccharide unit NDP-sugar is added to the sugar residues of various polysaccharides and glycoconjugates[26]. The biosynthesis pathway of polysaccharides in *P. grandiflorus* was inferred as shown in Supplementary Figure S1.

RNA-Seq has been widely used to analyze the pathway of secondary metabolite biosynthesis and regulation, excavate key enzyme genes and transcription factors in medicinal plants[31,32], which provides a scientific basis for the efficient accumulation and effective utilization of active ingredients in medicinal plants. Transcriptome sequencing have been completed in many medicinal plants such as *Taxus chinensis*[33], *Panax quinquefolius*[34], *Coptis chinensis*[35], *Panax zingiberensis*[15], and many other medicinal plants. Many functional genes are discovered and partial specific biosynthetic pathways of secondary metabolite are analyzed.

In this paper, calli of *P. grandiflorus* were induced by tissue culture techniques. The contents of PE, PD and PGPs in calli and each organ of original plant of *P. grandiflorus* were determined by high performance liquid chromatography (HPLC) and Phenol–sulfuric acid method, respectively. Based on the differences of the secondary metabolites between calli and organs of original plant *P. grandiflorus*, RNA-seq was performed between calli and leaves. Comparative analysis of the transcriptome data provides valuable resources for further studies of the molecular mechanisms of terpenoids, saponins and polysaccharides biosynthesis in *P. grandiflorus*. In this study, the differences between calli and organs of original plants were analyzed, and candidate key genes and transcription factors were identified to help our knowledge of the metabolism and regulation of secondary metabolites in *P. grandiflorus*.

Results

Calli induction of *Platycodon grandiflorus*. The effects of four explants, different basic media, and plant growth regulator combinations on calli induction of *P. grandiflorus* were studied by orthogonal experimental design of 4 factors and 3 levels ($L_9(3^4)$) without consideration of interactions among factors. Each treatment has

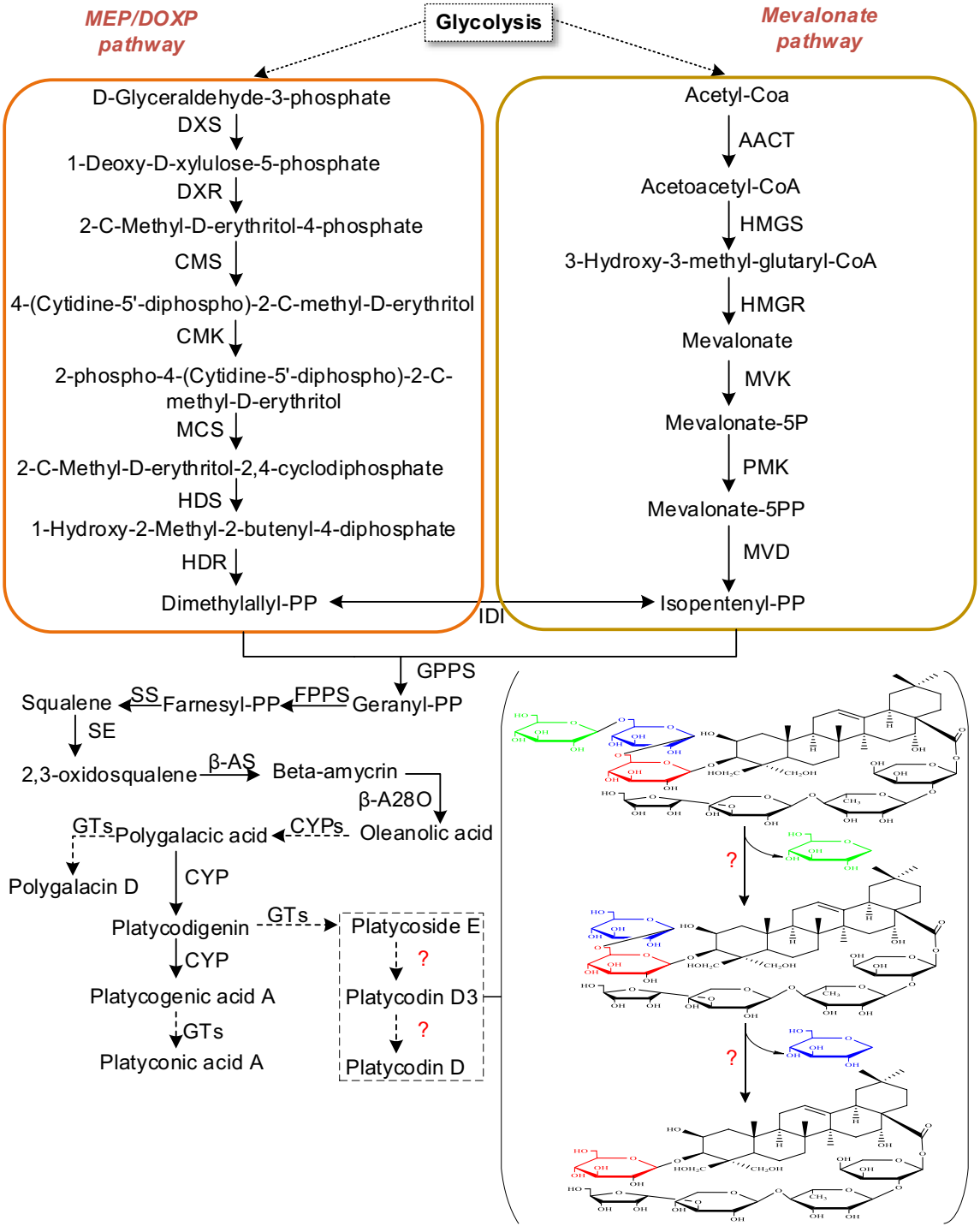

Figure 1. Triterpenoid saponin biosynthetic pathways predicted in *P. grandiflorus*. Arrows with solid lines represent the identified enzymatic reactions, and arrows with dashed lines represent multiple enzymatic reactions through multiple steps and putative enzymatic reactions.

60 independent duplicates (20 bottles with 3 pieces). The calli induced are shown in Supplementary Figure S2, and the results are shown in Table 1.

The induction rates of the calli were calculated with the number of the calli divided by the number of uncontaminated explants. The K value and R value were obtained by statistical analysis of orthogonal experiment data, as the K value was the average of the rates from each level, and the R value called extreme difference is the difference between the maximum and minimum average values of each factor.

By analysis of variance (ANOVA), a significant difference among these factors and their levels ($F = 8.67 > F_{0.01}(3, 4) = 6.59$) was identified. It can be concluded from the R value that the category of explants had a great influence on the calli induction, and stem with nodes is the optimal explants. It also can be concluded that the MS

	Factors				Inoculation Num (non-contaminated)	Calli Num	Induction rate (%)
	Media (A)	Explants (B)	NAA mg/L (C)	6-BA mg/L (D)			
1	B5	Leaves	0.1	0.5	56	4	7.14
2	B5	Stems with nods	0.2	1.0	58	22	37.93
3	B5	Stems	1.0	2.0	57	8	14.04
4	MS	Leaves	0.2	2.0	53	6	11.32
5	MS	Stems with nods	1.0	0.5	57	57	100.00
6	MS	stems	0.1	1.0	54	2	3.70
7	WPM	leaves	1.0	1.0	55	0	0.00
8	WPM	Stems with nods	0.1	2.0	58	23	39.66
9	WPM	stems	0.2	0.5	56	4	7.14
k1	59.11	18.46	50.50	114.28			
k2	115.02	177.59	56.39	41.63			
k3	46.80	24.88	114.04	65.02			
K1	19.70	6.15	16.83	38.09			
K2	38.34	59.20	18.80	13.88			
K3	15.60	8.29	38.01	21.67			
R	22.74	53.05	21.18	24.21			

Table 1. Orthogonal experiment ($L_9(3^4)$) analyzing the effects of different factors on callus induction of *P. grandiflorus*.

medium is better than the other two basic media and 6-BA could get calli induction more efficiently than NAA, and the optimal concentrations are 0.5 mg/L and 1.0 mg/L, respectively. Based on the above analysis, the best combination is A2B2C3D1. In order to verify whether the A2B2C3D1 is the best combination or not, a total of 30 stems with nodes were inoculated in 10 bottles with 3 duplicates per bottle, a total of 25 calli were induced after 50 days, and the induction rate is up to 83.33%.

It can be concluded from what is stated above that the best protocol for calli induction of *P.grandiflorus* is to use stem with nodes as the explants, and take MS + NAA 1.0 mg/L + 6-BA 0.5 mg/L as the optimal formula.

Comparative of PD and PE content in organs and calli. Two kinds of specific platyopsis saponins PD and PE were identified by comparing the retention time[34] and maximum uptake of the samples and standards at a wavelength of 210 nm by HPLC. The results showed that the concentrations of PD and PE are positively correlated with the area of the peak in the test range ($R^2 > 0.9999$, $R^2 > 0.9946$), and the content of each saponins was calculated from the peak area versus its own standard curve. The retention time and peak profile of PD and PE in standard and one sample are shown in Fig. 2. The content variance of PD and PE in callus and different organs in *P. grandiflorus* was shown in Supplementary Table S1 and Fig. 3.

ANOVA indicated that the distribution of PD content varied wildly among organs in plants of *P. grandiflorus* (Fig. 3A). It can be concluded that PD was most enriched in flower buds, followed by the roots, while PD content in leaves, stems and the whole plants is low. When it comes to the distribution of PD content in stems and roots, the contents in the bark and phloem are higher than that in the xylem. The results also showed that the content of PD in callus is higher than that of the whole plants, leaves, stems and roots, and it only lower than that in the flower buds. It can be inferred that PD could be obtained more efficiently from flowers and calli than from any other organs and tissues. Since the biomass of flower buds in the whole plants is very low, callus of *P. grandiflorus* may become an alternative material for PD production.

As for the distribution of PE content in *P. grandiflorus*, only the PE content in leaves is higher than that in the whole plant, while the PE content of flowers, roots and stems is much lower than that of the whole plant (Fig. 3B). It can be concluded from PE content distribution in each organ that the PE content of phloem and bark is higher than that of xylem which is consistent with the distribution of PD content distribution described above. The result also shows that the content of PE of calli is lower than that of all the organs in *P. grandiflorus*.

We found the content of PD in calli was higher than that in leaves which was sharply contrast to the highest PE and lower PD content in leaves. Experiments in vitro have shown that through enzymes catalysis, PE could convert to PD with higher pharmacological activities[18,19]. Therefore, we speculate the existence of putative enzymes converting PE to PD in *P. grandiflorus*. Based on the significant differences of PD and PE content between calli and leaves of *P. grandiflorus*, an RNA-Seq experiment was designed to identify the candidate genes.

Total polysaccharide content variation among calli and plant organs. A standard curve was drawn from the absorbance value versus 7 gradient concentrations of standard glucose work solution with blank at 486 nm following chromogenic reaction (Fig. 4A). The regression equation of the standard curve is $Y = 0.026X - 0.0226$, $R^2 = 0.9991$, showing that the quantification responds to linearity within the tested concentrations. The total polysaccharide content of callus and organs from the plant was detected by phenol-concentrated sulfuric acid method mentioned above (Fig. 4B) (Supplementary Table S1).

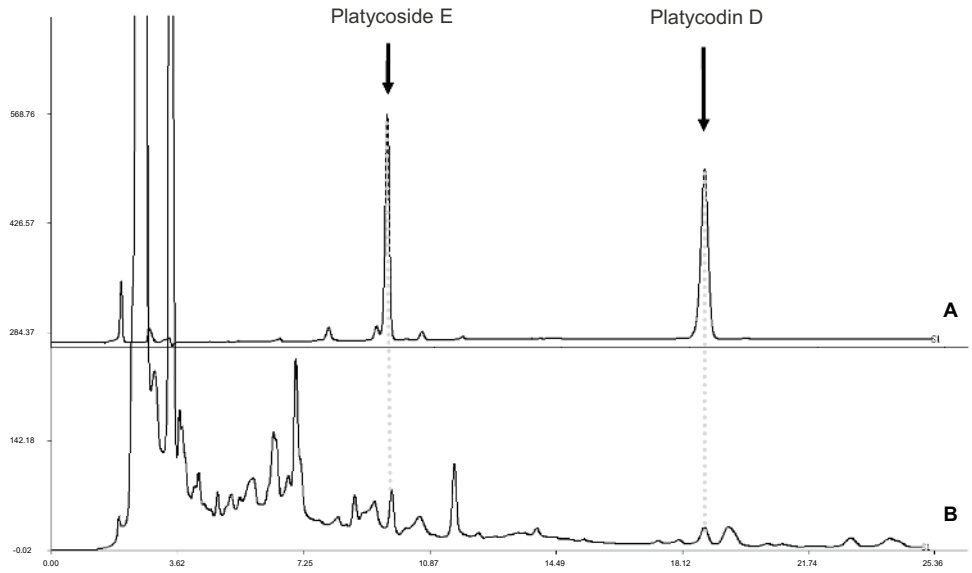

Figure 2. HPLC peak profiles of (**A**) standard products of PD and PE, and of (**B**) the *P. grandiflorus* extract. The X axis represents the retention time (min) of peaks, and the Y axis represents the height of the peak (mAU).

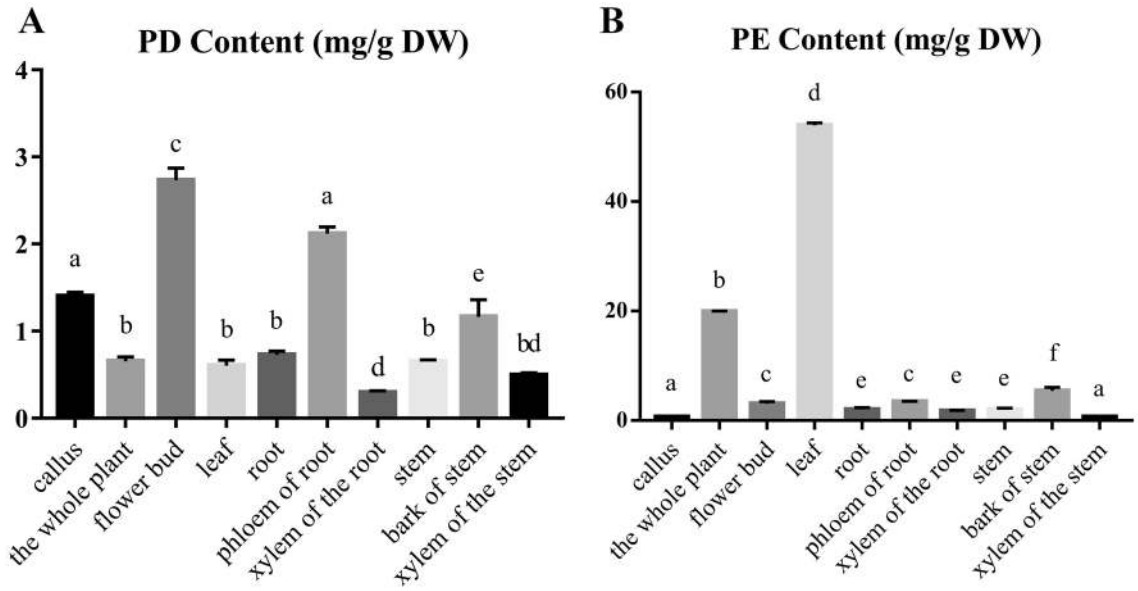

Figure 3. The contents of main triterpenoid saponins in callus and plant organs of *P. grandiflorus*. (**A**) Platycodin D (PD) content in different samples. (**B**) Platycoside E (PE) content in different samples. The X axis is the sample name and the Y axis is the content (mg/g DW), DW is dried weight. Bars with different letters within same histogram, represent significant difference at $p \leq 0.01$.

As for the distribution of polysaccharides in the *P. grandiflorus* also showed significant difference among samples by ANOVA, the content of polysaccharides of root is much higher than that of the whole plant which is consistent with the use of platycodi radix as medicinal part, which indicates that the roots of *P. grandiflorus* are an important place for the storage of polysaccharides. The results showed that the content of polysaccharides of the calli induced by this optimal medium is higher by comparing to the whole plants and organs in *P. grandiflorus*, implying callus might be a high-efficiency material and an important alternative for polysaccharides production than *P. grandiflorus* plants. At the same time, we also found that the polysaccharides content of root phloem was higher than that of the root xylem. This indicates that the root phloem is a more important place for storing polysaccharides. Moreover, we found that the content of polysaccharides of calli is much higher than that of leaves, which is in stark contrast to the PE content of callus and leaves.

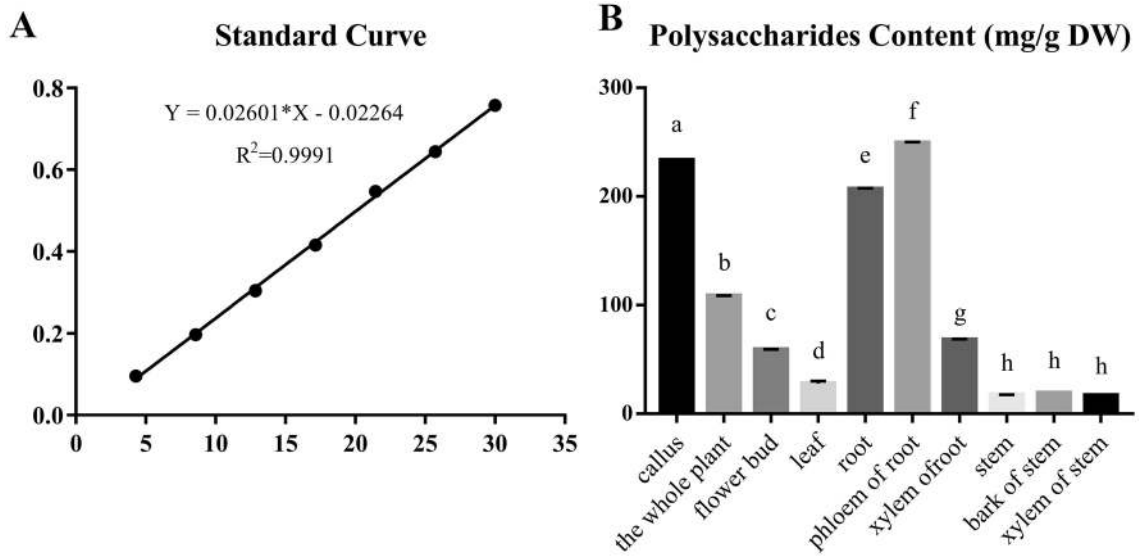

Figure 4. The contents of total polysaccharide among callus and different organ samples. (**A**) Polysaccharides content in different samples. The X axis is the sample name and the Y axis is the content (mg/g DW), and DW means dried weight. Bars with different letters within same histogram, represent significant difference at $p \leq 0.01$. (**B**) Standard curve, $R^2 = 0.9991$. The X axis is the standard concentration (μg/mL), and the Y axis is the absorbance value.

Sequencing and De novo assembly. A total of 39.44 Gb data were generated using the BGISEQ-500 platform by sequencing the established six libraries (leaf and callus each has three duplicates). The raw data were filtered by trimmomatic software (version 0.36, parameters are illuminaclip:2:30:10, leading:3, trailing:3, slidingwindow:4:15 minlen:50) to obtain clean reads by removing the reads containing connector, or unknown bases content more than 5%, or reads with low quality. The clean read numbers of each library were counted by SOAPnuke (version 1.4.0, parameters are -l 5 -q 0.5 -n 0.1). Trinity software was used to assemble clean reads, and then Tgicl software was used to cluster the transcripts to remove redundancy and obtain unigenes. Finally, 152,777 unigenes were obtained, which is higher than those (34,053 unigenes) obtained by Chunhua Ma et al[16]. The total length, average length, N50, and GC content were 228,154,936 bp, 1,493 bp, 2,514 bp, and 40.58%, respectively.

A single-copy ortholog database, BUSCO (https://busco.ezlab.org/) (Supplementary Figure S3), was used to evaluate the quality of the assembled transcripts. By comparing with conserved genes, the results showed the good integrity of the transcriptome assembly. A total of 80,826 coding sequences (N50 = 1,380) were identified, with a maximum length of 16,398 and a minimum length of 297 by TransDecoder software.

Unigenes functional annotation and expression analysis of unigenes. Unigenes were compared to the seven major functional databases to annotate. There were 97,878 (NR: 64.07%), 82,833 (NT: 54.22%), 73,371 (SwissProt: 48.02%), 77,858 (KOG: 50.96%), 78,197 (KEGG: 51.18%), 73,988 (GO: 48.43%), and 73,302 (Pfam: 47.98%) unigenes received functional annotations. 80,826 CDS were detected using Transdecoder software. At the same time, 77,478 SSRs were distributed among 50,171 unigenes, and 3,153 unigenes encoding transcription factors were predicted (Supplementary Table S2: X1, X2 and X3 are the three independent replicates of leaves, X4, X5 and X6 are the three independent replicates of callus). The sequencing and analyzing data had been stored in website: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE153777.

Identification of candidate genes involved in triterpenoid saponin biosynthesis by expression level analysis. KEGG analysis was applied to gain insight into pathways of unigenes. A total of 22,842 unigenes are annotated in the KEGG database[20]. In order to discover the most important biological pathways, the KEGG metabolic pathways involved in genes are divided into 5 branches: cellular processes, environmental information processing, genetic information processing, metabolism, and organismal systems (Supplementary Table S2), including 19 subcategories (135 routes). Eight pathways (ko00906, ko00908, ko00909, ko00904, ko00903, ko00902, ko00900 and ko00905) of "Metabolism of terpenoids and polyketides" containing 864 unigenes were analyzed. These genes encoding enzymes for terpenoid synthesis that are mainly distributed in upstream of MEP and MVP, while some are distributed in downstream (Fig. 1) (Table 2).

Expression level of putative genes encoding enzymes for triterpenoid saponins biosynthesis in *P. grandiflorus* between callus and leaf was measured by value of log2 (FC) of similar unigene from RNA-Seq data, and the similar unigene was obtained by aligning the amino acid sequences between putative genes and the unigene. Expression of genes encoding enzymes in the saponin biosynthetic pathway such as FPPS, HMGR, HMGS, IDI, MVD, MVK, SS, beta-A28O, beta-AS and beta-Glucosidase are up-regulated significantly (Fig. 5A). The expression level of a putative gene predicted to encode a de-glucosidase (ACM59590, β-Glucosidase) is consistent with the variance of PD and PE content, and two extracellular experiments indicate this enzyme may function

Enzyme name	EC number
AACT (acetoacetyl-coenzyme A)	2.3.1.9
HMGS (hydroxymethylglutaryl-CoA synthase)	2.3.3.10
HMGR (3-hydroxy-3-methylglutaryl-coenzyme A reductase)	1.1.1.34
MVK (mevalonate kinase)	2.1.7.36
PMK (phosphomevalonate kinase)	2.7.4.2
MVD (mevalonate diphosphate decarboxylase)	4.1.1.33
GPPS (geranylgeranyl pyrophosphate synthase)	2.5.1.29
FPPS (Farnesyl-diphosphate synthase)	2.5.1.1, 2.5.1.10
SS (squalene synthase)	2.5.1.21
SE (squalene epoxidase)	1.14.14.17
β-AS (beta-amyrin synthase)	5.4.99.39
β-A28O (isolate CYP716A140 beta-amyrin 28-oxidase)	1.14.13.-
LS (lupeol synthase)	5.4.99.41
CAS (cycloartenol synthase)	5.4.99.8
IDI (isopentenyl-diphosphate Delta-isomerase)	5.3.3.2
DXS (1-deoxy-D-xylulose 5-phosphate synthase)	2.2.1.7
DXR (1-deoxy-D-xylulose 5-phosphate reductoisomerase)	1.1.1.267
CMS (2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (4-diphosphocytidyl-2C-methyl-D-erythritol synthase))	2.7.7.60
CMK (4-diphosphocytidyl-2-C-methyl-D-erythritol kinase)	2.7.1.148
MCS (2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase)	4.6.1.12
HDS ((E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase)	1.17.7.1, 1.17.1.3
HDR (4-hydroxy-3-methylbut-2-enyl diphosphate reductase)	1.17.1.4
ACM59590 (β-Glucosidase)	3.2.1.21

Table 2. Putative key enzymes involved in the triterpenoid saponins biosynthesis pathway in *P. grandiflorus*.

as the β-Glucosidase, which showed that exogenous β-Glucosidase can catalyze the conversion of PE into PD through removing two glycosyl groups from PE in vitro[18,19]. From expression level analysis, 4 putative unigenes (CL4020.Contig1_All, Unigene 1627_All, CL3189.Contig2_All and Unigene7900_All) have high identity with the β-glucosidase from *Caldicellulosiruptor bescii* with amino acid sequence[18]. The result of real-time quantitative PCR indicated that the expression patterns of putative gene sequences encoding enzyme involved in converting PE to PD is consistent with the result from transcriptome analysis (Fig. 5B). Furthermore, to better select the putative sequences which might encode enzyme involved in converting PE to PD, the conserved motifs were analyzed using MEME (Fig. 5C). The result showed 3 unigenes (CL4020.Contig1_All, Unigene 1627_All and Unigene7900_All) were screened as candidate fragments of the target gene.

Gene expression variance involved in polysaccharide biosynthesis. Furthermore, fifteen pathways (ko00010, ko00500, ko00620, ko00020, ko00051, ko00562, ko00052, ko00630, ko00030, ko00040, ko00053, ko00520, ko00640, ko00650 and ko00660) of "Carbohydrate metabolism" were analyzed, including 4,441 unigenes. There are 1,148 unigenes involved in starch and sucrose metabolism, and 765 unigenes involved in amino sugar and nucleotide sugar metabolism. Based on KEGG pathway analysis, we developed a model to infer the biosynthetic pathway of polysaccharides in *P. grandiflorus* (Supplementary Figure S4) (Table 3).

The value of log2 (FC) was also used to analyze the expression of genes encoding enzymes for polysaccharide biosynthesis, and it showed that the expression levels of putative genes encoding AXS, GALE, GMDS, GPI, HK, PMM, SUS, TSTA3, UGDH, USP and UXE in the polysaccharide biosynthetic pathway were also upregulated significantly (Fig. 6). The result of gene expression quantity is consistent with the fact that the content of total polysaccharides in calli is higher than that of leaves.

Detection of transcription factors. Transcription factors (TFs) can temporarily and spatially regulate the activity of target genes, and play a key role in plant development and response to the external environment[36–38]. A total of 3,153 candidate TFs were identified in the *P. grandiflorus* transcriptome database and classified into 57 different TF families (Fig. 7A), among these families that MYB family (326 unigenes) accounted for the largest proportion of TF families, followed by mTERF (286 unigenes), bHLH (200 unigenes), AP2-EREBP (185 unigenes), C3H (160 unigenes), ABI3VP1 (152 unigenes), NAC (132 unigenes), WRKY (130 unigenes), and FAR1 (118 unigenes). In these target TFs 1,542 were up-regulated and 1,444 were down-regulated in the calli versus leaves (Fig. 7B). We found that 54 candidate genes among transcription factors are expressed only in leaves, and 23 are expressed only in calli (Fig. 7C) (Supplementary Table S3).

By KEGG pathway analysis it can be concluded that total 11 (4 Trihelix and 7 FHA) and 15 (3 bHLH, 6 C2H2, 3 Trihelix, 1 G2-like, 1 FHA and 1 VARL) TFs may relate to the biosynthesis of saponins and polysaccharides,

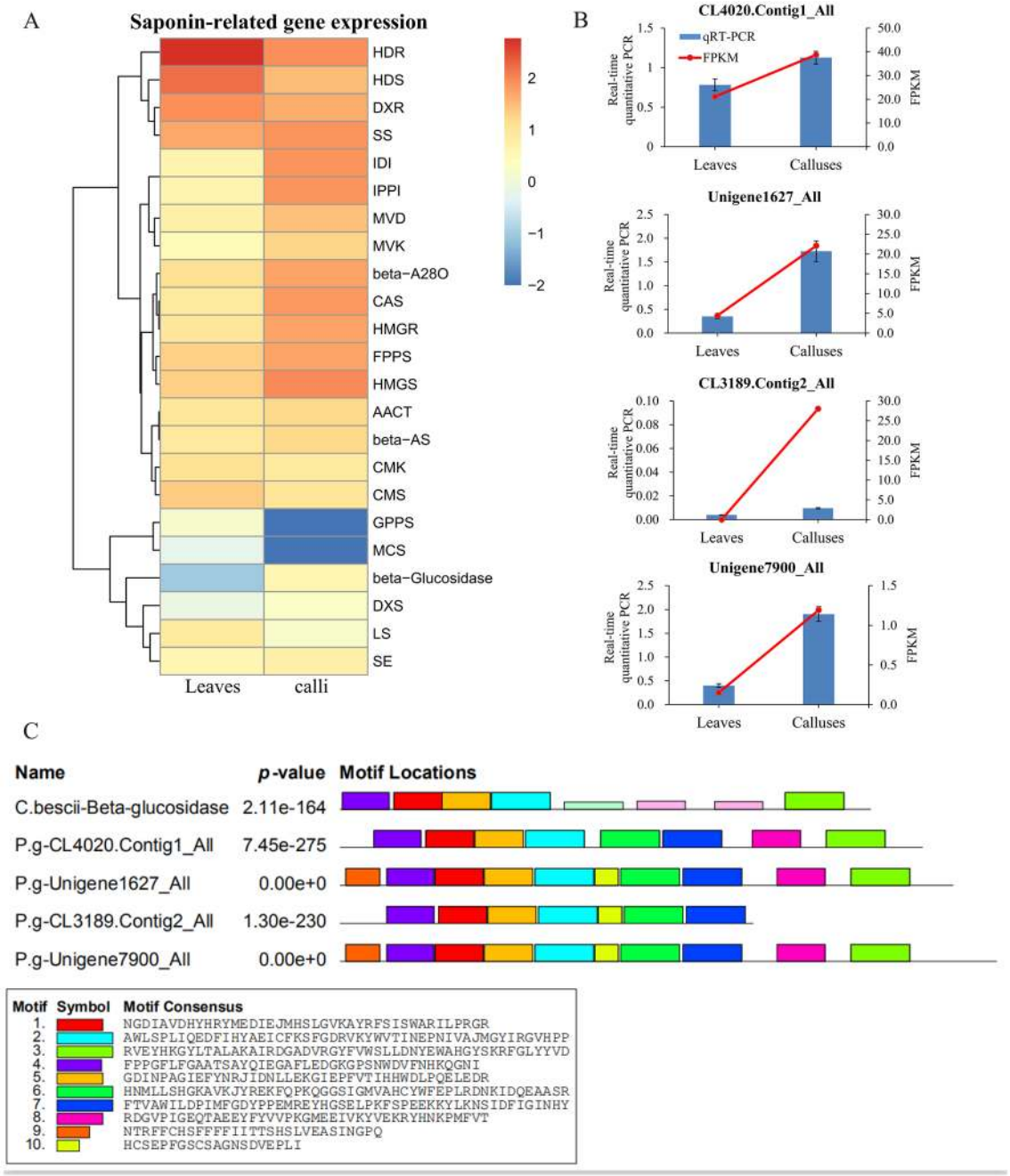

Figure 5. Expression level analysis of genes from triterpenoid saponins biosynthetic pathway in *P. grandiflorus* and conserve motif analysis of predictive β-glucosidase. (**A**) The expression levels of a single gene encoding an enzyme from each step of triterpenoid saponins biosynthetic pathway are shown. Red and green represent high and low expression levels, respectively. (**B**) Real-time quantitative PCR analysis of CL4020.Contig1_All, Unigene 1627_All, CL3189.Contig2_All and Unigene7900_All in leaves and calli. 18S rRNA as internal reference gene. Error bars indicate SD (n = 3). The blue bars represent the real-time quantitative PCR data and the red line represents the FPKM value from RNA-Seq. (**C**) Comparison of the conserve motifs among β-glucosidase from *C. bescii* and the amino acid sequences of putative orthologous unigenes from *P. grandiflorus*. Each block shows the position and strength of a motif site. The height of a block gives an indication of the significance of the site as taller blocks are more significant. The height is calculated to be proportional to the negative logarithm of the p-value of the site, truncated at the height for a p-value of 1e−10.

respectively (Supplementary Table S4). It also found a Trihelix TF gene (Unigene26781_All) is co-expressed with the putative de-glucosylation enzyme gene (β-Glucosidase).

Enzyme name	EC number
SUS (sucrose synthase)	2.4.1.13
INV (sucrose invertase)	3.2.1.26
HK (Hexokinase)	2.7.1.1
PGM (Phosphoglucomutase)	5.4.2.2
USP (UDP-sugar pyrophosphorylase)	2.7.7.64
UGP2 (UTP-glucose-1-phosphate Uridylyltransferase)	2.7.7.9
scrK (Fructokinase)	2.7.1.4
MPI (Mannose-6-phosphate isomerase)	5.3.1.8
UGDH (UDP-glucose 6-dehydrogenase)	1.1.1.22
UXS1 (UDP- glucuronate decarboxylase)	4.1.1.35
UXE (UDP-arabinose 4-epimerase)	5.1.3.5
RHM (UDP-glucose 4,6-dehydratase)	4.2.1.76
UER1(3,5-Epimerase-4-reductase)	5.1.3.-, 1.1.1.-
PMM (Phosphomannomutase)	5.4.2.8
GMPP (Mannose-1-phosphate Guanylyltransferase)	2.7.7.13
GMDS (GDP-mannose 4,6-dehydratase)	4.2.1.47
TSTA3 (GDP-$_L$-fucose synthase)	1.1.1.271
GPI (Glucose-6-phosphate isomerase)	5.3.1.9
GALE (UDP-glucose 4-epimerase)	5.1.3.2
UGE (UDP-glucuronate 4-epimerase)	5.1.3.6
AXS (UDP-apiose/xylose synthase)	AXS
Amylosucrase (1,4-α-D-glucan 4-α-D-glucosyltransferase-glucan)	2.4.1.14

Table 3. Putative key enzymes involved in polysaccharide biosynthesis in *P. grandiflorus*.

Discussion

Plant tissue culture is not only a method for plant rapid propagation, but also an ideal way for plant improvement, germplasm preservation and production of useful compounds. Studies have shown that compounds can be quickly obtained from callus of *Citrus junos* Siebold ex. Tanaka[39] and *Ocimum basilicum* L.[18,19]. It was reported that there was high content of pentacyclic triterpenoids with anti-inflammatory and antinociceptive activities in callus of *Chaenomeles japonica* (Thunb.) Lindl. ex Spach[40]. So far, there's no specific protocol for the callus induction of *P. grandiflorus*. The best callus induction plan of *P. grandiflorus* was screened out through experiments in this article, which filled the research gap, and provided a basis for the development and utilization of *P. grandiflorus* callus.

The contents of PE and PD in the phloem of *P. grandiflorus* is higher than that in the xylem like some other secondary compounds[41,42], indicating that the phloem of *P. grandiflorus* may have the function to transport and store PD and PE. In addition, we found that the content of PE is the highest in leaves, while that in callus was lower than that in other organs. So, obtaining PE from leaves of *P. grandiflorus* is probably the best choice.

Studies have shown that PE converted to PD by removing two molecules of glycosyl group under the action of deglycosylated enzymes[18,19]. Since callus has much higher content of polysaccharides and PD, and lower content of PE than leaves, then we highly speculate that the active glycosyltransferase in the callus will catalyze the conversion of PE into PD, and the glucose groups were released to participate in the biosynthesis of polysaccharides. This process has a positive effect on the accumulation of polysaccharides in callus.

The number of unigenes is 152,777 of us versus 34,053 of Ma's[16]. The average length and N50 values of genes in this study are also higher than those in Ma's results (with the total average length of 936 bp and N50 of 1,661 bp)[16], and the differences are mainly due to the different materials and sequencing platform used for RNA-Seq, and different sequencing depth. It was reported that endophytic bacteria are involved in secondary metabolite biosynthesis, which could be isolated from the interior of *P. grandiflorum*[43], which means there might be genes in the endophytic bacteria encoding key enzymes catalyzing the formation of secondary metabolites[44]. So, the calli and leaves of *P. grandiflorum* are used as materials for RNA-Seq to avoid the residual RNA from endophytes. Transcriptome analysis were quite important methods to identify new genes in triterpenoid saponin biosynthetic pathway in *P. grandiflorum*. Besides, the difference in plant materials between Ma[16] et al. and us. Many key enzymes involved in triterpenoid saponin biosynthesis were discovered in both studies, including HMGS, HMGR, FPPS, SS, SE, et al. However, UGTs cloned in Ma's study catalyze the glucosylation; in our work, 3 unigenes or orthologous sequences (CL4020.Contig1_All, Unigene 1627_All and Unigene7900_All) were screened as candidate gene which can catalyze degradation of glycosyl group and convert PE to PD. Our research could supply more transcriptome data, and it is the first time to identify the candidate genes which that converts PE to PD in *P. grandifloras*.

Previous studies have shown TFs, such as the bHLH transcription factor AabHLH1 and AaMYC2 in *Artemisia annua* L., have effects on the primary and secondary metabolites of plants, that can effectively regulate the biosynthesis of artemisinin[45,46], and the study has shown that the Trihelix family transcription factor BdTHX1

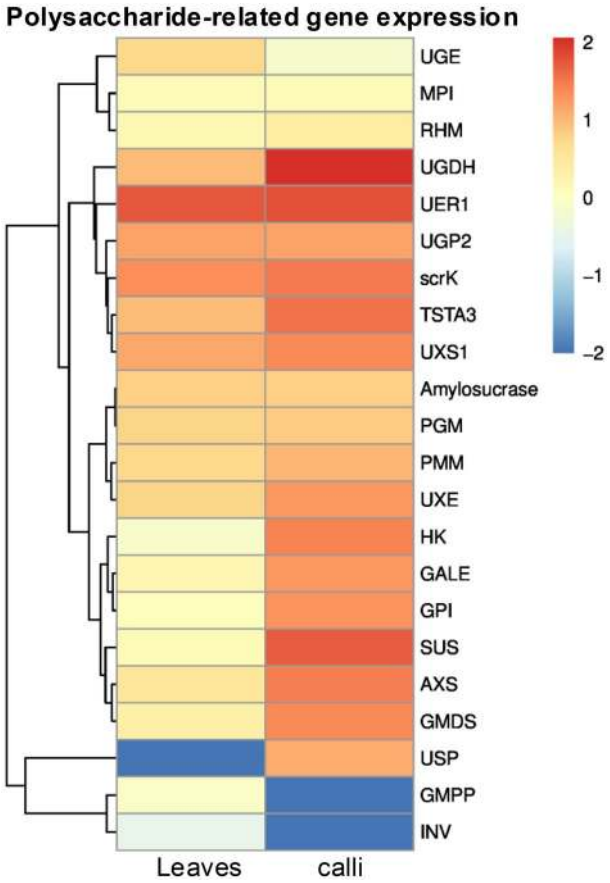

Figure 6. The expression levels of a single gene encoding an enzyme from each step of polysaccharides biosynthetic pathway in *P. grandiflorus* are shown. Red and green represent high and low expression levels, respectively.

likely plays an important role in mixed-linkage glucan biosynthesis and restructuring by regulating the expression polysaccharide synthase related genes[47]. It was reported that Trihelix TF, responding to light signals, regulates the expression of downstream gene by calcium-dependent phosphorylation and dephosphorylation[48]. The expression level of a Trihelix TF gene (Unigene26781_All) is co-expressed with the putative deglucosylation enzyme gene (β-Glucosidase) in this study, which demonstrated that this Trihelix TF might sense the light signal, and influence the expression of the putative gene encoding β-Glucosidase, then modify and regulate the conversion of PE to PD. The discoveries of these TFs may have great potential value and broad application prospects in studying the regulation of terpenoid saponin and polysaccharide bio-synthesis in *P. grandiflorus*.

Conclusions

A best protocol for calli induction of *P. grandiflorus* is that use the stem with nodes as explants, and take MS + NAA 1.0 mg/L + 6-BA 0.5 mg/L as the formula. We found that the content of PD and polysaccharide in callus is higher than that in the plant of *P. grandiflorus*, as well higher than that in the root which is traditional medicinal parts. Study also showed that PD content of calli is higher than that of leaves, which is in sharp contrast to PE content.

We performed comprehensive RNA-Seq analysis on the leaves and calli of *P. grandiflorus*, and were able to find the expression of many genes involved in triterpenoid saponins and polysaccharides which were co-expression with the content of corresponding metabolites. Three putative unigenes with high amino acid sequence identity were screened as orthologous sequences of candidate β-Glucosidase gene converting PE to PD, which is helpful to deeply understand the biosynthesis mechanism of triterpene saponins in *P. grandiflorus* at the molecular level. A total of 11 TFs were involved in regulating the biosynthesis of saponins, and 15 TFs were involved in carbohydrate metabolism were obtained.

In summary, our work may greatly help to promote the molecular biology research and improve the large-scale production of triterpene saponins and polysaccharides in *P. grandiflorus*.

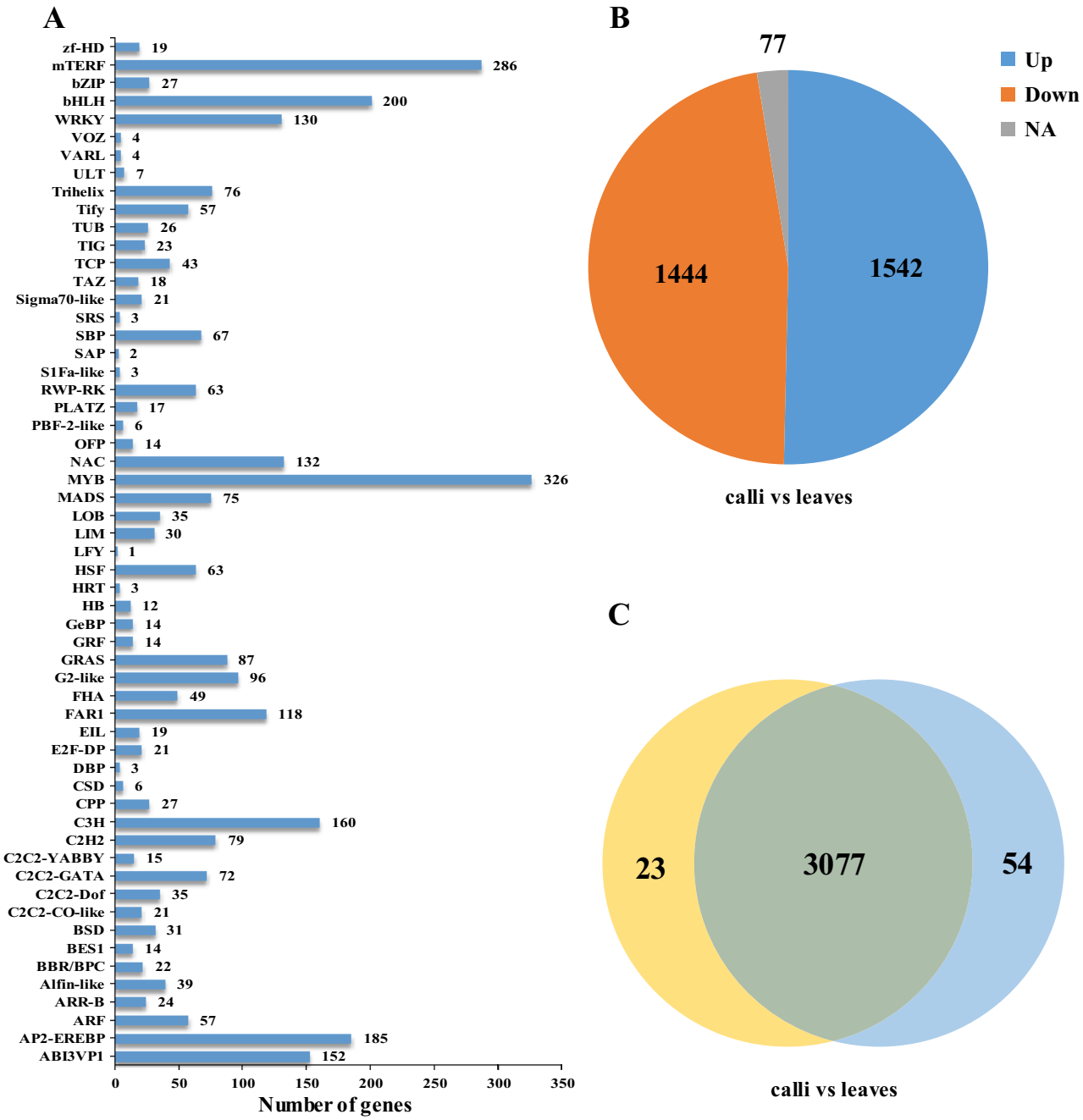

Figure 7. Transcription factors (TFs) expression analysis. (**A**) Classification of gene transcription factors' family. (**B**) The expression level of TFs gene is different in calli vs. leaves, NA stands for expression only in calli or leaves. (**C**) Venn diagram of TFs gene expression in leaves and calli.

Methods

Plant materials. Plants used in this experiment were identified as *Platycodon grandiflora* (Jacq.) A. DC.by expert who is major in plant identification, which were grown at 25 °C during day and 23 °C at night in the green house of Anhui University of Chinese Medicine, Anhui, China. The seedlings of *P. grandiflorus* were purchased from Bozhou Market of Traditional Chinese Medicine, Anhui Province, China. Specimen of *P. grandiflorus* in this study was deposited in the Herbarium of Anhui University of Chinese Medicine and the depository No. is 20,200,705. The explants used for callus induction and the materials used for metabolites determination and RNA-Seq are all from the same plantlets. Each sample in this study has three independent biological duplicates.

Chemicals. Standard compounds of platycodin D (PD, > 98% purity) and platycodin E (PE, > 98% purity) were purchased from Chengdu Desite Biotech Co., Ltd. and Chengdu Push Bio-technology CO., Ltd., respectively. CTAB-PBIOZOL reagent was purchased from Bioflux. Acetonitrile and Methanol (HPLC grade) were purchased from Oceanpak. HPLC-grade water was prepared using laboratory water purification system from

Pall Filter Co., Ltd. (Beijing, China). Plant growth regulators of 6-BA (6-Benzylaminopurine) and α-NAA (α-Naphthylacetic acid) were purchased from Beijing Solarbio Technology Co., Ltd. Chemicals for plant tissue culture were bought from Sinopharm group, and all other chemicals were of analytical grade.

Calli induction of *Platycodon grandifloras*. In order to screen the optimal processes of inducing calli of *P. grandiflorus* included optimum explants, basic medium and plant growth regulators combination, a $L_9(3^4)$ orthogonal experiment was carried out based on the documents and our previous study regardless of the interactions among factors (Supplementary Table S5). The best one was selected from basic media (Factor A) of B5, MS and WPM for *P. grandiflorus* calli induction. Leaves, stems with nodes and stems without nodes were used as explants (B), which are in the same batch with the materials used for content detection and RNA-Seq analysis. In order to select the optimal combination for calli induction of *P. grandiflorus*, plant growth regulator combinations with different concentrations of auxin and cytokinin were designed according to relevant literature. NAA (C) and 6-BA (D) were divided into three concentrations of 0.1 mg/L, 0.2 mg/L and 1.0 mg/L and 0.5 mg/L, 1.0 mg/L and 2.0 mg/L, respectively.

Basic medium B5 (Gamborg B5 Medium), MS (Murashige and Skoog Medium), WPM (Lloyd & McCown Woody Plant Basal Medium)[49] stock solutions were prepared. The full media formulas were prepared according to the different proportions of components, and different concentrations of plant growth regulators. Finally, adjusted the pH of media to 5.6 ~ 6.0, followed by the addition of sucrose (30 g/L) and agar (7 g/L), then the media was sterilized by autoclaving at 121 °C for 30 min.

Explants of *P. grandiflorus* were rinsed and surface-sterilized, and rinsed 3–5 times with sterile distilled water. The sterile explants were dried, leaves were cut into 0.5 cm^2 in size, and the stems were cut into pieces about 1 cm in length, then they were inoculated onto the media. Each group of explants were inoculated in 20 bottles with 3 pieces per bottle. The culture conditions were as follows: the photosynthetic photon flux density was 30 μmol/m^2/s for 12 h/d, the culture temperature was (25 ± 1) °C, and the culture lasted for 50 d.

Extraction of triterpenoid saponin D and E. Jaeyoung Kwon et al.[2] found that PD degraded in the minimum at 40 °C in drying process, and suitable for detection. In this paper, all the organs of wild plants and tissue culture materials of *P. grandiflorus* were dried at 40 °C to constant weight and pulverized by a ball mill. Then accurately weighed the sample powders (each sample 0.5 g) and transferred them to a 50 mL of centrifuge tubes. After adding of 25 mL methanol, each sample solution was adjusted to the same weight and recorded the final weight followed by ultrasonic extraction lasting for 50 min. And the loss of solution occurring during ultrasonic extraction was compensated for by adding a certain amount of 100% methanol to the same weight. Finally, 20 mL of continuous filtrated to 2 mL for further use.

HPLC quantitative analysis of saponin D and E from calli and organs of plants. All analyses of PD and PE contents were performed on an Agilent Series 1260 system (Agilent Technologies, Germany). Topsil C18 HPLC column (4.6 mm × 250 mm, 5 μm particle size) was used for chromatography. Elution was carried out using (A) water and (B) acetonitrile as a mobile phase. The ratio of A to B is 71: 29. The flow rate was 1.0 mL/min, the sample injection volume was 5 μL, and the column temperature was 30 °C. The Detection wavelength was 210 nm.

Extraction and determination of total polysaccharide content. Polysaccharides extraction from *P. grandiflorus* and determination were performed by the methods reported previously with a few modifications[50]. All the samples were dried at 40 °C to constant weight and milled to powder. Accurately weighed samples (0.5 g each sample), and added 50 mL distilled water to mix evenly. Subsequently, weighed the mixture solution and allowed it to stand for 30 min, and then refluxed and extracted in a boiling water bath for 1 h. When the solution was cold, the loss of the solutions was compensated by adding distilled water, then shook and filtered it. Precisely took 10 mL of the filtrate and evaporated it, added 2 mL distilled water to dissolve the evaporated sample and then mixed the solution with 10 mL of absolute ethanol to dissolve for 24 h. After centrifugation at 4000 g for 10 min, removed the supernatant and washed the precipitate 3 times with 85% ethanol. Then the precipitate was dissolved in 50 mL of water and shook intermittently and solution for test was obtained. 1 mL of each raw material was used to determine the total polysaccharide content by the improved sulfuric acid phenol method.

The standard glucose solution (0.3 mg/mL) was prepared as follows: glucose of 7.5 mg in chromatographic grade was dissolved in distilled water and diluted to a constant volume of 25 mL. Gradient solutions were prepared by transferring 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7 mL of standard glucose solution to the test tubes and added distilled water to a constant volume of 1 mL. Phenol sulfate method was used to determine and draw the standard curve, and the absorbance were measured at 486 nm using spectrophotometer.

RNA extraction. Total RNA was purified by CTAB-PBIOZOL reagent as extraction solution from leaves and calli of *P. grandiflorus*. Around 80 mg samples were ground into powder in liquid nitrogen and transferred to tubes containing 1.5 mL of preheated 65 °C CTAB-pBIOZOL reagents. The samples were incubated by Solarbio mixer for 15 min at 65 °C to make the nucleoprotein complexes in the samples completely dissolved. After centrifugation at 12,000 g at 4 °C for 5 min, the supernatants were mixed with 400 μL of chloroform per 1.5 mL of extraction solution and centrifuged at 12,000 g for 10 min at 4 °C. After centrifuge, the supernatants were transferred to a new 2.0 mL test tubes containing 700 μL acidic phenol and 200 uL chloroform followed by centrifuging at 12,000 g for 10 min at 4 °C. Subsequently, the aqueous phase of each sample was mixed with an equal volume of chloroform before centrifuging at 12,000 g for 10 min at 4 °C. The supernatants were mixed with an equal volume of isopropyl alcohol and placed in − 20 °C for 2 h to precipitate, then centrifuged at 12,000 g for

20 min at 4 °C, and the supernatants were removed. After washing twice with 1 mL of 75% ethanol, they were dried in the bio-safety cabinet, and dissolved with 50 µL of sterilized DEPC-treated water. Finally, total RNA of each samples was qualified and quantified by a Nano Drop and Agilent 2100 bioanalyzer (Thermo Fisher Scientific, MA, USA).

Construction of cDNA and RNA-Seq libraries and sequencing. Oligo(dT) attached magnetic beads were used to purified mRNA from leaves and calli of *P. grandiflorus* in three replicates by eliminating rRNA and tRNA in the total RNA. Purified mRNA was fragmented into small pieces with fragment buffer at appropriate temperature. Then, First-strand cDNA was generated by random hexamer-primed reverse transcription, followed by a second-strand cDNA synthesis. After that, RNA Index Adapters and A-Tailing Mix were added by incubating to end repair. The fragments of cDNA obtained from previous steps were amplified by PCR, and the PCR products were purified by Ampure XP Beads, and dissolved in EB solution. The products were validated on the Agilent Technologies 2100 bioanalyzer for quality control. The double stranded PCR products obtained from previous steps were denatured by heating and circularized by the splint oligo sequence to get the final library, and the single stranded circular DNA (ssCir DNA) was formatted as the final library.

The constructional final six libraries were further amplified with phi29 to make DNA nanoball (DNB), each molecular of which had more than 300 copies, DNBs were loaded into the patterned nano-array and paired-end of 150 bp base reads were generated on BGIseq 500 platform (BGI-Shenzhen, China).

De novo transcriptome assembly and unigenes annotation. After sequencing, raw data were obtained, and low-quality, joint contamination and unknown base N were filtered by software SOAPnuke (version 1.4.0) to generate clean data. The filtered data which is called clean reads were de novo transcriptome assembled using Trinity software (version 2.0.6). The acquired full-length transcripts for alternatively splicing isoforms were obtained by splicing transcripts corresponding to paralogous genes, then the redundant transcripts were removed by TGICL (version 2.06, parameters: – 1 30 –v 35) to acquire non-redundant sequences which called unigenes. TransDecoder software (Version 3.0.1, parameters: default) was used to identify candidate coding regions in unigene[51].

The assembled unigenes were subjected to databases as KEGG (Kyoto Encyclopedia of Genes and Genomes), NT (NCBI non-redundant nucleotide sequence), NR (NCBI non-redundant protein sequences), SwissProt (a manually annotated and reviewed protein sequence database) and KOG (Clusters of Eukaryotic Orthologous Groups) by Blast software (version 2.2.23)[52] with default parameters (under a threshold E-value ≤ 10–5) to get the functional annotations. Ultimately, GO (Gene Ontology) annotations and functional classifications were obtained using Blast2GO program (version 2.5.0, E-value ≤ 10–6)[53] based on NR annotations.

Identification of differentially expressed genes (DEGs). All clean reads from all samples were aligned to unigene sets using Bowtie2 (V2.2.5) with default settings[54]. The expression level of genes were calculated by RSEM (v1.2.8) [55] and normalized to fragments per kilobase of transcript per million (FPKM). DEseq2 (v1.4.5)[56] was used to detect DEGs (Different expressed genes) with Q value (adjust P value) < 0.001, Unigenes showing differential expression between two issue types (leaf vs callus) at fold change ≥ 2 or ≤ -2, and a false discovery rate (FDR) ≤ 0.001 was identified as DEGs using the Posisson distribution method[57]. The identified DEGs were subsequently carried into GO and KEGG enrichment with Phyper in R package by Q value ≤ 0.05 as default. The heatmap was drawn by pheatmap (v1.0.8)[56] according to the gene expression in different samples.

Analysis of genes involved in saponin and polysaccharide biosynthesis.. According to the KEGG annotation results and official classification, we classified the differential genes in biological pathways, and used the phyper function in R software to perform enrichment analysis, calculated the p-value, and then performed FDR correction on the p-value. Generally, functions with Q-value ≤ 0.05 are considered significant enrichment (https://en.wikipedia.org/wiki/Hypergeometric_distribution). The logarithm of normalization of average FPKM and 0.01 (avoiding the value of FPKM is zero) was used to measure the expression level of each gene in leaf and callus. The comparison of conserve motif between putative candidate gene and the target gene was performed based on the results of MEME (http://meme-suite.org/tools/meme) analysis.

Real-time quantitative PCR analysis. In order to verify RNA-Seq data, real-time quantitative PCR analysis was performed. Collected leaves and calli samples (each in three independent biological duplicates), which is same to RNA-Seq, and promptly frozen in liquid nitrogen and stored at − 80 °C. Total RNA was extracted according to the previous RNA extraction method. 0.5 µg of RNA was used to synthesize the cDNA using FastKing RT Kit (Tiangen, China). Real-time quantitative PCR analysis was conducted by using SuperReal PreMix Plus (Tiangen, China) on the LightCycle480 machine (Roche, Switzerland). The primer sequences of genes were designed by Primer Premier (version 5.0) (Supplementary Table S6). Housekeeping gene 18sRNA of *P. grandiflorus* was used as internal reference gene. The data from real-time quantitative PCR was statistical analyzed by the $2^{-\Delta\Delta Ct}$ approach.

Analysis of transcription factors. ORFs (open reading frames) of unigenes were mapped to TF protein domain in PlnTFDB (plant TF database) based on BlastTX (E-value ≤ 1e−5) using Hmmsearch method[53]. High-expressing IFs corresponding to the saponin and polysaccharide were selected by comparing the value of log2(FC) of each IF in samples.

Statistical analysis. All experiments in this study were performed with three independent biological duplicates including RNA-Seq and HPLC analysis, and the data are expressed as the mean ± standard deviation (SD) of three duplicates. The tissue culture data were analyzed by orthogonal experimental analysis, and the statistical differences of saponin D, E and polysaccharide contents among different samples, and the significance of difference from different factors on calli induction were analyzed using analysis of one-way and two-way variance (ANOVA) by SPSS software package (version 17.0), respectively. Values at $P \leq 0.01$ were considered statistically significant difference.

Data availability

The datasets generated and analyzed during the current study are available in the [GEO] repository, [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE153777].

Received: 28 October 2020; Accepted: 23 April 2021
Published online: 07 May 2021

References

1. Wu, J. T. *et al.* Anti-atherosclerotic activity of platycodin D derived from roots of *platycodon grandiflorum* in human endothelial cells. *Biol. Pharm. Bull.* **35**, 1216–1221. https://doi.org/10.1248/bpb.b-y110129 (2012).
2. Kwon, J. *et al.* Effect of processing method on platycodin D content in Platycodon grandiflorum roots. *Arch. Pharm. Res.* **40**, 1087–1093. https://doi.org/10.1007/s12272-017-0946-6 (2017).
3. Zhang, L. *et al.* Platycodon grandiflorus—An ethnopharmacological, phytochemical and pharmacological review. *J. Ethnopharmacol.* **164**, 147–161. https://doi.org/10.1016/j.jep.2015.01.052 (2015).
4. Xie, Y., Sun, H. X. & Li, D. Platycodin D is a potent adjuvant of specific cellular and humoral immune responses against recombinant hepatitis B antigen. *Vaccine* **27**, 757–764. https://doi.org/10.1016/j.vaccine.2008.11.029 (2009).
5. Ahn, K. S. *et al.* Inhibition of inducible nitric oxide synthase and cyclooxygenase II by *Platycodon grandiflorum* saponins via suppression of nuclear factor-kappaB activation in RAW 2647 cells. *Life Sci.* **76**, 2315–2328. https://doi.org/10.1016/j.lfs.2004.10.042 (2005).
6. Lee, E. J., Kang, M. & Kim, Y. S. Platycodin D inhibits lipogenesis through AMPKalpha-PPARgamma2 in 3T3-L1 cells and modulates fat accumulation in obese mice. *Planta Med.* **78**, 1536–1542. https://doi.org/10.1055/s-0032-1315147 (2012).
7. Luan, X. *et al.* Platycodin D inhibits tumor growth by antiangiogenic activity via blocking VEGFR2-mediated signaling pathway. *Toxicol. Appl. Pharmacol.* **281**, 118–124. https://doi.org/10.1016/j.taap.2014.09.009 (2014).
8. Li, W. *et al.* Platycoside N: A new oleanane-type triterpenoid saponin from the roots of Platycodon grandiflorum. *Molecules (Basel, Switzerland).* **15**, 8702–8708. https://doi.org/10.3390/molecules15128702 (2010).
9. Sheng, Y., Liu, G., Wang, M., Lv, Z. & Du, P. A selenium polysaccharide from *Platycodon grandiflorum* rescues PC12 cell death caused by H(2)O(2) via inhibiting oxidative stress. *Int. J. Biol. Macromol.* **104**, 393–399. https://doi.org/10.1016/j.ijbiomac.2017.06.052 (2017).
10. Zheng, P. *et al.* Characterization of polysaccharides extracted from *Platycodon grandiflorus* (Jacq.) A.DC. affecting activation of chicken peritoneal macrophages. *Int. J. Biol. Macromol.* **96**, 775–785. https://doi.org/10.1016/j.ijbiomac.2016.12.077 (2017).
11. Haralampidis, K., Trojanowska, M. & Osbourn, A. E. Biosynthesis of triterpenoid saponins in plants. *Adv. Biochem. Eng. Biotechnol.* **75**, 31–49. https://doi.org/10.1007/3-540-44604-4_2 (2002).
12. Kim, Y. K. *et al.* Enhanced accumulation of phytosterol and triterpene in hairy root cultures of *Platycodon grandiflorum* by overexpression of Panax ginseng 3-hydroxy-3-methylglutaryl-coenzyme A reductase. *J. Agric. Food Chem.* **61**, 1928–1934. https://doi.org/10.1021/jf304911t (2013).
13. Zhao, C. L., Cui, X. M., Chen, Y. P. & Liang, Q. Key enzymes of triterpenoid saponin biosynthesis and the induction of their activities and gene expressions in plants. *Nat. Prod. Commun.* **5**, 1147–1158 (2010).
14. Niu, Y. *et al.* Expression profiling of the triterpene saponin biosynthesis genes FPS, SS, SE, and DS in the medicinal plant Panax notoginseng. *Gene* **533**, 295–303. https://doi.org/10.1016/j.gene.2013.09.045 (2014).
15. Tang, Q. Y. *et al.* Transcriptome analysis of *Panax zingiberensis* identifies genes encoding oleanolic acid glucuronosyltransferase involved in the biosynthesis of oleanane-type ginsenosides. *Planta* **249**, 393–406. https://doi.org/10.1007/s00425-018-2995-6 (2019).
16. Ma, C. H. *et al.* Candidate genes involved in the biosynthesis of triterpenoid saponins in *Platycodon grandiflorum* identified by transcriptome analysis. *Front. Plant Sci.* **7**, 673. https://doi.org/10.3389/fpls.2016.00673 (2016).
17. Shin, K. C., Kim, D. W., Woo, H. S., Oh, D. K. & Kim, Y. S. Conversion of glycosylated platycoside E to deapiose-xylosylated platycodin D by cytolase PCL5. *Int. J. Mol. Sci.* https://doi.org/10.3390/ijms21041207 (2020).
18. Kil, T. G., Kang, S. H., Kim, T. H., Shin, K. C. & Oh, D. K. Enzymatic biotransformation of balloon flower root saponins into bioactive platycodin D by deglucosylation with Caldicellulosiruptor bescii β-Glucosidase. *Int. J. Mol. Sci.* https://doi.org/10.3390/ijms20163854 (2019).
19. Ahn, H. J. *et al.* Biocatalysis of platycoside E and platycodin D3 using fungal extracellular beta-glucosidase responsible for rapid platycodin D production. *Int. J. Mol. Sci.* https://doi.org/10.3390/ijms19092671 (2018).
20. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. https://doi.org/10.1093/nar/28.1.27 (2000).
21. Yu, Y., Shen, M., Song, Q. & Xie, J. Biological activities and pharmaceutical applications of polysaccharide from natural resources: A review. *Carbohydr. Polym.* **183**, 91–101. https://doi.org/10.1016/j.carbpol.2017.12.009 (2018).
22. Hong, S. M., Bahn, S. C., Lyu, A., Jung, H. S. & Ahn, J. H. Identification and testing of superior reference genes for a starting pool of transcript normalization in Arabidopsis. *Plant Cell Physiol.* **51**, 1694–1706. https://doi.org/10.1093/pcp/pcq128 (2010).
23. Zhang, G. Q. *et al.* The *Dendrobium catenatum* Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution. *Sci. Rep.* **6**, 19029. https://doi.org/10.1038/srep19029 (2016).
24. Wagner, E., Götz, F. & Brückner, R. Cloning and characterization of the scrA gene encoding the sucrose-specific Enzyme II of the phosphotransferase system from *Staphylococcus xylosus*. *Mol. Gen. Genet. MGG* **241**, 33–41. https://doi.org/10.1007/bf00280198 (1993).
25. Schmid, K., Schupfner, M. & Schmitt, R. Plasmid-mediated uptake and metabolism of sucrose by *Escherichia coli* K-12. *J. Bacteriol.* **151**, 68–76 (1982).
26. Yuan, Y. *et al.* Polysaccharide biosynthetic pathway profiling and putative gene mining of *Dendrobium moniliforme* using RNA-Seq in different tissues. *BMC Plant Biol.* **19**, 521. https://doi.org/10.1186/s12870-019-2138-7 (2019).
27. Bachmann, P. & Zetsche, K. A close temporal and spatial correlation between cell growth, cell wall synthesis and the activity of enzymes of mannan synthesis in *Acetabularia mediterranea*. *Planta* **145**, 331–337. https://doi.org/10.1007/bf00388357 (1979).

28. Yin, Y., Huang, J., Gu, X., Bar-Peled, M. & Xu, Y. Evolution of plant nucleotide-sugar interconversion enzymes. *PLoS ONE* **6**, e27995. https://doi.org/10.1371/journal.pone.0027995 (2011).
29. Liang, D. *et al.* Hydrogen cyanamide induces grape bud endodormancy release through carbohydrate metabolism and plant hormone signaling. *BMC Genom.* **20**, 1034. https://doi.org/10.1186/s12864-019-6368-8 (2019).
30. Wang, C. *et al.* Transcriptome analysis of *Polygonatum cyrtonema* Hua: Identification of genes involved in polysaccharide biosynthesis. *Plant Methods* **15**, 65. https://doi.org/10.1186/s13007-019-0441-9 (2019).
31. Gupta, V. *et al.* RNA-Seq analysis and annotation of a draft blueberry genome assembly identifies candidate genes involved in fruit ripening, biosynthesis of bioactive compounds, and stage-specific alternative splicing. *GigaScience* **4**, s13742-015 (2015).
32. Augustin, M. M. *et al.* Elucidating steroid alkaloid biosynthesis in *Veratrum californicum*: Production of verazine in Sf9 cells. *Plant J.* **82**, 991–1003 (2015).
33. Li, S. T. *et al.* Transcriptional profile of *Taxus chinensis* cells in response to methyl jasmonate. *BMC Genom.* **13**, 1–11 (2012).
34. Jayakodi, M. *et al.* Transcriptome profiling and comparative analysis of Panax ginseng adventitious roots. *J. Ginseng Res.* **38**, 278–288 (2014).
35. He, S. M. *et al.* Identification and characterization of genes involved in benzylisoquinoline alkaloid biosynthesis in coptis species. *Front. Plant Sci.* **9**, 731. https://doi.org/10.3389/fpls.2018.00731 (2018).
36. Hassani, D. *et al.* Parallel transcriptional regulation of artemisinin and flavonoid biosynthesis. *Trends Plant Sci.* **25**, 466–476. https://doi.org/10.1016/j.tplants.2020.01.001 (2020).
37. Pan, Q. *et al.* CrERF5, an AP2/ERF transcription factor, positively regulates the biosynthesis of bisindole alkaloids and their precursors in *Catharanthus roseus*. *Front. Plant Sci.* **10**, 931. https://doi.org/10.3389/fpls.2019.00931 (2019).
38. Ma, Y. N. *et al.* Jasmonate promotes artemisinin biosynthesis by activating the TCP14-ORA complex in *Artemisia annua*. *Sci. Adv.* **4**, eaas9357. https://doi.org/10.1126/sciadv.aas9357 (2018).
39. Adhikari, D., Panthi, V. K., Pangeni, R., Kim, H. J. & Park, J. W. Preparation, characterization, and biological activities of topical anti-aging ingredients in a *Citrus junos* callus extract. *Molecules* https://doi.org/10.3390/molecules22122198 (2017).
40. Kikowska, M. A. *et al.* Effect of pentacyclic triterpenoids-rich callus extract of *Chaenomeles japonica* (Thunb.) Lindl. ex spach on viability, morphology, and proliferation of normal human skin fibroblasts. *Molecules* https://doi.org/10.3390/molecules23113009 (2018).
41. Yang, F. *et al.* Illumination on "reserving phloem and discarding xylem" and quality evaluation of *Radix polygalae* by determining oligosaccharide esters, saponins, and xanthones. *Molecules* https://doi.org/10.3390/molecules23040836 (2018).
42. Jensen, K. H. Phloem physics: Mechanisms, constraints, and perspectives. *Curr. Opin. Plant Biol.* **43**, 96–100. https://doi.org/10.1016/j.pbi.2018.03.005 (2018).
43. Cho, S. J. *et al. Endophytic bacillus* sp. isolated from the interior of balloon flower root. *Biosci. Biotechnol. Biochem.* **66**, 1270–1275. https://doi.org/10.1271/bbb.66.1270 (2002).
44. Venugopalan, A. & Srivastava, S. Endophytes as in vitro production platforms of high value plant secondary metabolites. *Biotechnol. Adv.* **33**, 873–887. https://doi.org/10.1016/j.biotechadv.2015.07.004 (2015).
45. Li, L. *et al.* Jasmonic acid-responsive AabHLH1 positively regulates artemisinin biosynthesis in *Artemisia annua*. *Biotechnol. Appl. Biochem.* **66**, 369–375. https://doi.org/10.1002/bab.1733 (2019).
46. Shen, Q. *et al.* The jasmonate-responsive AaMYC2 transcription factor positively regulates artemisinin biosynthesis in *Artemisia annua*. *New Phytol.* **210**, 1269–1281. https://doi.org/10.1111/nph.13874 (2016).
47. Fan, M. *et al.* A trihelix family transcription factor is associated with key genes in mixed-linkage glucan accumulation. *Plant Physiol.* **178**, 1207–1221. https://doi.org/10.1104/pp.18.00978 (2018).
48. Maréchal, E. *et al.* Modulation of GT-1 DNA-binding activity by calcium-dependent phosphorylation. *Plant Mol. Biol.* **40**, 373–386. https://doi.org/10.1023/a:1006131330930 (1999).
49. Kajani, A. A., Moghim, S. & Mofid, M. R. Optimization of the basal medium for improving production and secretion of taxanes from suspension cell culture of *Taxus baccata* L. *Daru* **20**, 54. https://doi.org/10.1186/2008-2231-20-54 (2012).
50. Chen, F., Huang, G., Yang, Z. & Hou, Y. Antioxidant activity of Momordica charantia polysaccharide and its derivatives. *Int. J. Biol. Macromol.* **138**, 673–680. https://doi.org/10.1016/j.ijbiomac.2019.07.129 (2019).
51. Kim, H. S. *et al.* Identification of xenobiotic biodegradation and metabolism-related genes in the copepod *Tigriopus japonicus* whole transcriptome analysis. *Mar. Genom.* **24**(Pt 3), 207–208. https://doi.org/10.1016/j.margen.2015.05.011 (2015).
52. Miller, W., Myers, E. W. & Lipman, D. J. B. *Encyclopedia of Genetics, Genomics, Proteomics and Informatics* 221–221 (Springer, Berlin, 2008).
53. Conesa, A. *et al.* Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676. https://doi.org/10.1093/bioinformatics/bti610 (2005).
54. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359. https://doi.org/10.1038/nmeth.1923 (2012).
55. Li, B. & Dewey, C. N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **12**, 323. https://doi.org/10.1186/1471-2105-12-323 (2011).
56. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. https://doi.org/10.1186/s13059-014-0550-8 (2014).
57. Chen, Z. *et al.* Statistical methods on detecting differentially expressed genes for RNA-seq data. *BMC Syst. Biol.* **5**(Suppl 3), S1. https://doi.org/10.1186/1752-0509-5-s3-s1 (2011).

Acknowledgements

We would like to thank BGI-Shenzhen in China for RNA-Seq sequencing and some bioinformatics analysis, and we also thank Prof. Zhaohua Peng from Mississippi State University for revising the manuscript and Dr. Xiaoxi Meng from University of Minnesota for helpful comments on an earlier version of the manuscript.

Author contributions

Conceived, designed, and implemented the study: S.X., X.S., Z.W.; Statistics analysis: S.X., X.S.; Reagents/materials/analysis tools: S.X., D.P., N.Y., S.G., N.Y., F.M. and X.G.; Drafted the manuscript: X.S., S.X.; All authors edited the manuscript and approved the final version.

Funding

This work was supported by Anhui Province Natural Science Foundation of China (Grant No. 1908085MH268); Hunan Provincial Natural Science Foundation of China (Grant No. 2018JJ3008); Key Natural Science Research Projects in Anhui Universities (Grant No. KJ2019A0453); Returnee Program of Anhui People's Society Office (Grant No. DT18100035) and Key Project of Scientific Research Fund of Anhui University of Chinese Medicine (Grant No. 2018zrzd05).

Competing interests
The authors declare no competing interests.

Guidelines and legislation
The authors declare that the experimental research and field studies on plants, including the collection of plant material, complied with our institutional, national, and international guidelines and legislation.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-89294-1.

Correspondence and requests for materials should be addressed to S.X. or D.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.