De novo Assembly and Transcriptome Analysis of *Polygonatum odoratum* (Mill.) Druce : Detection of Potential Genes Involved in Polysaccharide Synthesis

shengxiang zhang

Anhui University of Traditional Chinese Medicine

yuanyuan shi

Anhui University of Traditional Chinese Medicine

luqi huang

Anhui University of Traditional Chinese Medicine

chenkai wang

Anhui University of Traditional Chinese Medicine

derui zhao

Anhui University of Traditional Chinese Medicine

kelong ma

Anhui University of Traditional Chinese Medicine

jiawen wu (✉ wujiawen@ahTCM.edu.cn)

Anhui University of Traditional Chinese Medicine <https://orcid.org/0000-0003-4407-6819>

daiyin peng

Anhui University of Traditional Chinese Medicine

Research article

Keywords: *Polygonatum odoratum*, transcriptome , polysaccharide synthesis, Differentially expressed gene, key enzyme genes

Posted Date: August 23rd, 2019

DOI: <https://doi.org/10.21203/rs.2.13418/v1>

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

[Read Full License](#)

Abstract

Background: *Polygonatum odoratum* (Mill.) Druce is a well-known traditional Chinese herb. Polysaccharide is one of the main bioactive components from *Polygonatum odoratum*, with broad pharmacological effects, including improving immunity, and is used in the treatment of rheumatic heart disease, cardiovascular disease, and diabetes. **Results:** This study identifies potential genes and transcription factors (TFs) that regulate polysaccharide synthesis in *Polygonatum odoratum* by using RNA sequencing data from leaf, stem and rhizome tissues. A total of 112,443 unigenes were de novo assembled, and 76,714 were annotated in public databases. Differentially expressed gene analysis showed that most upregulated and uniquely expressed unigenes were enriched in rhizome tissue compared with leaf or stem tissue. UV-spectrophotometry results showed that polysaccharide content was the greatest in the rhizome tissue. Additionally, 2,865 unigenes relevant to TF families were predicted, including 73 involved in polysaccharide synthesis. A few key enzyme genes were also verified by quantitative real-time PCR (qRT-PCR). Seven β -fructofuranosidases with different amino acid sequences showed similar spatial structures and all had well-conserved catalytic triads. **Conclusion:** This study substantially enlarges the public transcriptome datasets of this species, and provides insight into the detection of novel genes involved polysaccharide and other secondary metabolite synthesis. **Keywords:** *Polygonatum odoratum*, transcriptome, polysaccharide synthesis, Differentially expressed gene, key enzyme genes

Background

Polygonatum odoratum (Mill.) Druce (*P. odoratum*), a typical representative of the *Liliaceae* family, is a perennial herbaceous plant, that is widely distributed in East Asia and Europe[1]. Resources of this herb are diminishing due to uncontrolled excavation wild harvesting. *P. odoratum* rhizomes are regarded as the medicinal part of the plant, which have been used extensively to treat diseases, such as hyp immunity, rheumatic heart disease, cardiovascular diseases and diabetes [2, 3]. *P. odoratum* has been found to contain components with bioactive effects; its composition includes polysaccharides, steroidal glycosides, dipeptides, flavonoids, amino acids, and trace mineral elements [4, 5]. Polysaccharides are significant bioactive components of *P. odoratum*, with immunomodulatory, antidiabetic, antiaging, antitumor and antioxidant properties [6].

Polysaccharides are long-chain polymers, with more than 20 monosaccharide molecules linked by glycosidic bonds, which can be divided into homopolysaccharide and heteropolysaccharide types [7, 8]. Homopolysaccharides are composed of the same monosaccharide, such as starch and β -glucan, while heteropolysaccharides are comprised of two or more monosaccharides, such as glucomannan and pectin. *P. odoratum* polysaccharides are heteropolysaccharides, comprising of mannose, galactose, glucose, fructose, rhamnose, arabinose and galacturonic acid[9, 10]. Previous studies have shown that polysaccharides in *P. odoratum* can treat hyperthyroidism, osteoporosis and ameliorate the immunological disequilibria, thereby enhancing cell immunity and postponing the senility [11, 12].

Polysaccharides biosynthesis comprises of three main processes, Firstly, sucrose is converted to glucose 1-phosphate (Glc-1P) and fructose 6-phosphate (Fru-6P) [13]. In this reaction, sucrose forms glucose 6-phosphate (Glc-6P) and fructose, which is catalyzed by β -fructofuranosidase (sacA), and fructose is then converted into Fru-6P by hexokinase (HK) and fructokinase (scrk)[14, 15]. Phosphoglucosyltransferase (pgm) catalyzes the isomerization of Glc-6P to Glc-1P, and UDP-glucose (UDP-Glc) and guanosine diphosphate mannose (GDP-Man) are produced from Glc-1P and Fru-6P precursors, respectively [16, 17]. Secondly, several NDP-sugar interconversion enzymes (NSEs) catalyze the conversion of either UDP-Glc or GDP-Man to other NDP sugars [18]. Finally, the NDP-sugars are used in polysaccharide biosynthesis through various glycosyltransferases (GTs)[19, 20]. To date, there have been no comprehensive reports on the genes involved in polysaccharide metabolism in *P. odoratum*, and no reports exist on their expression patterns in different tissues.

RNA-sequencing can provide comprehensive information for gene expression to improve our understanding of gene regulation and metabolic pathways [21]. Some medicinal plant genomes have been sequenced, including *Artemisia argyi*[22], *Dendrobium officinale*[23], *Pueraria lobata*[24], *Gentiana rigescens*[25] and *Arisaema heterophyllum* Blume[26], and novel genes encoding key enzymes involved in specific metabolic pathways have been identified. However, genomic information for *P. odoratum* has not yet been reported, which hinders further research on this species.

In this study, we performed deep de novo transcriptome analysis on different tissues of *P. odoratum* using RNA sequencing. The numbers of potential genes that participate in polysaccharide pathways are identified in our transcriptome results. The transcriptome data from *P. odoratum* will be an important resource to investigate polysaccharide biosynthesis and other metabolic pathways in plants.

Methods

Plant material and RNA extraction.

Whole *Polygonatum odoratum* plants (identified by Professor Qingshan Yang,

Anhui University of Chinese Medicine) were harvested from the Anhui University of Chinese Medicine herb garden with permission of managers and Professional. Fresh plants was washed with sterile distilled water several times and wiped clean with filter paper. The tissues (leaf, rhizome and stem) were separately collected and placed in a 50 mL centrifuge tube, and quickly frozen in liquid nitrogen , and then stored in a refrigerator at -80°C for RNA extraction. The stems, leaves and rhizomes selected from three replicates were pooled together. Total RNA from the plants was extracted using the RNA Plant Kit (Aidlab Biotech, Beijing, China) according to the manufacturer's protocol.

Determination of total polysaccharides content

The dried *P. odoratum* samples (rhizomes, stems and leaves) were used to detect the content of polysaccharides with the Phenol-sulfuric acid method [27]. Dried powders (0.2 g) from each sample were mixed with 100mL distilled water and then extracted twice at 85 °C (1 h each time) in boiling distilled water. After that, precipitate after adding 95% ethanol (10 mL) and centrifuging was collected and dissolved with distilled water (50 mL), and 4% phenol (1 mL) and sulfuric acid (7 mL) were added for determine the absorbance. The absorbance of three tissues were measured in 490nm using UV-spectrophotometry (JASCO company, Japan). The standard curve of the relationship between the concentration and the absorbance was drawn through a standard of anhydrous glucose (**Supplementary Image S1**). The yield (%) of total polysaccharides was calculated using the equation:

cDNA library construction and sequencing (mRNA-Seq).

Following the total RNA extraction, The extracted RNA was checked using a NanoDrop 2000 (Thermo, CA, USA), and the RNA quality was evaluated using an Agilent 2100 BioAnalyzer to ensure that the RNA Integrity Number (RIN) values were above 7.0. mRNA was enriched using Oligo (dT). mRNA was split into small fragments using a fragmentation buffer, and cDNA synthesis was performed with random primers. PCR amplification was prepared for library construction according to manufacturer' instructions. In detail, each mRNA samples were then used for first-strand cDNA synthesis using reverse transcriptase and random hexamer primers, while second-strand cDNA was synthesized with RNase H and DNA Polymerase I by using NEBNext doublestranded cDNA (ds cDNA) Fragmentase (New England BioLabs). Short cDNA fragments were repaired with NEBNext End Repair Module (New England BioLabs), and a single nucleotide (adenine) was added to 3-ends using the NEBNext dA-Tailing Module (New England BioLabs). cDNA Fragments were ligated to the NEBNext Adaptor by using the NEBNext Quick Ligation Module. The quality of the each cDNA libraries was evaluated using the ABI StepOnePlus Real-Time PCR System and Agilent 2100 Bioanalyzer (Agilent RNA 6000 Nano Kit). The libraries were sequenced by the BGISEQ-500 at the Beijing Genomics Institute (BGI)(Shenzhen, Guangdong province, China)[28].

Transcriptome *de novo* assembly and unigene functional annotation.

After sequencing, raw data were received, and low quality reads (above 50% of bases with Q-value 20) were removed to obtain clean reads [29]. De novo assembly of the transcriptome was accomplished from clean reads with Trinity v2.0.6 [30]. The contigs were connected using Trinity, and obtain sequences termed as unigenes. The Unigenes were divided into two classes, including clusters and singletons. The prefix of clusters was CL with the cluster id behind it, The prefix of singletons was unigene.

The unigenes were used for BLAST and annotation against protein databases, including Nt (NCBI nucleotide sequences), Nr (NCBI non-redundant protein sequences), KOG (Clusters of euKaryotic

Orthologous Groups), KEGG (Kyoto Encyclopedia of Genes and Genome), SwissProt (A manually annotated and reviewed protein sequence database). With nr annotation, the GO(Gene Ontology) annotation of unigenes was obtained using the Blast2GO program[31]. With the help of the KEGG database, we can further excavate genes' biological complex behaviors, and pathway annotation for unigenes were obtained by KEGG [32].

Analysis of differentially expressed genes(DEGs).

Each unigene was normalized into FPKM (Fragments Per Kb per Million fragments) values. The unigenes abundance was calculated by the ratio of FPKM values ($|\log_2 \text{Ratio}|$). To analyze the differentially expressed genes (DEGs), clean reads to assembled unigenes were mapped using Bowtie2(version 2.2.5). DEGs were identified by PossionDis method, which was performed as described: The parameters of PossionDis was set as $|\text{Fold Change}| > 2.0$ -fold ($|\log_2 \text{FC}| > 1$) with a false discovery rate (FDR) < 0.001 [33]. For functional annotation clustering of DEGs, GO analysis was performed using GO-Term Finder software. DEGs were mapped to the terms of KEGG database for the pathway enrichment analysis.

Identification of transcriptome factor (TF)

TF, short for Transcription Factor, is a protein that binds to specific DNA sequences, thereby controlling the rate of transcription of genetic information from DNA to messenger RNA. To analyze TFs of *P. odoratum*, we tested the ORFs of unigenes and then aligned the ORF to the transcription factor protein domain using hmmsearch. Unigenes were annotated by using PlantTFDB(Plant transcription factor database) [34]. The unigenes encoding the transcription factors were identified by comparison with Pfam23.0 using the hmmsearch program [35].

qRT-PCR Analysis of key genes in polysaccharide biosynthesis.

The qRT-PCR was conducted on a Real-time Thermal Cycler 5100 PCR System using GoTaq qPCR Master Mix PCR kit(Promega). Specific primers were designed using Primer Premier 5.0 (**Additional files 12 : Table 4**). Total RNA was extracted from the leaves, stems, and rhizomes, and reverse-transcribed to cDNA with TransScript All-in-One cDNA supermix for qPCR (TransGen), following the manufacturer's instructions. All PCR conditions were performed as follows: 95 °C for 2 min, 40 cycles of 95°C for 15 s and 60°C for 1 min, and all qRT-PCR were repeated in three biological and three technical replications. Additionally, the actin gene (CL2254.Contig1) of *P. odoratum* was used as a reference. The relative expression levels of the selected genes were determined using the equation $2^{(-\Delta\Delta Ct)}$ method [36].

Analysis of the structure characteristics of sacA.

Total of 7 complete amino acid sequences were selected from 20 unigenes encode sacA (β -fructofuranosidase). The alignment of 7 sacA amino acid sequences was performed using DNAMAN software. 3D structures model of sacA proteins was analyzed using the SWISS-MODEL (<https://swissmodel.expasy.org/>) and PyMOL software [37].

Results

Total polysaccharides content in *P. odoratum* samples

Total polysaccharide was extracted from the dried rhizomes, stems and leaves of *P. odoratum*. Polysaccharide content was the highest in the rhizomes (3.11%) and the lowest in the leaves (1.15%) (**Additional files 1 : Figure 1**).

RNA-seq and *de novo* transcriptome assembly.

The number of raw reads for *P. odoratum* for the samples of leaf, stem and rhizome tissue was 120,415,666, 117,871,048 and 120,302,266, respectively; these were generated using a BGISEQ-500 high-throughput sequencing platform. The Q30 of each sample was greater than 89.33%, and the GC content of each sample was approximately 43.57%. After thorough quality control and filtering, transcriptomes were generated and a total of 112,443 unigenes were assembled and clustered using Trinity software. The average length of unigene was 1240 bp and N50 was 1937 bp. Of these unigenes, 47.35% (53,237) were longer than 1000 bp, while 66.81% (75,126) were longer than 500 bp (**Additional files 2 : Figure 2**). The quality of the assembled transcripts was assessed using a single copy orthologous database (BUSCO), and 97% of the unigenes were perfectly matched in the BUSCO database (**Additional files 3 : Figure 3**).

Unigenes functional annotation.

In order to predict and classify possible functions of unigenes, 112,443 unigenes were annotated by NR, NT, Swissprot, KEGG, Pfam, KOG, and GO databases. In total, 76,714 (68.22%) unigenes were mapped to least one public database. 72,709 (NR: 64.66%), 56,712 (NT: 50.44%), 55,223 (Swissprot: 49.11%), 56,755 (KEGG: 50.47%), 17,752 (GO: 15.79%), 58,103 (KOG: 51.67%), and 53,845 (Pfam: 47.89%) unigenes were annotated (**Table 1**). A total of 10,877 (9.67%) unigenes were annotated together by five databases (**Fig. 1A**). The majority of mapped unigenes revealed the closest homology was with those from *Asparagus officinalis* (62.71%) in the NR database. The second highest homology was with *Elaeis guineensis* (9.16%). *Phoenix dactylifera* (5.66%), *Ananas comosus* (2.13%), and *Musa acuminata* subsp. malaccensis (2.02 %) were other homologous species (**Fig. 1B**). In each tissue, the expression values of

all transcripts (FPKM > 1) were counted, and a total of 59,271, 52,184 and 45,893 unigenes were expressed in leaf, stem and rhizome tissues, respectively.

GO classification showed that 17,752 unigenes (15.79%) were divided into three categories: molecular function, cellular component, and biological processes. A detailed analysis of the biological processes group showed 'cellular process' (5,209 genes), 'biological regulation' (1,765 genes), 'localization' (1,260 genes), 'metabolic process' (1,161 genes), and 'cellular component organization or biogenesis' (1158 genes) were highly represented. In relation to molecular function, 'binding' (8,257 genes) and 'catalytic activity' (7,963 genes) were the top two GO terms. In the cellular component group, the top three GO terms were related to 'cell' (5,313 genes), 'membrane part' (5,259 genes), and 'organelle part' (2,301 genes) (**Additional files 4 : Figure 4**). For KOG analysis results, 58,103 unigenes (51.67%) were grouped into 25 functional categories: The top two categories were "general function prediction only" (12,145 genes), and "signal transduction mechanisms" (6,994 genes).

Overview of expression

In each tissue, the expression value of all transcripts (FPKM > 1) were counted, and a total of 41,811, 44,472 and 46,349 unigenes were expressed in leaf, rhizome and stem tissues, respectively (**Fig. 2A**). The overall level of gene expression was higher in rhizome tissue than in leaf and stem tissues (**Fig. 2B**).

Characterization of functional genes involved in polysaccharide biosynthesis via KEGG pathway analysis.

KEGG classification resulted in 56,755 unigenes (50.47%) being mapped into 20 pathways, and the most represented KEGG function category was global and overview maps (12,994 genes), which was followed by carbohydrate metabolism (4,829 genes), translation (4,537 genes), folding, sorting and degradation (3,572 genes), and transcription (3,238 genes) (**Additional files 5 : Figure 5**). The "carbohydrate metabolism" subcategory contained 15 pathways, including starch and sucrose metabolism (ko00500), amino sugar and nucleotide sugar metabolism (ko00520), glycolysis / gluconeogenesis (ko00010), citrate cycle (ko00020), pentose phosphate pathway (ko00030), pentose and glucuronate interconversions (ko00040), fructose and mannose metabolism (ko00051), galactose metabolism (ko00052), ascorbate and aldarate metabolism (ko00053), inositol phosphate metabolism (ko00562), pyruvate metabolism (ko00620), glyoxylate and dicarboxylate metabolism (ko00630), propanoate metabolism (ko00640), butanoate metabolism (ko00650), and c5-branched dibasic acid metabolism (ko00660) (**Fig. 3**). Among these pathways, we annotated 1108 unigenes involved in starch and sucrose metabolism and 942 unigenes involved in amino sugar and nucleotide sugar metabolism. Based on bioinformatics analysis, 211 unigenes encoding the enzymes involved in polysaccharide biosynthesis were identified, including beta-fructofuranosidase (sacA), hexokinase (HK), fructokinase (scrK), mannose-

6-phosphate isomerase (MPI), phosphomannomutase (PMM), mannose-1-phosphate guanylyltransferase (GMPP), GDP-mannose 4,6-dehydratase (GMDS), GDP-L-fucose synthase (TSTA3), glucose-6-phosphate isomerase (GPI), phosphoglucomutase (pgm), UTP-glucose-1-phosphate uridylyltransferase (UGP2), UDP-glucose 4-epimerase (GALE), UDP-glucose 6-dehydrogenase (UGDH), UDP-apiose/xylose synthase (AXS), UDP-arabinose 4-epimerase (UXE), UDP-glucose 4,6-dehydratase (RHM) and 3,5-epimerase/4-reductase (UER1) (**Table 2**). In the *P. odoratum* transcriptome, 7 subclasses of NDP-sugar interconversion enzymes (NSEs) were identified, namely RHM (17 unigenes), UER1 (2 unigenes), GMDS (6 unigenes), GALE (34 unigenes), UGDH (12 unigenes), UGE (1 unigenes) and UXE (5 unigenes). Based on our results that identified constituents and key enzymes in carbohydrate metabolism, we outlined potential biosynthetic pathways for polysaccharide formation (**Fig. 4**). We obtained key enzymes unigenes that were predicted to be the sacA, HK, scrK, MPI, PMM, GMPP, GMDS, TSTA3, GPI, pgm, UGP2, GALE, UGDH, AXS, UXE, UGE, RHM, and UER1 may involve in polysaccharide biosynthesis.

Validation and analysis of differentially expressed genes (DEGs) and specific expressed genes in *P. odoratum* tissues.

In *P. odoratum* tissues, 78,288 shared unigenes were identified, of which 3,760 unigenes were found to be involved in polysaccharide biosynthesis. Based on the FPKM values of all unigenes, 4,374, 3,343, and 3,109 unigenes showed unique expression in leaf, rhizome and stem tissues, respectively (**Fig. 5A**). Among these unique expression unigenes, 128, 82, and 79 unigenes are involved in carbohydrate metabolism, respectively (**Additional files 6 : Figure 6**).

Based on the KEGG enrichment analysis of DEGs, 38,085 unigenes (22,825 up-regulated unigenes and 15,260 down-regulated unigenes) were identified as significant DEGs in rhizome tissue compared to leaf tissue, including 810 unigenes involved in phenylpropanoid biosynthesis, 566 unigenes involved in starch and sucrose metabolism, 147 unigenes involved in carotenoid biosynthesis and 133 unigenes involved in flavonoid biosynthesis. Comparison of rhizome tissue and stem tissue revealed 23,016 DEGs (9,701 up-regulated unigenes and 13,315 down-regulated unigenes), including 557 unigenes involved in phenylpropanoid biosynthesis, 380 unigenes involved in starch and sucrose metabolism, 107 unigenes involved in flavonoid biosynthesis and 100 unigenes involved in carotenoid biosynthesis (**Additional files 7 : Figure 7**). Whereas comparison of the rhizome tissue with stem and leaf tissues showed that 16,655 DEGs, including 7,647 co-upregulated unigenes and 7,797 co-downregulated ones in rhizome tissue compared to stem and leaf tissues (**Fig. 5B**). Among the up-regulated and down-regulated genes, we focused more on those up-regulated unigenes involved in the pathway of biosynthesis. 7610 specific up-regulated unigenes were identified in rhizome tissue compared to stem and leaf tissues with log₂ (fold changes) > 1, and these unigenes were further estimated by KEGG classification and GO enrichment. Among these up-regulated unigenes, 394 unigenes involved in carbohydrate metabolism were identified by KEGG classification, while 42 unigenes were enriched to DNA binding transcription factor activity by GO enrichment (**Fig. 5C,D**).

The “carbohydrate metabolism” classification was enriched in DEGs. Of which, 2367 unigenes were up-regulated in rhizome tissue compared to leaf tissue, while 1821 unigenes were up-regulated in rhizome tissue compared to stem tissue (**Additional files 9 : Table 1**). The “Biosynthesis of other secondary metabolites” classification was enriched in DEGs. Of which, 1031 unigenes were also up-regulated in rhizome tissue compared to leaf tissue, while 809 unigenes were up-regulated in rhizome tissue compared to stem tissue (**Additional files 10 : Table 2**). In this study, these up-regulated unigenes in rhizome tissue were the candidate genes that can enhance the availability of the precursor for biosynthesis of polysaccharide, and have the potential to improve the production of polysaccharide.

Analysis of transcriptome factors involved in polysaccharide biosynthesis and other secondary metabolites

Transcription factors (TFs) participate in a variety of developmental and physiological roles in plants. An increasing number of TFs have been isolated and characterized for several plant secondary metabolic pathways. From our *Podoratum* database, 2,865 unigenes were annotated in the PlantTFDB and assigned to 58 TF families. Among these TF families, the most abundant one was the MYB family (343 unigenes), followed by bHLH (223 unigenes), WRKY (188 unigenes), AP2-EREBP (180 unigenes), NAC (149 unigenes), bHLH (150 unigenes), C3H (143 unigenes), FAR (135 unigenes), C2H2 (76 unigenes), and Trihelix (69 unigenes) (**Fig. 6A**). Through pathway classifications of all TF families, 73 unigenes involved in carbohydrate metabolism were identified, including MYB (6 unigenes), C2H2 (41 unigenes), Trihelix (7 unigenes), bHLH (4 unigenes), FAR1 (2 unigenes), and GeBD (2 unigenes) (**Fig. 6B**). Meanwhile, 49 unigenes involved in biosynthesis of other secondary metabolites were identified, including MYB (15 unigenes), Trihelix (8 unigenes), C2C2-Dof (5 unigenes), LoB (4 unigenes), C2H2 (3 unigenes), and BES1 (2 unigenes) (**Fig. 6C**). These unigenes have been identified in the regulation of polysaccharide biosynthesis pathways (**Additional files 11 : Table 3**).

As shown in the heat map of hierarchical clustering(**Additional files 8 : Figure 8**), the expression level of 73 transcriptome factors involved in carbohydrate metabolism was higher in rhizome tissue compared to leaf and stem tissues, including 47 upregulated unigenes in rhizome vs leaf tissue and 42 up-regulated unigenes in rhizome vs stem tissue. The expression level of 49 transcriptome factors involved in the biosynthesis of other secondary metabolites was higher in rhizome tissue compared to leaf and stem tissues, of which 37 up-regulated unigenes in rhizome vs leaf tissue and 27 up-regulated unigenes in rhizome vs stem tissue.

Validation of key enzyme genes using qRT-PCR

UGE, *UGP2*, *GMPP* and *sacA* genes were tested for their expression level using qRT-PCR assays. The relative expression level of *UGE*, *UGP2* and *sacA* was obtained for rhizome, stem and leaf tissues, and

rhizome tissue showed higher expression than stem or leaf tissues. The relative expression of the *GMPP* gene was greater in leaf tissue than stem or rhizome tissue, which is consistent with our transcriptional data (Fig. 7).

The structural characteristics of *sacA* involved in polysaccharide biosynthesis.

β -fructofuranosidase (*sacA*) is the first key enzyme in the polysaccharide synthesis pathway, which catalyzes the conversion of sucrose to glucose 6-phosphate (Glc-6P) and Fructose. The alignment of 7 *sacA* amino acid sequences revealed that the sequence identity is not high (59.57%), but the 7 *sacA*s showed similar spatial structures, and all had three well-conserved motifs. A 3D structural model of *sacA* (CL7969. Contig2) was constructed based on the crystal structure of 6-SST/6-SFT from *Pachysandra terminalis* (PDB ID: 3UGF) using SWISS-MODEL (<https://swissmodel.expasy.org/>) and PyMOL software. The spatial structure model consisted of one β -propeller domain and one β -sheet domain, which are connected by an α -helix. The β -propeller domain has five radially oriented blades (Fig. 8A, blades I–V, colored in orange, cyan, magenta, green, and yellow, respectively), each with one α -helix at the N-terminus and C-terminus. The β -sheet domain consists of two six-stranded antiparallel beta-sheets (Fig. 8A, colored in blue), forming a sandwich-like fold. A well-conserved catalytic triad is located in the deep central pocket of the beta-propeller domain (D95, D224 and E282) (Fig. 8C).

Discussion

Although polysaccharides are the major bioactive components of *P. odoratum*, with significant medicinal value, the molecular mechanisms behind their synthesis are still unknown. To further identify the genes encoding key enzymes and TFs modulating polysaccharide synthesis, we constructed a comprehensive genomic library of *P. odoratum* leaf, stem and rhizome tissues. This is the first report about transcriptome study of *P. odoratum*, providing adequate references to study on other plants with close relationship to *P. odoratum*. A total of 112,443 unigenes were assembled in our datasets, with an N50 of 1937 bp. 47.35% unigenes (53,237) were longer than 1000 bp, while 66.81% unigenes (75,126) were longer than 500 bp. Compared with transcript datasets reported for other medicinal plants, *P. odoratum* has more unigenes than *Polygonatum sibiricum* (76,717)[38] and *Platycodon grandiflorum* (34,053)[39], and the average unigene length (1937 bp) is longer than those from *Polygonatum sibiricum* (900 bp), *Pinellia ternata* (750 bp)[40], and *Platycodon grandiflorum* (1102 bp). These results strongly indicated a high quality of assembly and the reliability of our transcriptome database.

Analysis of the DEGs of KEGG pathways indicated that 38,085 DEGs were mapped to metabolic pathways in rhizome versus leaf tissue and rhizome versus stem tissue, which are mainly enriched in “starch and sucrose metabolism”, “plant hormone signal transduction”, and “phenylpropanoid biosynthesis”. Of which, 276 DEGs were up-regulated in starch and sucrose metabolism in rhizome

versus leaf tissue, while 123 DEGs were up-regulated in starch and sucrose metabolism in rhizome versus stem tissue. These results support the use of *P. odoratum* rhizomes as a Chinese medicinal material at the genetic level. KEGG annotations revealed 18 key enzymes responsible for polysaccharide biosynthesis, including sacA, HK, scrK, MPI, PMM, GMPP, GMDS, TSTA3, GPI, pgm, UGP2, GALE, UGDH, AXS, UXE, UGE, RHM, and UER1. The unigenes that encode sacA, UGE, and UGP2 enzymes showed higher expression in rhizome than leaf or stem tissue. Furthermore, the higher expression level of unigenes encoding UGP2 (CL6184.Contig2), UGE (Unigene7572), and sacA (CL7969.Contig2) in the rhizomes was authenticated by qRT-PCR data, which was consistent with the accumulation of total polysaccharide content in *P. odoratum* detected by UV-spectrophotometry. Studies of UGP2, UGE and sacA have indicated that they play important roles in regulating polysaccharide biosynthesis, and suggest that the levels of their expression may be rate-limiting on the accumulation of polysaccharide [41-43].

The seven sacAs with different amino acid sequences (identity 59.57%) showed similar spatial structures and all had well-conserved motifs. The active site of sacA is located in the deep central pocket of the β -propeller domain. A catalytic triad was identified as D95, D224 and E282 (part of the FMSPSG motif, FRDP motif and WECTD motif, respectively) by superposition and comparison with the active site of a plant fructan biosynthesis enzyme from *Pachysandra terminalis* [44] (**Fig. 8B**, shown as red spheres). These results indicate that the gene cluster of sacA may play roles in regulating polysaccharide biosynthesis in different tissues of *P. odoratum* rhizomes. The characterization of these unigenes can be used to gain deeper knowledge on the molecular mechanisms of polysaccharide biosynthesis. In the future research, we will identify the function of sacA and other candidate genes.

By analysis of TFs, 2,865 candidate genes were predicted using the PlantTFDB database, and assigned to 58 TF families, including MYB, C2H2, GeBP, bHLH, WRKY, AP2-EREBP, NAC, bHLH, C3H, FAR, and Trihelix. Total of 73 transcriptome factors involved in carbohydrate metabolism were identified, including 47 up-regulated unigenes in rhizome vs leaf tissue and 42 up-regulated unigenes in rhizome vs stem tissue. The expression levels of unigenes involved in carbohydrate metabolism was higher in rhizomes compared to either leaves or stems. Previous studies have shown that the overexpression of MYB genes results in a significant increase in mannan content in *Arabidopsis* [45], and C2H2 zinc finger transcription factors play an important role in plant tolerance to various environmental stresses such as drought, cold, osmotic stress, wounding, and mechanical loading [46, 47]. A total of 343 genes related to the MYB family and 76 genes related to the C2H2 family were examined in our transcriptional data. These results strongly indicated that these genes play critical roles in regulating the polysaccharide content.

This is the first study of large-scale transcriptome sequencing and analysis in *P. odoratum*, and these data provide a comprehensive genetic resource that will enable improvements in understanding how polysaccharide biosynthesis and accumulation are regulated at the molecular level.

Conclusions

In this study, the high-quality transcriptome sequencing data of leaf, root, and rhizome tissues of *P. odoratum* were obtained and the results of functional annotation, and expression profile were presented. Particularly, 18 key enzyme involved in the biosynthesis of polysaccharide were identified, and differentially expressed genes were analyzed. A few key enzyme genes validated using qRT-PCR were well in accordance with the expression data obtained by RNA-Seq. These studies will not only address the molecular mechanisms of polysaccharide biosynthesis in *P. odoratum*, but also provide valuable information for our complete understanding of the biosynthesis pathway on this species.

Abbreviations

P. odoratum: *Polygonatum odoratum* (Mill.) Druceis; TF: transcriptome factor; qRT-PCR: quantitative real-time PCR; DEG: differentially expressed gene; sacA: β -fructofuranosidase; KEGG: Kyoto Encyclopedia of Genes and Genomes; GO: Gene Ontology; KOG: clusters of euKaryotic Orthologous Groups; NR: NCBI non-redundant protein sequences; NT: NCBI nucleotide sequences; HK: hexokinase; scrK: fructokinase; MPI: mannose-6-phosphate isomerase; PMM: phosphomannomutase; GMPP: mannose-1-phosphate guanylyltransferase; GMDS: GDP-mannose 4,6-dehydratase; TSTA3: GDP-L-fucose synthase; GPI: glucose-6-phosphate isomerase; pgm: phosphoglucomutase; UGP2: UTP-glucose-1-phosphate uridylyltransferase; GALE: UDP-glucose 4-epimerase; UGDH: UDP-glucose 6-dehydrogenase; AXS: UDP-apiose/xylose synthase; UXE: UDP-arabinose 4-epimerase; RHM: UDP-glucose 4,6-dehydratase; UER1: 3,5-epimerase/4-reductase.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Funding

Financial support for the development and completion of this project was provided by startup funds through the National Natural Science Foundation of China (81373598) and by a Natural Science Foundation of Anhui Province of China (1608085MH177), a National students' platform for Innovation and Entrepreneurship Training Program (201810369050) providing experimental support. Publication costs were funded by startup funds through a Natural Science Research Grant of Higher Education of

Anhui Province (KJ2018ZD028) and a Project of Sustainable Utilization of Famous Traditional Chinese Medicine Resources (2060302).

Availability of data and materials

I can confirm I have included a statement regarding data and material availability in the declaration section of my manuscript. The RNA-seq datasets from three *P. odoratum* tissues were deposited within the NCBI Sequence Read Archive (SRA) database (Accession:SRP187533, <https://dataview.ncbi.nlm.nih.gov/object/PRJNA525641?reviewer=lih1q117q13bmpui01fq9uoik>).

Authors' contributions

Project design: J.W.W., D.Y.P. and L.Q.H. Experiments and data analysis: S.X.Z., Y.Y.S., C.K.W., D.R.Z., and K.L.M. Manuscript preparation: S.X.Z. Preparation of plant materials: Q.S.Y. All the authors read and approved the final manuscript.

Acknowledgements

The authors thank the Beijing Genomics Institute for assistance with experiments and Qingshan Yang for identifying plant materials.

References

1. Zhao P, Zhao C, Li X, Gao Q, Huang L, Xiao P, Gao W: **The genus *Polygonatum* : A review of ethnopharmacology, phytochemistry and pharmacology.** *Journal of Ethnopharmacology* 2017, **214**:274-291.
2. Deng Y, He K, Ye X, Chen X, Huang J, Li X, Yuan L, Jin Y, Jin Q, Li P: **Saponin rich fractions from *Polygonatum odoratum* (Mill.) Druce with more potential hypoglycemic effects.** *Journal of Ethnopharmacology* 2012, **141**(1):228-233.
3. Lin HW, Han GY, Liao SX: **[Studies on the active constituents of the Chinese traditional medicine *Polygonatum odoratum* (Mill.) Druce].** *Yao xue xue bao = Acta pharmaceutica Sinica* 1994, **29**(3):215.
4. Haiming L, Hong B, Wei L, Yuanshu W, Huanxin Z: **Study on chemical constituents of *Polygonatum odoratum* (Mill.) Druce.** *Food & Drug* 2010.
5. Quan LT, Wang SC, Zhang J: **Chemical constituents from *Polygonatum odoratum*.** *Biochemical Systematics & Ecology* 2015, **58**:281-284.
6. Jiang Q, Lv Y, Dai W, Miao X, Zhong D: **Extraction and bioactivity of polygonatum polysaccharides.** *International Journal of Biological Macromolecules* 2013, **54**(3):131-135.
7. Guo HZ, Chen CY, Zhao HR: **Determination of Polysaccharide and Their Monosaccharide Composition from *Polygonatum* and *Odoratum* by CZE.** *Chinese Journal of Experimental Traditional Medical Formulae* 2011.

8. Shi L: **Bioactivities, isolation and purification methods of polysaccharides from natural products: A review.** *International Journal of Biological Macromolecules* 2016, **92**:37-48.
9. Wang Q, Sheng-Yu LI, Yang F, Liu CM: **Isolation, Purification and Monosaccharide Analysis of a Neutral Polysaccharide from Polygonatum odoratum (Mill) Druce.** *Food Science* 2010.
10. Zhang X, Gereltu B, Zhaorigetu, Narisu: **[Determination of relative molecular mass and composition for Polygonatum sibiricum polysaccharide by high performance liquid chromatography].** *Chinese Journal of Chromatography* 2005, **23**(4):394.
11. Ying S, Dong J, Pan X, Li S, Wang W: **The effect of Polygonatum Odoratum polysaccharide on the immunological function of the aging model mice.** *Chinese Journal of Gerontology* 2007.
12. Shan Y, Jiang D, Pan XY, Jin YS, Li SH: **Immune effect of polygonatum odoratum polysaccharide on aging model mice.** *Chinese Journal of Clinical Rehabilitation* 2006.
13. Richez C, Boetzel J, Floquet N, Koteswar K, Stevens J, Badet B, Badet-Denisot MA: **Expression and purification of active human internal His6-tagged l-glutamine: d-Fructose-6P amidotransferase I.** *Protein Expr Purif* 2007, **54**(1):45-53.
14. Kelker NE, Hanson TE, Anderson RL: **Alternate pathways of D-fructose metabolism in Aerobacter aerogenes. A specific D-fructokinase and its preferential role in the metabolism of sucrose.** *Journal of Biological Chemistry* 1970, **245**(8):2060.
15. Kono T, Nishiki Y, Seki Y, Sato K, Akiba Y: **Insulin stimulates glucose transporter 1 (GLUT1) and hexokinase II (HK II) gene expression in chicken skeletal muscle.** In: *2006*; 2006.
16. Decker D, Meng M, Gornicka A, Hofer A, Wilczynska M, Kleczkowski LA: **Substrate kinetics and substrate effects on the quaternary structure of barley UDP-glucose pyrophosphorylase.** *Phytochemistry* 2012, **79**(1):39-45.
17. Marchase RB, Bounelis P, Brumley LM, Dey N, Browne B, Auger D, Fritz TA, Kulesza P, Bedwell DM: **Phosphoglucomutase in Saccharomyces cerevisiae is a cytoplasmic glycoprotein and the acceptor for a Glc-phosphotransferase.** *Journal of Biological Chemistry* 1993, **268**(11):8341-8349.
18. Yanbin Yin JH, Xiaogang Gu, Maor Bar-Peled, Ying Xu: **Evolution of Plant Nucleotide-Sugar Interconversion Enzymes.** *Plos One* 2011, **6**(11):e27995.
19. Breton C, Šnajdrová L, Jeanneau C, Koča J, Imberty A: **Structures and mechanisms of glycosyltransferases.** *Glycobiology* 2006, **16**(2):29R.
20. Gille, Sascha, Mansoori, Nasim, Xiong, Guangyan, Pauly, Markus, Souza D, Amancio: **Hemicellulose biosynthesis.** *Planta* 2013, **238**(4):627-642.
21. Saito KH, Masami Y, Yonekura-Sakakibara K: **Decoding genes with coexpression networks and metabolomics - 'majority report by precogs'.** *Trends in Plant Science* 2008, **13**(1):36-43.
22. Liu M, Zhu J, Wu S, Wang C, Guo X, Wu J, Zhou M: **De novo assembly and analysis of the Artemisia argyi transcriptome and identification of genes involved in terpenoid biosynthesis.** *Scientific reports* 2018, **8**(1):5824.

23. Chunmei H, Jianxia Z, Xuncheng L, Songjun Z, Kunlin W, Zhenming Y, Xiaojuan W, Silva JA, Teixeira Da, Zijian L, Jun D: **Identification of genes involved in biosynthesis of mannan polysaccharides in *Dendrobium officinale* by RNA-seq analysis.** *Plant Molecular Biology* 2015, **88**(3):219.
24. Han R, Takahashi H, Nakamura M, Yoshimoto N, Suzuki H, Shibata D, Yamazaki M, Saito K: **Transcriptomic landscape of *Pueraria lobata* demonstrates potential for phytochemical study.** *Frontiers in Plant Science* 2015, **6**:426.
25. Zhang X, Allan AC, Li C, Wang Y, Yao Q: **De Novo Assembly and Characterization of the Transcriptome of the Chinese Medicinal Herb, *Gentiana rigescens*.** *International Journal of Molecular Sciences* 2015, **16**(5):11550.
26. Wang C, Zhu J, Liu M, Yang Q, Wu J, Li Z: **De novo sequencing and transcriptome assembly of *Arisaema heterophyllum* Blume and identification of genes involved in isoflavonoid biosynthesis.** *Scientific reports* 2018, **8**(1):17643.
27. Kushwaha SC, Kates M: **Modification of phenol-sulfuric acid method for the estimation of sugars in lipids.** *Lipids* 1981, **16**(5):372-373.
28. Zhu FY, Chen MX, Ye NH, Qiao WM, Gao B, Law WK, Tian Y, Zhang D, Zhang D, Liu TY: **Comparative performance of the BGISEQ-500 and Illumina HiSeq4000 sequencing platforms for transcriptome analysis in plants.** *Plant Methods* 2018, **14**(1):69.
29. Cock PJA, Fields CJ, Naohisa G, Heuer ML, Rice PM: **The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants.** *Nucleic Acids Research* 2009, **38**(6):1767-1771.
30. Grabherr MG, Haas BJ, Moran Y, Levin JZ, Thompson DA, Ido A, Xian A, Lin F, Raktima R, Qiandong Z: **Full-length transcriptome assembly from RNA-Seq data without a reference genome.** *Nature Biotechnology* 2011, **29**(7):644.
31. Ana C, Stefan GT, Juan Miguel GG, Javier T, Manuel T, Montserrat R: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics* 2005, **21**(18):3674-3676.
32. Minoru K, Susumu G, Shuichi K, Yasushi O, Masahiro H: **The KEGG resource for deciphering the genome.** *Nucleic Acids Research* 2004.
33. Audic S, Claverie JM: **The significance of digital gene expression profiles.** *Genome Research* 1997, **7**(10):986-995.
34. Rice P, Longden I, Bleasby A: **EMBOSS: the European Molecular Biology Open Software Suite.** *Trends in Genetics* 2000, **16**(6):276-277.
35. Jaina M, Finn RD, Eddy SR, Alex B, Marco P: **Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions.** *Nucleic Acids Research* 2013, **41**(12):e121-e121.
36. Livak KJ, Schmittgen TD: **Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2^{-ΔΔCT} Method.** *Methods* 2001.
37. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L: **SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary**

- information. *Nucleic Acids Research* 2014, **42**(Web Server issue):W252.
38. Wang S, Wang B, Hua W, Niu J, Dang K, Qiang Y, Wang Z: **De Novo Assembly and Analysis of Polygonatum sibiricum Transcriptome and Identification of Genes Involved in Polysaccharide Biosynthesis.** *International Journal of Molecular Sciences* 2017, **18**(9):1950.
 39. Ma CH, Gao ZJ, Zhang JJ, Wei Z, Shao JH, Hai MR, Chen JW, Yang SC, Zhang GH: **Candidate Genes Involved in the Biosynthesis of Triterpenoid Saponins in Platycodon grandiflorum Identified by Transcriptome Analysis.** *Frontiers in Plant Science* 2016, **7**(3698).
 40. Zhang GH, Jiang NH, Song WL, Ma CH, Yang SC, Chen JW: **De novo Sequencing and Transcriptome Analysis of Pinellia ternata Identify the Candidate Genes Involved in the Biosynthesis of Benzoic Acid and Ephedrine.** *Frontiers in Plant Science* 2016, **7**.
 41. Luesink EJ, Marugg JD, Kuipers OP, Vos WM, De: **Characterization of the divergent sacBK and sacAR operons, involved in sucrose utilization by Lactococcus lactis.** *Journal of Bacteriology* 1999, **181**(6):1924.
 42. Mølthøj M, Verma R, Reiter WD: **The Biosynthesis of d-Galacturonate in Plants. Functional Cloning and Characterization of a Membrane-Anchored UDP-d-Glucuronate 4-Epimerase from Arabidopsis.** *Plant Physiology* 2004, **135**(3):1221-1230.
 43. Weissborn AC, Liu Q, Rumley MK, Kennedy EP: **UTP: alpha-D-glucose-1-phosphate uridylyltransferase of Escherichia coli: isolation and DNA sequence of the galU gene and purification of the enzyme.** *Journal of Bacteriology* 1994, **176**(9):2611-2618.
 44. Lammens W, Le RK, Yuan S, Vergauwen R, Rabijns A, Van LA, Strelkov SV, Van dEW: **Crystal structure of 6-SST/6-SFT from Pachysandra terminalis, a plant fructan biosynthesizing enzyme in complex with its acceptor substrate 6-kestose.** *Plant Journal for Cell & Molecular Biology* 2012, **70**(2):205-219.
 45. Kim WC, Reca IB, Kim YS, Park S, Thomashow MF, Keegstra K, Han KH: **Transcription factors that directly regulate the expression of CSLA9 encoding mannan synthase in Arabidopsis thaliana.** *Plant Molecular Biology* 2014, **84**(4-5):577.
 46. Xuan CN, Sun HK, Lee K, Kim KE, Liu XM, Han HJ, Hoang MHT, Lee SW, Hong JC, Moon YH: **Identification of a CH-type zinc finger transcription factor (ZAT10) from Arabidopsis as a substrate of MAP kinase.** *Plant Cell Reports* 2012, **31**(4):737-745.
 47. Ling J: **Identification and expression of C2H2 transcription factor genes in Carica papaya under abiotic and biotic stresses.** *Molecular Biology Reports* 2012, **39**(6):7105-7115.

Tables

Table 1 Summary statistics of annotations for *P. odoratum* unigenes via seven public databases.

Database	Number Annotated	Percentage (%)
Nr	72,709	64.66
Nt	56,712	50.44
Swissprot	55,223	49.11
KEGG	56,755	50.47
KOG	58,103	51.67
Pfam	53,845	47.89
GO	17,752	15.79
Overall	76,714	68.22

Table 2 Number of unigenes encoding the key enzymes involved in the biosynthesis of polysaccharide metabolism in *P. odoratum*.

Abbreviation	EC number	Unigene Number	Enzyme Name
sacA	3.2.1.26	20	beta-fructofuranosidase
HK	2.7.1.1	18	hexokinase
scrK	2.7.1.4	22	fructokinase
MPI	5.3.1.8	1	mannose-6-phosphate isomerase
PMM	5.4.2.8	6	phosphomannomutase
GMPP	2.7.7.13	21	mannose-1-phosphate guanylyltransferase
GMDS	4.2.1.47	6	GDP-mannose 4,6-dehydratase
TSTA3	1.1.1.271	1	GDP-L-fucose synthase
GPI	5.3.1.9	21	glucose-6-phosphate isomerase
pgm	5.4.2.2	4	phosphoglucomutase
UGP2	2.7.7.9	17	UTP-glucose-1-phosphate uridylyltransferase
GALE	5.1.3.2	34	UDP-glucose 4-epimerase
UGDH	1.1.1.22	12	UDP-glucose 6-dehydrogenase
AXS	AXS	4	UDP-apiose/xylose synthase
UXE	5.1.3.5	5	UDP-arabinose 4-epimerase
RHM	4.2.1.76	17	UDP-glucose 4,6-dehydratase
UER1	5.1.3.-, 1.1.1	2	3,5-epimerase/4-reductase
UGE	5.1.3.6	1	UDP-glucuronate 4-epimerase

Figures

Figure 1

Unigene functional annotation.

Figure 2

Overall expression profiles in *P. odoratum* leaf, rhizome and stem tissues. (A) The distributions of expressed unigene number in the three tissues. (B) Boxplot of unigenes expressed among the three tissues. The x-axis represents the three tissues, and y-axis represents log₁₀ (FPKM + 1) values.

Figure 3

Pathway classifcations for carbohydrate metabolism.

Figure 5

Unigenes expressed in *P. odoratum* tissues. (A) Venn diagram of unigenes expressed in leaf, stem and rhizome tissues. (B) Number of DEGs in the *P. odoratum* leaf, stem and rhizome tissues. A summary of unigene numbers that were up-regulated and down-regulated between two specified samples are shown. DEGs of highly expressed levels in rhizome tissue compared with in leaf or stem tissue were defined as “up-regulated”, while those with lower expression levels in rhizome tissue were defined as “down-regulated”. (C) KEGG classification of up-regulated in rhizome tissue compared to stem and leaf tissues. (D) GO enrichment of up-regulated in rhizome tissue compared to stem and leaf tissues.

Figure 6

Analysis of transcription factors. (A) Classification of transcription factor families. (B) Classification of transcription factors involved in carbohydrate metabolism. (C) Classification of transcription factors involved in biosynthesis of other secondary metabolites.

Figure 7

qRT-PCR analysis of four unigenes encoding enzymes involved in polysaccharide biosynthesis. Relative expression of CL6184.Contig2 (UGP2), Unigene7572 (UGE), CL7969. Contig2 (sacA) and Unigene7862 (GMPP) was analyzed using the actin gene (CL2254.Contig1) as the reference for normalization with technical triplicates. Root, rhizome and leaf samples are used as a normalizer for each experiment.

Figure 8

The structural model and active site of sacA. (A) The structural models of sacA. β-propeller domain (blades I–V, colored in orange, cyan, magenta, green, and yellow, respectively) and one β-sheet domain (colored in blue). (B) The active site of sacA. The catalytic triad (D95, D224 and E282) is depicted as spheres in red. (C) Alignment of seven amino acid sequences of sacAs. Black and red show identical and similar amino acids, respectively.

Supplementary Files

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

- [supplement1.docx](#)