



Article

De Novo Assembly and Analysis of *Polygonatum sibiricum* Transcriptome and Identification of Genes Involved in Polysaccharide Biosynthesis

Shiqiang Wang^{1,†}, Bin Wang^{1,2,†}, Wenping Hua³, Junfeng Niu¹, Kaikai Dang¹, Yi Qiang¹ and Zhezhi Wang^{1,*}

¹ National Engineering Laboratory for Resource Developing of Endangered Chinese Crude Drugs in Northwest of China, Key Laboratory of the Ministry of Education for Medicinal Resources and Natural Pharmaceutical Chemistry, College of Life Sciences, Shaanxi Normal University, Xi'an 710119, Shaanxi, China; wsq@snnu.edu.cn (S.W.); wangb@ecit.cn (B.W.); niujunfeng6829@126.com (J.N.); kkdang@snnu.edu.cn (K.D.); yqiang@snnu.edu.cn (Y.Q.)

² College of Chemistry, Biology and Materials Science, East China University of Technology, Nanchang 330013, Jiangxi, China

³ College of Life Sciences and Food Engineering, Shaanxi XueQian Normal University, Xi'an 710119, Shaanxi, China; huawenping@126.com

* Correspondence: zzwang@snnu.edu.cn; Tel.: +86-135-7285-8369

† These authors contributed equally to this work.

Received: 17 July 2017; Accepted: 8 September 2017; Published: 12 September 2017

Abstract: *Polygonatum sibiricum* polysaccharides (PSPs) are used to improve immunity, alleviate dryness, promote the secretion of fluids, and quench thirst. However, the PSP biosynthetic pathway is largely unknown. Understanding the genetic background will help delineate that pathway at the molecular level so that researchers can develop better conservation strategies. After comparing the PSP contents among several different *P. sibiricum* germplasms, we selected two groups with the largest contrasts in contents and subjected them to HiSeq2500 transcriptome sequencing to identify the candidate genes involved in PSP biosynthesis. In all, 20 kinds of enzyme-encoding genes were related to PSP biosynthesis. The polysaccharide content was positively correlated with the expression patterns of β -fructofuranosidase (*sacA*), fructokinase (*scrK*), UDP-glucose 4-epimerase (*GALE*), Mannose-1-phosphate guanylyltransferase (*GMPP*), and UDP-glucose 6-dehydrogenase (*UGDH*), but negatively correlated with the expression of Hexokinase (*HK*). Through qRT-PCR validation and comprehensive analysis, we determined that *sacA*, *HK*, and *GMPP* are key genes for enzymes within the PSP metabolic pathway in *P. sibiricum*. Our results provide a public transcriptome dataset for this species and an outline of pathways for the production of polysaccharides in medicinal plants. They also present more information about the PSP biosynthesis pathway at the molecular level in *P. sibiricum* and lay the foundation for subsequent research of gene functions.

Keywords: *Polygonatum sibiricum*; polysaccharide; transcriptome; polysaccharide biosynthesis; qRT-PCR

1. Introduction

Polygonatum sibiricum is a well-known traditional Chinese herb that has been widely applied for hundreds of years to treat many diseases in China, Korea, Japan, and other East Asian countries [1]. Now, it has become endangered due to the uncontrolled excavation of natural resources [2]. In the Chinese Pharmacopoeia, "*P. sibiricum*" is prescribed as the dried rhizome of *Polygonatum sibiricum* Red., *Polygonatum cyrtoneuma* Hua, and *Polygonatum kingianum* Coll. et Hemsl. within the genus *Polygonatum*. Polysaccharides isolated from *P. sibiricum* (PSPs) are major contributors to sweet taste and have many important biological functions [3]. The results of animal and cell experiments have shown that these

compounds can promote the proliferation and enhance the viability of mesenchymal stem cells in bones [4], help prevent Alzheimer's disease [5] and osteoporosis [6], serve as hypolipidemics and antiatherosclerotics [7], and improve immunologic functions [8]. Their polysaccharide structures and composition are closely related to their biological functions. Therefore, it is important to understand the biosynthesis, metabolism, and regulation of PSPs in *P. sibiricum*.

A large volume of sequences relevant to the biosynthesis of several natural products has been identified via next-generation sequencing technology [9]. However, little research has been reported about polysaccharide biosynthesis. Using boiling water extractions and alcohol precipitation, the result demonstrated that PSPs are composed of fructose [10], but genetic information for genes related to their biosynthesis remains unknown.

In this study, we conducted a comparative analysis of the transcriptomes for two germplasms of *P. sibiricum* that contrast sharply in their PSP contents. After identifying candidate genes involved in PSP biosynthesis, we verified the quality of our dataset and confirmed these putative PSP genes through quantitative real-time PCR (qRT-PCR). Our results provide insight into the biosynthesis of polysaccharides in this species.

2. Results

2.1. PSP Content Varies among *Polygonatum sibiricum* Germplasms

Crude water-soluble polysaccharide was extracted from the dried rhizomes of different germplasms of *Polygonatum sibiricum*. As shown in Figure 1, the PSP content was the highest in Lueyang, Shaanxi (SXLY) (13.33%) and the lowest in Luoyang, Henan (HNLY) (6.37%). Because of this wide gap, we chose these two germplasms for transcriptome library preparation and RNA-sequencing (RNA-seq).

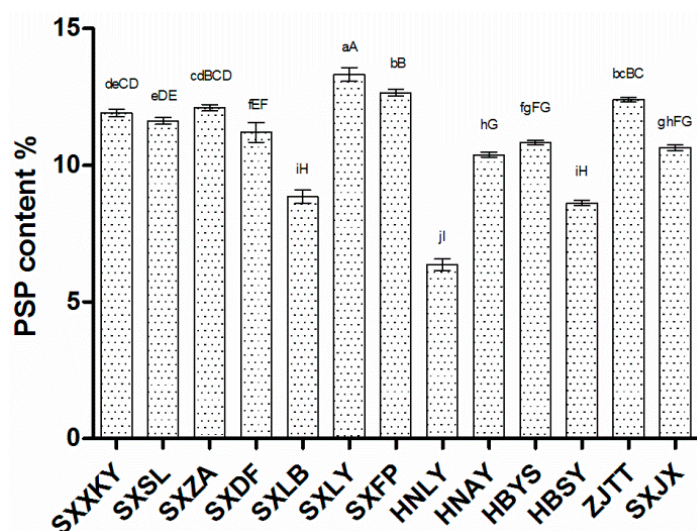


Figure 1. *Polygonatum sibiricum* polysaccharide (PSP) contents in different germplasms of *Polygonatum sibiricum*. SXXKY: Xiakou Yi, Shaanxi, China; SXSL: Shangluo, Shaanxi; SXZA: ZhenAn, Shaanxi; SXDF: Danfeng, Shaanxi; SXLB: Liuba, Shaanxi; SXLY: Lueyang, Shaanxi; SXFP: Foping, Shaanxi; HNLY: Luoyang, Henan, China; HNAY: Anyang, Henan; HBYS: Yingshan, Hubei, China; HBSY: Shiyan, Hubei; ZJTT: Tiantai, Zhejiang, China; and SXJX: Jiangxian, Shanxi, China. For the different germplasms, PSP content is significantly different between germplasms at $p < 0.05$ (different lowercase letters) and $p < 0.01$ (different capital letters).

2.2. Illumina Sequencing, De Novo Assembly, and Assessment of Assembly Program

We obtained 72.06 Gb of clean data from seven samples via the Illumina system. The mixed materials (PSABC) used to provide the genetic background comprised a pool of rhizomes, stems, and leaves. The PS01-03 samples represented the SXLY rhizomes, while the PS04-06 samples contained the HNLY rhizomes. Each sample had more than 6.79 Gb of clean data, with Q30 values greater than 93.31% (Table 1). After adaptor sequences, ambiguous reads, and low-quality reads were removed, the sequence data were assembled with Trinity software to produce 23,733,072 contigs, with an N50 of 45 bp. Based on paired-end information and the similarity of contig sequences, we obtained 133,121 transcripts, with an N50 of 1532 bp. Finally, 74,130 unigenes were generated with an N50 of 1364 bp. The size distribution was homogeneous (Figure 2); 30.36% of the unigenes were longer than 1000 bp, while 57.34% were longer than 500 bp (Table 2). These findings indicated that the assembly was successful [11].

Table 1. Output statistics from sequencing of *Polygonatum sibiricum* transcriptome.

ID ¹	Number of Reads	Number of Bases	Guanine-Cytosine (GC) Content	N Percentage	% ≥Q30
PSABC	48,703,742	14,514,388,398	50.21%	0.00	93.57
PS01	29,810,908	8,876,080,296	49.81%	0.00	93.40
PS02	30,941,196	9,196,437,754	49.21%	0.00	93.44
PS03	36,392,600	10,740,869,676	49.37%	0.00	93.55
PS04	40,885,773	12,098,862,416	49.70%	0.00	93.31
PS05	22,762,445	6,786,142,412	48.66%	0.00	94.31
PS06	32,982,501	9,849,781,232	48.96%	0.00	93.35

¹ PSABC: mixed sample of rhizomes, stems, and leaves; PS01-03: rhizomes from Lueyang, Shaanxi (SXLY); PS04-06: rhizomes from Luoyang, Henan (HNLY).

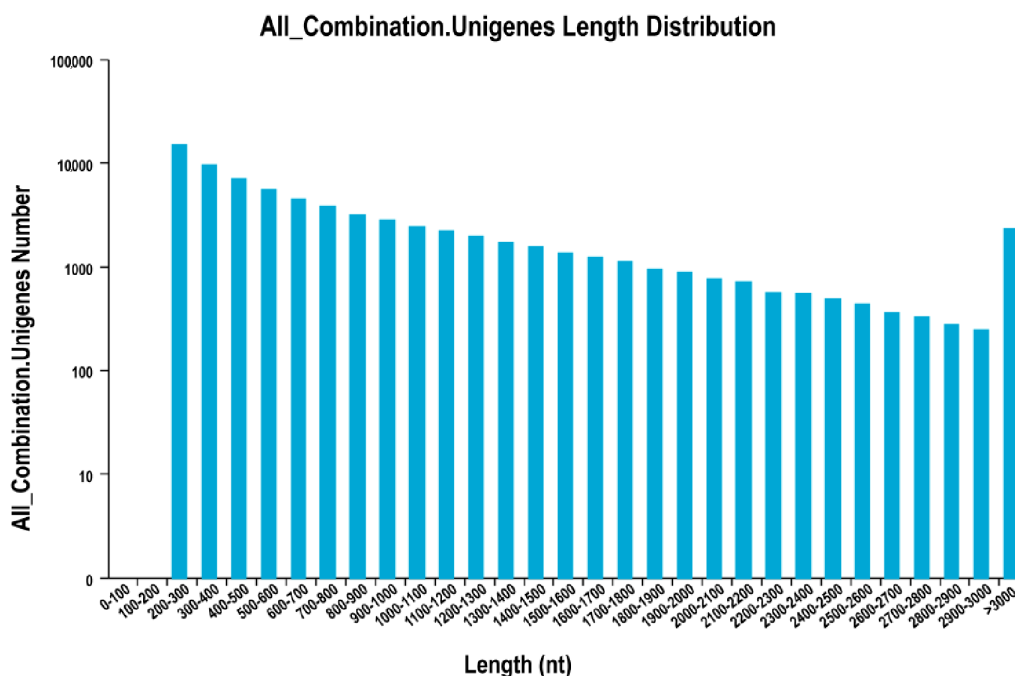


Figure 2. Distribution of unigene lengths from *Polygonatum sibiricum*.

Table 2. Output statistics from assembly of *Polygonatum sibiricum* transcriptome sequencing.

Length Range (bp)	Number of Contigs	Transcript Abundance	Number of Unigenes
200–300	23,592,360 (99.41%)	19,417 (14.59%)	15,057 (20.31%)
300–500	72,139 (0.30%)	26,138 (19.63%)	16,631 (22.43%)
500–1000	42,020 (0.18%)	37,052 (27.83%)	19,930 (26.89%)
1000–2000	19,425 (0.08%)	33,660 (25.29%)	15,422 (20.80%)
>2000	7128 (0.03%)	16,854 (12.66%)	7090 (9.56%)
Total number	23,733,072	133,121	74,130
Total length	1,025,259,786	138,445,867	66,783,395
N50 length	45	1532	1364
Mean length	43.20	1040.00	900.90

2.3. Functional Annotations

All unigenes generated by HiSeq 2500 were aligned to information in the non-redundant (NR), Clusters of Orthologous Groups (COG), euKaryotic Orthologous Groups (KOG), Gene Ontology (GO), Swiss-Prot, Protein family (Pfam), and Kyoto Encyclopedia of Genes and Genomes (KEGG) protein databases by BLASTx and BLASTn. After all of the amino acid sequences were predicted, we compared them with the Pfam database to obtain annotation information. In total, 35,793 unigenes (48.28%) were aligned to homologous sequences in those public databases. Of them, 14,618 (19.72%) were 300 to 1000 bp long, while 17,140 (23.12%) were longer than 1000 bp (Table 3).

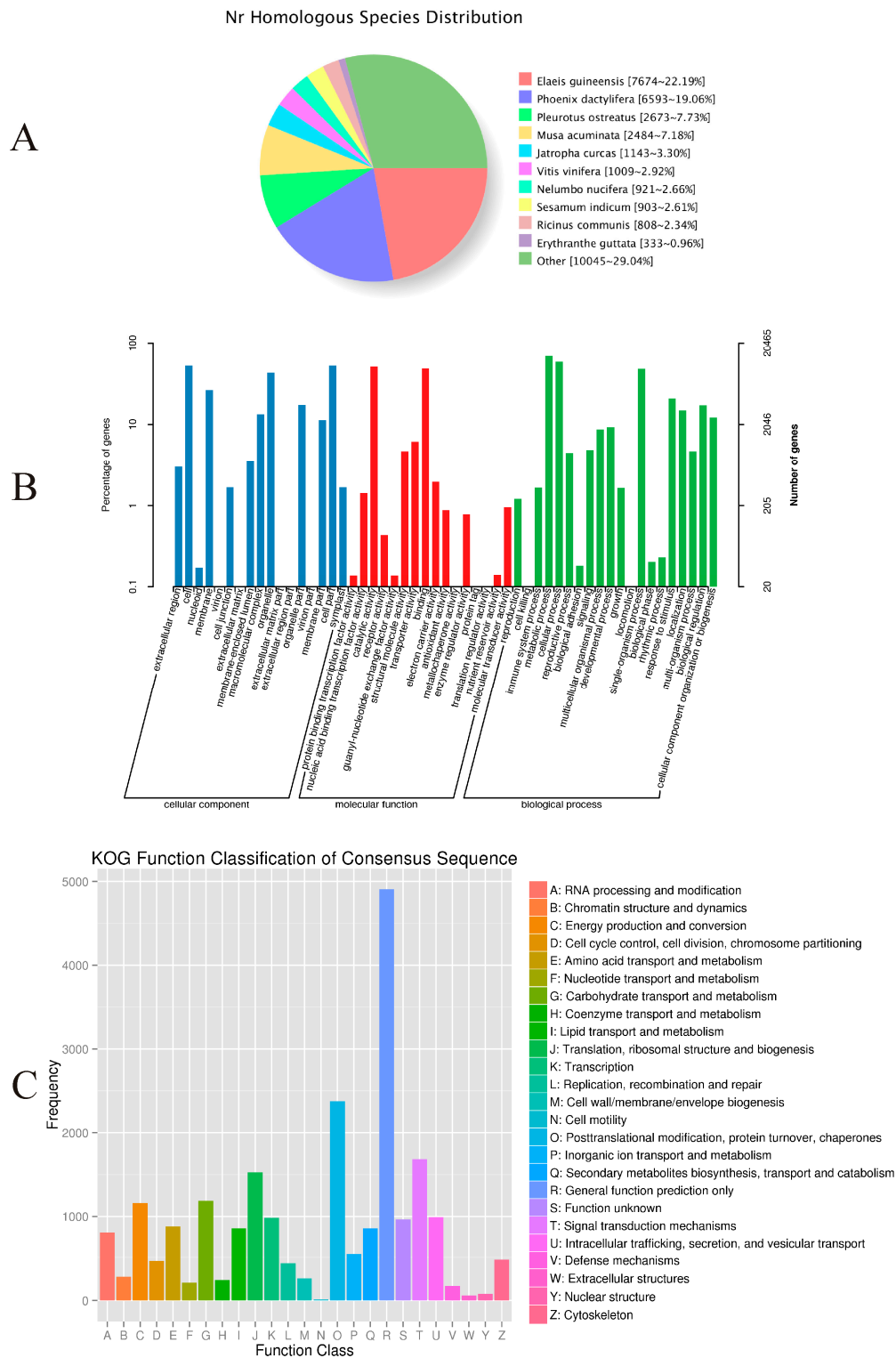
Table 3. Summary statistics of functional annotations for *Polygonatum sibiricum* unigenes via public databases.

Database	Number Annotated	300 bp < Length Number of Unigenes < 1000 bp	Number Longer than 1000 bp
Clusters of Orthologous Groups (COG)	12,175 (16.42%)	4011 (5.41%)	6892 (9.30%)
Gene Ontology (GO)	20,465 (27.61%)	8253 (11.13%)	10,444 (14.09%)
Kyoto Encyclopedia of Genes and Genomes (KEGG)	14,267 (19.25%)	5748 (7.75%)	6797 (9.17%)
euKaryotic Orthologous Groups (KOG)	20,240 (27.30%)	7785 (10.50%)	10,681 (14.41%)
Protein family (Pfam)	24,732 (33.36%)	8700 (11.74%)	14,264 (19.24%)
Swiss-Prot	21,965 (29.63%)	8574 (11.57%)	11,727 (15.82%)
non-redundant (NR)	34,601 (46.68%)	14,135 (19.07%)	17,043 (22.99%)
All annotated	35,793 (48.28%)	14,618 (19.72%)	17,140 (23.12%)

Approximately 46.68% (34,601) of the unigenes were annotated with reference to the NR database. As shown in Figure 3A, 22.19% had top matches to the sequences from *Elaeis guineensis*, 19.06% to the sequences from *Phoenix dactylifera*, and 7.73% to the sequences from *Pleurotus ostreatus*.

Among the 34,601 most significant BLASTx hits against the NR database, 20,465 unigenes could be assigned to one or more GO terms. These were assigned to 53 functional groups (Figure 3B). Among those categorized in “cellular components”, 10,797 (52.76%) were related to “cell part” (GO: 0044464), followed by “membrane” (GO: 0016020; 5395, or 26.36%). Within the molecular functions, the assignments were mostly enriched in “catalytic activity” (GO: 0003824; 10,444) and “binding” (GO: 0005488; 9905). The GO terms of biological processes were mainly grouped into “metabolic process” (GO: 0008152; 14,065) and “cellular process” (GO: 0009987; 11,945).

The COG assignments were used for further evaluation of the completeness of our *P. sibiricum* transcriptome library and the reliability of the annotation process. Overall, 12,175 (16.42%) unigenes were assigned to 25 COG categories (Figure 3C). Among these, the most numerous were unigenes belonging to the “General function prediction only” (2965, or 24.35% of the total), followed by “Replication, recombination, and repair” (1441, 11.83%), and “Translation, ribosomal structure, and biogenesis” (1426, 11.74%). Another 946 unigenes (7.77%) were assigned to “carbohydrate transport and metabolism”.



8014 (15.57%) were 1000 to 2000 bp long, 1636 (3.19%) were 2000 to 3000 bp long, and 629 (1.22%) were longer than 3000 bp.

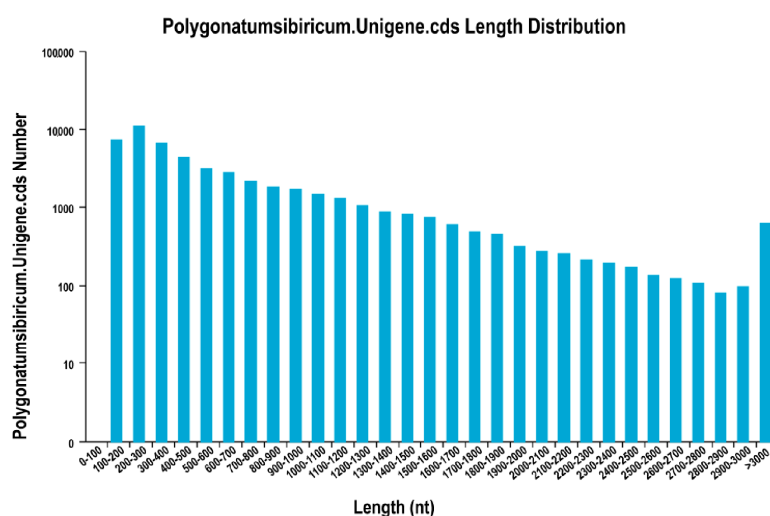


Figure 4. Distribution of lengths for coding sequences (CDSs) predicted from *Polygonatum sibiricum* unigenes.

2.4. Kyoto Encyclopedia of Genes and Genomes Pathway Analysis

To understand the biological processes in *P. sibiricum*, we mapped the unigenes annotated to the KEGG database and assigned 14,267 unigenes to 128 KEGG pathways (Table S1). Among them, 891 were mapped to ribosome pathways, 657 to “Carbon metabolism” pathways, 578 to pathways for “Biosynthesis of amino acids”, 473 for “Protein processing in endoplasmic reticulum”, and 284 for “Starch and sucrose metabolism” (Table 4). Furthermore, 1860 unigenes were mapped to the relevant pathways of polysaccharide biosynthesis, including glycolysis/gluconeogenesis, starch and sucrose metabolism, pyruvate metabolism, and amino sugar and nucleotide sugar metabolism (Table 5).

Table 4. The 20 most-represented KEGG pathways (KO, KEGG Orthology).

Pathway	Number of Unigenes	KO Entry
Ribosome	891 (6.25%)	KO03010
Carbon metabolism	657 (4.61%)	KO01200
Biosynthesis of amino acids	578 (4.05%)	KO01230
Protein processing in endoplasmic reticulum	473 (3.32%)	KO04141
Spliceosome	402 (2.82%)	KO03040
Oxidative phosphorylation	397 (2.78%)	KO00190
RNA transport	365 (2.56%)	KO03013
Glycolysis/Gluconeogenesis	304 (2.13%)	KO00010
Plant hormone signal transduction	295 (2.07%)	KO04075
Starch and sucrose metabolism	284 (1.99%)	KO00500
Plant-pathogen interaction	282 (1.98%)	KO04626
RNA degradation	245 (1.72%)	KO03018
Purine metabolism	243 (1.70%)	KO00230
Carbon fixation in photosynthetic organisms	236 (1.65%)	KO00710
Phagosome	233 (1.63%)	KO04145
Ubiquitin-mediated proteolysis	232 (1.63%)	KO04120
Endocytosis	229 (1.61%)	KO04144
Pyruvate metabolism	227 (1.59%)	KO00620
mRNA surveillance pathway	224 (1.57%)	KO03015
Cysteine and methionine metabolism	215 (1.51%)	KO00270

Table 5. The number of unigenes involved in polysaccharide biosynthesis in *Polygonatum sibiricum* database.

Pathway	Number of Unigenes	KO Entry
Glycolysis/Gluconeogenesis	304	KO00010
Starch and sucrose metabolism	284	KO00500
Pyruvate metabolism	227	KO00620
Amino sugar and nucleotide sugar metabolism	215	KO00520
Citrate cycle (Tricarboxylic Acid (TCA) cycle)	170	KO00020
Pentose phosphate pathway	128	KO00030
Galactose metabolism	128	KO00052
Fructose and mannose metabolism	113	KO00051
Pentose and glucuronate interconversions	98	KO00040
N-Glycan biosynthesis	74	KO00510
Other glycan degradation	65	KO00511
Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	31	KO00563
Glycosaminoglycan degradation	23	KO00531

2.5. Candidate Genes Involved in PSP Biosynthesis

To understand the biosynthesis of PSP and to identify the responsible genes, we annotated the transcripts related to the KO00500 and KO00520 pathway. Mainly based on the KEGG database, we determined the key enzyme genes involved in these pathways (Table 6).

Table 6. Number of unigenes involved in the biosynthesis of starch and sucrose in *Polygonatum sibiricum*.

Enzyme Code	Enzyme Name	Abbreviation	Number	FPKM
2.7.1.211	Phosphotransferase System	<i>scrA</i>	3	11.08
3.2.1.26	β -fructofuranosidase	<i>sacA</i>	31	58.64
5.4.2.2	Phosphoglucomutase	<i>pgm</i>	5	71.13
2.7.7.9	Uridine-diphosphate glucose pyrophosphorylase	<i>galU</i>	8	1.25
5.1.3.2	UDP-glucose 4-epimerase	<i>GALE</i>	4	5.59
1.1.1.-	UDP-D-galactose dehydrogenase	<i>UGD</i>	4	30.33
5.1.3.6	UDP-glucuronate 4-epimerase	<i>UGE</i>	8	0.66
1.1.1.22	UDP-glucose 6-dehydrogenase	<i>UGDH</i>	9	2.18
AXS	UDP-apiiose/xylose synthase	<i>AXS</i>	3	1.75
5.1.3.5	UDP-arabinose 4-epimerase	<i>UXE</i>	5	17.67
4.2.1.76	UDP-glucose 4,6-dehydratase	<i>RHM</i>	10	0.54
5.1.3.-, 1.1.1.-	3,5-epimerase-4-reductase	<i>UER1</i>	4	0.46
5.3.1.9	Glucose-6-phosphate isomerase	<i>GPI</i>	7	0.73
2.7.1.1	Hexokinase	<i>HK</i>	13	53.76
2.7.1.4	Fructokinase	<i>scrK</i>	8	4.45
5.3.1.8	Mannose-6-phosphate isomerase	<i>MPI</i>	3	38.76
5.4.2.8	Phosphomannomutase	<i>PMM</i>	1	10.15
2.7.7.13	Mannose-1-phosphate guanylyltransferase	<i>GMPP</i>	8	144.93
4.2.1.47	GDP-mannose 4,6-dehydratase	<i>GMDS</i>	3	0.28
1.1.1.271	GDP-L-fucose synthase	<i>TSTA3</i>	1	9.57

FPKM: Fragments Per Kilobase of transcript per Million mapped reads.

Plant polysaccharides are formed by the active nucleotide-diphospho-sugar (NDP-sugar) precursors, which are added to the residues of polysaccharides and glycoconjugates by the action of various glycosyltransferases (GTs) [12,13]. According to previous studies [12–20], PSP biosynthesis can be divided into three main stages. Firstly, sucrose is converted to Glc-1P and Fru-6P. During these processes, many enzymes play their roles, such as the β -fructofuranosidase (encoded by *sacA*) [14] that converts sucrose to Glc-6P and Fructose, and Phosphoglucomutase (encoded by *pgm*) that isomerizes Glc-6P to Glc-1P [15]. Hexokinase (encoded by *HK*) [16] and Fructokinase (encoded by *scrK*) [17] also take part in the biosynthesis of Fru-6P. Secondly, uridine diphosphate glucose (UDP-Glc) is derived from Glc-1P immediately [18], and Fru-6P is converted to GDP-Man indirectly [19]. Based on UDP-Glc and guanosine diphospho mannose (GDP-Man), other NDP sugars are further converted through the action of NDP-sugar interconversion enzymes (NSEs) [20], such as UDP-glucose 4-epimerase (*GALE*), UDP-D-galactose dehydrogenase (*UGD*), UDP-glucuronate 4-epimerase (*UGE*), UDP-glucose

6-dehydrogenase (*UGDH*), UDP-arabinose 4-epimerase (*UXE*), UDP-glucose 4,6-dehydratase (*RHM*), 3,5-epimerase-4-reductase (*UER1*), and GDP-mannose 4,6-dehydratase (*GMDS*). Finally, various NDP-sugars form growing polysaccharide chains by the action of glycosyltransferases (GTs).

In plants, UDP-Glc is a key precursor of nucleoside diphosphate (NDP)-sugars, and it is derived from 1-P-glucose (Glu-1-P) by the action of Uridine-diphosphate glucose pyrophosphorylase (encoded by *galU*) [18,21]. In the *P. sibiricum* library, Unigene 124,676 was identified as a homologue of *galU*. This 1153 bp unigene contained a complete open reading frame, and shared the highest amino acid identity (86%) with *galU* from *Ricinus communis* (Accession Number: XM_002526548.2) (Figures S1 and S2). Sugars produced by plants, such as the hexose sugars glucose and fructose, serve as the basic material for the production and accumulation of organic compounds in plants. The first step of hexose sugars' metabolism is to phosphorylate them. Currently, only two kinds of plant enzymes, hexokinases (HKs) and fructokinases (*scrKs*), have been found to catalyze the conversion of glucose and fructose. Here, we found that the 2017 bp Unigene 137,389 was a *HK* homologue that shared the highest amino acid identity (72%) with *HK* from *Phoenix dactylifera* (Accession Number: XM_008797620.2) (Figures S1 and S3). Furthermore, the 1803 bp Unigene 107,236 was a homologue of *scrK* from *Elaeis guineensis* (Accession number: XM_010934340.2) (Figures S1 and S4).

An examination of the monosaccharide composition (Table S2) showed that PSPs are formed by the polymerization of NDP-sugars such as UDP-Glc, uridine diphosphate galactose (UDP-Gal), uridine diphosphate arabinose (UDP-Ara), uridine diphosphate rhamnose (UDP-Rha), GDP-Man, and guanosine diphosphate fucose (GDP-Fuc). UDP-Glc is derived from Glc-1P directly, and GDP-Man from Fru-6P indirectly, and the others are derived from UDP-Glc and GDP-Man through the actions of NSEs [20]. We identified eight subclasses of NSEs in *P. sibiricum*, including *GALE*, *UGD*, *UGE*, *UGDH*, *UXE*, *RHM*, *UER1*, and *GMDS*. In all, 47 transcripts encoding NSEs were found in this species (Table 6, Table S3). By comparison of the fragments per kilobase of transcript per Million mapped reads (FPKM) from different RNA-Seq libraries, the most abundant transcript was for mannose-1-phosphate guanylyltransferase (*GMPP*), followed by *pgm* (phosphoglucomutase) and *sacA* (β -fructofuranosidase).

Protein encoded by *pgm* is responsible for the production of Glc-1P, using Glc-6P as a precursor. The enzyme encoded by *UGDH* converts UDP-Glu into UDP-GluA, while the enzyme encoded by *GMPP* converts Man-1P into GDP-Man. *SacA* converts sucrose into fructose and Glc-6P. The relatively higher abundance of such transcripts determined here suggests that the biological process of conversion from Glc-6P to Glc-1P should be very active.

2.6. Analysis of Differential Gene Expression in Lueyang, Shaanxi (SXLY) and Luoyang, Henan (HNLY) Germplasms

We used DESeq to analyze differences in gene expression between SXLY and HNLY. In the screening process, the false discovery rate (FDR) was set at <0.01 and the fold change (FC) was >2. The Volcano Plot presented a clear visual of the relationship between the FDR and FC for all genes of interest, enabling us to view expression levels quickly. As shown in Figure 5A, 435 genes were upregulated and 494 were downregulated. These upregulated and downregulated (929) genes were assigned to 50 functional groups (Figure 5B). Among those categorized in "biological processes", 132 were related to "metabolic process" (GO: 0008152), followed by "cellular process" (GO: 0009987; 108). Within the molecular functions, the assignments were mostly enriched in "catalytic activity" (GO: 0003824; 146) and "binding" (GO: 0005488; 112). The GO terms of cellular components were mainly grouped into "cell part" (GO: 0044464; 88) and "organelle" (GO: 0043231; 60).

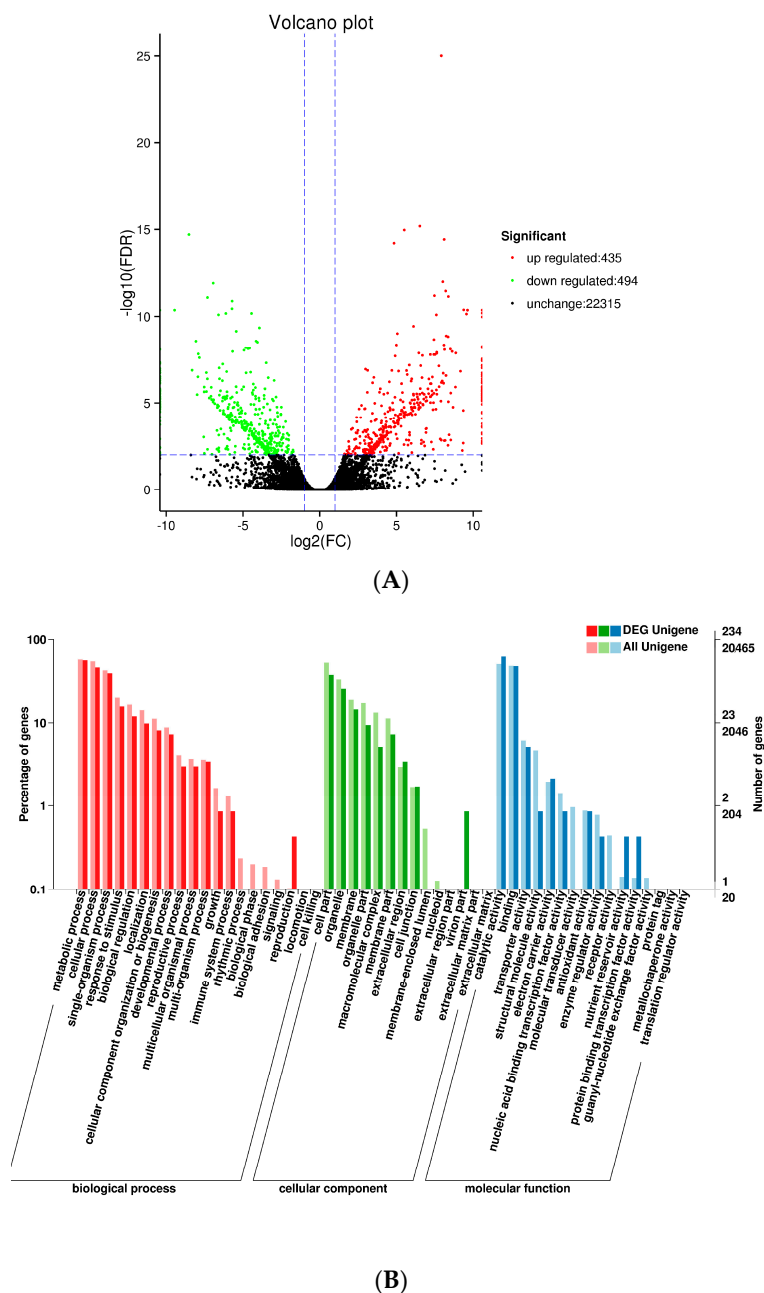


Figure 5. Volcanic map and GO classifications of differentially expressed genes. **(A)** Volcanic plot map of unigenes; **(B)** GO classifications of DEG unigenes. FDR: false discovery rate; DEG: differentially expressed gene.

Among these upregulated and downregulated genes, we focused more on those involved in the pathway of PSP biosynthesis. Based on the abundance of their transcripts, as determined through RNA-seq, we selected eight candidate unigenes: *sacA*, *GALE*, *UGDH*, *UXE*, *RHM*, *scrK*, *GMPP*, and *HK* (Figure 6). Those sharing the least identity with their homologs in the dataset were *scrK* (Unigene 107,236; 53% identity) and *UGDH* (Unigene 153,987; 64% identity). Percent identities for the others—*sacA* (Unigene 157,362), *GALE* (Unigene 51,042), *UXE* (Unigene 153,844), *RHM* (Unigene 142,186), *HK* (Unigene 137,389), and *GMPP* (Unigene 151,473)—ranged from 72% to 99%. Details for all of these alignments are shown in Table S4. This analysis primarily confirmed that our chosen potential unigenes are rather conservative and responsible for PSP biosynthesis.

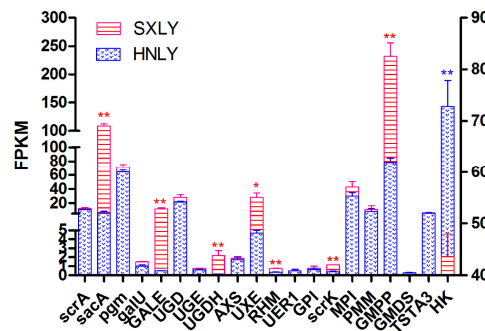


Figure 6. FPKM analysis of candidate unigenes involved in PSP biosynthesis. For a given gene, expression is significantly different between germplasms at * $p < 0.05$ and ** $p < 0.01$. Values on y-axis along right side of figure pertain to *HK*; values along left side are for all other genes.

2.7. Analysis of PSP Biosynthetic Pathway

Based on our results that identified constituents in the polysaccharide biosynthesis pathway and monosaccharide composition (Table S2), we outlined potential biosynthetic pathways for PSP formation from sucrose (Figure 7). In this scheme, the biosynthesis of PSP comprises three main steps. First, Glc-1P is converted to UDP-Glc immediately, and Fru-6P to GDP-Man indirectly, and then converted in the second step into other NDP-sugars by NSEs. Finally, the NDP-sugars are added to the polysaccharide by several GTs. Various color blocks represent the logarithms of FPKM values for different samples, using normalized data processing and plotting on a heatmap produced with R software. The first three colors in that figure are for SXLY samples; the latter three, for HNLY samples. The key enzymes are highlighted in red and are responsible for the biosynthesis of Glc-6P (*sacA*), Fru-6P (*scrK*), UDP-Gal (*GALE*), UDP-GlcA (*UGDH*), UDP-Ara (*UXE*), UDP-4-keto-6-Deoxu-D-Glc (*RHM*), fructose (*HK*), and GDP-Man (*GMPP*). Whereas expression of *sacA*, *scrK*, *GALE*, *UGDH*, *RHM*, *GMPP*, and *UXE* are positively correlated to the PSP content, that of *HK* is negatively correlated.

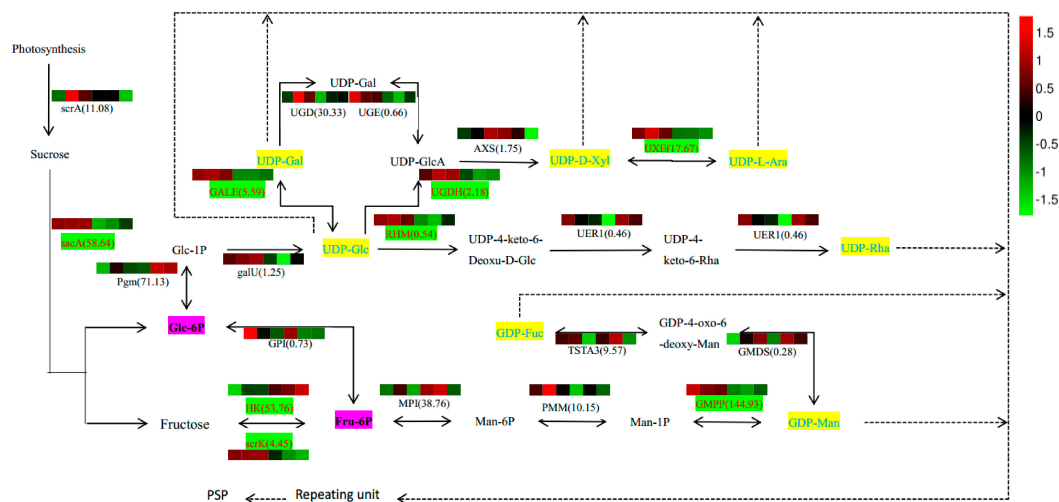


Figure 7. Proposed pathways for polysaccharide biosynthesis in *Polygonatum sibiricum*. Activated monosaccharide units, marked in blue with yellow background; key enzymes, marked in red with green background. Bolded text with pink background indicates key intermediates, while corresponding FPKM values are shown in brackets. Various color blocks represent logarithm of FPKM values. First three colors are SXLY samples; the latter three, HNLY samples. The real line arrows represent the identified enzymatic reactions, and the dashed line arrows represent multiple enzymatic reactions by multiple steps.

Various NDP-sugars form a growing polysaccharide chain by the action of glycosyltransferases (GTs) (EC: 2.4.x.y). In the *P. sibiricum* library, we identified 380 unigenes encoding GTs, based on KGG annotations. Among them, 18 unigenes are UDP glycosyltransferases (UGTs) (Table S5). In general, approximately 48% of UGTs have a plant secondary product glycosyltransferase (PSPG) box [22]. This motif contains 44 conserved amino acids. We used WebLogo (<http://weblogo.berkeley.edu/logo.cgi>) to examine the typical PSPG amino acid sequences for these 18 UGTs. As shown in Figure 8, the font size represents the intensity of the conserved properties for amino acid residues. That is, the 1st and 22nd tryptophan (W); 3rd, 34th, and 39th proline (P); 4th and 44th glutamine (Q); 8th leucine (L); 14th, 21st, and 32nd glycine (G); and 19th histidine (H) are more conserved than any of the other amino acids in the motif. A previous study of the structure-activity relationship in UGTs has shown that the N-terminal domain is the acceptor binding site and the C-terminal domain is the donor response site [23].

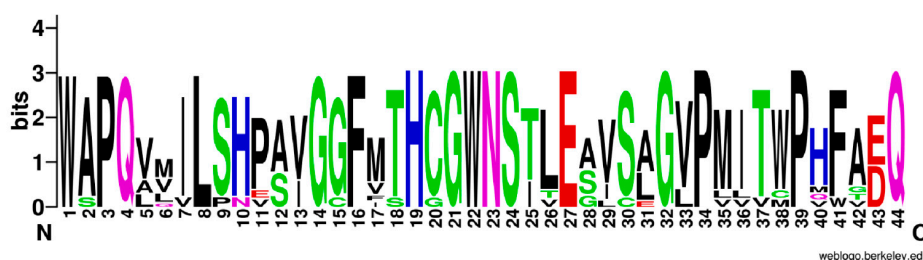


Figure 8. Consensus sequence defining glycosyltransferase plant secondary product glycosyltransferase (PSPG) motif of 18 UDP glycosyltransferases (UGTs) in *P. sibiricum*. The larger font size the more conserved residues.

2.8. Real-Time PCR Analysis

Performing a qRT-PCR analysis enabled us to determine the abundance of transcripts of genes for enzymes related to PSP biosynthesis. After selecting the reference genes in *P. sibiricum* (Table S6), we chose β -Tubulin (*TUB*) as our internal reference. We then screened eight genes as candidate unigenes based on the significantly broad differences in their levels of expression between germplasms (Figure 6). Among these candidate unigenes, we predicted that those encoding *sacA*, *scrK*, and *HK* should function upstream of the proposed pathway for PSP biosynthesis (Figure 7), while those encoding *GALE*, *RHM*, *UGDH*, *UXE*, and *GMPP* should act downstream.

We also examined the relative expression of these eight candidate genes in four germplasms: SXLY, HNLY, SXLB, and SXDF (Figure 9). That of *sacA*, *GALE*, *UGDH*, *UXE*, *RHM*, and *scrK* was positively correlated with polysaccharide content, while the relative expression of *GMPP* and *HK* was negatively correlated. By comparing the results of qRT-PCR and RNA-seq, we found that only the expression pattern of *GMPP* was inconsistent between the two results. The enzymes encoded by *GMPP* can convert Man-1P into GDP-Man in both directions. Maybe too much accumulation of PSP can cause *GMPP* negative feedback regulation, leading to the opposite results between the RNA-seq and qRT-PCR.

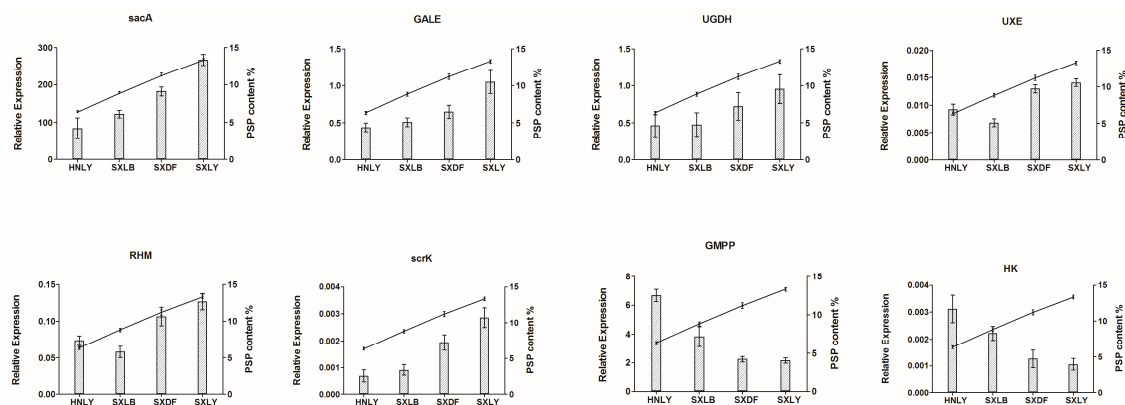


Figure 9. The expression of candidate unigenes involved in PSP biosynthesis between different germplasms. *y*-Axis on left side of each chart indicates the level of relative expression based on results from real-time PCR. Error bars show SD ($n = 9$). *y*-Axis on right side of each chart indicates % PSP content.

3. Discussion

As far as we know, this is the first time *de novo* sequencing of *Polygonatum sibiricum* has been conducted that has utilized the Illumina HiSeq 2500 platform. Our sequencing produced 72.06 Gb of clean data and 74,130 unigenes after assembly, making it possible to understand PSP biosynthesis at the molecular level. When compared with datasets published for other medicinal plants, *P. sibiricum* has more unigenes than *Neolamarckia cadamba* (55,432), but fewer than *Erigeron breviscapus* (182,527) and *Gentiana rigescens* (76,717) [24–26]. However, the unigenes obtained here average 900.9 bp, which is longer than those from *N. cadamba* (803.2 bp), *E. breviscapus* (738 bp), and *G. rigescens* (757 bp). Notably, the size distribution is homogeneous and 30.36% of the unigenes (22,512) are longer than 1000 bp. This demonstrates a higher quality of assembly results and the reliability of Illumina HiSeq-2500 technology for studying medicinal plants.

Through the analysis of bioinformatics, we identified the unigenes of PSP biosynthesis. The results indicated that sucrose was the precursor of PSP. The real-time PCR results indicated that *GMPP* was more abundant than those of the other candidates. This suggested that, in *P. sibiricum*, this pathway functions as a key branch of PSP biosynthesis. When compared with the molar ratio of monosaccharide composition in PSP biosynthesis (Table S2), we found that Man had a higher ratio (0.478) than that computed for Glc (0.285), Gal (0.139), Ara (0.064), Rha (0.020), and Fuc (0.014). This showed that Man is the most important component in PSP biosynthesis. The enzyme encoded by *GMPP* converts Man-1P into GDP-Man. Our results also revealed that *GMPP* transcripts were the most abundant in the *P. sibiricum* library. It has been reported that reducing the activity of *GMPP* will seriously affect the accumulation of mannose content, and ultimately destroy the plant leaf cell wall biosynthesis [27]. So, further studies are needed to identify the molecular mechanism that modulates PSP content and monosaccharide composition.

As a key enzyme that functions upstream of PSP biosynthesis, β -fructofuranosidase (encoded by *sacA*) is responsible for the conversion of sucrose to Glc-6P and fructose. Therefore, the high abundance of Unigene 157,362 may be related to the relatively high amount of PSP. The consistent abundance pattern of those transcripts in all tested germplasms confirmed the conservation and importance of this enzyme in PSP production. Studies have shown that the increase of β -fructofuranosidase activity would help improve fructooligosaccharides (FOS) productivity and reduce the time needed for completion of FOS biosynthesis [28]. Future research should focus on elucidating the relationships among transcript abundance, enzyme activity, and PSP content.

Both hexokinase and fructokinase are key enzymes involved in PSP biosynthesis, catalyzing the conversion of fructose to Fru-6P. In tea (*Camellia sinensis*), hexokinase and fructokinase act as important

parts in the sugar signaling, carbohydrate metabolism, and response to abiotic stresses. [29]. Although hexokinase (encoded by *HK*) and fructokinase (encoded by *scrK*) function in the same branches of the PSP biosynthesis pathway, they exhibit opposite patterns of expression, thereby showing that they cooperate in that pathway. Additional investigation will enable us to improve our understanding of the mechanism(s) by which PSP biosynthesis and accumulations are regulated at the molecular level.

4. Materials and Methods

4.1. Ethics Statement

Experimental materials were collected across China, but the field studies did not involve endangered or protected species. This study was conducted at the National Engineering Laboratory for Resource Development of Endangered Chinese Crude Drugs in Northwest China, Xi'an, China.

4.2. Plant Materials

Thirteen different germplasms of *Polygonatum sibiricum* were collected throughout China in 2015. The origins of the collected samples are specified in Table S7, and all of the germplasms were authenticated by Prof. Yaping Xiao (Key Laboratory of the Ministry of Education for Medicinal Resources and Natural Pharmaceutical Chemistry, Shaanxi Normal University). The samples were deposited at the National Engineering Laboratory for Resource Development of Endangered Chinese Crude Drugs in Northwest China. For isolating and detecting the polysaccharides, we used rhizomes from these 13 germplasms. Our mRNA-sequencing analysis utilized rhizomes collected from Lueyang, Shaanxi (SXLY) and Luoyang, Henan (HNLY). Each germplasm had three biological repetitions. The germplasm samples were first washed with tap water, then dried on filter paper and cut into small pieces with low temperature before freezing with liquid nitrogen. They were stored at $-70\text{ }^{\circ}\text{C}$ prior to library preparation and RNA-sequencing (RNA-seq), which used the Illumina HiSeqTM 2500 system (BioMarker Technologies Co., Ltd, Beijing, China). Validation by qRT-PCR involved not only the SXLY and HNLY germplasms required for transcription, but also rhizomes collected from Danfeng (SXDF) ($33^{\circ}68'24''\text{ N}$, $110^{\circ}29'66''\text{ E}$) and Liuba (SXLB) ($33^{\circ}37'27''\text{ N}$, $106^{\circ}54'55''\text{ E}$), Shaanxi Province. All qRT-PCR were repeated in three biological and three technical replications. These rhizome samples were also washed and stored as described above.

4.3. Isolation and Detection of Polysaccharides

The polysaccharides were isolated from *P. sibiricum* rhizomes with a similar method which has been reported earlier [30]. Dried powders (100 g) from rhizomes of individual germplasms were mixed with petroleum ether to remove lipids and then extracted twice at $80\text{ }^{\circ}\text{C}$ (3 h each time) in boiling distilled water (1:30, *w/v*). The water extract was collected and concentrated by a rotary evaporator. Afterward, a four-fold volume of 95% ethanol was added. The mixture was held overnight at $4\text{ }^{\circ}\text{C}$ to precipitate the polysaccharide before percolation. Protein was removed from the sediment through repeated freeze-thawing before the mixture was dialyzed and lyophilized to obtain the crude polysaccharide. The weight was recorded and the content of PSP was calculated per 100 g of dry powder.

4.4. Library Preparation and Sequencing (mRNA-Seq)

Total RNA was obtained from each sample, using a total RNA Isolation Kit (TIANGEN, Beijing, China) according to the manufacturer's instructions. Qualified RNAs were used to construct an RNA library for sequencing on the HiSeq 2500 platform.

4.5. De Novo Assembly and Functional Annotation

After obtaining high-quality sequencing data, de novo assembly was performed with Trinity software [31]. We then conducted BLAST [32] to compare our unigene sequences with those

available from the databases of Non-Redundant (NR) [33]; Swiss-Prot [34]; Gene Ontology (GO) [35]; Clusters of Orthologous Groups (COG) [36]; euKaryotic Orthologous Groups (KOG) [37]; and Kyoto Encyclopedia of Genes and Genomes (KEGG) [38]. We then applied KOBAS 2.0 software [39] to obtain the unigene KEGG orthologies. Using HMMER software (Howard Hughes Medical Institute, San Francisco, CA, USA) [40], we compared our output with the Pfam database [41]. After setting the BLAST E-value parameter at less than 1×10^{-5} and the HMMER parameter E-value at less than 1×10^{-10} , we obtained the unigene annotations. Software for TransDecoder (version 2.0.1, Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia) [42] was also used to predict coding sequences and their corresponding amino acid sequences.

4.6. Analysis of Differential Gene Expression

Bowtie software (version 1.1.2, Johns Hopkins University, Baltimore, MD, USA) [43] was used to compare sequenced reads with the unigene database. Based on those results, RSEM [44] was run to estimate expression levels. We then computed the values for Fragments Per Kilobase of transcript per Million mapped reads, or FPKM [45] to determine the abundance of corresponding unigene transcripts. Software for DESeq [46] was used to compare patterns of gene expression between the two germplasm groups, SXLY and HNLY. For purposes of screening, the FDR was set at less than 0.01 and the FC at more than 2.

4.7. Identification of Genes Involved in PSP Biosynthesis

Genes involved in PSP biosynthesis were identified based on the KEGG pathway annotation dataset. This included the metabolic pathways for amino sugar, nucleotide sugar, glucose, and starch. Neighbor-Joining (NJ) phylogenetic analysis of homologs for candidate genes was completed with MEGA 6 software. All potential transcripts annotated as glycosyltransferases (GTs) were classified via BLASTx. Then we used WebLogo (<http://weblogo.berkeley.edu/logo.cgi>) to examine the typical plant secondary product glycosyltransferase (PSPG) amino acid sequences for those UDP glycosyltransferases (UGTs).

4.8. Real-Time PCR

Reference genes were selected with geNORM, NormFinder, and BestKeeper. Specific primers were designed by Premier 3.0 [47]. The details of 20 candidate genes and the necessary primers are presented in Table S8. Total RNA was extracted from the SXLY, HNLY, SXDF, and SXLB rhizomes and reverse-transcribed to cDNA with Prime-Script RT Master Mix (TaKaRa, Kyoto, Japan), following the manufacturer's instructions. Real-time PCR was performed using a 2× Sybr Green qPCR Mix (Aidlab, Beijing, China) and a Light Cycler 96 Instrument (Roche, Germany). Relative expression was calculated by the $2^{-\Delta\Delta Ct}$ method [48,49]. All qRT-PCR were repeated in three biological and three technical replications. The relative expressions were analyzed as means \pm standard deviation (SD).

5. Conclusions

To summarize, we found that *sacA*, *HK*, and *GMPP* are the key genes that encode enzymes necessary for the PSP metabolic pathway in *Polygonatum sibiricum*. Our results provide new information about PSP biosynthesis in this species at the molecular level. Furthermore, our qRT-PCR analysis of selected unigenes indicated that these candidates function within separate, but equally important, branches of that pathway. These findings should help scientists devise an approach that achieves efficient and sustainable production of PSPs and related bioactive natural products.

Supplementary Materials: Supplementary materials can be found at www.mdpi.com/1422-0067/18/9/1950/s1.

Acknowledgments: This study was financially supported by the National Natural Science Foundation of China (Nos. 32170338 and 31670299); the Innovation Funds of Graduate Programs, Shaanxi Normal University

(No. 2015CXB010); and the Shaanxi Provincial Science and Technology Department of Social Development (No. 2015SF309).

Author Contributions: Shiqiang Wang and Zhezhi Wang conceived and designed the experiments. Bin Wang, Wenping Hua, and Junfeng Niu performed the experiments. Shiqiang Wang and Bin Wang analyzed the data. Shiqiang Wang, Kaikai Dang, and Yi Qiang contributed reagents/materials/analytical tools. Shiqiang Wang drafted the manuscript, and Zhezhi Wang revised it. All authors discussed and commented on the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhao, X.; Li, J. Chemical constituents of the genus *Polygonatum* and their role in medicinal treatment. *Nat. Prod. Commun.* **2015**, *10*, 683–688. [[PubMed](#)]
2. Virk, J.K.; Kumar, S.; Singh, R.; Tripathi, A.C.; Saraf, S.K.; Gupta, V.; Bansal, P. Isolation and characterization of quinine from *Polygonatum verticillatum*: A new marker approach to identify substitution and adulteration. *J. Adv. Pharm. Technol. Res.* **2016**, *7*, 153–158. [[PubMed](#)]
3. Liu, L.; Dong, Q.; Dong, X.T.; Fang, J.N.; Ding, K. Structural investigation of two neutral polysaccharides isolated from rhizome of *Polygonatum sibiricum*. *Carbohydr. Polym.* **2007**, *70*, 304–309.
4. Zong, S.; Zeng, G.; Zou, B.; Li, K.; Fang, Y.; Lu, L.; Xiao, D.; Zhang, Z. Effects of *Polygonatum sibiricum* polysaccharide on the osteogenic differentiation of bone mesenchymal stem cells in mice. *Int. J. Clin. Exp. Pathol.* **2015**, *8*, 6169–6180. [[PubMed](#)]
5. Zhang, H.; Cao, Y.; Chen, L.; Wang, J.; Tian, Q.; Wang, N.; Liu, Z.; Li, J.; Wang, N.; Wang, X.; et al. A polysaccharide from *Polygonatum sibiricum* attenuates amyloid- β -induced neurotoxicity in PC12 cells. *Carbohydr. Polym.* **2015**, *117*, 879–886. [[CrossRef](#)] [[PubMed](#)]
6. Zeng, G.F.; Zhang, Z.Y.; Lu, L.; Xiao, D.Q.; Xiong, C.X.; Zhao, Y.X.; Zong, S.H. Protective effects of *Polygonatum sibiricum* polysaccharide on ovariectomy-induced bone loss in rats. *J. Ethnopharmacol.* **2011**, *136*, 224–229. [[CrossRef](#)] [[PubMed](#)]
7. Yang, J.X.; Wu, S.; Huang, X.L.; Hu, X.Q.; Zhang, Y. Hypolipidemic activity and antiatherosclerotic effect of polysaccharide of *Polygonatum sibiricum* in rabbit model and related cellular mechanisms. *Evid. Based Complement. Altern. Med.* **2015**, *2015*, 391065. [[CrossRef](#)] [[PubMed](#)]
8. Wu, L.H.; Lv, G.Y.; Li, B.; Zhang, Y.L.; Su, J.; Chen, S.H. Study on effect of *Polygonatum sibiricum* on Yin deficiency model rats induced by long-term overload swimming. *Zhongguo Zhong Yao Za Zhi* **2014**, *39*, 1886–1891. (In Chinese) [[PubMed](#)]
9. Xiao, M.; Zhang, Y.; Chen, X.; Lee, E.J.; Barber, C.J.; Chakrabarty, R.; Desgagné-Penix, I.; Haslam, T.M.; Kim, Y.B.; Liu, E.; et al. Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *J. Biotechnol.* **2013**, *166*, 122–134. [[CrossRef](#)] [[PubMed](#)]
10. Zhang, X.; Gereltu, B.; Zhaorigetu; Narisu. Determination of relative molecular mass and composition for *Polygonatum sibiricum* polysaccharide by high performance liquid chromatography. *Chin. J. Chromatogr.* **2005**, *23*, 394–396. [[PubMed](#)]
11. Tang, Q.; Ma, X.; Mo, C.; Wilson, I.W.; Song, C.; Zhao, H.; Yang, Y.; Fu, W.; Qiu, D. An efficient approach to finding *Siraitia grosvenorii* triterpene biosynthetic genes by RNA-seq and digital gene expression analysis. *BMC Genom.* **2011**, *12*, 1–13. [[CrossRef](#)] [[PubMed](#)]
12. Wang, Y.; Alonso, A.P.; Wilkerson, C.G.; Keegstra, K. Deep EST profiling of developing fenugreek endosperm to investigate galactomannan biosynthesis and its regulation. *Plant Mol. Biol.* **2012**, *79*, 243–258. [[CrossRef](#)] [[PubMed](#)]
13. Pauly, M.; Gille, S.; Liu, L.; Mansoori, N.; de Souza, A.; Schultink, A.; Xiong, G. Hemicellulose biosynthesis. *Planta* **2013**, *238*, 627–642. [[CrossRef](#)] [[PubMed](#)]
14. Xie, Y.; Zhou, H.; Liu, C.; Zhang, J.; Li, N.; Zhao, Z.; Sun, G.; Zhong, Y. A molasses habitat-derived fungus *Aspergillus tubingensis* XG21 with high β -fructofuranosidase activity and its potential use for fructooligosaccharides production. *AMB Express* **2017**, *7*, 1–9. [[CrossRef](#)] [[PubMed](#)]
15. Uematsu, K.; Suzuki, N.; Iwamae, T.; Inui, M.; Yukawa, H. Expression of Arabidopsis plastidial phosphoglucomutase in tobacco stimulates photosynthetic carbon flow into starch synthesis. *J. Plant Physiol.* **2012**, *169*, 1454–1462. [[CrossRef](#)] [[PubMed](#)]

16. Zhang, Y.; Zhen, L.; Tan, X.; Li, L.; Wang, X. The involvement of hexokinase in the coordinated regulation of glucose and gibberellin on cell wall invertase and sucrose synthesis in grape berry. *Mol. Biol. Rep.* **2014**, *41*, 7899–7910. [[CrossRef](#)] [[PubMed](#)]
17. Perez-Cenci, M.; Salerno, G.L. Functional characterization of *Synechococcus amylosucrase* and fructokinase encoding genes discovers two novel actors on the stage of cyanobacterial sucrose metabolism. *Plant Sci.* **2014**, *224*, 95–102. [[CrossRef](#)] [[PubMed](#)]
18. Park, J.I.; Ishimizu, T.; Suwabe, K.; Sudo, K.; Masuko, H.; Hakozaiki, H.; Nou, I.S.; Suzuki, G.; Watanabe, M. UDP-glucose pyrophosphorylase is rate limiting in vegetative and reproductive phases in *Arabidopsis thaliana*. *Plant Cell Physiol.* **2010**, *51*, 981–996. [[CrossRef](#)] [[PubMed](#)]
19. 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* **1979**, *145*, 331–337. [[CrossRef](#)] [[PubMed](#)]
20. Yin, Y.; Huang, J.; Gu, X.; Bar-Peled, M.; Xu, Y. Evolution of plant nucleotide-sugar interconversion enzymes. *PLoS ONE* **2011**, *6*, e27995. [[CrossRef](#)] [[PubMed](#)]
21. Soares, J.S.; Gentile, A.; Scorsato, V.; Lima, A.C.; Kiyota, E.; Dos, S.M.L.; Piattoni, C.V.; Huber, S.C.; Aparicio, R.; Menossi, M. Oligomerization, membrane association, and in vivo phosphorylation of sugarcane UDP-glucose pyrophosphorylase. *J. Biol. Chem.* **2014**, *289*, 33364–33377. [[CrossRef](#)] [[PubMed](#)]
22. Mamoon, R.H.; Amjad, N.M.; Bao, L.; Hussain, S.Z.; Lee, J.M.; Ahmad, M.Q.; Chung, G.; Yang, S.H. Genome-wide analysis of family-1 UDP-glycosyltransferases in soybean confirms their abundance and varied expression during seed development. *J. Plant Physiol.* **2016**, *206*, 87–97. [[CrossRef](#)] [[PubMed](#)]
23. Gachon, C.M.; Langlois-Meurinne, M.; Saindrenan, P. Plant secondary metabolism glycosyltransferases: The emerging functional analysis. *Trends Plant. Sci.* **2005**, *10*, 542–549. [[CrossRef](#)] [[PubMed](#)]
24. Chen, R.B.; Liu, J.H.; Xiao, Y.; Zhang, F.; Chen, J.F.; Ji, Q.; Tan, H.X.; Huang, X.; Feng, H.; Huang, B.K.; et al. Deep sequencing reveals the effect of MeJA on scutellarin biosynthesis in *Erigeron breviscapus*. *PLoS ONE* **2015**, *10*, e0143881. [[CrossRef](#)] [[PubMed](#)]
25. Zhang, X.; Allan, A.C.; Li, C.; Wang, Y.; Yao, Q. *De novo* assembly and characterization of the transcriptome of the chinese medicinal herb, *Gentiana rigescens*. *Int. J. Mol. Sci.* **2015**, *16*, 11550–11573. [[CrossRef](#)] [[PubMed](#)]
26. Ouyang, K.; Li, J.; Zhao, X.; Que, Q.; Li, P.; Huang, H.; Deng, X.; Singh, S.K.; Wu, A.M.; Chen, X. Transcriptomic analysis of multipurpose timber yielding tree *Neolamarckia cadamba* during xylogenesis using RNA-Seq. *PLoS ONE* **2016**, *11*, e0159407. [[CrossRef](#)] [[PubMed](#)]
27. Keller, R.; Renz, F.S.; Kossman, J. Antisense inhibition of the GDP-mannose pyrophosphorylase reduces the ascorbate content in transgenic plants leading to developmental changes during senescence. *Plant J.* **1999**, *19*, 131–141. [[CrossRef](#)] [[PubMed](#)]
28. Zhang, J.; Liu, C.; Xie, Y.; Li, N.; Ning, Z.; Du, N.; Huang, X.; Zhong, Y. Enhancing fructooligosaccharides production by genetic improvement of the industrial fungus *Aspergillus niger* ATCC 20611. *J. Biotechnol.* **2017**, *249*, 25–33. [[CrossRef](#)] [[PubMed](#)]
29. Li, N.N.; Qian, W.J.; Wang, L.; Cao, H.L.; Hao, X.Y.; Yang, Y.J.; Wang, X.C. Isolation and expression features of hexose kinase genes under various abiotic stresses in the tea plant (*Camellia sinensis*). *J. Plant Physiol.* **2017**, *209*, 95–104. [[CrossRef](#)] [[PubMed](#)]
30. Li, F.; Gao, J.; Xue, F.; Yu, X.; Shao, T. Extraction optimization, purification and physicochemical properties of polysaccharides from *Gynura medica*. *Molecules* **2016**, *21*, 397. [[CrossRef](#)] [[PubMed](#)]
31. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [[CrossRef](#)] [[PubMed](#)]
32. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [[CrossRef](#)] [[PubMed](#)]
33. Deng, Y.; Jianqi, L.I.; Songfeng, W.U.; Zhu, Y.; Chen, Y.; Fuchu, H.E. Integrated nr database in protein annotation system and its localization. *Comput. Eng.* **2006**, *32*, 71–74.
34. Apweiler, R.; Bairoch, A.; Wu, C.H.; Barker, W.C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; et al. UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* **2004**, *32*, 115–119. [[CrossRef](#)] [[PubMed](#)]

35. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–29. [[CrossRef](#)] [[PubMed](#)]
36. Tatusov, R.L.; Galperin, M.Y.; Natale, D.A.; Koonin, E.V. The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **2000**, *28*, 33–36. [[CrossRef](#)] [[PubMed](#)]
37. Koonin, E.V.; Fedorova, N.D.; Jackson, J.D.; Jacobs, A.R.; Krylov, D.M.; Makarova, K.S.; Mazumder, R.; Mekhedov, S.L.; Nikolskaya, A.N.; Rao, B.S.; et al. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* **2004**, *5*, 1–28. [[CrossRef](#)] [[PubMed](#)]
38. Kanehisa, M.; Goto, S.; Kawashima, S.; Okuno, Y.; Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **2004**, *32*, 277–280. [[CrossRef](#)] [[PubMed](#)]
39. Xie, C.; Mao, X.; Huang, J.; Ding, Y.; Wu, J.; Dong, S.; Kong, L.; Gao, G.; Li, C.Y.; Wei, L. KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* **2011**, *39*, 316–322. [[CrossRef](#)] [[PubMed](#)]
40. Eddy, S.R. Profile hidden Markov models. *Bioinformatics.* **1998**, *14*, 755–763. [[CrossRef](#)] [[PubMed](#)]
41. Finn, R.D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; et al. Pfam: The protein families database. *Nucleic Acids Res.* **2014**, *42*, 222–230. [[CrossRef](#)] [[PubMed](#)]
42. Kim, H.S.; Hwang, D.S.; Lee, B.Y.; Park, J.C.; Lee, Y.H.; Lee, J.S. De novo assembly and annotation of the marine mysid (*Neomysis awatschensis*) transcriptome. *Mar. Genom.* **2016**, *28*, 41–43. [[CrossRef](#)] [[PubMed](#)]
43. Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **2009**, *10*, 1–10. [[CrossRef](#)] [[PubMed](#)]
44. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 1–16. [[CrossRef](#)] [[PubMed](#)]
45. Trapnell, C.; Williams, B.A.; Pertea, G.; Mortazavi, A.; Kwan, G.; van Baren, M.J.; Salzberg, S.L.; Wold, B.J.; Pachter, L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **2010**, *28*, 511–515. [[CrossRef](#)] [[PubMed](#)]
46. Anders, S.; Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **2010**, *11*, 1–12. [[CrossRef](#)] [[PubMed](#)]
47. Rozen, S.; Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **2000**, *132*, 365–386. [[PubMed](#)]
48. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_t}$ Method. *Methods.* **2001**, *25*, 402–408. [[CrossRef](#)] [[PubMed](#)]
49. Schmittgen, T.D.; Livak, K.J. Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* **2008**, *3*, 1101–1108. [[CrossRef](#)] [[PubMed](#)]

