Identification and analysis of sucrose synthase gene family associated with polysaccharide biosynthesis in *Dendrobium catenatum* by transcriptomic analysis

Min Jiang [Corresp., Equal first author, 1, 2] , **Shangyun Li** [Equal first author, 1] , **Changling Zhao** [1] , **Mingfu Zhao** [1] , **Shaozhong Xu** [1] , **Guosong Wen** [Corresp. 1]

[1] Research & Development Center for Heath Product, College of Agronomy and Biotechnology, Yunnan Agricultural University, Kunming, China

[2] Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, Institute of Eco-Chongming (IEC), School of Life Sciences, Fudan University, Shanghai, China

Corresponding Authors: Min Jiang, Guosong Wen
Email address: yijinsha@126.com, wengs@ynau.edu.cn

Abstract

Background : *Dendrobium catenatum* is a valuable traditional medicinal herb with high commercial value. *D. catenatum* stems contain abundant polysaccharides which are one of the main bioactive components. However, although some genes related to the synthesis of the polysaccharides have been reported, more key genes need to be further elucidated. **Results**: In this study, the contents of polysaccharides and mannose in *D. catenatum* stems at four developmental stages were compared, and the stems' transcriptomes were analyzed to explore the synthesis mechanism of the polysaccharides. Many genes involved in starch and sucrose metabolisms were identified by KEGG pathway analysis. Further analysis found that sucrose synthase (SUS ; EC 2.4.1.13) gene maybe participated in the polysaccharide synthesis. Hence, we further investigated the genomic characteristics and evolution relationships of the *SUS* family in plants. The result suggested that the SUS gene of *D. catenatum* (*DcSUS*) had undergone the expansion characterized by tandem duplication which might be related to the enrichment of the polysaccharides in *D. catenatum* stems. Moreover, expression analyses of the *DcSUS* displayed significant divergent patterns in different tissues and could be divided into two main groups in the stems with four developmental stages. **Conclusion**: In general, our results revealed that *DcSUS* is likely involved in the metabolic process of the stem polysaccharides, providing crucial clues for exploiting the key genes associated with the polysaccharide synthesis.

1 **Identification and analysis of sucrose synthase gene family associated with**

2 **polysaccharide biosynthesis in *Dendrobium catenatum* by transcriptomic**

3 **analysis**

4 Min Jiang[1,2,#,*], Shangyun Li[1,#], Changling Zhao[1], Mingfu Zhao[1], Shaozhong Xu[1], Guosong Wen[1*]

5 [1]Research & Development Center for Heath Product, College of Agronomy and Biotechnology,
6 Yunnan Agricultural University, Kunming 650201, Yunnan, China
7 [2]Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering,
8 Institute of Eco-Chongming (IEC), School of Life Sciences, Fudan University, Shanghai 200438,
9 China
10 * Correspondence: Min Jiang: (20110700001@fudan.edu.cn) and Guosong Wen
11 (wengs@163.com)

12 # These authors contributed equally to this work.

13 **Author E-mail:**

14 Min Jiang: 20110700001@fudan.edu.cn, yijinsha@126.com, ORCID: 0000-0002-0757-9217

15 Shangyun Li: lsy2574357728@163.com

16 Changling Zhao: zhaoplumblossom7@163.com

17 Mingfu Zhao: zhaomingfu@163.com

18 Shaozhong Xu: 2737308675@qq.com

19 Guosong Wen: wengs@ynau.edu.cn

20 **Running Head:** Transcriptomic and phylogenetic analysis of *SUS* in *Dendrobium catenatum*

21 **Abstract**
22 **Background**: *Dendrobium catenatum* is a valuable traditional medicinal herb with high
23 commercial value. *D. catenatum* stems contain abundant polysaccharides which are one of the
24 main bioactive components. However, although some genes related to the synthesis of the
25 polysaccharides have been reported, more key genes need to be further elucidated.

26 **Results**: In this study, the contents of polysaccharides and mannose in *D. catenatum* stems at four
27 developmental stages were compared, and the stems' transcriptomes were analyzed to explore the
28 synthesis mechanism of the polysaccharides. Many genes involved in starch and sucrose
29 metabolisms were identified by KEGG pathway analysis. Further analysis found that sucrose
30 synthase (SUS; EC 2.4.1.13) gene maybe participated in the polysaccharide synthesis. Hence, we
31 further investigated the genomic characteristics and evolution relationships of the *SUS* family in
32 plants. The result suggested that the SUS gene of *D. catenatum* (*DcSUS*) had undergone the

33 expansion characterized by tandem duplication which might be related to the enrichment of the
34 polysaccharides in *D. catenatum* stems. Moreover, expression analyses of the *DcSUS* displayed
35 significant divergent patterns in different tissues and could be divided into two main groups in the
36 stems with four developmental stages.

37 **Conclusion**: In general, our results revealed that *DcSUS* is likely involved in the metabolic process
38 of the stem polysaccharides, providing crucial clues for exploiting the key genes associated with
39 the polysaccharide synthesis.

40 **Keywords:** SUS, Mannose, RNA-Seq, Evolution, Expression analysis, *Dendrobium catenatum*

41 **Introduction**
42 The genus *Dendrobium* is a perennial medicinal herb with about 1450 species, and widely
43 distributes in tropical and subtropical areas such as Australia and New Guinea (Xiang et al., 2013,
44 Zhang et al., 2016a). *D. catenatum*, an endangered orchid in the wild, has been served as a folk
45 medicine nourishing "Yin", relieving fevers and stomach upsets and enhancing immunity for
46 hundreds of years in China (Jiang et al., 2017, Ma et al., 2018). *D. catenatum* is also called
47 *Dendrobium officinale* under the Chinese name "Tiepishihu" and was recorded in the 2010 edition
48 of the Chinese Pharmacopoeia (Li et al., 2017).

49 The fleshy stems, i.e., the main medicinal parts, of *D. catenatum* have abundant polysaccharides
50 which are considered to be the main bioactive ingredients of the stems (Zhang et al., 2016b), and
51 possess immunomodulatory, antioxidant and hepatoprotective activities (Ng et al., 2012).
52 Therefore, the content of the polysaccharides is the main market indicator of the stem quality
53 (Meng et al., 2013). The polysaccharides consist of mannose, glucose and arabinose, and exist in
54 the stems with the form of 2-O-acetylglucomannan (Hua et al., 2004). Indeed, fructose and
55 mannose are the basic building units for the polysaccharide synthesis. Cellulose synthase (CESA)
56 gene is associated with the mannan synthesis in *D. catenatum* ((He et al., 2015), while sucrose
57 synthase (SUS; EC 2.4.1.13) is the main enzyme which is involved in sucrose metabolism (Koch
58 et al., 1996, N'tchobo et al., 1999), and catalyzes the reversible conversion of sucrose and UDP to
59 UDP-glucose and fructose (Nunez et al., 2008, Barratt et al., 2009). Moreover, SUS interacts with
60 CESA as a complex to supply UDP-glucose for cell wall synthesis (Haigler et al., 2001, Ruan et
61 al., 2003). The activities of sucrose invertase and sucrose-phosphate synthase (SPS) are correlated
62 with polysaccharide levels (Wang et al., 2013), and sucrose breakdown is largely catalyzed by
63 SUS and invertase (Huber and Huber, 1996). Although this is a general principle, rather than
64 specific to *D. catenatum*, in this case, SUS and invertase should hold some correlations with the
65 polysaccharide levels in *D. catenatum*. What's more, *SPS* and *SUS* genes have been reported to be
66 related to polysaccharide generation (Yan et al., 2015), and, in higher plants, sucrose degradation
67 catalyzed by SUS provides the glycosyl needed in the polysaccharide synthesis (Bar-Peled and
68 O'Neill, 2011). Hence, it is reasonable to speculate that the *SUS* is likely involved in the
69 polysaccharide synthesis in *D. catenatum*.

70 SUS is encoded by a small multiple gene family that exhibits distinct, partially overlapping
71 expression patterns and functional divergences. Identification and characterization of the *SUS*s in

plants such as *Arabidopsis thaliana* (Bieniawska et al., 2007), rice (Hirose et al., 2008), *Populus* (Zhang et al., 2011), apple (Tong et al., 2018), and pear (Abdullah et al., 2018), are helpful for understanding the physiological roles and metabolic processes of the plants. The six SUSs encoded by Arabidopsis genome are divided into three groups based on their phylogenetic relationship and genomic structures (Baud et al., 2004, Bieniawska et al., 2007). In detail, the *AtSUS1* and *AtSUS2* display differential ABA-independent expressions at sugar/osmoticum levels (Dejardin et al., 1999), and the *SUS1* and *SUS4* are all involved in the tolerance under hypoxic conditions (Bieniawska et al., 2007). During the maturation phase, the *SUS2* has highly specific expression in seeds and co-localized with the plastids in embryos (Nunez et al., 2008). Moreover, the SUS2 and SUS3 can also alter sucrose/hexose homeostasis and affect carbon partitioning and storage in developing seeds (Angeles-Nunez and Tiessen, 2010), and the SUS5 and SUS6 have C-terminal extensions relative to other isoforms and play an important role in callose synthesis in sieve plates (Barratt et al., 2009).

To date, although the key enzyme genes underlying the polysaccharide synthesis and metabolic pathway have been reported in several transcriptomes of *D. catenatum* (He et al., 2015, Meng et al., 2016, Zhang et al., 2016b, Shen et al., 2017), the related information of the genes remain limited. In this study, we investigated the mannose and polysaccharide contents in *D. catenatum* stems with four developmental stages and established the corresponding transcription databases. We also identified and characterized the *SUS* family involved in the polysaccharide synthesis. Eventually, the expression profiles of the *SUS*s in the stems were also surveyed. Our research could be helpful to further decipher the molecular mechanism of the bioactive polysaccharide biosynthesis and utilize genetic engineering to obtain abundant bioactive polysaccharides from *D. catenatum*.

Material and Methods

Plant materials and growth conditions

At the plantation of Yunnan Honghe Qunxin Shihu Planting Co. Ltd. of China (22°36'7" N, 103°27'36" E; Average altitude: 1300 m; Row spacing: 12 cm ×15 cm), the stems with four developmental stages, i.e., S1 (about 2-3 months after sprouting), S2 (about 5-6 months after sprouting), S3 (about 8-9 months after sprouting) and S4 (about 11-12 months after sprouting) (Table 1; Fig. S1), were sequentially collected from the plant population of *D. catenatum* 'Hongxin 6#' which were from the clone of a wild *D. catenatum* plant. The stems were divided into 2 batches. The first used for the determination of polysaccharide and mannose contents were dried at 105 °C in an oven for 10 h and then triturated by a DFT-50 pulverizer and the second, together with the roots, leaves, pedicels, dorsal sepals, lateral sepals, lips and columns, used for the RNA extraction were rapidly frozen in liquid nitrogen and then stored at -80℃.

Determination of polysaccharide and mannose contents in the stems

The polysaccharide contents in the stems with the four developmental stages detected by SGS (SGS-CSTC Standard Technical Services (Shanghai) Co., Ltd, China) were determined by using the phenol-sulfuric acid method described by the Pharmacopoeia Committee of the People's

111 Republic of China with glucose solutions (18, 36, 54, 72 and 90 μg/mL) as standards. 0.3 g stem
112 powder was added with 1 mL of 5% phenol and 5 mL concentrated sulfuric acid, heated in a
113 boiling water bath for 20 min. Finally, the absorbance at 488 nm of the reaction solution was
114 determined with a UV-6000 spectrophotometer (Shanghai Metash, Shanghai, China). Meanwhile,
115 the reaction solution was diluted to 2 mL with distilled water as the calibration standard. Each
116 sample was assayed as three replicates.

117 For the determination of mannose contents, high performance liquid chromatography (HPLC)
118 assays were executed with octadecylsilyl (ODS) as the filler described by the Pharmacopoeia
119 Committee of the People's Republic of China. 0.12 g stem powder was pre-extracted with 80%

120 ethanol at 80℃ for 4 h. Then, 100 mL double-distilled water were added to the residues and mixed

121 with 1 mL internal standard (12 mg/mL, d-glucosamine hydrochloride, chromatographically pure,
122 Sigma-Aldrich) at 100 °C for 1 h. Subsequently, 0.5 mL HCl (3.0 M) was mixed with the former

123 solution and hydrolyzed at 110℃ for 1 h. Then, 1-phenyl-3-methyl-5-pyrazolone (PMP) was used

124 for derivation, incubated at 70℃ for 110 min and neutralized by 0.5 mL HCl (0.3 M). The PMP

125 labeling reaction solution was prepared by mixing with 0.3 M NaOH solution (0.4 mL) and 0.5 M
126 PMP methanol solution (0.4 mL). To remove proteins, the sample was extracted with 2 mL
127 chloroform for three times, mixed thoroughly for 2 min and centrifuged at 12,000 rpm for 5 min.
128 Finally, the aqueous phase was gathered and detected by HPLC under the following conditions:
129 ZORBAX SB-Aq C (18) column (4.6 mm × 250 mm, 5 μm); acetonitrile–0.5 % ammonium acetate
130 solution (20:80, v/v) as the mobile phase; flow rate = 1.0 mL min^{-1}; detection wave length = 250
131 nm. Three biological replicates for each sample were used for the determination.

132 **Digital gene expression library construction, sequencing and assembly**

133 Total RNA (3 μg) from the stems were extracted by using Trizol reagent (Invitrogen, CA, USA)
134 following the manufacturer's recommendations, and the RNA integrity was measured by using
135 Bioanalyzer 2100 system with RNA Nano 6000 LabChip Kit (Agilent Technologies, CA, USA)
136 with RIN number >7.0. Sequencing libraries were constructed by using NEBNext® Ultra™ RNA
137 Library Prep Kit for Illumina® (NEB, USA) according to manufacturer's protocol and index codes
138 were appended to attribute sequences to each sample. We performed only two biological replicates
139 on each sample. The libraries were named S1-1, S1-2, S2-1, S2-2, S3-1, S3-2, S4-1, S4-2,
140 respectively. Subsequently, the index-coded samples were clustered by cBot Cluster Generation
141 System and then the libraries were sequenced by using Illumina Hiseq platform and 150 bp paired-
142 end reads were generated following the manufacturer's protocol. All raw data were cleaned by
143 eliminating the reads containing the adapters and ploy-N and the reads with low quality, which
144 were mapped to reference genome of *D. catenatum* downloaded in the Herbal Medicine Omics
145 Database (http://herbalplant.ynau.edu.cn/)(Yan et al., 2015). Index of the reference genome was
146 built by using Bowtie v2.2.3 (Langmead and Salzberg, 2012) and paired-end clean reads were
147 aligned to the reference genome by using TopHat v2.0.12. We selected TopHat as the mapping

148   tool because TopHat can generate a database of splice junctions based on the gene model
149   annotation file and, thus, produce a better mapping result than other non-splice mapping tools. The
150   rate of unique mapping all reached to 60% in all samples (Table 2). The square of correlation
151   coefficients between replicates of each sample were more than 75% (Fig. S2). These raw data
152   generated in this study had been deposited in the National Center for Biotechnology Information
153   (NCBI) Short Read Archive (SRA) under BioProject ID PRJNA668448.

154   **Differentially expressed genes (DEGs) and functional annotation**

155   The calculation of the statistical power for our RNA-seq data was performed by RNASeqPower
156   Calculator (Ching et al., 2014). The abundance of gene expression was calculated by FPKM
157   (Fragments Per Kilobase of transcript sequence per Million mapped reads) (Wagner et al., 2012).
158   The differential expression analysis between two randomly samples of developmental stages was
159   executed by using the DESeq R package (1.18.0)(Anders and Huber, 2010). The DEGs were
160   screened by EdgeR packege with fold change ≥2 and the false discovery rate (FDR) adjusted p-
161   value <0.05 as the threshold (Robinson et al., 2010). For gene annotation, Gene Ontology (GO)
162   enrichment analysis of DEGs was implemented by the GOseq R package with the corrected p-
163   values less than 0.05 as cut off (Young et al., 2010). In addition, the Kyoto Encyclopedia of Genes
164   and Genomes (KEGG) pathway enrichment analysis of DEGs was performed by using the KOBAS
165   software (Kanehisa et al., 2008). Likewise, the corrected p-value <0.05 was considered as
166   statistically significant difference.

167   **Identification of the *SUS* family members in plants**

168   The 15 *SUS* information from the *D. catenatum* were downloaded from the Herbal Medicine
169   Omics Database (http://herbalplant.ynau.edu.cn/) (Yan et al., 2015). BLASTP searches was
170   proposed against orthologous protein sequences using Arabidopsis SUSs (Bieniawska et al., 2007)
171   as queries in public PLAZA (http://bioinformatics.psb.ugent.be/plaza/)(Van Bel et al., 2018).
172   Meanwhile, all retrieved gene sequences were only considered as the candidates and subjected to
173   domain analyses by scanning in InterProscan software (de Castro et al., 2006). Importantly, the
174   proteins only harboring sucrose synthase domain (IPR000368) rather than sucrose-phosphatase
175   domain (IPR006380) were considered as the SUS ones. Moreover, redundancy and any alternative
176   splice variants of sequences were eliminated. The SUSs in given species was designated according
177   to their order on the chromosomes.

178   **Gene structure, sequence alignment and phylogenetic analysis**

179   All SUS protein sequences were compared with those of Arabidopsis and *D. catenatum* by using
180   the Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) with default settings. The
181   theoretical pI (isoelectric point) and Mw (molecular weight) of the SUSs were calculated by using
182   Compute pI/Mw tool online software (http://web.expasy.org/compute_pi/).The exon/intron
183   structures of the *SUS* candidates were generated by using the online Gene Structure Display Server
184   (GSDS 2.0: http://gsds.cbi.pku.edu.ch) (Hu et al., 2015) with output in accordance with their
185   phylogenetic tree. Phylogenetic trees of the SUSs were constructed by using MEGA 6.0 based the

186 maximum-likelihood (ML) method with a Jones-Taylor-Thornton (JTT) model same as previous
187 research (Tamura et al., 2013). Test of phylogeny was assessed by bootstrap method with 2000
188 iterations test and all positions with 95% site coverage were eliminated.

189 **Quantitative real-time PCR (qRT-PCR) analyses**

190 To assess the veracities of the RNA-Seq data, qRT−PCR amplification of 44 genes involved in the

191 stem polysaccharide synthesis in four developmental stages was performed. The total RNA was
192 extracted by using Trizol reagent and reverse-transcribed into cDNA by using PrimeScript RT
193 Master Mix Perfect Real Time (TaKaRa) followed the manufacturer's instructions. The qRT-PCR
194 was executed by using SYBR®Premix Ex TaqTMII (TaKaRa) and each sample was repeated in
195 triplicate independently. 10 μL reaction systems containing 5 -50 ng of cDNA products (4 μL), 5
196 pmol of each primer (0.4 μL), 5 μL SYBR green master mix (2X), 0.2 μL ROX normalized
197 fluorescent signal. The procedure for amplification was set as follows: initial activation at 95 °C
198 for 10 min, followed by 45 cycles of 95 °C for 30 s, 60 °C for 30 s, and 72 °C for 30 s. Melting
199 curves ranging from 60 °C to 95 °C followed by 0.5 °C/min were detected. The constitutively
200 expression gene, *D. catenatum* 18S rRNA (Dendrobium_GLEAN_10067105), was used as an
201 internal control. The primer sets were listed in Table S1. Moreover, the expression profiles of the
202 *DcSUSs* in different tissues and four developmental stages were also investigated. The
203 housekeeping gene *GAPDH* (NCBI accession number: KP719976) was used to normalize the
204 relative expression levels (Jiang et al., 2017). The *SUS*-specific primers were shown in Table S2.

205 The methods and conditions of the qRT−PCRs were performed as mentioned above. The relative

206 expression levels of candidate genes were calculated by using the $2^{-\Delta\Delta Ct}$ method. The up-regulated
207 genes were defined as the ones whose fold-changes greater than 2 with the p values of < 0.05, and
208 the genes with the fold changes of 0.5 or less and the p values of < 0.05 were defined as the down-
209 regulated ones.

210 **Results**

211 **Polysaccharide and mannose contents of the stems with four developmental stages**

212 To provide a scientific data for taking advantage of the endangered wild *Dendrobium* resources,
213 the contents of total polysaccharides and mannose, which are the most abundant neutral
214 monosaccharide in various developmental stages of *D. catenatum*, were determined. *D. catenatum*
215 plants had the highest polysaccharide (431 mg/g) in S4, while the highest concentration of mannose
216 in S3 (49.6%) (Table 1). From S1 to S3, the mannose accumulated gradually, which was consistent
217 with the variation trend of the polysaccharide. Interestingly, mannose content showed decreased
218 in S4, rather than a sustained increase like that of the polysaccharide (Table 1), indicating that the
219 presence of other monosaccharides had a greater impact on the polysaccharide accumulation in the
220 later stages of *D. catenatum* development.

221 **Overview of RNA-seq analysis**

222 To elucidate the genes and metabolic pathways involved in the polysaccharide synthesis, eight
223 transcriptomes from S1–S4 samples were studied. Due to irresistable factors, we performed only
224 two biological replicates on each sample instead of three replicates. A total of 128.58 million,
225 137.76 million, 135.88 million and 140.5 million raw reads were obtained for S1, S2, S3 and S4,
226 respectively (Table 2). Likewise, after filtering, 18.95 Gb, 20.28 Gb, 20.09 Gb and 20.73 Gb clean
227 reads were generated, respectively (Table 2). In addition, the clean reads were mapped to *D.*
228 *catenatum* reference genome database, and a total of 16384 novel genes and their expression data
229 were obtained (Tables 3 and S4). The statistical power of our RNA-seq data calculated by
230 RNASeqPower Calculator was 0.832.

Go classification and KEGG analysis of the DEGs among different developmental stages

232 Pairwise comparisons were made to identify the candidate genes associated with the
233 polysaccharide synthesis during development. Based on the RNA-Seq data, the distribution rang
234 of FPKM in S1-S4 were mostly located between of 1 and 60 (Fig S3; Table S4). The detailed
235 information of the FPKM values was enclosed in Table S5. In addition, a total of 1762 significant

236 DEGs were screened with the p values of ⩽0.05 and the 2-fold change differences as a criteria

237 (Fig.1a; Fig. S4). The DEGs were shown in Table S6. To investigate the major trends between the
238 different developmental samples in *D. catenatum*, these DEGs were clustered into six groups by
239 K-means methods (Fig.1b and Fig.1c). Moreover, the five data sets from different comparisons
240 were exhibited by using a Venn diagram (Fig.1d). For example, 106 DEGs were identified in both
241 S1 vs S3 and S1 vs S4 comparisons, while only 18 DEGs were identified in both S1 vs S3 and S3
242 vs S4 comparisons (Fig. 1d). To further gain insight into the functions of DEGs, the GO annotation,
243 KEGG pathway and enrichment analyses were performed. GO annotation analysis showed that
244 these DEGs were distributed into 70 functional terms, mainly associated with metabolic and
245 biosynthetic process (Fig. S5 and Table S7). Meanwhile, KEGG pathway and enrichment analyses
246 exhibited that DEGs were obviously involved in starch- and sucrose- related metabolism,
247 phenylpropanoid biosynthesis and circadian rhythm (Table 3; Table S8).

Validation and expression analysis of selected key enzyme genes related to polysaccharide
biosynthesis by qRT-PCR

250 To verify the reliability of the RNA-seq results, we performed qRT-PCR analysis on 16 candidate
251 genes implicated in the polysaccharide synthesis (Fig. 2). Among these, 13 genes showed high
252 expressions with RPKM _ 10 in most stages, other genes displayed low expressions with RPKM
253 _ 10 in all four stages. The genes selected for qRT-PCR compared with the RNA-seq data were
254 shown (Table S1). The results showed that RNA-seq and qRT-PCR expression patterns of 16 genes
255 were consistent, and the positive correlation coefficients (r) were all greater than 0.9 (Fig. 2). For
256 example,   the   r   values   of   Dendrobium_GLEAN_10077593   (*GALE*),
257 Dendrobium_GLEAN_10128035 (*GAE*), and Dendrobium_GLEAN_10026326 (*ASD*) were
258 0.999, 0.999 and 0.977, respectively, indicating that the RNA-Seq data were reliable for the
259 differential gene expression profiles in the *D. catenatum* stems with four developmental stages.

Putative pathway for polysaccharide biosynthesis based on KEGG analysis in *D. catenatum*

Based on KEGG pathways, a total of 29 unigenes encoding 11 key enzymes involved in the polysaccharide metabolism were identified and a detailed metabolic map with the expression patterns of these genes was constructed (Fig. 3a). The detailed information of each enzyme was showed (Table S9). The largest number of unigenes (4 unigenes) were identified as beta-glucosidase (BGLU, EC:3.2.1.21) genes; the second largest number of unigenes (three) were annotated as sucrose synthase (SUS, E2.4.1.13) genes; while the third largest number of unigenes (two) were annotated as trehalose 6-phosphate phosphatase (TPP, EC:3.1.3.12) and UDP-glucuronate 4-epimerase (GAE, EC:5.1.3.6) genes. Furthermore, several unigenes were annotated as encoding UDP glucose 6-dehydrogenase (UGDH, EC:1.1.1.22), maltase-glucoamylase (MGAM, EC:3.2.1.20) and beta-fructofuranosidase (INV, EC:3.2.1.26) genes. In addition, we also showed the expression profiles of 12 genes which had differentially expressed at the four stages (Fig. 3b). Dendrobium_GLEAN_10005633 (UGDH) was highly expressed in S2 and low in S4. However, Dendrobium_GLEAN_10048133 (UGDH) was highly expressed in S3, and low in other stages (Fig. 3b), suggesting that there may be functional divergence among the members of different gene families.

Identification, evolutionary and protein domain analysis of the *SUS* family in *D. catenatum*

The polysaccharide content, transcriptome and metabolic pathway were comprehensively taken into account, we selected the *SUS* for further analysis. To investigate the evolution of these protein architectures across plant species, identification analyses of *SUS* family were conducted in 16 publicly available plant genomes. At last, 105 non-redundant *SUS*s were retrieved in total (Table S10). We found that the *D. catenatum* genome had the largest numbers of *SUS*s, which had subjected to prominent expansion through tandem duplication. Moreover, the *Malus domestica* and tetraploid soybean genomes all contained 12 *SUS*s, whereas a few numbers in other plant genomes were discovered, even those containing only one *SUS* (Table S10). Furthermore, the gene expansions might favor the generation and accumulation of polysaccharides in *D. catenatum*, as was similar with previous study (Yan et al., 2015). During this study, we found that the *SUS*s generally containing above 10 introns occupied the main points (74%), while only one *SUS* harbored a maximum of 23 introns (*CrSUS1*) (Table S10). Afterwards, the Mw and pI of the SUS proteins identified were further determined by using the online version of Compute pI/Mw tool. The Mws ranged from 10.53 (DcSUS8) to 226.02 (CrSUS1) kDa and the pIs varied from 4.72 (DcSUS8) to 9.91 (DcSUS13) (Table S10). It was noteworthy that the pIs of most SUSs (81.9%) were slightly acidic.

Then the ML tree was reconstructed based on the full-length amino-acid sequences of the 105 SUS proteins. Phylogenetic analysis showed that these SUS proteins could be divided into four families (Fig. 4). The SUS I family contained the members from green alga and seed plants, the SUS II and SUS III families gathered in angiosperm and gymnosperm, while the SUS IV family only existed in fern and moss (Fig. 4), indicating that SUS I, SUS II and SUS III families originated in common ancestral genomic contexts before the divergency of the green alga and terrestrial plants and,

299 subsequently, non-seed and seed plant SUSs followed two distinct tracks to evolve.

300 In order to search the evolutionary relationship of the *DcSUS* family, a phylogenetic tree was
301 derived from the alignments of the full-length nucleotide sequences by using the ML method by
302 MEGA6.0 (Fig. 5a). The syntenic analyses showed that there were many duplicate pairs in the
303 *DcSUS* family, including 14 dispersed duplication (DSD) and 2 tandem duplication (TD) ones
304 (Table S11). The evolutionary relationship among the *DcSUS*s was inconsistent with that revealed
305 in their protein phylogenetic tree. In addition, the exon/intron organization analyses of the *DcSUS*s
306 were performed to examine the further genesis. Most *DcSUS*s had no or one intron except the
307 *DcSUS2, DcSUS4, DcSUS5* and *DcSUS11* (Fig. 5b), which varied significantly from the intron
308 numbers of other plant *SUS*s. The results were probably related to the incomplete annotation of
309 the Dendrobium genome sequence. Furthermore, we further investigated the conserved domains
310 of the DcSUSs by using InterProScan database. Generally, SUS proteins harbored a typical sucrose
311 synthase domain (IPR000368) (Fig. 5c), but many SUS proteins also contained other ones, such
312 as the protein kinase-like, Zinc finger and tetratricopeptide-like helical domains (Fig. 5c).

313 **Expression profiles of the *DcSUS*s in different tissues and the stems with four developmental**
314 **stages**

315 Expression profiles could provide clues for their functional divergence among all members of a
316 gene family (Whittle and Krochko, 2009). In the present research, the roots, stems, leaves,
317 pedicels, dorsal sepals, lateral sepals, lips and columns of *D. catenatum* were used to measure the
318 relative expression levels of the 15 *DcSUS*s by qRT-PCR. All 15 *DcSUS*s expressed in the eight
319 tissues. However, most *DcSUS* transcripts showed tissue-specific abundance patterns. For
320 instance, *DcSUS11, DcSUS13* and *DcSUS14* displayed lower expressions in the pedicels, dorsal
321 sepals and lateral sepals, and moderate expressions in other tissues. By contrast, *DcSUS2, DcSUS5*
322 and *DcSUS12* expressed highly in the stems, pedicels, dorsal sepals and lateral sepals, while low
323 expressed in other tissues (Fig. 6a). *DcSUS6, DcSUS10* and *DcSUS15* had low expressions only
324 in the stems, while high expressions in other tissues. *DcSUS3* and *DcSUS4* showed moderate
325 expressions in the pedicels, dorsal sepals and lateral sepals, but high expressions in other tissues.
326 It was interesting that the expression pattern of the *DcSUS8* was just opposite. In addition, the
327 *DcSUS1* and *DcSUS7* displayed a relatively lower expressions in all tissues, while the *DcSUS9*
328 displayed relatively high expressions in all tissues (Fig. 6a). Therefore, the *DcSUS*s might exhibit
329 function divergences in the corresponding tissues.

330 To survey the development-dependent expression difference of the *DcSUS*s, we also investigated
331 their expression patterns in the stems with four developmental stages (Table S12). Their expression
332 profiles could be mainly divided into two cases. One had low expressions in four developmental
333 stages including the *DcSUS1, DcSUS5, DcSUS6, DcSUS10* and *DcSUS15*. Another displayed
334 relative high expression including *DcSUS2, DcSUS3, DcSUS7, DcSUS8, DcSUS9, DcSUS12,*
335 *DcSUS13* and *DcSUS14* (Fig. 6b). Moreover, the *DcSUS11* showed high expression only in S1
336 stage, and low expressions in other three stages. While the *DcSUS4* had a moderate expression
337 only in S1 stage, and high expressions in other stages (Fig. 6b). Thus, the *DcSUS*s might hold

338 distinct functions in different development stages of the stems.

Discussion

Potential candidate genes involved in polysaccharide synthesis

The dominating bioactive components of *D. catenatum* were soluble polysaccharides (Ng et al., 2012), which were synthesized from monosaccharides, such as glucose, mannose, galactose, rhamnose, arabinose, xylose, and so on (Zha et al., 2007, Fan et al., 2009). It was indicated in the current research that, the longer the plants grow, the higher the content of soluble polysaccharides (Table 1). In addition, the transcriptomes of the stems with the four developmental stages revealed 1762 DEGs (Fig. 1). Moreover, further analysis of KEGG pathway displayed that many genes encoding the key enzymes involved in starch and sucrose metabolism were identified. The results showed that beta-fructofuranosidase (INV, EC:3.2.1.26) expressed differentially in the developmental process of *D. catenatum* stems (Fig. 3a). A previous study has reported that the elongation of *A. thaliana* roots was regulated by vacuolar invertase (INV) that can degrade sucrose to produce glucose and fructose (Sergeeva et al., 2006). Moreover, the SUS (EC 2.4.1.13) gene associated with sucrose metabolism had significant differentially expressed among developmental stages (Fig. 3a). Moreover, Os4bglu12 β-glucosidase (EC:3.2.1.21) had high exoglucanase activity and consistent with a role in cell wall metabolism (Opassiri et al., 2006), which similarly had distinct expression profiles in our study(Fig. 3a). Likewise, a previous research reported that UDP-arabinose 4-epimerase (UXE, EC:5.1.3.5) with distinct expression patterns (Fig. 3) played key roles in the synthesis of arabinosylated cell wall components (Burget et al., 2003). Furthermore, UDP-glucose dehydrogenase (UGDH, EC 1.1.1.22) catalyzes the formation of UDP-GlcA from UDPGlc, and then synthesizes UDP-Xyl (Oka and Jigami, 2006) in which d-xylose is mainly present in the form of cell wall polysaccharides and N-glycan (Mercx et al., 2017). However, there were also some genes that had no obvious differentially expressions in the present study, but they had been previously proved to be involved in polysaccharide synthesis. For example, mannose-1-phosphate_guanylyltransferase (EC:2.7.7.13) was a key enzyme associated with the cellulose biosynthesis (Kim et al., 2014), which had been identified to be involved in the synthesis of mannan polysaccharides in *D. officinale* (*D. catenatum*) (He et al., 2015). In addition, mannose-6-phosphate isomerase (MPI, EC:5.3.1.8) catalyzed the reversible isomerization between D-fructose 6-phosphate and D-mannose 6-phosphate and participated in hexose metabolic process (Wang et al., 2014). Interestingly, MPI encoding genes had highly expression levels in the *D. catenatum* stems with four developmental stages (Fig. 3b). In a word, these enzymes might contribute to the growth and development of *D. catenatum* and their genes could be used as the candidate ones involved in the synthesis of the polysaccharides.

Tandem duplication contributed to *DcSUS* expansion

Tandem duplications generating the duplicates that are closely adjacent to each other (generally separated by 10 or fewer genes) has facilitated greatly to the expansion of plant gene families (Rizzon et al., 2006, Freeling, 2009). Our analysis indicated that there were obvious gene expansions in the *DcSUS*s holding the most gene family members (Table S11). Further syntenic

377 analyses displayed that the multiple copies of the *DcSUS* resulted from tandem duplications or
378 segmental duplications (Fig. 5a; Table S11), which was similar to those of *MKK* family (Jiang and
379 Chu, 2018, Jiang et al., 2021). Moreover, previous studies also reported that the *DcSUS*s maybe
380 underwent expansion through tandem duplication (Yan et al., 2015, Qiao et al., 2019). What,s
381 more, these tandem duplicates generally played a vital role in the plant adaptation to respond
382 rapidly changing environments (Hanada et al., 2008). Similarly, the expansions of *SPS* and *SUS*
383 (*SuSy*) had been believed to contribute to the polysaccharide richness in *D. officinale* (i.e., *D.*
384 *catenatum*) (Yan et al., 2015). Meanwhile, just as the synthesis of the polysaccharide in
385 *Arabidopsis thaliana* was beneficial to the drought-resisting of *A. thaliana* (Balsamo et al., 2015),
386 the polysaccharide synthesis and the expressions of the related genes including the *DcSUS* in *D.*
387 *catenatum* stems contributed to the adaptability of cultivated *D. catenatum* to various stresses, e.g.,
388 drought, salt and osmotic ones (Huang et al., 2021). As a result, since the genes giving plants the
389 ability to respond rapidly changing environments have been verified to hold expansion tendency
390 including tandem duplication (Hanada et al., 2008), the richness of polysaccharides in *D.*
391 *catenatum* might be related to the expansion of these genes including the *DcSUS*, which was in
392 accordance with that reported in previous studies (Yan et al., 2015).

393 **Sequence evolution, domain organization and expression patterns of *DcSUS***

394 Compared with other species, there were more large *SUS* numbers in *Glycine max*, *M. domestica*
395 and *D. catenatum* genome (Table S10). Generally, the increase in genes is associated with
396 polyploidy or ancient polyploidization events (Jiao et al., 2011). For example, the 12 *SUS*s existed
397 in recently duplicated tetraploid *G. max* genomes which had occurred two rounds of whole genome
398 duplications (WGDs). In addition, except the sucrose synthase domain, other domain organizations
399 were detected, e.g., the protein kinase-like and tetratricopeptide-like helical domains (Fig. 5c). The
400 glycosyltransferase (GT) gene family affect many aspects of plant growth and development. For
401 instance, the loss of GT genes affect the production of recombinant proteins beta-1, 2-xylose and
402 core alpha-1,3-fucose in *Nicotiana benthamiana* (Jansing et al., 2019), and the UDP-
403 glycosyltransferase gene regulates the ginsenoside synthesis in *Panax ginseng* and *Panax*
404 *quinquefolius* (Lu et al., 2017). Moreover, *Magnaporthe oryzae* Chitinase MoChia1 interacts with
405 OsTPR that is a tetratricopeptide repeat protein to counteract the function of this fungal chitinase
406 and regain immunity (Yang et al., 2019). These results indicated that the *SUS*s might also have
407 divergent functions due to the domain organization change.

408 In a number of cases, the changes in tissue-specific expression of genes may also lead to the
409 changes in the functions of the gene paralogs (Zhang and Ma, 2012). *VvSUS3* was the most highly
410 expressed gene in the berries, which closely affected the sugar content in the berries and the
411 changes in SUS activity (Ren et al., 2020). The carrot *DcSus* were higher expressed in the leaf
412 blades than those in the roots and petioles, which showed strong negative correlation both with the
413 sucrose and soluble sugar contents (Liu et al., 2018). In addition, most *MdSUS*s displayed
414 decreased expressions during fruit development, whereas the expression profiles of *MdSUS2.1* and
415 *MdSUS1.4* were opposite, which indicated that *MdSUS*s might play distinct functions in the sugar
416 utilization and sink-source sugar cycle in apple (Tong et al., 2018). Therefore, we investigated the

417 expression profiles of *DcSUS*s in different *D. catenatum* tissues. Most *DcSUS*s displayed tissue-
418 specific abundance patterns, especially in the stems (Fig. 6a), implying that, in sucrose
419 metabolism, the *DcSUS*s might play distinct roles different from those in other plants. For example,
420 *DcSUS7* showed higher expression in stems than those in other tissues (Fig. 6a), indicating that
421 the gene was most likely related to the synthesis of the sucrose in *D. catenatum* stems.
422 Furthermore, *DcSUS10* had higher expressions in S3 and S4 stages than that in S1 (Fig. 6b), which
423 was similar to the polysaccharide and mannose contents in the stems with the specific stages (Table
424 1), suggesting that the sucrose synthesis in developing stems might be closely associated with the
425 *DcSUS*s.

426 **Conclusion**
427 In this study, we detected and compared the polysaccharide and mannose contents of the *D.*
428 *catenatum* stems with four developmental stages, and conducted RNA-seq analysis of the
429 corresponding stems. A total of 16,384 genes were detected. Several DEGs correlated with the
430 metabolic and biosynthetic process were identified. Further analysis showed that the DEGs were
431 mainly enriched in starch and sucrose metabolism pathway. More importantly, we observed the
432 *SUS* encoding polysaccharide biosynthase in 16 representative plants, and studied its genomic
433 characteristics and evolutionary relationships. The results suggested that the expansions in *DcSUS*s
434 were caused by tandem duplications. Moreover, the 15 *SUS*s showed two different expression
435 patterns at the four developmental stages and were significantly regulated in different *D.*
436 *catenatum* tissues. In general, these results not only provided gene resources for the genetic
437 improvement of *D. catenatum*, but also laid a foundation for further understanding of the molecular
438 mechanism of polysaccharide biosynthesis.

439 **Data Availability Statement (DAS)**
440 Raw data were deposited in SRA (https://www.ncbi.nlm.nih.gov/sra/) under the accession
441 SRR12805741- SRR12805748.

442 **References**
443 Abdullah M, Cao YP, Cheng X, Meng DD, Chen Y, Shakoor A, Gao JS, and Cai YP. 2018. The sucrose synthase
444   gene family in Chinese pear (*Pyrus bretschneideri* rehd.): structure, expression, and evolution. Molecules
445   **23**:1144.
446 Anders S, and Huber W. 2010. Differential expression analysis for sequence count data. Genome Biology **11**:R106.
447 Angeles-Nunez JG, and Tiessen A. 2010. Arabidopsis sucrose synthase 2 and 3 modulate metabolic homeostasis and
448   direct carbon towards starch synthesis in developing seeds. Planta **232**:701-718.
449 Balsamo R, Boak M, Nagle K, Peethambaran B, and Layton B. 2015. Leaf biomechanical properties in *Arabidopsis*
450   *thaliana* polysaccharide mutants affect drought survival. Journal of Biomechanics **48**:4124-4129.
451 Bar-Peled M, and O'Neill MA. 2011. Plant nucleotide sugar formation, interconversion, and salvage by sugar
452   recycling. Annual Review of Plant Biology **62**:127-155.
453 Barratt DHP, Derbyshire P, Findlay K, Pike M, Wellner N, Lunn J, Feil R, Simpson C, Maule AJ, and Smith AM.
454   2009. Normal growth of Arabidopsis requires cytosolic invertase but not sucrose synthase. Proceedings of
455   the National Academy of Sciences of the United States of America **106**:13124-13129.

456   Baud S, Vaultier MN, and Rochat C. 2004. Structure and expression profile of the sucrose synthase multigene
457   family in Arabidopsis. Journal of Experimental Botany **55**:397-409.
458   Bieniawska Z, Barratt DHP, Garlick AP, Thole V, Kruger NJ, Martin C, Zrenner R, and Smith AM. 2007. Analysis
459   of the sucrose synthase gene family in Arabidopsis. Plant Journal **49**:810-828.
460   Burget EG, Verma R, Molhoj M, and Reiter WD. 2003. The biosynthesis of L-arabinose in plants: Molecular
461   cloning and characterization of a Golgi-localized UDP-D-xylose 4-epimerase encoded by the *MUR4* gene
462   of Arabidopsis. Plant Cell **15**:523-531.
463   Ching T, Huang S, and Garmire LX. 2014. Power analysis and sample size estimation for RNA-Seq differential
464   expression. RNA **20**:1684-1696.
465   de Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, and Hulo N.
466   2006. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and
467   structural residues in proteins. Nucleic Acids Research **34**:W362-W365.
468   Dejardin A, Sokolov LN, and Kleczkowski LA. 1999. Sugar/osmoticum levels modulate differential abscisic acid-
469   independent expression of two stress-responsive sucrose synthase genes in Arabidopsis. Biochemical
470   Journal **344**:503-509.
471   Fan YJ, He XJ, Zhou SD, Luo AX, He T, and Chun Z. 2009. Composition analysis and antioxidant activity of
472   polysaccharide from *Dendrobium denneanum*. International Journal of Biological Macromolecules **45**:169-
473   173.
474   Freeling M. 2009. Bias in plant gene content following different sorts of duplication: tandem, whole-genome,
475   segmental, or by transposition. Annual Review of Plant Biology **60**:433-453.
476   Haigler CH, Ivanova-Datcheva M, Hogan PS, Salnikov VV, Hwang S, Martin K, and Delmer DP. 2001. Carbon
477   partitioning to cellulose synthesis. Plant Molecular Biology **47**:29-51.
478   Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, and Shiu SH. 2008. Importance of lineage-specific expansion of
479   plant tandem duplicates in the adaptive response to environmental stimuli. Plant Physiology **148**:993-1003.
480   He CM, Zhang JX, Liu XC, Zeng SJ, Wu KL, Yu ZM, Wang XJ, da Silva JAT, Lin ZJ, and Duan J. 2015.
481   Identification of genes involved in biosynthesis of mannan polysaccharides in *Dendrobium officinale* by
482   RNA-seq analysis. Plant Molecular Biology **88**:219-231.
483   Hirose T, Scofield GN, and Terao T. 2008. An expression analysis profile for the entire sucrose synthase gene
484   family in rice. Plant Science **174**:534-543.
485   Hu B, Jin JP, Guo AY, Zhang H, Luo JC, and Gao G. 2015. GSDS 2.0: an upgraded gene feature visualization
486   server. Bioinformatics **31**:1296-1297.
487   Hua YF, Zhang M, Fu CX, Chen ZH, and Chan GYS. 2004. Structural characterization of a 2-O-acetylglucomannan
488   from *Dendrobium officinale* stem. Carbohydrate Research **339**:2219-2224.
489   Huang H, Wang H, Tong Y, and Wang YH. 2021. Identification and characterization of HD-Zip genes reveals their
490   roles in stresses responses and facultative crassulacean acid metabolism in *Dendrobium catenatum*. Scientia
491   Horticulturae **285:** 110058.
492   Huber SC, and Huber JL. 1996. Role and regulation of sucrose-phosphate synthase in higher plants. Annual Review
493   of Plant Physiology and Plant Molecular Biology **47**:431-444.
494   Jansing J, Sack M, Augustine SM, Fischer R, and Bortesi L. 2019. CRISPR/Cas9-mediated knockout of six
495   glycosyltransferase genes in *Nicotiana benthamiana* for the production of recombinant proteins lacking
496   beta-1,2-xylose and core alpha-1,3-fucose. Plant Biotechnology Journal **17**:350-361.

497 Jiang M, and Chu ZQ. 2018. Comparative analysis of plant *MKK* gene family reveals novel expansion mechanism
498   of the members and sheds new light on functional conservation. Bmc Genomics **19**:407.
499 Jiang M, Li P, and Wang W. 2021. Comparative analysis of *MAPK* and *MKK* gene families reveals differential
500   evolutionary patterns in *Brachypodium distachyon* inbred lines. Peerj **9**:e11238.
501 Jiang M, Xu SZ, Wen GS, and Zhao CL. 2017. Validation of seven housekeeping genes as reference ones for qRT-
502   PCR normalization in *Dendrobium catenatum*. Russian Journal of Plant Physiology **64**:497-508.
503 Jiao YN, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang HY,
504   Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, and dePamphilis
505   CW. 2011. Ancestral polyploidy in seed plants and angiosperms. Nature **473**:97-100.
506 Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T,
507   and Yamanishi Y. 2008. KEGG for linking genomes to life and the environment. Nucleic Acids Research
508   **36**:D480-D484.
509 Kim WC, Reca IB, Kim Y, Park S, Thomashow M, Keegstra K, and Han KH. 2014. Transcription factors that
510   directly regulate the expression of CSLA9 encoding mannan synthase in *Arabidopsis thaliana*. Plant
511   Molecular Biology **84**:577-587.
512 Koch KE, Wu Y, and Xu J. 1996. Sugar and metabolic regulation of genes for sucrose metabolism: Potential
513   influence of maize sucrose synthase and soluble invertase responses on carbon partitioning and sugar
514   sensing. Journal of Experimental Botany **47**:1179-1185.
515 Langmead B, and Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods **9**:357-359.
516 Li O, Xiao R, Sun LH, Guan CL, Kong DD, and Hu XF. 2017. Bacterial and diazotrophic diversities of endophytes
517   in *Dendrobium catenatum* determined through barcoded pyrosequencing. Plos One **12**:e0184717.
518 Liu YJ, Wang GL, Ma J, Xu ZS, Wang F, and Xiong AS. 2018. Transcript profiling of sucrose synthase genes
519   involved in sucrose metabolism among four carrot (*Daucus carota* L.) cultivars reveals distinct patterns.
520   Bmc Plant Biology **18**.
521 Lu C, Zhao SJ, Wei GN, Zhao HJ, and Qu QL. 2017. Functional regulation of ginsenoside biosynthesis by RNA
522   interferences of a UDP-glycosyltransferase gene in *Panax ginseng* and *Panax quinquefolius*. Plant
523   Physiology and Biochemistry **111**:67-76.
524 Ma YN, Hou ZN, Liang ZS, and Liu L. 2018. Quality evaluation of the rare medicinal plant *Dendrobium officinale*
525   based on volatile constituents, methanol extracts and polysaccharides. Current Pharmaceutical Analysis
526   **14**:121-132.
527 Meng LZ, Lv GP, Hu DJ, Cheong KL, Xie J, Zhao J, and Li SP. 2013. Effects of polysaccharides from different
528   species of Dendrobium (Shihu) on macrophage function. Molecules **18**:5779-5791.
529 Meng YJ, Yu DL, Xue J, Lu JJ, Feng SG, Shen CJ, and Wang HZ. 2016. A transcriptome-wide, organ-specific
530   regulatory map of *Dendrobium officinale*, an important traditional Chinese orchid herb. Scientific Reports
531   **6**:18864.
532 Mercx S, Smargiasso N, Chaumont F, De Pauw E, Boutry M, and Navarre C. 2017. Inactivation of the beta(1,2)-
533   xylosyltransferase and the alpha(1,3)-fucosyltransferase genes in *Nicotiana tabacum* BY-2 cells by a
534   multiplex CRISPR/Cas9 strategy results in glycoproteins without plant-specific glycans. Frontiers in Plant
535   Science **8**:403.

536 N'tchobo H, Dali N, Nguyen-Quoc B, Foyer CH, and Yelle S. 1999. Starch synthesis in tomato remains constant
537   throughout fruit development and is dependent on sucrose supply and sucrose synthase activity. Journal of
538   Experimental Botany **50**:1457-1463.
539 Ng TB, Liu JY, Wong JH, Ye XJ, Sze SCW, Tong Y, and Zhang KY. 2012. Review of research on Dendrobium, a
540   prized folk medicine. Applied Microbiology and Biotechnology **93**:1795-1803.
541 Nunez JGA, Kronenberger J, Wuilleme S, Lepiniec L, and Rochat C. 2008. Study of AtSUS2 localization in seeds
542   reveals a strong association with plastids. Plant and Cell Physiology **49**:1621-1626.
543 Oka T, and Jigami Y. 2006. Reconstruction of de novo pathway for synthesis of UDP-glucuronic acid and UDP-
544   xylose from intrinsic UDP-glucose in *Saccharomyces cerevisiae*. Febs Journal **273**:2645-2657.
545 Opassiri R, Pomthong B, Onkoksoong T, Akiyama T, Esen A, and Cairns JRK. 2006. Analysis of rice glycosyl
546   hydrolase family I and expression of Os4bglu12 beta-glucosidase. Bmc Plant Biology **6**:33.
547 Qiao X, Li QH, Yin H, Qi KJ, Li LT, Wang RZ, Zhang SL, and Paterson AH. 2019. Gene duplication and evolution
548   in recurring polyploidization-diploidization cycles in plants. Genome Biology **20**:38.
549 Ren R, Yue X, Li J, Xie S, Guo S, and Zhang Z. 2020. Coexpression of sucrose synthase and the SWEET
550   transporter, which are associated with sugar hydrolysis and transport, respectively, increases the hexose
551   content in *Vitis vinifera* L. grape berries. Frontiers in Plant Science **11**:321.
552 Rizzon C, Ponger L, and Gaut BS. 2006. Striking similarities in the genomic distribution of tandemly arrayed genes
553   in Arabidopsis and rice. Plos Computational Biology **2**:989-1000.
554 Robinson MD, McCarthy DJ, and Smyth GK. 2010. edgeR: a Bioconductor package for differential expression
555   analysis of digital gene expression data. Bioinformatics **26**:139-140.
556 Ruan YL, Llewellyn DJ, and Furbank RT. 2003. Suppression of sucrose synthase gene expression represses cotton
557   fiber cell initiation, elongation, and seed development. Plant Cell **15**:952-964.
558 Sergeeva LI, Keurentjes JJB, Bentsink L, Vonk J, van der Plas LHW, Koornneef M, and Vreugdenhil D. 2006.
559   Vacuolar invertase regulates elongation of *Arabidopsis thaliana* roots as revealed by QTL and mutant
560   analysis. Proceedings of the National Academy of Sciences of the United States of America **103**:2994-
561   2999.
562 Shen CJ, Guo H, Chen HL, Shi YJ, Meng YJ, Lu JJ, Feng SG, and Wang HZ. 2017. Identification and analysis of
563   genes associated with the synthesis of bioactive constituents in *Dendrobium officinale* using RNA-Seq.
564   Scientific Reports **7**:187.
565 Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S. 2013. MEGA6: Molecular evolutionary genetics
566   analysis version 6.0. Molecular Biology and Evolution **30**:2725-2729.
567 Tong XL, Wang ZY, Ma BQ, Zhang CX, Zhu LC, Ma FW, and Li MJ. 2018. Structure and expression analysis of
568   the sucrose synthase gene family in apple. Journal of Integrative Agriculture **17**:847-856.
569 Van Bel M, Diels T, Vancaester E, Kreft L, Botzki A, Van de Peer Y, Coppens F, and Vandepoele K. 2018. PLAZA
570   4.0: an integrative resource for functional, evolutionary and comparative plant genomics. Nucleic Acids
571   Research **46**:D1190-D1196.
572 Wagner GP, Kin K, and Lynch VJ. 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure
573   is inconsistent among samples. Theory in Biosciences **131**:281-285.
574 Wang QX, Yan N, Ji DG, Li SY, and Hu H. 2013. *In vitro* growth and carbon utilization of the green-leaved orchid
575   *Dendrobium officinale* are promoted by mycorrhizal associations. Botanical Studies **54**:23.

576 Wang XH, Zhang S, Hu D, Zhao XJ, Li Y, Liu TK, Wang JJ, Hou XL, and Li Y. 2014. BcPMI2, isolated from non-
577   heading Chinese cabbage encoding phosphomannose isomerase, improves stress tolerance in transgenic
578   tobacco. Molecular Biology Reports **41**:2207-2216.
579 Whittle CA, and Krochko JE. 2009. Transcript profiling provides evidence of functional divergence and expression
580   networks among ribosomal protein gene paralogs in *Brassica napus*. The Plant Cell **21**:2203-2219.
581 Xiang XG, Schuiteman A, Li DZ, Huang WC, Chung SW, Li JW, Zhou HL, Jin WT, Lai YJ, Li ZY, and Jin XH.
582   2013. Molecular systematics of Dendrobium (Orchidaceae, Dendrobieae) from mainland Asia based on
583   plastid and nuclear sequences. Molecular Phylogenetics and Evolution **69**:950-960.
584 Yan L, Wang X, Liu H, Tian Y, Lian JM, Yang RJ, Hao SM, Wang XJ, Yang SC, Li QY, Qi S, Kui L, Okpekum M,
585   Ma X, Zhang JJ, Ding ZL, Zhang GJ, Wang W, Dong Y, and Sheng J. 2015. The genome of *Dendrobium*
586   *officinale* illuminates the biology of the important traditional Chinese orchid herb. Molecular Plant **8**:922-
587   934.
588 Yang C, Yu YQ, Huang JK, Meng FW, Pang JH, Zhao QQ, Islam MA, Xu N, Tian Y, and Liu J. 2019. Binding of
589   the *Magnaporthe oryzae* chitinase MoChia1 by a rice tetratricopeptide repeat protein allows free chitin to
590   trigger immune responses. Plant Cell **31**:172-188.
591 Young MD, Wakefield MJ, Smyth GK, and Oshlack A. 2010. Gene ontology analysis for RNA-seq: accounting for
592   selection bias. Genome Biology **11**:R14.
593 Zha XQ, Luo JP, Luo SZ, and Jiang ST. 2007. Structure identification of a new immunostimulating polysaccharide
594   from the stems of *Dendrobium huoshanense*. Carbohydrate Polymers **69**:86-93.
595 Zhang DQ, Xu BH, Yang XH, Zhang ZY, and Li BL. 2011. The sucrose synthase gene family in Populus: structure,
596   expression, and evolution. Tree Genetics & Genomes **7**:443-456.
597 Zhang GQ, Xu Q, Bian C, Tsai WC, Yeh CM, Liu KW, Yoshida K, Zhang LS, Chang SB, Chen F, Shi Y, Su YY,
598   Zhang YQ, Chen LJ, Yin YY, Lin M, Huang HX, Deng H, Wang ZW, Zhu SL, Zhao X, Deng C, Niu SC,
599   Huang J, Wang MN, Liu GH, Yang HJ, Xiao XJ, Hsiao YY, Wu WL, Chen YY, Mitsuda N, Ohme-Takagi
600   M, Luo YB, Van de Peer Y, and Liu ZJ. 2016a. The *Dendrobium catenatum* Lindl. genome sequence
601   provides insights into polysaccharide synthase, floral development and adaptive evolution. Scientific
602   Reports **6**:19029.
603 Zhang JX, He CM, Wu KL, da Silva JAT, Zeng SJ, Zhang XH, Yu ZM, Xia HQ, and Duan J. 2016b. Transcriptome
604   analysis of *Dendrobium officinale* and its application to the identification of genes associated with
605   polysaccharide synthesis. Frontiers in Plant Science **7**:5.
606 Zhang LS, and Ma H. 2012. Complex evolutionary history and diverse domain organization of SET proteins suggest
607   divergent regulatory interactions. New Phytologist **195**:248-263.
608

Polysaccharide and mannose content in four development stages of *D. catenatum.*

1 **Table 1 Polysaccharide and mannose content in four development stages of *Dendrobium***
2 ***catenatum.***
3

Sample name	Tissue description	Polysaccharide content(mg/g)	Mannose content(%w/w)
S1	2-3 months	146	27.0
S2	5-6 months	213	24.3
S3	8-9 months	347	49.6
S4	11-12 months	431	36.1

Sequencing and assembly statistics for the 8 transcriptome data of four developmental stages in *D. catenatum*.

1 **Table 2 Sequencing and assembly statistics for the 8 transcriptome data of four**
2 **developmental stages in *D.catenatum*.**

Sample ID	Raw reads (M)	Clean reads (M)	Q30 (%)	GC content (%)	No.of mapped reads(M)	Uniquely mapped reads (M)
S1-1	63.47	62.33	95.53	46.49	44.24(70.98%)	39.03 (62.61%)
S1-2	65.11	64.03	95.81	46.12	46.51(72.65%)	41.17 (64.3%)
S2-1	76.23	74.71	95.08	46.05	52.82(70.7%)	46.69(62.5%)
S2-2	61.53	60.48	95.72	46.07	43.79(72.39%)	38.64(63.88%)
S3-1	63.69	62.79	95.85	45.87	45.64(72.68%)	40.06(63.79%)
S3-2	72.19	71.13	95.96	45.68	50.78(71.39%)	44.87 (63.08%)
S4-1	64.66	63.62	96.11	45.76	43.53(68.43%)	38.41(60.37%)
S4-2	75.84	74,.59	96.17	45.76	53.62(71.88%)	47.09(63.13%)

Top 20 enriched KEGG pathways with the highest representation of DEGs among four development stages.

1 **Table 3 Top 20 enriched KEGG pathways with the highest representation of DEGs among**
2 **four development stages.**

Pathway	Pathway ID	DEGs genes	Background number
Starch and sucrose metabolism	ko00500	29	353
Phenylpropanoid biosynthesis	ko00940	24	226
Protein processing in endoplasmic reticulum	ko04141	24	355
Circadian rhythm - plant	ko04712	24	95
Plant hormone signal transduction	ko04075	22	327
Plant-pathogen interaction	ko04626	19	294
Glycine, serine and threonine metabolism	ko00260	16	126
Phenylalanine metabolism	ko00360	16	156
Pentose and glucuronate interconversions	ko00040	14	125
Tyrosine metabolism	ko00350	14	69
Glutathione metabolism	ko00480	14	153
Isoquinoline alkaloid biosynthesis	ko00950	13	42
Toxoplasmosis	ko05145	12	159
Cyanoamino acid metabolism	ko00460	11	77
Cutin, suberine and wax biosynthesis	ko00073	10	46
MicroRNAs in cancer	ko05206	10	131
Metabolism of xenobiotics by cytochrome P450	ko00980	9	78
Drug metabolism-cytochrome P450	ko00982	9	100
Ribosome	ko03010	9	444
Phagosome	ko04145	9	139

Figure 1

Analysis of changes in gene expression among four developmental stages in *D. catenatum*.

(a) The number of differentially expressed genes (DEGs) was obtained from comparisons of S1 versus S2, S1 versus S3, S1 versus S4, S2 versus S3, S2 versus S4 and S3 versus S4. (b) All DEGs were classified into six clusters by short time-series expression miner (STEM, P value < 0.05). (c) Heat map illustrating the expression profiles of the developmental-differentially expressed genes. (d) VennDiagrams of the DEGs in different comparisons.
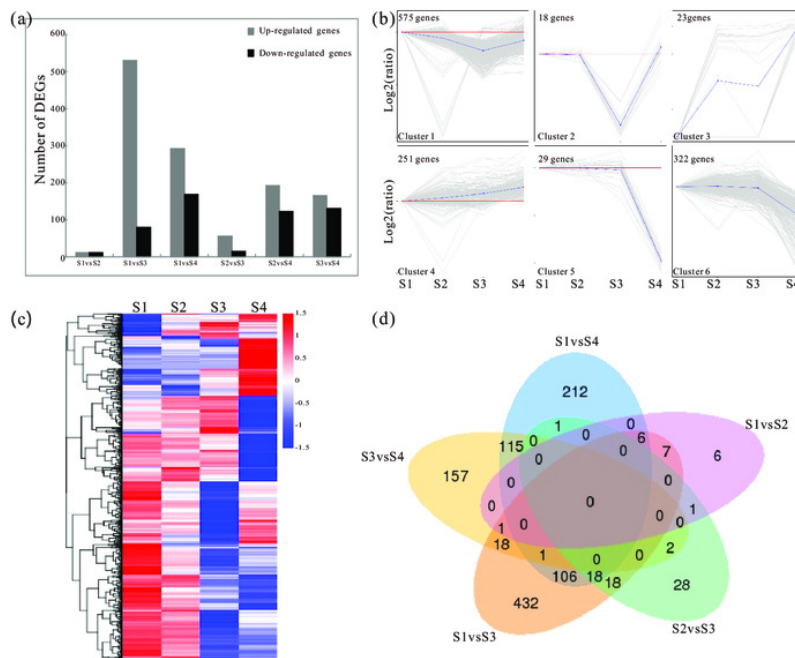
Figure 2

Result of qRT-PCR analysis.

The left Y axis represents RPKM value of each gene using RNA-Seq analysis. The right Y axis represents log2 transformed relative transcript amount obtained by qRT-PCR. The correlation co-efficient(r) between the two expression profiles is also showed.
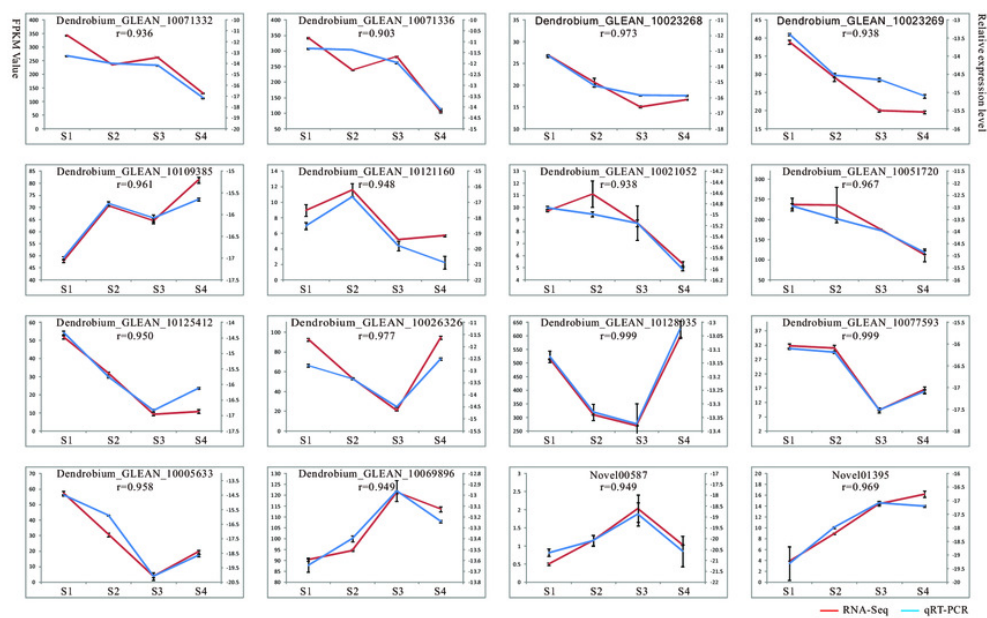
Figure 3

Biosynthesis pathways of polysaccharides and related gene expression status with four developmental stages in *D. catenatum*.

(a) Schematic representation of the polysaccharide biosynthetic pathway based on KEGG enrichment analysis. Enzymes were highlighted and marked in orange and blue according to up-regulated or down-regulated expression in different development stages, respectively. (b) Heat map of polysaccharide biosynthetic pathway related genes in four developmental stages of *D. catenatum*.
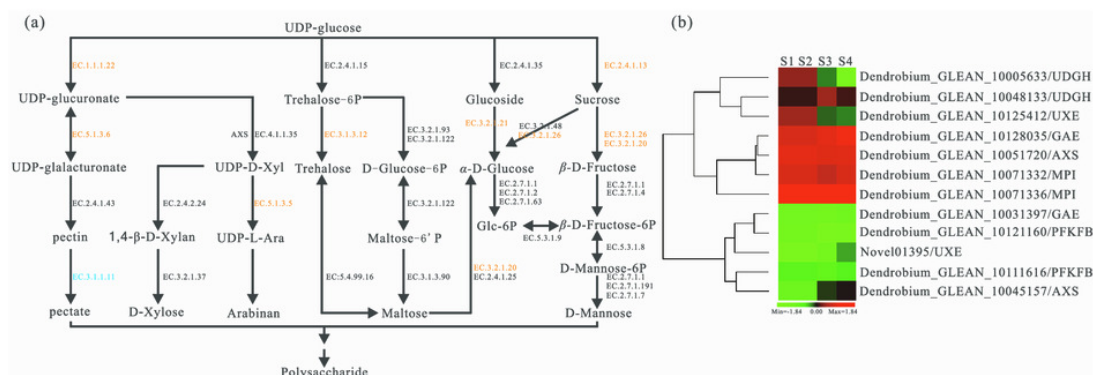
Figure 4

Maximum Likelihood phylogenetic tree of sucrose synthase gene family from 16 plant species.

Phylogenetic analysis was carried with protein sequences for 105 SUS proteins from 16 plant species identified in this study.
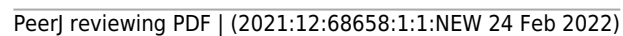
Figure 5

Phylogenetic relationships, gene and protein structure analyses in SUS proteins from *D. catenatum*.

(a)The phylogenetic tree was constructed from the amino acid sequences using the ML program from MEGA 6, representing relationships among 15 SUS proteins from *D. catenatum*. Two tandem duplicated genes were marked by blue and green, respectively. (b) The exon/intron structure of each *DcSUS* gene was proportionally showed based on the scale at the bottom. (c) Structure of SUS proteins in *D. catenatum*.
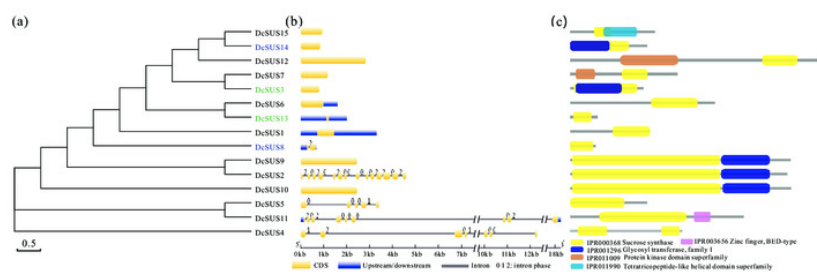
Figure 6

Expression profiles of *D. catenatum DcSUS* genes in different tissues and developmental stages.

(a) Expression patterns of *DcSUS* genes in eight tissues including root, stem, leaf, pedicel, dorsal sepal, lateral sepal, petal, lip and column. (b) Expression patterns of *DcSUS* genes in four developmental stages of *D. catenatum*.