Species-Specific Gene Expansion of the Cellulose synthase Gene Superfamily in the Orchidaceae Family and Functional Divergence of Mannan Synthesis-Related Genes in Dendrobium officinale

Yunzhu Wang¹, Kunkun Zhao¹, Yue Chen¹, Qingzhen Wei², Xiaoyang Chen³, Hongjian Wan² and Chongbo Sun¹*

¹ Institute of Horticulture Research, Zhejiang Academy of Agricultural Sciences, Hangzhou, China, ² Institute of Vegetable Research, Zhejiang Academy of Agricultural Sciences, Hangzhou, China, ³ Seed Management Terminal of Zhejiang, Hangzhou, China

Citation: Wang Y, Zhao K, Chen Y, Wei Q, Chen X, Wan H and Sun C (2022) Species-Specific Gene Expansion of the Cellulose synthase Gene Superfamily in the Orchidaceae Family and Functional Divergence of Mannan Synthesis-Related Genes in Dendrobium officinale. Front. Plant Sci. 13:777332. doi: 10.3389/fpls.2022.777332

Plant Cellulose synthase genes constitute a supergene family that includes the Cellulose synthase (CesA) family and nine Cellulose synthase-like (Csl) families, the members of which are widely involved in the biosynthesis of cellulose and hemicellulose. However, little is known about the Cellulose synthase superfamily in the family Orchidaceae, one of the largest families of angiosperms. In the present study, we identified and systematically analyzed the CesA/Csl family members in three fully sequenced Orchidaceae species, i.e., Dendrobium officinale, Phalaenopsis equestris, and Apostasia shenzhenica. A total of 125 Cellulose synthase superfamily genes were identified in the three orchid species and classified into one CesA family and six Csl families: CslA, CslC, CslD, CslE, CslG, and CslH according to phylogenetic analysis involving nine representative plant species. We found species-specific expansion of certain gene families, such as the CslAs in D. officinale (19 members). The CesA/Csl families exhibited sequence divergence and conservation in terms of gene structure, phylogeny, and deduced protein sequence, indicating multiple origins via different evolutionary processes. The distribution of the DofCesA/DofCsl genes was investigated, and 14 tandemly duplicated genes were detected, implying that the expansion of DofCesA/DofCsl genes may have originated via gene duplication. Furthermore, the expression profiles of the DofCesA/DofCsl genes were investigated using transcriptome sequencing and quantitative Real-time PCR (qRT-PCR) analysis, which revealed functional divergence in different tissues and during different developmental stages of D. officinale. Three DofCesAs were highly expressed in the flower, whereas DofCslD and DofCslC family genes exhibited low expression levels in all tissues and at all developmental stages. The 19 DofCslAs were differentially expressed in the D. officinale stems at different developmental stages, among which six DofCslAs were expressed at low levels or not at all. Notably, two DofCslAs (DofCslA14 and DofCslA15) showed significantly high expression in the stems.
INTRODUCTION

cellulose, hemicellulose, and pectin are three major types of polysaccharides in plant cell walls and play vital roles in controlling cell shape, expansion, overall development, and interactions with the environment (McFarlane et al., 2014). Cellulose is a paracrystalline polymer of β-1,4 glucan chains and is the major determinant of the load-bearing capacity of cell walls constituting approximately one-third of the total plant mass (Nishiyama, 2009; McFarlane et al., 2014). Hemicelluloses are polysaccharides that have equatorial β-(1→4)-linked backbones, including xyloglucans, xylans, mannans, glucomannans, and β-(1→3,1→4)-glucans (Scheller and Ulvskov, 2010). Hemicellulose and pectin polysaccharide matrix fill in the gaps between cellulose microfibrils.

The Cellulose synthase (CesA) superfamily comprises the CesA family and nine CesA-like (Csl) families, which all belong to the glycosyltransferase-2 (GT2) superfamily, typically with a catalytic domain containing a DDDQXXRW motif (Pear et al., 1996). CesA family members include xylanases, mannanases, and glucomannanases, and β-(1→3,1→4)-glucanases (CesA genes). CesA/Csl genes in green algae to up to 50 genes in terrestrial angiosperms have been identified in many plant species, including Arabidopsis (Persson et al., 2007), tomato (Song et al., 2018), rice (Shu et al., 2007), and barley (Burton and Taniguchi, 2015), Dendrobium, and P. bretschneideri. CesA/C-like genes in green algae, including multicellularity, terrestrialization, and algal lineages, are closely linked to major events in the evolution of plant angiosperms (Keegstra and Walton, 2006; Suzuki et al., 2006; Popper et al., 2011; Yin et al., 2014). Of these, CesA and CslD are commonly found in all land plants, and CesA and CslD are common to angiosperms (Yin et al., 2009, 2014; Popper et al., 2011). CesA and CslD are specific to grasses, whereas CesA and CslG are found in non-grass angiosperms (Keegstra and Walton, 2006; Suzuki et al., 2006; Popper et al., 2011; Yin et al., 2014). CesA family members are involved in the synthesis of various cell wall polysaccharides (Kaur et al., 2017). CesA/Csl genes encode proteins with both mannan and glucomannan synthase activity (Liepman et al., 2010). In Arabidopsis, the stems of csa1-knockout mutants have no glucomannan, indicating that CesA family genes have an exclusive role in mannan biosynthesis (Goubet et al., 2009). The members of the Csl family are involved in the synthesis of the β-1,4-linked glucan backbone of xyloglucan in Tropaeolum majus (Cocuron et al., 2007). CslDs in Arabidopsis are involved in mannan and Cellulose synthase, especially in tip-growing root hairs and pollen tubes (Park et al., 2011). In tobacco, CslD-overexpressing plants have high mannan synthase activity (Verhertbruggen et al., 2011). The rice OsCSLD1 gene is required for root hair morphogenesis (Kim et al., 2007). The CesA/Csl genes are involved in the synthesis of mixed-linkage glucan polymers in grasses such as rice and barley (Burton et al., 2006; Dobkin et al., 2009; Fincher, 2009). However, the roles of the remaining Csl family members (CslB, CslD, CslE, CslG) remain unclear.

Orchidaceae is one of the largest families of angiosperms whose members exhibit vastly different morphotypes, lifestyles, and remarkable adaptations to environmental conditions. Dendrobium is one of the largest genera of the Orchidaceae family, containing approximately 1,450 species (Zhang et al., 2016); these species are characterized by bioactive ingredients with immunomodulatory hepatoprotective activities, such as dendrobine and polysaccharides (Ng et al., 2012). Dendrobium officinale contains abundant polysaccharides in flesh stems, primarily glucomannan (GM) and galactoglucomannan (GGM) (Xing et al., 2014). Previous studies have shown that CesA/Csl genes, especially CesA/CsLs, play important roles in the synthesis of GM and GGM. However, little is known about the evolutionary pattern and functional diversification of the CesA/Csl genes involved in polysaccharide content accumulation. In this study, we performed a global analysis of the identification and characterization of the CesA/Csl family in three fully sequenced...
Phylogenetic Analysis of the CesA/Csl Proteins

The amino acid sequences of CesA/Csls were retrieved for nine species, including algae, mosses, lycophytes, monocots, and dicots: *Chlamydomonas reinhardtii* (Cre), *Volvox carteri* (Vca), *Physcomitrella patens* (Ppa), *Selaginella moellendorffii* (Smo), *Oryza sativa* (Osa), *A. shenzhenica* (As), *D. officinale* (Do), *P. equestris* (Peq), and *Arabidopsis thaliana* (AT). The protein sequences were downloaded from the Pfam database⁶. HMMsearch from the HMMER suite (version 3.1; Finn et al., 2011) was used to search for cellulose synthase proteins in *D. officinale*, *P. equestris*, and *A. shenzhenica* with a cutoff E-value of 1e⁻⁴. The molecular weight and theoretical isoelectric point (pI) of the proteins were predicted using the ExPASy website⁷. Transmembrane domains and subcellular locations were predicted using TMHMM⁸ and WoLF PSORT⁹, respectively.

MATERIALS AND METHODS

Identification and Characterization of Cellulose synthases in the Orchidaceae Family

The genome sequences of three Orchidaceae species, *D. officinale* (Zhang et al., 2016), *P. equestris* (Cai et al., 2014), and *A. shenzhenica* (Zhang et al., 2017) were downloaded from the NCBI database¹. The Cellulose synthase family proteins contain two different pfam domains: PF03552 for CesAs and PF00535 for Csls. The seed sequences of the two domains were then downloaded from the TIGRFAMs database². HMMsearch from the HMMER suite (version 3.1; Finn et al., 2011) was used to search for cellulose synthase proteins in *D. officinale*, *P. equestris*, and *A. shenzhenica* with a cutoff E-value of 1e⁻⁴. The molecular weight and theoretical isoelectric point (pI) of the proteins were predicted using the ExPASy website⁷. Transmembrane domains and subcellular locations were predicted using TMHMM⁸ and WoLF PSORT⁹, respectively.

Gene Structure, Distribution, and Protein Sequence Analyses

The Gene Structure Display Server tool¹⁰ (v2.0; Hu et al., 2015) was used to analyze the gene structure of all the CesA/Csls identified in the three Orchidaceae species. Gene distribution in the genome of *D. officinale* was visualized using the TBtools program (Chen et al., 2020), and conserved domains were identified with the NCBI Web CD-Search tool¹¹. MEME software¹² (v4.11.0) was used to search for sequences of motifs of CesA/Csl proteins, with a motif window length from 10 to 100 bp (Bailey et al., 2009), with the maximum number of motifs set at 10, and where motifs present in at least three proteins were considered true motifs. Multiple alignments of CesA/Csls in the three orchid species were conducted using DNAMAN software (version 9; Lynnon Biosoft Company, Quebec, QC, Canada).

Expression Analysis of CesA/Csl Genes in Different Organs and at Different Developmental Stages of Dendrobium officinale

Two transcriptome sequencing analyses were performed using four different organs of a 3-year-old *D. officinale* and the stems of three developmental stages of *D. officinale*. The four different organs of the 3-year-old *D. officinale* used in the present study include flower, leaf, flower, and root. Three different *D. officinale* developmental stages were investigated: 1-year old (Y1), 2 years old (Y2), and 3 years old (Y3). The plants were grown in glasshouses at the Mulberry Field Station of Zhejiang Academy of Agriculture Science (Hangzhou, China). The tissues were collected and frozen in liquid nitrogen and stored at −80°C until use. For each tissue, five plants were treated as an independent biological replicate, and three biological replicates were performed. Total RNA was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, United States) according to the manufacturer’s instructions and sequenced on an Illumina HiSeq 2000 platform. Subsequently, the expression profiles of all *D. officinale* genes were obtained with FPKM (fragments per kilobase of exon per million fragments mapped) values using cufflinks software¹³ (v2.2.1) according to annotated gene models with a GFF file. The expression profiles of the CesA/Csl genes from each sample were analyzed using the HemI program¹² with the average hierarchical clustering method.

Determination of Total Water-Soluble Polysaccharide Contents

Stems of *D. officinale* plants at five different growth stages, i.e., 3 months (3M), 9 months (9M), 1 year (Y1), 2 years (Y2), and 3 years (Y3), were collected (three replicates for each sample) and dried in an oven at 105°C until constant weight. The

---

¹https://www.ncbi.nlm.nih.gov
²http://tigrfams.jcvi.org/cgi-bin/index.cgi
³https://web.expasy.org/compute_pi/
⁴http://meme.nbcr.net/meme/
⁵https://wolfpsort.hgc.jp/
⁶https://phytozone.jgi.doe.gov/
⁷https://itol.embl.de/index.shtml
⁸https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi
⁹http://gsds.cbi.pku.edu.cn/
¹₀https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi
¹¹http://meme.nbcr.net/meme/
¹²http://cole-trapnell-lab.github.io/cufflinks
¹³http://hemi.biocuckoo.org/
3-M and 9-M *D. officinale* seedlings were cultured on half-strength Murashige and Skoog (MS) (Murashige and Skoog, 1962) media containing 0.1% activated carbon, 2% sucrose, and 0.6% agar (pH 5.4) in a growth chamber (25.5 ± 1°C, 45 μmol m-2 s-1 irradiance, 12-h light/dark photoperiod, 60% relative humidity). Y1, Y2, and Y3 *D. officinale* plants were grown in glasshouses at the Mulberry Field Station of Zhejiang Academy of Agriculture Science (Hangzhou, China). The samples were shattered into fine powder independently by a mixing mill (MM 400, Retsch). The total polysaccharide was extracted using the water extraction and alcohol precipitation method, and the contents of total polysaccharides were measured using the phenol–sulfuric acid method as described by Wang et al. (2021). Glucose was used as a reference for subsequent calculations in which total polysaccharides (µg/g dry weight) = (A + 0.0037)×7.981×V1×V2×V3÷W ×1000 = 626. 49×(A + 0.0037)÷W, where V1 represents the redissolved volume after alcohol precipitation (1 mL), V2 represents the volume of alcohol precipitation (0.2 mL), V3 represents the volume of water added during extraction (1 mL), and W represents sample weight in grams (1,000 g), serving as a coefficient converting milligrams to micrograms.

**Determination of Monosaccharide Contents**

After drying, the samples from the *D. officinale* stems of the five growth stages were shattered into fine powder independently by a mixing mill (MM 400, Retsch). The constituting monosaccharides, including mannose, glucose, and galactose, were extracted using the GC–MS/MS method (Sun et al., 2016). Briefly, 20 mg of powder was diluted in a 500 µL solution and the extracts were centrifuged at 14,000 rpm at 4°C for 3 min. Then, the supernatants were mixed, evaporated, and freeze-dried. The residue was used for further derivatization. The small-molecule carbohydrates and 100 µL of a solution of methoxyamine hydrochloride in pyridine (15 mg/mL) were mixed together and incubated at 37°C for 2 h. Then, 100 µL of BSTFA was added and the mixture was maintained at 37°C for 30 min after being vortexed. The mixture was subsequently diluted and analyzed by GC–MS/MS according to the methods of Gómez-González et al. (2010) and Sun et al. (2016) with modifications. Agilent 7890B gas chromatograph coupled to a 7000D mass spectrometer equipped with a DB-5MS column (30 m length × 0.25 mm i.d. × 0.25 µm film thickness, J&W Scientific, Santa Clara, CA, United States) was employed for GC–MS/MS analysis of the monosaccharides.

**RNA Extraction and Quantitative Real-Time PCR Analysis**

Total RNA was extracted from all the samples of *D. officinale* stems at the five growth stages (3M, 9M, Y1, Y2, and Y3) using TRIzol reagent (Invitrogen, Carlsbad, CA, United States). The potential contaminating genomic DNA was eliminated with DNase I. The quality of the RNA samples was checked with a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Beijing, China) and 1% denaturing agarose gels. The RNA was used as a template for first-strand cDNA synthesis using PrimeScript reverse transcriptase (TaKaRa Biotechnology, Dalian, China). Gene-specific primers were designed with the Primer Premier 5.0 program. The *DnActin* (comp205612_c0) gene was used as an internal standard for normalizing the gene expression data. The *DofCesA/DofCsl* expression levels were analyzed using a quantitative Real-time PCR (qRT–PCR) assay, which was performed by using a SYBR Green qPCR Kit (TaKaRa Biotechnology, Dalian, China) and a Stratagene MX3000P thermocycler (Agilent, Santa Clara, CA, United States). The PCR program was as follows: 95°C for 5 min followed by 40 cycles at 95°C for 15 s and then 60°C for 30 s. The relative gene expression levels were calculated with the 2−ΔΔCT method (Livak and Schmittgen, 2001). Three biological replicates were subjected to the analysis, each with three technical replicates. The expression levels in the different tissues were visualized using GraphPad Prism (v. 8.4.3).

**Statistical Analyses**

Statistical analysis was performed, and the average values and standard errors of the three replicates were calculated. SPSS software (v. 16.0) was used to determine the significant differences in polysaccharide and monosaccharide contents between the different developmental stages using a one-way ANOVA procedure and post hoc analysis. The value p < 0.05 indicates a significant difference and is represented by an asterisk (*) in the figures, and p < 0.01 indicates a very significant difference and is represented by two asterisks (**) in the figures. (****) indicates p value < 0.001 and (****) indicates p value < 0.0001.

**RESULTS**

**Genome-Wide Characterization of the Cellulose synthase Superfamily in Orchidaceae Species**

To investigate the potential roles of *Cellulose synthase* superfamily genes in orchids, we performed a genome-wide identification and characterization of CesA/Csl genes in three sequenced Orchidaceae species: *D. officinale* (Zhang et al., 2016), *P. equestris* (Cai et al., 2014), and *A. shenzhenica* (Zhang et al., 2017). A total of 125 *Cellulose synthase* superfamily genes were identified in the three orchid species, which were designated ‘DofCesA/DofCsl’ for *D. officinale*, ‘PeqCesA/PeqCsl’ for *P. equestris*, and ‘AsCesA/AsCsl’ for *A. shenzhenica* (Supplementary Table S1). The CesA superfamily members in the Orchidaceae family were classified into seven families: CesA, CslA, CslC, CslD, CslE, CslG, and CslH according to the sequence similarities with that of *A. thaliana* and *O. sativa* (Figure 1, Supplementary Figure S1, and Supplementary Table S1).

We identified 54 CesA/Csl proteins from the *D. officinale* genome, which comprised 11 CesAs and 43 Csls (Table 1). The 54 proteins were further classified into DofCesA (11 members) and six different Csl families, i.e., DofCslA (19 proteins), DofCslC (5 proteins), DofCslD (8 proteins), DofCslE (7 proteins), DofCslG...
Phylogenetic tree of the Cellulose synthase superfamily in D. officinale (Dof), rice (Osa) and Arabidopsis (At). The phylogenetic tree was constructed using MEGA 6.0 with the neighbor-joining (NJ) method and 1,000 bootstrap replicates. The CesA/Csl proteins were grouped into one CesA family and eight Csl families: CesA, CslA, CslB, CslC, CslD, CslE, CslF, CslG, and CslH. The families are marked by different arc lines and branch colors, and individual species are distinguished by triangles, squares, or rhombuses in different colors.
TABLE 1 | Physical and molecular characteristics of Cellulose synthase superfamily genes in *D. officinale*.

| Name                  | Gene_Protein_ID          | Start       | End         | Exon | Intron | CDS length (bp) | Size (aa) | pl     | MW       | TM domains | Subcellular localization |
|-----------------------|-------------------------|-------------|-------------|------|--------|-----------------|-----------|--------|----------|------------|--------------------------|
| DenCslE5 rna-XM_028698806.1 | 385324                | 356661      | 10          | 9    | 2214   | 737             | 6.27      | 82859.64| plas, vacu | 6                       |                          |
| DenCslE4 rna-XM_020841487.2 | 214425                | 3169725     | 16          | 15   | 3222   | 1073            | 6.80      | 120382.50| plas       | 6                       |                          |
| DenCslE3 rna-XM_020841488.2 | 345956                | 356662      | 11          | 10   | 2607   | 868             | 5.79      | 97737.54| plas       | 6                       |                          |
| DenCslE2 rna-XM_020841489.2 | 171675                | 3037542     | 13          | 12   | 3246   | 1081            | 6.62      | 122499.99| plas, golg_plas | 8                       |                          |
| DenCslE1 rna-XM_020839332.2 | 682541                | 6879711     | 9           | 8    | 1596   | 531             | 8.76      | 60970.18| plas, golg_plas | 5                       |                          |
| DenCslE10 rna-XM_020841490.2 | 714120                | 720825      | 10          | 9    | 1692   | 563             | 9.11      | 64700.67| plas, vacu | 6                       |                          |
| DenCslE9 rna-XM_020841491.2 | 768048                | 768638      | 11          | 10   | 2715   | 904             | 8.68      | 103177.55| plas       | 8                       |                          |
| DenCslE8 rna-XM_020841492.2 | 2037852               | 20384906    | 3           | 2    | 3423   | 1140            | 6.76      | 127811.67| plas       | 8                       |                          |
| DenCslE7 rna-XM_020841493.2 | 318232                | 321190      | 1           | 0    | 2376   | 791             | 8.84      | 88877.51| plas, vacu | 6                       |                          |
| DenCslE6 rna-XM_020841494.2 | 107776                | 1082979     | 5           | 4    | 2223   | 740             | 5.75      | 82838.96| plas       | 2                       | chloro_mito              |
| DenCslE5 rna-XM_020841495.2 | 492781                | 496772      | 2           | 1    | 2592   | 863             | 8.57      | 97991.25| plas       | 7                       |                          |
| DenCslE4 rna-XM_020841496.2 | 1151327               | 1156560     | 3           | 2    | 3462   | 1153            | 6.50      | 129061.48| plas       | 8                       |                          |
| DenCslE3 rna-XM_020841497.2 | 379240                | 379399      | 5           | 4    | 3519   | 1172            | 5.68      | 130380.35| plas       | 6                       |                          |
| DenCslE2 rna-XM_020841498.2 | 406655                | 418669      | 7           | 6    | 3543   | 1180            | 7.89      | 131733.04| plas       | 6                       |                          |
| DenCslE1 rna-XM_020841499.2 | 357017                | 363143      | 8           | 7    | 2190   | 729             | 8.12      | 83067.58| plas       | 8                       |                          |
| DenCslE0 rna-XM_020841500.2 | 366228                | 371432      | 7           | 6    | 1842   | 613             | 7.86      | 70609.91| plas       | 6                       |                          |
| DenCslE9 rna-XM_020841501.2 | 366228                | 371432      | 8           | 7    | 2190   | 729             | 8.50      | 83808.67| plas       | 8                       |                          |
| DenCslE8 rna-XM_020841502.2 | 390655                | 396656      | 8           | 7    | 2193   | 730             | 7.94      | 83696.62| plas       | 8                       |                          |
| DenCslE7 rna-XM_020841503.2 | 378096                | 395930      | 7           | 6    | 1395   | 464             | 7.17      | 54328.52| plas       | 2                       | nucl, cyto               |
| DenCslE6 rna-XM_020841504.2 | 378070                | 396512      | 8           | 7    | 2226   | 741             | 8.20      | 85761.01| plas       | 8                       |                          |

(Continued)
evolved from a separate cyanobacterial endosymbiotic event (Yin et al., 2009). CSLB was found only in the dicot species Arabidopsis, whereas CSLH was specific to the four monocots (Figure 2C). Notably, CSLG family members were found in all three Orchidaceae species.

The number of CesAs was rather conserved among angiosperms, with each generally having 8–13 members (except for S. moellendorfii), whereas the members in the other families greatly varied. P. patens had the highest number of CesAs (13 members) and CSLs (10 members), whereas S. moellendorfii had the lowest (five SmoCesAs and three SmoCSls). The CSLs were most abundant in S. moellendorfii (10 SmoCSls); however, the family had greatly contracted in other species, especially in P. equestris (containing only PeqCSla1). Notably, the CSLA family had significantly expanded in D. officinale (with 19 DofCSls) as compared with the remaining species (Figure 2B), which may have resulted in functional redundancy or innovation.

### Structural Conservation and Diversity of CesA/CSls in Orchidaceae Species

The sequences of conserved domains of the 125 CesA/CSl proteins in three Orchidaceae species were searched and analyzed. Five conserved domains: Cellulose synthase domain (cellulose_synt), zf-UDP, zf-RING_4, and two glycosyltransferase family domains (Glyco_transf_2, Glyco_transf_3) were found (Figure 3 and Supplementary Figures S3, S4). We found that the Cellulose synthase domain was present in CesA and CSLD/E/G/H, whereas all the CSLA/CSlC proteins were found to contain two glycosyltransferase domains (Table 2), which support the CSLA/CSlC families evolved from an independent cyanobacterial endosymbiotic event (Yin et al., 2009). The number of CesA and CSLD proteins that contain zf-UDP/zf-RING_4 domains varied among the three Orchidaceae species. For example, the zf-UDP domain was found in all nine AsCesAs proteins but was absent from three DofCSls and one PeqCesA. The motif patterns also varied among different species and families; however, we found species-specific motifs and similar conserved motif patterns within the same CesA/CSl family (Table 2, Figure 3 and Supplementary Figures S3, S4). These results suggested that CesA/CSl proteins among different families in Orchidaceae species might have different functional properties. Furthermore, multiple alignments of the predicted Cellulose synthase amino acid sequences showed that 18 of the 125 CesA/CSl proteins had no “D,D,D,QxxRW” integrated active site amino acid sequence (Supplementary Figures S5–S7), implying possible functional redundancy.

---

**Table 1 (Continued)**

| Name          | Gene Protein ID          | Start | End   | Exon | Intron | CDS length (bp) | Size (aa) | pl  | MW (kDa) | TM domains | Subcellular localization |
|---------------|-------------------------|-------|-------|------|--------|-----------------|-----------|-----|----------|-------------|--------------------------|
| DenCSlg1      | ma-XM_020827350.2       | 9834  | 20089 | 7    | 6      | 2139            | 712       | 7.18| 80943.69 | 7           | plas                     |
| DenCSlg2      | ma-XM_020836250.2       | 10246 | 120034| 6    | 5      | 2142            | 713       | 7.16| 81316.00 | 7           | plas                     |
| DenCSlg3      | ma-XM_020836251.2       | 68669 | 74378 | 5    | 4      | 1248            | 415       | 6.24| 47275.30 | 1           | cho                      |
| DenCSlh1      | ma-XM_028695005.1       | 18470706 | 18483281 | 9   | 8      | 2310            | 769       | 8.80| 87093.48 | 5           | plas                     |
| Group     | Gene_Protein ID      | Size (aa) | Domain 1         | Position (aa) | Domain 2   | Position (aa) |
|-----------|----------------------|----------|------------------|---------------|------------|---------------|
| DenCesA1  | rna-XM_020830109.2   | 1086     | cellulose_synt   | 359–1076      | zf-UDP     | 29–106        |
| DenCesA2  | rna-XM_02698906.1    | 737      | cellulose_synt   | 11–726        |            |               |
| DenCesA3  | rna-XM_026987805.1   | 758      | cellulose_synt   | 32–747        |            |               |
| DenCesA4  | rna-XM_020847107.2   | 868      | cellulose_synt   | 142–857       | zf-UDP     | 8–65          |
| DenCesA5  | rna-XM_020847069.2   | 1090     | cellulose_synt   | 357–1092      | zf-UDP     | 29–106        |
| DenCesA6  | rna-XM_020829579.2   | 1091     | cellulose_synt   | 358–1083      | zf-UDP     | 29–106        |
| DenCesA7  | rna-XM_020822139.2   | 904      | cellulose_synt   | 139–897       |            |               |
| DenCesA8  | rna-XM_020841487.2   | 1073     | cellulose_synt   | 345–1067      | zf-UDP     | 12–84         |
| DenCesA9  | rna-XM_020816047.2   | 1081     | cellulose_synt   | 375–1074      | zf-UDP     | 58–134        |
| DenCesA10 | rna-XM_020847106.2   | 980      | cellulose_synt   | 254–969       | zf-UDP     | 8–60          |
| DenCesA11 | rna-XM_026986667.1   | 541      | cellulose_synt   | 368–520       | zf-UDP     | 27–106        |
| DenCslD1  | rna-XM_020828005.2   | 528      | Glycos_transf_2  | 97–253        | Glyco_trans_2_3 | 186–375 |
| DenCslD2  | rna-XM_020817384.2   | 536      | Glycos_transf_2  | 105–263       | Glyco_trans_2_3 | 194–384 |
| DenCslD3  | rna-XM_020838280.2   | 559      | Glycos_transf_2  | 127–287       | Glyco_trans_2_3 | 215–408 |
| DenCslD4  | rna-XM_020838281.2   | 552      | Glycos_transf_2  | 120–281       | Glyco_trans_2_3 | 206–395 |
| DenCslD5  | rna-XM_020839292.2   | 572      | Glycos_transf_2  | 139–284       | Glyco_trans_2_3 | 227–444 |
| DenCslD6  | rna-XM_020830743.2   | 550      | Glycos_transf_2  | 99–259        | Glyco_trans_2_3 | 187–404 |
| DenCslD7  | rna-XM_026989893.1   | 516      | Glycos_transf_2  | 95–254        | Glyco_trans_2_3 | 182–387 |
| DenCslD8  | rna-XM_020830740.2   | 549      | Glycos_transf_2  | 98–258        | Glyco_trans_2_3 | 186–394 |
| DenCslD9  | rna-XM_020844197.2   | 487      | Glycos_transf_2  | 74–223        | Glyco_trans_2_3 | 163–371 |
| DenCslE1  | rna-XM_02867189.1    | 863      | cellulose_synt   | 94–856        |            |               |
| DenCslE2  | rna-XM_020841486.2   | 823      | Glycos_transf_2  | 168–323       | Glyco_trans_2_3 | 257–446 |
| DenCslE3  | rna-XM_020830967.2   | 693      | Glycos_transf_2  | 237–411       | Glyco_trans_2_3 | 320–519 |
| DenCslE4  | rna-XM_020829767.2   | 681      | Glycos_transf_2  | 231–386       | Glyco_trans_2_3 | 319–514 |
| DenCslE5  | rna-XM_020848683.2   | 690      | Glycos_transf_2  | 237–411       | Glyco_trans_2_3 | 325–520 |
| DenCslE6  | rna-XM_020830334.2   | 547      | Glycos_transf_2  | 116–276       | Glyco_trans_2_3 | 205–410 |
| DenCslE7  | rna-XM_020830335.2   | 546      | Glycos_transf_2  | 125–285       | Glyco_trans_2_3 | 214–419 |
| DenCslE8  | rna-XM_020839342.2   | 547      | Glycos_transf_2  | 131–290       | Glyco_trans_2_3 | 219–426 |
| DenCslE9  | rna-XM_020833422.2   | 564      | Glycos_transf_2  | 232–387       | Glyco_trans_2_3 | 320–514 |
| DenCslE10 | rna-XM_028698939.1   | 729      | cellulose_synt   | 98–388, 417–706 |            |               |
| DenCslE11 | rna-XM_020837355.2   | 613      | cellulose_synt   | 98–396, 435–601 |            |               |
| DenCslE12 | rna-XM_020837352.2   | 727      | cellulose_synt   | 98–394, 408–715 |            |               |
| DenCslE13 | rna-XM_020837351.2   | 729      | cellulose_synt   | 98–396, 409–717 |            |               |
| DenCslE14 | rna-XM_020837350.2   | 730      | cellulose_synt   | 98–393, 406–671 |            |               |
| DenCslE15 | rna-XM_026986599.1   | 464      | cellulose_synt   | 109–404       |            |               |
| DenCslE16 | rna-XM_020837350.2   | 741      | cellulose_synt   | 109–404, 416–882 |            |               |
| DenCslE17 | rna-XM_020827350.2   | 712      | cellulose_synt   | 87–383, 394–703 |            |               |
| DenCslE18 | rna-XM_020836250.2   | 713      | cellulose_synt   | 87–386, 394–706 |            |               |
| DenCslE19 | rna-XM_020836251.2   | 415      | cellulose_synt   | 86–385        |            |               |
| DenCslE20 | rna-XM_020836251.2   | 769      | cellulose_synt   | 98–360, 414–763 |            |               |
FIGURE 2 | Phylogenetic relationships between Cellulose synthase proteins from nine representative species: two chlorophyte algae (light green), a moss and spike moss (dark green), three orchid species (red), rice (yellow), and Arabidopsis (blue). The CesA/Csl proteins were grouped into nine families: CesA, CslA, CslB, CslC, CslD, CslE, CslF, CslG, and CslH. (A) The phylogenetic tree contains all 273 CesA/Csl protein sequences; (B) phylogenetic tree showing only the CslA cluster; (C) phylogenetic tree showing only the CslB/H/E/G cluster. The sequences in (B, C) were extracted from (A) and realigned.

To investigate the sequence diversity of the CesA/Csl proteins in the three orchid species, putative motifs were identified using MEME Suite 5.0.2; a total of 10 conserved motifs were found (Figure 3C and Supplementary Figures S3, S4). Although the motifs varied among the different species and families, the CesA/Csl proteins in each family in the same species shared several unique motifs. For example, in D. officinale, motif4, motif8, and motif2 were all observed in the DofCesA, DofCslD, DofCslG, and DofCslE families but not in the DofCesA11 family (Figure 3C and Supplementary Figure S5A). The DofCslA and DofCslC family proteins shared five common motifs: motif4, motif7, motif1, motif6, and motif5; however, DofCslA10 lacked motif 4. Among these motifs, motif7 and motif1 were only observed in DofCslA and DofCslC families. In P. equestris, the conserved motifs in PeqCesA, PeqCslD, PeqCslE, PeqCslH, and PeqCslG were similar; except for four proteins, all contained motif1–motif7 and motif10 (Supplementary Figure S3). Moreover, motif9 was specific to members of the PeqCslA and PeqCslC families. In A. shenzhenica, motif1, motif5, and motif2 were uniformly observed in all CesA/Csl proteins, with the exceptions of AsCesA9 and AsCslA2. Members of AsCesA, AsCslD, AsCslE, AsCslH, and AsCslG also shared similar motifs (Supplementary Figure S4), and motif 10 was specifically present in the AsCslA and AsCslC proteins. Taken together, these results revealed species-specific motifs and similar motifs shared among certain families. Thus, CesA/Csl proteins among different families and species might have different functional properties.

Gene Structure, Distribution, and Duplication Analyses
To gain more information on the evolutionary patterns of Cellulose synthase genes, the gene structures of the CesA/Csls in D. officinale, P. equestris, and A. shenzhenica were analyzed (Supplementary Table S1). The intron–exon structure of the CesA/Csls was highly diverse among the different families within the same species (Figures 4A, B). Nonetheless, the same family members have similar exon/intron numbers among the three Orchidaceae species. For example, except for four genes, AsCesAs and PeqCesAs usually have 13 or 14 exons. The exon number in DenCesAs varied more significantly compared to the other two orchids and ranged from 8 to 16 exons. CslAs generally contain 5–10 exons, and most have 9 or 10 exons.
The exon–intron structures and intron phases of the genes were relatively conserved within the same family (Figure 4B and Supplementary Figures S3, S4). In *D. officinale*, all of the DofCslCs had 5 exons and 4 introns. The gene structure varied the most within the CslD genes, from 0–7 or 8 introns/exons. These results suggested that possible functional diversity occurred within the same family and evolutionary conservation in Orchidaceae species.

Based on the physical positions of the *Cellulose synthase* genes in the *D. officinale* genome (Zhang et al., 2016), the distribution of 54 DofCesA/DofCsl genes was mapped to 30 scaffolds (Figure 4C and Table 1). Among them, 23 scaffolds had only one *Cellulose synthase* gene and two scaffolds each had two genes. Apart from the alternative splicing genes (indicated with green scale bars and asterisks), 14 tandemly duplicated genes (with two or more homologous genes ≤ 100 kb apart) were detected on six scaffolds (Figure 4C). A maximum number of 10 genes were detected on scaffold NW_021503689.1, all of which were DofCslA members (DofCslA10-19); after merging alternative splicing genes, there were four tandemly duplicated genes. DofCslA3, -4, -5 and DofCslA6, -7, -8 were also tandem duplicated genes located on scaffold NW_021319309.1 and NW_021319871.1, respectively. In addition, the DofCslEs also contain tandemly duplicated members that clustered on scaffold NW_02131965.1 (Figure 4C). Moreover, we found that the structures of alternative splicing genes, such as the number of exons and the length of introns, varied in different degrees among the splicing variants of the same gene (Supplementary Figure S8 and Supplementary Table S2). These results revealed that the expansion of the DofCesA/DofCsl genes mainly originated through gene duplication.

**Diverse Expression Patterns of DofCesA/DofCsl Genes in Dendrobium officinale**

To analyze the roles of the CesA/Csl genes in *D. officinale*, their expression patterns were investigated in different organs in *D. officinale* at different developmental stages. Transcriptome sequencing was performed on four organs of *D. officinale* (i.e., the leave, stem, flower, and root) and the stems of three different stages (Y1, Y2, and Y3). The expression levels of the DofCesA/DofCsl genes are presented in heatmaps in Figure 5 and Supplementary Table S3). The results revealed diverse
expression patterns of the DofCesA/DofCsl genes, suggesting functional divergence had occurred after gene expansion.

In the DofCesA family, three genes (DofCesA1, DofCesA5, and DofCesA8) were consistently expressed across different organs and developmental stages and were all significantly expressed in the flowers. Among them, DofCesA8 also had high expression levels in the roots and low expression in the leaves and stems; both DofCesA1 and DofCesA5 were expressed at low levels in the leaves, roots, and stems. Notably, DofCesA8 was also highly expressed in the Y1, Y2, and Y3 stems of D. officinale. The DofCslD and DofCslC family genes exhibited similar expression patterns, with low expression across the four organs and three developmental stages. The DofCslEs showed organ-specific expression patterns; for example, DofCslE1 and DofCslE4 exhibited moderate expression in the roots and leaves, respectively, whereas DofCslE6 was expressed in the leaves, roots, and stems across all three growth stages. DofCslH1 was expressed in three organs (flowers, leaves, and roots), with the
highest expression level in the roots. Moreover, the DofCslG family genes showed similar expression patterns; all exhibited higher expression levels in the flowers than in the roots, leaves, and stems. DofCslG1 and DofCslG2 also exhibited moderate expression in Y2 and Y3 stems. CslAs are known for their participation in the synthesis of polysaccharides, especially mannans and glucomannans. Interestingly, the DofCslAs exhibited apparent organ- and development-specific expression in *D. officinale*. Most of these genes were predominantly expressed in the stems and/or flowers, with very low levels in the roots and leaves. For example, DofCslA2, DofCslA14, and DofCslA15 had significantly higher expression levels in the stems than in the other organs (especially DofCslA15); DofCslA2 and DofCslA14 were also expressed in the flowers. However, several genes had almost no expression in any of the four organs. In addition, DofCslA6 was specifically expressed in the flowers. Among the three growth stages, some genes showed similar expression patterns, such as DofCslA7 and DofCslA8, both of which were expressed at higher levels in the Y2 stems. Notably, DofCslA14 and DofCslA15 had significantly high expression in Y3 stems. However, five of the 19 DofCslAs showed low levels of or no expression in the Y1, Y2, and Y3 stems. DofCslA2 was expressed at moderate levels in the Y3 stems and at low levels in the Y1 stems, whereas DofCslA3 showed a consistent expression pattern across all three stages, with the highest expression levels occurring in Y2 stems. These results revealed the possible functional specialization of DofCslAs in the polysaccharide synthesis in *D. officinale*.

**Water-Soluble Polysaccharide Contents in Dendrobium officinale**

A previous study showed that polysaccharide accumulation in *D. officinale* occurred predominantly in the stems (He et al., 2015). To further investigate the polysaccharide synthesis and functions of Cellulose synthase superfamily genes in Orchidaceae species, we measured the contents of total water-soluble polysaccharides, and their component monosaccharides (glucose, mannose, and galactose) in the stems of *D. officinale* at five different growth stages, i.e., 3 months (3M), 9 months (9M), 1 year (Y1), 2 years (Y2), and 3 years (Y3) (*Figure 6*). The results showed that the total polysaccharide content varied significantly among the different stages (*Figure 6D* and *Table 3*); it was lowest in the 3M stems (~18.06 mg/g) and highest in the Y3 stems.
FIGURE 6 | Histogram of the contents of water-soluble polysaccharides and monosaccharide components (mg/g) in the stems of *D. officinale* in different growth stages, including 3M, 9M, Y1, Y2, and Y3. (A) Mannose content; (B) glucose content; (C) galactose content; (D) total polysaccharide content. *p value < 0.05, **p value < 0.01, ***p value < 0.001, ****p value < 0.0001.

(~93.02 mg/g). The total polysaccharide contents varied slightly between the 3M and 9M stems and between Y2 and Y3 stems, but increased significantly from 3M/9M to Y2/Y3.

To determine the composition and contents of monosaccharides in the stems, the neutral monosaccharide composition was analyzed at the five growth stages. The results revealed that mannose was the most abundant neutral monosaccharide in the stems of *D. officinale* across all five growth stages, followed by glucose and galactose (Figure 6). The contents of glucose and mannose ranged from 2.40 to 20.01 mg/g and from 3.09 to 69.13 mg/g, respectively, with the highest detected in the Y3 stems and the lowest in the 3M stems. Mannose and glucose in the stems increased remarkably during *D. officinale* development, corresponding to the increases in total polysaccharides. Interestingly, the galactose content was highest in the 3M stems; however, this content decreased as *D. officinale* developed, lowest in the Y1 stems, and then increased slightly from Y2 to Y3.

### qPCR Analyses of DofCslA Genes in the Stems of *Dendrobium officinale* at Different Developmental Stages

To better understand the roles of CslAs in polysaccharide synthesis in *D. officinale*, qPCR analyses of the 19 DofCslAs were performed on *D. officinale* stems at five different growth stages: 3M, 9M, Y1, Y2, and Y3. The results for Y1, Y2, and Y3 were generally consistent with the transcriptome sequencing data. In general, most of the DofCslAs exhibited high expression levels in Y2 and/or Y3 stems (Figure 7 and Supplementary Table S4), which was in agreement with the accumulation of water-soluble polysaccharides and monosaccharides. Among them, five genes, i.e., *DofCslA2, DofCslA6, DofCslA11, DofCslA14*, and *DofCslA15*, exhibited significantly high expression levels in the Y3 stems (especially *DofCslA15*). Four genes, i.e., *DofCslA3, DofCslA4, DofCslA7*, and *DofCslA8*, were highly expressed in the Y2 stems compared with the Y3 stems. Notably, several genes, such as *DofCslA2, DofCslA3*, and *DofCslA7*, were also expressed in the 9M stems; *DofCslA8* also exhibited moderate expression in the 3M stems. In addition, six DofCslAs, such as *DofCslA16-19*, showed low expression levels. The primer sequences for qPCR analysis are provided in Supplementary Table S5. Taken together, these results suggest that DofCslA genes may exhibit functional redundancy and may be specialized for activity at different growth stages.

### DISCUSSION

Owing to its important roles in the biosynthesis of plant cell wall polysaccharides, primarily cellulose and hemicellulose, the *Cellulose synthase* superfamily has been the subject of intense
TABLE 3 | Contents of water-soluble polysaccharides and constituting monosaccharide (mg/g) in the stems of D. officinale at different growth stages, including 3M, 9M, Y1, Y2, and Y3.

| Sample | Mannose (mg/mL) Mean | Galactose (mg/mL) Mean | Glucose (mg/g FW) Mean | Total (mg/g FW) Mean |
|--------|----------------------|------------------------|------------------------|----------------------|
| 3M     | 3.0796               | 1.0831                 | 2.3171                 | 18.1847              |
| 3.0656 |                      |                        | 2.5021                 | 17.9997              |
| 3.1299 |                      | 1.0362                 | 2.3860                 | 18.0077              |
| 2.5761 |                      | 0.8282                 | 1.9033                 | 18.6568              |
| 2.6111 |                      | 0.9844                 | 2.1444                 | 18.5024              |
| 2.5794 |                      | 0.9453                 | 1.9782                 | 18.8652              |
| 9M     | 35.5856              | 34.4407                | 16.2450                | 42.4609              |
| 35.2837|                      | 0.2770                 | 0.2827                 | 42.2383              |
| 32.4527|                      | 0.2910                 | 17.3928                | 40.8961              |
| Y1     | 67.5208              | 66.1652                | 17.7653                | 88.3224              |
| 68.2550|                      | 0.4390                 | 0.4383                 | 88.2296              |
| 62.7199|                      | 0.4350                 | 18.2186                | 86.0851              |
| Y2     | 73.0744              | 69.1258                | 0.5370                 | 96.1469              |
| 70.4218|                      | 0.4960                 | 19.3982                | 94.5241              |
| 63.8812|                      | 0.5030                 | 18.7632                | 88.4010              |

FIGURE 7 | Expression levels of DofCslA genes in the stems of D. officinale in different growth stages (including 3M, 9M, Y1, Y2, and Y3), as determined by qRT-PCR analysis. The results are shown as the means ± SDs of three independent experiments. The presented expression levels are relative to the expression of the reference gene. *p value < 0.05, **p value < 0.01, ***p value < 0.001, ****p value < 0.0001.
studies. Much effort has been devoted to the identification of the CesA/Csl family in plants, including mosses, lycophytes, monocots, and dicots, as well as in green algae (Persson et al., 2007; Yin et al., 2014; Takata and Taniguchi, 2015; Li et al., 2017, 2020; Song et al., 2018; Cui et al., 2020). However, knowledge of the Cellulose synthase superfamily is still lacking in the family Orchidaceae. In the present study, we performed genome-wide characterizations of CesA/Csl family members in three fully sequenced Orchidaceae species (Cai et al., 2014; Zhang et al., 2016, 2017), which revealed the evolution, structural divergence, and expression profiles of CesA/Csls.

**Species- and Family-Specific Expansion of CesA/Csl Proteins**

The Cellulose synthase superfamily belongs to the glycosyltransferase-2 superfamily, which is widely distributed throughout the plant kingdom (Kim et al., 2007; Cantarel et al., 2009). In the present study, we identified a total of 125 Cellulose synthase superfamily members from the three sequenced orchids using two Pfam domains (PF03552/PF00535; Table 1 and Supplementary Table S1). Among them, 54 DofCesAs/DofCsls, 37 PeqCesAs/PeqCsls, and 34 AsCesAs/AsCsls were identified in the genomes of *D. officinale*, *P. equestris*, and *A. shenzhenica*, respectively. The results revealed species-specific expansion of the Cellulose synthase superfamily in *D. officinale*.

In land plants, the Cellulose synthase superfamily can be further classified into a CesA family and nine Csl families: CslA-CslH and CslJ (Fincher, 2009; Yin et al., 2014; Schwerdt et al., 2015). Previous studies indicate that PF03552 encompasses the CesA family and seven Csl families (CslB/D/E/F/G/H/J), whereas PF03552 is absent from the CslA/CslC families (Yin et al., 2009). In the present study, the CesA superfamily genes in Orchidaceae were grouped into one CesA family and six Csl families: CslA, CslC, CslD, CslE, CslG, and CslH (Figure 1 and Supplementary Figure S1). The number of CesAs in the three orchid species was largely the same, while the number of Csls greatly varied, from 25 AsCsIs to 43 DofCsls. Interestingly, the Csl proteins exhibited family-specific expansion among the three orchid species; for example, there were 19 DofCslAs, almost twice that in the other two orchids. The results of similar expansion were also observed in CesA families such as PeqCslDs and PeqCshs. The expansion of CesA/Csl families likely resulted in functional redundancy or innovation. To further investigate the evolution of CesA/Csls, a phylogenetic tree based on 273 amino acid sequences from nine representative plant species was constructed (Figure 2). CesA/Csls were classified into one CesA family and eight Csl families: CslA, CslB, CslC, CslD, CslE, CslF, CslG, and CslH. These results are consistent with those of previous findings in which CesA, CslC, and CslD were found to be conserved in all land plants (Farrokhi et al., 2006; Yin et al., 2014). CslF and CslH may be restricted to monocots, and CslFs were absent from the three Orchidaceae species. CslG members were previously thought to be confined to eudicots (Fincher, 2009). However, CslG family members were found in all three Orchidaceae species, supporting the proposal that CslG should no longer be considered a dicot-specific family (Yin et al., 2014). In addition, our results strongly supported previous findings that the numbers of CesA/Csls vary greatly among different plant species (Paterson et al., 2009; Takata and Taniguchi, 2015; Song et al., 2018; Zou et al., 2018; Li et al., 2020). The number of CesAs is largely conserved among angiosperms, most of which have 8–13 members (except for *S. moellendorffii*), whereas the members in other families greatly varied among different species, indicating that extensive expansion and diversification have occurred.

We identified three proteins in *P. equestris* and *A. shenzhenica* that had extremely short amino acid sequences (less than 300 aa). The molecular weights of the remaining 122 CesA/Csls in the three orchid species ranged from 42.84 to 146.86 kDa. The longest protein was AsCslD5-1392 aa, which is shorter than the 2038 aa sequence in tomato (Song et al., 2018). Previous studies have indicated that cellulose is synthesized by plasma membrane-localized Cellulose synthase complexes with access to the cytosolic GDP-glucose pool (Taylor, 2008; Endler and Persson, 2011; McFarlane et al., 2014; Schneider et al., 2016). In the present study, subcellular localization predictions showed that most of the CesA/Csls were located on the plasma membrane.

**Sequence Conservation and Diversification of CesA/Csl Proteins in Orchids**

The GT2 superfamily is characterized by conserved cytosolic substrate binding and catalytic residues. These residues are positioned in a loop between TM domains 2 and 3 and contain a D,D,D,QXXRW motif (Pear et al., 1996). In addition, plant CesA proteins also contain an extended N-terminal zinc-finger domain and two insertions in the catalytic loop, which are presumably involved in specific functions in higher plants, such as multimerization (Pear et al., 1996; Sethaphong et al., 2013). In the present study, five conserved domains were identified from the 125 orchid CesA/Csls (Figure 3 and Supplementary Figures S3, S4). We found that the Cellulose synthase domain was present in CesA and CslD/E/G/H, whereas all the CslA/CslC proteins were found to contain two glycosyltransferase domains, which support the CslA/CslC families evolved from an independent cyanobacterial endosymbiotic event (Yin et al., 2009). The number of CesA and CslD proteins that contain zf-UDP/zf-RING_4 domains varied among the three Orchidaceae species. For example, the zf-UDP domain was found in all nine AsCesAs proteins but was absent from three DofCesAs and one PeqCesA. The motif patterns also varied among different species and families; however, we found species-specific motifs and similar conserved motif patterns within the same CesA/Csl family (Figure 3 and Supplementary Figures S3, S4). These results suggested that CesA/Csl proteins among different families in Orchidaceae species might have different functional properties. Furthermore, multiple alignments of the predicted Cellulose synthase amino acid sequences showed that 18 of the 125 CesA/Csl proteins...
had no “D,D,D,QxxRW” integrated active site amino acid sequence (Supplementary Figures S5–S7), implying possible functional redundancy.

Gene structure and duplication analysis can facilitate the understanding of gene evolution, structural divergence, and functional conservation within a gene family (Worberg et al., 2008; Xu et al., 2012; Nawaz et al., 2017). We found highly diverse structures among the different families but similar exon/intron numbers in the same family among the three orchid species (Figure 4B). The results revealed that there was functional diversity within the same family and that evolutionary conservation occurred among different orchid species. Moreover, a total of 14 tandemly duplicated genes were detected (Figure 4C and Table 1), indicating that the expansion of DofCesA/DofCsl genes might result from tandem duplication. These results provide insight into the evolutionary divergence and origins of CesA/Csl in Orchidaceae species.

### Functional Divergence of DofCesA/DofCsl Genes in Different Organs and Growth Stages of Dendrobium officinale

Cellulose synthase superfamily genes are involved in the synthesis of the subunits of cellulose and hemicellulose (Burton et al., 2004; Scheller and Ulvskov, 2010; Takata and Taniguchi, 2015). The expansion and diversification of the plant CesA superfamily are linked tightly with major events in the evolution of plant and algal lineages, including multicellularity, terrestrialization, and vascularization (Popper et al., 2011). To understand the functions of the CesA/Csl genes in orchids, we investigated gene expression profiles in different organs of D. officinale and at different developmental stages (Figure 5). The results revealed diverse expression patterns of the DofCesA/DofCsl genes, suggesting functional divergence had occurred after gene expansion.

Three DofCesAs (DofCesA1, DofCesA5, and DofCesA8) were significantly expressed in the flowers of D. officinale. DofCesA8 also exhibited high expression in the roots and moderate expression in the leaves and stems. Except for DofCesA6, which exhibited low expression in the flowers, the expression levels of the remaining DofCesAs were minimal in all the tested organs. In addition, DofCesA8 was also highly expressed in Y1 stems and moderately expressed in the Y2 and Y3 stems. In the Arabidopsis genome, ten AtCesAs were identified and found to be responsible for primary (CesA1-3, -5, -6, and -9) and secondary (AtCesA4, -7, and -8) cell wall synthesis (Taylor et al., 2003; Burton et al., 2004; Takata and Taniguchi, 2015). Mutations in secondary cell wall AtCesAs result in collapsed vasculature or irregular xylem phenotypes (Turner and Somerville, 1997; Taylor et al., 2003). Moreover, AtCesA1 and AtCesA3 also play roles in flowering; mutations in atcesa1 and atcesa3 result in gametic lethal phenotypes, whereas single-knockout AtCesA2, -5, -6, and -9 mutants exhibit more moderate phenotypes (Scheible et al., 2001; Caño-Delgado et al., 2003). The AtCesA1 was clustered in the same subgroup with DofCesA1, and AtCesA3 was clustered with DofCesA8. Therefore, the results, combined with the expression patterns, suggested that the three DofCesAs likely play roles in flower organ development, and that DofCesA8 may play additional roles in cellulose deposition in the stems of D. officinale.

In plants, the synthesis of hemicellulose occurs in the Golgi membrane by proteins encoded by Csl genes (Kaur et al., 2017). In land plants, CslAs and CslGs encode proteins that are involved in the synthesis of mannan and the β-1,4-linked glucan backbone of xyloglucan, respectively (Liemann et al., 2005, 2010; Cocuron et al., 2007). Moreover, CslD members have been reported to have mannan and Cellulose synthase activities, especially in tip-growing root hairs and pollen tubes of plant species such as rice and Arabidopsis. In rice, the OsCSDL1 gene is required for root hair morphogenesis (Kim et al., 2007). AtCslD3 also plays a role in Cellulose synthesis, as atcsld3 mutants exhibit defects in polarized growth of root hairs and abnormal distribution of cellulose and xyloglucan (Pauly et al., 2001; Park et al., 2011). CslGs are involved in the biosynthesis of the β-1,4-linked glucan backbone of xyloglucan (Cocuron et al., 2007). However, the expression of the DofCslD and DofCslC family genes was either low or absent from the four organs and across the three developmental stages, indicating minimal effects on polysaccharide synthesis in D. officinale. Interestingly, most of the DofCslAs were predominantly expressed in the stems and/or flowers. DofCslA2, -14, and -15 had significantly high expression levels in the stem; DofCslA6 was specifically expressed in the flowers. Moreover, DofCslAs were differentially expressed among the three growth stages. For example, DofCslA14 and DofCslA15 exhibited significantly high expression in Y3 stems, while DofCslA7 and DofCslA8 were expressed at higher levels in Y2 stems. Five of the 19 DofCslAs exhibited low expression or none in the stems of plants at all three growth stages. These results revealed functional divergence and possible redundancy of the DofCslAs involved in polysaccharide synthesis in D. officinale. The functional role of the CslB/E/G families remains unclear, but CslF/H genes have been suggested to be involved in the synthesis of mixed-linkage glucans (MLGs) (Prasad et al., 2005; Burton et al., 2006). We found that the DofCslGs all presented higher expression levels in the flowers than in the rest of the organs. DofCslG1 and DofCslG2 also had moderate expression in Y2 and Y3 stems. Our phylogenetic study showed that DofCslE had significantly expanded in D. officinale compared with other species (Figure 2). However, the expression of most DofCslEs was very low in different organs and growth stages. Thus, further investigations are needed to elucidate the exact roles of DofCslE/G/H family genes.

### Polysaccharide Accumulation and Roles of DofCslAs in the Stems of Dendrobium officinale

Containing approximately 1,450 species, Dendrobium is one of the largest genera of the Orchidaceae family (Zhang et al., 2016). Dendrobium species are characterized by
their diverse growth habits and bioactive constituents with immunomodulatory hepatoprotective activities, such as dendrobine and polysaccharides (Ng et al., 2012). *D. officinale* stems contain abundant polysaccharides, primarily those consisting of GM and GGM (Xing et al., 2014). The high amount of polysaccharides in *D. officinale* stems supposedly helps maintain osmotic pressure and improve drought tolerance, as this species is epiphytic in natural habitats (Zotz and Tyree, 1996; Wu et al., 2009). CslAs encode proteins with both mannann and glucomannan synthase activity (Liepman et al., 2005, 2010). In *Arabidopsis*, the stems of *csla* knockout mutants have no glucomannan, indicating an exclusive role in mannann biosynthesis of the CslA family genes (Goulet et al., 2009).

To determine the roles of the 19 DofCslAs in polysaccharide accumulation in *D. officinale* stems across different growth stages, we performed qRT-PCR and measured the content of total polysaccharides and their monosaccharide components in the stems of 3M, 9M, Y1, Y2, and Y3 plants. The total polysaccharide content was lowest in the 3M stems and highest in the Y3 stems (Figure 6 and Table 3). The content varied slightly between 3M and 9M stems but increased significantly in the mature stems. Monosaccharide composition analysis showed that mannose was the most abundant component across all development stages, supporting previous findings in *Dendrobium* species (Meng et al., 2013). The amounts of glucose and mannose in the stems also experienced remarkable increases from 3M to Y3. In agreement with the changes in total polysaccharide and monosaccharide component contents, most of the DofCslAs exhibited high expression levels in the Y2 and/or Y3 stems (Figure 7). Among them, five genes exhibited significantly high expression levels in Y3 stems (especially DofCslA15); four genes were highly expressed in Y2 stems. Thus, these genes likely play roles in the biosynthesis of mannann and glucomannan in the stems of *D. officinale*. In addition, six genes showed low expression levels, suggesting that possible redundancy occurred after gene expansion. These results imply that the DofCslAs may experience functional specialization with respect to polysaccharide accumulation in different growth stages.

**DATA AVAILABILITY STATEMENT**

The raw data of RNA-seq experiment is deposited in Sequence Read Archive (NCBI): PRJNA680456 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA680456) and PRJNA762115 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA762115). All data and material used in this study are available from the corresponding author upon reasonable request.

**AUTHOR CONTRIBUTIONS**

YW and CS conceived and designed the experiments, performed the experiments, analyzed the data, prepared the figures and/or tables, authored or reviewed the drafts of the study, and approved the final draft. KZ and YC analyzed the data, authored or reviewed the drafts of the study, and approved the final draft. XC, QW, and HW analyzed the data, contributed reagents, materials, and analysis tools, authored or reviewed the drafts of the study, and approved the final draft. CS conceived the experiments, authored or reviewed the drafts of the study, and approved the final draft. All authors contributed to the article and approved the submitted version.

**FUNDING**

This study was funded by the Zhejiang Provincial Natural Science Foundation (grant number LQ20C150002), Zhejiang Science and Technology Major Program on Agricultural New Variety Breeding (grant number 2021C02071-5).

**SUPPLEMENTARY MATERIAL**

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.777332/full#supplementary-material

**Supplementary Figure S1 | Phylogenetic tree of the Cellulose synthase superfamily members in *D. officinale* (Do), *A. shenzhenica* (Apo), and *P. equestris* (Peq).** The phylogenetic tree was constructed using MEGA 6.0 with the neighbor-joining (NJ) method and 1000 bootstrap replicates. The CesA/Csl proteins were grouped into one CesA family and six Csl families: CslA, CslC, CslD, CslE, CslF, and CslH. The families are marked by different arc lines and branch colors. The *D. officinale*, *A. shenzhenica*, and *P. equestris* are distinguished by red, green and blue rhombuses, respectively.

**Supplementary Figure S2 | Phylogenetic tree of *Cellulose synthase* superfamily members from nine representative plant species.** The phylogenetic tree was constructed using MEGA 6.0 with the neighbor-joining (NJ) method and 1000 bootstrap replicates. The CesA/Csl proteins were grouped into nine families: CesA, CslA, CslB, CslC, CslD, CslE, CslF, CslG, and CslH. The species included (with the gene code prefixes shown in parentheses) are as follows: *Chlamydomonas reinhardtii* (Cre), *Volvox carteri* (Vca), *Physcomitrella patens* (Ppa), *Selaginella moellendorffii* (Smr), *A. shenzhenica* (Apo), *D. officinale* (Den), *P. equestris* (Peq), *Oryza sativa* (Osa), and *Arabidopsis thaliana* (AT). The families are marked by different arc lines and branch colors.

**Supplementary Figure S3 | Phylogenetic, conserved domain and motif and gene structure analyses of *Cellulose synthase* synthetases in *P. equestris* (Peq).** (A) Phylogenetic relationships and conserved domains of the PeqCesA/PeqCsl proteins. The scale bar (aa) indicates the amino acid position in the corresponding conserved domain. (B) Phylogenetic relationships, conserved motifs and gene structure of PeqCesA/PeqCsl genes. UTRs and exons are indicated using green and yellow rectangles, respectively. The solid lines indicate introns. The numbers above the solid lines represent the intron phase. Scale bars (bp/aa) indicate the length/amino acid position of corresponding genes/proteins; (C) Sequence logos of the 10 motifs.

**Supplementary Figure S4 | Phylogenetic, conserved domain and motif and gene structure analyses of *Cellulose synthase* synthetases in *A. shenzhenica* (Apo).** (A) Phylogenetic relationships and conserved domains of the AsCesA/AsCsl proteins. The scale bar (aa) indicates the amino acid position in the corresponding conserved domain. (B) Phylogenetic relationships, conserved motifs and gene structure of the AsCesA/AsCsl genes. Exons are indicated in yellow rectangles. The solid lines indicate introns. The numbers above the solid lines represent the intron phase. The scale bars (bp/aa) indicate the length/amino acid position of corresponding genes/proteins; (C) sequence logos of the 10 motifs.
REFERENCES

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic. Acids Res. 37, W202–W208. doi: 10.1093/nar/gkp355

Burton, R. A., Shirley, N. J., Harvey, A. J., and Fincher, G. B. (2004). The CesA gene family of barley. Quantitative analysis of transcripts reveals two groups of co-expressed genes. Plant Physiol. 134, 224–236. doi: 10.1104/pp.103.032904

Burton, R. A., Wilson, S. M., Hrmova, M., Harvey, A. J., Shirley, N. J., Medhurst, A., et al. (2006). Cellulose synthase-like CslF genes mediate the synthesis of cell wall (1,3;1,4)-β-D-glucans. Science 311, 1940–1942. doi: 10.1126/science.1122975

Cai, J., Liu, X., Vanneste, K., Proost, S., Tsai, W. C., Liu, K. W., et al. (2014). The Cellulose synthase-like-like CslF genes mediate the synthesis of cell wall (1,3;1,4)-β-D-glucans. Science 311, 1940–1942. doi: 10.1126/science.1122975

Cai, J., Liu, X., Vanneste, K., Proost, S., Tsai, W. C., Liu, K. W., et al. (2014). Conserved cysteine residues are marked by red frames. Nat. Genet. 47, 65–72.

Cao-Delgado, A., Penfield, S., Smith, C., Cai, J., Liu, X., Vanneste, K., Proost, S., Tsai, W. C., Liu, K. W., et al. (2014). The Cellulose synthase-like-like CslF genes mediate the synthesis of cell wall (1,3;1,4)-β-D-glucans. Science 311, 1940–1942. doi: 10.1126/science.1122975

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Lombard, V., and Brunak, S. (2009). Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic. Acids Res. 37, D233–D238. doi: 10.1093/nar/gkn663

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., and Xia, R. (2020). T TBtools, an integrative toolkit developed for interactive analyses of big biological data. Mol. Plant 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Cocuron, J. C., Lerouxel, O., Drakaki, G., Alonso, A. P., Liepman, A. H., Keegstra, K., et al. (2007). A gene from the Cellulose synthase-like C family encodes a beta-1,4 glucan synthase. Proc. Natl. Acad. Sci. U. S. A. 104, 8550–8555. doi: 10.1073/pnas.0703131104

Cui, Z., Sun, G., and Zhao, Q. (2020). Phylogenetic and functional analysis of CesA genes in cotton. Res. Square [Preprint]. doi: 10.21203/rs.3.rs.101509/v1

Desprez, T. T., Juraniec, M. M., Crowell, E. F. E., Jouy, H. H., Pochylava, Z., Parcy, F., et al. (2007). Organization of Cellulose synthase complexes involved in primary cell wall synthesis in Arabidopsis thaliana. Proc. Natl. Acad. Sci. U. S. A. 104, 15527–15532. doi: 10.1073/pnas.0706591104

Doblin, M. S., Pettolino, F. A., Wilson, S. M., Campbell, R., Burton, R. A., and Fincher, G. B. (2009). A barley Cellulose synthase-like-like CSLH gene mediates a beta-1,4 glucan synthase. Proc. Natl. Acad. Sci. U. S. A. 104, 8550–8555. doi: 10.1073/pnas.0703131104

Endler, A., and Persson, S. (2011). Cellulose synthases and synthesis in Arabidopsis. Mol. Plant 4, 199–211.

Farrokhni, N., Burton, R. A., Brownfield, L., Hrmova, M., Wilson, S. M., Bacic, A., et al. (2006). Plant cell wall biosynthesis: genetic, biochemical and functional genomics approaches to the identification of key genes. Plant Biotechnol. J. 4, 145–167. doi: 10.1111/j.1467-7652.2005.00169.x

Fincher, G. B. (2009). Evolutionary times in our understanding of cell wall biosynthesis and remodeling in the grasses. Plant Physiol. 149, 27–37. doi: 10.1104/pp.110.130096

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic. Acids Res. 39, 29–37. doi: 10.1093/nar/gkq367

Gómez-González, S., Ruiz-Jiménez, J., Priego-Capote, F., and Luque de Castro, M. D. (2010). Quantitative and qualitative sugar profiling in olive fruits, leaves, and stems by Gas Chromatography-Tandem Mass Spectrometry (GC-MS/MS) after ultrasound-assisted Leaching. J. Agric. Food Chem. 58, 12292–12299. doi: 10.1021/jf102350s

Goubet, F., Barton, C. J., Mortimer, J. C., Yu, X., Zhang, Z., Miles, G. P., et al. (2009). Cell wall glucanmannan in Arabidopsis is synthesised by CSLA glycosyltransferases, and influences the progression of embryogenesis. Plant J. 60, 527–538. doi: 10.1111/j.1365-313X.2009.03977.x

He, C., Zhang, J., Liu, X., Zeng, S., Wu, K., Yu, Z., et al. (2015). Identification of genes involved in biosynthesis of mannan polysaccharides in Dendrobium officinale by RNA-seq analysis. Plant Mol. Biol. 88, 219–231.

Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., and Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics 31, 1296–1297. doi: 10.1093/bioinformatics/btu817

Kaur, S., Dhugga, K. S., Beech, R., and Singh, J. (2017). Genome-wide analysis of the Cellulose synthase-like-like (CslF) gene family of barley. Quantitative analysis of transcripts reveals two groups of genes. UTRs and exons are indicated using blue and yellow rectangles, respectively. The solid lines indicate introns. The scale bar (bp) indicates the length of the corresponding transcripts.

Keeble, D. officinale genome sequence of the orchid Phalaenopsis equestris (1,3;1,4)-β-D-glucan synthase. Proc. Natl. Acad. Sci. U. S. A. 103, 69–94. doi: 10.1146/annurev-arplant-052109-090351

Meng, L. Z., Lv, G. P., Hu, D. J., Cheong, K. L., Xie, J., Zhao, J., et al. (2013). Effects of ultrasound-assisted Leaching. J. Agric. Food Chem. 61, 241–252. doi: 10.1021/jf303052z

Ng, T. B., Liu, J., Wong, J. H., Ye, X., Sze, S. C. W., Tong, Y., et al. (2012). Review data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. J. Plant Physiol. 175, 551–556. doi: 10.1016/j.jplph.2012.04.009

Nishiyama, Y. (2009). Structure and properties of the cellulose microfibril. J. Wood Sci. 55, 241–249.

Park, S., Szmuelski, A. L., Gu, F., Guo, F., and Nielsen, E. (2011). A role for CSLD3 during cell-wall synthesis in apical plasma membranes of tip-growing root-hair cells. Nat. Cell Biol. 13, 973–980. doi: 10.1038/ncb2294

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., et al. (2009). The Sorghum bicolor genome and the diversification of grasses. Nature 457, 551–556. doi: 10.1038/nature07723
