Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

Tissue-Specific Transcriptome Analysis Reveals Lignocellulose Synthesis Regulation in Elephant Grass (*Pennisetum purpureum* Schum.)

Wenqing Zhang
Qilu University of Technology
Shengkui Zhang
Qilu University of Technology
Xianqin Lu
Qilu University of Technology
Can Li
Qilu University of Technology
Xingwang Liu
Qilu University of Technology
Geyu Dong
Qilu University of Technology
Tao Xia (txia@qlu.edu.cn)
Qilu University of Technology

Research article

Keywords: Elephant Grass (Pennisetum purpureum Schum.), RNA-Seq, KEGG (Kyoto Encyclopedia of Genes and Genomes), WGCNA (Weighted Gene Co-Expression Network Analysis), Cellulose, Lignin

Posted Date: August 25th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-37358/v2

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

Version of Record: A version of this preprint was published at BMC Plant Biology on November 19th, 2020. See the published version at https://doi.org/10.1186/s12870-020-02735-3.

Abstract

Background: The characteristics of elephant grass, especially its stem lignocellulose, are of great significance for its quality as feed or other industrial raw materials. However, the research on lignocellulose biosynthesis pathway and key genes is limited because the genome of elephant grass has not been deciphered.

Results: In this study, RNA sequencing (RNA-seq) combined with lignocellulose content analysis and cell wall morphology observation using elephant grass stems from different development stages as materials were applied to reveal the genes that regulate the synthesis of cellulose and lignin. A total of 3852 differentially expressed genes (DEGs) were identified in three periods of T1, T2, and T3 through RNA-seq analysis. Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis of all DEGs showed that the two most abundant metabolic pathways were phenylpropane metabolism, starch and sucrose metabolism, which were closely related to cell wall development, hemicellulose, lignin and cellulose synthesis. Through weighted gene co-expression network analysis (WGCNA) of DEGs, a 'blue' module highly associated with cellulose synthesis and a 'turquoise' module highly correlated with lignin synthesis were exhibited. A total of 43 candidate genes were screened, of which 17 had function annotations in other species. Besides, by analyzing the content of lignocellulose in the stem tissues of elephant grass at different developmental stages and the expression levels of genes such as *CesA*, *PAL*, *CAD*, *C4H*, *COMT*, *CCoAMT*, *F5H* and *CCR*, it was found that the content of lignocellulose was related to the expression level of these structural genes.

Conclusions: This study provides a basis for further understanding the molecular mechanisms of cellulose and lignin synthesis pathways of elephant grass, and offers a unique and extensive list of candidate genes for future specialized functional studies which may promote the development of high-quality elephant grass varieties with high cellulose and low lignin content.

Background

Fiber, which is mainly composed of three biological macromolecules of cellulose, lignin, and hemicellulose, plays an essential role in plant growth and stress responses. The formation of fibers and deposition of components strengthen special types of cells, such as fiber cells and vessel cells, which form mechanical tissue to provide structural support and protection for plant cells, and produce negative pressure gradient to protect plant cells during transpiration. Fiber formation is a complex process, which requires the coordination and balance of multiple metabolic pathways [1-2].

Cellulose is the most essential component of fiber. In the process of cellulose synthesis, cellulose synthase (CesA) monomers form cellulose synthase complex (CSC), which catalyzes the synthesis of the dextran chain of cellulose by related substrates. Generally, there are two or more CesA proteins involved in the synthesis of cellulose. To date, *CesA* has been cloned in bread wheat (*Triticum aestivum*, L.), *Arabidopsis thaliana*, maize (*Zea mays* L.), poplar (*Populust remuloides*) and some other plants, their

function has been clarified [3-5]. In poplar, *PtrcesA2* and *PtrcesA1* were homologous to *Arabidopsis* mutants *IRX1* and *IRX3*, respectively. They were all expressed in the process of xylem secondary wall formation of poplar. It was speculated that these two genes may be related to cell wall formation [6].

In addition to cellulose synthase genes, other genes also play an important role in cellulose biosynthesis, such as *KORRIGAN*, sucrose phosphate synthase, cytoskeleton protein, lipid transfer protein, oxidized protein [7-8]. *KORRIGAN* encodes 1,4-β-D-glucosidase, and its mutant exhibited a decrease in cellulose and an increase in pectin synthesis, which led to excessive callus formation. In addition, it has also been found that MYB transcription factor can regulate the expression level of cellulose synthase gene, thereby changing the cellulose content and stalk strength of plants. Researchers used candidate gene genetic transformation and map-based cloning technology to study the regulation mechanism of transcription factor (OsMYB103L) on cell wall synthesis in rice (*Oryza sativa* L.) [9].

Hemicellulose is another major component of the secondary wall. It can form a network with cellulose to make the cell wall more compact. The known hemicelluloses include xyloglucan, xylan, β-1,3 (1,4) -d-glucan and mannan. Although hemicellulose is a kind of heteropolysaccharide, most hemicellulose has a single skeleton. Except for xylan, all the main chains of hemicellulose are synthesized by cellulose synthase-like (CSL) [10]. Xylan also takes β- 1,4-glycosidic bond as the main chain, but it is synthesized by the related family members of glycosyltransferase (GT) family [11].

Lignin, the second abundant component in plant cell walls, mainly plays roles in increasing plant cell wall strength and stem bending resistance [1]. Lignin is a kind of complex phenolic polymer, its monomer synthesis is derived from the phenylpropane pathway and the lignin-specific pathway [12]. Lignin monomer synthase genes mainly include *PAL*, *C4H*, *C3H*, *4CL*, *COMT*, *CCoAMT*, *F5H*, *CAD* and *CCR*, etc, which have been studied in maize [13], *Arabidopsis thaliana* [14], poplar (*Populust remuloides*) [15], ryegrass (*Lolium perenne* L.) [16], switchgrass (*Panicum virgatum* L.) and other plants [17]. The content and composition of lignin can be changed when the expression levels of these genes were up-regulated or down-regulated. By down-regulating the expression of the lignin biosynthetic pathway gene *Pt4CL1* in poplar, the deposition of lignin and cellulose can be regulated in a compensatory fashion, which may contribute to metabolic flexibility and a growth advantages to sustain the long-term structural integrity of woody perennials [1-2]. In *Arabidopsis*, NAC, MYB, and WRKY transcription factors were also found to be involved in the regulation of lignin biosynthesis.

Elephant grass (*Pennisetum purpureum* Schum.) is a perennial Poaceae C4 plant, originated in tropical Africa, and then widely distributed in tropical and subtropical climate regions of Asia, Africa and America. It is one of the highest biomass plants in the world, with plant height up to 3 ~ 5m and annual biomass up to 4500kg/hm^{-2} [18]. The main biological characteristics of elephant grass are high photosynthesis, high yield, and resistance to biotic and abiotic stress. Elephant grass is not only a high-quality forage for livestock and poultry [19], an ideal plant for soil and water conservation, a high-quality paper pulp raw material, the raw material for biofuels preparation, but also an ideal lignocellulosic energy plant [20-22]. The characteristics of elephant grass, especially its stem lignocellulose, are of great significance for its

quality as feed or other industrial raw materials. Improving the cellulose content and reducing the lignin content can promote the feed quality, the conversion and utilization efficiency of lignocellulose.

Elephant grass is the allotetraploid crop (A'A'BB, 2n=4x=28) with a complex genome. The species is primarily cross-pollinated due to its androgynous flowering behavior, resulting in high heterozygosity and broad genetic diversity which can be utilized in breeding programs. Although the elephant grass genome has not been deciphered yet, its genetic research has focused on the evaluation of genetic diversity by constructing molecular markers and fingerprints and determining genetic relationships [23-25]. Meanwhile, researchers have revealed the biosynthetic pathway of anthocyanins and the molecular mechanism of early response to cadmium in elephant grass roots and leaves through RNA-seq analysis [26-27]. Some progress has also been made in improving the agronomic traits of elephant grass by using traditional breeding methods based on phenotypic selection [28-29]. Although the regulatory network of lignocellulose synthesis has been reported in other species such as *Arabidopsis thaliana*, maize, poplar and switchgrass, research in elephant grass is still scarce.

In this study, 12 samples of elephant grass at different stem development stages were used as materials for RNA-seq analysis, lignocellulose composition analysis and cell wall morphology observation. Combined with GO, KEGG, WGCNA and Q-PCR correlation analysis, the study identified some candidate genes and provided valuable genomic data for molecular mechanism of fiber formation in elephant grass, which was helpful to the development and industrial application of high-quality elephant grass varieties with high cellulose and low lignin.

Results

Changes of cellulose, hemicellulose and lignin content in elephant grass stem at different development stages

The content of cellulose, hemicellulose and lignin in different stem segments of T1, T2, T3 phase were measured. It was found that the content of cellulose and hemicellulose increased first and then decreased. For example, the content of cellulose and hemicellulose in T1-S2 was higher than that in T1-S1, but the content of T1-S3 was lower than that of T1-S2, whereas the content of lignin fell gradually. The similar changes also appeared in different stem nodes at T2 and T3 stages. Meanwhile, by analyzing the content change of the same stem node in different development times, it was also found that with the increase of growth time of elephant grass, the content of cellulose and hemicellulose decreased, while the content of lignin increased (Fig. 1b).

By analyzing the ratio of the primary cell wall to the secondary cell wall, it was found that the secondary cell wall increases as the cell development time increases while the rate of the cell wall to the primary cell wall keeps rising, which might be the reason of the gradually increases of lignin content to maintain and support the strength of elephant grass stems (Fig. S1).

Characteristics of the cell wall in different developmental stages of the elephant grass stem

The changes of cell wall morphology at different development stages of elephant grass, especially the primary and secondary cell wall were observed (Fig. 2a). The ratio of secondary cell wall (sw) thickness to primary cell wall (pw) thickness in T1-S1 and T1-S2 was 1.18 and 0.85, the ratio of sw/pw thickness in T2-S1 and T2-S2 was 1.67 and 0.92, and the ratio of sw/pw thickness in T3-S1 and T3-S3 was 2.15 and 1.25, respectively (Fig. 2b). The above data showed that the change trend of sw/pw thickness of T1, T2, and T3 was consistent, that is, with the increase of development time, the development speed of secondary cell wall (sw) was faster than that of primary cell wall (pw). The ratio of sw/pw in S1 node of three different development stages was also analyzed. It was found that with the increase of development time, the thickness ratio of sw/pw in T1-S1 T2-S1, T3-S1 increased subsequently. The change of sw/pw thickness ratio in S2 and S3 stages of three different development stages showed the same trends.

To further understand the morphological changes of cell wall during the development of elephant grass, we selected S1 and S5 stem segments with a longer development time span at T3 stage as samples for micro-CT observation. The stem of elephant grass is composed of epidermis, parenchyma cells and vascular bundles. Vascular bundles, which scattered in parenchyma cells and cannot be thickened, are composed of phloem and xylem without cambium. With the development of stem tissue, S1 vascular bundle were arranged regularly and compactly (Fig. S1). The content of cellulose and hemicellulose in vascular bundle decreased gradually, while the content of lignin increased gradually (Fig. 1b) to meet the need of mechanical support during the maturation of elephant grass stems need in the process of stem maturity of elephant grass.

Differential gene expression during the development of elephant grass stems

36 cDNA libraries were constructed from different stages of elephant grass stems (three biological replicates for each tissue). Totally, 1.32 billion raw reads (396.96 Gb) were obtained, 1.29 billion cleaned reads (388.57Gb) were acquired after filtering with 6.56-21.07 Gb in each sample. The error rate was 0.03% (Q20 and Q30 values were more than 93% and 90%, respectively), which met the requirements of gene discovery (Table S1). De novo assembly generated 77,435 cluster sequences from 12 representative samples with the largest sequencing depth. Finally, we got a non-redundant transcript clusters include 230572 unique genes with the average length of 961.35bp and an N50 of 1435bp, an N90 of 423bp (Fig S2). Transcriptome de novo assembly was carried out with short reads assembling program-Trinity [36]. The Pearson correlation coefficient based on the expression value of each library indicated that there was a high correlation between sample replicates (Fig. S3). Cluster analysis among samples showed that the development time of elephant grass was the main factor affecting the clustering. The DEGs in three developmental stages of elephant grass stems were analyzed, a total of 15611, 10235 and 27389 DEGs were identified in T1, T2, and T3, respectively (Fig. 3). The DEGs of different stem segments at three development stages were also analyzed, it was found that 147 genes were expressed in three segments of T1, 54 in four segments of T2, and 91 in five segments of T3 (Fig. 3). The intersection of all DEGs at three different developmental stages was compared to determine the shared core set. It was found that 3852 genes were differentially expressed at three developmental stages (Fig. 3d).

3852 DEGs were then subjected to enrichment analysis of GO functions and KEGG pathways. 4 of the top 10 enriched GO annotation functions were related to cell composition such as apoplast, cell wall, extracellular region, plant-type cell wall, four were related to molecular functions such as peroxidase activity, xyloglucan, and xyloglucosyl transferase activity, heme-binding, xyloglucan–specific endo–beta–1,4–glucanase activity, and two were related to biological processes such as cell wall macromolecule catabolic process, hydrogen peroxide catabolic process (Fig. 4a) (Table S2). Based on KEGG pathway analysis, all DEGs were enriched to 9 pathways (Table S3), of which the two most significant pathways were phenylpropane metabolism (23 DEGs) and starch and sucrose metabolism (23 DEGs) (Fig. 4b).

Genes highly correlated with the synthesis of cellulose, hemicellulose and lignin by WGCNA analysis

Weighted gene co-expression network analysis (WGCNA) was performed on 3852 DEGs at three stages, and the network was divided into three modules. The analysis of module-trait relationship showed that the 'blue' module was highly correlated with the synthesis of cellulose (r=0.67, $P=6.0\times10^{-6}$) and hemicellulose (r=0.51, P=0.001), whereas the 'turquoise' module was related to lignin synthesis (r=0.68, $P=5.0\times10^{-6}$) (Fig. S4).

The 'blue' module was filtered according to Module membership > 0.9, the absolute value of the correlation coefficient between the 'turquoise' module and lignin was greater than 0.75, and 20 and 23 genes remained in the 'blue' and 'turquoise' module respectively. The WGCNA gene significance (GS) (i.e., related to traits) showed that the genes with highest GS in 'blue' and 'turquoise' modules were Cluster-55067.0 (0.659) and Cluster-17353.3 (0.798), respectively. Six of 20 genes in 'blue' module were known to be functional, such as *GTL1* transcription factor, O-methyl transferase (*OMT*), expansin-likeA2, alpha-humulene synthase, probable galactinol sucrose, *GhGalT1*. Meanwhile, 14 genes with unknown functions were also covered by 'blue' module, which needs further study. The GO function annotations of these 20 genes in 'blue' module included the genes which were related to extracellular region, C-4 methyl sterol oxidase activity, terpene biosynthesis, etc. The expression levels of these genes were shown in Fig. 5.

Among 23 genes in 'turquoise' module, there were eight coding genes such as *GTL1*, *MYB2* transcription factors, threonine-protein kinase ERECTA, probable methionine-tRNA ligase, alcohol holding hydrogenase-like2, zinc finger protein GIS3, protein slr0074, proline-rich receptor-like protein kinase PERK8. The GO function annotations of these 23 genes in 'turquoise' module included genes that were related to cell growth regulation, secondary cell wall formation regulation, cell wall composition regulation, etc. 15 genes remained with unknown functions (Table S4).

Lignin and cellulose synthesis pathway during stem development of elephant grass

Cellulose synthase (CesA) is the most important enzyme in the cellulose synthesis pathway. It can directly utilize UDPG produced by starch and sucrose metabolism to synthesize cellulose. In elephant grass, 27 *CesA* genes were identified. At T1 stage, the expression level of *CesA1* - *CesA6* was higher in the whole stem, but decreased in T2 and T3 stages, while the expression level of *CesA7* - *CesA27* in tender

stems was much higher than that in mature stems at T1, T2, and T3 stages. This indicated that *CesA* gene mainly synthesizes cellulose in the tender stem tissue. When the cellulose accumulated to a certain level, its expression level gradually decreased (Fig. 6a).

Lignin synthesis is one of the most important pathways of phenylpropane metabolism. In the lignin metabolism pathway, *CAD* (21 unigenes), *4CL9* (21 unigenes), *C4H* (8 unigenes), *PAL* (16 unigenes), *CCR* (25 unigenes), *F5H* (6 unigenes), *CCoAOMT* (4 unigenes) were identified to be related. Expression analysis found that most members of these gene families showed higher expression levels in the mature stem. In contrast, a few gene members were continuously expressed in the whole tissue (Fig. 6b).

The results of qRT-PCR demonstrated that the expression trends of these genes were consistent with that of RNA-seq data (Fig. S5, Table S6). Overall, as the stem development, the expression level of cellulose synthesis genes and lignin synthesis related genes showed the opposite trend. On the other hand, the changes of stem cellulose, hemicellulose and lignin content, as well as the changes of primary stem wall and secondary cell wall thickness, were positively correlated with the expression of these two types of genes.

Discussion

Elephant grass is one of the highest biomass forage grasses on the earth, which is widely used in feed and bioenergy related industries. It is of great significance to further understand the genes and important metabolic pathways related to lignocelluloses synthesis for molecular breeding the new elephant grass varieties and make them more suitable for industrial application.

In this study, 3852 differentially expressed genes (DEGs) were identified by transcriptome analysis on 12 stem samples at different development stages of elephant grass. The GO enrichment analysis of DEGs found that the top 10 enriched GO annotation functions included three functions that were directly related to cell wall development such as cell wall, plant-type cell wall, cell wall macromolecule catabolic process. KEGG pathway analysis discovered that sucrose and starch metabolism and phenylpropane metabolism were the most significant metabolic pathways of differential gene enrichment. Lignin synthesis is one of the important branches in phenylpropane metabolism [43], the substrate for cellulose synthesis can be provided by starch and sucrose metabolic pathways [44-45], which indicated that these DEGs might affect the synthesis of cellulose and lignin.

In the process of cellulose synthesis, sucrose is the starting substrate, and UDPG produced by its decomposition can directly synthesize the dextran chain under the action of *CesA* gene [46]. 27 *CesA* genes were identified in elephant grass, which were more abundant than other species such as wheat (14 species) [4], *Arabidopsis thaliana* (10 species) [3], maize (10 species) [5], poplar (16 species) [8]. *CesAs* were mainly expressed in tender stem tissue (Fig.6a). As the stem development, the cellulose content accumulated gradually (Fig.1). The high copy of *CesAs* in elephant grass makes it has a high cellulose synthesis potential. The cellulose content in mature stem of elephant grass is significantly higher than that of corn, wheat, reed and other plants [47].

Totally, 101 lignin-synthesizing related genes such as *PAL*, *C4H*, *4CL*, *CCoAMT*, *F5H*, *CAD*, and *CCR* were identified in elephant grass. PAL is the critical enzyme in phenylpropane metabolism pathway, its inhibition will reduce lignin content and affect plant growth and development. 4CL is the key rate-limiting enzyme for the production of G or S-lignin monomers. Different expressions of 4CL can regulate the content of three kinds of lignin monomers, promoting or inhibiting the expression of *4CL* gene can significantly regulate the relative proportion of lignin/cellulose [48]. In elephant grass, 16 *PAL* genes and 21 *4CL* genes were mainly expressed in mature stem tissues which led to the lignin content in the mature stem of elephant grass was higher. It was also found that inhibition or overexpression of these genes in tobacco would affect the lignin content [49-50].

To further understand the relationship between these DEGs and lignocellulose components, WGCNA was conducted and found that there were 20 genes were related to the synthesis of cellulose and hemicellulose, 23 genes were associated with the synthesis of lignin. Among these 43 genes, 14 of which have been identified with precise functions by GO annotation, some of them such as *OMT* and *GalT1* were directly related to the synthesis of lignocelluloses, some of them played important roles in growth and development and cell wall formation. There were 28 genes with unknown functions, which need further functional verification. In tobacco, it was reported that the sense or antisense expression of sequences encoding O-methyltransferase (*OMT*) could regulate enzyme activity of lignin synthesis [51]. The fiber length of transgenic cotton overexpressing *GhGalT1* was shorter than that of wild type, while in *GhGalT1* silenced line, the fiber length was significantly increased than that of wild type [52].

Generally, our work indicated the dynamic changes in cell wall composition and morphology during the stem development of elephant grass were consistent with the changes and expression of cellulose and lignin related genes. These data provided the new and extensive list of candidate genes for more specialized functional studies in the future, and an essential theoretical basis for the genetic improvements of elephant grass lignocellulose synthesis as well.

Conclusion

RNA-seq, lignocellulose content and cell wall morphology of elephant grass stem were conducted and analyzed in this study, which provided a basis for further revealing the mechanism of lignocellulose synthesis and accumulation of elephant grass. A total of 3852 common DEGs were identified. KEGG analysis showed that the two most abundant metabolic pathways were phenylpropane metabolism (23 DEGs), starch and sucrose metabolism (23 DEGs), among which phenylpropane metabolism functioned as an important pathway for lignin synthesis, while the latter produced UDPG for cellulose synthesis. 27 *CesA* genes for cellulose synthesis and 101 related genes for lignin synthesis were identified, respectively. *CesA* genes had higher expression levels in young stems while the lignin-related genes had higher expression levels in mature stems. In addition, a total of 43 candidate genes were screened by WGCNA, of which 17 had function annotations in other species. Among them, the *GTL1* transcription factor and O-Methyl transferase (*OMT*) gene have been proved to regulate the synthesis of lignocellulose in other plant species.

Methods

Experimental materials

Elephant grass was cultured in the greenhouse of Qilu University of Technology, Jinan City, Shandong Province. Stalks at seedling stage 40 days (T1 period), 80 days (T2 period), 120 days (T3 period) were sampled. The lowest node of the stem was taken and labeled as S1, samples from every other stem node were labeled as S2, S3, S4 and S5 respectively. Three samples (stem nodes) were taken at T1 period (T1-S1, T1-S2, T1-S3), four samples at T2 period (T2-S1, T2-S2, T2-S3, T2-S4) and five samples at T3 period (T3-S1, T3-S2, T3-S3, T3-S4, T3-S5) (Fig.1a). A total of 36 samples were collected, each including three biological replicates. All samples were immediately frozen in liquid nitrogen and stored at -80 °C before total RNA extraction.

Determination of the content of cellulose, hemicellulose and lignin

The sample to be tested was naturally air-dried or placed in an oven to dry (temperature not exceeding 50°C) until the moisture is less than 10%, crushed and screened by the grinder. Took a portion between 20-80 mesh, cooled, and stored in a sealed bag for further analysis. Content analysis was conducted by high-performance liquid chromatographic (HPLC) according to NERL method [30-31]. The reducing sugar was analyzed by HPLC (Shimadzu, Kyoto, Japan) with a Shimadzu LC-10AD detector. The HPLC was performed in a Bio-Rad HPX-87H column with 10 uL injected volume at 60°C with 5 mM H_2SO_4 as eluent at a flow rate of 0.4 mL/min [32]. The content of cellulose, hemicellulose and lignin are the percentage of the dry weight of the sample.

Cell wall morphology observation

The periods with the highest and lowest cellulose content at three different development stages of elephant grass stems were selected as samples, then cut the fresh stalks into 1cm×1cm pieces and placed in 2.5% glutaraldehyde phosphate buffer (0.1mol/L, pH=7.0). Samples were treated as reported [33] and observed under EM-420 transmission electron microscope (Philips Electronics, Holland).

Micro CT observation of elephant grass stem

The stalks of elephant grass were cut into 1cm×1cm pieces and put them into 2.5% glutaraldehyde phosphate buffer (0.1mol/L, pH=7.0), fixed in a 4°C storage cabinet for 3h, washed the fixed tissue twice with 1mol/L phosphate buffer, and placed the sample in a 40°C oven for 24h. Cross-section of the processed sample were observed under SkyScan 2211 micro-CT (Bruker, Belgium).

RNA extraction and transcriptome sequencing

Total RNA was extracted from stems using HiPure Plant RNA Kit (Magen, Guangzhou, China) according to the manufacturer’s instructions. A total of 3μg of RNA per sample was used for library preparation with

insert sizes of 350bp and sequenced on Novaseq 6000 (Illumina, USA). RNA-seq analysis was conducted using the Illumina platform according to the standard protocols [34].

Quality control of RNA-seq and transcriptome assembly

Raw data (raw reads) of fastq format were firstly processed through Trim Galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) [35]. At this step, clean data (clean reads) were obtained by removing reads containing adapter, ploy-N and low quality reads from raw data. At the same time, Q20, Q30, GC-content and sequence duplication level of the clean data were calculated. All the downstream analyses were based on clean data with high quality. 12 samples with the largest amount of sequencing were selected, assembled them with Trinity-v2.9.1 software [36], and then used cd-hit-estv4.8.1 software for cluster analysis to obtain representative sequences [37]. Salmonv1.1.0 software was employed to map the sequencing data back. On the representative sequence [38], the corset software was adopted to obtain all the unigenes [39].

Gene function annotation

Gene function was annotated based on the following databases: Nr (NCBI non-redundant protein sequences) (https://www.ncbi.nlm.nih.gov/); Swiss-Prot (A manually annotated and reviewed protein sequence database) (http://www.gpmaw.com/html/swiss-prot.html); KO (KEGG Ortholog database) (https://www.kegg.jp/); GO (Gene Ontology) (http://geneontology.org/).

Differential expression and enrichment analysis

Differential expression analysis was performed using the R package DESeq [28]. The p-values were adjusted by the Benjamini-Hochberg (BH) method. Genes with an adjusted p-value < 0.05 were assigned as DEGs. GO enrichment and KEGG pathway enrichment analysis of 3852 differentially DEGs, which were identified in three periods of T1, T2, and T3 was implemented by R Package Goseq [40]. The population set is a set with all the annotated genes, and the study set consisted of the DEGs in the population set. The p-values were adjusted as differential expression analysis did. Adjusted p-value < 0.05 for GO enrichment analysis and adjusted p-value < 0.1 for KEGG pathway enrichment analysis were considered significant.

Weighted gene co-expression network analysis

Gene expression patterns for common differentially expressed genes in T1, T2, T3 period were used to construct a co-expression network by WGCNA. The genes that were not detected to be expressed in all tissues were removed before analysis. Soft thresholds were set based on the scale-free topology criterion [41-42].

qRT-PCR analysis

DNase-treated RNA (2 μg) was reverse transcribed using High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, USA). Gene-specific primers were designed using Primer Express (v3.0, Applied Biosystems). Quantitative reverse transcription PCR (qRT-PCR) assays were performed using SYBR Green I Master Mix (Roche, Indianapolis, USA). Three biological and three technical replicates for each reaction were analyzed on LightCycler 480 instrument (Roche, USA) with the first step of 95 °C for 5 min followed by 40 cycles of 95 °C for 10 s, 60 °C for 10 s, and 72 °C for 20 s. Melting curves were generated using the following program: 95 °C for 15 s, 60 °C for 15 s, and 95 °C for 15 s. 18S rRNA was used as an internal control. Data analysis was calculated by $2^{-\Delta\Delta CT}$ method. Significant differences between different samples were tested with IBM SPSS Statistics 19.0 software.

Declarations

Ethics approval and consent to participate

There is no ethics approval and consent to participate in this manuscript.

Consent to publication

Not applicable.

Availability of data and materials

The elephant grass RNA sequences have been deposited into BIG Sub GSA database, the GSA accession number is CRA003077. The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no conflict of interests.

Funding

This study was financially supported by the Integration of Science and Education Program Foundation for the Talents by Qilu University of Technology (No. 2018-81110268), Foundation of State Key Laboratory of Biobased Material and Green Papermaking (No. 2419010205 and No. 23190444). The funding bodies provided the financial support to this research, including experimental design and implementation, sampling and data analysis. No funder played the role in data collection and analysis and writing the manuscript.

Authors' contributions

WQZ SKZ performed the experiments, data analysis and manuscript writing. XQL, CL, XWL, GYD participated in some experiments and data analysis. TX designed the project and contributed to the writing of the manuscript and approved the final manuscript.

Acknowledgments

Thanks Shenzhen Aimigene Institution (Shenzhen,China) for providing consultation on bioinformatics methods.

Abbreviations

RNA-seq: RNA sequencing

DEGs: Differentially expressed genes

KEGG: Kyoto Encyclopedia of Genes and Genomes

WGCNA: Weighted gene co-expression network analysis

CesA: Cellulose synthase

CSL: Cellulose synthase-like

GT: Glycosyltransferase

sw: secondary cell wall

pw: primary cell wall

GS: Gene significance

OMT: O-methyl transferase

GO: Gene Ontology

KO: KEGG Ortholog database

BH: Benjamini-Hochberg

UDPG: Uridine diphosphoglucose

PAL: Phenylalanine Ammonia-Lyase

C4H: Cinnamic Acid4-Hydroxylase

CCR: Cinnamoyl-CoA Reductase

CAD: Cinnamyl Alcohol Dehydrogenase

F5H: FerulicAcid5-Hydroxylase

CoAOMT: Caffeoyl-CoAO-methytransferase

COMT: 5-Hydroxyferulic Acid O-Methyltransferase/Bispecific Caffeic Acid

HPLC: High Performance Liquid Chromatography

References

1. Hu WJ, Harding SA, Lung J. Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. Nature Biotechnology. 17 (8) (1999) 808-812; doi:1038/11758
2. Zhang CB, Chen LH, Jiang J. Why fine tree roots are stronger than thicker roots: The role of cellulose and lignin in relation to slope stability. Geomorphology. 206 (FEB.1) (2014) 196-202; doi:1016/j.geomorph.2013.09.024
3. Samuga, Joshi CP. A new cellulose synthase gene (*PtrCesA2*) from aspen xylem is orthologous to Arabidopsis *AtCesA7* (irx3) gene associated with secondary cell wall synthesis. Gene. 296 (1-2) (2020) 37-44; doi:1016/s0378-1119(02)00864-8
4. Kaur S, Dhugga KS, Gill K. Novel Structural and Functional Motifs in cellulose synthase (*CesA*) Genes of Bread Wheat (*Triticum aestivum*, L.). Plos One. 11 (2016) 1-18; doi: 1371/journal.pone.0147046
5. Appenzeller L, Doblin M, Barreiro R. Cellulose synthesis in maize: isolation and expression analysis of the cellulose synthase (*CesA*) gene family. Cellulose. 11 (3-4) (2004) 287-299; doi:1023/b:cell.0000046417.84715.27
6. Taylor NG, Laurie S, Turner SR. Multiple Cellulose Synthase Catalytic Subunits Are Required for Cellulose Synthesis in Arabidopsis. The Plant Cell. 12 (12) (2001) 2529; doi:10.2307/3871246
7. Taylor NG, Scheible WR, Cutler S. The *irregular xylem3* locus of Arabidopsis encodes a cellulose synthase required for secondary cell wall synthesis. The Plant Cell. 11 (5) (1999) 769-780; doi: 2307/3870813
8. Bhandari S, Fujino T, Thammanagowda S. Xylem-specific and tension stress-responsive coexpression of *KORRIGAN* endoglucanase and three secondary wall-associated cellulose synthase genes in aspen trees. Planta. 224 (4) (2006) 828-837; doi: 2307/23389483
9. Yang C, Li D, Liu X. OsMYB103L, an R2R3-MYB transcription factor, influences leaf rolling and mechanical strength in rice (*Oryza sativa*). BMC Plant Biology. 14 (1) (2014) 158; doi: 10.1186/1471-2229-14-158
10. Roberts JA, Evan D, Mcmanus MT. Annual Plant Reviews || Glycosyltransferases of the GT47 Family. 41 (2018) 265-283; doi: 1002/9781119312994.apr0439
11. Zeng W, Lampugnani ER, Picard KL. Asparagus *IRX9*, *IRX10*, and *IRX14A* are components of an active xylan backbone synthase complex that forms in the Golgi apparatus. Plant Physiology. 171 (1) (2016) 93-109; doi:1104/pp.15.01919
12. Bonawitz ND, Chapple C. The Genetics of Lignin Biosynthesis: Connecting Genotype to Phenotype. Annual Review of Genetics. 44 (1) (2010) 337-363; doi:1146/annurev-genet-102209-163508

13. Silvia F, Montserrat C. Altered Lignin Biosynthesis Improves Cellulosic Bioethanol Production in Transgenic Maize Plants Down-Regulated for Cinnamyl Alcohol Dehydrogenase. Molecular Plant. 5 (4) (2012) 817-830; doi:1093/mp/ssr097
14. Acker R, Vanholme R, Véronique S. Lignin biosynthesis perturbations affect secondary cell wall composition and saccharification yield in *Arabidopsis thaliana*. Biotechnology for Biofuels. 6 (1) (2013) 46-47; doi: 1186/1754-6834-6-46
15. Pilate G, Guiney E, Holt K. Field and pulping performances of transgenic trees with altered lignification. Nature Biotechnology. 20 (6) (2020) 607-612; doi: 1038/nbt0602-607
16. Tu Y, Rochfort S, Liu Z. Functional Analyses of *Caffeic Acid O-Methyltransferase* and *Cinnamoyl-CoA-Reductase* Genes from Perennial Ryegrass (*Lolium perenne*). Plant Cell. 22 (10) (2020) 3357-3373; doi: 1105/tpc.109.072827
17. Wang Z, Li R, Xu J. Sodium hydroxide pretreatment of genetically modified switchgrass for improved enzymatic release of sugars. Bioresource Technology. 110 (2012) 364-370; doi: 1016/j.biortech.2012.01.097
18. José C, Prinsen P, Rencoret J. Structural characterization of the lignin in the cortex and pith of elephant grass (*Pennisetum purpureum*) stems. Journal of Agricultural and Food Chemistry. 60 (14) (2012) 3619-3634; doi: 1021/jf300099g
19. Nyambati E, Nyambati M, Sollenberger L. Feed intake and lactation performance of dairy cows offered napiergrass supplemented with legume hay. Livestock Production Science. 83 (3-2) (2003) 179-189; doi: 1016/s0301-6226(03)00094-0
20. Strezov V, Evans TJ, Hayman C. Thermal conversion of elephant grass (*Pennisetum Purpureum*) to bio-gas, bio-oil and charcoal. Bioresour Technol. 99 (17) (2008) 8394-8399; doi:10.1016/j.biortech.2008.02.039
21. Liu X, Shen Y, Lou L. Copper tolerance of the biomass crops Elephant grass (*Pennisetum purpureum* Schumach), Vetiver grass (*Vetiveria zizanioides*) and the upland reed (*Phragmites australis*) in soil culture. Biotechnology Advances. 27 (5) (2009) 633-640; doi: 1016/j.biotechadv.2009.04.017
22. Somerville, Youngs H, Taylor C. Feedstocks for Lignocellulosic Biofuels. Science. 329 (5993) (2010) 790-792; doi: 1126/science.1189268
23. Kawube G, Alicai T, Wanjala B. Genetic Diversity in Napier Grass (*Pennisetum purpureum*) Assessed by SSR Markers. Journal of Agricultural Science. 7 (7) (2015) 147-155; doi: 5539/jas.v7n7p147
24. Bhandari P, Sukanya DH, Ramesh CR. Application of Isozyme Data in Fingerprinting Napier Grass (*Pennisetum purpureum*) for Germplasm Management. Genetic Resources and Crop Evolution. 53 (2) (2006) 253-264; doi: 10.1007/s10722-004-6120-2
25. Harris K, Anderson W, Malik R. Genetic relationships among napiergrass (*Pennisetum purpureum*) nursery accessions using AFLP markers. Plant Genetic Resources. 8 (1) (2010) 63-70; doi: 10.1017/S1479262109990165
26. Zhou S, Chen J, Lai J, Yin G, Chen P. Integrative analysis of metabolome and transcriptome reveals anthocyanins biosynthesis regulation in grass species *Pennisetum purpureu* Industrial Crops &

Products. 138 (2019) 111470; doi: 10.1016/j.indcrop.2019.111470

27. Zhao J, Xia B, Meng Y, Yang Z, Pan L. Transcriptome Analysis to Shed Light on the Molecular Mechanisms of Early Responses to Cadmium in Roots and Leaves of King Grass (*Pennisetum americanum × purpureum*). International Journal of Molecular Sciences. 20 (10) (2019) 1-18; doi: 10.3390/ijms20102532
28. Bhandari P, Sukanya DH, Ramesh CR. Application of Isozyme Data in Fingerprinting Napier Grass (*Pennisetum purpureum* Schum.) for Germplasm Management. Genetic Resources and Crop Evolution. 53 (2) (2006) 253-264; doi: 1007/s10722-004-6120-2
29. Jakob K, Zhou F, Paterson AH. Genetic improvement of C4 grasses as cellulosic biofuel feedstocks. Vitro Cellular & Developmental Biology Plant. 45 (3) (2009) 291-305; doi: 1007/978-1-4419-7145-6_7
30. Lu X, Zheng X, Li X. Adsorption and mechanism of cellulase enzymes onto lignin isolated from corn stover pretreated with liquid hot water. Biotechnology for Biofuels. 9 (1) (2016) 1-12; doi: 1186/s13068-016-0531-0
31. Du J, Cao Y, Liu G. Identifying and overcoming the effect of mass transfer limitation on decreased yield in enzymatic hydrolysis of lignocellulose at high solid concentrations. Bioresource Technology. 229 (2017) 88-95; doi: 11016/j.biortech.2017.01.011
32. Sluiter A, Hanes B, Ruiz R, Scarlata C, Sluiter J. Determination of structural carbohydrates and lignin in biomass. National Renewable Energy Laboratory (NREL) Laboratory Analytical Procedures (LAP) for standard biomass analysis. 2007.
33. Liu Y, Muhammad R, Yan L, Zeng Y, Jiang C. Boron and calcium deficiency disturbing the growth of trifoliate rootstock seedlings (Poncirus trifoliate L.) by changing root architecture and cell wall. Plant Physiology and Biochemistry. 144 (2019) 345-354; doi: 1016/j.plaphy.2019.10.007
34. Jiang C. Efficient extraction of RNA from various Camellia species rich in secondary metabolites for deep transcriptome sequencing and gene expression analysis. African Journal of Biotechnology. 144 (2019) 345-354; doi: 10.5897/AJB11.235
35. Kanno M, Kijima A. Quantitative and Qualitative Evaluation on the Color Variation of the Japanese Sea Cucumber Stichopus japonicus. Aquaculture Science. 50 (1) (2002) 63-69; doi: 10.11233/aquaculturesci1953.50.63
36. Grabher R, Manfred G. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology. 29 (7) (2011) 644-652; doi: 10.1038/nbt.1883
37. Limin F, Beifang N, Zhu Z. CD-HIT: accelerated for clustering the next generation sequencing data. Bioinformatics. 28 (23) (2012) 3150-3152; doi: 10.1093/bioinformatics/bts565.
38. Patro R. Salmon provides fast and bias-aware quantification of transcript expression. Nature methods. 14 (4) (2017) 417-419; doi: 10.1038/nmeth.4197
39. Davidson NM, Oshlack A. Corset: Enabling differential gene expression analysis for de novoassembled transcriptomes. Genome Biology. 15 (7) (2014) 410-432; doi: 10.1186/s13059-014-0410-6

40. Young MD, Wakefield MJ, Smyth GK. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biology. 11 (2) (2010) 1-12; doi: 1186/gb-2010-11-2-r14
41. Zhang B, Horvath S. A general framework for weighted gene coexpression network analysis. Statal Applications in Genetics and Molecular Biology. 4 (1) (2005) Article17; doi: 2202/1544-6115.1128
42. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 9 (559) (2008) 1-13; doi: 10.1186/1471-2105-9-559
43. Douglas CJ. Phenylpropanoid metabolism and lignin biosynthesis: From weeds to trees. Trends in Plant Science. 1 (6) (1996) 171-178; doi: 1016/1360-1385(96)10019-4
44. Kleczkowski LA. Glucose activation and metabolism through UDP-glucose pyrophosphorylase in plants. Phytochemistry. 37 (6) (1994) 1507-1515; doi: 1016/s0031-9422(00)89568-0
45. Martin LK, Haigler CH. Cool temperature hinders flux from glucose to sucrose during cellulose synthesis in secondary wall stage cotton fibers. Cellulose. 11 (3-4) (2004) 339-349; doi: 1023/b:cell.0000046420.10403.15
46. John W, Downton S, Hawker JS. Enzymes of starch and sucrose metabolism in *Zea mays* Phytochemistry. 12 (7) (1973) 1551-1556; doi: 10.1016/0031-9422(73)80366-8
47. Lu X, Li C, Wang X, Zhang W, Xia T. Enzymatic sugar production from elephant grass and reed straw through pretreatments and hydrolysis with addition of thioredoxin-His-S. Biotechnology for Biofuels. 12 (297) (2019) 1-11; doi:1186/s13068-019-1629-y
48. Xu B, Luis L, Escamilla T, Sathitsuksanoh N. Silencing of 4-coumarate: coenzyme A ligase in switchgrass leads to reduced lignin content and improved fermentable sugar yields for biofuel production. New Phytologist. 192 (3) (2011) 611-625; doi: 1111/j.1469-8137.2011.03830.x
49. Dixon R, Sewalt V, Howles P. Genetic manipulation of the phenylpropanoid pathway in transgenic tobacco: new fundamental insights and prospects for crop improvement. Biotechnology & Biotechnological Equipment. 1996 (4) (1996) 45-51; doi: 1016/S0734-9750(96)00033-X
50. Paul H, Sameer A, Masoud JW. Overexpression of L-phenylalanine ammonia-lyase and cinnamate 4-hydroxylase in tobacco cell suspension cultures. Plant Biotechnology and In Vitro Biology in the 21st Century. 20 (1999) 297-301; doi: 10.1007/978-94-011-4661-6_69
51. Boerjan W, Ralph J, Baucher M. Lignin Biosynthesis. Annual Review of Plant Biology. 54 (8) (1999) 519-546.
52. Qin L, Qin X, Chen Y, Zeng W. The cotton β-galactosyltransferase 1 (*GalT1*) that galactosylates arabinogalactan-proteins participates in controlling fiber development. Plant Journal. 89 (2016) 957-971; doi: 1111/tpj.13434

Figures

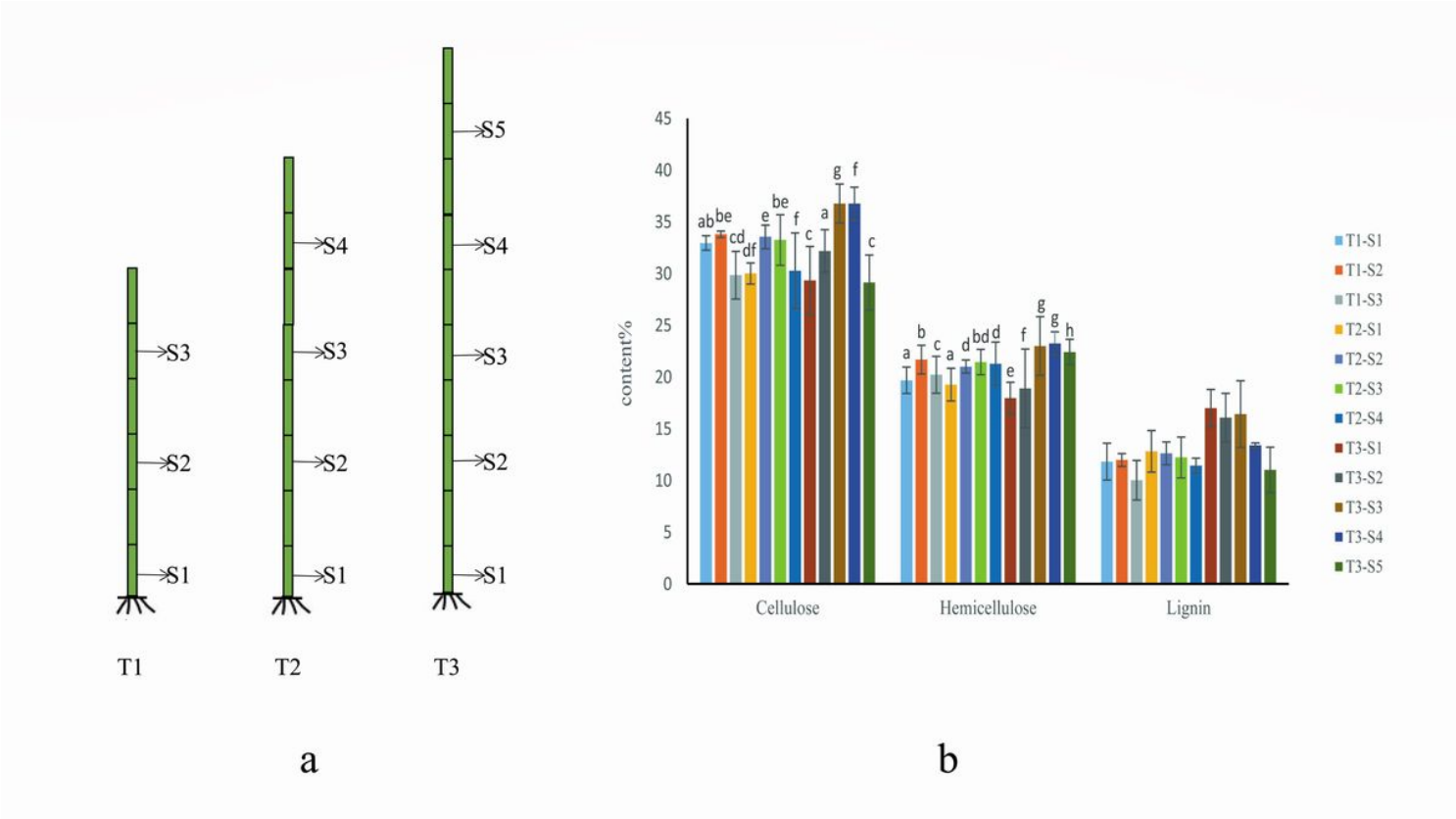

Figure 1

Contents of cellulose, hemicellulose and lignin at different development stages of elephant grass stems. Different letters indicate statistically significant differences (ANOVA, Duncan < 0.05). (a) Sampling period and position. (b) Cellulose, hemicellulose and lignin contents

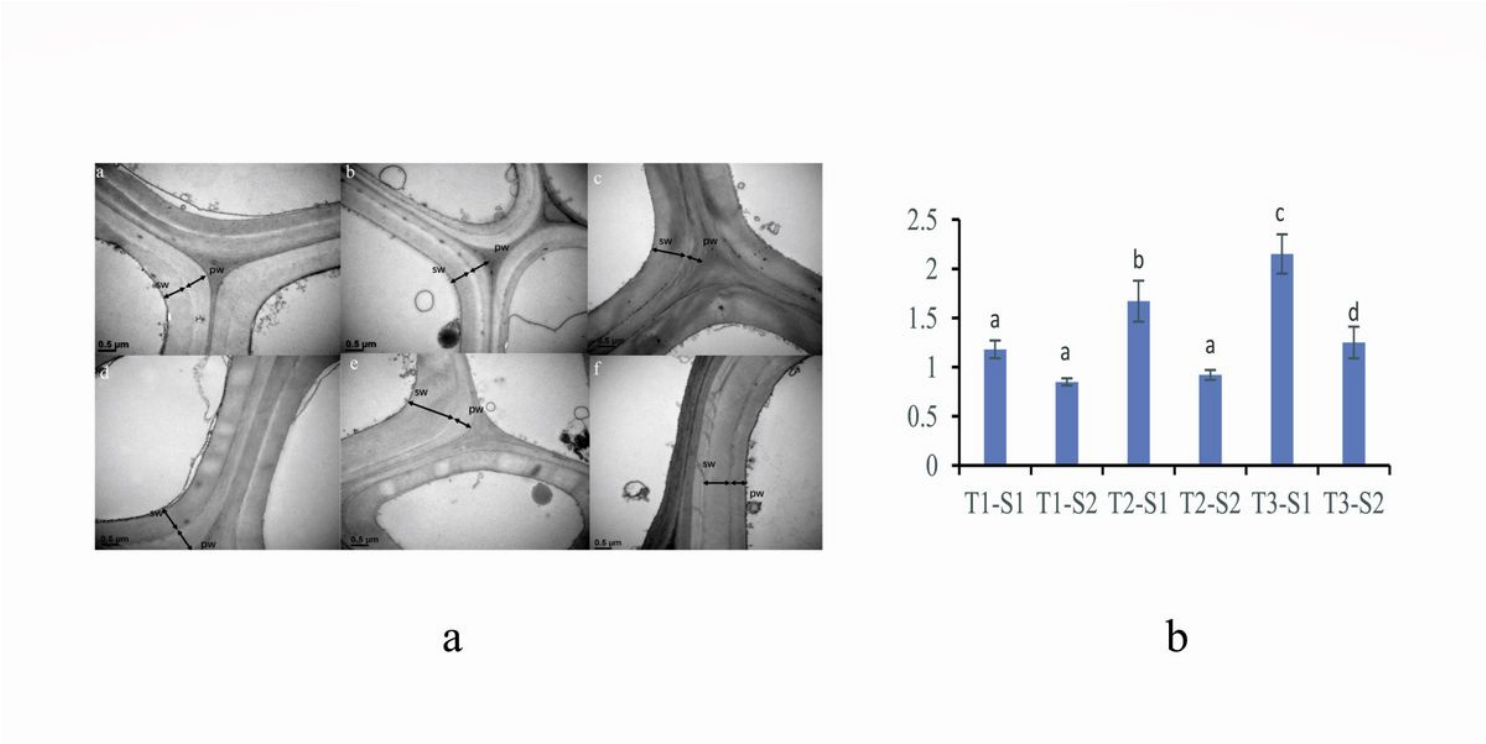

Figure 2

The cell wall morphology and thickness changes at different development stages of elephant grass stems. Different letters indicate statistically significant differences (ANOVA, Duncan < 0.05). (a) a.T1-S1, b. T1-S2, c. T2-S1, d. T2-S2, e. T3-S1, f. T3-S3. SW: Secondary cell wall, PW: Primary cell wall. (b) thickness ratio of secondary cell wall to primary cell wall.

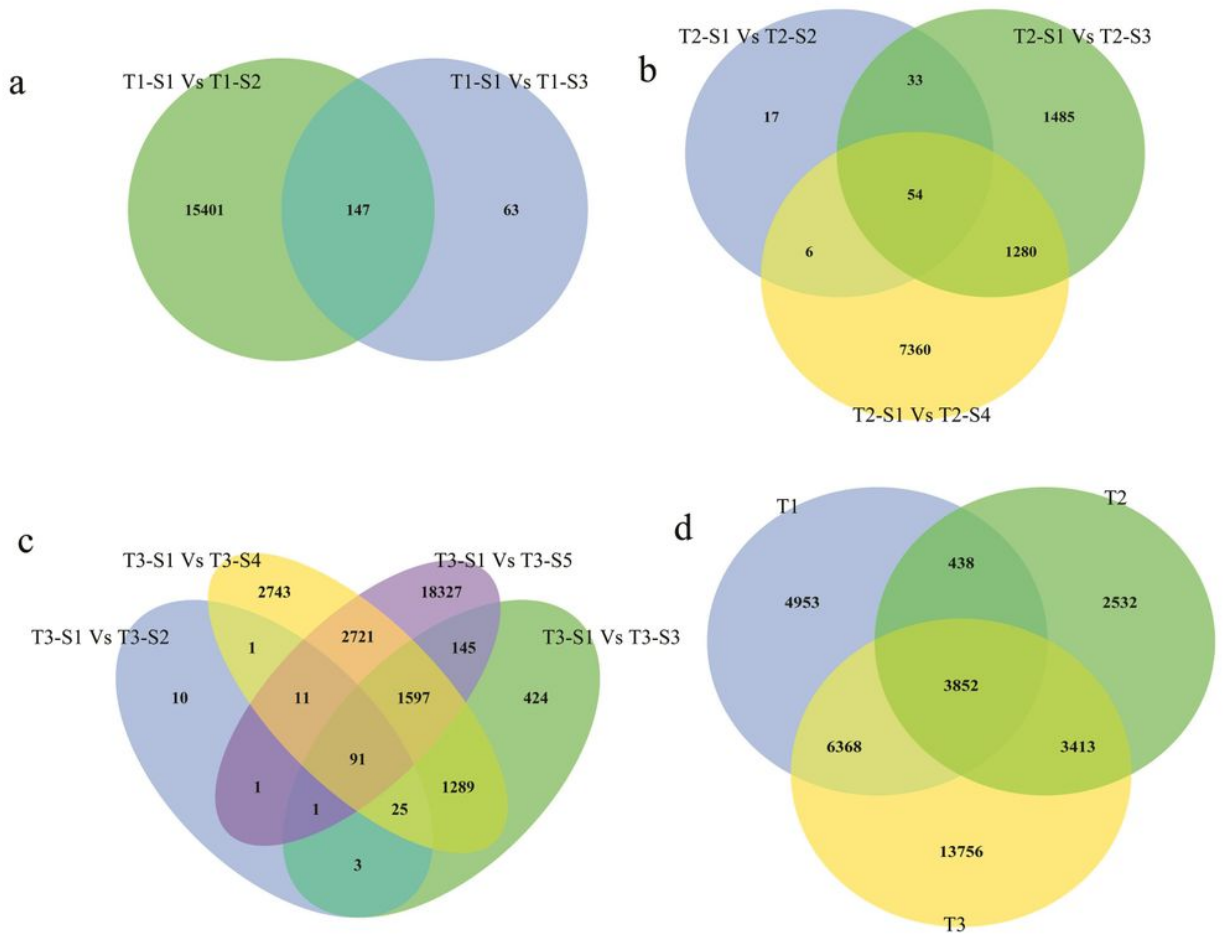

Figure 3

The differentially expressed genes (DEGs) identified by RNA sequence analysis in stem tissues of development stages of elephant grass. (a) The number of DEG between three adjacent stages (T1-S1, T1-S2, T1-S3). green: T1-S1 Vs T1-S2, blue: T1-S1 Vs T1-S3. (b) The number of DEG between the four adjacent stages (T2-S1, T2-S2, T2-S3, T2-S4). green: T2-S1 Vs T2-S2, blue: T2-S1 Vs T2-S3, yellow: T2-S1 Vs T2-S4. (c) The number of DEG between the five adjacent stages (T3-S1, T3-S2, T3-S3, T3-S4). Purple: T3-S1 Vs T3-S5, green: T3-S1 Vs T3-S2, blue: T3-S1 Vs T3-S3, yellow: T3-S1 Vs T3-S4. (d) DEGs between T1, T2 and T3 stages. blue: T1 (DEGs), green: T2 (DEGs), yellow: T3 (DEGs).

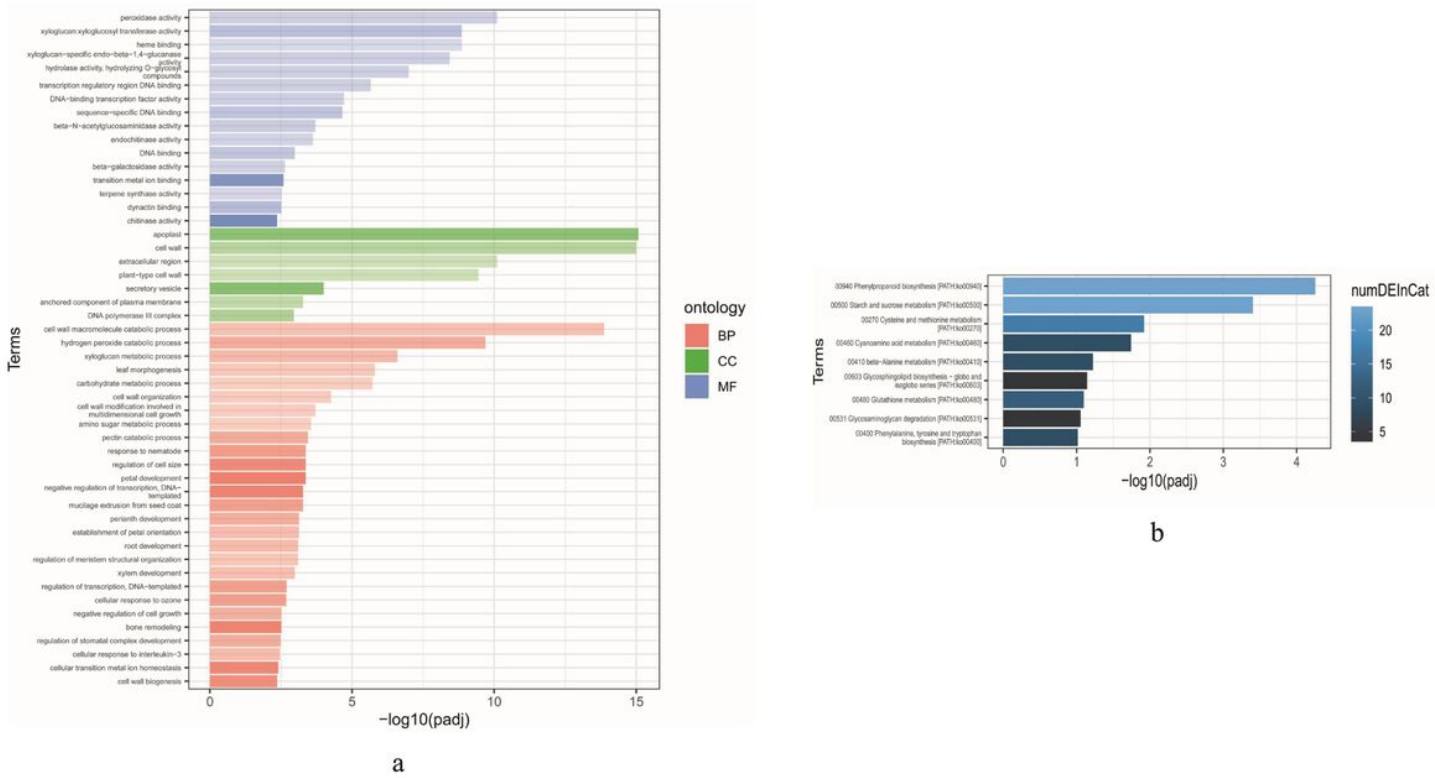

Figure 4

The 3852 DEGs co-expressed at three developmental stages of T1, T2, and T3 were enriched by GO and KEGG. (a) GO enrichment analysis. (b) KEGG enrichment analysis.

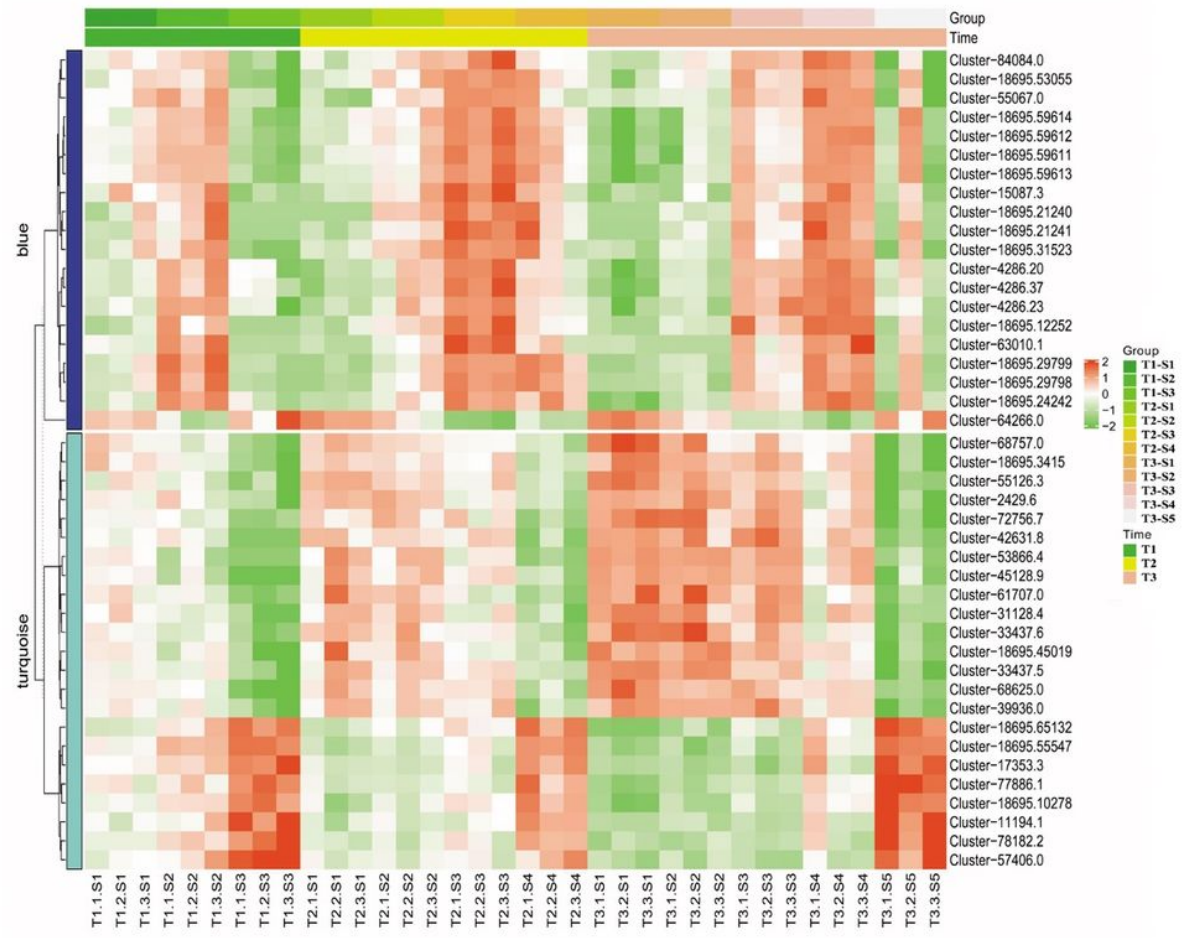

Figure 5

The heat maps of the expression pattern of 43 genes in WGCNA analysis at different development stages of elephant grass stems.

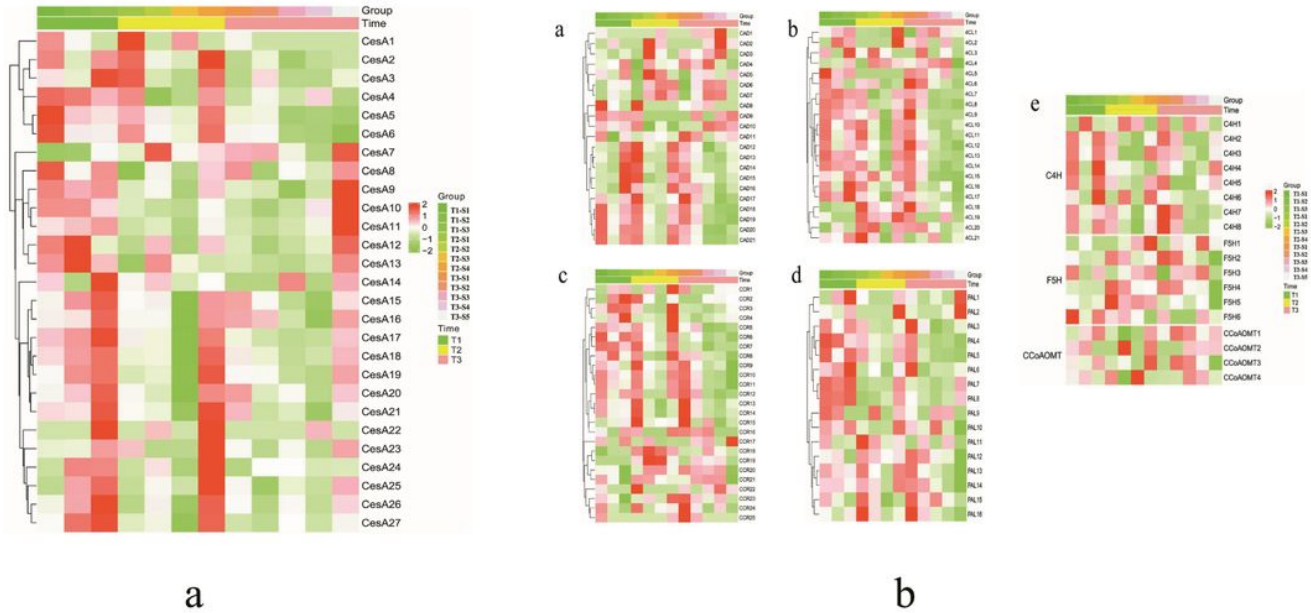

Figure 6

The heat maps of the expression pattern of cellulose synthase genes andlignin synthesis-related genes at different developmental stages of elephant grass stems. (a) Heat map of the expression pattern of 27 cellulose synthase genes. (b) Heat map of the expression pattern of some lignin synthesis-related genes. The grid with four different colors shows the absolute expression of genes, which are represented by different scale levels 2, 1, 0, -1, and -2, respectively. PAL:PhenylalanineAmmonia-Lyase;C4H: CinnamicAcid4-Hydroxylase;CCR:Cinnamoyl-CoAReductase;CAD:CinnamylAlcoholDehydrogenase;F5H:FerulicAcid5-Hydroxylase;CoAOMT:Caffeoyl-CoAO-methytransferase; COMT:5-HydroxyferulicAcidO-MethyltransferaseBispecificCaffeicAcid.

Supplementary Files

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

- Additionalfile1.tif
- Additionalfile2.tif
- Additionalfile3.tif
- Additionalfile4.tif
- Additionalfile6.xlsx
- Additionalfile8.xlsx

- Additionalfile7.xlsx
- Additionalfile11.xlsx
- Additionalfile9.xlsx
- Additionalfile10.rtf
- Additionalfile5.tif