Comparative Transcriptome and WGCNA Reveal Key Genes Involved in Lignocellulose Degradation in *Sarcomyxa Edulis*

Chao Duan  
Jilin Agricultural University

Feng-Hua Tian  
Guizhou University

Lan Yao  
Jilin Agricultural University

Jian-Hua Lv  
Jilin Agricultural University

Chuan-Wen Jia  
Jilin Agricultural University

Chang-Tian Li (lct@jlau.edu.cn)  
Jilin Agricultural University

Research Article

**Keywords:** Sarcomyxa edulis, Lignocellulose degradation, WGCNA, Transcriptome, Key genes

**Posted Date:** January 12th, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-1165785/v1

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

In order to explore the molecular mechanism of *Sarcomyxa edulis* response to lignocelluloses degradation, the developmental transcriptomes was analyzed for six stages covering the whole developmental process, including mycelium growing to half bag (B1), mycelium in cold stimulation after full bag (B2), mycelium in primordia appearing (B3), primordia (B4), mycelium at the harvest stage (B5) and mature fruiting body (B6). A total of 6 samples were used for transcriptome sequencing, with three biological replicates. Based on the above transcriptome data, we constructed a co-expression network of weighted genes associated with extracellular enzyme physiological traits by WGCNA, and obtained 19 gene co-expression modules closely related to lignocelluloses degradation. In addition, a number of key genes involved in lignocelluloses degradation pathways were discovered from the four modules with the highest correlation with target traits. These results provide clues for further study on the molecular genetic mechanisms of *Sarcomyxa edulis* lignocelluloses degradation.

Introduction

*Sarcomyxa edulis* belonging to Basidiomycota, Mycenaceae, Sarcomyxa, degrades and utilizes forestry and agricultural residues which are rich in lignocellulose for its fruiting body development (Dai et al., 2003; Saito et al., 2014; Tian et al., 2021). Under appropriate environment, the mycelia absorb nutrients from substrates for fruiting body reproduction of the mushroom, which significantly hinges on the yield and quality of edible mushrooms (Zhou et al., 2016). The capacity for degrading lignocellulose, the key gene and enzymes related to decomposing and lignocelluloses degradation pathways in *S. edulis* are not well known, though many studies have been carried out on the model fungus, e.g. *Phanerochaete chrysosporium* (Kirk et al., 1978; Gold et al., 1993). It is therefore, the study on the degradation and utilization of lignocellulose by *S. edulis* would provide theoretical support for edible mushroom industry improvement.

In recent years, thanks to the continuous breakthrough of high-throughput sequencing technology and the continuous reduction of cost, transcriptome sequencing of multiple copies have been gradually applied to the systematic study of life science problems and have dramatically improved the efficiency of gene discovery (Tian et al., 2009; Cheng et al., 2013; Yu et al., 2012). The traditional comparative analysis of a small number of samples have been unable to effectively deal with massive biological information data. Therefore, bioinformatics analysis tools and methods came into being (Dillies et al., 2013). Among them, the weighted gene co-expression network analysis (WGCNA) can specifically screen the genes highly associated with the target traits, and carry out module classification to obtain a series of co-expression modules with high biological significance, and then screen the key genes (Wang et al., 2014), as an accurate and efficient bioinformatics and biological data mining, WGCNA algorithm is widely used in various fields of biology. It has played an important role in the analysis of multiple transcriptome data of many plants (brassica rapa (Greenham et al., 2017), strawberry (Hollender et al., 2014) ) and fungi (such as Chinese cordyceps (Xiao et al., 2019) ).
In the present study, fungal samples of six developmental stages were harvested from the artificial cultivation and the developmental transcriptomes were analyzed. Using WGCNA, we identified modules of co-expressed genes and candidate key genes for each developmental stage. In particular, we focused on the key genes involved in lignocelluloses degradation pathways discovering from the four modules with the highest correlation with target traits. This study provides important insights into the molecular networks underlying development and the molecular genetic mechanisms of *Sarcomyxa edulis* lignocelluloses degradation.

**Results**

**Global transcriptomic analysis**

To describe the gene expression patterns during the development, 18 libraries were constructed using samples from six growth stages of *Sarcomyxa edulis* (Fig. 1), with three biological replicates per sample. A total of 774.59 million raw reads were generated by Illumina paired-end sequencing. After cleaning and quality checks, 742.29 million clean reads were obtained, with an average of 41.23 million reads per copy (Table S1). More than 77.56% of the reads per replicate could be mapped to the *S. edulis* genome. The Q30 percentage of all sequences in the 18 libraries were over 91%.

**Gene expression level analysis**

By calculating the FPKM value of different genes, the number of genes and the expression level of single gene under different expression levels were analyzed. Generally, it is considered that the value of FPKM is greater than 1, indicating that the gene is expressed, and the higher the value of FPKM, the higher the level of gene expression. The results of gene expression level analysis showed that the median of 18 samples was relatively concentrated, and the value of log10 (FPKM) ranged from 3.82 to 4.59. The highest expression in B2 was 16.13 (Fig. 2, Table S2)

**Identification of differentially expressed genes across various developmental stages**

Differentially expressed genes (DEGs) across six stages were identified based on FPKM values with a Q-value of 0.05 and |log 2 (Fold Change)| of 1 set as thresholds (the adjustment range of the difference multiple was defined as greater than or equal to 2, and the adjustment range of the significance threshold was defined as less than or equal to 0.05). As shown in Fig. 3A, the largest number of differentially up-regulated genes were identified between the B5 and B6 stages (3171), followed by B4 vs. B5 (2478) and B1 vs. B2 (2243) (Fig. 3a). These results revealed that the expression profiles of the primordium (B4) and mature fruiting body (B6) are significantly different from mycelium (B5). Probably because they're different organizations. The greatest number of unique DEGs were found in the comparison between B5 and B6 (903), while only 153 DEGs were unique to the B3 vs. B4 comparison (Fig. 3B). There were 215 shared DEGs among all five comparisons of six stages and these were enriched primarily with the response to catalytic activity (GO:0003824) (Fig. S2).
Functional classification of differentially expressed genes

All DEGs were classified into three categories: biological process (BP), cellular component (CC) and molecular function (MF). The significantly enriched terms were almost the same for the B1 vs. B2, B2 vs. B3, B3 vs. B4, B4 vs. B5 and B5 vs. B6 comparisons (Fig. 4A-E). All of the significantly enriched terms were assigned into metabolic process (BP), cell (CC) and molecular function (MF).

The DEGs were mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database and tested for enrichment to further research their functions. The significantly enriched pathways were almost the same for the B1 vs. B2, B2 vs. B3, B3 vs. B4, B4 vs. B5 and B5 vs. B6 comparisons. All of the significantly enriched pathways were assigned into Global and overview maps and Carbohydrate metabolism, they are all part of the Metabolism (Table S3).

Gene co-expression networks construction

Because the results of network module analysis are easily affected by outliers, outliers must be removed before building the network module to ensure the accuracy of the results. There were 18 samples in this study, which were clustered by calculating the correlation coefficient of each sample expression level (Figure 5A). The Pearson's correlation coefficient of gene expression between replicates for each sample was shown in Figure S2. Samples with low correlation, or samples that cannot be clustered on the tree graph, are outlier samples, such as B6-1, B4-2 and B3-2. After the outliers are removed, the clustering tree of the remaining samples is shown in Figure 5B.

For the selection of power value, in general, we will take the minimum power value when the correlation coefficient reaches the plateau period (or greater than 0.8) as the parameter for subsequent analysis (as shown on the left of Figure 6), and we will count the changes of average gene connectivity under different power values (as shown on the right of Figure 6). According to the results of Figure 6, the power value parameter selected for this analysis is 8. The selected characteristic parameters for this analysis are shown in the figure below.

This analysis identified 19 different modules (marked with different colors, Fig. 7A), in which each branch of the gene clustering tree corresponds to one module. The 19 modules correlated with different development stages due to stage-specific expression profiles. The number of genes in each module is shown in Fig. 7B. The turquoise module, with 2312 identified genes, has the most genes. The grey module has the least genes, with 1 identified gene.

Using the correlation analysis between module eigenvalues and specific traits and phenotype data, find out the modules most related to traits and phenotypes for further analysis. The matrix of four physiological and biochemical traits (Laccase, acidic xylanase (ACX), cellulase (CL) and lignin peroxidase (Lip) activities) of the samples obtained in the previous step was used for correlation analysis with the above modules under different treatment time (Table S4), among which some modules were highly correlated with physiological and biochemical traits (Fig. 8). For example, the blue module had significant
positive correlation with CL and ACX \( (r = 0.96, r = 0.88, \text{respectively}) \). There was a significant positive correlation between dark orange modulus and laccase \( (r = 0.93) \), laccase and ACX \( (r = 0.82, r = 0.86) \), salmon modulus and lip \( (r_{128} = 0.91) \), salmon module and Lip \( (r = 0.91) \). Next, the genes in these four modules will be further studied.

**GO annotation of target modules**

In order to further explore the function of the target module, we map each module gene to each term of GO database(\url{http://www.geneontology.org/}), and calculate the number of genes in each term, so as to get the gene list and gene number statistics with a certain go function. The results show that these four modules can be significantly enriched into several go pathways under the three categories of biological processes, molecular functions and cellular components (Fig. 9). As shown in Fig. 8, the four modules can be enriched into catalytic activity genes and binding genes, indicating that WGCNA can be effectively constructed into biologically significant co-expression modules. These modules can be the focus of this study.

**Screening and functional analysis of key genes in target modules**

In order to obtain the key genes in the above four modules, Cytoscape software was used to visualize the gene regulation network and screen out the genes with high connectivity in the modules(Fig. 10). From the 4 modules with the highest correlation, 17 key genes that may be related to lignocellulose degradation were screened (Table 1). In these networks, each node represents a gene, the nodes are linked by lines, and genes at both ends of the line are generally assumed to have the same biological function. Nine key genes were screened out from the blue module, two key genes were screened out from the dark orange module, three key genes were screened out from the steelblue module, and three key genes were screened out from the salmon module.

Seventeen key genes were described in Table 1. At the same time, they can all be annotated to the CAZy database, indicating that they all belong to carbohydrate enzymes. It shows that our analysis is correct. Among them, six genes belonged to AAs family( Auxiliary Activities : Redox enzymes that act in conjunction with Cazymes), nine genes belonged to GHs family( Glycoside Hydrolases : hydrolysis and/or rearrangement of glycosidic bonds ), one gene belonged to CBMs family( Carbohydrate-Binding Modules: adhesion to carbohydrates ), and one gene belonged to PLs family( Polysaccharide Lyases: non-hydrolytic cleavage of glycosidic bonds ).

**Differential expression of lignocellulose degradation related genes**

The differential expression of these 17 genes was analyzed (Fig. 11). SE.1A4887, SE.1A4757, SE.1A2186, SE.1A5997, SE.1A7714 and SE.1A4339 were clustered together, and the expression level were high in all the six stages. SE.1A3039, SE.1A1616, SE.1A5542, SE.1A8861 and SE.1A1551 were clustered together, and the expression level were low. Among them, SE.1A3690 showed as the highest expression in the B1
stage, indicating its relation to lignocellulose degradation significantly. B4 and B6 are clustered together because they are all fruiting body. B2 and B3 cluster together because they are at a lower temperature.

**Validation of DEGs by Quantitative Real-Time PCR (qRT-PCR)**

In order to further verify the reliability of the transcriptome data, Four core genes and four other DEGs related to lignocelluloses degradation were selected for qRT-PCR (Fig. 12). The primers used in qPCR were listed in Table S5. The expression patterns of 8 genes detected by qRT-PCR were consistent with the transcriptome expression trend during the six development but a few genes were at some stage showed different expression trends. The above disparities may have been caused by differences between transcriptome sequencing and qRT-PCR as detection methods, a certain degree of inconsistency (about 30-40%) (Chen et al., 2020). The result was normal and reasonable. The results showed that the qRT-PCR results of most of the genes were consistent with the transcriptome data.

**Discussion**

The degradation and utilization of lignocellulose is a very important prerequisite for the formation of yield and quality of edible fungi. In addition, most of the agronomic traits are highly correlated with the whole population structure. However, the current bioinformatics and data analysis methods can be used to deal with them through co-expression modularization. Thousands of genes with similar biological functions are assigned to the same module. By studying the biological significance of the module, the function of genes in the module can be further mined, which can better solve the problem of insufficient analysis of characters of populations with complex genetic background. Therefore, it is widely used in the research of big data omics. Therefore, this paper studied the transcriptome analysis of six materials from different growth stages of *Sarcomyxa edulis*, in order to obtain new research clues related to lignocellulose degradation of *S. edulis*. At the same time, the physiological and biochemical indexes of lignocellulose degrading enzymes in each sample at different growth stages were also detected (ACX, Laccase, CL, and Lip).

After obtaining a large number of transcriptome data and physiological data, how to mine the biological significance of transcriptome data has become the primary consideration. Network analysis is widely used in genome, transcriptome and metabonomics data mining. In order to find out the genes related to lignocellulose degradation, WGCNA analysis method was used in this study. The genes related to the target traits could be specifically screened out and modularized to obtain the co-expression modules with high biological significance, which has proved to be an effective data mining method. The degradation of lignocellulose requires a complex process involving a variety of enzymes. ACX, laccase, CL, and Lip, as extracellular enzymes, can directly reflect the degradation of lignocellulose, and can be used as target traits in WGCNA analysis. Therefore, we enriched 19 co-expression modules associated with physiological traits by WGCNA analysis method, four of which were highly correlated with these traits. Through further network analysis, we found the key genes of these four modules. These results indicate
that WGCNA can be constructed into co-expression modules and pathways with biological significance, which will lay a foundation for further research.

The results show that WGCNA analysis can be used to dig out the gene modules and key genes which are highly related to the biological significance of the target characters, and provide a new research idea for the exploration of target genes, and provide reference for the analysis of complex agronomic characters. In this study, we focus on four high correlation regulatory modules, blue, darkorange, steelblue and salmon. Although the other gene modules are not discussed in detail, they may also contain pathways related to lignocellulose degradation, which can further explore the biological significance of these modules.

**Conclusion**

In this study, a co-expression network of weight genes associated with the physiological characteristics of lignocellulose degradation was constructed. Nineteen gene modules closely related to lignocellulose decomposition were obtained and the go pathway was enriched. Four gene modules with the best correlation degree with the target traits were selected for further analysis. Among them, 9 key genes were selected in blue module, 2 key genes were selected in darkorange module, 3 key genes were selected in steelblue module, 3 key genes were selected in salmon module, and through functional annotation, it was found that key genes were closely related to lignocellulose degradation. The results can provide clues for the molecular mechanism of decomposition and utilization of lignocellulose, and provide theoretical support for the cultivation of new varieties of high-yield edible fungi.

**Methods**

**Collection of fungal samples at different developmental stages**

*The Sarcomyxa edulis* strain (No. 2016120327) were used in this study. It's maintained in the Engineering Research Center of the Chinese Ministry of Education for Edible and Medicinal Fungi, Jilin Agricultural University, China. The cultivation substrates using polypropylene bags as containers contained 1200 g medium including 29.64% oak sawdust, 7.6% wheat bran, 0.38% gypsum, and 0.38% lime.

The samples were collected during the growth period of mycelium growing to half bag (B1), mycelium in cold stimulation after full bag (B2), mycelium in primordia appearing (B3), primordia (B4), mycelium at the harvest stage (B5), and mature fruiting body (B6). Take 3 bags each time, mix well according to the same part, and store the sample at -80°C.

**RNA isolation, library construction and sequencing**

Total RNA was extracted from the samples using RNA prep Pure PlantPlus Kit (Tiangen Biotech, Beijing, China). RNA degradation and contamination were monitored on 1% agarose gels and RNA purity and
concentration were measured using the BIOSPEC-NANO® (230 V) spectrophotometer (SHIMADZU). Three biological replicates were used for each sample. 18 libraries were sequenced using Illumina HiSeq 4000 by Gene Denovo Biotechnology Co. (Guangzhou, China).

**Sequencing data analysis**

High quality clean reads were obtained from the sequencing machines containing raw reads containing adapters or low-quality bases which should be filtered out. The rRNA removed reads of each sample were then mapped to the reference genome by TopHat2 (version2.0.3.12). The reconstruction of the transcripts was performed using Cufflinks software, which together with TopHat2. Gene abundance was quantified by RSEM software. Our laboratory has uploaded the reference genome sequence data to NCBI database (BioProject: PRJNA483858) (Tian et al., 2021). The gene expression level was normalized using the FPKM method. To identify differentially expressed genes across samples or groups, the edge R package was used. We identified genes with a fold change $\geq 2$ and a false discovery rate (FDR) $<0.05$ in comparison as significant DEGs. DEGs were then subjected to enrichment analysis of GO (http://www.geneontology.org/) and KEGG databases (https://www.kegg.jp/)(Kanehisa et al., 2017; Kanehisa et al., 2016; Kanehisa et al., 2000).

**Analysis of key genes in the gene co-expression network**

The filtered data was used to construct the network and identify modules. In this study, the lignocellulose degradation genes in *Sarcomyxa edulis* were combined with the data of physiological and biochemical indexes of four growth period in mycelium (Table S1). WGCNA analysis was performed using the OmicShare tools, a free online platform for data analysis (https://www.omicshare.com/tools). Through WGCNA analysis and gene set phenotype association analysis, the key genes were deeply excavated. The genetic network map, which was drawn according to the kME values, was created using Cytoscape software (Qi et al., 2020; Hollender et al., 2014).

**Validation of DEGs by qRT-PCR**

The expression levels of 8 genes were screened by qRT-PCR analysis, and these genes showed significant changes in the transcriptional expression data of 6 samples. Gene-specific primers were designed using Primer v5.0. ITS gene was used as an internal control for gene expression (Table S5). The gene expression data were analyzed using the $2^{-\Delta\Delta Ct}$ method (Xiao et al., 2019).

**Declarations**

**Authors’ contributions**

Provided experimental material: FHT. Conceived and designed the experiments: CD (Chao Duan), and CTL. Performed the experiment: CD, LY and CWJ. Wrote the manuscript: CD. Analyzed data and submitted data: CD and LY. Modified the manuscript: CTL, LY, and JHL.

**Funding**
This work was supported by the National Key Research and Development Program of China under Grant [2020YFD1000304], National Key R&D Program of China [2020YFD1000300], China Agriculture Research System [CARS-20] and the National Key Research and Development Program of China under Grant [2017YFD0601002]

Availability of data and materials

The raw data have been submitted under BioProject number PRJNA739377 to the Sequence Read Archive (SRA) database at NCBI (https://dataview.ncbi.nlm.nih.gov/object/PRJNA739377?reviewer=dlp50ehjesb7chmkqn39ofhk6).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Cheng CK, Au CH, Wilke SK, Stajich JE, Zolan ME, Pukkila PJ, et al. 5'-Serial Analysis of Gene Expression studies reveal a transcriptomic switch during fruiting body development in Coprinopsis cinerea. BMC Genomics. 2013

2. Dai YC, Niemela T, Qin GF. Changbai wood-rotting fungi. A new pleurotoid species Panellus edulis[J]. Ann. Bot. Fennici., 2003, 40:107-112.

3. Dillies Marie-Agnès, Andrea, R., Julie, A., Christelle, H. A., Marine, J. , & Nicolas, S. , et al. A comprehensive evaluation of normalization methods for illumina high-throughput rna sequencing data analysis. Briefings in Bioinformatics(6), 671-683.

4. Fenghua Tian , Changtian Li and Yu Li. Genomic Analysis of Sarcomyx a edulis Reveals the Basis of Its Medicinal Properties and Evolutionary Relationships. Front. Microbiol., 2021.652324

5. Gold M H, Alic M. Molecular biology of the lignin-degrading basidiomycete Phanerochaete chrysosporium.[J]. Microbiological reviews, 1993, 57(3):605-622.

6. Greenham K, Guadagno C R, Gehan M A, et al.. Temporal network analysis identifies early physiological and transcriptomic indicators of mild drought in brassica rapa. eLife, 2017, 6: e29655.
8. Hollender C A, Kang C, Darwish O, Geretz A, Matthews B F, Slovin J, Alkharouf N, Liu Z. Floral transcriptomes in woodland strawberry uncover developing receptacle and anther gene networks. Plant Physiol, 2014, 165: 1062–1075.

9. Kanehisa, Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 45, D353–D361 (2017).

www.nature.com/scientificreports/12 Scientific RepoRts | 7: 9266 | DOI:10.1038/s41598-017-08049-z.

10. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28, 27–30 (2000).

11. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44, D457–D462 (2016).

12. Kirk T K, Schultz E, Connors W J, et al. Influence of culture parameters on lignin metabolism by Phanerochaete chrysosporium[J]. Archives of Microbiology, 1978, 117(3):277-285.

13. Qi, Shen D, Zhang T Y, et al. Comparative transcriptome and co-expression analysis reveal key genes involved in leaf margin serration in Perilla frutescens[J]. Chinese Herbal Medicines, 2020, v.12(03):66-73.

14. Saito, T., Tonouchi, A., & Harada, Y. (2014). Biological characteristics and molecular phylogeny of sarcomyxa edulis comb. nov. and s. serotina. 

15. Tian, J., K. Ma and I. Saaem (2009). Advancing high-throughput gene synthesis technology. Molecular Biosystems.5(7): 714.

16. Tian, F., Li, C., & Li, Y. (2021). Genomic analysis of sarcomyxa edulis reveals the basis of its medicinal properties and evolutionary relationships. Frontiers in Microbiology, 12, 652324.

17. Wang J, Zhang X, Shi M, et al. Metabolomic analysis of the salt-sensitive mutants reveals changes in amino acid and fatty acid composition important to long-term salt stress in Synechocystis sp. PCC 6803[J]. Functional & Integrative Genomics, 2014, 14(2):431-440.

18. Xiao Li, Fen Wang, Qing Liu, et al.. Developmental transcriptomics of Chinese cordyceps reveals gene regulatory network and expression profiles of sexual development-related genes[J]. BioMed Central,2019,20(1).

19. X Chen, X Yang, J Xie, Ding, W., Li, Y., & Yue, Y., et al. (2020). Biochemical and comparative transcriptome analyses reveal key genes involved in major metabolic regulation related to colored leaf formation in osmanthus fragrans 'yinbi shuanghui' during development. Biomolecules, 10(4).

20. Yu GJ, Wang M, Huang J, Yin YL, Chen YJ, Jang S, et al. Deep Insight into the Ganoderma lucidum by Comprehensive Analysis of Its Transcriptome. PLoS One. 2012

21. Zhou S, Ma F, Zhang X, et al. Carbohydrate changes during growth and fruiting in Pleurotus ostreatus[J]. Fungal Biology, 2016:852-861.

Tables
Table 1: Core gene information table (SE.1A3690, SE.1A4757, SE.1A4339, and SE.1A2186 were selected for qRT-PCR).

| Module       | Gene ID   | Symbol | Family in CAZy database | Description                        |
|--------------|-----------|--------|-------------------------|------------------------------------|
| blue         | SE.1A3690 | GLX    | AA5_1                   | copper radical oxidase              |
|              | SE.1A4757 | –      | GH128                   | glycoside hydrolase                |
|              | SE.1A1422 | ARMGADRAFT_1018426 | AA3_2 | Pyranose dehydrogenase3 |
|              | SE.1A3039 | exgA   | GH5_9                   | glycoside hydrolase family 5 protein |
|              | SE.1A8947 | POX1   | AA1_1                   | Laccase-2                           |
|              | SE.1A1616 | exgB   | GH5_15                  | glycoside hydrolase family 5 protein |
|              | SE.1A5866 | vpl2   | AA2                     | evidence of expression at protein level |
|              | SE.1A4339 | AFUB_029980 | GH16_1 | glycoside hydrolase family 16 protein |
|              | SE.1A2186 | –      | CBM13                   | carbohydrate-binding module family 13 protein |
| darkorange   | SE.1A6315 | bglJ   | GH3                     | glycoside hydrolase family 3 protein |
|              | SE.1A7714 | gba-2  | GH30_3                  | glucan endo-1,6-beta-glucosidase     |
| steelblue    | SE.1A1136 | CDH-1  | AA8                     | cellobiose dehydrogenase1           |
|              | SE.1A1551 | agn1   | GH71                    | glycoside hydrolase family 71 protein |
|              | SE.1A8861 | abfC   | GH51                    | glycoside hydrolase family 51 protein |
| salmon       | SE.1A5542 | ARB_07870 | AA2 | heme peroxidase                |
|              | SE.1A4887 | btgC   | GH17                    | glycoside hydrolase family 17 protein |
|              | SE.1A5997 | –      | PL14_5                  | polysaccharide lyase family 14 protein |

Figures
Figure 1

Developmental stages of *Sarcomyxa edulis*. A: Mycelium growing to half bag(B1). B: Mycelium in cold stimulation after full bag(B2). C: Mycelium in primordia appearing(B3) and primordial(B4), D: Mycelium at the harvest stage(B5) and mature fruiting body(B6).
Figure 2

Violin map of horizontal distribution of gene expression (FPKM) in different samples. The abscissa is the name of the sample, and the ordinate is log10 (FPKM). The values from top to bottom represent the maximum, the upper quartile, the median, the lower quartile and the minimum in turn. The width of each violin represents the number of genes under the same expression.
Figure 3

Differentially expressed genes across various developmental stages. **A** Venn diagrams of differentially up-regulated genes. **B** Histogram of differentially expressed genes.
Figure 4

GO functional classification of differentially expressed genes. The green bars represent biological processes; yellow bars represent cellular components; blue bars represent molecular functions. Only the significant Go terms (P< 0.005) were shown.
Figure 5

Dendrogram of samples. A: dendrogram of all 18 samples. B: dendrogram of samples after removing outlier.
Figure 6

Graph of Power value. Left: the abscissa represents the power value, the ordinate represents the correlation coefficient, the blue line represents the correlation coefficient of 0.8, and the red line represents the correlation coefficient of 0.9. Right: the abscissa represents the power value, and the ordinate represents the average connectivity of genes.
Nineteen different modules identified. A: Gene co-expression network gene clustering number and modular cutting. Dynamic Tree Cut is the module divided according to clustering results. Merged Dynamic is the module division of merged modules with similar expression patterns according to module similarity. The subsequent analysis is conducted in accordance with merged modules. In the case of trees, the vertical distance represents the distance between two nodes (between genes), and the
Figure 8

Association analysis of gene co-expression network modules with physiological and biochemical traits. The horizontal axis represents different characteristics, and the vertical axis represents each module. The red lattice represents a positive correlation between the physiological traits with the module, while the green lattice represents a negative correlation.
Figure 9

Gene GO annotation in the target modules. **A:** blue module. **B:** darkorange module. **C:** steelblue module. **D:** salmon module.
Figure 10

Gene co-expression network and key genes in four modules. **A**: blue module. **B**: darkorange module. **C**: salmon module. **D**: steelblue module.
Figure 11

Heatmap clustering of key genes in the six stages. The levels of expression are represented by log 2 (FPKM) values after centralization correction. Genes with similar patterns of expression are clustered together.
Figure 12

The qRT-PCR analysis of gene expression compared to the RNA-seq data.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.
• Supplementalfigures.docx
• TableS1.xls
• TableS2.xlsx
• TableS3.xlsx
• TableS4.xlsx
• TableS5.doc