Pol. J. Environ. Stud. Vol. 25, No. 6 (2016), 2327-2338
DOI: *10.15244/pjoes/63297*

Original Research

Genetic Diversity and RNA-seq Transcriptome Analysis of *Tricholoma matsutake* from Sichuan Province, China

Xiang Ding[1, 2], Panpan Wang[2], Yiling Hou*[2], Mei Wang[3], Wanru Hou[2], Yuming Li[3]

[1]Ecological Security and Protection Key Laboratory of Sichuan Province, Mianyang Normal University, Mianyang, 621000, China
[2]Key Laboratory of Southwest China Wildlife Resources Conservation (Ministry of Education), College of Life Sciences, China West Normal University, Nanchong, 637009, China
[3]National Center for Sweet Potato Improvement Centre of Nanchong, Nanchong Academy of Agricultural Sciences, Nanchong, 637001, China

Received: 16 March 2016
Accepted: 19 May 2016

Abstract

The sequences of the internal transcribed spacer (ITS) regions of ectomycorrhizal fungi collected from Sichuan Province were analyzed using a PCR primer pair specific to *T. matsutake*. The amplified fragments were sequenced and compared with each other to build a phylogenetic tree. The mRNA deep sequencing approach was adopted to identify differentially expressed *T. matsutake* genes among the transcriptomes developed from a Xiaojin sample. A phylogenetic analysis of the aligned sequences was performed using maximum-likelihood (ML) and neighbor-joining (NJ) analyses. The results clearly showed that the KD (KM657344) and BT (KM657342) strains were more closely related to each other than to other strains. Moreover, *T. matsutake* from Sichuan differed from those specimens derived from Heilongjiang, Yunnam, and Guizhou provinces of China, Finland, and Japan. Furthermore, there was extremely high homology among these *T. matsutake* samples, despite some genetic variation. In addition, the genome of *T. matsutake* was sequenced using Illumina sequencing technology (RNA-seq). In all, a total of 24,549,990 reads were obtained that yielded 18,266,492 high-quality clean reads. The quality reads were excluded later. The BLAST analysis of the sequence reads against the NR database indicated that *T. matsutake* shared a high number of contigs with *Laccaria bicolor*. The results also indicated that catalytic activity, metabolic processes, metabolic pathways, and biosynthesis of secondary metabolites were the main functions identified by gene ontology (GO) and the *Kyoto Encyclopedia of Genes and Genomes* (KEGG) pathway analyses. Phylogenetic analysis showed that *T. matsutake* growing in Sichuan differed from samples growing in other regions. The differences in secondary metabolites between the Sichuan and Xiaojin samples may be due to differences in metabolic pathways. Thus the study provides a foundation for understanding *T. matsutake* biogeography and

*e-mail: starthlh@126.com

origins, and identifies DEGs in the Xiaojin sample to help elucidate the molecular mechanisms in secondary metabolite synthesis.

Keywords: *Tricholoma matsutake*, ITS, phylogenetic, RNA-seq, differentially expressed gene, Sichuan, China

Introduction

Tricholoma matsutake (S. Ito et Imai), is one of the most valuable fungi in the world and it belongs to the genus Tricholoma (class Basidiomycota, family Tricholomaceae). *T. matsutake* is found in low-latitude regions of Asia, such as Sichuan, Yunnan, and Guizhou provinces of China, and in mid-latitude regions such as Japan, Korea, Finland, Canada, and Jilin Province of China [1-4]. *T. matsutake* usually grow in coniferous forests and broadleaf forests. In general, the range of *T. matsutake* approximately matches the distribution of coniferous genera such as *Pinaceae* and is less similar to the distribution of *Fagaceae*. Principal hosts include *Pinus densiflora*, *Pinus thunbergii*, and *Pinus taiwanensis Hayata*. However, *Quercus aquifolioides* and *Quercus pannosa* are the principal hosts in Sichuan Province [5-9]. Nonetheless, several forms of *T. matsutake*, such as the matsutake harvested from Xiaojin in Sichuan Province, are frequently found in mixed forests in association with *Quercus*. A large number of bioactive substances extracted from the fruiting bodies of *T. matsutake*, such as volatile compounds, polysaccharides, and polysaccharide-protein complex fractions, have been found to have immunomodulating and antioxidant properties. Despite their importance, their biogeography and origins have remained elusive, and little is known about the molecular mechanisms associated with the differences in secondary metabolite synthesis [10-14].

In recent years, molecular methods have been developed to investigate genetic variations within fungi [15-20]. Internal transcribed spacer (ITS) regions have been widely used in phylogenetic studies at different taxonomic levels because they are accessible with universal primers and their DNA sequences are variable [21-24]. Genome, transcriptome, proteome, and metabolome analyses have also played increasingly important roles in studies of the regulation of physiological activity and metabolic mechanisms. RNA-seq (transcriptome analysis) is used for transcriptome quantification and structural analysis. Transcriptome analysis lays the foundation for research into gene structure and function [25-27].

In our previous study, among several fruiting bodies from different geographical origins in southwestern China, the Xiaojin sample was found to have the best health effects and was delicious and nontoxic [28]. Furthermore, a novel heteropolysaccharide was isolated from the fruiting bodies of the Xiaojin sample [10]. Thus, the Xiaojin sample is a good model to investigate genes related to secondary metabolite synthesis. Here, the sequences of the ITS region of ectomycorrhizal fungi collected from Sichuan Province were analyzed using a specific PCR primer pair for *T. matsutake*. The amplified fragments were sequenced and compared with each other to build a phylogenetic tree. An mRNA deep sequencing (RNA-seq) approach was also used to identify differentially expressed genes (DEGs) in the transcriptomes of the Xiaojin sample. The goal of this study was to identify the biogeography and origins of this fungus and to identify differentially expressed genes DEGs in the Xiaojin sample to elucidate the molecular mechanisms associated with differences in secondary metabolite synthesis.

Materials and Methods

Fungal Material

T. matsutake is distributed in alpine valleys in Sichuan, located in upstream areas of the Jinsha, Yalong, and Dadu Rivers. Some samples were obtained from Jiulong, Daocheng, Yajiang, Litang, Batang, and Kangding in the Ganzi Tibetan Autonomous Prefecture, and from Rangtang, Xiaojin, and Barkam in the Aba Tibetan and Qiang Autonomous Prefecture. Other samples were obtained from Muli, Yanyuan, Yanbian, Huidong, and Mianning in the Liangshan Tibetan Autonomous Prefecture (Fig. 1). All field studies were permitted and did not involve endangered or protected species, and no samples were obtained from endangered or protected species protection zones. The names of the 37 samples are presented in Table 1 (see stock Nos. 1-37). These voucher specimens were preliminarily authenticated by Prof. Zhirong Yang

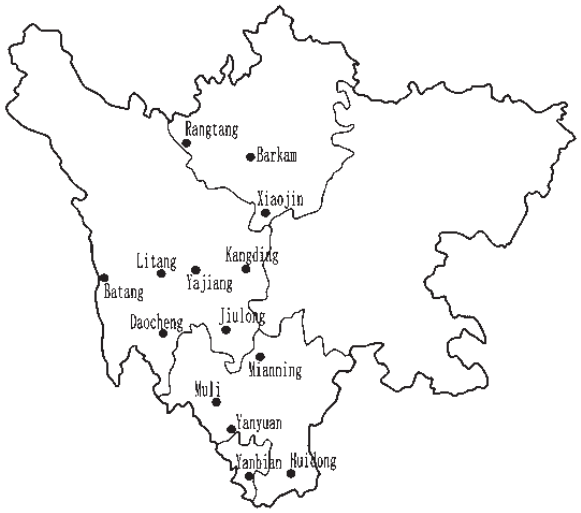

Fig. 1. Biogeographical distribution of *Tricholoma matsutake* varieties in Sichuan Province.

Table 1. *Tricholoma matsutake* strains used in this study.

Stock number	Collection number	Location	GenBank accession number
01	BT	Batang, Sichuan, China	KM657342
02	RT	Rangtang, Sichuan, China	KM657343
03	KD	Kangding, Sichuan, China	KM657344
04	YJ	Yajiang, Sichuan, China	KM657345
05	JL	Jiulong, Sichuan, China	KM657346
06	DC	Daocheng, Sichuan, China	KM657347
07	TLwpp	Litang, Sichuan, China	KM073272
08	TM07	Barkam, Sichuan, China	KM102537
09	TX07	Xiaojin, Sichuan, China	KM102538
10	YJ1-1	Yajiang, Sichuan, China	KM581365
11	YJ2-1	Yajiang, Sichuan, China	KM581366
12	YJ3-1	Yajiang, Sichuan, China	KM581367
13	YJ4-1	Yajiang, Sichuan, China	KM581368
14	YJ5-1	Yajiang, Sichuan, China	KM581369
15	ML1-1	Muli, Sichuan, China	KM581370
16	ML2-1	Muli, Sichuan, China	KM581371
17	ML3-1	Muli, Sichuan, China	KM581372
18	ML4-1	Muli, Sichuan, China	KM581373
19	ML5-1	Muli, Sichuan, China	KM581374
20	YY1-1	Yanyuan, Sichuan, China	KM581375
21	YY2-1	Yanyuan, Sichuan, China	KM581376
22	YY5-1	Yanyuan, Sichuan, China	KM581377
23	YY5-3	Yanyuan, Sichuan, China	KM581378
24	YB2-1	Yanbian, Sichuan, China	KM581379
25	YB3-1	Yanbian, Sichuan, China	KM581380
26	YB5-2	Yanbian, Sichuan, China	KM581381
27	XJ1-1	Xiaojin, Sichuan, China	KM581382
28	XJ1-6	Xiaojin, Sichuan, China	KM581383
29	XJ2-2	Xiaojin, Sichuan, China	KM581384
30	XJ2-5	Xiaojin, Sichuan, China	KM581385
31	XJ3-1	Xiaojin, Sichuan, China	KM581386
32	HD	Huidong, Sichuan, China	KM581387
33	HD-2	Huidong, Sichuan, China	KM581388
34	HD-4	Huidong, Sichuan, China	KM581389
35	MN-1	Mianning, Sichuan, China	KM581390
36	MN-2	Mianning, Sichuan, China	KM581391
37	MN-3	Mianning, Sichuan, China	KM581392
38	jl02	Sichuan, China	EU552797
39	lt02	Sichuan, China	EU552800
40	XJ02	Sichuan, China	KF954507
41	YY1	Sichuan, China	KF961006
42	YB1	Sichuan, China	KF961007
43	ML1	Sichuan, China	KF961009
44	MB1	Sichuan, China	KF961010
45	HD1	Sichuan, China	KF961011
46	TF46	Diqing, Yunnan, China	DQ323063
47	Unknown	Guizhou, China	EU051918
48	Unknown	Heilongjiang, China	GU357477
49	Unknown	Finland	JF908736
50	Unknown	Finland	GU373503
51	Tm1	Japan	AF204868
52	TmA-5	Japan	AF202772
53	Tm-33	Japan	AF204806

(College of Life Sciences, Sichuan University, Chengdu, China) and preserved at the Key Laboratory of Southwest China Wildlife Resources Conservation (Ministry of Education), College of Life Sciences, China West Normal University.

DNA Extraction

Total genomic DNA was isolated from each sample according to the CTAB method, with slight modifications. After being triturated with liquid nitrogen, fruiting bodies of *T. matsutake* were transferred to 1.5 ml microcentrifuge tubes. Subsequently, 0.5 ml of extraction buffer was added (100 mM NaCl, 200 mM Tris-HCl pH 8.0, 25 mM EDTA, 0.5% SDS [w/v]). The tube was then centrifuged at 12,000 rpm for 6 min after incubation at 37ºC for 1 h. The upper aqueous layer was moved to a new tube, and 2× CTAB buffer (100 mM Tris-HCl pH 8.0, 2% CTAB [w/v], 1.4 M NaCl) was then added. The mixture was placed in a 65ºC water bath for 30 min. The upper aqueous layer was extracted at least twice with an equal volume of a 1:24:25 (v:v:v) mixture of isoamyl alcohol, chloroform, and phenol. The mixture was then centrifuged at 15,000 rpm for 12 min, followed by extraction with isoamyl alcohol and chloroform at a ratio of 1:24 (v:v). The upper layer was transferred to a new tube, and DNA was precipitated with an equal volume of isopropanol at -20ºC for 10 h. The precipitate was washed with 70% (v:v) ethanol and absolute ethanol, dried, and dissolved in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). The concentration of the DNA was measured using a UV spectrophotometer (BIO-RAD).

PCR Amplification and Sequencing of rDNA ITS Region

Primers were synthesized by Shanghai Invitrogen Corporation, China. The ITS region of nuclear rDNA was amplified using the primers ITS5 (5'-GGAAGTAAAAGTCGTAACAAGG-3') and ITS4 (5'-TCCTCCGCTTATTGATATGC-3') [29, 30]. The DNA sequence amplified by this pair of primers included partial sequences of 18S rDNA and 28S rDNA and the entire sequences of ITS1, 5.8S rDNA, and ITS2. The PCR amplification was performed in a 50 µl reaction volume containing 100 ng of template DNA, 5.0 µl of 10× PCR buffer (Mg^{2+} free), 0.2 mM of each dNTP, 0.2 µM of each primer, 1.5 mM $MgCl_2$, and 1.5 U of Taq DNA polymerase. The PCR thermal cycling conditions were an initial denaturation at 94ºC for 5 min; then cycles of 40 s at 94ºC, 30 s at 55ºC, and 90 s at 72ºC; followed by extension at 72ºC for 5 min. The PCR products were visualized with TAE-buffered 1.5% agarose gel electrophoresis and were then purified and sequenced by the Shanghai Invitrogen Corporation. Negative controls (no template DNA) were used in each experiment to test for the presence of nonspecific PCR products.

Sequence Alignments and Analysis

BLAST searches were conducted using the sequences of these fungi to identify homologous sequences. The DNA sequences of several strains used in this study were submitted to GenBank, and the accession numbers are shown in Table 1. Corresponding sequences from representative species were included in the phylogenetic analyses. The sequences were manually edited using DNAMAN and were aligned in Mega 5.2 using the ClustalW algorithm with the default parameters. Preliminary multiple alignments of all the sequence data were performed using ClustalX (1.8). A phylogenetic analysis of the aligned sequences was performed using the MEGA (4.0) Kimura 2-Parameter Distance Model based on ML or NJ analyses. The NJ and ML analyses included bootstrapping using 1,000 replicates.

Basic Statistics of the Quality of the Original Sequencing Reads

Transcriptome sequencing was conducted at Beijing GGT Co., Ltd. RNA Solexa sequencing yielded the sample data. Image data were transformed into sequence data by base calling. The data were provided in FASTQ format, including the names of the sequencing reads, sequences, and quality measures. Each FASTQ-formatted file is described by four lines: the first and third rows are the sequence names generated by the sequencing, the second line is the sequence, and the fourth line is the sequence quality.

Quality Pretreatment of Reads

Linker sequences, low-quality reads, and contaminating reads were removed from the original data to obtain clean sequence data. The length distribution statistics were then computed as follows:

1. Remove low-quality reads: quality threshold 20 (error rate = 1%), proportional threshold = 40%.
2. Remove reads with proportion of the larger sequence of N; threshold = 4%.
3. Remove the linker sequence.
4. Compute the length distribution statistics of the clean sequence.

Assembly Analysis

Trinity is a novel method for the efficient and robust de novo reconstruction of transcriptomes from RNA-seq data. Trinity combines three independent software modules: Inchworm, Chrysalis, and Butterfly, which are sequentially applied to process large volumes of RNA-seq reads. Trinity partitions the sequence data into many individual de Bruijn graphs (each representing the transcriptional complexity at a given gene or locus) and then processes each graph independently to extract full-length splicing isoforms to tease apart transcripts derived from paralogous genes. Briefly, the process consists of the following:

1. Inchworm assembles the RNA-seq data into the unique sequences of transcripts, often generating full-length transcripts for a dominant isoform, but then reports only the unique portions of alternatively spliced transcripts.
2. Chrysalis clusters the Inchworm contigs into clusters and constructs complete de Bruijn graphs for each cluster. Each cluster represents the full transcriptional complexity for a given gene (or sets of genes that share sequences in common). Chrysalis then partitions the full read set among these disjointed graphs.
3. Butterfly then processes the individual graphs in parallel, tracing the paths that read and pairs of reads taken within the graph, ultimately reporting full-length transcripts for alternatively spliced isoforms, and teasing apart transcripts that correspond to paralogous genes.

Trinity Results and Common Database Comparison

BLASTX was performed against the COG, EST, NR, Pfam, UniProt, and other protein databases to identify all the unigenes, The results, which provided the corresponding database and degree of homology standard (E-value), were used to functionally annotate the genes. The annotation returned a value of NULL when no homologous genes were identified in the database.

GO Annotation and KEGG Enrichment Analyses

All genes were annotated for protein function using InterProScan (*www.ebi.ac.uk/interpro*) and BLASTX against the NCBI NR database. The resulting InterPro and BLAST annotations were converted into GO annotations, and all GO terms were mapped to GO slim categories. The significance of the functional GO slim enrichment was evaluated using Fisher's exact test within Blast2GO with a false discovery rate (FDR) of <0.05. Significantly enriched KEGG pathways were identified with KOBAS 2.0 using a hypergeometric test and the Benjamini-Hochberg FDR correction.

Results and Discussion

Phylogenetic Analyses Based on the Sequence of the ITS Region

A part of the ITS region of *T. matsutake* samples of various geographical origins was amplified using PCR and a *T. matsutake*-specific primer pair. BLAST searches were conducted based on the sequences of 37 *T. matsutake* samples from Sichuan Province (486,000 km^2) to identify homologous sequences (Fig. 1). The ITS of the fungus was sequenced and submitted to GenBank (Table 1). A phylogenetic tree of aligned sequences was constructed using maximum-likelihood (ML) (Figs 2 and 3) and neighbor-joining (NJ) analyses (Figs 4 and 5). Both trees clearly showed two phylogenetic lineages arising from the root. The results showed differences between the ML tree and the NJ tree.

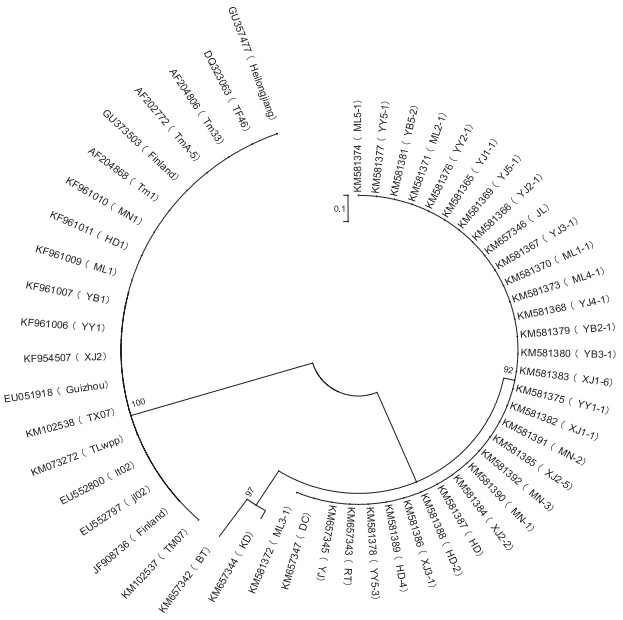

Fig. 2. Topology of *Tricholoma matsutake* obtained using the maximum-likelihood (ML) method.

The first clade at the top of these two trees consisted of 34 individuals of *T. matsutake*, but strains KD (KM657344) and BT (KM657342) were shown to be more closely related to each other than to other strains. The *T. Matsutake* samples from the three regions of Sichuan (shown in Fig. 1) had a closer genetic relationship in adjacent regions. For example, samples from Jiulong, Daocheng, Yajiang, Litang, Batang, and Kangding in the Ganzi Tibetan Autonomous Prefecture had a close genetic relationship, whereas samples from Rangtang, Xiaojin, and Barkam of the Aba Tibetan and Qiang Autonomous Prefecture had a closer genetic relationship. However, the

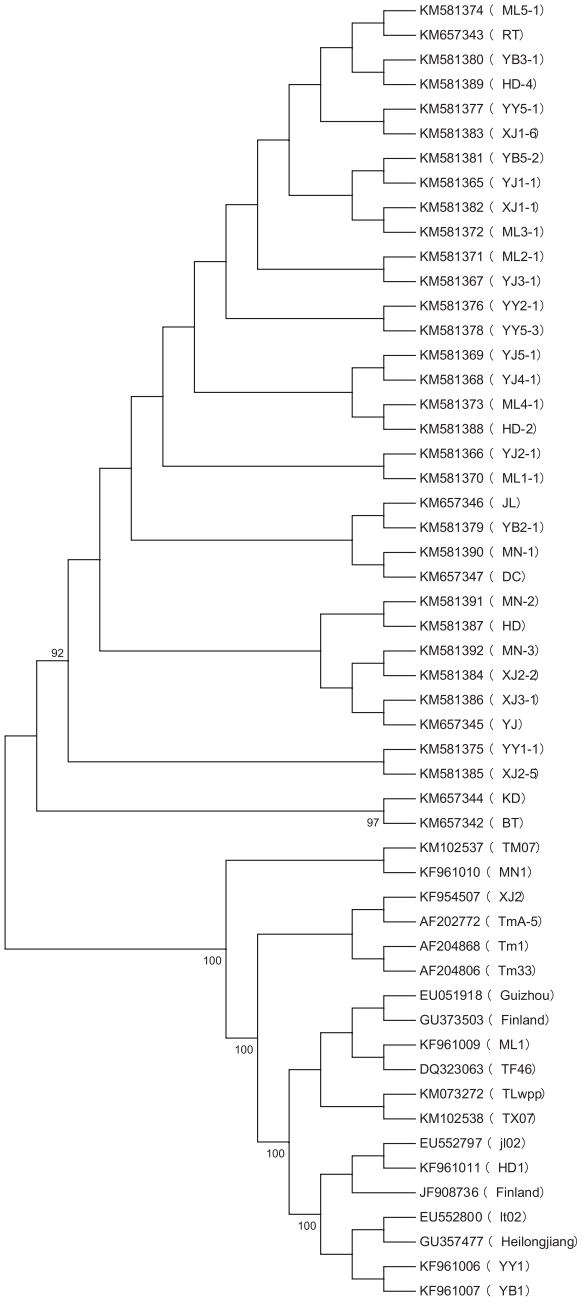

Fig. 3. Maximum-likelihood (ML) tree generated for *Tricholoma matsutake* obtained by the ML search. The number below each internode represents the percentage of 1,000 bootstrap replicates supporting binary partition.

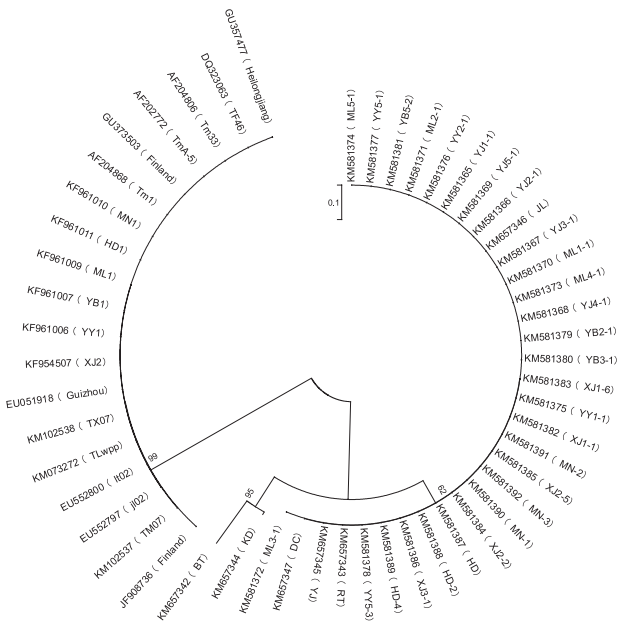

Fig. 4. Topology of *Tricholoma matsutake* obtained using the neighbor-joining (NJ) method.

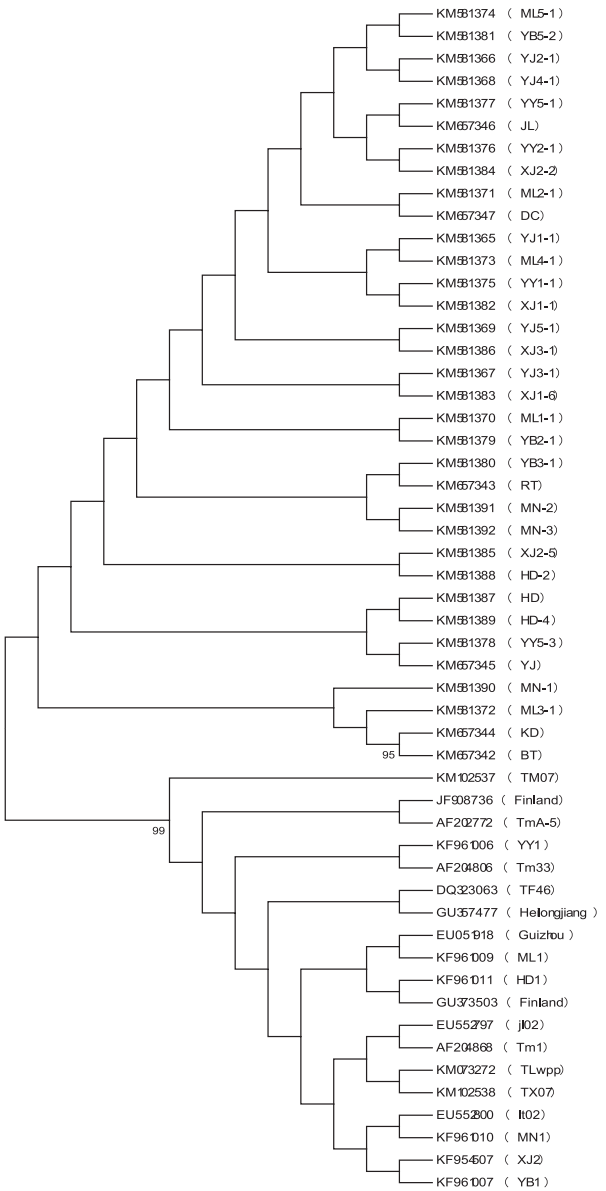

Fig. 5. Neighbor-joining (NJ) tree generated for *Tricholoma matsutake* obtained by the NJ search. The number below each internode represents the percentage of 1,000 bootstrap replicates supporting that binary partition.

relationship between *T. matsutake* from the Sichuan area and samples from other places was distant because the first clade at the top of these two trees consisted of 34 samples of *T. matsutake* that were all from Sichuan Province.

The second clade at the bottom of the tree consisted of 19 strains, all of which were closely related, including samples from the Heilongjiang, Yunnan, Guizhou, Finland, and Japan. *T. matsutake* from northeastern China is known as *Pinus densiflora*, whereas *T. matsutake* from Japan is mainly known as *Pinus thunbergii Parl.* and *Pinus densiflora*. However, *Quercus aquifolioides* and *Q. pannosa* are the major host plants for Sichuan *T. matsutake*. Notably, TM07 (KM102537) from Barkam (Sichuan, China) was distinct in the NJ trees. Analyses of both lineages indicated that the genetic relationship among samples KM102537, KM102538, and KM073272 was closer than that between other samples from Sichuan Province, which are in the second clade at the bottom of these two trees. Overall, individuals from Sichuan showed a high level of homogeneity. However, *T. matsutake* from Sichuan differed from samples from Heilongjiang, Yunnan, and Guizhou provinces of China, plus Finland and Japan. It was concluded that the homology among these *T. matsutake* samples was extremely high, despite some genetic variation.

Analysis of the Transcriptome Database

Based on the merits of transcriptome analyses, *T. matsutake* samples from Xiaojin were subjected to RNA-seq on the Illumina sequencing platform. The quality of the original read data from the samples was good, and the base quality exceeded 32 in all cases, which fulfilled the requirements for follow-up analysis (Fig. 6). A total of 24,549,990 reads were obtained by sequencing, and 18,266,492 high-quality and clean reads were selected after low-quality reads were excluded (Table 2). Trinity software generated clean data, and 32,098 contigs were obtained by assembling the reads to yield an N50 contig size of 812 bp (i.e., 50% of the genomic fragments were this length or longer; Table 3). There were 8,909 contigs (27.82%) with lengths between 201 and 300 bp. Moreover, 5,185, 3,677, 2,752, 2,114, 1,734, 1,362, 1,147, 1,625, 1,180, and 2,336 contigs had lengths of 301-400 bp, 401-500 bp, 501-600 bp, 601-700 bp, 701-800 bp, 801-900 bp, 901-1,000 bp, 1,001-1,200 bp, 1,201-1,400 bp, and greater than 1,400 bp, respectively (Table 4, Figs 7-8). Furthermore, 26,729 unigenes were obtained by assembling the scaffolds, and the proportions of A, T, C, and G were almost equal (Table 5).

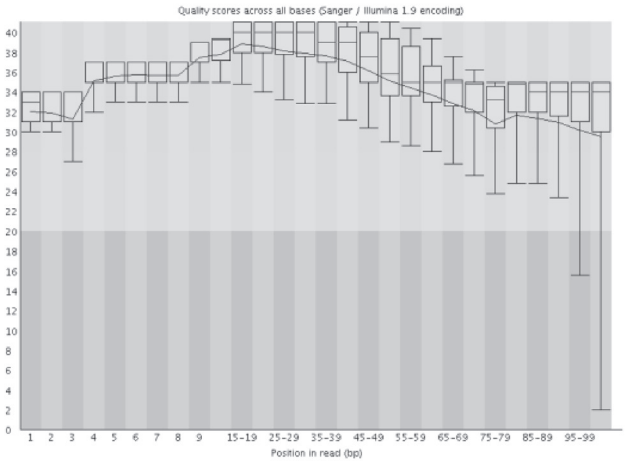

Fig. 6. Boxplot graph of sequence quality of Xiaojin *Tricholoma matsutake* genome.

Functional Annotation and Classification

Based on a comparison of the unigenes with the COG, EST, non-redundant (NR), Pfam, and UniProt databases, genome annotations and E-values were obtained. The results showed that there were 9,487, 6,856, 21,482, 18,537, and 12,715 matches in the COG, EST, NR, Pfam, and UniProt databases, respectively. The Venn diagram is shown in Fig. 9 and Table 6.

The identified unigenes with BLASTX hits in the NR database showed the highest homology to *Laccaria bicolor* S238N-H82, with a contig number of 6,559 (30.33%). Furthermore, 1,647, 1,540, and 1,018 contigs were obtained with *Coprinopsis cinerea okayama* 7#130, *Serpula lacrymans var. lacrymans* S7.9 and *Pseudomonas fluorescens* WH6, respectively (Table 7).

Moreover, the BLAST analysis against the NR database indicated that *T. matsutake* shared a high number of contigs with *Laccaria bicolor*. *L. bicolor* grows in mixed broadleaf conifer forests in Sichuan, the Tibet Autonomous Region of Southern China, and Yunnan Province. The similarity to the host of *T. matsutake* might be responsible for the high contig number.

GO and KEGG Analyses

The Gene Ontology (GO) Consortium categorizes genes based on molecular function, cellular component, and biological process.

All the unigenes in the transcriptome of *T. matsutake* were examined for GO functional prediction and classification. All the unigenes were broadly divided into 55 categories according to their function. The largest set

Table 3. Features of the Xiaojin *Tricholoma matsutake* genome.

Item	Contig length	Unique gene length
N75	458	418
N50	813	741
N25	1,334	1,248
Minimum	201	201
Maximum	5,879	5,879
Average	620	578
Count	32,098	26,729
Total length	1,992,214	15,469,839

Table 4. Length distribution of assembled contigs and unigenes of Xiaojin *Tricholoma matsutake* genome.

Nucleotide length (bp)	Contig number	Unique genes
75-100	0	0
101-200	0	0
201-300	8,909	8,183
301-400	5,185	4,577
401-500	3,677	3,139
501-600	2,752	2,237
601-700	2,114	1,652
701-800	1,734	1,354
801-900	1,362	1,034
901-1,000	1,147	852
1,001-1,200	1,625	1,191
1,201-1,400	1,180	841
1,401-1,600	714	524
1,601-1,800	500	346
1,801-2,000	322	233
2,001-2,500	486	301
2501-3000	195	104
>3,000	119	88

Table 2. RNA-seq statistical data of Xiaojin *Tricholoma matsutake.*

Sample	Raw Data		Valid Data		Valid
	Read	Length	Read	Length	Ratio
LCF-song	24,549,990	100	18,266,492	90.03972	74.41%

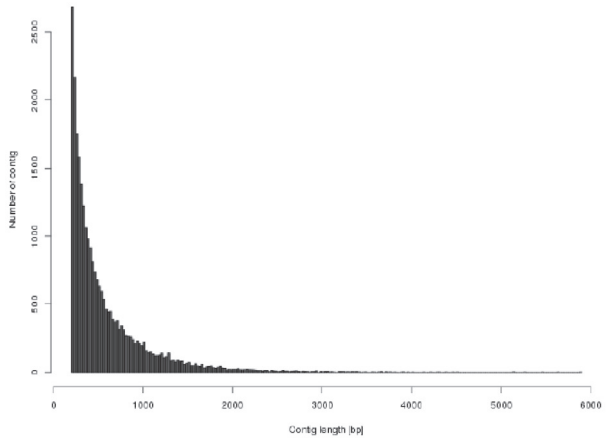

Fig. 7. Histogram of length distribution of assembled contigs derived from the Xiaojin *Tricholoma matsutake* genome.

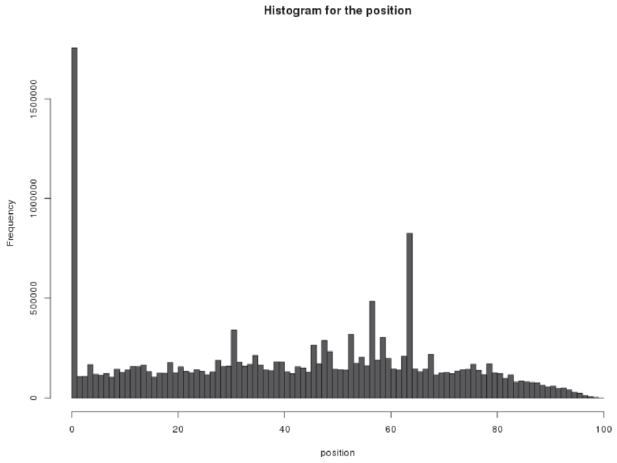

Fig. 8. Results of Kolmogorov-Smirnov test on genome of Xiaojin *Tricholoma matsutake.*

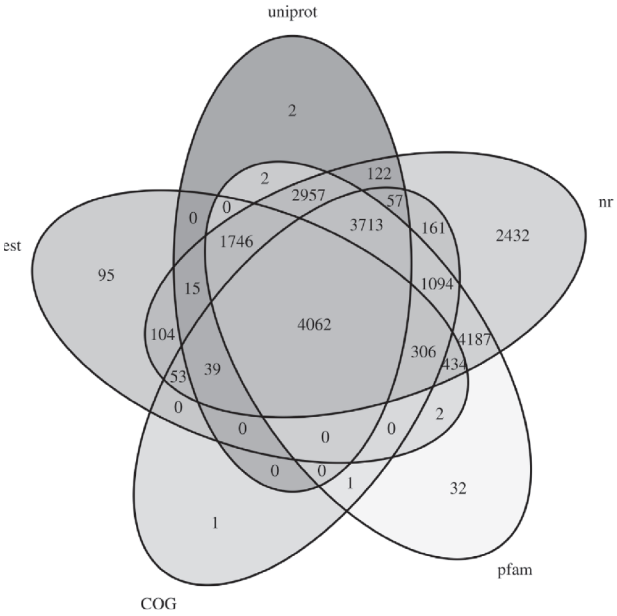

Fig. 9. Venn diagram of Trinity BLAST comparison of genome of Xiaojin *Tricholoma matsutake* with other databases.

Table 6. Trinity BLAST comparison of genome of Xiaojin *Tricholoma matsutake* with other databases.

Sample name	Database type	Match number
LCF-song	COG	9,487
LCF-song	EST	6,856
LCF-song	NR	21,482
LCF-song	Pfam	18,536
LCF-song	UniProt	12,715

was assigned to the cluster "catalytic activity" (22,304; 19.96% of total unigenes), followed by "metabolic process" (20,093; 17.98%), "binding" (17,897; 16.02%), "cell part" (7478; 6.69%), "organelle" (7031; 6.29%), "regulation of biological process" (3941; 3.53%), "membrane part" (3698; 3.31%), "cellular process" (3663; 3.28%), and "establishment of localization" (3537; 3.17%) (Table 8 and Fig. 10).

Table 5. Nucleotide data of Xiaojin *Tricholoma matsutake* genome.

Nucleotide	Count	Frequency
Adenine (A)	4,875,123	24.43%
Cytosine (C)	5,106,053	25.59%
Guanine (G)	5,101,894	25.57%
Thymine (T)	4,839,077	24.25%
Any nucleotide (N)	32,098	0.16%

Table 7. Species homology analysis of Xiaojin *Tricholoma matsutake* and other species.

Species	Contig number
Laccaria bicolor S238N-H82	6,559
Coprinopsis cinerea okayama 7#130	1,647
Serpula lacrymans var. lacrymans S7.9	1,540
Pseudomonas fluorescens WH6	1,018
Pseudomonas fluorescens SBW25	837
Schizophyllum commune H4-8	721
Pseudomonas extremaustralis 14-3 substr. 14-3b	718
Moniliophthora perniciosa FA553	536
Serpula lacrymans var. lacrymans S7.3	528
Fibroporia radiculosa	514
Other	6,855

Table 8. Gene ontology functional classification of Xiaojin *Tricholoma matsutake* genome (Count > 1,000).

Go_term	Count	Description
GO:0003824	22,304	catalytic activity
GO:0005215	1,820	transporter activity
GO:0005488	17,897	binding
GO:0008152	20,093	metabolic process
GO:0009987	3,663	cellular process
GO:0016020	2,934	membrane
GO:0032991	3,126	macromolecular complex
GO:0043226	7,031	organelle
GO:0044422	1,137	organelle part
GO:0044425	3,698	membrane part
GO:0044464	7,478	cell part
GO:0044699	2,628	single-organism process
GO:0050789	3,941	regulation of biological process
GO:0050896	2,385	response to stimulus
GO:0051234	3,537	establishment of localization
GO:0071840	2,096	cellular component organization or biogenesis

The active biological pathways in *T. matsutake* were mapped to the annotated sequences by reference to pathways in KEGG. The unigenes were divided into 190 pathways. The largest set was "metabolic pathways" (812; 16.75%), followed by "biosynthesis of secondary metabolites" (505; 10.42%), "microbial metabolism" (451; 9.30%), and "ABC transporters" (192; 3.96%). In our previous study, organic solvent extraction-distillation followed by GC-MS combined with quantification by peak area normalization was used to analyze volatile compounds in fresh fruiting bodies of *T. matsutake* collected from different geographical origins in southwestern China. Twenty-seven volatile compounds were identified from 66 species, including seven alcohols, three carbonyls, 11 acids and esters, three terpene hydrocarbons, and three miscellaneous components. The volatile compounds in the Xiaojin sample differed from those in *T. matsutake* collected from different geographical origins in southwestern China, and the Xiaojin sample had the best health effects and was delicious and nontoxic. Additionally, a novel heteropolysaccharide (TMP-A) was isolated from the fruiting bodies of *T. matsutake* from Xiaojin County. TMP-A has a backbone of 1,4-β-glucopyranose that branches at O-6 into an (1→3)-α-galactopyranose residue and terminates at an α-xylopyranose residue. The difference in the secondary metabolites from the Xiaojin sample may be due to differences in metabolic pathways.

Table 9. KEGG analysis of Xiaojin *Tricholoma matsutake* genome (Count > 100).

Pathway	Count	Description
map00970	125	Aminoacyl-tRNA biosynthesis
map01100	812	Metabolic pathways
map01110	505	Biosynthesis of secondary metabolites
map01120	451	Microbial metabolism in diverse environments
map02010	192	ABC transporters
map02020	143	Two-component system
map03010	126	Ribosome
map03018	125	RNA degradation
map03040	142	Spliceosome

Many pathways are involved in fungal growth, development, differentiation, and death. Three signaling pathways were identified in *T. matsutake*: the MAPK, PI3K, and Notch pathways. A total of 41 unigenes were identified. Moreover, 9 and 32 genes were involved in the MAPK signaling pathway (Map ID: 04010) and the yeast MAPK signaling pathway (Map ID: 04011), respectively. Three genes and one gene were involved in the PI3K-Akt (Map ID: 04151) and Notch signaling pathways (Map ID: 04330). Notably, two tumor-related pathways were identified in *T. matsutake*: p53 (Map ID: 04115) and Wnt (Map ID: 04310), plus xamples of alanine, aspartate, and glutamate metabolism (Map ID: 00250) (Table 9 and Fig. 11).

The analysis of the *T. matsutake* transcriptome provided information for further research to elucidate the physiology and medicinal value of this fungus.

Conclusion

This study analyzed the homology of *T. matsutake* based on ITS sequences, and the results were congruent with those of previous studies. In total, *T. matsutake* samples from Sichuan showed a high level of homology. However, the *T. matsutake* samples from Sichuan differed from samples from Heilongjiang, Yunnan, and Guizhou provinces of China, and Finland and Japan. *T. matsutake* from northeastern China is known as *Pinus densiflora*, whereas *T. matsutake* from Japan is mainly known as *Pinus thunbergii Parl.* and *Pinus densiflora*. However, *Quercus aquifolioides* and *Q. pannosa* are the major host plants for *T. matsutake* in Sichuan. Moreover, a BLAST analysis against the NR database indicated that *T. matsutake* shared a high number of contigs with *Laccaria bicolor*. *L. bicolor* grows in mixed broadleaf conifer forests in Sichuan, the Tibet Autonomous Region of Southern China, and Yunnan. The similarity in hosts between *T. matsutake*

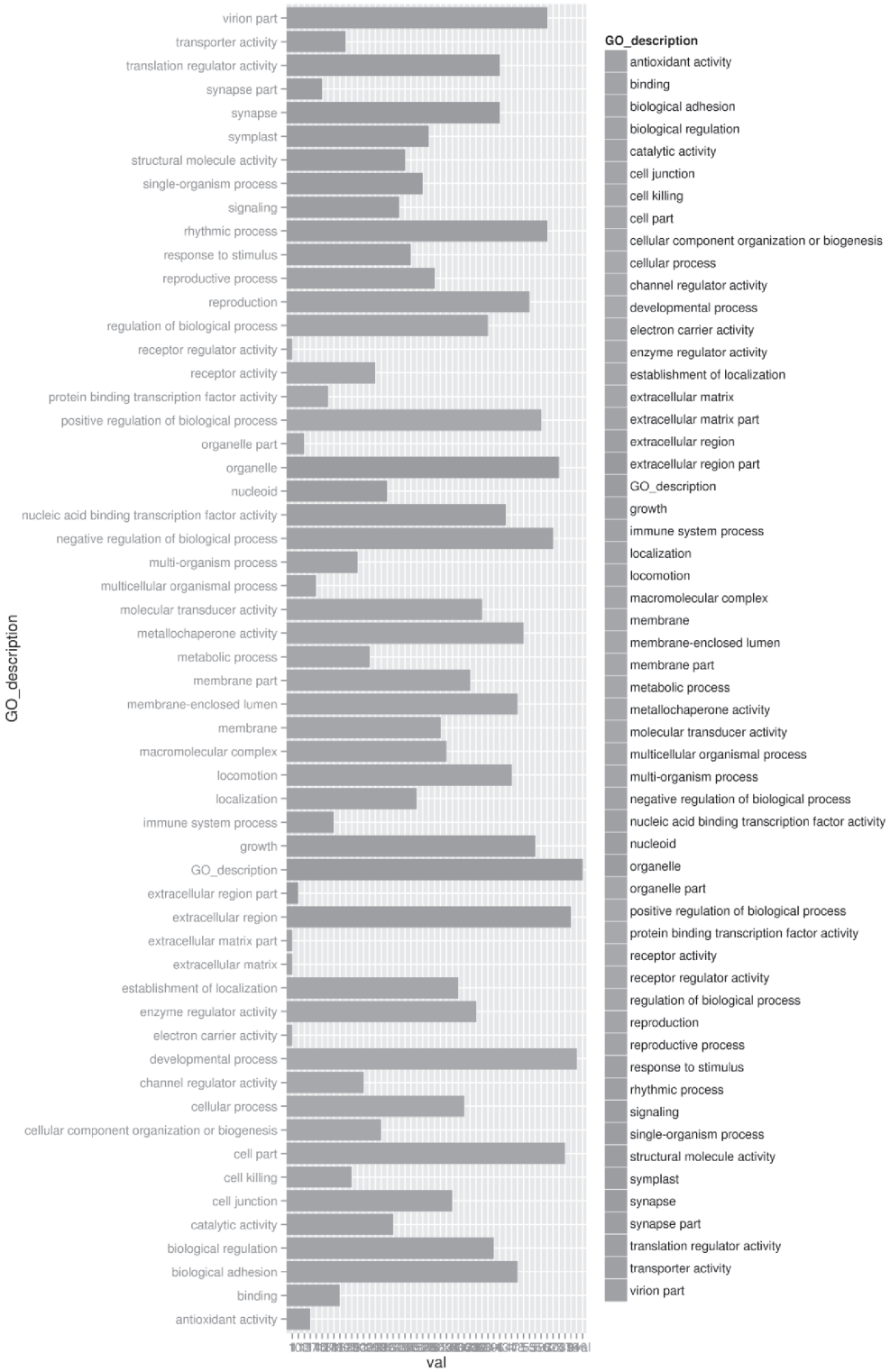

Fig. 10. GO analysis of Xiaojin *Tricholoma matsutake* genome.

and *L. bicolor* may be partly responsible for the high contig number. Notably, the influence of differences in host plants on genetic differentiation among different *T. matsutake* individuals remains to be fully explored. In the present study, we also used next-generation sequencing technology for the de novo assembly of the transcriptome of *T. matsutake* from Xiaojin. The transcriptome analysis of Xiaojin *T. matsutake* provided information for further research to elucidate the physiology and medicinal value of this fungus.

Abbreviations

RNA-seq: RNA sequencing
GO: Gene ontology

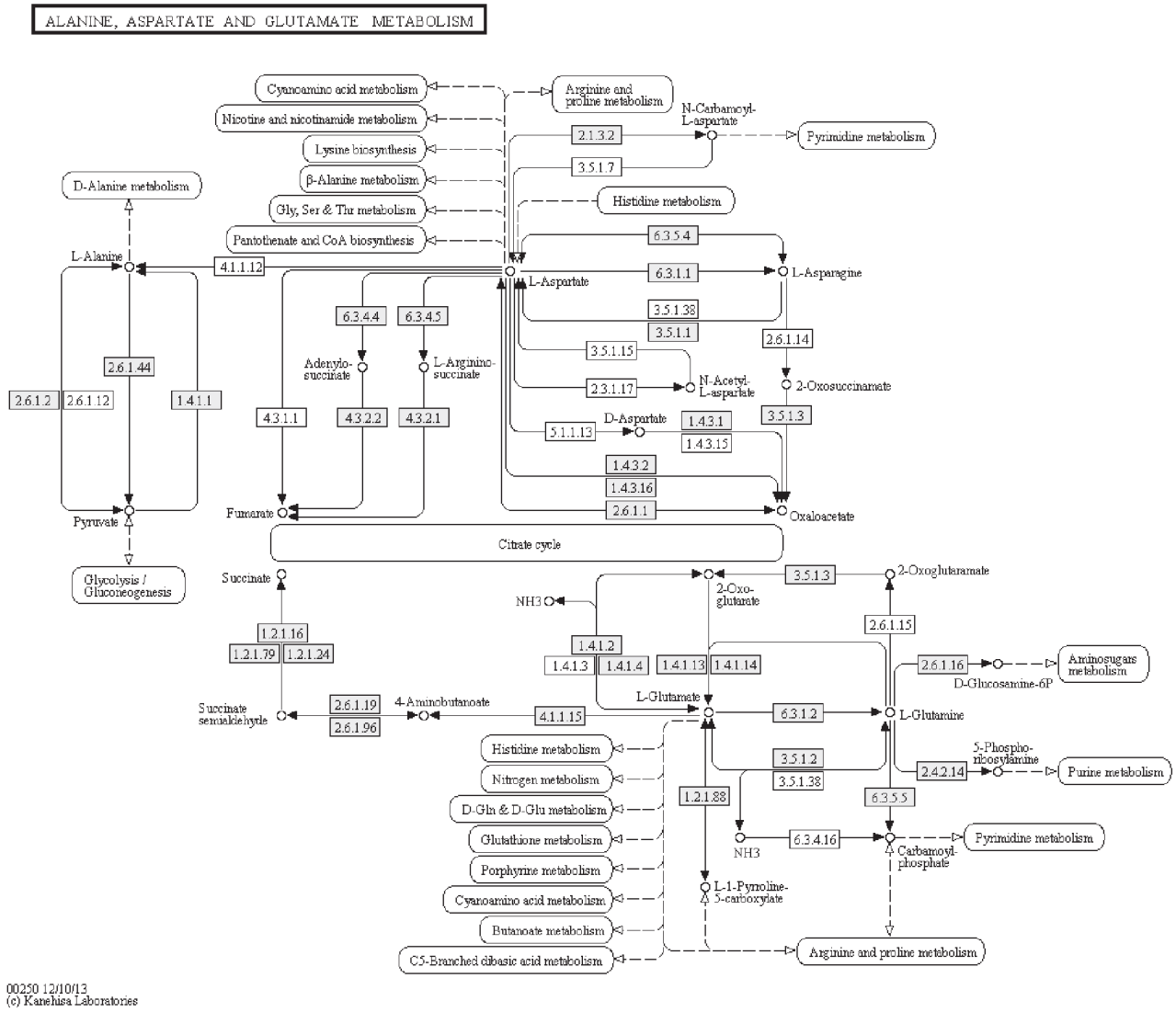

Fig. 11. KEGG analysis of Xiaojin *Tricholoma matsutake* genome. Core genes are colored gray; uncolored genes do not exist in the genome.

KEGG: Kyoto Encyclopedia of Genes and Genomes
BLAST: Basic local alignment search tool
DEG: Differentially expressed gene

Authors' Contributions

XD and YH designed the project. XD and MW performed the sample collection and RNA isolation. XD, YH, WH, YL, and PW performed the bioinformatics analysis. XD and YH drafted the manuscript, and YH edited it. All the authors read and approved the final manuscript.

Acknowledgements

This project was supported by the National Natural S cience Foundation of China (31400016 and 31200012), the Key Laboratory Fund Project of Ecological security and protection of Sichuan Province (ESP201307 and ESP1408), the Application Foundation Project of Sichua n Province (2013JY0094), the Science and Technology S upport Project of Sichuan Province (2014SZ0020 and 201 4FZ0024), the Cultivate Major Projects of Sichuan Provi nce (14CZ0016 and 16CZ0018), and the Open Foundation of Microbial Resources and Drug Development of Key Laboratory and of Guizhou Province (GZMRD-2014-00 2).

References

1. BAO D., KOIKE A., YAO F., YAMANAKA K., AIMI T., KITAMOTO Y. Analyses of the genetic diversity of matsutake isolates collected from different ecological environments in Asia. J. Wood. Sci. **53** (4), 344, **2007**.
2. BWEGIUS N., DANELL E. The Swedish matsutake (Tricholoma nauseosum syn. *T. matsutake*): distribution,

abundance and ecology. Scan. J. Forest. Res. **15** 3), 318, **2000**.
3. SHA T., ZHANG H., DING H., LI Z., CHENG L., ZHAO Z., ZHANG Y. Genetic diversity of *Tricholoma matsutake* in Yunnan Province. Chin. Sci. Bull. **52** (9), 1212, **2007**.
4. YANG X., HE J., LI C., MA J., YANG Y., XU J. Matsutake trade in Yunnan Province, China: an overview. Econ. Bot. **62** (3), 269. **2008**.
5. GILL W.M., GUERIN-LAGUETTE A., LAPEYRIE F., SUZUKI K. Matsutake – morphological evidence of ectomycorrhiza formation between *Tricholoma matsutake* and host roots in a pure *Pinus densiflora* forest stand. New. Phytol. **147** (147), 381, **2000**.
6. VAARIO L., HEINONSALO J., SPETZ P., PENNANEN T., HEINONEN J., TERVAHAUTA A., FRITZE H. The ectomycorrhizal fungus *Tricholoma matsutake* is a facultative saprotroph in vitro. Mycorrhiza. **22** (6), 409, **2012**.
7. KUSUDA M., NAGAI M., HUR T., UEDA M., TERASHITA T. Purification and some properties of α-amylase from an ectomycorrhizal fungus, *Tricholoma matsutake*. Mycoscience. **44** (4), 311, **2003**.
8. VISNOVSKY S.B., CUMMINGS N., GUERIN-LAGUETTE A., WANG Y., YAMADA A., KOBAYASHI H., KAWAI M., PITMAN A.R. Detection of the edible ectomycorrhizal fungus *Lyophyllum shimeji* colonising seedlings of cultivated conifer species in New Zealand. Mycorrhiza **24** (6), 453, **2014**.
9. KATAOKA R., SIDDIQUI Z. A., KIKUCHI J., ANDO M., SRIWATI R., NOZAKI A., FUTAI K. Detecting nonculturable bacteria in the active mycorrhizal zone of the pine mushroom *Tricholoma matsutake*. J. Microbiol. **50** (2), 199, **2012**.
10. XIANG D., YILING H., WANRU H. Structure feature and antitumor activity of a novel polysaccharide isolated from Lactarius deliciosus Gray, Carbohyd. Polym. **89** (2), 397, **2012**.
11. XIANG D., YILING H., YUANXIU Z., PANPAN W., LEI F., HONGQING Z., NAN Z., HANG Q., WEI Q., FANG W., WANRU H. Structure elucidation, anticancer and antioxidant activities of a novel polysaccharide from *Gomphus clavatus* Gray, Oncol. Rep. **33** (6), 3162, **2015**.
12. FANG W., YILING H., XIANG D., WANRU H., BO S., TING W., JIAN L., YICHUN Z. Structure elucidation and antioxidant effect of a novel polysaccharide from *Lactarius camphoratum* (Bull.) Fr., Int. J. Biol. Macromol. **62** (10), 131, **2013**.
13. XIANG D., YILING H., WANRU H. Structure elucidation and antioxidant activity of a novel polysaccharide isolated from *Boletus speciosus* Forst, Int. J. Biol. Macromol. **50** (3), 613, **2012**.
14. DING X., FENG S., CAO M., LI M., TANG J., GUO C., JIE Z., QUN S., ZHIRONG Y., JIAN Z. Structure characterization of polysaccharide isolated from the fruiting bodies of *Tricholoma matsutake*. Carbohydr. Polym. **81** (4), 942, **2010**.
15. GUERIN-LAGUETTE A., MATSUSHITA N., KIKUCHI K., IWASE K., LAPEYRIE F., SUZUKI K. Identification of a prevalent *Tricholoma matsutake* ribotype in Japan by rDNA IGS1 spacer characterization. Mycol. Res. **106** (2), 435, **2002**.
16. LIAN C., HOGETSU T., MATSUSHITA N., GUERIN-LAGUETTE A., SUZUKI K., YAMADA A. Development of microsatellite markers from an ectomycorrhizal fungus, *Tricholoma matsutake*, by an ISSR-suppression-PCR method. Mycorrhiza. **13** (1), 27, **2003**.
17. MURATA H., BABASAKI K., SAEGUSA T., TAKEMOTO K., YAMADA A., OHTA A. Traceability of Asian Matsutake, specialty mushrooms produced by the ectomycorrhizal basidiomycete *Tricholoma matsutake*, on the basis of retroelement-based DNA markers. Appl. Environ. Microbiol. **74** (7), 2023, **2008**.
18. XU J., CADORIN M., LIANG Y., YANG Z.L. DNA-based geographic typing of the gourmet mushroom *Tricholoma matsutake* traded in China. Mycoscience. **51** (3), 248, **2010**.
19. HWANG S., KIM J. Secondary structural and phylogenetic implications of nuclear large subunit ribosomal RNA in the ectomycorrhizal fungus *Tricholoma matsutake*. Curr. Microbio. **40** (4), 250, **2000**.
20. MATSUSHITA N., KIKUCHI K., SASAKI Y., GUERIN-LAGUETTE A., LAPEYRIE F., VAARIO L., SUZUKI K. Genetic relationship of *Tricholoma matsutake* and T.nauseosum from the Northern hemisphere based on analyses of ribosomal DNA spacer regions. Mycoscience. **46** (2), 90, **2005**.
21. CHAPELA I.H., GARBELOTTO M. Phylogeography and evolution in matsutake and close allies inferred by analyses of ITS sequences and AFLPs. Mycologia. **96** (4), 730, **2004**.
22. DUNSTAN W.A., DELL B., MALAJCZUK N., IWASE K. Detection of the ectomycorrhizal fungus *Tricholoma matsutake* and some related species with specific ITS primers. Mycoscience. **41** (1), 33, **2000**.
23. WU Q., MUELLER G.M., LUTZONI F.M., HUANG Y., GUO S. Phylogenetic and biogeographic relationships of eastern Asian and eastern North American disjunct Suillus species (fungi) as inferred from nuclear ribosomal RNA ITS sequences. Mol. Phylogenet. Evol. **17** (1), 37, **2000**.
24. KIKUCHI K., MATSUSHITA N., GUERIN-LAGUETTE A., OHTA A., SUZUKI K. Detection of *Tricholoma matsutake* by specific ITS primers. Mycol. Res. **104** (104), 1427, **2001**.
25. ZAJC J., LIU Y., DAI W., YANG Z., HU J., GOSTINČAR C. Genome and transcriptome sequencing of the halophilic fungus Wallemia ichthyophaga: haloadaptations present and absent. BMC Genomics. **14** (18), 617, **2013**.
26. LEE S.S., HONG S.W., CHUNG H.C., SUNG C.K., KIM J.H., KA K.H., KIM H.J. The specific probes confirming the genomic DNA of *Tricholoma matsutake* in Korea. Korean. J. Mycol. **27** (1), 20, **1999**.
27. YU M., YU J., GU C., NIE Y., CHEN Z., YIN X., LIU Y. De novo sequencing and transcriptome analysis of Ustilaginoidea virens by using Illumina paired-end sequencing and development of simple sequence repeat markers. Gene. **547** (2), 202, **2014**.
28. XIANG D., YILING H. Identification of Genetic characterization and volatile compounds of *Tricholoma matsutake* from different geographical origins, Biochem. Syst. Ecol. **44** (44), 233, **2012**.
29. GARDES M., BRUNS T.D. ITS primers with enhanced specificity for basidiomycetes – application to the identification of mycorrhizae and rusts. Mol. Ecol. **2** (2), 113, **1993**.
30. AMICUCCI A., ZAMBONELLI A., GIOMARO G., POTENZA L., STOCCHI V. Identification of ectomycorrhizal fungi of the genus tuber by species-specific ITS primers. Mol. Ecol. **7** (3), 273, **1998**.