De novo assembly and characterization of the transcriptome of a wild edible mushroom *Leucocalocybe mongolica* and identification of SSR markers

Tie Lu and Tolgor Bau

Engineering Research Center of Chinese Ministry of Education for Edible and Medicinal Fungi, Institute of Mycology, Jilin Agricultural University, Changchun, PR China

**ABSTRACT**

*Leucocalocybe mongolica* is a prestigious rare wild edible mushroom in Northeast Asia. It is the unique species of the genus and the studying of its phenotype and genome is crucial to genus and species classification. Beyond that, it has high application and ecological value in the industry of food and atmosphere investigation. On account of the fact that transcriptomic and genomic data of *L. mongolica* lacked in the biological information database, that is a limitation to further study. The transcriptome data were obtained by virtue of Illumina paired-end sequencing technology: 42,622,958 clean reads were achieved and 37,302 contigs were generated. These contigs were subsequently assembled into 13,821 unigenes. These unigenes were annotated within 7 public databases. The 3914 unigenes were associated with a COG classification. Throughout all of the unigenes, 6642 were classified as three functional groups; 3110 unigenes were selected from KEGG pathways, and taken to further clustering analysis for 5 main categories; 57 genes, potentially involved in terpenoid, steroid, and unsaturated fatty acids biosynthesis were identified and selected for further research. The total number of carbohydrate-active enzymes of *L. mongolica* is 446 and the number of carbohydrate binding module (CBM) is relatively low. CE11, GT19, GT51, GT56, GH131, GH133, GH135 constitute the characteristic carbohydrate-active enzymes subfamily compared to other edible mushrooms. The characteristic carbohydrate enzymes relative to other mushrooms could play a vital role on the metabolism of nutrients. In these generated sequences, 1860 SSRs were identified and characterized as molecular candidate markers existing *L. mongolica*.

**Introduction**

*Leucocalocybe mongolica* (S. Imai) X.D. Yu & Y.J. Yao is a prestigious rare wild edible mushroom in Northeast Asia. Its former Latin name was *Tricholoma mongolicum* S. Imai [1] and its Chinese name is ‘Bai-mo’ or ‘Kou-mo’. *T. mongolicum* was proposed to constitute a new genus, *Leucocalocybe*, based on morphological and molecular evidence [2], which was later verified by other experimenters [3]. *L. mongolica* is the unique member of a single-species genus. The species has an isolated taxonomical status, which is of great importance in scientific research from the perspective of biodiversity [4].

In a global context, it is an endemic species in some specific areas of temperate steppe in Northeast Asia. It is distributed in the grassland of Inner Mongolia of China, Mongolia, and the Far East of Russia. Due to its highly desirable flavour and short fruiting season, *L. mongolica* is collected by indigenes and mycophiles. Since ancient times, *L. mongolica* has been traded as a natural delicacy. Besides that, it has a potential for application in the treatment of a number of diseases, such as stomach disorders. Lectin in *L. mongolicais* is known to inhibit cancer and lower both blood pressure and cholesterol [5].

The mushroom is a native folk medicine for local minorities. Mongolians believe that it can be applied in the treatment of botulism and wounds, and for detoxification. Evenks boiled it with mutton to fortify their health and to relieve blood-heat syndrome [6]. The bioactivities of *L. mongolica* have attracted intense interest from scientists over recent years, as the basidiomata has been detected, such as anti-proliferative effect on human cervical carcinoma cells and hepatoma cells [7]. Some new biological activities have also been detected, such as anti-proliferative effect on human cervical carcinoma cells and hepatoma cells [12]. *L. mongolica* has a high antioxidant activity and can decrease the risk of atherosclerosis [13].
contains carbohydrate-binding proteins – lectins \cite{14,15} which possess a number of physiological activities in tests on mice, such as immunomodulatory, antitumour \cite{16–19}, antiproliferative \cite{20} and vasorelaxant activity \cite{21}. The main types of chemical constituents found in the medicinal fungi include terpenoids, steroids, alkaloids, phenolic constituents, sphingolipids, pigments and polysaccharides \cite{22}. However, for more comprehensive activity screening, the research has been restricted by the rareness of \textit{L. mongolica}, so the potential applications of the biologically active components of \textit{L. mongolica} remain to be exploited.

Nowadays, there is still a lack of a positive procedure for artificial cultivation of \textit{L. mongolica}. \textit{L. mongolica} is frequently collected, wild resources are in the state of predatory development and their habitats are vulnerable to destruction like grassland desertification, overgrazing and trampling. The mushroom is highly endangered. Therefore, more attention should be drawn to this unique member of the single-species genus \textit{Leucocalocybe}.

With the advancements of technology, the next-generation sequencing technologies have opened up exciting possibilities for whole genome sequencing of macro-fungi. Macro-fungi belonging to 64 families (seven species are incertae sedis) have been reported with their genome sequences in the biology informatics databases. The advances in Macro-fungal Genomes Sequencing stimulate the development of basic research, conservation and utilization of macro-fungi. On account of the fact that there are no transcriptomic data or genomic data of \textit{L. mongolica} in the biological information databases, it is necessary to obtain transcriptomic and genomic data. The further understanding of genomic information is indispensable for more far-reaching protection of the germplasm diversity. We gained a deeper understanding of the complex transcriptome involved in the biosynthesis of active substances and conducted a new plan for further research. We can identify candidate genes involved in the biosynthesis of key constituents.

Microsatellites (SSRs) as essential molecular markers have been used in forensics, population and conservation genetics phylogeography, genomic mapping and determining parentage \cite{23–25} due to their high polymorphism, rich informativeness, simplicity and spread through the genome. The key research topic concerning the conservation of \textit{L. mongolica} populations is the genetic diversity. In this study, SSR markers were sought on the basis of the transcriptome sequences to explore the population genetic diversity of \textit{L. mongolica}.

**Materials and methods**

**Strain isolation and identification**

The strain was originally isolated from basidiomata which was collected in Old Barag Banner, Inner Mongolia Autonomous Region, China (GPS coordinates: N49.34762449; E119.2343394) at an altitude of 625 m in 2013 (Online supplementary Figure 1S). The strain was maintained in Strains Laboratory of Fungi at the Engineering Research Center of the Chinese Ministry of Education for Edible and Medicinal Fungi in Jilin Agricultural University. The voucher specimen was preserved in Herbarium Mycology of Jilin Agriculture University (HMJAU) under number MCCJLAU2015YKS. It was preserved on potato dextrose agar (PDA) medium at 4 °C. The sequence (ITS1-5.8S-ITS2 of nuclear ribosomal DNA) was amplified from the mycelium (annealing Tm. is 50 °C) and 639 bp of the DNA fragment were sequenced (Sangon Co., Ltd., Shanghai, China). The sequence of the strain (GenBank accession number: KY070316) is identical to that of the voucher specimen.

**Fungal culture conditions**

The strain was transferred into Petri dishes by the roll-tube technique. The media components included soluble starch 2%, beef extract 0.2%, CaCO$_3$ 1%, agar 2% and distilled water. The medium was autoclaved at 121 °C for 30 min. When mycelia had grown all over the Petri dishes, 1 cm (diameter) homogeneous pieces were punched around concentric circles using a puncher. Five homogeneous pieces were transferred into 250 mL flasks with fermentation medium. They were propagated on a rotary shaker (150 rpm) at 25 °C. The formula of the fermentation medium contained 20% corn extract. Mycelium pellets were collected by the filtering method. After that, mycelium pellets were stored at −80 °C.

**RNA extraction, cDNA library construction and sequencing**

Total RNA was extracted from frozen mycelium pellets by using the Transzol plant (TransGen Biotech, Inc.) following the manufacturer’s instructions. The mycelium pellets samples from three fermentation periods (8 days, 14 days and 20 days) were mixed together to prepare total RNA. The RNA samples that met the requirements were used to construct transcriptome sequence libraries. The RNA samples were tested for concentration, purity and RNA integrity with NanoDrop 2000 microvolume spectrophotometer (Thermo Scientific, Waltham, MA, USA) and Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara,
CA, USA). The total RNA amount of 12.2 µg was equally pooled for cDNA library preparation. cDNA library construction and Illumina sequencing of the sample was conducted at Annoroad Gene Technology Co Ltd., Beijing, China. The specific method of cDNA library construction was done as described in [26]. IlluminaHiseq 2000 system with sequencing PE150 strategy was implemented with regard to RNA-sequencing.

**Transcriptome data processing and de novo assembly**

The image data originally generated by the sequenator were converted to raw data which were processed with Perl scripts to ensure the quality of data. The raw data were filtered by the following criteria: (1) removing the adaptor-polluted reads, (2) removing the low-quality reads which account for more than 15% of total bases, (3) removing reads with a number of N bases accounting for more than 5%. As for paired-end sequencing data, both reads would be filtered out if any read of the paired-end reads was adaptor-polluted. The obtained Clean Data after filtering were analysed statistically based on quantity and quality, including Q30 (those with a base quality value less than 30), data quantity and base content statistics, etc.

The reference genome of *L. mongolica* has not been reported in biology informatics databases. The software Trinity was used for de novo assembly [27]. Trinity represents a novel method for efficient and robust de novo reconstruction of transcriptomes from RNA-sequencing data.

**Gene annotation and analysis**

Trinotate version 20140717 (http://trinotate.github.io/) was used for performing the functional annotation of unigenes [28]. Trinotate has an integrated annotation procedure, especially for de novo assembled transcriptomes. Trinity was run for functional annotation by virtue of a series of different well referenced programs. Homology search in bioinformatics databases which consist of the National Center for Biotechnology Information (NCBI) non-redundant (NR) protein database, NCBI nucleotide sequence database (NT), Swissprot, Pfam (http://pfam.xfam.org/) and EggNOG (http://egg nog.embl.de/version_4.0.beta/) was implemented. BlastX (https://blast.ncbi.nlm.nih.gov/Blast.cgi) analysis was further carried out with an E-value cut-off of 1.0e-5 based on the results of the protein database annotation. On account of an unknown functional gene, the cluster of orthologous group (COG) method was utilized [29].

The Gene Ontology (GO) (http://geneontology.org/) classification was applied to classify the annotated sequences. The vital biological biosynthesis pathways of *L. mongolica* were identified by Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.kegg.jp/).

Annotation of carbohydrate-related enzymes was implemented by CAZy database (http://www.cazy.org/). The six CAZy modules for glycoside hydrolases (GH), glycosyltransferases (GT), polysaccharide lyases (PL), carbohydrate esterases (CE) and carbohydrate-binding modules (CBMs) were applied to classify the results in accordance with the CAZymes database [30].

**SSRs identification**

SSR in the unigenes was identified using MicroSAtellite software (MISA) (http://pgrc.ipk-gatersleben.de/misa/) based on the arguments about the threshold of mono-, di-, tri-, tetra-, penta- and hexanucleotide motifs with a minimum of 10, 6, 5, 5 and 5. SSR was identified by setting 100-bp flanking sequences upstream and downstream of the simple sequence repeat.

**Results and discussion**

**RNA sequencing and de novo assembly**

A total of 46,624,842 raw reads were generated for *L. mongolica* with a paired-end sequencing method by Illuminasequenator. Based on the filter of the low-quality reads, adapters, poly-A tails and reads containing more than 5% unknown nucleotides, the raw data were processed in order to satisfy the quality control. Approximately 42,622,958 clean reads were obtained. A general summary of the statistical data on sequencing and assembly is demonstrated in Table 1. The Q30 percentage and N percentage were 96.39% and 0.01%, respectively.

| Database                  | Number         |
|---------------------------|----------------|
| Total raw reads           | 46,624,842     |
| Total clean reads         | 42,622,958     |
| Raw bases number          | 6,993,726,300  |
| Clean bases number        | 6,393,443,700  |
| Q30 percentage            | 96.3900        |
| Ns reads rate (%)         | 0.0100         |
| Number of contigs         | 37,302         |
| Number of unigenes        | 13,821         |
| N50 of contig             | 3559           |
| N50 of unigenes           | 3580           |
| GC percentage             | 49.6200        |
| Max length of unigenes    | 43,076         |
| Min length of unigene     | 201            |

Note: Q30 percentage is the proportion of nucleotides with quality value larger than 30 in reads; Ns reads-rate is the proportion of unknown nucleotides in clean reads; GC percentage is the proportion of guanidine and cytosine nucleotides among the total nucleotides; N50 length is defined as the shortest sequence length at 50% of the genome.
respectively. The genome data of *L. mongolica* has not been reported; therefore, de novo assembly was chosen (Table 1). As a consequence, 37,302 contigs were generated, which were subsequently assembled into 13,821 unigenes. In addition, the GC percentage of these unigenes is 49.62%. The maximum length of unigenes is 43,076 and the minimum length is 201. In brief, all relevant data revealed a high quality and sequencing depth generated by Illuminasequenator.

**Functional annotation by searching against public databases**

The reference genome of *L. mongolica* was not present in biology informatics databases. The de novo transcriptome assembly was annotated by Trinity: 138,201 non-redundant unigenes, 72.65% had powerful homology with the aligned Nr protein database. Of all resoundingly annotated unigenes in public databases (Table 2). The maximum number of 9278 unigenes (67.13%) matches the Nr database. Followed by the SwissProt (6764 unigenes / 48.94%), GO (6642 unigenes / 48.06%), Pfam (6362 unigenes / 46.03%), EggNOG (4618 unigenes / 33.41%), KEGG (4019 unigenes / 29.08%), COG (3914 unigenes / 28.32%) and Nr (1790 unigenes / 12.95%).

A total of 9278 unigenes were annotated in the NCBI Nr protein database. Of all resoundingly annotated unigenes, 72.65% had powerful homology with the aligned proteins (E-value <1E-45) (Figure 1(A)). The similarity distribution showed that 54.39% of the unigenes had a similarity higher than 60% (Figure 1(B)). Distribution analysis by Blast based on the Nr protein database showed that 12.68% of the annotated unigenes have homologs in *Laccaria amethystina* (Figure 1(C)), indicating that the annotated *L. mongolica* unigenes were not highly homologous with species in this database.

The Venn diagram (Figure 2) illustrates the interrelation of Blast hits for unigenes against three databases. BlastX represents an optimal gene alignment in the Uniport database. BlastP represents an optimal open-reading frame (ORF) alignment in the Uniprot database. The results revealed 1373 unigenes were considered as orthologous genes across all these three databases, as shown in the Venn diagram.

### Table 2. Summary statistics of functional annotation of *Leucoagaricus mongolica* unigenes in public databases.

| Public protein database | Number of unigenes | Percentage (%) |
|-------------------------|--------------------|----------------|
| Nr                      | 9278               | 67.13%         |
| Nt                      | 1790               | 12.95%         |
| SwissProt               | 6764               | 48.94%         |
| KEGG                    | 4019               | 29.08%         |
| Pfam                    | 6362               | 46.03%         |
| EggNOG                  | 4618               | 33.41%         |
| COG                     | 3914               | 28.32%         |
| GO                      | 6642               | 48.06%         |
| Total unigenes          | 13,821             | 100%           |

**Functional classification by COG, GO and KEGG**

To further obtain better understanding of the functions of the unigenes, the BlastX method was implemented by comparing to the COG database (Figure 3). A total of 3,914 unigenes were associated with a COG categorization. With regard to the 25 COG categorizations, the R category (general function prediction only) represented the maximal group (439; 11.22%), followed by the E category (amino acid transport and metabolism), the C category (energy production and conversion), the G category (carbohydrate transport and metabolism) and the J category (translation, ribosomal structure and biogenesis).

GOntology classification was applied to classify the predicted functions of *L. mongolica* genes. Within the molecular function category, 6642 unigenes were distributed among 67 functional groups in Gene Ontology (Figure 4). Three dominant groups, and namely biological processes, cellular components and molecular functions, were shown to include 52,744,137 and 3097 unigenes, respectively. Under the category of biological processes, cellular processes (1236, 18.61%), metabolic processes (1209, 18.20%) and single-organism processes (1005, 15.13%) were conspicuous and considered to be involved in vital cell processes and metabolic activities of *L. mongolica*. Regarding the cellular components, there were two prominent categories: cell (1277 unigenes/19.23%) and organelle (689 unigenes/10.37%). In the categorization of molecular function, the first and second largest groups were catalytic activity (1283 unigenes / 19.32%) and binding (1174 unigenes/17.68%), respectively. The rest of the categories, such as transporter, electron carrier, molecular function regulator, nucleic acid binding, enzyme regulator, structural molecule, molecular transducer, antioxidant, protein-binding transcription factor, guanyl-nucleotide exchange factor and nutrient reservoir accounted for 638 unigenes, representing only 9.61%.

The annotation of the KEGG pathway database can promote the further understanding on the biological functions and interactions of genes. Out of the 4019 unigenes, 3110 had prominent matches in the KEGG database and were classified into five main groups, including 35 KEGG pathways (Figure 5). Relative to the five main categories, metabolism was the largest (1389, 44.66%), followed by genetic information (620, 19.94%), organismal systems (452, 14.53%), cellular processes (360, 11.58%) and environmental information processing (289, 9.29%). The results indicate that stirring metabolic processes were currently taking place.

From Figure 5, it can be seen that KEGG metabolism contained 13 categories, for instance, amino acid metabolism, biosynthesis of other secondary metabolites,
carbohydrate metabolism and energy metabolism. The KEGG database annotation was also conducive to revealing the essential biological pathways operating in *L. mongolica*.

Candidate genes involved in bioactivation in the three categories of significant biosynthesis

Mushrooms are usually applied as adaptogens and immunostimulants to balance and promote a healthy body. Most mushroom products possess beneficial health effects owing to the synergistic action of the present bioactive molecules and can be used on a regular basis without harm [31]. So far, terpenoids and steroids have been extracted from various mushroom species [32]. Terpenoids and steroids are important pharmacologically active ingredients. Like the Ganoderma-triterpenoids, peroxyergosterol is recognized worldwide. In addition, unsaturated fatty acids play a crucial role in healthy products that are becoming a hot spot of the wellness industry.

We investigated the genes that control the terpenoid, steroid and unsaturated fatty acids biosynthesis in *L. mongolica*. As shown in Table 3, 27 genes related to
terpenoid biosynthesis (15 involved in the terpenoid backbone, 12 involved in ubiquinone and others involved in terpenoid-quinone synthesis), 20 genes related to the steroid biosynthesis (9 involved in steroid hormone synthesis and 11 involved in steroid synthesis) and 10 genes related to unsaturated fatty acids biosynthesis in *L. mongolica* were identified. *L. mongolica* is also a folk medicine for local native minorities. It is originally attractive for its high nutritional and healthy function. Secondary metabolites have aroused the interest of more and more researchers.

**Carbohydrate-active enzyme (CAZy) database annotation**

The transcriptome of *L. mongolica* was annotated with the carbohydrate-active enzyme (CAZy) database that contributes to promote the deeper understanding of the correlation between nutritional modes and ecological environment. Carbohydrate-active enzymes play an important role in substrate degradation processes [33]. Mushrooms are important decomposers in the ecosystem returning nutrients to the soil by putrefaction and decay. We can obtain information on the distribution of the major classes of carbohydrate-active enzymes in *L. mongolica* from Figure 6. GH constitute a reasonable

---

**Figure 2.** Venn diagram of Blast hits for unigenes against three databases.

**Figure 3.** Histogram of clusters of COG classification.
amount of 44.17%, the following is auxiliary activities 26.46%, GT 14.13%, CE 10.09% and PL 3.14%; the least represented class is that of CBM 2.02%. *L. mongolica* possesses a wide spectrum of CAZymes responsible for the biosynthesis, degradation and modification of oligo- and polysaccharides and of glycoconjugates.

Nine other mushrooms were selected as references. Some model mushroom species possess representative ecological characteristics. For example, grass-rotting, wood-rotting and ECM fungi form mutualistic symbioses with many tree species. The three ecotypes are rather typical. Among the nine mushrooms, *Agaricus bisporus*, *Volvariella volvacea* and *Coprinus cinereus* are grass-rotting fungi, *Pleurotus ostreatus*, *Lentinus edodes*, *Trametes versicolor*, *Schizophyllum commune* and *Serpula lacrymans* are wood-rotting fungi and *Laccaria bicolor* is an ectomycorrhizal fungus. The above species have reference genomes. As shown in [Figure 7](#), the total number of CAZymes in *L. mongolica* is 446, more than those of *A. bisporus* and *L. bicolor*. The number of carbohydrate-binding modules was the lowest relative to other mushrooms. The distribution of 446 CAZyme genes in *L. mongolica* is presented in the Online supplementary Table 1S.

Referenced to more related literature [33–37] on carbohydrate-active enzymes of mushrooms, CE11, GT19, GT51, GT56, GH131, GH133, and GH135 are the most characteristic carbohydrate-active enzymes subfamily compared to other edible mushrooms.

**Identification and characterization of SSR markers**

Using SSRs, strains have been identified of some important basidiomycetes in commercial trade, for example *P. ostreatus*, *L. edodes*, *Phellinus linteus*, *Auricularia auricula-judae* and *Flammulina velutipes* [38–44]. SSRs are molecular markers that have a very important contribution in scientific research and various applications.

Out of the 13,821 unigenes that were examined, an aggregate number of 1860 SSRs were identified in *L. mongolica*. Statistical analysis of the identified SSRs is shown in Table 4. The quantities of sequences containing SSRs which is equal and greater than 1 were 1407 and 286, respectively. The numbers of mono-, di-, tri-, tetra-, penta- and hexarepeats were 1252, 200, 383, 14, 7 and 4, respectively. These SSRs are very informative in forensics, population and conservation genetics phylo geography, genomic mapping and determining parentage in *L. mongolica*. From Table 5 it can be seen that the number of potential SSRs with 10 repeats was 478, followed by 11 (274), 5 (237) and 6 (207). There were 147
potential SSRs with more than 15 repeats, and all these motifs were mono- or dinucleotide repeats.

The frequency distribution of the SSRs depending on the motif type is shown in Figure 8. The ACC/GGT trinucleotide repeat was the most abundant motif (comprising 16.12%), followed by AC/GT (11.35%), AT/TA (11.02%), AGC/CTG (9.70%), ACG/CGT (7.89%) and AG/CT (7.73%). The six types of SSRs referenced above delineated about 63.16%; nevertheless, the other 224 types of SSRs only aggregated to 36.84%.

Relative to whole-genome sequencing of Eukaryota, transcriptome sequencing is a faster and more low-cost approach to develop molecular SSR markers for genetic diversity, cultivar identification and breeding programmes. SSR markers have been used to trace pedigree relationships of cultivars [45].

Figure 5. Pathway assignment based on KEGG: cellular processes (A); environmental information processing (B); genetic information processing (C); metabolism (D); organism systems (E).
| Pathway                                      | Enzyme name                                                                 | Symbol/Gene name               | Number of annotated sequence transcripts |
|---------------------------------------------|------------------------------------------------------------------------------|--------------------------------|------------------------------------------|
| **Terpenoid backbone biosynthesis**         |                                                                              |                                |                                          |
|                                            | Farnesyl-diphosphate synthase                                                | FDPS, ISPA                     | 2                                        |
|                                            | Acetyl-CoA C-acetyltransferase                                                | E2.3.1.9, atoB                  | 4                                        |
|                                            | Diposphomevalonate decarboxylase                                             | MVD, mvaD                      | 1                                        |
|                                            | Hydroxymethylglutaryl-CoA synthase                                           | E2.3.3.10                      | 4                                        |
|                                            | 1-deoxy-D-xylulose-5-phosphate synthase                                      | DXS                            | 1                                        |
|                                            | Isopentenyl-diphosphate Delta-isomerase                                      | id, IDI                        | 1                                        |
|                                            | 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase                       | ispH, lytB                     | 1                                        |
|                                            | Protein farnesyltransferase subunit beta                                     | FNTB                           | 1                                        |
|                                            | Hexaprenyl-diphosphate synthase                                              | hexPS, COQ1                    | 2                                        |
|                                            | Protein farnesyltransferase/geranylgeranyltransferase type-1 subunit alpha    | STE24                          | 1                                        |
|                                            | STE24 endopeptidase                                                          |                                |                                          |
|                                            | Prenyl protein peptidase                                                     | RCE1, FACE2                    | 1                                        |
|                                            | Ditrans, polycis-polypropenyl diphosphate synthase                           | DHDDS, RER2, SRT1              | 1                                        |
|                                            | Prenylcysteine alpha-carboxyl methyltransferase                              | IMCL2                          | 1                                        |
|                                            | 2-polypropenyl-6-hydroxyphenyl methylase / 3-demethylubiquinone-9 3-methyltransferase | ubiG                        | 1                                        |
|                                            | Tyrosine aminotransferase                                                    | TAT                            | 1                                        |
|                                            | 4-coumarate–CoA ligase                                                       | menB                           | 1                                        |
|                                            | O-succinylbenzoic acid–CoA ligase                                            | menE                           | 6                                        |
|                                            | 2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylate synthase       | menD                           | 3                                        |
|                                            | 4-hydroxybenzoxoate polypropenyltransferase                                  | ubiA                           | 1                                        |
|                                            | Ubiquinone biosynthesis protein                                              | ubiB, aarF                     | 1                                        |
|                                            | 4-hydroxybenzoxoate polypropenyltransferase                                  | COQ2                           | 8                                        |
|                                            | Ubiquinone biosynthesis monoxygenase Coq6                                   | Coq6                           | 1                                        |
|                                            | Ubiquinone biosynthesis monoxygenase Coq7                                    | Coq7                           | 1                                        |
|                                            | 3-beta-hydroxy-Delta-5-steroid dehydrogenase/stereoid Delta-isomerase        | HSD3B                          | 1                                        |
|                                            | Catechol O-methyl-transferase                                                | COMT                           | 1                                        |
|                                            | Glucuronosyl transferase                                                     | UGT                            | 3                                        |
|                                            | Cytochrome P450, family 1, subfamily B, polypeptide 1                         | CYP1B1                         | 1                                        |
|                                            | Cytochrome P450, family 3, subfamily A                                       | CYP3A                          | 9                                        |
|                                            | 25/26-hydroxycholesterol 7-alpha-hydroxylase                                 | CYP7B                          | 1                                        |
|                                            | 17-beta-estradiol 17-dehydrogenase / very-long-chain 3-oxoacyl-CoA reductase | HSD17B12, KAR, IFA38           | 2                                        |
|                                            | 3-oxo-5-alpha-steroid 4-dehydrogenase 1                                      | SRD5A1                         | 1                                        |
|                                            | Corticosteroid 11-beta-dehydrogenase isozyme 1                              | HSD11B1                        | 1                                        |
|                                            | Delta14-sterol reductase                                                     | TM/5F2, ERG24                  | 2                                        |
|                                            | Delta24(24(1))-sterol reductase                                              | ERG4                           | 1                                        |
|                                            | Delta7-sterol 5-desaturase                                                    | SCSDL, ERG3                    | 2                                        |
|                                            | Sterol 24-(23)-methyltransferase                                             | E2.1.1.41, SMT1, ERG6          | 2                                        |
|                                            | Sterol O-acetyltransferase                                                    | E2.3.1.26, SOAT                 | 1                                        |
|                                            | Sterol 14-demethylase                                                        | CYP51                          | 1                                        |
|                                            | Sterol-Aldophenoxyblate 3-desaturated (decarboxylating)                      | E1.1.1.170, NSDHL, ERG26       | 4                                        |
|                                            | Methyldiol monoxygenases                                                     | E1.14.13.72, SC4MOL, ERG25     | 2                                        |
|                                            | 3-keto steriod reductase                                                      | ERG27                          | 1                                        |
|                                            | C-8 sterol isomerase                                                         | ERG2                           | 1                                        |
|                                            | Sterol 22-desaturase                                                         | ERG5, CYP61A                   | 1                                        |
|                                            | 3-oxoacyl-[acyl-carrier protein] reductase                                    | fabG                           | 4                                        |
| **Steroid biosynthesis**                    |                                                                              |                                |                                          |
| **Biosynthesis of unsaturated fatty acids** |                                                                              |                                |                                          |
|                                            | Acetyl-CoA oxidase                                                           | E1.3.3.6, ACOX1, ACOX3         | 2                                        |
|                                            | Stearoyl-CoA desaturase (Delta-9 desaturase)                                 | SCD, desC                      | 2                                        |
|                                            | 3-hydroxyacyl-CoA dehydrogenase/3-hydroxybutyryl-CoA epimerase/CoA isomerase | fadB                           | 1                                        |
|                                            | Acetyl-CoA acytransferase 1                                                  | ACAA1                          | 5                                        |
|                                            | Acetyl-CoA (8-3)-desaturase (Delta-5 desaturase)                              | FADS1                          | 2                                        |
|                                            | Acetyl-CoA 6-desaturase (Delta-6 desaturase)                                 | FADS2                          | 2                                        |
|                                            | Elongation of very long chain fatty acids protein 2                          | ELOVL2                         | 1                                        |
|                                            | 17beta-estradiol 17-dehydrogenase/very-long-chain 3-oxoacyl-CoA reductase    | HSD17B12, KAR, IFA38           | 2                                        |
|                                            | Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase                           | PHS1, PAS2                     | 2                                        |
markers to identify wild populations which were rich in diversity and private alleles and good candidates for conservation [46].

Conclusions

In this study, we comprehensively characterized the L. mongolica transcriptome by next-generation Illumina sequencing and de novo analysis without a reference genome. The results showed that Illumina paired-end sequencing by the strategy PE150 is feasible. Candidate genes potentially related to terpenoid, steroid and unsaturated fatty acids biosynthesis were identified. Seven characteristic carbohydrate-active enzymes subfamily compared to other edible mushroom were discovered. Furthermore, in all of the sequences, a total number of 1860 SSRs were identified as molecular markers in L. mongolica. The annotation information associated with the genomic sequences will expedite the further study of the physiology and ecology of this specie.

Acknowledgement

We would like to express our gratitude to professor Mei Ya and doctor Liang Guo in the Xilin Gol institute of bioengineering who helped us during the writing of the thesis.
Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
This project was sponsored by the Program for ‘Changjiang Scholars and Innovative Research Team in University’ of Ministry of Education of China [grant number IRT15R25].

References
[1] Imai S. On an edible Mongolian fungus "Pai-mo-ku.". Proc Imper Acad Tokyo. 1937;13(7):280–282.
[2] Yu XD, Deng H, Yao YJ. Leucocalocybe, a new genus for Tricholomamongolicum (Agaricales, Basidiomycota). Afr J Microbiol Res. 2011;5(31):5750–5756.
[3] Dong D, Bau T. A study on the taxonomic position of Tricholoma mongolicum. J Fungal Res. 2013;11(3):172–5175.
[4] Krishnamurthy KV. Textbook of biodiversity. Enfield (NH): Science Publishers, Inc.; 2003.
[5] Wu EQ, Bau T. Research advancement on Tricholoma mongolicum. Edible Fungi Chin. 2007;26:3–5.
[6] Wu N, Chun L, Khasbagan. [Ewenki folk medicinal plants and its comparison with Mongolian medicine]. Chin J Ethnomed Ethnopharm. 2009;17:156–158. Chinese.
[7] Bau S, Bao HY, Bau T, et al. Anti-tumor activity of Tricholoma mongolicum fruit bodies. Food Sci. 2012;33:280–284.
[8] Wang DW, Shan YL, Bau T. Effect of extraction rate of supercritical CO2 extraction on Mongolia mushroom polysaccharide. Food Sci. 2006;27:107–110.
[9] Hou Z, Zhang N, Wang DW. Study on technique of microwave extraction of Tricholoma mongolicum Imai polysaccharide. Food Sci. 2008;29:252–255.
[10] Wu XT, Xu RH, Ren QW, et al. Factors affecting extracellular and intracellular polysaccharide production in submerged cultivation of Tricholoma mongolicum. Afr J Microbiol Res. 2012;6:909–916.
[11] You QH, Yin XL, Zhang SN, et al. Extraction, purification, and antioxidant activities of polysaccharides from Tricholoma mongolicum Imai. Carbohydr Polym. 2014;99:1–10.
[12] Ge SM, Yu YH, Zhang YF. Study on the extraction and anti-tumor activity from Tricholoma mongolicum polysaccharide. Mod Prev Med. 2009;36:3708–3711.
[13] Bao L, Bai HP, Sa CF, et al. Antioxidant activity and vascular protecting effect of Tricholoma mongolicum polysaccharide. J Inner Mongolia Univ. 2014;45:498–501.
[14] Wang HX, Ng TB, Ooi VEC. Lectin activity in fruiting bodies of the edible mushroom Tricholoma mongolicum. Iubmb Life. 1998;44(1):135–141.
[15] Yao QZ, Wang F, Qu ZZ, et al. Study on the extraction technology of lectins from the mycelium produced by submerged fermentation of Tricholoma mongolicum. Acta Agr Boreali-Sinica. 2007;22:79–82.
[16] Wang HX, Liu WK, Ng TB, et al. The immunomodulatory and antitumor activities of lectins from the mushroom Tricholoma mongolicum. Immunopharmacology. 1996;31:205–211.
[17] Liu F, Ng TB, Wang HX, et al. Lectin from Tricholoma mongolicum S. Imai (Agaricomycetidae) mycelia stimulates gene expression of immunomodulating cytokines in mouse peritoneal macrophages and splenocytes. Int J Med Mushroom. 2005;7:243–248.
[18] Wang HX, Ooi VEC, Chang ST, et al. A polysaccharide–peptide complex from cultured mycelia of the mushroom Tricholoma mongolicum with immunoenhancing and antitumor activities. Biochem Cell Biol. 1996;74:95–100.
[19] Wang HX, Ng TB, Ooi VEC, et al. Actions of lectins from the mushroom Tricholoma mongolicum on macrophages, splenocytes and life-span in sarcoma-bearing mice. Anticancer Res. 1997;17:419–424.
[20] Wang HX, Ng TB, Liu WK, et al. Isolation and characterization of two distinct lectins with antiproliferative activity from the cultured mycelium of the edible mushroom Tricholoma mongolicum. Int J Pept Res Ther. 1995;46:508–513.
[21] Zhao YX, Wu XL, Huang SZ. Advance on chemical constituents and bioactivity of the medicinal fungi in China. Guizhou Sci. 2013;31(1):18–27.
[23] Ellegren H. Microsatellite mutations in the germline: implications for evolutionary inference. Trends Genet. 2000;16:551–558.

[24] Esselink GD, Nybom H, Vosman B. Assignment of allelic configuration in polyploids using the MAC-PR (microsatellite DNA allele counting–peak ratios) method. Theor Appl Genet. 2004;109:402–408.

[25] Kalia RK, Rai MK, Kalia S, et al. Microsatellite markers: an overview of the recent progress in plants. Euphytica. 2011;177:309–334.

[26] Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families. Science. 1997;278(5338):631–637.

[27] Grabherr MG, Haas BJ, Yassour M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644–652.

[28] Shu S, Chen B, Zhou M, et al. De novo sequencing and transcriptome analysis of *Wolfiporia cocos* to reveal genes related to biosynthesis of triterpenoids. PLoS One. 2013 [cited 2017 Feb 16];8(8):e71350 [10 p.]. DOI:10.1371/journal.pone.0071350

[29] Yan X, Zhang X, Lu M, et al. De novo sequencing analysis of the *Rosa roxburghii* fruit transcriptome reveals putative ascorbate biosynthetic genes and EST-SSR markers. Gene. 2015;561(1):54–62.

[30] Lombard V, Ramulu HG, Drula E, et al. The carbohydrate-active enzymes database (CAZy) in 2013. Nucl Acids Res. 2013;42(D1):D490–D495.

[31] Badalyan SM. Potential of mushroom bioactive molecules to develop healthcare biotech products, In: Singh M, editor. Proceedings of the 8th International Conference on Mushroom Biology and Mushroom Products (ICMBMP8); New Delhi: Yugantar Prakashan Pvt. Ltd., 2014 p. 373–378.

[32] Duru ME, Cayan GT. Biologically active terpenoids from mushroom origin: a review. Rec Nat Prod. 2015;9:456–483.

[33] Chen B, Gui F, Xie B, et al. Composition and expression of genes encoding carbohydrate-active enzymes in the straw-degrading mushroom *Volvariella volvacea*. PLoS One. 2013 [cited 2017 Feb 16];8(3):e58780 [9 p.]. DOI:10.1371/journal.pone.0058780

[34] Chen L, Gong Y, Cai YL, et al. Genome sequence of the edible cultivated mushroom *Lentinula edodes* (Shiitake) reveals insights into lignocellulose degradation. PLoS One. 2016 [cited 2017 Feb 16];11(8):e0160336 [20 p.]. DOI:10.1371/journal.pone.0160336

[35] Bing WU, Zhang XL, Cui BK, et al. Comparative genomic analysis of edible (medicinal) fungi reveals different ecological habitats. Mycosystema. 2015;34(4):742–760.

[36] Ohm RA, Jong JF, Lugones LG, et al. Genome sequence of the model mushroom *Schizophyllum commune*. Nat Biotechnol. 2010;28:957–963.

[37] Martin F, Aerts A, Ahren D, et al. The genome sequence of the basidiomycete fungus *Laccaria bicolor* provides insights into the mycorrhizal symbiosis. Nature. 2008;452:88–92.

[38] Xiao Y, Liu W, Dai Y, et al. Using SSR markers to evaluate the genetic diversity of *Lentinula edodes*’ natural germplasm in China. World J Microb Biot. 2010;26:527–536.

[39] Ma KH, Lee GA, Lee SY, et al. Development and characterization of new microsatellite markers for the oyster mushroom (*Pleurotus ostreatus*). J Microbiol Biotechnol. 2009;19:851–857.

[40] Xiang X, Li C, Li L, et al. Genetic diversity and population structure of Chinese *Lentinula edodes* revealed by InDel and SSR markers. Mycol Prog. 2016;15:1–13.

[41] Zhang R, Hu D, Zhang J, et al. Development and characterization of simple sequence repeat (SSR) markers for the mushroom *Flammulina velutipes*. J Biosci Bioeng. 2010;110:273–275.

[42] Zhang RY, Hu DD, Gu JG, et al. Genetic diversity and breeding history of winter mushroom (*Flammulina velutipes*) in China uncovered by genomic SSR markers. Genet 2016;597:227–235.

[43] Liu XB, Feng B, Li J, et al. Genetic diversity and breeding history of winter mushroom (*Flammulina velutipes*) in China uncovered by genomic SSR markers. Genet. 2016;591(1):227–235.

[44] Huang Y, Wu X, Jian D, et al. De novo transcriptome analysis of a medicinal fungi *Phellinus linteus* and identification of SSR markers. Biotechnol Biotechnol Equip. 2015;29(2):395–403.

[45] Tan LQ, Peng M, Xu LY, et al. Fingerprinting 128 Chinese clonal tea cultivars using SSR markers provides new insights into their pedigree relationships. Tree Genet Genome. 2015 [cited 2017 Feb 16];11(5):90. DOI:10.1007/s11295-015-0914-6

[46] Lusini I, Velichkov I, Pollegioni P, et al. Assessing the potential of mushroom bioactive molecules. In: Singh M, editor. Proceedings of the 8th International Conference on Mushroom Biology and Mushroom Products (ICMBMP8); New Delhi: Yugantar Prakashan Pvt. Ltd., 2014 p. 373–378.