Mycosphere 13(2): 86–133 (2022) www.mycosphere.org ISSN 2077 7019
Article – Special issue
Doi 10.5943/mycosphere/si/1f/4

Large-scale genome investigations reveal insights into domestication of cultivated mushrooms

Fu YP[1,#], Dai YT[1#], Chethana KWT[21], Li ZH[4], Sun L[1], Li CT[1], Yu HL[5], Yang RH[5], Tan Q[5], Bao DP[5], Deng YJ[6], Wang SX[7], Wang YF[8], Tian FH[9], Qi LL[10], Shu LL[11], Jia PS[1,l2], Chen LC[13], Chen MY[14], Hu QX[15], Tan H[16], Song TT[17], Zhang ZW[1,18], Bonito G[19], Zervakis GI[20], Xiao SJ[3*], Hyde KD[21*], Li Y[1*], and Yuan XH[2*]

[1]International Cooperation Research Center of China for New Germplasm Breeding of Edible Mushrooms, Jilin Agricultural University, Changchun 130118, China
[2]College of Plant Protection, Jilin Agricultural University, Changchun 130118, China
[3]Jiaxing Key Laboratory for New Germplasm Breeding of Economic Mycology, Jiaxing, Zhejiang 314000, China
[4]Shouxiangu Botanical Drug Institute Co., Ltd., Jinhua 321000, China
[5]Institute of Edible Fungi, Shanghai Academy of Agricultural Sciences, Shanghai 201403, China
[6]College of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, China
[7]Institute of Plant and Environment Protection, Beijing Academy of Agriculture and Forestry Sciences, Beijing Engineering Research Center for Edible Mushroom, Beijing 100097, China
[8]Institute of Edible Fungi, Mudanjiang Branch, Heilongjiang Academy of Agricultural Sciences, Mudanjiang 157041, China
[9]College of Agriculture, Guizhou University, Guiyang 550025, China
[10]Microbiology Research Institute, Guangxi Academy of Agricultural Sciences, Nanning 530007, China
[11]College of Horticulture, Shenyang Agricultural University, Shenyang 110866, China
[12]Institute of Plant Protection of Xinjiang Academy of Agricultural Sciences, Xinjiang Autonomous Region 830011, China
[13]Sylvan International Biotechnology Co., Ltd., Huai'an 223001, China
[14]Edible Fungi Institute of Fujian Academy of Agricultural Sciences, Fuzhou 350014, China
[15]Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China
[16]Sichuan Institute of Edible Fungi, Sichuan Academy of Agricultural Sciences, Chengdu 610000, China
[17]Institute of Horticulture, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China
[18]Department of Crop and Soil Sciences, Washington State University, Washington 99164, USA
[19]Department of Plant, Soil, and Microbial Sciences, Michigan State University, East Lansing 48824, USA
[20]Agricultural University of Athens, Laboratory of General and Agricultural Microbiology, Athens 11855, Greece
[21]Center of Excellence in Fungal Research, School of Science, Mae Fah Luang University, Chiang Rai 57100, Thailand

Fu YP, Dai YT, Chethana KWT, Li ZH, Sun L, Li CT, Yu HL, Yang RH, Tan Q, Bao DP, Deng YJ, Wang SX, Wang YF, Tian FH, Qi LL, Shu LL, Jia PS, Chen LC, Chen MY, Hu QX, Tan H, Song TT, Zhang ZW, Bonito G, Zervakis GI, Xiao SJ, Hyde KD, Li Y, Yuan XH 2022 – Large-scale genome investigations reveal insights into domestication of cultivated mushrooms. Mycosphere 13(2), 86–133, Doi 10.5943/mycosphere/si/1f/4

Abstract

Domestication and cultivation of edible and medicinal mushrooms are ongoing ~1,400-year-old evolutionary experiments. There are roughly 55 commercially cultivated mushrooms

Submitted 22 Apr 2022, Accepted 19 Aug 2022, Published online 21 Sep 2022, Published 29 Dec 2022
# These two authors contributed equally to this work and shared the first author
Corresponding Author: Shi-Jun Xiao – e-mail – shijun_xiao@163.com,
Kevin D. Hyde – e-mail – kdhyde3@gmail.com,
Yu Li – e-mail –yuli966@126.com,
Xiao-Hui Yuan – e-mail –yuanxiaohui@whut.edu.cn

worldwide; however, genomes for many species are lacking, as is our understanding of mushroom domestication. In this work, 22 high-quality reference genomes, along with 63 transcriptome and 381 re-sequencing data, were reported for the first time. Combined with public genome resources, an integrated omics database (MushDB, http://mushroomomics.com) were constructed including 50 reference genomes, 265 transcriptome and 621 whole-genome re-sequencing data, covering ~90% of the worldwide commercially cultivated mushroom species. Using multi-omics data in MushDB, whole-genome variations of the representative wild and cultivated populations were used for selective sweep analysis to identify putative functional genes contributing to the mushroom domestication. Key genes in the starch and sucrose metabolism and mitogen-activated protein kinase signaling pathway, such as *chb1*, *cdc24* and *hog1*, were putatively selected in *Auricularia cornea*, *Lentinula edodes*, *Pleurotus eryngii* and *Pleurotus tuoliensis*, indicating those genes might play important roles during mushroom domestications. The function of *hog1* in the low temperature stimulation adaptation and short growth period in cultivation environment, which could facilitate the successful domestication, was validated using CRISPR/Cas9 system. Our work offers valuable and abundant omics open resource for wide research communities and lays solid foundations for multi-purposes in mushroom evolution, genetics, and breeding studies.

Keywords – database – domestication – genome – multi-omics – mushroom

Introduction

Mushrooms play an important role in the forests eco-system and represent a major economic source worldwide. Mushrooms are widely used for food, medicine, industry, and ornamentals (Phan & Sabaratnam 2012, Wu et al. 2019). Wild mushrooms were utilized as far back as at least 6,800 years ago (Li 2018). As the first writing record of mushroom cultivation, the genus *Auricularia* (Wood ear) was domesticated and cultivated in China around 1,400 years ago (Wu et al. 2015). Until now, there are roughly 55 kinds of edible and medical mushrooms that have been commercially cultivated worldwide (Miles & Chang 2004), while production has increased 700-fold since 1978 and annual production is worth more than 3,000 million yuan in China (Li 2018). Despite their essential contribution to maintain the ecological biodiversity in many ecosystems and growing commercial importance, most work of cultivated mushrooms to date has focused on the taxonomy, phylogeny, ecology, development, cultivation, and morphological evolution (Hibbett & Binder 2002, Hibbett 2004, Varga et al. 2019). Genome resource and genomic investigations for mushrooms still lag far behind the research on plants and animals. For example, many key functional genes contributing to yield, metabolite synthesis and disease/environment stress tolerance for plants and animals have been identified with large-scale genome studies and their functions have been validated to illuminate the mechanism of domestications (Liu et al. 2020, Pan et al. 2021, Qin et al. 2021). However, functional genes underlying the domestication of mushrooms remain largely unknown (Sun et al. 2019).

With advances in sequencing technologies, the application of omics data has deeply expanded our understanding for the development and adaptation of mushrooms. Multi-gene sequences and 104 genomic data were combined to inferred ages and broad patterns of speciation and extinction in mushrooms (Varga et al. 2019). The evolution and multicellular development of *Armillaria* was investigated based on the comparative genomics analyses (Sipos et al. 2018). Meanwhile, previous studies have shown that improved cultivars of edible and medicinal mushrooms have become highly adapted to cultivation substrates and environments, and characterized by higher yield, favorable precocity, and better fruiting body quality than those of wild strains (Zhang et al. 2021, Yu et al. 2022). Therefore, the knowledge of key genes contributing to the important agriculture traits would widen our understanding of the mushroom domestication, and could benefit the artificial breeding of new strains. Mushroom omics data are essential resource for domestication studies. However, the lack of genomic resource, including reference genome, whole-genome re-sequencing and transcriptome data has restricted the identification of key functional genes responsible for the successful cultivation of edible and medicinal mushrooms (Sipos et al. 2018).

In the last decade, many mushroom genomes have been reported and available in public database; however, many genomes of cultivated edible and medicinal mushrooms are highly fragmented (contig N50 < 0.5 Mb and contig number over 2000) (Yap et al. 2014, Kurata et al. 2016). The sequence fragmentation spoiled the continuity and completeness of the genomes (Floudas et al. 2012), and could lead to inaccurate and deficient gene annotations, obstructing the functional gene identification and genomic comparison among species (Denton et al. 2014, Rhie et al. 2021). More importantly, the genome and transcriptome data for edible and medicinal mushrooms are dispersed into several databases, such as the National Centre for Biotechnology Information (NCBI) and Joint Genome Institute (JGI). The majority of databases merely provided sequencing raw data without data processing, hindering a ready-to-use of multi-omic data and direct comparison cross databases. Therefore, it has become essential to develop an integrated database of the genome, re-sequencing, and transcriptome data for cultivated mushrooms, facilitating the genomic investigation for domestication and molecular breeding of those important species.

In this study, we generated 22 new high-quality genome sequences, 381 re-sequencing and 63 transcriptome data for the cultivated mushrooms. An integrated mushroom database (MushDB, http://mushroomomics.com) was constructed to cover ~90% of the world-wide commercially cultivated mushroom species, including 50 high-quality genomes, 621 whole genome resequencing and 265 transcriptome data. Based on those multi-omics data in MushDB, we performed genome-wide selective sweep analysis between the wild and cultivated populations, and identified putatively key functional genes likely contributing to the domestication and cultivation for different kinds of mushrooms. Our work offers valuable and comprehensive omics open resource for the functional genomic investigation and breeding research of cultivated mushrooms.

Materials & Methods

Fungal genome sequencing and *de novo* assembly

All monokaryons using for *de novo* genome sequencing were isolated and cultured from mushroom dikaryotic strains as described previously (Table 1) (Dai et al. 2017). Genomic DNA was extracted from mycelia using NuClean Plant Genomic DNA Kits (CWBIO, Beijing, China). DNA qualities, purities, and concentrations were then evaluated using agarose gel electrophoresis, NanoDrop spectrophotometry (Thermo Fisher Scientific, MA, USA), and Qubit fluorometry (Life Technologies, CA, USA), respectively.

The 10-20 kb SMRTbell DNA long-read libraries were constructed and sequenced following the manufacturer's protocols on a PacBio RSII and Sequel platforms at Washington state university (Pullman, WA, USA), Tianjin biochip corporation (Tianjin, China), and the Engineering Research Center of the Chinese Ministry of Education for Edible and Medicinal Fungi (CMEEMF, Changchun, China). Genomic DNA short reads were also sequenced using Illumina HiSeq sequence platform at Novogene Co., LTD (Tianjin, China).

Genomes of all monokaryon strains were assembled with SMARTdenovo (https://github.com/ruanjue/smartdenovo) using long reads generated from the PacBio platform. Then, the long and short reads produced by PacBio and HiSeq platform of all monokaryon strains were used for two-rounds of error correction by Arrow (https://github.com/ruanjue/smartdenovo) and Pilon (Walker et al. 2014), respectively. Finally, the completeness of assembled genomes were evaluated using Core Eukaryotic Genes Mapping Approach (CEGMA) (Parra et al. 2007) and Benchmarking Universal Single-Copy Orthologs (BUSCO) (Simão et al. 2015) with the database of fungi-odb10.

Whole-genome re-sequencing and transcriptome sequencing for representative cultivated mushrooms

All samples of different developmental stages of mushrooms (mycelium, primordium, and fruiting body) were provided by the mushroom base of CMEEMF, Changchun, China. RNA

extraction and quality control were performed following the processes in our previous studies (Fu et al. 2016). The 150 bp paired-end cDNA libraries were constructed and then sequenced using an Illumina HiSeq platform at Novogene, Tianjin, China.

All strains used for whole-genome resequencing were provide by CMEEMF, Changchun, China. Strains were cultured and genomic DNA were extracted from cultured mycelia and quality controlled as above method for the PacBio long-read sequencing. Meanwhile, genomic DNA were also used to construct 350 bp insert length libraries and were sequenced with 150 bp paired-end mode using an Illumina HiSeq platform.

Gene prediction and functional annotation in genome sequences

Repeat element, functional genes, and non-coding RNAs were predicted using the identical method in our publication (Sun et al. 2019, Jiang et al. 2021). Firstly, homology-based and *de novo*-based methods were used to identify the Repeat element. Repbase (version 2017-01-27) was used to build the homology repeat library (Jurka et al. 2005). Then, RepeatMasker v4.1.0 (Tempel 2012) was used to search the *de novo* repeat library that were built using RepeatModeler v2.0.1 (Jurka et al. 2005). Secondly, *de novo* and homologous predictions strategies were used to annotation the genomes. For homologous prediction, *Agaricus bisporus*, *Coprinopsis cinerea*, *Pleurotus ostreatus*, *Schizophyllum commune*, *Ganoderma lucidum* and *Ganoderma boninense* were used as reference species. For *de novo* prediction, software Augustus, Genescan, GlimmerHMM, and SNAP were used to predict functional genes. Finally, noncoding RNA genes were annotated by tRNAscan-SE (version 1.3.1) (Lowe & Eddy 1997) for tRNA, BLASTN (version 2.10.1) (Altschul et al. 1990) for rRNA and INFERNAL (version 1.1.2) (Nawrocki et al. 2009) for miRNA and snRNA. The gene function were annotated by homolog searching with an E-value cut-off of 1×10^{-6} against public databases, including NCBI non-redundant (nr) sequence database (Pruitt et al. 2007), SwissProt database (Schneider et al. 2009), PFAM protein family database (Finn et al. 2014), InterPro database (Quevillon et al. 2005), and the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa & Goto 2000). The carbohydrate-active enzymes (CAZyme) in the newly sequenced and downloaded genomes (Table 1) were identified by mapping annotated protein sequences to the CAZy database (http://www.cazy.org/) using BLASTP (cut-off E-value $\leq 1e^{-5}$, identity>=40% and coverage>=40%). Secondary metabolite gene clusters were predicted with fungal AntiSMASH 3.0 (https://fungismash.secondarymetabolites.org/) (Weber et al. 2015) with the default parameters.

Standard pipeline for gene expression analysis and whole-genome variation detection using transcriptome and re-sequencing data

For new and public transcriptome data, quality control of original reads of each library was mainly processed using software fastp (version 0.20.0) (Chen et al. 2018) and FastQC (v0.11.8, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). First, sequencing adapters and low-quality reads were clipped using fastp (parameter: -q 15 -u 40 -n 10 -l 75) software. Then software FastQC was used to assess the data quality with default parameters. Cleaned new and public transcriptome data were mapped to corresponding reference genomes of cultivated mushroom using HISAT2 (version 2.1.0) (Kim et al. 2015). The mapping results were further quantified by FPKM (Fragments Per Kilobase Million) values using Stringtie (version 1.3.3b) (Anders et al. 2015).

New and public whole-genome resequencing data were cleaned by the identical method. Cleaned reads were used to detect whole-genome variations employing the pipeline of Genome Analysis Toolkit (GATK) (Poplin et al. 2017) v4.1.8.1 with default parameters. SNPs and INDELs were filtered by suggested parameters in GTAK hard-filtering Tutorials from GATK team: QUAL (phred quality) > 30, QD (quality score divided by depth to comprehensively evaluate the quality and depth) > 2, DP (read depth) > 5 and MQ (mapping quality to evaluate read alignment) > 40.

MushDB database construction for cultivated mushroom

MushDB (http://mushroomomics.com) was developed using JSP (Java Server Pages) language and open source software MySQL (Widenius et al. 2002) plus MongoDB as its database management system. MySQL was used major database for system management and MongoDB was applied to maintain the large dataset, including gene sequence, expression profiles and whole-genome variation data. The database adopted the MVC design pattern based on Spring (Walls & Breidenbach 2005), SpringMVC (Mak 2008), and MyBatis (SSM) (Ho 2012), which has the advantage of low developmental cost and high developmental efficiency. Web front-end was developed with the AJAX (Woychowsky 2006) technology based on jQuery (Bibeault & Katz 2008) front-end framework, which can reduce page refresh time and enhance user experience. The data visualization, such as bar, heatmap and pot plot for gene expression and variation data, was implemented by using open source Highcharts package (https://github.com/highcharts/highcharts). The BLAST (Camacho et al. 2009) component and Primer 3 software (Rozen & Skaletsky 2000) were integrated in the homepage of MushDB. For each species, taxon was checked according to Index Fungorum (http://www.indexfungorum.org) and brief introduction was citated from Wikipedia and a book "The book of Fungi" (Roberts & Evans 2014). The Mozilla Firefox and Google Chrome with JavaScript options were recommended to access the website.

Phylogenetics analysis using whole genome information

The phylogenetic analysis of 84 species was performed using the predicted protein sequences. These species included Agaricales (45), Polyporales (13), Hypocrealesares (7), Russulales (4), Auriculariales (4), Pezizales (3), Tremellales (2), and Boletales (2), respectively. *Phycomyces blakesleeanus* (PBL) in Mucorales was selected as the outgroup.

The "all against all" BLASTP searches were performed with a cutoff E-value of 1×10^{-5} for proteins from all species. Gene families were clustered using OrthoMCL software (version 2.0.9) (Li et al. 2003). The single copy genes were aligned using software mafft (version 7.490), then software protest (version 3.4.2) was used to find the best model of amino acid replacement of the single copy gene alignment. The phylogenetic tree was constructed using Randomized Accelerated Maximum Likelihood (RAxML) software (version 8.2.12) (Stamatakis 2014) with the maximum likelihood (ML) algorithm (parameters: -f a -m PROTGAMMAIJTTF -N 1000) with bootstrap parameter of 1000. The divergence time among species was estimated using the MCMCtree module in Phylogenetic Analysis by Maximum Likelihood (PAML) (Yang 2007) with the calibration times according to Floudas et al. (2012), and Zhao et al. (2017). Based on the calibration of divergence times from TimeTree (http://timetree.org/), the following constraints were used for time calibrations as input parameters in PAML: (i) the divergence time of *Phycomyces blakesleeanus* and *Scheffersomyces stipitis* was 605.2-1049.0 Mya; (ii) the divergence time of *Dacryopinax primogenitus* and *Auricularia subglabra* was 265.2-371.5 Mya; (iii) the divergence time of *Pl. populinus* and *Amanita muscaria* was 18.8-217.9 Mya.

Gene families inferred from OrthoMCL (Li et al. 2003) and the phylogenetic tree constructed previously were used to calculate the expansion or contraction of the gene families in each lineage using CAFE (Version: v.4.2.1, parameters: -p 0.05 -filter) (De Bie et al. 2006), which accounted for phylogenetic history and provided a statistical basis for evolutionary inference. *P* values of 0.05 were used to estimate the likelihood of the observed sizes given average rates of gain and loss and to determine expansion or contraction for individual gene families for each node.

Identification of key genes contributing to the domestication of cultivated mushrooms

The population whole-genome resequencing data for four representative cultivated mushroom species were analyzed in this work, including *Auricularia cornea* (AC): 6 cultivated and 10 wild strains, *Lentinus edodes* (LE): 12 cultivated and 8 wild strains, *Pleurotus eryngii* (PE): 15 cultivated and 5 wild strains, and *Pleurotus tuoliensis* (PT): 9 cultivated and 45 wild strains. Whole-genome SNPs were filtered and the ML tree were constructed using SNPhylo pipeline (Lam et al. 2010). The principal component analysis (PCA) was performed using the smartpca program

in Eigensoft (https://github.com/DReichLab/EIG), and the population genetic structure was analyzed using Structure with 10,000 iterations per run.

Using allele frequency information for whole-genome variations, the F_{st} value of the wild and cultivated populations of each species was calculated using PoPoolation2 software (Kofler et al. 2011), which was then used to estimate the selective fragments in these two populations. According to the criterion to identify selected genomic regions and genes in previous studies (Akey et al. 2002), we identified genomic regions with the highest 5% F_{st} values to be significantly selected, and genes resided in these genomic regions were putatively selected functional genes. Following the method for selection sweep analysis validation in previous literature (Li et al. 2019), we employed SweeD (version 3.1) (Pavlidis et al. 2007) to confirm the reliability of selected regions. Finally, we found that most of the selective regions identified with F_{st} method (82% for AC, 84% for LE, 81% for PE and 86% for PT) also exhibited selective signals in SweeD. The overlap of F_{st} method and SweeD was comparable with previous study (Li et al. 2019), illuminating the reliability of identified selected regions for mushrooms in this work.

Functional validation of *pthog1* gene in *Pl. tuoliensis* using CRISPR/Cas9 system

The genomic DNA and RNA of PT were extracted as above and the gene *pthog1* was amplified with primers 5'-ATGTCGTTTGTCAAGCTC-3' and 5'-TTACTCGGATTTCGGTTC-3'. Southern blot was carried out using DIG High Prime DNA Labeling and Detection Starter Kit I (F. Hoffmann-La Roche Ltd., Mannheim, Germany). The recombination of plasmid pCAMBIA1302-*pthog* was constructed and fused with the reporter gene *gfp*, which was then transformed into onion epidermal cells to perform subcellular localization by observing green fluorescence under fluorescence microscope.

CRISPR/Cas9 system was used to generate *pthog1* mutations for PT. The target RNA sequence (5'-AATCCAAGATCCCCAAATGA-3') was designed using the CRISPR-P program (Lei et al. 2014). The revised plasmid pP1C.8 with a hygromycin resistance cassette was used to construct genome editing plasmid pP1C.8-△*pthog1*. The transformation of pP1C.8-△*pthog1* into *Agrobacterium tumefaciens* EHA 105 strain was carried out by the heat shock method (Höfgen & Willmitzer 1988) and the positive transformants were identified by PCR amplification of the hygromycin resistant gene with primers 5'-GTCCGTCAGGACATTGTTGGAGCC-3' and 5'-GTCTCCGACCTGATGCAGCTCTCGG-3'.

To screen hygromycin resistance for PT, the mycelia were cultured on MYG solid medium with 0, 5, 10, 15, 20, 25, 30 and 35 mg/L hygromycin, respectively. The *Agr. tumefaciens* mediated method was used to transfer the plasmid pP1C.8-△*pthog1* to PT mycelia cultured in Malt Yeast Extract Glucose (MYG) liquid medium for 7 days. The positive mutants were screened by MYG medium with 35mg/L hygromycin and identified by pthog1-specific primer. The sequencing results from the mutants were blast with the gene sequence of *pthog1*.

To investigate the function of *pthog1* in the response to osmotic stress and adaptation to low temperature stimulation, the positive mutant ΔM5 and the wide type (WT) strains were separately cultured on MYG medium containing gradients of mannitol concentration (0-0.9mol/L) and under 15 ℃ for 7-15 days. The growth rates of mycelia for ΔM5 and WT strains were monitored every day. Furthermore, to investigate the function of *pthog1* for growth period of primordium and fruit body formation, the mycelia of ΔM5 and WT strains were cultured at 25 ℃ for 7-15 days, plus 4 ℃ for 3 days, and then subjected to 13 ℃ to observe the primordium initiation. Then, the mycelia of ΔM5 and WT strains were grown in cultivation bottles to form fruiting body. The development of fruiting body of ΔM5 and WT strains were observed and monitored in the same condition.

Results & Discussion

Genome assembly and annotation for the cultivated mushrooms

Reference genomes provide fundamental genetic resources for genetic analysis of cultivated mushrooms. According to the record, there are roughly 55 commercially cultivated mushrooms

worldwide (Li 2018). To date, there are 28 genomes available in the NCBI database. In this work, we generated 22 new high-quality genome sequences of the cultivated mushrooms (Supplementary Fig. 1, Table 1). For new sequenced species, a high coverage long sequencing data (71–318×) using PacBio sequel platform and short sequencing data (~70–150×) using Illumina platform (Supplementary Fig. 1) were generated. Genomic sizes of *de novo* assembled genome ranged from 33.1 to 74.8 Mb (Table 1). CEGMA and BUSCO analyses showed that 94%–98% of the core eukaryotic genes and 90%–98% single-copy genes were identified in the newly assembled genomes, respectively (Supplementary Table 1). Notably, the contig N50 of new genomes were up to 1.4–4.4 Mb (Supplementary Fig. 1, Table 1) and contig numbers were reduced to 19–70 (Supplementary Fig. 1, Table 1), exhibiting higher continuity than previous mushroom genomes (Supplementary Fig. 1). Overall, the newly assembled genomes exhibited high-quality both on completeness and continuity.

We identified 9.18–26.43% repeat contents in the newly released genomes (Supplementary Fig. 1, Supplementary Table 2), containing 0.01–1.33% tandem repeats and 5.95–24.24% transposons. Most of the predicted TEs were long terminal repeats (LTRs), representing 3.93–20.60% of the genomes. We predicted 5,799 to 17,803 protein-coding genes in the newly assembled genomes (Supplementary Fig. 1) and functionally annotated 83.6–99.9% homologs in seven public databases (Supplementary Table 3). Among these protein-coding genes, the average gene length ranged from 1,874 to 3,326 bp, while coding sequences (CDS) length range from 1,328 to 1,600 bp (Supplementary Table 4). The standard deviation of gene length (368.7 bp) among mushrooms species was much larger than that of CDS length (62.9 bp), indicating the difference of gene length distribution among mushroom could be largely attributed to non-coding sequences, such as the intron regions. Copy number of tRNA, rRNA and tRNA genes is one of interesting direction in biology studies (Iben & Maraia 2012). For non-coding RNAs in mushrooms, the new genomes possessed 167–756 tRNAs, 4–69 rRNAs, and 14–29 snRNAs (Supplementary Table 5).

Table 1 Major genome features for 50 cultivated mushrooms deposited in MushDB.

Scientific name	Genome size (Mb)	N50 (Mb)	Contig number	Genome data source	Note
Agaricus bisporus	30.0	2.30	26	NCBI	Brown strain
Agaricus bisporus	30.2	2.33	29	JGI	White strain
Auricularia cornea	74.8	4.36	37	*de novo*	Black-brown strain
Auricularia cornea	73.5	4.35	24	NCBI	White strain
Auricularia heimuer	47.2	3.90	20	NCBI	--
Armillaria gallica	85.3	1.04	319	NCBI	--
Coprinus comatus	33.1	2.36	22	*de novo*	--
Cyclocybe chaxingu	51.2	2.94	29	*de novo*	--
Cordyceps militaris	32.3	4.55	32	NCBI	--
Cyclocybe parasitica	53.9	1.92	42	*de novo*	--
Flammulina filiformis (as "*Flammulina velutipes*")	35.6	3.90	11	NCBI	Yellow strain
Flammulina filiformis (as "*Flammulina velutipes*")	40.4	2.80	19	*de novo*	White strain
Grifola frondosa	39.3	2.55	30	*de novo*	--
Ganoderma sichuanense	47	3.57	16	NCBI	--
Ganoderma sinense	54.5	3.25	71	NCBI	--
Gloeostereum incarnatum	38.7	3.5	20	NCBI	--
Hericium coralloides	42.4	1.43	62	*de novo*	--
Hericium erinaceus	39.8	2.64	39	*de novo*	--
Hypsizygus marmoreus	41.9	1.23	178	NCBI	Brown strain
Hypsizygus marmoreus	40.9	0.82	115	NCBI	White strain
Lyophyllum decastes	43	0.28	491	NCBI	--

Table 1 Continued.

Scientific name	Genome size (Mb)	N50 (Mb)	Contig number	Genome data source	Note
Lyophyllum shimeji	52.4	0.83	306	NCBI	--
Lentinula edodes	46	2.76	28	NCBI	--
Hymenopellis raphanipes	61.2	2.08	70	*de novo*	--
Ophiocordyceps sinensis	101.1	0.35	618	NCBI	--
Pholiota adiposa	43.6	3.47	20	*de novo*	--
Pholiota microspora	36.3	1.67	53	*de novo*	--
Pleurotus abieticola	40.8	3.49	24	*de novo*	--
Pleurotus citrinopileatus	36.3	2.48	32	*de novo*	--
Pleurotus djamor	41.4	3.63	23	*de novo*	--
Pleurotus eryngii var. *eryngii*	53.6	3.20	48	NCBI	--
Pleurotus eryngii var. *ferulae*	48	2.30	51	NCBI	--
Pleurotus floridanus	34.5	1.86	46	*de novo*	--
Pleurotus giganteus	39.8	3.19	27	*de novo*	--
Pleurotus nebrodensis	49.4	2.54	39	NCBI	--
Pleurotus ostreatus	40	2.65	21	*de novo*	--
Pleurotus populinus	41	3.44	23	*de novo*	--
Pleurotus pulmonarius	42.2	2.95	31	*de novo*	--
Pleurotus sapidus	37.2	2.36	28	*de novo*	--
Pleurotus tuoliensis	43.3	2.45	34	*de novo*	--
Schizophyllum commune	38.5	2.55	36	NCBI	--
Sarcomyxa edulis	35.7	1.77	41	NCBI	--
Sparassis latifolia	35.7	0.35	184	JGI	--
Stropharia rugosoannulata	47.8	2.96	33	*de novo*	--
Taiwanofungus camphoratus	32.2	1.03	360	NCBI	--
Tremella fuciformis	28.1	2.71	14	NCBI	--
Tremella mesenterica	28.6	1.62	45	NCBI	--
Tuber melanosporum	124.9	0.64	398	NCBI	--
Volvariella volvacea	35.7	0.93	62	NCBI	--
Wolfiporia cocos	50.5	2.54	348	NCBI	--

To probe the phylogenetic relationship and the speciation divergence time among the commercially cultivated mushrooms, the species in MushDB and their relative species with relative higher quality genomes in the NCBI database were used for the phylogenomic analysis. As a result, gene families were clustered using all protein-coding genes for 84 fungal species, and 32 single-copy orthologous genes among all species were obtained to construct the phylogenetic tree with *Phy. blakesleeanus* as the outgroup taxon (Fig. 1). The topology of the phylogenetic tree generally recovered the evolutionary relationships of cultivated mushroom species. The species were strictly clustered into clades, which were consistent with their classification at the order level. The divergence times of species were also coincident with previous literatures (Floudas et al. 2012, Zhao et al. 2017). For example, we estimated the divergence time of basidiomycetes was 510.52 (499.4–528.7) Mya, which was similar with 521 (402.79–664.85) Mya in Floudas et al. (2012) and 530 Mya in Zhao et al. (2017). The estimated divergence time of Polyporales (151.69 (133.7–163.3) Mya) and Gloenphyllales (123.91 (107–134) Mya) were also similar to those of 158 Mya and 124 Mya, respectively. In addition, we estimated that Eurotiales diverged from their common ancestor of Pleosporales and Hypocreales about 255.02 (166.2–392.4) Mya, 178.52 (71.8–370.3)

Mya, respectively, which was also consistent with previous studies (Sharpton et al. 2009, Prieto & Wedin 2013).

Previous studies have illuminated copy number variations of functional genes that might play an important role in adaptations of ecological niches and domestication among fungi species (Hu et al. 2014). We performed functional annotation of the protein-coding gene families using CAZy database, and found that many functional genes involved in the carbohydrate metabolism showed clade-specific expansion patterns. The CAZy families of GH75, GH18, and CE5 were significantly expanded in many species, especially for Cordycipitaceae. Given the fact that the Cordycipitaceae species utilized animal tissues, the expansion of those functional genes, especially for CE5, might be related with the adaptive requirement to hydrolyze chitin and cutin (Floudas et al. 2012). The GH7, AA5, AA9, PL1, CMB1, CMB13 families were significantly larger in Agaricales, especially in *Pleurotus* (Fig. 1). GH7 is a glycoside hydrolases for cellulosis and AA5 and AA9 both are typical auxiliary proteins to boost the enzymatic hydrolysis of cellulose (Levasseur et al. 2013). Cellulose binding modules (CBM) promote the association of the enzyme and cellulose substrate and also play a central role in cellobiohydrolase (Boraston et al. 2004). Compared with *Sarcomyxa*, the expansion of those functional genes involved in cellulose degradation may meet nutrient requirements and facilitate domestication of *Pleurotus* species. The whole genome data in MushDB provided the research community useful reference for the functional gene studies.

MushDB database construction

To facilitate the data sharing and functional analysis of the commercially cultivated mushrooms by an interactive visualization platform, MushDB (http://mushroomomics.com) database was then constructed using multi-omics data from this work and previous studies (Fig. 2a). Besides 22 newly assembled genomes in this study, 12 previously published genomes in our team and 16 genomes from NCBI and JGI database of the cultivated mushrooms were collected in MushDB database (Supplementary Fig. 1), covering ~90% of commercially cultivated mushrooms. Among these 50 genomes, four species have more than one genome representing different color strains that are commercially cultivated, including *Ag. bisporus* (white and brown strain), *Flammulina filiformis* (as "*Flammulina velutipes*", white and yellow strain), *Au. cornea* (white and black-brown strain), and *Hypsizygus marmoreus* (white and brown strain, Table 1). To unify the genome information for cross-species comparisons, gene models were predicted with the standard gene prediction pipeline for all 50 genomes and released in MushDB. The final genome sequences, genome annotation and CDS/protein sequences of predicted genes were available for all 50 genomes in MushDB homepage, which can be downloaded and used for further studies (Supplementary Fig. 2). The genome data have also been deposited into CNGB Sequence Archive (CNSA) of China National GeneBank DataBase (CNGBdb) with accession number CNP0003221.

The current release of MushDB also included 621 whole genome resequencing data (Fig. 2b) with 17-48 × depth (240 previously published by our research team) and 265 transcriptome data (Fig. 2c) for those commercially cultivated mushrooms. Among which, 381 re-sequencing and 63 transcriptome data were reported for the first time. Using 265 transcriptome data and the standardized analysis pipelines for RNA-seq, we integrated RNA-seq-based gene expression profile data of each species from major mushroom tissues (mycelium, primordium and fruiting body, Fig. 2c), different biological processes such as low temperature stimulation, different developmental stages such as young and mature fruiting body, and different cultivars (Supplementary Fig. 3). We also manually curated pairwise comparison to identify differentially expressed genes in response to environmental stresses, in the different developmental stages and genetic mutations. Meanwhile, using resequencing data with the standardized analysis pipelines, a dedicated variation database was constructed to provide whole-genome SNP/INDEL variation information for each species available in MushDB. The variation database of each species is varied in size depending on the sample number and sequencing data amount available for each species (Supplementary Fig. 3).

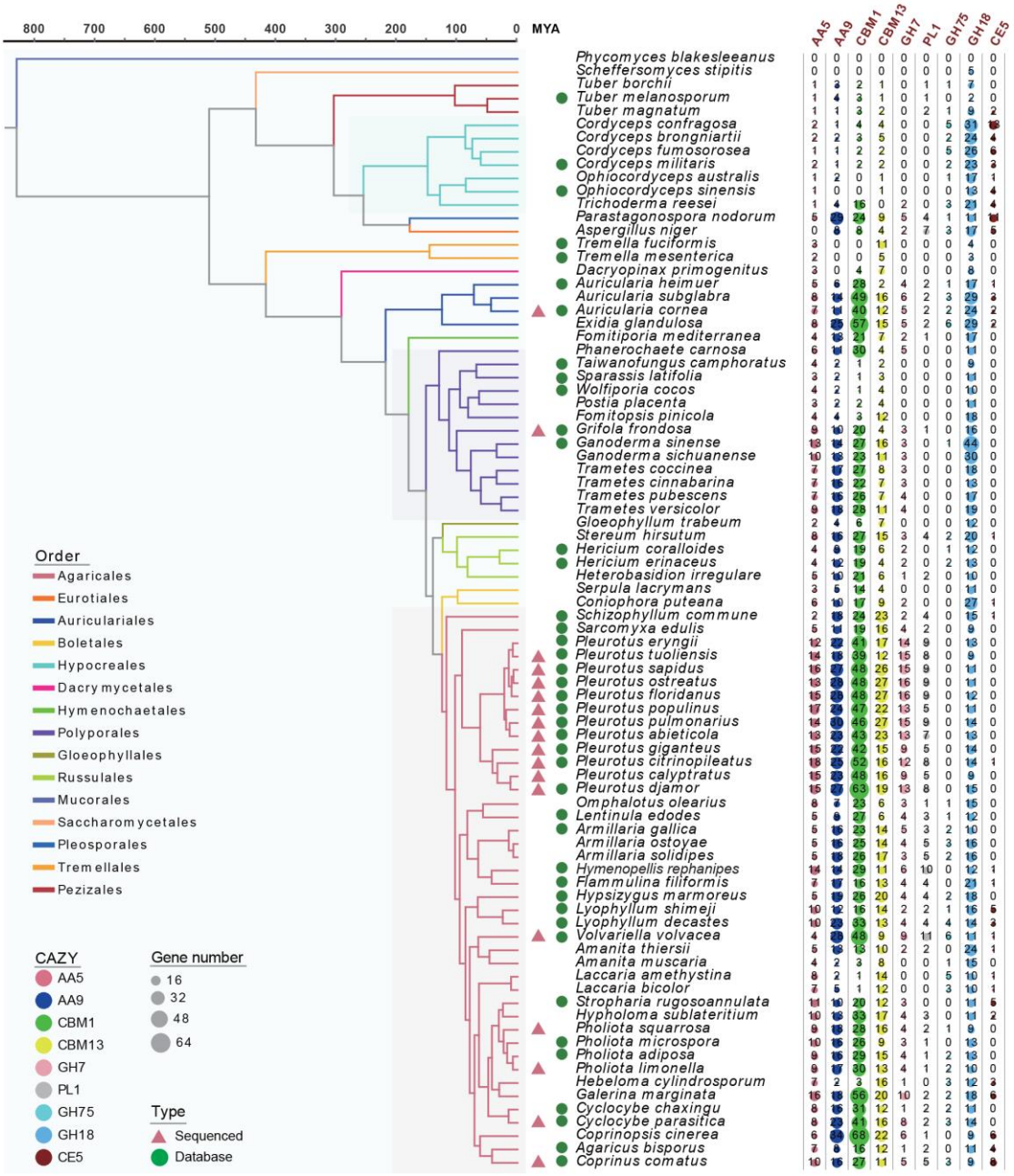


Figure 1 – The phylogenetic analysis of cultivated mushrooms with other fungal species. The colors of branches in the tree represented the taxonomy on the order level. The genomes reported in this study and included in the database were labeled by triangles and dots before the species name, respectively. Specific CAZy terms and gene numbers were shown by the color and size of dots, respectively.

In summary, MushDB database has the following features to facilitate the genomic studies of cultivated mushrooms compared with general public databases. First, MushDB provides the most abundant and comprehensive genome sequencing, gene expression and variation data specifically for cultivated mushrooms (Supplementary Figs 1-3), covering two phyla, four classes, seven orders, 20 families, 27 genera and 46 species (four species each had two genomes that represented the different color strains) in the kingdom fungi. Secondly, MushDB emphasizes the functionality of genomic and transcriptome date. Rather than storing raw datasets, MushDB have processed all sequencing data by uniform and standardized pipelines, and the results were structurally maintained

in the database, making the gene sequence, expression, and variant data in MushDB easily queried and compared directly. Thirdly, MushDB also provides advanced functional modules in the integrated omics portal to facilitate systematical and flexible gene investigations using comprehensive information, including annotations, sequences, expressions, and variations (Supplementary Fig. 3). For example, users only need to input either the gene ID, the name, or the located region in the search engine of the interested species to access all gene information, including the statistical information, sequences, biological functions, expression profiles, and variations in/around the gene. In addition, we also provide the unique representations and a powerful and comprehensive interface visualizations for expression and variation data (Supplementary Fig. 3). Specifically, gene expression profiles for different tissues and treatments could be visualized by interactively heatmaps (Supplementary Fig. 3). Different types of SNP/INDEL variations around a given gene could be displayed on a flexible visualization panel. More importantly, MushDB also allows users to perform customized population analysis to compare the frequency of variations between two groups using the existing whole-genome data, such as the different type (wild and cultivated group) or cap color (white and yellow group, Supplementary Fig. 3). The relevant table and figure of the general features of gene expression profiles and SNP/INDEL loci could also be downloaded as publishable high-quality figures (Supplementary Fig. 3).

Functional genes and pathways in the domestication of cultivated mushrooms

Functional genes play critical roles in the domestication of many crops and live stock (Doebley et al. 2006, Lin et al. 2014, Chen et al. 2018). Understanding the domestication mechanism could not only help us to understand the breeding history of modern strains, but also provide important hints to prompt the domestication efficiency for un-cultivated mushroom species. To identify putative functional genes and pathways that involved in the domestication of mushrooms, we compared the wild and cultivated populations for four representative species: *Au. cornea* (AC), *Le. edodes* (LE), *Pl. eryngii* (PE) and *Pl. tuoliensis* (PT) (Supplementary Table 6). Based on the whole-genome sequencing data, we defined those genomic regions with 5% highest F_{st} value which were significantly selected and genes resided in those genomic regions were putatively selected functional genes for AC, LE, PE and PT (Fig. 3a, b, Supplementary Fig. 4, Supplementary Tables 7-10).

To investigate the major biological function of putatively selected genes for those four species, the KEGG pathway annotations were assigned for those putatively selected genes of AC, LE, PE and PT. As a result, 1,033 KEGG Ortholog (KO) terms were assigned to be putatively selected in four species. We found that 21 KO terms were significantly selected in all four species, and 85 and 370 KO terms were selected in at least three and two species (Fig. 3c), respectively. Remarkably, we identified gene encoding cellulose 1,4-beta-cellobiosidase (gene symbol: *cbh1*, KO term: K01225) as one of the 21 independently selected KO terms in all species (Fig. 3c). The c*bh1* gene is a central in the cellulose hydrolysis to catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in cellulose to release the disaccharide cellobiose (Tiquia-Arashiro & Grube 2019). Based on the fundamental function of *cbh1* in the nutrition digestive absorption for mushroom, we proposed that *cbh1* might play an essential role in the successful domestication of cultivated mushrooms. Meanwhile, we also identified the gene encoding cell division control protein 24 (gene symbol: *cdc24*, KO term: K11236) was selected in four species. The *cdc24* gene is an important gene in MAPK signaling pathway and involved in wide biological functions of the development, mating, and stress response (Zhang et al. 2016, Pérez et al. 2020).

To analyze the function and pathway of independently selected genes during the domestication of four cultivated mushrooms, we performed an enrichment analysis for genes that selected in at-least two species (Supplementary Table 11). We found that independently selected genes were significantly enriched in the starch and sucrose metabolism (Fig. 3d, p-value = 0.002) and MAPK signaling pathway (Fig. 3d, p-value = 0.04). Previous studies illuminated the importance of MAPK signaling pathway in the development, pathogenicity, environment

adaptation, stress response and mating of fungi (Bahn 2008, Saito 2010). In the MAPK signaling pathway, we identified that 23 of all 43 KO terms were selected (Fig. 3e). Massive independently selected genes in the MAPK signaling pathway among species revealed that the pathway might be essential for successful domestication of mushrooms. The majority of MAPK, MAPKK and MAPKKK genes were selected in at least two species, including *hog1*, *ass1*, *fus3*, *ste11*, *ssk2/22*, *mkk1/2* and *bck1* genes (Fig. 3e). Recent gene functional analysis uncovered genes in the MAPK pathway might also be involved in the light sensation and signaling (Esquivel-Naranjo et al. 2016, Fu et al. 2016), indicating the MAPK pathway might also be involved in the rhythm in mushroom. The rhythm is one of the domestication directions for wild and cultivated mushrooms due to the changes of the growth and development environments, such as light, temperature, humidity, and CO_2 concentration (Xiao et al. 2016). Therefore, independently selected genes in MAPK pathway provide valuable information to investigate the function of rhythm-regulation related genes and their potential contribution to mushroom domestication. In the starch and sucrose metabolism, we identified 17 KO terms were assigned to genes for the four species. Among them, 12 KOs were selected in at-least two species. Besides the above-mentioned *cbh1* gene, many other important genes involved in the carbohydrate metabolism were also independently selected in multiple species, including glucan 1,3-beta-glucosidase (*bgl2*), beta-glucosidase (*bglx*, EC:3.2.1.21) and cellulase (*celb*, EC:3.2.1.4). Strikingly, genes encoding hydrolase in the cellulose degradation were all selected by multiple species (Fig. 3f), indicating functional genes in the cellulose hydrolysis also likely contributed to the domestication of mushroom species.

Functional analyses of MAPK gene *pthog1* in *Pl. tuoliensis* using CRISPR/Cas9 system

To validate the gene function of above-mentioned MAPK gene *hog1* putative associated with domestication, we used the CRISPR/Cas9 system to knock out *hog1* in *Pl. tuoliensis* (PT). We found the ORF length of the hog1 gene *pthog1* in PT was 1,149 bp with four introns (Supplementary Fig. 5), which is a conserved ortholog gene in fungi with single copy in PT (Supplementary Fig. 6). Based on the modified pP1c.8 expression vector with the hygromycin resistance gene, we constructed the CRISPR/Cas9 knockout vector for *pthog1* with the target sites predicted by the CRISPR-P program (Supplementary Fig. 7). The *Agr. tumefaciens* mediated transformation method was selected for the mycelia transformation of PT, which used the 35 mg/mL hygromycin in MYG medium to screen the positive transformants (Supplementary Fig. 7). The sequencing results of these transformants showed two deletions near the target sites in *pthog1* (Fig. 4a).

The hog-MAPK signaling pathway was first discovered in *Saccharomyces cerevisiae* and its hog gene is responsible for responding to osmotic stress in regulating cells (Posas et al. 1996). Therefore, we firstly performed the functional validation of these positive transformants involving in the mycelia response to the osmotic stress. Along with the increasing mannitol concentration in MYG medium (Fig. 4b), the mutant strain ΔM5 showed much slower growth rate comparing to WT strain, and even stopped growing under 0.9 mol/L of mannitol while the WT strain could still grow. It implied that the gene *pthog1* played an important role in the tolerance of osmotic stress, which was consistent with the previous studies in yeast (Posas et al. 1996, Chen & Thorner 2007).

Recent studies have shown that the *hog* gene is crucial for the adaptation of external stress, such as low temperature. Panadero et al. (2006) found that the hog-MAPK pathway was activated when yeast cells were stimulated by cold stress, and could enhance the resistance of cells to cold stress. For domestication and cultivation of PT, cold stimulation of mycelia is the main trigger for primordia initiation, mushroom formation, and yields (Fu et al. 2016). Based on our previously transcriptomic analysis during different temporal stages of cold stimulation in PT, we also found *pthog1* was critical for the response to cold stimulation (Fu et al. 2016). In this study, we found the mycelia growth rate of ΔM5 was remarkably lower under a low culture temperature than that of WT strain (Fig. 4c), indicating *pthog1* might contribute to the short growth period of PT during the low temperature stimulation. Furthermore, the mycelia of ΔM5 almost failed to initiate the primordium or to form fruiting body after cold stimulation, while the mycelia of WT grew well in

identical culture conditions (Fig. 4d, Supplementary Fig. 7). Our CRISPR-Cas9 based gene function analysis revealed that *pthog1* might play an important role in the successful adaptation of cold stimulation and short growth period during domestication of *Pl. tuoliensis* in the cultivation environment.

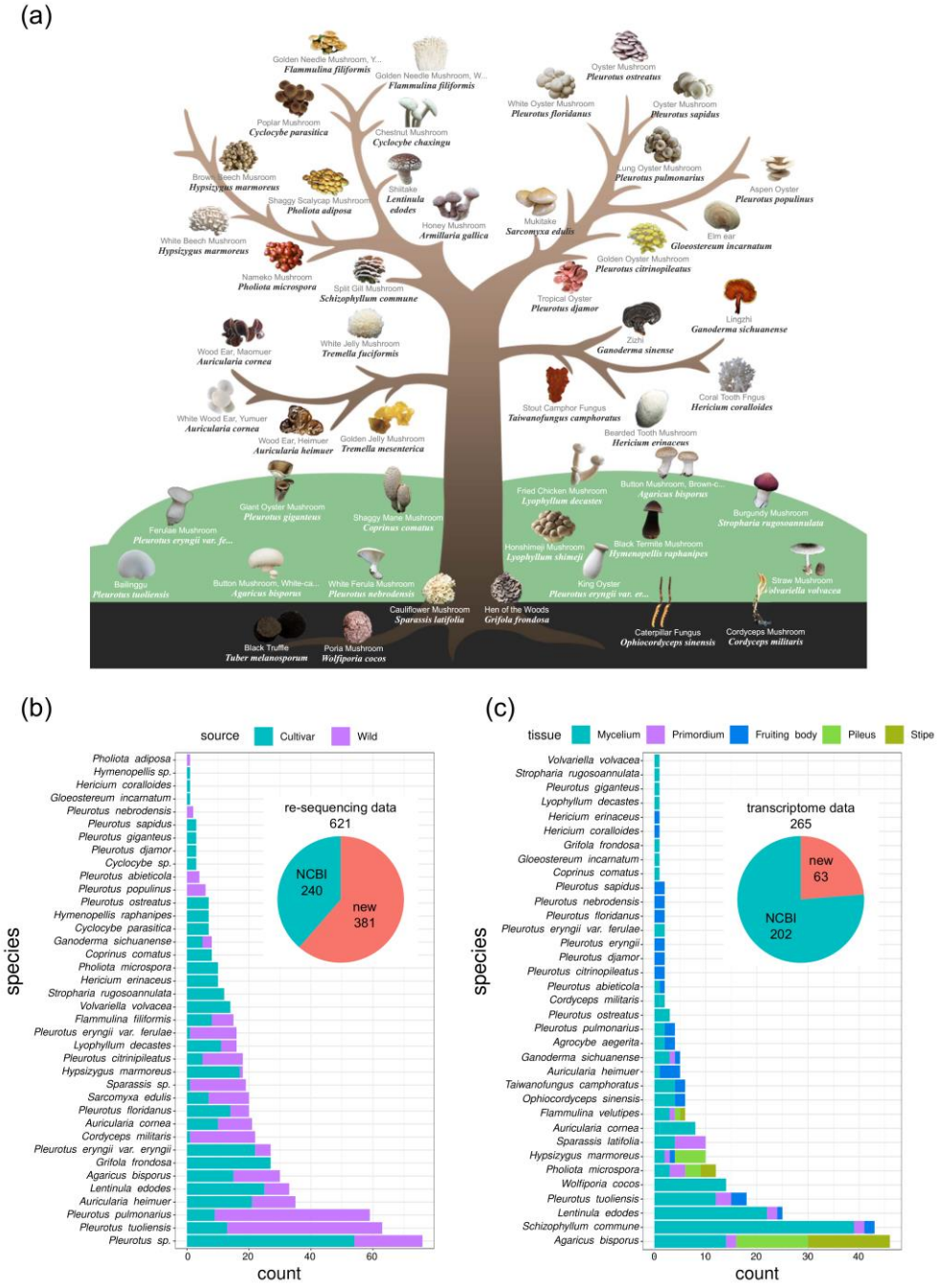


Figure 2 – Species available in MushDB database and the statistics of whole-genome re-sequencing and transcriptome data. a Integrated omics database of 50 kinds of cultivated mushrooms. b The whole-genome re-sequencing data in MushDB from wild and cultivated strains. The inner panel shows the number of newly generated and public re-sequencing data.

c The transcriptome data in MushDB for different tissues. The inner panel shows the number of newly generated and public transcriptome data.

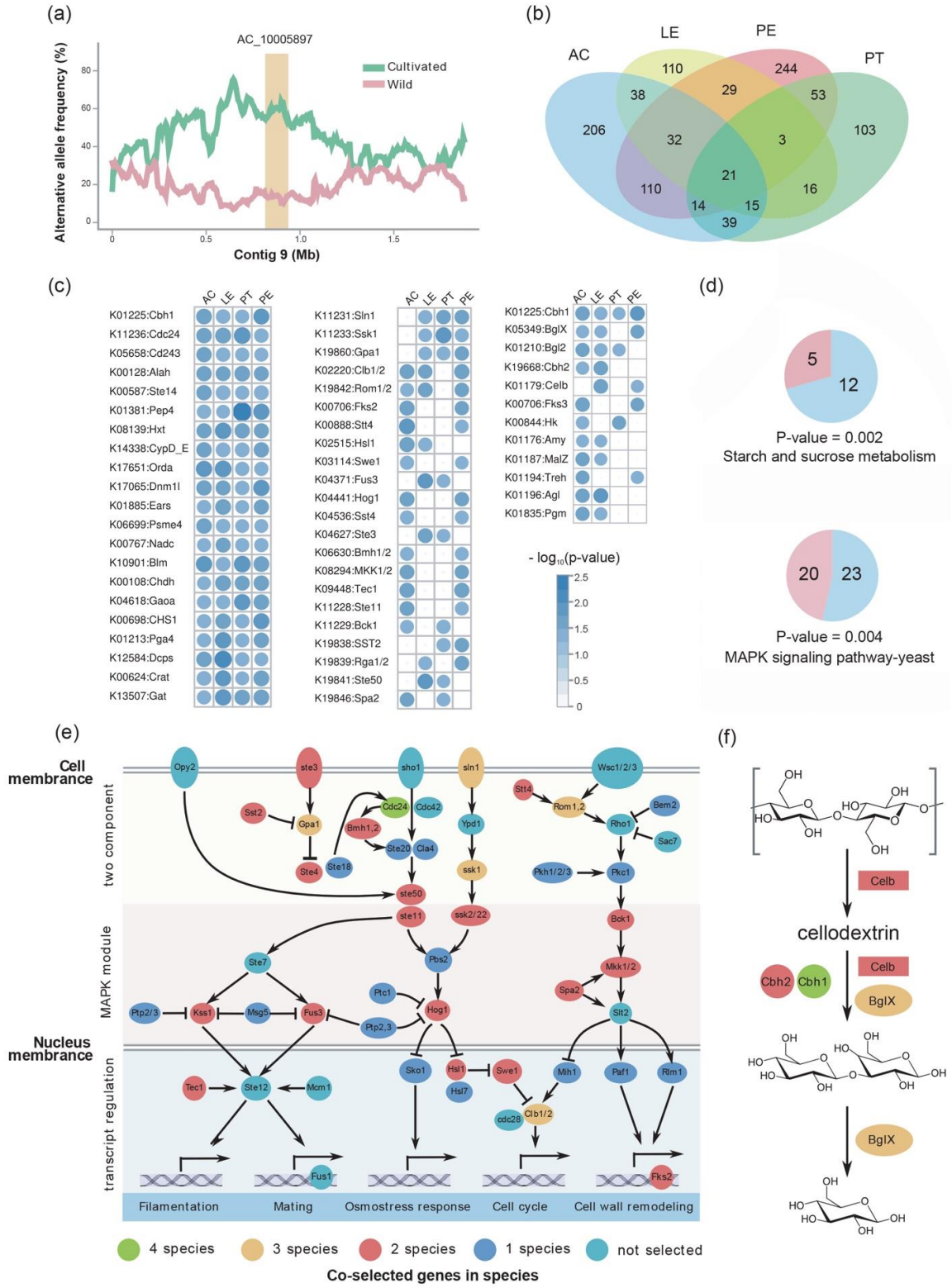


Figure 3 – Putative functional genes contributing to the domestication of *Au. cornea* (AC), *Le. edodes* (LE), *Pl. eryngii* (PE) and *Pl. tuoliensis* (PT). a The allele frequency differences between wild and cultivated populations to identified the selective genome regions. b The distribution of putative selected genes among AC, LE, PE and PT. c Significance of selected KO

terms among species. Genes selected in all species were shown in the left panel. Genes that selected in at least two species from MAPK signaling pathway and starch and sucrose metabolism were shown in the middle and right panel. d The enrichment of pathways of putatively selected genes in starch and sucrose metabolism, and MAPK signaling pathway. Selected genes from AC, LE, PE and PT in MAPK signaling pathway (e) and cellulose hydrolysis (f).

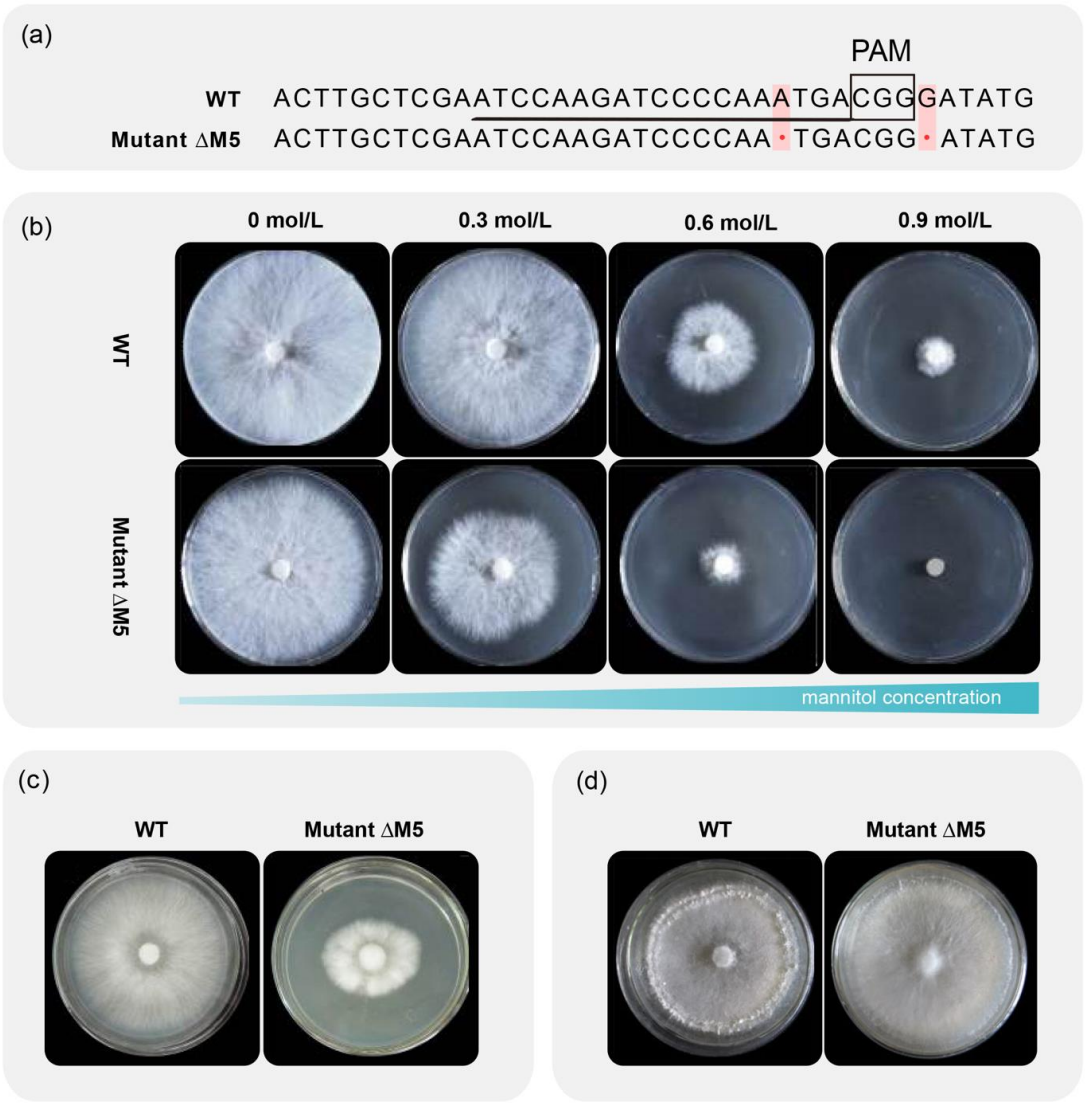


Figure 4 – CRISPR-Cas9 based functional validation of *pthog1* for its contribution to domestication of *Pleurotus tuoliensis*. a The sequencing result of the mutant strain ΔM5 validate the deletion near the target sites in *pthog1*. b The growth difference between WT and ΔM5 cultured in the different mannitol concentration. c The growth rate between WT and ΔM5 cultured at low temperature. d The comparison of primordium initiation between WT and ΔM5.

Conclusions

Cultivated mushrooms have been instrumental sources of food and medicine for humans, and are also critical in environmental maintenance and industry development. Functional genomics analysis of cultivated mushrooms using whole-genome information would help us to widen understanding of molecular mechanism of domestication of cultivated mushrooms. In this study, 22 new high-quality genome sequences, 381 re-sequencing and 63 transcriptome data for the cultivated mushrooms were reported for the first time. Using the *de novo* genomes, whole-genome

re-sequencing, and transcriptome data in this study, we constructed an integrated multi-omics database MushDB. Putative functional genes for the successful domestication of AC, LE, PE and PT were identified by wild and cultivated population comparison. Key genes in the starch and sucrose metabolism and MAPK signaling pathway, including *cbh1*, *cdc24* and *hog1*, were putatively selected and might contribute to the domestication of those species. To validate the biological function of selected genes, we applied the CRISPR-Cas9 system to demonstrate the important roles of *pthog1* for the successful domestication of PT in the adaptation of low temperature stimulation and short growth period. Our study offers important genetic resource for domestication studies of cultivated mushroom and set a solid foundation for the following identification of functional gene for important economic traits and molecular breeding of cultivated mushrooms.

Acknowledgements

This work was financed by Joint Funds of the National Natural Science Foundation of China (No. U20A204), Scientific and Technological Planning Project of Jilin Province (No. 20210402047GH), Special Fund for Agro-scientific Research in the Public Interest (No. 201503137), National Science and Technology Fundamental Resources Investigation Program of China (No. 2021FY100900), New Germplasm Breeding Program for Lingzhi (No. GF20190034), the Project for Seed Industry Revitalization of Jiangsu Province (No. JBGS-2021-088), the Fellowship of China Postdoctoral Science Foundation (No. 2021M693835), Science Research Project of Education Department of Jilin Province (No. JJKH20220355KJ), Science and Technology Research Project of Jiashan (No. 2021A21), National-level International Joint Research Centre (No. 2017B01011), "111" Project (No. D17014), National Research Council of Thailand (NRCT) (No. N42A650547), Thailand Science Research and Innovation (TSRI) (No. DBG6280009) and Zhejiang Science and Technology Major Program on Agricultural New Variety Breeding (2021C02073-9).

References

Akey JM, Zhang G, Zhang K, Jin L et al. 2002 – Interrogating a high-density SNP map for signatures of natural selection. Genome Research 12(12), 1805–1814. Doi 10.1101/gr.631202

Altschul SF, Gish W, Miller W, Myers EW et al. 1990 – Basic local alignment search tool. Journal of Molecular Biology 215, 403–410. Doi 10.1016/S0022-2836(05)80360-2

Anders S, Pyl PT, Huber W. 2015 – HTSeq – a Python framework to work with high-throughput sequencing data. Bioinformatics 31(2), 166–169. Doi 10.1101/002824

Bahn YS. 2008 – Master and commander in fungal pathogens: the two-component system and the HOG signaling pathway. Eukaryot Cell 7(12), 2017–2036. Doi 10.1128/EC.00323-08

Bibeault B, Katz Y. 2008 – jQuery in Action. Dreamtech Press, New Delhi.

Boraston AB, Bolum DN, Gilbert HJ, Davies GJ. 2004 – Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochemical Journal 382, 769–781. Doi 10.1042/BJ20040892

Camacho C, Coulouris G, Avagyan V, Ma N et al. 2009 – BLAST+: architecture and applications. BMC Bioinformatics 10, 421. Doi 10.1186/1471-2105-10-421

Chen NB, Cai YD, Chen QM, Li R et al. 2018 – Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. Nature Communications 9(1), 2337. Doi 10.1038/s41467-018-04737-0

Chen RE, Thorner J. 2007 – Function and regulation in MAPK signaling pathways: lessons learned from the yeast *Saccharomyces cerevisiae*. Biochim Biophys Acta 1773(8), 1311–1340. Doi 10.1016/j.bbamcr.2007.05.003

Chen S, Zhou Y, Chen Y, Gu J. 2018 – fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34(17), i884–i890. Doi 10.1093/bioinformatics/bty560

Dai YT, Su WY, Yang CT, Song B et al. 2017 – Development of novel polymorphic EST-SSR markers in Bailinggu (*Pleurotus tuoliensis*) for crossbreeding. Genes 8(11), 325. Doi 10.3390/genes8110325

De Bie T, Cristianini N, Demuth JP, Hahn MW. 2006 – CAFE: a computational tool for the study of gene family evolution. Bioinformatics 22(10), 1269–1271. Doi 10.1093/bioinformatics/btl097

Denton JF, Lugo-Martinez J, Tucker AE, Schrider DR et al. 2014 – Extensive error in the number of genes inferred from draft genome assemblies. PLoS Computational Biology 10(12), e1003998. Doi 10.1371/journal.pcbi.1003998

Doebley JF, Gaut BS, Smith BD. 2006 – The molecular genetics of crop domestication. Cell 127(7), 1309–1321. Doi 10.1016/j.cell.2006.12.006

Esquivel-Naranjo EU, García-Esquivel M, Medina-Castellanos E, Correa-Pérez VA et al. 2016 – A Trichoderma atroviride stress-activated MAPK pathway integrates stress and light signals. Molecular Microbiology 100(5), 860–876. Doi 10.1111/mmi.13355

Finn RD, Bateman A, Clements J, Coggill P et al. 2014 – Pfam: the protein families database. Nucleic Acids Research 42(D1), D222–D230. Doi 10.1093/nar/gkt1223

Floudas D, Binder M, Riley R, Barry K et al. 2012 – The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. Science 336(6089), 1715–1719. Doi 10.1126/science.1221748

Fu YP, Liang Y, Dai YT, Yang CT et al. 2016 – De novo sequencing and transcriptome analysis of *Pleurotus eryngii* subsp. *tuoliensis* (Bailinggu) mycelia in response to cold stimulation. Molecules 21(5), 560. Doi 10.3390/molecules21050560

Hibbett DS. 2004 – Trends in morphological evolution in homobasidiomycetes inferred using maximum likelihood: a comparison of binary and multistate approaches. Systematic Biology 53(6), 889–903. Doi 10.1080/10635150490522610

Hibbett DS, Binder M. 2002 – Evolution of complex fruiting-body morphologies in homobasidiomycetes. Proceedings of the Royal Society London B 269, 1963–1969. Doi 10.1098/rspb.2002.2123

Ho C. 2012 – Using MyBatis in Spring. In: Pro Spring 3. Apress, Heidelberg, New York.

Höfgen R, Willmitzer L. 1988 – Storage of competent cells for Agrobacterium transformation. Nucleic Acids Research 16(20), 9877. Doi 10.1093/nar/16.20.9877

Hu X, Xiao G, Zheng P, Shang Y et al. 2014 – Trajectory and genomic determinants of fungal-pathogen speciation and host adaptation. Proceedings of the National Academy of Sciences 111(47), 16796. Doi 10.1073/pnas.1412662111

Iben JR, Maraia RJ. 2012 – tRNAomics: tRNA gene copy number variation and codon use provide bioinformatic evidence of a new anticodon:codon wobble pair in a eukaryote. RNA 18(7), 1358–1372. Doi 10.1261/rna.032151.111

Jiang N, Hu S, Peng B, Li Z et al. 2021 – Genome of *Ganoderma* species provides insights into the evolution, conifers substrate utilization, and terpene synthesis for *Ganoderma tsugae*. Frontiers in Microbiology 12, 724451. Doi 10.3389/fmicb.2021.724451

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P et al. 2005 – Repbase update, a database of eukaryotic repetitive elements. Cytogenetic and Genome Research 110, 462–467. Doi 10.1159/000084979

Kanehisa M, Goto S. 2000 – KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Research 27(1), 29–34. Doi 10.1093/nar/28.1.27

Kim D, Langmead B, Salzberg SL. 2015 – HISAT: a fast spliced aligner with low memory requirements. Nature Methods 12(4), 357–360. Doi 10.1038/nmeth.3317

Kofler R, Pandey RR, Schlötterer C. 2011 – PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics 27(24), 3435–3436. Doi 10.1093/bioinformatics/btr589

Kurata A, Fukuta Y, Mori M, Kishimoto N et al. 2016 – Draft genome sequence of the Basidiomycetous fungus *Flammulina velutipes* TR19. Genome Announcements 4(3), e00505–00516. Doi 10.1128/genomeA.00505-16

Lam HM, Xu X, Liu X, Chen WB et al. 2010 – Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nature Genetics 42(12), 1053–1059. Doi 10.1038/ng.715

Lei Y, Lu L, Liu HY, Li S et al. 2014 – CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants. Molecular Plant 7(9), 1494–1496. Doi 10.1093/mp/ssu044

Levasseur A, Drula E, Lombard V, Coutinho PM et al. 2013 – Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. Biotechnology for Biofuels and Bioproducts 6(1), 41. Doi 10.1186/1754-6834-6-41

Li L, Jr CJS, Roos DS. 2003 – OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Research 13(9), 2178–2189. Doi 10.1101/gr.1224503

Li Y. 2018 – The status, opportunities and challenges of edible fungi industry in China: develop with Chinese characteristics, realize the dream of powerful mushroom industrial country. Journal of Fungal Research 3, 125–1231.

Li Y, Cao K, Zhu G, Fang W et al. 2019 – Genomic analyses of an extensive collection of wild and cultivated accessions provide new insights into peach breeding history. Genome Biology 2019 20(1), 36. Doi 10.1186/s13059-019-1648-9.

Lin T, Zhu GT, Zhang JH, Xu XY et al. 2014 – Genomic analyses provide insights into the history of tomato breeding. Nature Genetics 46(11), 1220–1226. Doi 10.1038/ng.3117

Liu YC, Du HL, Li PC, Shen YT et al. 2020 – Pan-genome of wild and cultivated soybeans. Cell 182(1), 162–176. Doi 10.1016/j.cell.2020.05.023

Lowe TM, Eddy SR. 1997 – tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research 25, 955–964. Doi 10.1093/nar/25.5.955

Mak G. 2008 – Spring MVC Framework. In: Spring Recipes. Apress, Heidelberg, New York.

Miles PG, Chang S. 2004 – Mushrooms: cultivation, nutritional value, medicinal effect, and environmental impact. 2nd edition. CRC press, Florida.

Nawrocki EP, Kolbe DL, Eddy SR. 2009 – Infernal 1.0: inference of RNA alignments. Bioinformatics 25, 1335–1337. Doi 10.1093/bioinformatics/btp157

Pan ZY, Yao YL, Yin HW, Cai ZX et al. 2021 – Pig genome functional annotation enhances the biological interpretation of complex traits and human disease. Nature Communications 12, 5848. Doi 10.1038/s41467-021-26153-7

Panadero J, Pallotti C, Rodríguez-Vargas S, Randez-Gil F et al. 2006 – A downshift in temperature activates the high osmolarity glycerol (HOG) pathway, which determines freeze tolerance in *Saccharomyces cerevisiae*. Journal of Biological Chemistry 281(8), 4638–4645. Doi 10.1074/jbc.M512736200

Parra G, Bradnam K, Korf I. 2007 – CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23(9), 1061–1067. Doi 10.1093/bioinformatics/btm071

Pavlidis P, Živkovic D, Stamatakis A, Alachiotis N et al. 2007 – SweeD: likelihood-based detection of selective sweeps in thousands of genomes. Molecular Biology and Evolution 30(9), 2224–2234. Doi 10.1093/molbev/mst112

Pérez P, Soto T, Gómez-Gil E, Cansado J. 2020 – Functional interaction between Cdc42 and the stress MAPK signaling pathway during the regulation of fission yeast polarized growth. International Microbiology 23(1), 31–41. Doi 10.1007/s10123-019-00072-6

Phan CW, Sabaratnam V. 2012 – Potential uses of spent mushroom substrate and its associated lignocellulosic enzymes. Applied Microbiology and Biotechnology 96(4), 863–873. Doi 10.1007/s00253-012-4446-9

Poplin R, Ruano-Rubio V, Depristo MA, Fennell TJ. 2017 – Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv 201178, Cold Spring Harbor Laboratory.

Posas F, Wurgler-Murphy SM, Maeda T, Witten EA et al. 1996 – Yeast HOG1 MAP kinase cascade is regulated by a multistep phosphorelay mechanism in the SLN1-YPD1-SSK1 "two-component" osmosensor. Cell 86(6), 865–875. Doi 10.1016/s0092-8674(00)80162-2

Prieto M, Wedin M. 2013 – Dating the diversification of the major lineages of Ascomycota (Fungi). PLoS One 8(6), e65576. Doi 10.1371/journal.pone.0065576

Pruitt KD, Tatusova T, Maglott DR. 2007 – NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Research 35, D61–D65. Doi 10.1093/nar/gkl842

Qin P, Lu HW, Du HL, Wang H et al. 2021 – Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. Cell 184(13), 3542–3558. Doi 10.1016/j.cell.2021.04.046

Quevillon E, Silventoinen V, Pillai S, Harte N et al. 2005 – InterProScan: protein domains identifier. Nucleic Acids Research 33, W116–120. Doi 10.1093/nar/gki442

Rhie A, McCarthy SA, Fedrigo O, Damas J et al. 2021 – Towards complete and error-free genome assemblies of all vertebrate species. Nature 592(7856), 737–746. Doi 10.1038/s41586-021-03451-0

Roberts P, Evans S. 2014 – The book of fungi: a life-size guide to six hundred species from around the world. University of Chicago Press, Chicago.

Rozen S, Skaletsky H. 2000 – Primer3 on the WWW for general users and for biologist programmers. Bioinformatics Methods and Protocols 132, 365–386. Doi 10.1385/1-59259-192-2:365

Saito H. 2010 – Regulation of cross-talk in yeast MAPK signaling pathways. Current Opinion in Microbiology 13(6), 677–683. Doi 10.1016/j.mib.2010.09.001

Schneider M, Lane L, Boutet E, Lieberherr D et al. 2009 – The UniProtKB/Swiss-Prot knowledgebase and its plant proteome annotation program. Journal of Proteomics 72(3), 567–573. Doi 10.1016/j.jprot.2008.11.010

Sharpton TJ, Stajich JE, Rounsley SD, Gardner MJ et al. 2009 – Comparative genomic analyses of the human fungal pathogens Coccidioides and their relatives. Genome Research 19(10), 1722–1731. Doi 10.1101/gr.087551.108

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV et al. 2015 – BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19), 3210–3212. Doi 10.1093/bioinformatics/btv351

Sipos G, Prasanna AN, Walter MC, O'Connor E et al. 2018 – Genome expansion and lineage-specific genetic innovations in the forest pathogenic fungi *Armillaria*. Nature Ecology & Evolution 2, 577. Doi 10.1038/s41559-017-0347-8

Stamatakis A. 2014 – RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9), 1312–1313. Doi 10.1093/bioinformatics/btu033

Sun L, Fu YH, Yang Y, Wang XX et al. 2019 – Genomic analyses reveal evidence of independent evolution, demographic history, and extreme environment adaptation of Tibetan Plateau *Agaricus bisporus*. Frontiers in Microbiology 10, 1786. Doi 10.3389/fmicb.2019.01786

Tempel S. 2012 – Using and understanding RepeatMasker. Methods in Molecular Biology 859, 29–51. Doi 10.1007/978-1-61779-603-6_2

Tiquia-Arashiro SM, Grube M. 2019 – Fungi in extreme environments: ecological role and biotechnological significance. Springer Cham, Heidelberg, New York.

Varga T, Krizsán K, Földi C, Dima B et al. 2019 – Megaphylogeny resolves global patterns of mushroom evolution. Nature Ecology & Evolution 3(4), 668–678. Doi 10.1038/s41559-019-0834-1

Walker BJ, Abeel T, Shea T, Priest M et al. 2014 – Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9(11), e112963. Doi 10.1371/journal.pone.0112963

Walls C, Breidenbach R. 2005 – Spring in action. Dreamtech Press, New Delhi.

Weber T, Blin K, Duddela S, Krug D et al. 2015 – antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Research 43(W1), W237–W243. Doi 10.1093/nar/gkv437

Widenius M, Axmark D, Arno K. 2002 – MySQL reference manual: documentation from the source. O'Reilly Media, Inc, Boston.

Woychowsky E. 2006– AJAX: creating web pages with asynchronous JavaScript and XML. 1st edition. Prentice Hall, Hoboken, New Jersey.

Wu F, Yuan Y, He SH, Bandara AR et al. 2015 – Global diversity and taxonomy of the *Auricularia auricula-judae* complex (Auriculariales, Basidiomycota). Mycological Progress 14, 95. Doi 10.1007/s11557-015-1113-4

Wu F, Zhou LW, Yang ZL, Bau T et al. 2019 – Resource diversity of Chinese macrofungi: edible, medicinal and poisonous species. Fungal Diversity 98, 1–76. Doi 10.1007/s13225-019-00432-7

Xiao Y, Cheng XJ, Liu J, Li C et al. 2016 – Population genomic analysis uncovers environmental stress-driven selection and adaptation of *Lentinula edodes* population in China. Scientific Repports 6, 36789. Doi 10.1038/srep36789

Yang ZH. 2007 – PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24(8), 1586–1591. Doi 10.1093/molbev/msm088

Yap H-YY, Chooi Y-H, Fau-Firdaus-Raih M, Firdaus-Raih M et al. 2014 – The genome of the tiger milk mushroom, *Lignosus rhinocerotis*, provides insights into the genetic basis of its medicinal properties. BMC Genomics 15, 635. Doi 10.1186/1471-2164-15-635

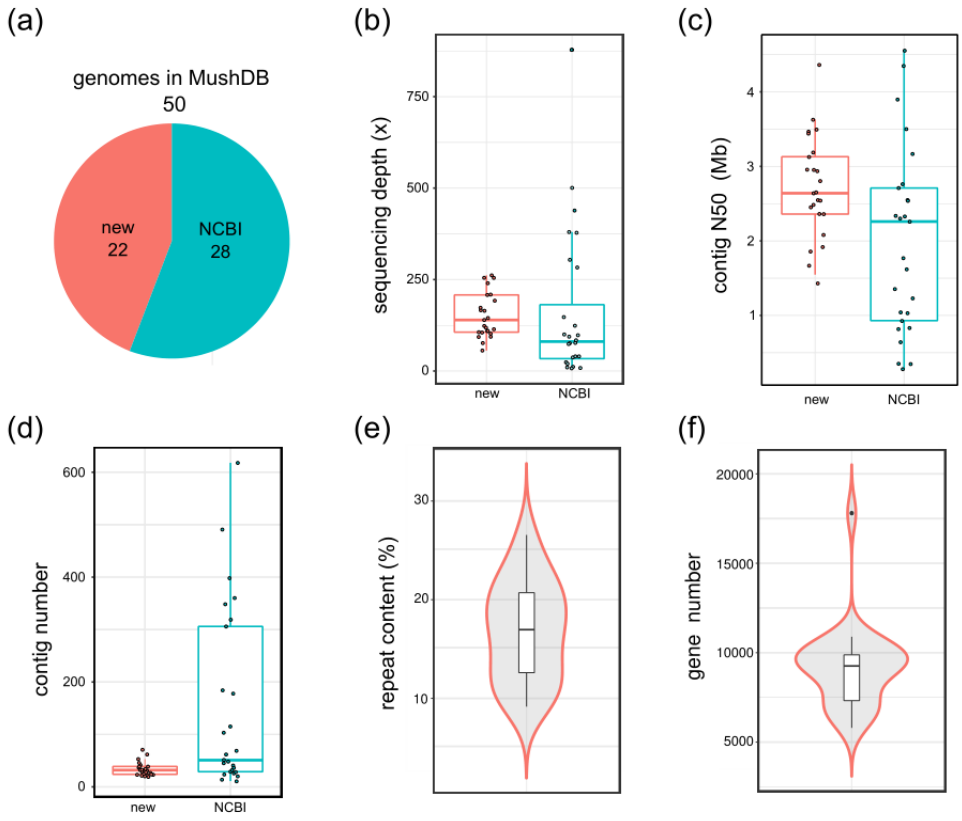
Yu H, Zhang L, Shang X, Peng B et al. 2022 – Chromosomal genome and population genetic analyses to reveal genetic architecture, breeding history and genes related to cadmium accumulation in *Lentinula edodes*. BMC Genomics 23, 120. Doi 10.1186/s12864-022-08325-x

Zhang HM, Hua Y, Li R, Kong DC. 2016 – Cdc24 is essential for long-range end resection in the repair of double-stranded DNA breaks. Journal of Biologic Chemistry 291(48), 24961–24973. Doi 10.1074/jbc.M116.755991

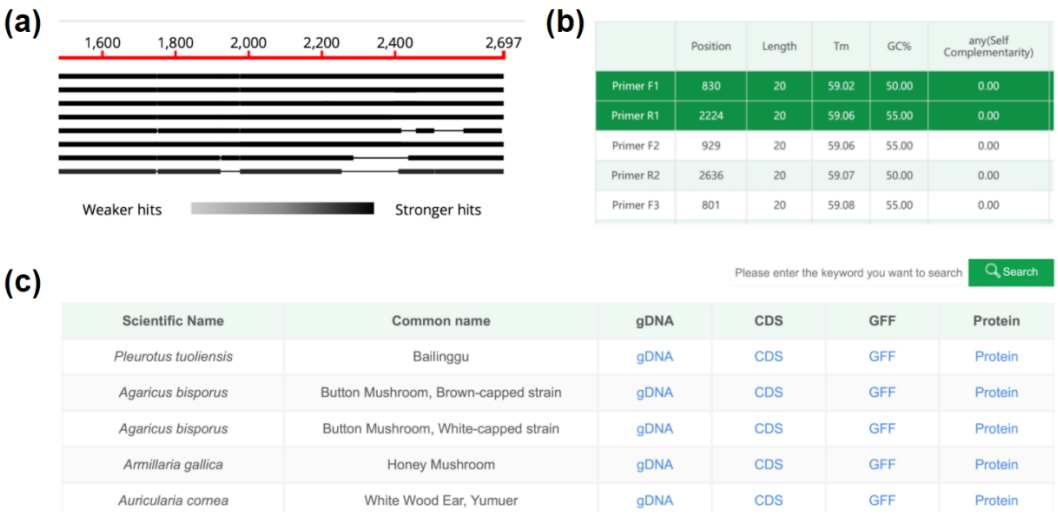
Zhang JC, Shen N, Li C, Xiang XJ et al. 2021 – Population genomics provides insights into the genetic basis of adaptive evolution in the mushroom-forming fungus *Lentinula edodes*. Journal of Advanced Research 38, 91–106. Doi 10.1016/j.jare.2021.09.008

Zhao RL, Li GJ, Sánchez-Ramírez S, Stata M et al. 2017 – A six-gene phylogenetic overview of Basidiomycota and allied phyla with estimated divergence times of higher taxa and a phyloproteomics perspective. Fungal Diversity 84, 43–74. Doi 10.1007/s13225-017-0381-5

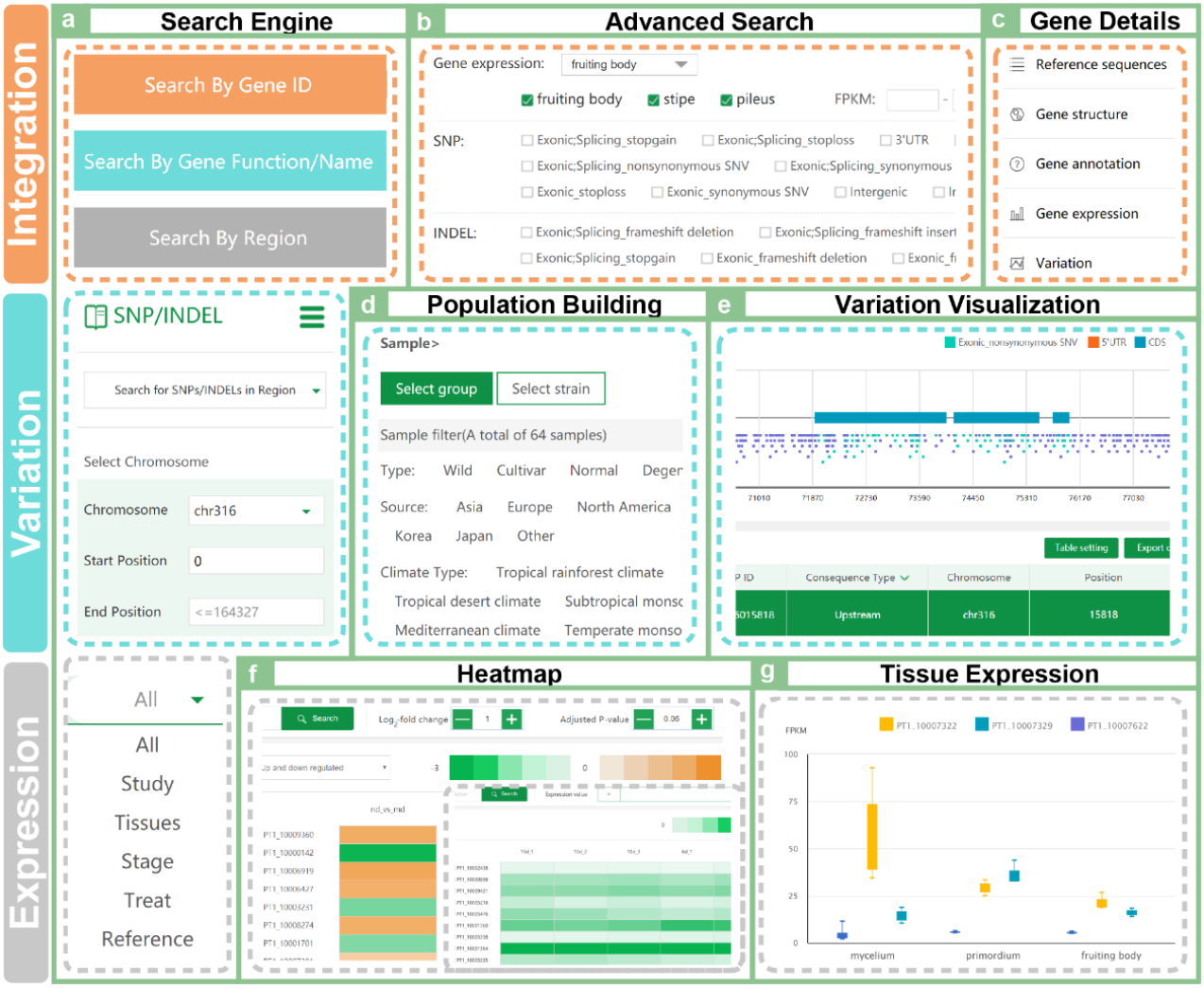
Supplementary materials



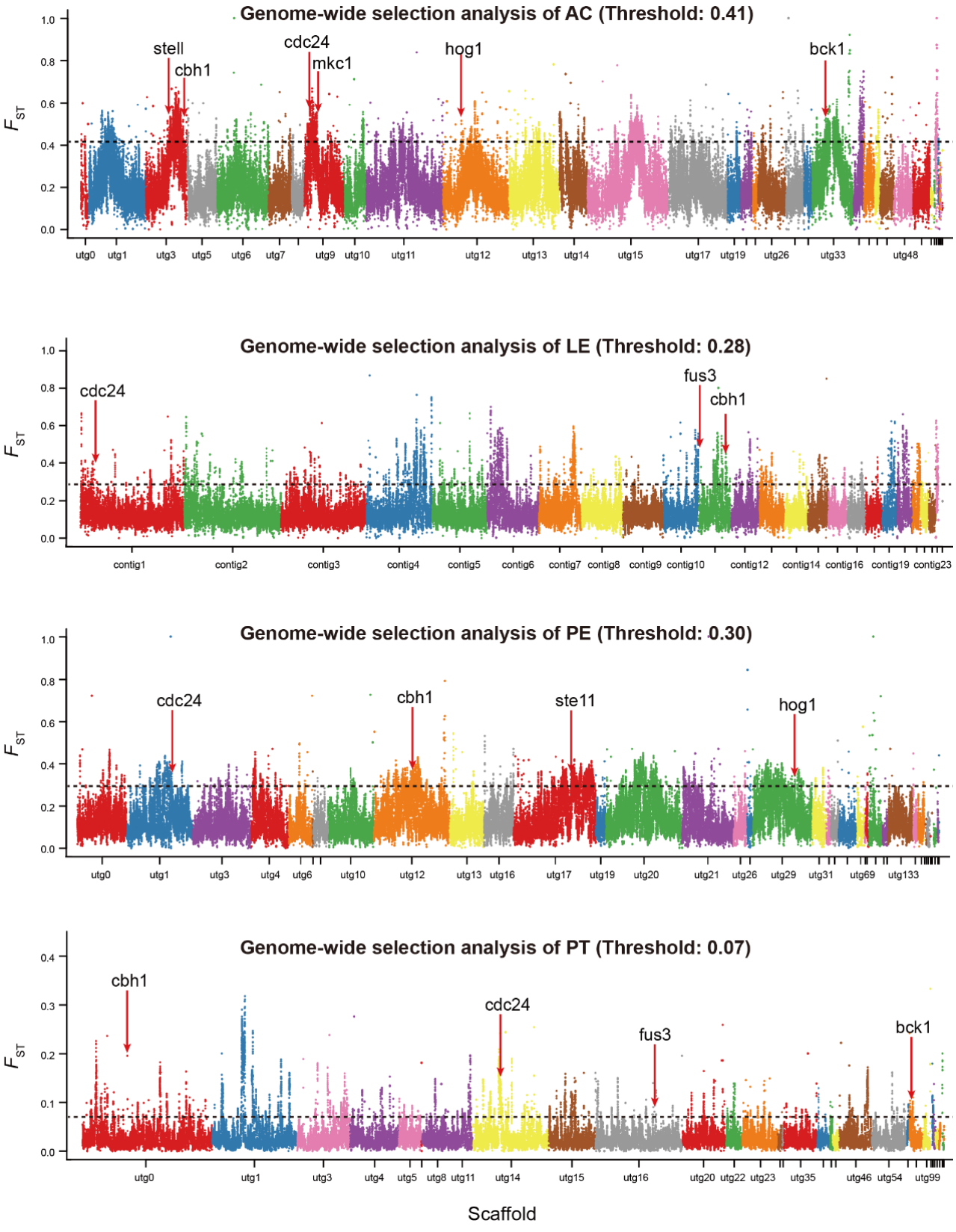
Supplementary Figure 1 – Newly assembled cultivated macrofungal genomes for MushDB, comparing to the previous public genomes. a The number of newly assembled and public genomes in MushDB. b The distribution of sequencing depth for newly assemblies and public genomes in NCBI. c The distribution of contig N50 length for newly assemblies and public genomes in NCBI. d The distribution of contig number for newly assemblies and public genomes in NCBI. e Repeat content distribution for newly assembled genomes. f Gene number annotated for newly assembled genomes.



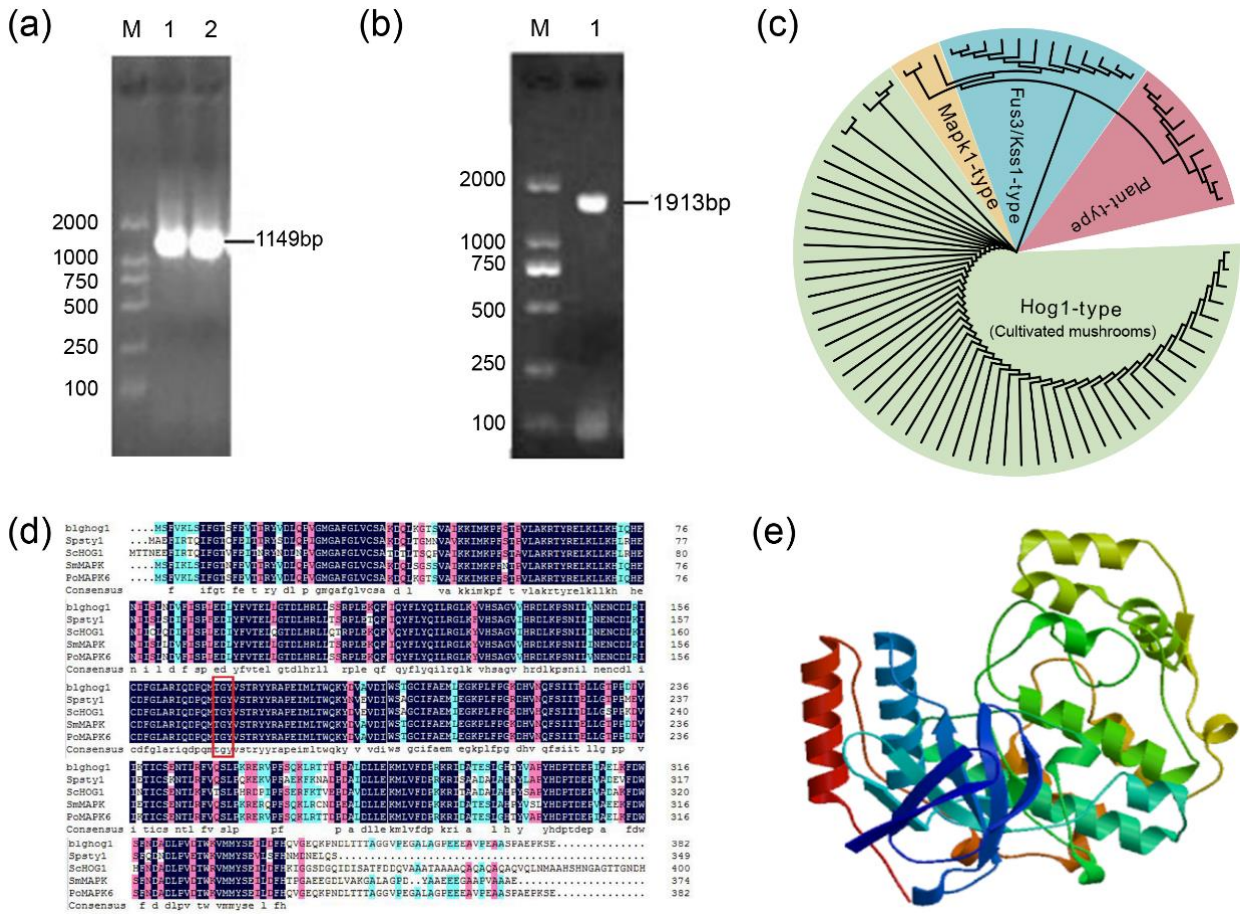
Supplementary Figure 2 – Construction and content of MushDB database. a BLAST engine for sequence alignment. b Primer engine for primer design. c FTP server for download of genomics information.



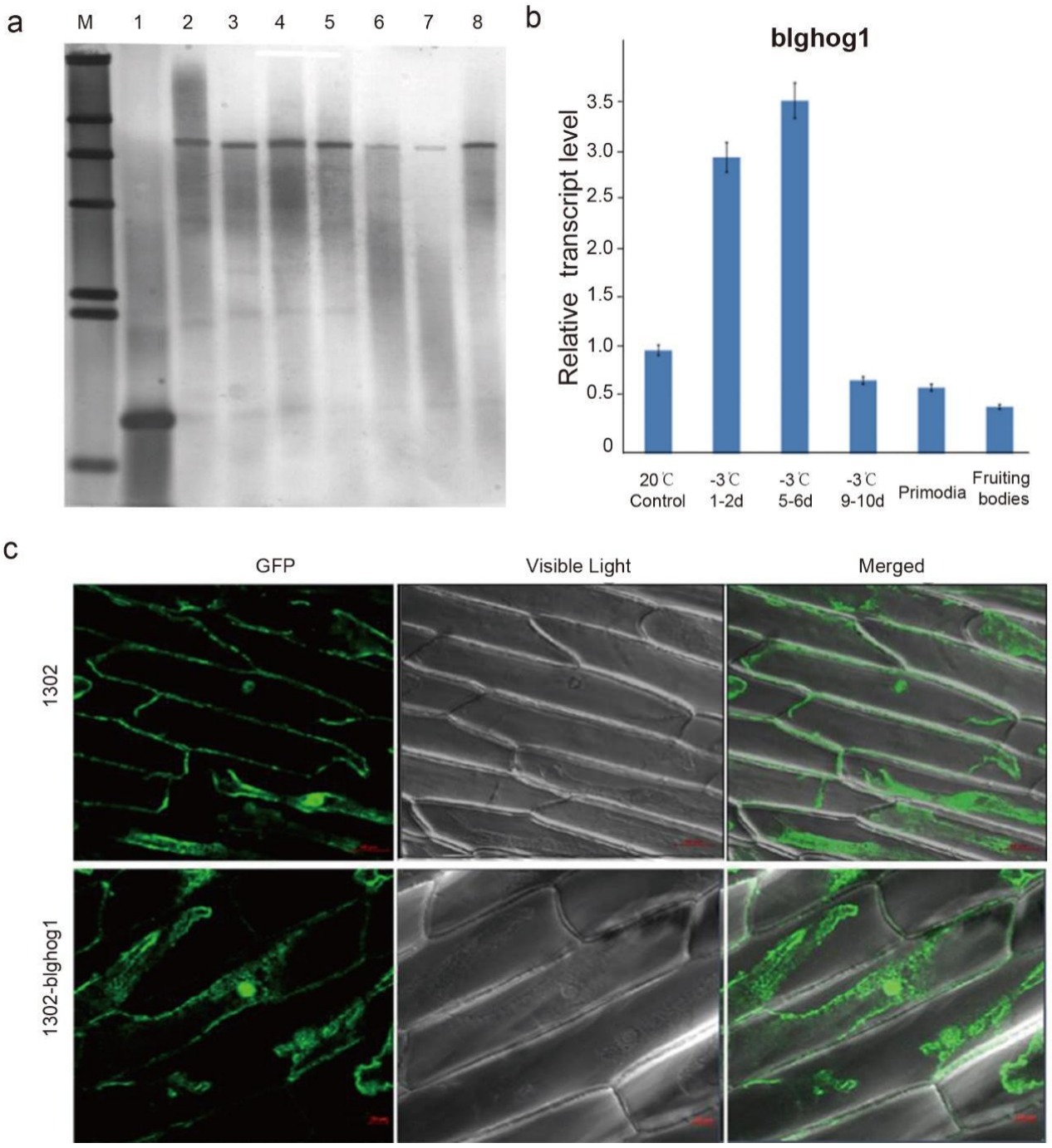
Supplementary Figure 3 – Interface and general functions of MushDB database. a Search engine of MushDB database. Various search modes have been designed according to the characteristics of different data; users can obtain their information of interest flexibly by choosing an applicable mode. b Advanced search engine for Integration database. Users can easily use the information of the basic database to filter the genes of interest. c Summary of the basic gene information. This page included gene sequence, gene structure, gene annotation, gene expression level, and gene mutation for each gene in the macrofungi genome. d A population building module for user to construct one or more populations by sample attributes or sample individuals. e Visualization of variations in a genome region. f Heatmap of gene expression/fold change of differential expressed genes. g Gene expression profile module to display gene expression pattern in different mushroom tissues.



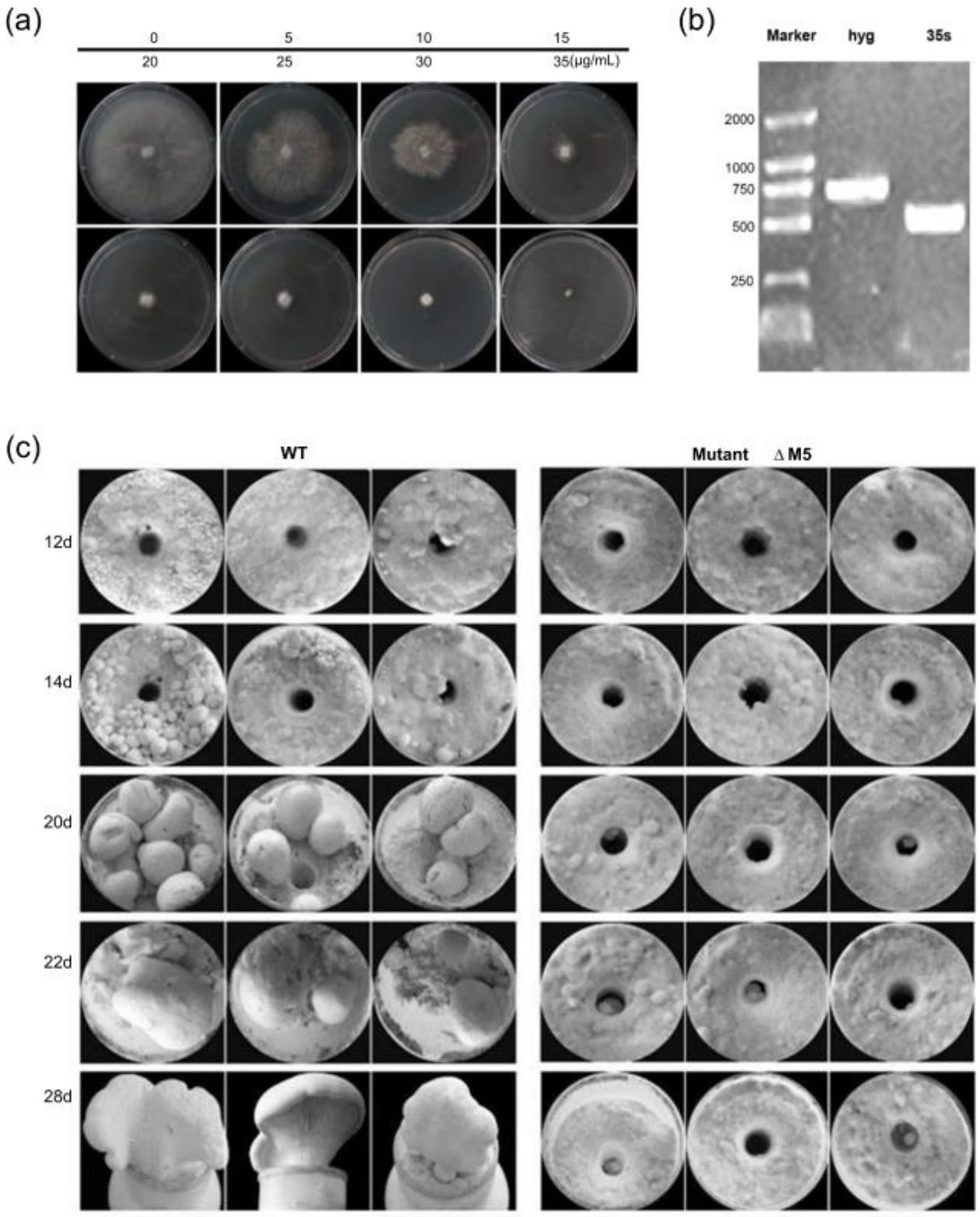
Supplementary Figure 4 – Genome-wide selection analysis of *Auricularia cornea* (AC), *Lentinula edodes* (LE), *Pleurotus eryngii* var. *eryngii* (PE) and *Pl. tuoliensis* (PT) populations. A dashed horizontal line indicates the cut-off (top 5%) used for extracting outliers.



Supplementary Figure 5 – Sequence cloning, functional prediction and evolutionary analysis of *hog1* Gene. a Sequences cloning of *hog1* gene from cDNA. b Sequences cloning of *hog1* gene from DNA. c Evolution analysis of MAPK gene families. d Sequencing alignment of *hog1* gene. The sequence 'TGY' required for kinase activation is marked by red box. *hog1* from *Pleurotus tuoliensis*; *Spsty1* from *Schizosaccharomyces pombe*; *ScHOG1* from *S. cerevisiae*; *PoMAPK1* from *Pl. ostreatus*. e Tertiary structure prediction of *hog1* gene.



Supplementary Figure 6 – Detection of copy number of *hog1* gene and subcellular localization. a Southern blot of *hog1* gene. b qRT-PCR showed that *hog1* gene expression was the highest in hyphae treated at low temperature for 5-6 days. c Subcellular localization of *hog1* gene.



Supplementary Figure 7 – Construction of genetic transformation system of *hog1* gene and fruiting experiment of transformants. a Hygromycin resistance experiment. b The enzyme digestion of knockout vector. c Fruiting experiment.

Supplementary Table 1 BUSCO and CEGMA of new sequenced genomes.

Scientific name	CEGMA (%)	BUSCO (%)
Auricularia cornea (the black-brown strain)	96.0	95.8
Coprinus comatus	96.8	94.5
Cyclocybe chaxingu	94.0	96.0
Cyclocybe parasitica	96.4	93.4
Flammulina filiformis (the white strain)	96.8	93.1
Grifola frondosa	96.0	95.5
Hericium coralloides	94.0	96.2
Hericium erinaceus	96.8	96.6
Hymenopellis raphanipes	94.8	95.5
Pholiota adiposa	97.6	93.7
Pholiota microspora	94.8	93.1
Pleurotus abieticola	97.0	95.0
Pleurotus citrinopileatus	97.6	96.2
Pleurotus djamor	96.0	96.9
Pleurotus floridanus	96.8	94.2
Pleurotus giganteus	96.8	96.6
Pleurotus ostreatus	96.8	95.9
Pleurotus populinus	97.0	95.0
Pleurotus pulmonarius	94.8	91.0
Pleurotus sapidus	97.2	94.8
Pleurotus tuoliensis	94.8	90.4
Stropharia rugosoannulata	97.2	93.4

Supplementary Table 2 Repeat annotation of new sequenced genomes.

Scientific name	DNA	LINE	SINE	LTR	Satellite	Simple_repeat	Other	Unknown	Total
Pleurotus citrinopileatus	1.77	0.15	0.00	7.53	0.01	0.60	0.00	3.91	13.05
Pleurotus djamor	1.73	0.30	0.00	3.93	0.06	0.04	0.00	4.70	10.14
Pleurotus floridanus	1.11	0.57	0.00	6.51	0.01	0.43	0.00	1.79	9.83
Pleurotus giganteus	1.15	0.19	0.04	6.36	0.01	0.12	0.00	3.81	11.39
Pleurotus ostreatus	1.43	0.60	0.00	11.37	0.03	0.03	0.00	4.20	17.34
Pleurotus populinus	1.98	0.40	0.02	16.29	0.22	0.06	0.00	4.21	22.42
Pleurotus pulmonarius	1.94	0.71	0.00	9.33	0.02	0.33	0.00	5.09	16.53

Supplementary Table 2 Continued.

Scientific name	DNA	LINE	SINE	LTR	Satellite	Simple_repeat	Other	Unknown	Total
Pleurotus tuoliensis	4.15	0.77	0.00	16.75	0.30	0.22	0.00	4.87	25.35
Pleurotus abieticola	2.98	0.65	0.01	20.60	0.08	0.37	0.00	2.57	26.43
Auricularia cornea (the black-brown strain)	0.81	2.17	0.02	10.24	0.05	0.70	0.00	9.36	21.90
Coprinus comatus	1.15	0.27	0.01	9.40	0.15	0.05	0.00	1.93	12.47
Cyclocybe chaxingu	1.89	2.79	0.01	12.72	0.03	0.29	0.00	3.96	20.57
Cyclocybe parasitica	2.12	2.65	0.01	12.53	0.04	0.68	0.00	5.96	22.39
Grifola frondosa	2.38	0.29	0.00	8.76	0.08	0.03	0.00	7.41	18.41
Hericium coralloides	2.75	1.11	0.00	12.86	0.20	0.70	0.00	5.43	21.13
Hericium erinaceus	1.47	1.29	0.00	10.65	0.45	0.88	0.00	4.30	16.83
Hymenopellis raphanipes	2.11	0.80	0.33	9.30	0.11	0.36	0.00	7.56	19.35
Pholiota adiposa	1.07	0.76	0.02	7.30	0.01	0.14	0.00	2.80	11.49
Pholiota squarrosa	3.29	0.96	0.00	10.12	0.03	0.06	0.00	5.61	19.42
Pholiota microspora	2.14	2.59	0.00	7.45	0.00	0.03	0.00	3.61	15.46
Stropharia rugosoannulata	0.92	0.48	0.00	12.63	0.00	0.01	0.45	4.26	18.24
Flammulina filiformis (the white strain)	0.40	0.08	0.17	5.30	0.00	0.00	0.04	3.31	9.18

Supplementary Table 3 Gene annotation in new sequenced genomes.

Scientific name	Predicted	Annotated	Percentage	Kegg	Go	Nr	Nt	Swissprot	Pfam	KOG
Pleurotus citrinopileatus	8,143	7,311	89.78%	3,986	3,836	7,272	1,217	3,578	5,096	3,537
Pleurotus djamor	10,143	8,525	84.05%	4,280	4,362	8,428	1,346	3,830	5,798	3,705
Pleurotus floridanus	8,969	8,582	95.69%	4,203	4,041	8,557	1,599	3,703	5,423	3,714
Pleurotus giganteus	9,880	8,453	85.56%	4,233	4,186	8,395	1,354	3,696	5,591	3,640
Pleurotus ostreatus	9,868	9,162	92.85%	4,339	4,164	9,149	1,615	3,830	5,629	3,784
Pleurotus populinus	7,320	7,313	99.90%	4,202	3,851	7,312	1,471	3,685	5,243	3,715
Pleurotus pulmonarius	10,515	9,342	88.84%	4,391	4,213	9,318	1,521	3,875	5,671	3,797
Pleurotus tuoliensis	10,891	10,385	95.35%	5,677	2,314	9,906	9,283	5,577	6,147	3,795
Pleurotus abieticola	9,371	8,696	92.80%	4,238	4,064	8,684	1,502	3,756	5,429	3,721
Auricularia cornea (the black-brown strain)	17,803	16,656	93.56%	7,319	2,510	16,519	16,611	6,916	9,018	5,073
Coprinus comatus	6,936	6,277	90.50%	3,424	3,179	6,238	1,027	2,906	4,206	2,957

Supplementary Table 3 Continued.

Scientific name	Predicted	Annotated	Percentage	Kegg	Go	Nr	Nt	Swissprot	Pfam	KOG
Cyclocybe chaxingu	10,362	8,896	85.85%	4,363	4,284	8,859	1,742	3,792	5,742	3,775
Cyclocybe parasitica	7,248	6,507	89.78%	3,198	2,982	6,486	1,244	2,682	3,932	2,595
Grifola frondosa	9,190	8,721	94.90%	4,500	4,052	8,711	2,243	3,821	5,433	3,811
Hericium coralloides	9,258	7,993	86.34%	4,390	4,168	7,946	2,535	3,743	5,534	3,834
Hericium erinaceus	9,351	8,179	87.47%	4,491	4,304	8,138	2,483	3,852	5,727	3,918
Hymenopellis raphanipes	9,080	8,960	98.68%	5,584	4,969	8,949	1,696	4,761	6,637	4,705
Pholiota adiposa	6,152	6,145	99.89%	4,053	3,554	6,144	1,688	3,604	4,851	3,569
Pholiota squarrosa	5,799	5,796	99.95%	3,907	3,466	5,793	1,613	3,466	4,664	3,465
Pholiota microspora	6,815	6,227	91.37%	3,349	2,992	6,203	1,115	2,826	3,981	2,764
Stropharia rugosoannulata	9,837	8,651	87.94%	4,466	4,137	8,627	1,455	3,988	5,662	3,758
Flammulina filiformis (the white strain)	9,774	8,170	83.59%	4,391	4,111	8,131	1,359	3,681	5,391	3,674

Supplementary Table 4 Statistics of gene predictions in new sequenced genomes.

Scientific name	Average gene length (bp)	Average coding sequence size (bp)	Average number of exons per gene	Average exon size (bp)	Average intron size (bp)
Pleurotus citrinopileatus	2870.06	1450.53	7.56	312.26	77.36
Pleurotus djamor	2716.48	1420.71	7.15	314.63	75.91
Pleurotus floridanus	2540.66	1412.73	7.09	294.78	74.25
Pleurotus giganteus	2627.48	1416.64	7.08	307.99	73.37
Pleurotus ostreatus	2497.64	1394.47	6.96	294.85	74.73
Pleurotus populinus	1874.08	1453.79	6.26	234.24	77.45
Pleurotus pulmonarius	2410.53	1328.14	6.75	287.99	80.98
Pleurotus tuoliensis	1889.09	1389.84	5.72	243.1	105.84
Pleurotus abieticola	2591.14	1387.5	6.78	317.27	76.09
Auricularia cornea (the black-brown strain)	2531.95	1600.67	5.19	308.57	222.4
Coprinus comatus	2671.94	1484.76	6.31	346.73	91.33
Cyclocybe chaxingu	2798.88	1409.07	6.66	348.86	84.02
Cyclocybe parasitica	3326.08	1435.01	6.42	429.71	104.43
Grifola frondosa	2516.13	1330.83	7.07	288.65	78.38
Hericium coralloides	2709.94	1352.35	6.61	338.56	84.27
Hericium erinaceus	2562.39	1426.29	6.54	324.72	78.95

Supplementary Table 4 Continued.

Scientific name	Average gene length (bp)	Average coding sequence size (bp)	Average number of exons per gene	Average exon size (bp)	Average intron size (bp)
Hymenopellis raphanipes	1902.68	1402.79	5.62	262.88	92.22
Pholiota adiposa	1992.12	1516.44	6.45	235.06	87.23
Pholiota squarrosa	2016.03	1506.37	6.63	228.12	89.37
Pholiota microspora	2920.36	1425.93	6.45	373.43	93.91
Stropharia rugosoannulata	2574.83	1381.89	6.6	317.83	84.98
Flammulina filiformis (the white strain)	2137.71	1344.71	5.24	331.39	94.21

Supplementary Table 5 Annotation of ncRNA in new sequenced genomes.

Scientific name	ncRNA		
	tRNA	rRNA	snRNA
Pleurotus citrinopileatus	183	13	17
Pleurotus djamor	189	30	18
Pleurotus floridanus	182	21	16
Pleurotus giganteus	756	38	20
Pleurotus ostreatus	195	32	16
Pleurotus populinus	221	11	17
Pleurotus pulmonarius	181	29	18
Pleurotus tuoliensis	199	47	14
Pleurotus abieticola	210	26	16
Auricularia cornea (the black-brown strain)	167	69	21
Coprinus comatus	250	7	17
Cyclocybe chaxingu	183	11	16
Cyclocybe parasitica	203	5	17
Grifola frondosa	203	5	24
Hericium coralloides	209	7	29
Hericium erinaceus	222	27	29
Hymenopellis raphanipes	482	9	26
Pholiota adiposa	206	46	18
Pholiota squarrosa	262	32	17
Pholiota microspora	206	4	19
Stropharia rugosoannulata	204	37	14
Flammulina filiformis (the white strain)	324	23	22

Supplementary Table 6 The strains information used in AC, LE, PE and PT populations.

Scientific name	Preservation ID	Type	Source
Auricularia cornea	CCMJ2539	Cultivar	China
Auricularia cornea	CCMJ2544	Cultivar	China
Auricularia cornea	CCMJ2545	Cultivar	China
Auricularia cornea	CCMJ2546	Cultivar	China
Auricularia cornea	CCMJ2567	Cultivar	China
Auricularia cornea	CCMJ2829	Cultivar	China
Auricularia cornea	CCMJ2895	Wild	China
Auricularia cornea	CCMJ2896	Wild	China
Auricularia cornea	CCMJ2897	Wild	China
Auricularia cornea	CCMJ2898	Wild	China
Auricularia cornea	CCMJ2899	Wild	China
Auricularia cornea	CCMJ2900	Wild	China
Auricularia cornea	CCMJ2901	Wild	China
Auricularia cornea	CCMJ2902	Wild	China
Auricularia cornea	CCMJ2903	Wild	China
Auricularia cornea	CCMJ2904	Wild	China
Lentinula edodes	cAZ3210	Cultivar	Australian
Lentinula edodes	cgsJD	Cultivar	China
Lentinula edodes	cgsL135	Cultivar	China
Lentinula edodes	cgsL241n4	Cultivar	China
Lentinula edodes	cgsL9015	Cultivar	China
Lentinula edodes	cgsSY10	Cultivar	China
Lentinula edodes	cL241	Cultivar	China
Lentinula edodes	cL939	Cultivar	China
Lentinula edodes	cL9608	Cultivar	China
Lentinula edodes	cQK20	Cultivar	China
Lentinula edodes	cSX16	Cultivar	China
Lentinula edodes	cSX18	Cultivar	China
Lentinula edodes	wCbs1	Wild	China
Lentinula edodes	wCbs1t	Wild	China
Lentinula edodes	wCbs2	Wild	China
Lentinula edodes	wCbs2t	Wild	China
Lentinula edodes	wCbs3	Wild	China
Lentinula edodes	wCbs3t	Wild	China
Lentinula edodes	wLnxy1	Wild	China
Lentinula edodes	wLnxy2	Wild	China
Pleurotus eryngii var. *eryngii*	CCMJ656	Cultivar	China
Pleurotus eryngii var. *eryngii*	CCMJ1725	Cultivar	China
Pleurotus eryngii var. *eryngii*	CCMJ1066	Cultivar	China
Pleurotus eryngii var. *eryngii*	CCMJ1072	Cultivar	China
Pleurotus eryngii var. *eryngii*	CCMJ645	Cultivar	China
Pleurotus eryngii var. *eryngii*	CCMJ1683	Cultivar	China
Pleurotus eryngii var. *eryngii*	CCMJ1707	Cultivar	China
Pleurotus eryngii var. *eryngii*	CCMJ655	Cultivar	China
Pleurotus eryngii var. *eryngii*	CCMJ1691	Cultivar	China
Pleurotus eryngii var. *eryngii*	CCMJ1699	Cultivar	China
Pleurotus eryngii var. *eryngii*	CCMJ1719	Cultivar	China
Pleurotus eryngii var. *eryngii*	CCMJ1730	Cultivar	China
Pleurotus eryngii var. *eryngii*	CCMJ1058	Cultivar	China
Pleurotus eryngii var. *eryngii*	CCMJ2365	Cultivar	China
Pleurotus eryngii var. *eryngii*	CCMJ1678	Cultivar	China
Pleurotus eryngii var. *eryngii*	CCMJ2130	Wild	Europe
Pleurotus eryngii var. *eryngii*	CCMJ2131	Wild	Europe
Pleurotus eryngii var. *eryngii*	CCMJ2132	Wild	Europe
Pleurotus eryngii var. *eryngii*	CCMJ2349	Wild	Pakistan
Pleurotus eryngii var. *eryngii*	CCMJ2133	Wild	Europe

Supplementary Table 6 Continued.

Scientific name	Preservation ID	Type	Source
Pleurotus tuoliensis	CCMJ822	Cultivar	China
Pleurotus tuoliensis	CCMJ1044	Cultivar	China
Pleurotus tuoliensis	CCMJ1077	Cultivar	China
Pleurotus tuoliensis	CCMJ1119	Cultivar	China
Pleurotus tuoliensis	CCMJ1123	Cultivar	China
Pleurotus tuoliensis	CCMJ2366	Cultivar	China
Pleurotus tuoliensis	CCMJ1668	Cultivar	China
Pleurotus tuoliensis	CCMJ1669	Cultivar	China
Pleurotus tuoliensis	CCMJ1670	Cultivar	China
Pleurotus tuoliensis	CCMJ964	Wild	China
Pleurotus tuoliensis	CCMJ965	Wild	China
Pleurotus tuoliensis	CCMJ966	Wild	China
Pleurotus tuoliensis	CCMJ967	Wild	China
Pleurotus tuoliensis	CCMJ968	Wild	China
Pleurotus tuoliensis	CCMJ969	Wild	China
Pleurotus tuoliensis	CCMJ974	Wild	China
Pleurotus tuoliensis	CCMJ975	Wild	China
Pleurotus tuoliensis	CCMJ976	Wild	China
Pleurotus tuoliensis	CCMJ977	Wild	China
Pleurotus tuoliensis	CCMJ978	Wild	China
Pleurotus tuoliensis	CCMJ979	Wild	China
Pleurotus tuoliensis	CCMJ980	Wild	China
Pleurotus tuoliensis	CCMJ981	Wild	China
Pleurotus tuoliensis	CCMJ982	Wild	China
Pleurotus tuoliensis	CCMJ983	Wild	China
Pleurotus tuoliensis	CCMJ987	Wild	China
Pleurotus tuoliensis	CCMJ994	Wild	China
Pleurotus tuoliensis	CCMJ1097	Wild	China
Pleurotus tuoliensis	CCMJ995	Wild	China
Pleurotus tuoliensis	CCMJ996	Wild	China
Pleurotus tuoliensis	CCMJ997	Wild	China
Pleurotus tuoliensis	CCMJ998	Wild	China
Pleurotus tuoliensis	CCMJ999	Wild	China
Pleurotus tuoliensis	CCMJ1000	Wild	China
Pleurotus tuoliensis	CCMJ1001	Wild	China
Pleurotus tuoliensis	CCMJ1002	Wild	China
Pleurotus tuoliensis	CCMJ1003	Wild	China
Pleurotus tuoliensis	CCMJ1160	Wild	China
Pleurotus tuoliensis	CCMJ1161	Wild	China
Pleurotus tuoliensis	CCMJ1162	Wild	China
Pleurotus tuoliensis	CCMJ1163	Wild	China
Pleurotus tuoliensis	CCMJ1165	Wild	China
Pleurotus tuoliensis	CCMJ1166	Wild	China
Pleurotus tuoliensis	CCMJ1167	Wild	China
Pleurotus tuoliensis	CCMJ1168	Wild	China
Pleurotus tuoliensis	CCMJ1169	Wild	China
Pleurotus tuoliensis	CCMJ1170	Wild	China
Pleurotus tuoliensis	CCMJ1171	Wild	China
Pleurotus tuoliensis	CCMJ1172	Wild	China
Pleurotus tuoliensis	CCMJ1173	Wild	China
Pleurotus tuoliensis	CCMJ1175	Wild	China
Pleurotus tuoliensis	CCMJ1176	Wild	China
Pleurotus tuoliensis	CCMJ1679	Wild	China
Pleurotus tuoliensis	CCMJ1682	Wild	China

Supplementary Table 7 Functional gene categories enriched for genes selected in AC population.

Category ID	Category description	Involved gene number	*P* value (< 0.01)
GO:0005623	cell	132	3.37E-09
GO:0044464	cell part	131	3.80E-09
GO:0044424	intracellular part	118	3.01E-08
GO:0005622	intracellular	124	4.58E-08
GO:0043226	organelle	92	2.99E-07
GO:0043229	intracellular organelle	92	2.99E-07
GO:0003774	motor activity	5	2.52E-04
GO:0043231	intracellular membrane-bounded organelle	67	5.62E-04
GO:0043228	non-membrane-bounded organelle	36	5.84E-04
GO:0043232	intracellular non-membrane-bounded organelle	36	5.84E-04
GO:0043227	membrane-bounded organelle	67	8.20E-04
GO:0030684	preribosome	5	8.93E-04
GO:0003777	microtubule motor activity	4	1.56E-03
GO:0009987	cellular process	168	1.75E-03
GO:0005634	nucleus	45	1.95E-03
GO:0016817	hydrolase activity, acting on acid anhydrides	26	2.31E-03
GO:0017111	nucleoside-triphosphatase activity	25	2.51E-03
GO:0044422	organelle part	47	2.85E-03
GO:0044446	intracellular organelle part	47	2.85E-03
GO:0065007	biological regulation	45	2.97E-03
GO:0016462	pyrophosphatase activity	25	3.32E-03
GO:0016818	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	25	3.80E-03
ko03040	Spliceosome	28	3.98E-03
GO:0015698	inorganic anion transport	3	4.00E-03
ko04626	Plant-pathogen interaction	4	5.88E-03
GO:0033180	proton-transporting V-type ATPase, V1 domain	3	1.00E-02
GO:0015935	small ribosomal subunit	5	1.21E-02
GO:0044763	single-organism cellular process	67	1.26E-02
GO:0051651	maintenance of location in cell	3	1.41E-02
GO:0000478	endonucleolytic cleavage involved in rRNA processing	3	1.41E-02
GO:0000479	endonucleolytic cleavage of tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	3	1.41E-02
GO:0006807	nitrogen compound metabolic process	89	1.41E-02

Supplementary Table 7 Continued.

Category ID	Category description	Involved gene number	*P* value (< 0.01)
GO:0050794	regulation of cellular process	34	1.44E-02
GO:0006928	movement of cell or subcellular component	4	1.48E-02
GO:0007018	microtubule-based movement	4	1.48E-02
GO:0007017	microtubule-based process	6	1.49E-02
GO:0031981	nuclear lumen	15	1.57E-02
GO:0034641	cellular nitrogen compound metabolic process	78	1.63E-02
GO:0022626	cytosolic ribosome	4	1.71E-02
ko00500	Starch and sucrose metabolism	26	1.83E-02
GO:0004673	protein histidine kinase activity	2	1.88E-02
GO:0038023	signaling receptor activity	2	1.88E-02
GO:0000155	phosphorelay sensor kinase activity	2	1.88E-02
GO:0060089	molecular transducer activity	2	1.88E-02
GO:0015103	inorganic anion transmembrane transporter activity	2	1.88E-02
GO:0004872	receptor activity	2	1.88E-02
GO:0016775	phosphotransferase activity, nitrogenous group as acceptor	2	1.88E-02
GO:0003984	acetolactate synthase activity	2	1.88E-02
GO:0016805	dipeptidase activity	2	1.88E-02
GO:0009058	biosynthetic process	68	1.94E-02
GO:0046483	heterocycle metabolic process	64	1.97E-02
GO:0050789	regulation of biological process	36	1.97E-02
GO:0030686	90S preribosome	2	2.00E-02
ko00564	Glycerophospholipid metabolism	17	2.05E-02
GO:0016782	transferase activity, transferring sulfur-containing groups	3	2.07E-02
GO:0022627	cytosolic small ribosomal subunit	3	2.25E-02
GO:0004386	helicase activity	10	2.32E-02
GO:0090151	establishment of protein localization to mitochondrial membrane	2	2.53E-02
GO:0000160	phosphorelay signal transduction system	2	2.53E-02
GO:0009129	pyrimidine nucleoside monophosphate metabolic process	2	2.53E-02
GO:0009130	pyrimidine nucleoside monophosphate biosynthetic process	2	2.53E-02
GO:1901607	alpha-amino acid biosynthetic process	8	2.67E-02
GO:0090304	nucleic acid metabolic process	51	2.76E-02
GO:0061024	membrane organization	5	2.98E-02
GO:0044802	single-organism membrane organization	5	2.98E-02

Supplementary Table 7 Continued.

Category ID	Category description	Involved gene number	*P* value (< 0.01)
GO:0051235	maintenance of location	3	3.11E-02
GO:0000462	maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	3	3.11E-02
GO:0000469	cleavage involved in rRNA processing	3	3.11E-02
GO:0061025	membrane fusion	3	3.11E-02
GO:0044801	single-organism membrane fusion	3	3.11E-02
GO:0030490	maturation of SSU-rRNA	3	3.11E-02
GO:0008652	cellular amino acid biosynthetic process	10	3.32E-02
GO:1901576	organic substance biosynthetic process	62	3.93E-02
GO:0000322	storage vacuole	3	4.03E-02
GO:0000323	lytic vacuole	3	4.03E-02
GO:0000324	fungal-type vacuole	3	4.03E-02
GO:0005856	cytoskeleton	9	4.07E-02
GO:0044249	cellular biosynthetic process	62	4.71E-02
GO:0006260	DNA replication	7	4.77E-02
GO:0044391	ribosomal subunit	5	4.89E-02
GO:0044428	nuclear part	18	4.89E-02
GO:0030529	intracellular ribonucleoprotein complex	19	4.93E-02
GO:1990904	ribonucleoprotein complex	19	4.93E-02
GO:0046394	carboxylic acid biosynthetic process	11	4.94E-02

Supplementary Table 8 Functional gene categories enriched for genes selected in LE population.

Category ID	Category description	Involved gene number	*P* value (< 0.01)
ko00380	Tryptophan metabolism	22	9.01E-05
GO:0033036	macromolecule localization	25	5.75E-04
GO:0008104	protein localization	23	9.17E-04
GO:0099568	cytoplasmic region	7	2.28E-03
GO:0005938	cell cortex	7	2.28E-03
GO:0016471	vacuolar proton-transporting V-type ATPase complex	3	4.19E-03
GO:0000145	exocyst	3	4.19E-03
GO:0005622	intracellular	124	4.84E-03
GO:0034613	cellular protein localization	19	5.31E-03

Supplementary Table 8 Continued.

Category ID	Category description	Involved gene number	*P* value (< 0.01)
GO:0043227	membrane-bounded organelle	84	5.40E-03
GO:0015031	protein transport	17	5.63E-03
GO:0071702	organic substance transport	22	6.04E-03
GO:0005787	signal peptidase complex	2	6.10E-03
GO:0000220	vacuolar proton-transporting V-type ATPase, V0 domain	2	6.10E-03
GO:0046983	protein dimerization activity	7	6.18E-03
GO:0044446	intracellular organelle part	58	6.63E-03
GO:0044424	intracellular part	119	6.77E-03
GO:0044422	organelle part	58	6.92E-03
GO:0070727	cellular macromolecule localization	19	7.06E-03
GO:0043231	intracellular membrane-bounded organelle	83	7.13E-03
ko00600	Sphingolipid metabolism	10	7.71E-03
GO:0051601	exocyst localization	2	7.82E-03
GO:0005980	glycogen catabolic process	2	7.82E-03
GO:0044437	vacuolar part	7	8.34E-03
GO:0045184	establishment of protein localization	17	8.43E-03
GO:0004133	glycogen debranching enzyme activity	2	9.45E-03
GO:0004134	4-alpha-glucanotransferase activity	2	9.45E-03
GO:0004135	amylo-alpha-1,6-glucosidase activity	2	9.45E-03
GO:0004042	acetyl-CoA: L-glutamate N-acetyltransferase activity	2	9.45E-03
GO:0031090	organelle membrane	18	1.04E-02
GO:0044444	cytoplasmic part	72	1.10E-02
GO:0008080	N-acetyltransferase activity	4	1.15E-02
GO:1902578	single-organism localization	19	1.20E-02
ko00330	Arginine and proline metabolism	11	1.23E-02
ko00680	Methane metabolism	9	1.35E-02
GO:0005623	cell	125	1.40E-02
GO:0031967	organelle envelope	16	1.49E-02
GO:0031975	envelope	16	1.49E-02
GO:0016469	proton-transporting two-sector ATPase complex	5	1.63E-02
GO:0016410	N-acyltransferase activity	4	1.67E-02
GO:0005934	cellular bud tip	2	1.73E-02
GO:0044464	cell part	124	1.82E-02

Supplementary Table 8 Continued.

Category ID	Category description	Involved gene number	*P* value (< 0.01)
GO:0072599	establishment of protein localization to endoplasmic reticulum	3	1.83E-02
GO:0044275	cellular carbohydrate catabolic process	3	1.83E-02
GO:0045047	protein targeting to ER	3	1.83E-02
GO:0000786	nucleosome	4	1.95E-02
GO:0006465	signal peptide processing	2	2.21E-02
GO:0006560	proline metabolic process	2	2.21E-02
GO:0006561	proline biosynthetic process	2	2.21E-02
GO:0007266	Rho protein signal transduction	2	2.21E-02
GO:0044765	single-organism transport	17	2.21E-02
GO:0005773	vacuole	7	2.52E-02
GO:0051641	cellular localization	21	2.53E-02
GO:0005774	vacuolar membrane	6	2.57E-02
GO:0044281	small molecule metabolic process	37	2.69E-02
GO:0005737	cytoplasm	81	2.71E-02
GO:0070972	protein localization to endoplasmic reticulum	3	2.75E-02
ko00071	Fatty acid degradation	15	2.76E-02
GO:0033177	proton-transporting two-sector ATPase complex, proton-transporting domain	3	2.78E-02
GO:0044448	cell cortex part	5	2.93E-02
ko00561	Glycerolipid metabolism	15	3.04E-02
GO:0015077	monovalent inorganic cation transmembrane transporter activity	7	3.07E-02
GO:0015078	hydrogen ion transmembrane transporter activity	7	3.07E-02
GO:0044815	DNA packaging complex	4	3.13E-02
GO:0005739	mitochondrion	23	3.38E-02
GO:0015672	monovalent inorganic cation transport	7	3.51E-02
ko02010	ABC transporters	10	3.56E-02
GO:0098796	membrane protein complex	10	3.58E-02
ko00053	Ascorbate and aldarate metabolism	5	3.59E-02
GO:0006631	fatty acid metabolic process	4	3.75E-02
GO:0033176	proton-transporting V-type ATPase complex	3	3.75E-02
ko00052	Galactose metabolism	8	3.80E-02
GO:0044710	single-organism metabolic process	66	3.84E-02
GO:0098588	bounding membrane of organelle	10	3.91E-02
GO:0005635	nuclear envelope	6	3.95E-02

Supplementary Table 8 Continued.

Category ID	Category description	Involved gene number	*P* value (< 0.01)
GO:0005789	endoplasmic reticulum membrane	6	3.95E-02
GO:0042175	nuclear outer membrane-endoplasmic reticulum membrane network	6	3.95E-02
GO:0016407	acetyltransferase activity	4	4.01E-02
GO:1901564	organonitrogen compound metabolic process	43	4.03E-02
GO:0098876	vesicle-mediated transport to the plasma membrane	2	4.16E-02
GO:0006893	Golgi to plasma membrane transport	2	4.16E-02
GO:0046037	GMP metabolic process	2	4.16E-02
GO:0009251	glucan catabolic process	2	4.16E-02
GO:0009262	deoxyribonucleotide metabolic process	2	4.16E-02
GO:0009263	deoxyribonucleotide biosynthetic process	2	4.16E-02
GO:0044247	cellular polysaccharide catabolic process	2	4.16E-02
GO:0034067	protein localization to Golgi apparatus	2	4.16E-02
GO:0015914	phospholipid transport	2	4.16E-02
GO:0045053	protein retention in Golgi apparatus	2	4.16E-02
GO:0043226	organelle	90	4.17E-02
GO:0043229	intracellular organelle	90	4.17E-02
GO:0044724	single-organism carbohydrate catabolic process	4	4.66E-02
ko00230	Purine metabolism	20	4.69E-02
GO:0099023	tethering complex	3	4.87E-02
GO:0098805	whole membrane	9	4.88E-02
GO:0016783	sulfurtransferase activity	2	4.96E-02
GO:0004499	N,N-dimethylaniline monooxygenase activity	2	4.96E-02
GO:0090599	alpha-glucosidase activity	2	4.96E-02

Supplementary Table 9 Functional gene categories enriched for genes selected in PE population.

Category ID	Category description	Involved gene number	*P* value (< 0.01)
GO:0009987	cellular process	266	3.53E-07
GO:0005622	intracellular	197	8.80E-06
GO:0044464	cell part	200	1.17E-05
GO:0005623	cell	200	1.92E-05
GO:0009987	cellular process	266	3.53E-07
GO:0005622	intracellular	197	8.80E-06

Supplementary Table 9 Continued.

Category ID	Category description	Involved gene number	*P* value (< 0.01)
GO:0044464	cell part	200	1.17E-05
GO:0005623	Cell	200	1.92E-05
GO:0044424	intracellular part	186	5.15E-05
GO:0006520	cellular amino acid metabolic process	37	8.69E-05
GO:0044237	cellular metabolic process	212	1.11E-04
ko00290	Valine, leucine and isoleucine biosynthesis	10	1.60E-04
GO:0016874	ligase activity	25	1.70E-04
GO:0016070	RNA metabolic process	67	2.07E-04
ko04120	Ubiquitin mediated proteolysis	27	2.31E-04
ko03013	RNA transport	41	3.14E-04
GO:0052689	carboxylic ester hydrolase activity	6	9.33E-04
GO:0044428	nuclear part	35	9.76E-04
GO:0044260	cellular macromolecule metabolic process	137	1.38E-03
GO:0090304	nucleic acid metabolic process	78	1.52E-03
GO:0005524	ATP binding	48	1.57E-03
GO:0032559	adenyl ribonucleotide binding	48	1.57E-03
GO:0030554	adenyl nucleotide binding	48	1.57E-03
GO:0017076	purine nucleotide binding	59	1.60E-03
GO:0030684	preribosome	5	1.78E-03
ko01230	Biosynthesis of amino acids	42	1.89E-03
GO:0031981	nuclear lumen	26	1.95E-03
GO:0032549	ribonucleoside binding	59	2.04E-03
GO:0001882	nucleoside binding	59	2.04E-03
GO:0032550	purine ribonucleoside binding	58	2.44E-03
GO:0032555	purine ribonucleotide binding	58	2.44E-03
GO:0035639	purine ribonucleoside triphosphate binding	58	2.44E-03
GO:0001883	purine nucleoside binding	58	2.44E-03
GO:0006807	nitrogen compound metabolic process	137	2.62E-03
GO:0005730	nucleolus	11	2.89E-03
GO:0006082	organic acid metabolic process	47	3.28E-03
GO:0003723	RNA binding	26	3.71E-03
ko00770	Pantothenate and CoA biosynthesis	11	3.73E-03

Supplementary Table 9 Continued.

Category ID	Category description	Involved gene number	*P* value (< 0.01)
GO:0034660	ncRNA metabolic process	25	3.76E-03
GO:0046483	heterocycle metabolic process	100	3.89E-03
GO:1901605	alpha-amino acid metabolic process	21	4.02E-03
GO:0016875	ligase activity, forming carbon-oxygen bonds	10	4.17E-03
GO:0016876	ligase activity, forming aminoacyl-tRNA and related compounds	10	4.17E-03
GO:0004812	aminoacyl-tRNA ligase activity	10	4.17E-03
GO:0043436	oxoacid metabolic process	46	4.36E-03
GO:0019752	carboxylic acid metabolic process	46	4.36E-03
GO:0016740	transferase activity	77	4.55E-03
GO:0097367	carbohydrate derivative binding	59	5.03E-03
GO:0032553	ribonucleotide binding	58	5.37E-03
ko03040	Spliceosome	33	5.52E-03
GO:0019787	ubiquitin-like protein transferase activity	3	6.96E-03
GO:0004842	ubiquitin-protein transferase activity	3	6.96E-03
GO:0006399	tRNA metabolic process	16	7.52E-03
GO:0016410	N-acyltransferase activity	5	8.41E-03
GO:0034399	nuclear periphery	3	8.62E-03
GO:0000460	maturation of 5.8S rRNA	4	8.78E-03
GO:0000466	maturation of 5.8S rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	4	8.78E-03
GO:0006725	cellular aromatic compound metabolic process	99	9.09E-03
GO:0043038	amino acid activation	10	1.00E-02
GO:0043039	tRNA aminoacylation	10	1.00E-02
GO:0006418	tRNA aminoacylation for protein translation	10	1.00E-02
GO:0003676	nucleic acid binding	51	1.03E-02
GO:1901363	heterocyclic compound binding	128	1.03E-02
GO:0097159	organic cyclic compound binding	128	1.03E-02
GO:0044282	small molecule catabolic process	9	1.05E-02
GO:0050794	regulation of cellular process	57	1.09E-02
GO:1901360	organic cyclic compound metabolic process	101	1.10E-02
ko03022	Basal transcription factors	12	1.14E-02
GO:0099513	polymeric cytoskeletal fiber	5	1.15E-02

Supplementary Table 9 Continued.

Category ID	Category description	Involved gene number	*P* value (< 0.01)
GO:0099512	supramolecular fiber	5	1.15E-02
GO:0099081	supramolecular polymer	5	1.15E-02
GO:0099080	supramolecular complex	5	1.15E-02
GO:0006139	nucleobase-containing compound metabolic process	89	1.22E-02
GO:0008652	cellular amino acid biosynthetic process	18	1.22E-02
GO:0010467	gene expression	78	1.26E-02
GO:0034641	cellular nitrogen compound metabolic process	116	1.28E-02
GO:0005635	nuclear envelope	6	1.30E-02
GO:0043168	anion binding	67	1.38E-02
GO:0016840	carbon-nitrogen lyase activity	4	1.43E-02
ko00660	C5-Branched dibasic acid metabolism	4	1.43E-02
GO:0036094	small molecule binding	88	1.47E-02
GO:0050789	regulation of biological process	60	1.58E-02
GO:0008509	anion transmembrane transporter activity	8	1.75E-02
GO:0043170	macromolecule metabolic process	142	1.77E-02
GO:0043226	organelle	130	1.84E-02
GO:0043229	intracellular organelle	130	1.84E-02
GO:0044281	small molecule metabolic process	62	1.85E-02
ko01210	2-Oxocarboxylic acid metabolism	14	2.00E-02
GO:0006396	RNA processing	25	2.07E-02
GO:0005634	nucleus	66	2.10E-02
GO:0098687	chromosomal region	6	2.16E-02
GO:0000166	nucleotide binding	85	2.22E-02
GO:1901265	nucleoside phosphate binding	85	2.22E-02
GO:0016053	organic acid biosynthetic process	23	2.33E-02
GO:0000049	tRNA binding	3	2.39E-02
GO:0002161	aminoacyl-tRNA editing activity	3	2.39E-02
GO:0019842	vitamin binding	3	2.39E-02
GO:0016841	ammonia-lyase activity	3	2.39E-02
GO:0044267	cellular protein metabolic process	72	2.50E-02
GO:0051345	positive regulation of hydrolase activity	5	2.62E-02
GO:0070647	protein modification by small protein conjugation or removal	5	2.62E-02
GO:1901607	alpha-amino acid biosynthetic process	14	2.66E-02

Supplementary Table 9 Continued.

Category ID	Category description	Involved gene number	*P* value (< 0.01)
GO:0046394	carboxylic acid biosynthetic process	22	2.68E-02
GO:0005488	binding	185	2.69E-02
GO:0032259	methylation	10	2.72E-02
GO:0044422	organelle part	75	2.75E-02
GO:0044446	intracellular organelle part	75	2.75E-02
GO:0022892	substrate-specific transporter activity	21	2.76E-02
GO:0008565	protein transporter activity	4	2.84E-02
GO:0008080	N-acetyltransferase activity	4	2.84E-02
GO:0044248	cellular catabolic process	26	2.86E-02
GO:2000112	regulation of cellular macromolecule biosynthetic process	27	2.88E-02
GO:0065007	biological regulation	66	2.90E-02
GO:0032991	macromolecular complex	78	3.03E-02
GO:0044763	single-organism cellular process	109	3.30E-02
GO:0010556	regulation of macromolecule biosynthetic process	27	3.32E-02
GO:0006383	transcription from RNA polymerase III promoter	3	3.33E-02
GO:0006450	regulation of translational fidelity	3	3.33E-02
GO:0015075	ion transmembrane transporter activity	16	3.36E-02
GO:0043233	organelle lumen	28	3.56E-02
GO:0031974	membrane-enclosed lumen	28	3.56E-02
GO:0070013	intracellular organelle lumen	28	3.56E-02
GO:0005515	protein binding	24	3.59E-02
GO:0000775	chromosome, centromeric region	5	3.61E-02
GO:0005874	microtubule	4	3.63E-02
GO:0042085	5-methyltetrahydropteroyltri-L-glutamate-dependent methyltransferase activity	2	3.65E-02
GO:0008308	voltage-gated anion channel activity	2	3.65E-02
GO:0005244	voltage-gated ion channel activity	2	3.65E-02
GO:0005253	anion channel activity	2	3.65E-02
GO:0004329	formate-tetrahydrofolate ligase activity	2	3.65E-02
GO:0061733	peptide-lysine-N-acetyltransferase activity	2	3.65E-02
GO:0022832	voltage-gated channel activity	2	3.65E-02
GO:0004402	histone acetyltransferase activity	2	3.65E-02
GO:0022836	gated channel activity	2	3.65E-02

Supplementary Table 9 Continued.

Category ID	Category description	Involved gene number	*P* value (< 0.01)
GO:0016744	transferase activity, transferring aldehyde or ketonic groups	2	3.65E-02
GO:0004488	methylenetetrahydrofolate dehydrogenase (NADP+) activity	2	3.65E-02
GO:0034212	peptide N-acetyltransferase activity	2	3.65E-02
GO:0004573	mannosyl-oligosaccharide glucosidase activity	2	3.65E-02
GO:0004826	phenylalanine-tRNA ligase activity	2	3.65E-02
GO:0003871	5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase activity	2	3.65E-02
GO:0050113	inositol oxygenase activity	2	3.65E-02
GO:0008172	S-methyltransferase activity	2	3.65E-02
GO:0031326	regulation of cellular biosynthetic process	27	3.80E-02
GO:0009889	regulation of biosynthetic process	27	3.80E-02
GO:0051171	regulation of nitrogen compound metabolic process	27	3.80E-02
GO:0008380	RNA splicing	9	4.11E-02
GO:0009081	branched-chain amino acid metabolic process	6	4.12E-02
GO:0005666	DNA-directed RNA polymerase III complex	2	4.22E-02
GO:0036452	ESCRT complex	2	4.22E-02
GO:0000815	ESCRT III complex	2	4.22E-02
GO:0030686	90S preribosome	2	4.22E-02
GO:0030687	preribosome, large subunit precursor	2	4.22E-02
GO:0043547	positive regulation of GTPase activity	4	4.25E-02
GO:0009082	branched-chain amino acid biosynthetic process	5	4.34E-02
GO:0044283	small molecule biosynthetic process	26	4.34E-02
GO:0015630	microtubule cytoskeleton	10	4.45E-02
GO:0071704	organic substance metabolic process	209	4.51E-02
GO:0016788	hydrolase activity, acting on ester bonds	22	4.55E-02
GO:0043162	ubiquitin-dependent protein catabolic process via the multivesicular body sorting pathway	2	4.62E-02
GO:0043328	protein targeting to vacuole involved in ubiquitin-dependent protein catabolic process via the multivesicular body sorting pathway	2	4.62E-02
GO:0090503	RNA phosphodiester bond hydrolysis, exonucleolytic	2	4.62E-02
GO:0000469	cleavage involved in rRNA processing	2	4.62E-02
GO:0006891	intra-Golgi vesicle-mediated transport	2	4.62E-02
GO:0000749	response to pheromone involved in conjugation with cellular fusion	2	4.62E-02

Supplementary Table 9 Continued.

Category ID	Category description	Involved gene number	*P* value (< 0.01)
GO:0019236	response to pheromone	2	4.62E-02
GO:0019310	inositol catabolic process	2	4.62E-02
GO:0048015	phosphatidylinositol-mediated signaling	2	4.62E-02
GO:0048017	inositol lipid-mediated signaling	2	4.62E-02
GO:1901616	organic hydroxy compound catabolic process	2	4.62E-02
GO:0046164	alcohol catabolic process	2	4.62E-02
GO:0046174	polyol catabolic process	2	4.62E-02
GO:0071444	cellular response to pheromone	2	4.62E-02
GO:0043632	modification-dependent macromolecule catabolic process	10	4.66E-02
GO:0000375	RNA splicing, via transesterification reactions	8	4.67E-02
GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	8	4.67E-02
GO:0000398	mRNA splicing, via spliceosome	8	4.67E-02
GO:0031323	regulation of cellular metabolic process	31	4.76E-02
GO:0010468	regulation of gene expression	30	4.81E-02
GO:0044238	primary metabolic process	195	4.93E-02

Supplementary Table 10 Functional gene categories enriched for genes selected in PT population.

Category ID	Category description	Involved gene number	*P* value (< 0.01)
GO:0022804	active transmembrane transporter activity	10	1.83E-05
GO:0043227	membrane-bounded organelle	17	3.63E-05
GO:0005622	intracellular	28	8.03E-05
GO:0043231	intracellular membrane-bounded organelle	16	8.95E-05
GO:0044424	intracellular part	27	1.21E-04
GO:0005634	nucleus	12	1.60E-04
GO:0044464	cell part	28	2.84E-04
GO:0005623	cell	28	2.84E-04
GO:0005215	transporter activity	18	6.16E-04
GO:0043492	ATPase activity, coupled to movement of substances	6	8.96E-04
GO:0016820	hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances	6	1.44E-03
GO:0044422	organelle part	12	1.50E-03

Supplementary Table 10 Continued.

Category ID	Category description	Involved gene number	*P* value (< 0.01)
GO:0044446	intracellular organelle part	12	1.50E-03
GO:0022857	transmembrane transporter activity	15	1.60E-03
GO:0042623	ATPase activity, coupled	7	1.64E-03
GO:0043226	organelle	17	1.64E-03
GO:0043229	intracellular organelle	17	1.64E-03
GO:0016887	ATPase activity	8	2.16E-03
GO:0015291	secondary active transmembrane transporter activity	4	2.72E-03
GO:0042626	ATPase activity, coupled to transmembrane movement of substances	5	4.29E-03
GO:0044428	nuclear part	7	5.71E-03
GO:0015399	primary active transmembrane transporter activity	5	6.41E-03
GO:0015405	P-P-bond-hydrolysis-driven transmembrane transporter activity	5	6.41E-03
GO:0030880	RNA polymerase complex	2	7.51E-03
GO:0000428	DNA-directed RNA polymerase complex	2	7.51E-03
GO:0016591	DNA-directed RNA polymerase II, holoenzyme	2	7.51E-03
GO:0055029	nuclear DNA-directed RNA polymerase complex	2	7.51E-03
GO:0061695	transferase complex, transferring phosphorus-containing groups	2	7.51E-03
GO:0032991	macromolecular complex	11	9.20E-03
ko00520	Amino sugar and nucleotide sugar metabolism	19	9.56E-03
GO:0005315	inorganic phosphate transmembrane transporter activity	2	1.01E-02
GO:0004576	oligosaccharyl transferase activity	2	1.01E-02
GO:0004579	dolichyl-diphosphooligosaccharide-protein glycotransferase activity	2	1.01E-02
ko00524	Neomycin, kanamycin and gentamicin biosynthesis	2	1.13E-02
GO:0015075	ion transmembrane transporter activity	10	1.15E-02
GO:0016817	hydrolase activity, acting on acid anhydrides	12	1.19E-02
GO:0022891	substrate-specific transmembrane transporter activity	11	1.80E-02
GO:0042625	ATPase coupled ion transmembrane transporter activity	4	1.86E-02
GO:0022853	active ion transmembrane transporter activity	4	1.86E-02
GO:0019829	cation-transporting ATPase activity	4	1.86E-02
GO:0022892	substrate-specific transporter activity	12	1.99E-02
GO:0016462	pyrophosphatase activity	11	2.04E-02
GO:0017111	nucleoside-triphosphatase activity	11	2.04E-02

Supplementary Table 10 Continued.

Category ID	Category description	Involved gene number	*P* value (< 0.01)
GO:0016818	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	11	2.04E-02
GO:0070013	intracellular organelle lumen	4	2.04E-02
GO:0043233	organelle lumen	4	2.04E-02
GO:0031974	membrane-enclosed lumen	4	2.04E-02
GO:0031981	nuclear lumen	4	2.04E-02
GO:0044433	cytoplasmic vesicle part	2	2.13E-02
GO:0044451	nucleoplasm part	2	2.13E-02
GO:0017069	snRNA binding	2	2.82E-02
GO:0019201	nucleotide kinase activity	2	2.82E-02
GO:0016776	phosphotransferase activity, phosphate group as acceptor	2	2.82E-02
GO:0009987	cellular process	13	2.95E-02
GO:0015078	hydrogen ion transmembrane transporter activity	5	3.52E-02
GO:0098805	whole membrane	3	3.63E-02
GO:0016763	transferase activity, transferring pentosyl groups	3	3.84E-02
GO:0044444	cytoplasmic part	9	3.88E-02
GO:0005654	nucleoplasm	2	4.02E-02
GO:1990234	transferase complex	2	4.02E-02
GO:0044431	Golgi apparatus part	2	4.02E-02
GO:0097708	intracellular vesicle	2	4.02E-02
GO:0031410	cytoplasmic vesicle	2	4.02E-02
GO:0031982	vesicle	2	4.02E-02
ko03040	Spliceosome	18	4.25E-02
GO:0031090	organelle membrane	4	4.28E-02
GO:0015077	monovalent inorganic cation transmembrane transporter activity	5	4.32E-02
GO:0098588	bounding membrane of organelle	3	4.87E-02

Supplementary Table 11 KEGG pathways enriched for selected genes in AC, LE, PE, PT populations.

Pathway ID	KEGG Pathway	Involved gene number	*P* value (< 0.01)
ko00500	Starch and sucrose metabolism	12	2.42E-03
ko04011	MAPK signaling pathway - yeast	23	4.05E-03
ko03010	Ribosome	8	5.23E-03

Supplementary Table 11 Continued.

Pathway ID	KEGG Pathway	Involved gene number	*P* value (< 0.01)
ko04657	IL-17 signaling pathway	5	5.65E-03
ko04139	Mitophagy - yeast	12	2.03E-02
ko04915	Estrogen signaling pathway	5	2.18E-02
ko04371	Apelin signaling pathway	6	2.31E-02
ko04621	NOD-like receptor signaling pathway	6	2.31E-02
ko01210	2-Oxocarboxylic acid metabolism	10	2.68E-02
ko04914	Progesterone-mediated oocyte maturation	8	3.50E-02
ko04217	Necroptosis	6	4.47E-02
ko04550	Signaling pathways regulating pluripotency of stem cells	3	4.53E-02
ko04668	TNF signaling pathway	3	4.53E-02
ko00982	Drug metabolism - cytochrome P450	3	4.53E-02
ko00980	Metabolism of xenobiotics by cytochrome P450	3	4.53E-02
ko04659	Th17 cell differentiation	3	4.53E-02
ko04137	Mitophagy - animal	5	4.93E-02
ko04013	MAPK signaling pathway - fly	5	4.93E-02