

# 利用加权基因共表达网络分析方法挖掘香菇发育不同阶段的关键基因

宋婷婷<sup>1</sup> 沈颖越<sup>1</sup> 徐飞<sup>2</sup> 范丽军<sup>1</sup> 蔡为明<sup>1\*</sup>

①浙江省农业科学院园艺研究所 浙江 杭州 310021

②浙江省农业科学院中心试验室 浙江 杭州 310021

**摘 要:** 香菇是世界产量第二大食用菌,栽培历史悠久。在木屑袋料栽培模式下,香菇发育可以分为菌丝生长期(G)、菌丝褐化期(B)、原基形成期(P)以及出菇期(FB) 4个阶段。褐化期和原基形成期是香菇从营养生长期到生殖生长两个关键发育阶段,对香菇子实体产量和质量至关重要。本研究以3种不同栽培材料为重复,对香菇发育的前3个阶段进行了转录组分析。主成分分析和相似性分析表明,基因随着发育进程的推进,不同栽培基质样本的基因表达特征相似。以菌丝生长阶段的转录本为参照,通过基因差异表达分析,获得与菌丝褐化成熟和原基形成相关的基因,并对这些基因进行GO和KEGG功能富集分析;其次,对9个转录本数据进行加权基因共表达网络分析(WGCNA),分别获得了与菌丝生长、褐化阶段及原基形成各阶段高度相关的黑色、蓝色及黄色基因模块,并利用网络节点分析获得了与菌丝褐化成熟和原基形成得到7个关键基因;最后,结合差异基因和基因模块分析,得到了菌丝生长阶段的17个重要基因、褐化阶段的167个重要基因以及原基形成阶段的67个重要基因。通过多分析手段结合为筛选候选基因提供了更为高效的方法。

**关键词:** 香菇, 加权基因共表达网络分析, 关键基因, 发育相关基因

[引用本文] 宋婷婷, 沈颖越, 徐飞, 范丽军, 蔡为明, 2021. 利用加权基因共表达网络分析方法挖掘香菇发育不同阶段的关键基因. 菌物学报, 40(7): 1761-1775

Song TT, Shen YY, Xu F, Fan LJ, Cai WM, 2021. Hub gene of *Lentinula edodes* at different developmental stages identified by weighted gene coexpression network analysis. Mycosystema, 40(7): 1761-1775

\* Corresponding author. E-mail: caiwm527@126.com

Received: 2020-12-29, accepted: 2021-02-28

# Hub gene of *Lentinula edodes* at different developmental stages identified by weighted gene coexpression network analysis

SONG Ting-Ting<sup>1</sup> SHEN Ying-Yue<sup>1</sup> XU Fei<sup>2</sup> FAN Li-Jun<sup>1</sup> CAI Wei-Ming<sup>1\*</sup>

①Institute of Horticulture, Zhejiang Academy of Agricultural Sciences, Hangzhou, Zhejiang 310021, China

②Central Laboratory, Zhejiang Academy of Agricultural Sciences, Hangzhou, Zhejiang 310021, China

**Abstract:** *Lentinula edodes* ranks second in production of edible mushroom in the world and has a long cultivation history. The development of *L. edodes* can be divided into four stages, growth stage (G), light-induced browning (B), primordial formation (P) and fruiting body development (FB). Mycelial browning and primordial formation are very important stage for transformation of *L. edodes* from vegetative stage to reproductive stage, which are crucial to the yield and quality. Transcription of *L. edodes* in the preceding three development stages was analyzed by using three different culture materials as replicates. Principal component and correlation analysis showed that the expression characteristics of the genes were similar within the same stage. Using transcriptome of growth stage as control, the genes related to mycelial browning maturation and primordial stage were obtained, and these genes were analyzed by GO and KEGG. WGCNA was used to analyze the differentially expressed genes, hub gene and the gene modules related to mycelial browning transformation and primordial formation. As a result, 17 important genes in mycelial growth stage, 167 in browning stage and 67 in primordium formation stage were obtained by combining differential gene and gene module analysis. The combination of multiple analysis provides us with a more efficient method for screening candidate genes.

**Key words:** *Lentinula edodes*, weighted gene coexpression network analysis (WGCNA), hub gene, development-related gene

香菇 *Lentinula edodes* (Berk.) Pegler 隶属于担子菌门、伞菌纲、伞菌目、光茸菌科，通常生长在壳斗科树木倒木或树桩上，是白腐菌中重要的大型食药真菌之一（戴玉成和杨祝良 2008；戴玉成等 2010）。作为世界第二大食用菌，香菇的栽培历史悠久，深受亚洲国家人民喜爱。我国香菇年产量约为 769 万吨，占到全球总产量 70%以上，居世界首位（耿建利 2015）。近几十年来，香菇的栽

培模式不断发展。从上世纪的 60 年代至今，全国以香菇木屑袋料的栽培模式为主。栽培香菇选用的最佳材料是硬质杂木，以阔叶树青冈属的栎木为佳。

我国香菇栽培技术已经相当成熟，作为典型的变温结实型真菌，整个栽培周期可分为 4 个阶段：（1）菌丝生长期（mycelial growth stage, G），即生长初期，菌丝以无性繁殖的方式在菌棒中大量繁殖直到培养基

质被充分利用; (2) 菌丝褐化期(light-induced browning, B), 即菌丝长满基质后菌棒由于光线增加、打孔通氧、温度变化及营养成分的变化等环境因素刺激进入生殖发育阶段, 在该阶段, 菌棒表面会逐渐出现绒毛状菌丝, 这类菌丝倒伏聚集形成菌膜(被称为人工树皮), 同时分泌褐色色素, 使菌膜呈深褐, 这是菌丝成熟的标志; (3) 原基形成期(primordial formation stage, P), 通过菌棒排场、降低环境温度、增加通风和光照, 促进菌丝纽结形成原基, 进入子实体发育阶段; (4) 出菇期(fruiting body development, FB), 即子实体形成阶段, 历时 6 个月(张锋 2008; Tang *et al.* 2013)。香菇的褐化期和原基形成期是香菇生长发育中至关重要的两个阶段, 在香菇生产中褐化期和原基形成期的管理是香菇品质和产量高低的关键阶段, 两个阶段均受到环境因素的诱导和刺激, 是研究香菇生长发育过程的两个重要时期。

伴随着高通量测序技术的完善, 大量物种全基因组测序陆续完成。在海量基因信息被获得的同时, 大量未知基因需要进行分类和功能鉴定。转录组测序技术(RNA-seq)的应用为此提供了最为便捷、有效的途径。同基因芯片技术一样, RNA-seq 是基于第二代高通量测序技术的基因表达鉴定技术, 它以全基因组序列为背景, 可以获得更丰富的信息。加权基因共表达网络分析(weighted gene coexpression network analysis, WGCNA) 是一种分析共表达基因模块(coexpressed modules) 和关键基因(hub gene)的有效方法(Langfelder & Horvath 2008)。基于配对关联分析(pairwise correlations)的方法, 有着同样表达趋势的相关基因利用该方法可以被分到 1 个组或 1 个模块。WGCNA 已经被成功地应用到植物中, 用于识别一些调控网络和对植物性状调

控的关键基因(Du *et al.* 2017; Garg *et al.* 2017); 在医学领域, 它已被应用于人类肝细胞癌变过程中关键通路和基因的动态变化分析等多个领域(Yin *et al.* 2018)。在真菌中, Li *et al.* (2019) 利用 WGCNA 方法分析了蛹虫草的发育进化过程。近年来, 大量食用菌的基因组也完成了测序(Chen *et al.* 2016; Zhou *et al.* 2016), 尤其是香菇基因组的测序的完成, 为研究食用菌的发育生物学提供了良好的遗传学背景, 这为食用菌发育转录本的研究提供序列基础。

以 3 种不同培养基质为实验重复, 对香菇品种浙香 6 号从香菇菌丝到原基形成的 3 个重要阶段(G、B 和 P)的样品进行转录组分析。基于 WGCNA 方法, 获得了香菇由菌丝生长期、褐化期和原基阶段关键基因模块, 这为我们对香菇生长发育的基本网络有了一个认识, 为如何改善和提高人工栽培香菇的条件提供了有利的分子生物学基础。

## 1 材料与方法

### 1.1 材料

**1.1.1 供试菌株:** 供试香菇菌株 Le949 来自浙江省农业科学院园艺研究所内部材料, 该菌株是浙香 6 号多孢自交的 1 个菌株。供试菌株接种于 PDA 改良培养基中 25℃ 培养 15d 活化, 用于后续实验的分析和使用。

**1.1.2 培养基质配方:** PDA 培养基配方: 200g 马铃薯洗净去皮切碎, 加水 1 000mL 煮沸 30min, 纱布过滤获得马铃薯浸提液, 再加 20g 葡萄糖和 20g 琼脂。香菇 3 种主材料代料栽培配方: 杂木屑/桑枝屑/棉杆 78%、麸皮 20%、石膏粉 1%、红糖 1%, 按 1:1.2 的料水比将上述原料拌匀, 采用  $\phi 15\text{cm} \times 55\text{cm}$  塑料袋装料。每袋装干料约 900g, 高压锅灭菌,

冷却后接种 3 点接种。

**1.1.3 样品的取样阶段:** 接种后的菌丝在培养室中生长 90–100d。由于前 10d 菌丝量少无法取样, 故从 20d 开始, 每隔 10d 在距离菌棒接种口 5cm 的范围内进行取样, 保证取样的均一性。取样点为 3 个, 取样后的样品立即冷冻干燥, 处理好的菌丝和基质混合物置于 -80℃ 冰箱内, 将菌丝生长定殖阶段的 20、30、40d 的菌丝和基质混合物和菌丝生长 70、80、90d 的菌丝和基质混合物进行混样作为菌丝生长阶段和菌丝褐化阶段的样品, 以及排场后 130d 取样的原基作为原基形成阶段的样品 (图 1), 进行后续 RNA 的提取和文库构建。

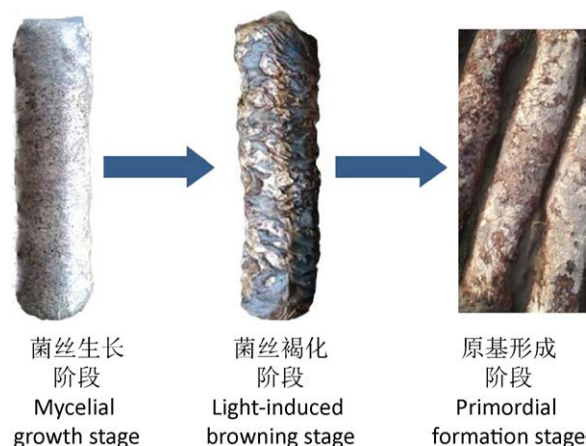


图 1 转录组分析的香菇发育的 3 个取样阶段示意图

Fig. 1 Three sampling stage for transcriptome analysis of *Lentinula edodes* development.

### 1.2 RNA 提取、文库构建和测序

总 RNA 使用 E.Z.N.A.<sup>TM</sup> 植物 RNA 试剂盒 (Omega, Stamford, CT, USA) 提取。分别采用 Nanodrop 和 Qubit 2.0 荧光 (Life Technologies, CA, USA) 以及 Agilent 2100 (Agilent Technologies, CA, USA) 方法检测 RNA 样品的纯度、浓度和完整性, 以保证使

用合格的样品进行转录组测序。样品检测合格后, 进行文库构建, 主要流程如下: 用带有 Oligo(dT) 的磁珠富集真核生物 mRNA; 加入 fragmentation buffer 将 mRNA 进行随机打断; 以 mRNA 为模板, 用 6 碱基随机引物 (random hexamers) 合成第 1 条 cDNA 链, 然后加入缓冲液、dNTPs、RNase H 和 DNA polymerase I 合成第 2 条 cDNA 链, 利用 AMPure XP beads 纯化 cDNA; 纯化的双链 cDNA 再进行末端修复、加 A 尾并连接测序接头, 然后用 AMPure XP beads 进行片段大小选择; 最后通过 PCR 富集得到 cDNA 文库。文库构建完成库检合格后, 用 HiSeq2500 进行高通量测序, 测序读长为 PE100。

### 1.3 基因比对与差异表达分析

基因组序列 (LDAT00000000) 以及注释信息从网站 <https://genome.jgi.doe.gov/portal/> 获得。该基因组信息有 340 个拼接片段 (>2kb, N50≈4.1Mb), 大小为 41.83Mb, 并预测获得 14 889 个蛋白质编码基因 (Chen *et al.* 2016)。将每个样本的转录本使用 HISAT2 (V2.5.5) 比对到参考基因组上 (Trapnell *et al.* 2009; Kim *et al.* 2015)。基因表达量计算采用 FPKM (fragments per kilobase of transcript per million mapped reads) 方法, 即每百万 reads 中比对到某一基因每千碱基长度的 reads 数目, 是转录组测序数据分析中常用的基因表达水平估算方法。差异表达分析 (differential expression analysis) 使用 DESeq2 r 软件包 (1.16.1) 对差异表达基因 (differentially expressed gene, DEG) 进行分析 (Benjamini & Hochberg 1995)。最终在筛选过程中, 将 FDR<0.05 且差异倍数 FC (fold change) ≥1 作为筛选标准。富集分析用于基因的本体论 (gene ontology, GO) 功能注释和代谢通路 (kyoto encyclopedia of genes and genomes, KEGG) 注释分析。

1.4 DEGs 的功能注释

差异表达基因的基因本体论 (gene ontology, GO) 功能注释富集分析由 R 包实现, GO 的注释分为生物过程、分子功能和细胞组成 3 个部分, 该分析过程中基因的长度被矫正为相同长度。群集分析器 R 用来测试统计学上差异的 KEGG 途径中表达的基因。

1.5 构建基因表达网络和候选关键基因的捕获

利用 Rstudio 中的 WGCNA 包 (Langfelder & Horvath 2008) 进行加权共表达网络构建。首先, 将所有基因通过  $\log_2(\text{FPKM}+1)$  进行表达矩阵标准化, 计算每个基因的中位数绝对偏差 (median absolute deviation, MAD), 选取所有基因进行 WGCNA 分析。使用 pickSoftThreshold 计算合适的加权系数  $\beta$ , 以  $\beta=28$  作为权重值, 通过邻接矩阵计算拓扑重叠矩阵, 并采用动态树切割算法对模块进行分割, 按照默认参数设置 (Chin et al.

2014)。使用 9 个样本的分组矩阵作为性状进行特异性模块鉴定。

2 结果与分析

2.1 转录组测序统计

为了阐述香菇在菌丝生长、菌丝褐化成熟和原基形成 3 个发育阶段的基因表达模式, 本研究构建了 3 个发育时期在 3 种栽培基质上的 9 个转录组文库, 9 个样品的数据产出统计见表 1。在本次分析中, 9 个样本通过质量检测和过滤, 共获得 531.7million 的 cleanreads, 平均每个样本约 6.5Gb 的 cleandata。各样品 Q30 碱基百分比均不小于 86.06%。利用软件将 cleanreads 与参考基因组进行序列比对, 获得测序样品特有的序列特征信息及在参考基因组或基因上的位置信息。根据比对结果, 各样品的 reads 与香菇指定参考基因组的比对效率在 80.1%–88.51%。

表 1 样品测序数据与参考基因组的序列比对结果统计

Table 1 Statistics of experimental data and alignment with reference on genome sequence

样品 Samples	总片读数 Total reads	比对率 Mapping rate (%)	匹配基因 Mapped to gene (%)	GC 含量 GC content (%)
菌丝生长阶段 CS/Growth_CS	56133768	87.41	97.56	47.75
菌丝生长阶段 CW/Growth_CW	53016146	87.74	97.95	47.91
菌丝生长阶段 MB/Growth_MB	54764330	87.31	97.69	47.60
菌丝褐化阶段 CS/Browning_CS	56477388	80.12	88.83	47.9
菌丝褐化阶段 CW/Browning_CW	52941312	85.47	80.785	46.57
菌丝褐化阶段 MB/Browning_MB	57353856	83.08	90.34	48.00
原基形成阶段 CS/Primordia_CS	53871596	88.49	97.13	48.73
原基形成阶段 CW/Primordia_CW	50055710	88.73	97.48	48.77
原基形成阶段 MB/Primordia_MB	57419516	88.51	96.65	48.72

Note: Growth: Mycelial growth stage; Browing: Light-induce browning stage; Primordia: Primordial formation stage; CS: Cotton straw; CW: Cyclobalanopsis sawdust; MB: Mulberry sawdust. The same below.

## 2.2 基于 9 个转录本的聚类分析和相关性分析

利用相关性分析,对 9 个转录本的样品进行分析。结果显示不同发育阶段的 3 个样品的聚类要显著性高于不同材料的样本(图 2A),说明后续分析的可信度高。其中原基形成阶段的转录本的相关性最好,相关系数为 0.99,菌丝生长阶段的相关性次之,相关系数为 0.82,褐化阶段的 3 个重复样品的相关性比较低,相关系数为 0.78,但均显著高于组间相关性( $P<0.05$ )。

主成分分析显示,3 个阶段不同栽培的 9 个转录组样本,被有效地分为 3 个组别,即菌丝生长期、菌丝褐化期和原基形成期(图 2B),PC1 和 PC2 总的贡献率高达 68.3%。这一结果说明不同发育阶段对转录组的影响远远大于不同栽培基质对转录组的影响。菌丝褐化期在 3 种培养材料上的差异最大,而在原基形成阶段不同栽培材料对相关原基的转录组影响最小,菌丝生长阶段不同材料对转录组的影响居中,在虫草的发育转录组学的报道中科研人员发现 DF 阶段、YF 阶段以及 MF 阶段和 PR 阶段的表达模式更为相似,这与本研究的结果有着相似性,推测大型食用菌可能有着相似的发育模式(Li *et al.* 2019)。

## 2.3 三个发育阶段差异基因的分析

以聚类分析结果为依据,将不同培养基质同一阶段的香菇转录组作为发育阶段的组内重复,基于 FPKM 数值,以校正  $P<0.05$  和差异倍数大于 1 为阈值,鉴定 3 个发育阶段的差异表达基因(DEGs)。在菌丝生长阶段和原基形成阶段之间,差异基因数量为 3 574 个,菌丝褐化阶段和原基形成阶段的差异基因数为 3 609 个,这两个阶段的差异基因总数为占香菇基因组的 31.9%(图 3)。比较菌丝褐化和原基形成阶段,发现 DEGs 明显减少。这

些结果表明,菌丝营养生长阶段与菌丝褐化阶段和原基形成阶段在表达上存在非常大的差异,而菌丝褐化阶段和原基形成阶段表达差异比较小。

## 2.4 差异基因在不同发育阶段的功能富集分析

对上述 3 个不同阶段的差异基因进行功能注释和富集分析,以菌丝阶段的转录组作为参照分别对菌丝褐化成熟阶段的转录组和原基形成的转录组进行比较。分析发现,与菌丝发育阶段相比,菌丝褐化阶段和原基形成阶段分别有 814 个和 1 418 个基因出现了上调表达,其中 COG 注释结果显示在褐化阶段上调表达的基因主要集中在氨基酸代谢、次级产物合成、转运与分解等功能类别里;在原基形成阶段上调表达的基因主要参与了氧化胁迫调控、细胞膜转运、有丝分裂等功能类别,而两阶段下调基因主要参与蛋白、脂质体和碳水化合物等大分子转运和代谢,细胞壁和膜的合成相关基因,防御机制相关基因以及次级代谢物合成性关基因。差异基因的功能富集结果与本研究观察到的生物学现象相吻合,即在菌丝褐化阶段菌丝最明显的变化是形成表面褐色分泌物,而原基形成阶段主要是与菌丝扭结和菌丝分泌孢外物质为基础。GO 富集结果显示菌丝褐化和原基形成期相关的基因主要参与代谢过程、催化活性等 14 项生物学过程,而 KEGG 富集分析则表明褐化阶段和原基阶段的差异基因主要参与脂肪酸合成、氨基酸与叶酸生物合成等途径(图 4)。

## 2.5 权重基因共表达网络构建

本研究基于 WGCNA 对 9 个转录本的所有表达基因进行聚类分析(剔除缺失值剩 11 814)。根据基因表达量的相关性构建基因聚类树(图 5A),树的 1 个分支对应一表达量高度相关的基因集。利用动态剪切树法对分支进



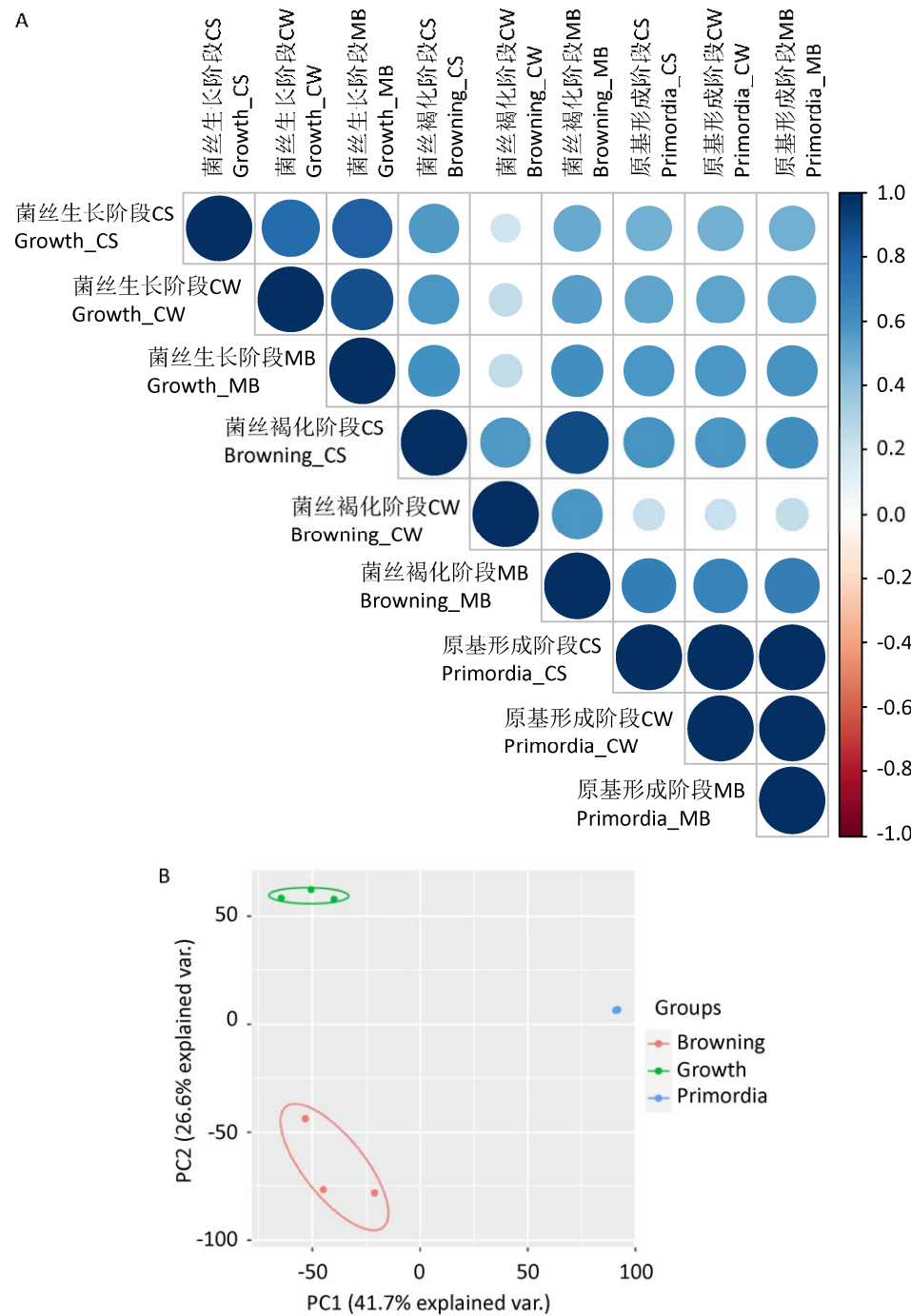


图 2 不同阶段不同栽培材料 9 个转录本的相关性分析 (A) 与聚类分析 (B)

Fig. 2 Correlation analysis (A) and cluster analysis (B) of 9 transcriptomes from different cultivation materials under different stages.

行剪切区分，产生不同的模块，根据模块相似度高于 0.75 则对表达模式相似的模块进行合并后的模块划分，最终获得 14 个共表达模块（图 5B），每种颜色代表一个模块，灰色的模块代表无法归入任何 1 个模块的基因。其中基因数目最多的前 5 个模块分别是青绿色（turquoise）模块、蓝色（blue）模块、棕色（brown）模块、黄色（yellow）模块以及绿色

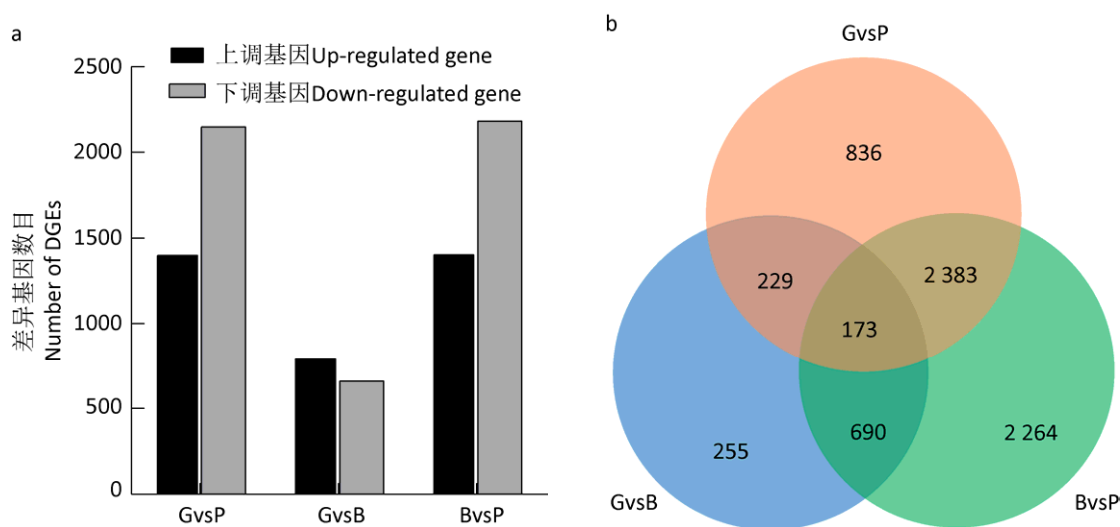


图 3 菌丝生长阶段 (G)、菌丝成熟褐化阶段 (B) 以及原基形成阶段 (P) 的差异基因两两比较分析 (a) 以及差异基因维恩图 (b)

Fig. 3 Differentially expressed gene (DEG) analysis (a) and the venn diagram for DGEs (b) at three stages: mycelial growth stage (G), light-induced browning stage (B), and primordial formation stage (P).

(green) 模块, 棕褐色 (tan) 模块和肉桂粉色 (salmon) 所含基因较少, 分别为 96 个和 86 个 (图 5B)。

利用不同色块基因与 3 个阶段的表型进行相关性分析 (图 5B), 鉴定到黑色模块的基因与菌丝生长阶段相关性高, 可能是菌丝生长期的关键基因。黑色模块中的 361 个基因中有 31 个基因与细胞膜质的形成有关, 有 10 个水解酶基因在这个模块中被聚类, 还有 13 个核酸结合蛋白被聚类到红色模块。蓝色模块基因与香菇菌丝褐化成熟阶段高度相关。对蓝色模块中的 1383 个基因进行进一步注释分析, 发现菌丝褐化阶段除了一些与菌丝细胞的生长有关的基因外, 聚类了与色素形成相关的基因 66 个 (占蓝色基因块的 4%), 其中 P450 单氧化酶类 23 个; 聚类了 40 个与脂肪合成代谢相关的基因; 16 个与信号转导相关的蛋白激酶存在于蓝色模块中。青绿色模块和黄色模块的基因与原基的形成显著相关。黄色模块和青绿色模块分别有 1022 和 5235 个基因, 基因涉及了信号

转导、翻译后修饰以及分泌蛋白等, 值得注意的是以黄色模块为例, 聚类到的 1022 个基因中, 有 486 个基因是未知蛋白, 比例达到 47.5%。

选择与菌丝褐化相关的蓝色模块和与子实体形成相关的黄色模块中权重值大于 0.5 的 200 个连接进行可视化, 并将模块内 degree 的值大于 10 的基因定义为枢纽基因用粉红色方型标记出 (图 6 和表 2)。

与褐化相关的 3 个关键基因有 1 个分泌蛋白 (oxidoreductase activity, GO:0016491)、1 个水解蛋白和 1 个水解酶, 这与在栽培过程中观察到香菇菌丝成熟转色分泌褐色物质的表观结果相一致; 而与原基形成的 4 个关键基因中 3 个均为位置蛋白, 已知的 LE01Gene13703 是核酸复制和修复的重要蛋白, 该基因 GO 富集到了分子功能的核酸绑定 (nucleic acid binding, GO:0003676)、解螺旋酶活性 (helicase activity, GO:0004386) 以及 ATP 绑定活性的途径 (ATP binding molecular function, GO:0005524)。



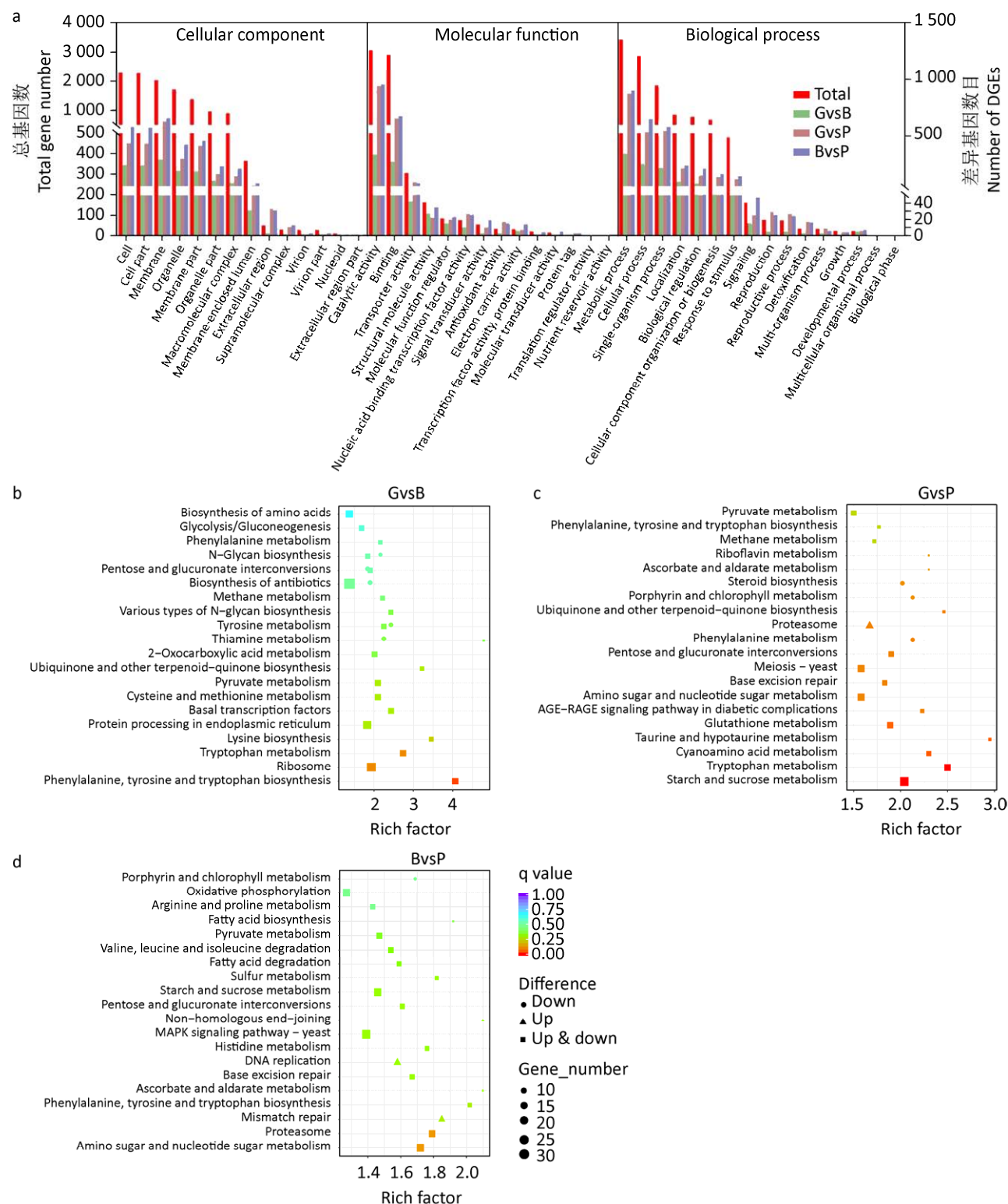


图 4 菌丝生长阶段 (G)、菌丝成熟褐化阶段 (B) 以及原基形成阶段 (P) 两两比较差异基因的 GO 分析 (a) 和 KEGG 分析 (b-d)

Fig. 4 GO functional classification (a) and KEGG analysis (b-d) of differentially expressed genes at three stages by group pairwise comparison: mycelial growth stage (G), light-induced browning stage (B), and primordial formation stage (P).

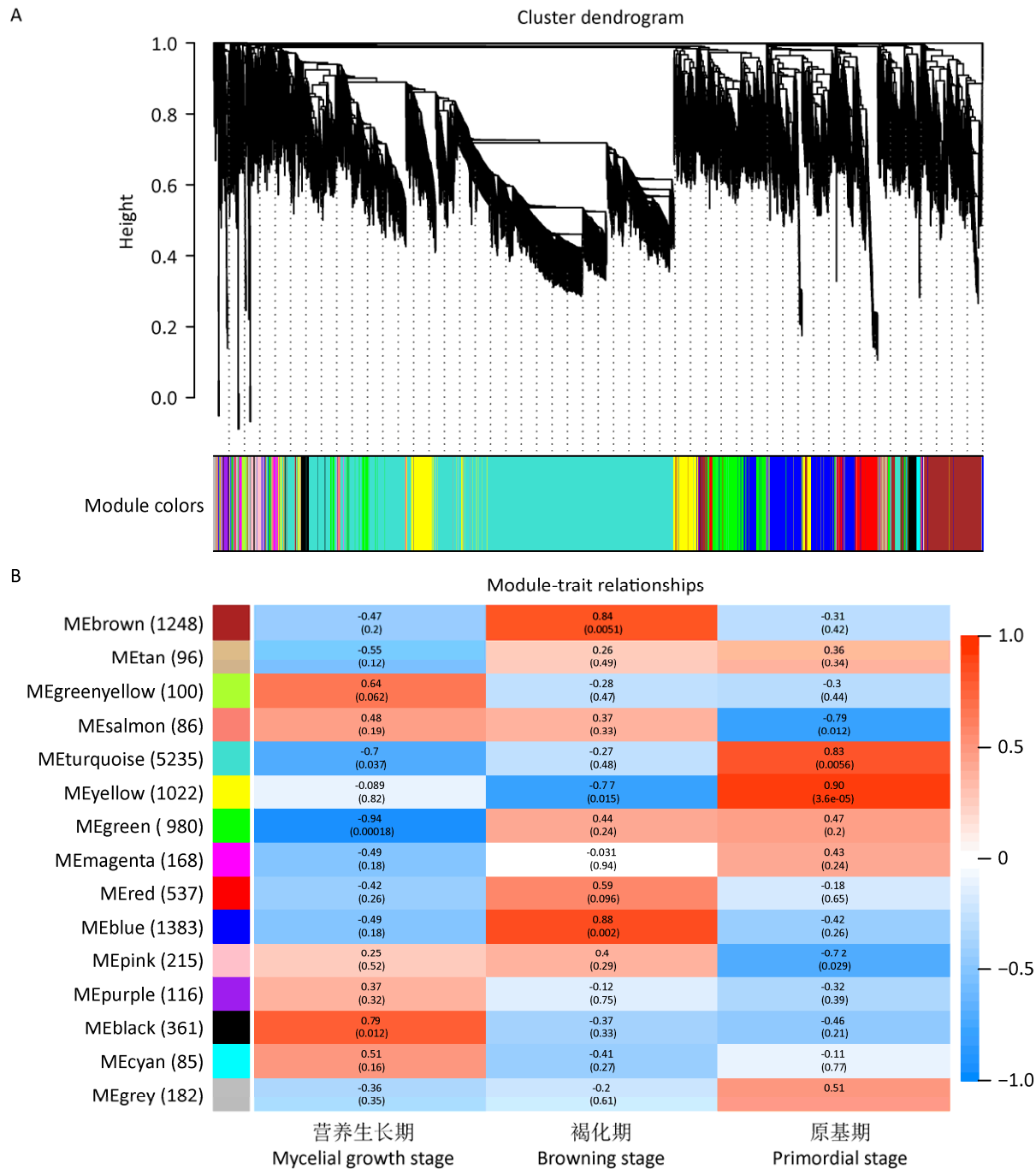


图 5 香菇不同发育阶段的基因网络共线性表达分析 A: 层次聚类树显示由 WGCNA 标识的共表达模块。每个分支代表一个基因，树的分支聚类为 14 个组别，用不同颜色标；B: 一致性模块特征基因与不同阶段的关系。表中的每一行都对应 1 个颜色的共表达模块与表 A 中的分组相对应，表中的 3 列代表了香菇的 3 个不同发育阶段

Fig. 5 WGCNA of genes in different developmental stages of *Lentinula edodes*. A: Hierarchical cluster tree showing coexpression modules identified by WGCNA. Each leaf on the tree represents one gene. The major tree branches constitute 14 modules, labeled with different colors. B: Relationships between consensus module eigengenes and different stages. Each row in the table corresponds to a consensus module, labeled with a color as in (A) and each column represents a developmental stage.

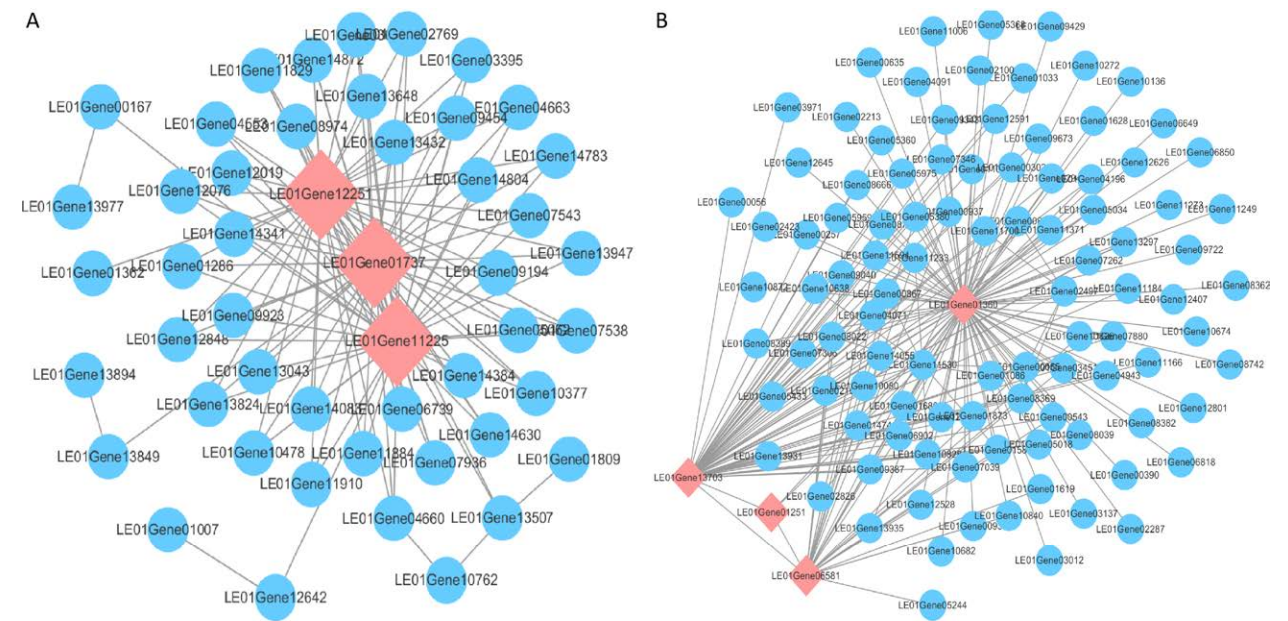


图 6 蓝色模块（A）与黄色模块（B）权重大于 0.5 的前 200 个基因的可视化结果  
Fig. 6 Visualization result of the top 200 genes for Weight ( $\geq 0.5$ ) of blue (A) and yellow (B) module connection.

表 2 与菌丝褐化和原基形成相关的关键基因筛选信息

Table 2 Screening information of hub genes related to Light-induced browning (B) and primordium stage (P)

基因 ID	基因名称 <sup>a</sup>	度	压力	COG 分类
Gene ID	Gene name	Degree	Stress	Cluster of orthologous groups of proteins (COG)
LE01Gene12251	Secreted protein (B)	38	4 242	Coenzyme transport and metabolism
LE01Gene01737	Alpha beta-hydrolase (B)	38	3 440	Lipid transport and metabolism
LE01Gene11225	Tetrapyrrole methylase (B)	38	2 926	Coenzyme transport and metabolism
LE01Gene01360	Hypothetical protein(P)	97	9 666	—
LE01Gene13703	DEAD-domain-containing protein (P)	61	4 024	Replication, recombination and repair
LE01Gene01251	Hypothetical protein (P)	30	1 052	—
LE01Gene06581	Hypothetical protein (P)	12	102	—

注：<sup>a</sup> 基因名称后括号内的字母缩写代表相关的发育阶段

Note：<sup>a</sup> The abbreviations in parentheses after the gene represent the relevant developmental stages.

进一步结合差异基因分析和网络模型共线性分析的结果，基于 3 个阶段中两两比较在不同阶段 2 倍差异上调表达的基因分别与黑色模块（菌丝生长相关）、蓝色模块（菌丝褐化相关）以及黄色模块（原基形成相关）进行了一个交集分析，得到了菌丝生长阶段的 17 个关键基因、褐化阶段的 167 个关键基因以及原基形成阶段的 67 个显著差异基因。

挑选了差异显著的 10–15 个基因利用标准化后的数据分别进行热图的可视化分析(图 7)，从图中可以直观看到，相关模块内的差异基因在其调控的阶段表达量高于其他两个发育阶段。同时，基因表达呈现了连续性，或者缓慢上调或者缓慢下降，整个基因表达的变化受环境调节动态变化。在菌丝阶段，有 4 个基因是水解酶活性基因占关键基因的 1/4，

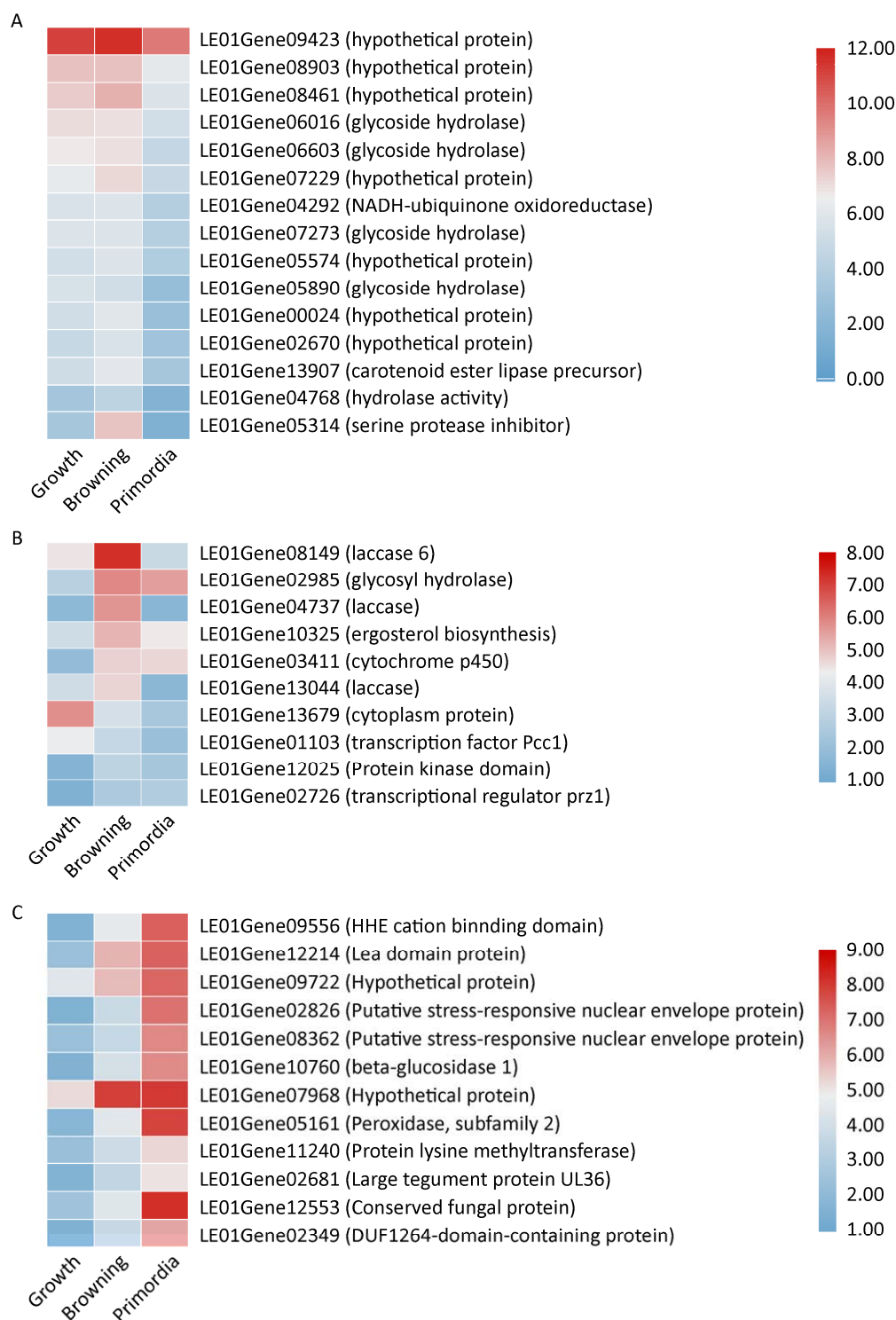


图 7 与 3 个发育阶段相关的模块中差异表达关键基因的相对表达量分析 A: 菌丝表达相关的红色模块基因中显著上调 15 个基因的热图; B: 菌丝褐化相关的蓝色模块中显著上调 10 个基因的热图; C: 原基阶段与绿色模块相关 10 个基因的热图

Fig. 7 The analysis of differentially expressed genes (DEGs) in different stages related to gene module. A: Heatmap of mycelial growth-related hub genes; B: Browning stage-related hub genes; C: Heatmap of primordial formation related genes.

菌丝褐化阶段的关键基因富集了 3 个漆酶, 1 个水解酶和 1 个 P450, 漆酶主要与木质素降解相关, 而 P450 是与色素形成相关基因。菌丝褐化阶段主要是由于环境中光刺激诱导, 本研究的关键基因中聚类到了 2 个与环境胁迫相关的转录因子 (Pcc1 和 Prz1) 和信号转导的蛋白激酶类基因。原基形成期的聚类到了 2 个胁迫相关的因子和过氧化物酶类, 但多数关键基因是一些功能未知的蛋白。

### 3 讨论

WGCNA 的聚类方法通常使用在大数据样本中 (通常大于 15 个样本)。本研究在较少的 9 个转录本中利用 WGCNA 进行了数据的分析尝试, 很好地将基因聚类, 获得一些重要的模块与特定的生长发育时期显著相关。

本研究通过对香菇 3 个阶段的 3 种不同栽培基质的 9 个转录组进行分析, 共获得了 3 609 个差异基因。将 9 个转录组利用主成分分析成功地将数据按照香菇发育的 3 个阶段即菌丝生长期、菌丝褐化成熟期以及原基形成期进行了聚类, 相关性系数也表明了组建重复性较好。结果显示随着香菇发育进程推进栽培基质对香菇的转录组的影响逐渐减弱。将同一时期 3 个不同材料的 RNA-seq 数据作为发育同一阶段的转录组重复, 利用菌丝生长阶段的数据作为参照, 挖掘了菌丝生长阶段、成熟阶段和原基形成时期的特异表达基因, 并对这些基因进行了 GO 和 KEGG 的分类。进一步利用了 WGCNA 的权重网络表达分析, 将所有差异基因进行模块化分析, 找到与 3 个阶段相关的重要基因模块, 并利用网络分析手段获得了 7 个关键基因。最后尝试结合 GO 和 KEGG 分析获得的差异基因, 对聚类的模块基因与不同阶段上调 2 倍的表达基因做交集, 获得了与香菇菌丝生长, 褐

化和性发育早期相关的重要基因。菌丝阶段主要以营养生长为主, 该阶段的菌丝细胞大量表达了水解酶, 用于分解基质中的碳源, 给自己的生长提供能量; 菌丝褐化阶段菌丝继续分解基质, 此时菌丝会转录翻译更多的漆酶和过氧化物酶, 这是因为木质纤维素晶体是纤维素和半纤维素包裹在木质素外的复杂晶体, 香菇菌丝前期降解大量的纤维素和半纤维素使得木质素与菌丝接触, 诱导菌丝产生木质纤维素降解酶降解木质素。同时菌丝褐化产生大量的褐色物质, 推测这些褐色物质是由一些前体物质氧化获得, 故在菌丝褐化阶段大量的氧化酶 (P450) 表达 (Tang *et al.* 2013)。菌丝褐化需要光刺激诱导, 在褐化阶段相关的蓝色模块发现了 2 个与环境胁迫相关的转录因子的高表达 (Pcc1 和 Prz1), Pcc1 转录因子在酵母和鬼笔灰伞中报道, 它是 1 个与细胞发育周期相关的调控转录因子 (Romanova *et al.* 2006; Murata & Kamada 2009), Prz1 是 1 个响应蓝光诱导和钙离子调控的转录因子 (Hirayama *et al.* 2003), 香菇的发育需要光的调节, 通过光刺激诱导香菇菌丝褐化, Prz1 的一个同源基因 (LE01Gene11723) 同时聚类到了转色密切相关的褐色模块 (未分析), 实验结果与实际香菇栽培人工调控相对应。香菇原基形成初期通过聚类分析获得了大量的未知功能蛋白, 由于香菇基因组注释信息相对不完善, 很多基因的功能需要我们日后进一步研究, 但是通过 WGCNA 分析和差异基因分析结合的方法, 我们将这些未知功能的基因进行了模块划分, 这为日后研究香菇的生长发育提供了重要的理论依据。

食用菌的发育是一个复杂的生命过程, 营养菌丝体要分化成为生殖菌丝再发育形成原基和子实体, 受外界环境条件和自身遗传条件的影响, 但归根结底是由遗传物质所决



定。子实体分化发育需要一系列基因的协同表达, 基因的差异表达对于子实体的形成具关键调控作用。在大样本的分析下传统的分析方法无法对基因进行有效地分类, 不能充分地挖掘海量信息背后的关键生物学意义, 新的共表达网络分析可以对复杂的数据进行归纳和整理, 高效的研究基因的整体表达规律, 同时能够系统地反馈样本间相互作用模式。WGCNA 可以通过基因模块化分类, 呈现样本的基因全局表达模式。寻找与样本相关的关键模块和核心基因, 尤其对于食用菌这种基因注释率较低的物种, 可以很好地推测未知基因的功能, 为下游生物学实验提供了有利的线索。因此, WGCNA 在注释比率较低的食用菌中分析多样本数据发挥着重要的意义。

## [REFERENCES]

- Benjamini Y, Hochberg Y, 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of The Royal Statistical Society Series B-Statistical Methodology*, 57(1): 289-300
- Chen LF, Guo YH, Cai YL, Liu W, Zhou Y, Xiao Y, Xu ZY, Liu Y, Lei XY, Wang GZ, Guo MP, Ma XL, Bian YB, 2016. Genome sequence of the edible cultivated mushroom *Lentinula edodes* (shiitake) reveals insights into lignocellulose degradation. *PLoS One*, 11(8): e0160336
- Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY, 2014. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology*, 8(Suppl. 4): S11-S11
- Dai YC, Yang ZL, 2008. A revised checklist of medicinal fungi in China. *Mycosystema*, 27: 801-824 (in Chinese)
- Dai YC, Zhou LW, Yang ZL, Wen HA, Bao T, Li TH, 2010. A revised checklist of edible fungi in China. *Mycosystema*, 29: 1-21 (in Chinese)
- Du J, Wang SD, He CM, Zhou B, Ruan YL, Shou H, 2017. Identification of regulatory networks and hub genes controlling soybean seed set and size using RNA sequencing analysis. *Journal of Experimental Botany*, 68(8): 1955-1972
- Garg R, Singh VK, Rajkumar MS, Kumar V, Jain M, 2017. Global transcriptome and coexpression network analyses reveal cultivar-specific molecular signatures associated with seed development and seed size/weight determination in chickpea. *Plant of Journal*, 91(6): 1088-1107
- Geng JL, 2015. The analysis for statistical survey of national mushroom in 2014. <http://www.Cefa.org.cn> (in Chinese)
- Hirayama S, Sugiura R, Lu Y, Maeda T, Kawagishi K, Yokoyama M, Tohda H, Giga-Hama Y, Shuntoh H, Kuno T, 2003. Zinc finger protein Prz1 regulates  $\text{Ca}^{2+}$  but not  $\text{Cl}^-$  homeostasis in fission yeast. Identification of distinct branches of calcineurin signaling pathway in fission yeast. *Journal of Biological Chemistry*, 278: 18078-18084
- Kim D, Langmead B, Salzberg SL, 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, 12(4): 357-360
- Langfelder P, Horvath S, 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 29(9): 559
- Li X, Wang F, Liu Q, Li QP, Qian ZM, Zhang XL, Li K, Li WJ, Dong CH, 2019. Developmental transcriptomics of Chinese cordyceps reveals gene regulatory network and expression profiles of sexual development-related genes. *BMC Genomics*, 20: 337
- Murata Y, Kamada T, 2009. Identification of new mutant alleles of pcc1 in the homobasidiomycete *Coprinopsis cinerea*. *Mycoscience*, 50(2): 137-139
- Romanova EK, Lopreiato R, Baillieu AB, Rousselle JC, Ilan L, Hofmann K, Namane A, Mann C, Libri D, 2006. Yeast homolog of a cancer-testis antigen defines a new transcription complex. *The EMBO Journal*, 25: 3576-3585



- Tang LH, Jian HH, Song CY, Bao DP, Shang XD, Wu DQ, Tan Q, Zhang XH, 2013. Transcriptome analysis of candidate genes and signaling pathways associated with light-induced brown film formation in *Lentinula edodes*. *Applied Microbiology and Biotechnology*, 97: 4977-4989
  - Trapnell C, Pachter L, Salzberg SL, 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 25(9): 1105-1111
  - Yin L, Cai ZH, Zhu BA, Xu CS, 2018. Identification of key pathways and genes in the dynamic progression of HCC based on WGCNA. *Genes*, 9(2): 92
  - Zhang F, 2008. Research progress on cultivation techniques of *Lentinula edodes*. *Food Engineering*, 2(4): 28-30 (in Chinese)
  - Zhou S, Zhang JS, Ma FY, Tang CH, Tang QJ, Zhang XY, 2016. Investigation of lignocellulolytic enzymes during different growth phases of *Ganoderma lucidum* strain G0119 using genomic, transcriptomic and secretomic analyses. *PLoS One*, 13(5): e0198404
- [附中文参考文献]
- 戴玉成, 杨祝良, 2008. 中国药用真菌名录及部分名称的修订. *菌物学报*, 27: 801-824
  - 戴玉成, 周丽伟, 杨祝良, 文华安, 图力古尔, 李泰辉, 2010. 中国食用菌名录. *菌物学报*, 29: 1-21
  - 耿建利, 2015. 对 2014 年度全国食用菌统计调查结果的 analysis. <http://www.Cefa.org.cn>
  - 张锋, 2008. 香菇栽培技术研究进展. *食品工程*, 2(4): 28-30

(本文责编: 王敏)