



## 研究报告

## 基于三代测序的大秃马勃的 rDNA ITS 鉴定及其基因组生物信息学分析

史延华 林维平\* 侯典海\*

潍坊医学院生命科学与技术学院 山东 潍坊 261053

**摘要:**【背景】马勃是著名的传统可食用药用真菌，具有多种药理作用，包括抗菌、抗炎、止咳、抗肿瘤和抑制细胞增殖等。目前关于马勃的研究多集中在活性物质的分离提取及其药理药效上，而对马勃基因组遗传信息知之甚少。【目的】获得马勃真菌的基因组遗传信息，挖掘潜在具有工业应用价值的功能基因。【方法】联合使用 Illumina 二代测序技术和 PacBio 三代测序技术进行基因组测序，并综合使用多种数据库进行基因组结构分析和基因注释，包括 GO 注释。基于基因组信息，使用 rDNA ITS 序列进行马勃真菌物种鉴定，并使用 dbCAN 和 antiSMASH 进行碳水化合物活性酶和次级代谢物基因簇的鉴定。【结果】测序获得大小为 46.04 Mb 基因组，有 11 903 个蛋白编码基因，包括 6 634 个 GO 注释基因。基于基因组中的 rDNA ITS 序列信息将该菌鉴定为大秃马勃(*Calvatia gigantea*)，分析了麦角甾醇和萜类合成相关基因 25 个，鉴定出 317 个碳水化合物酶和 18 个次级代谢物生物合成基因簇，并发掘出 1 个潜在的、可分泌的抗丹宁  $\alpha$ -淀粉酶。【结论】获得一株马勃真菌基因组，并发掘出潜在的食品工业应用价值的抗丹宁  $\alpha$ -淀粉酶和多个新的次级代谢物合成基因簇。这将为马勃真菌遗传进化、功能基因组学研究，以及其在食品工业应用和次级代谢物的药用价值开发提供重要的遗传信息基础。

关键词：马勃，基因组，三代测序，次级代谢物基因簇

rDNA ITS identification and genomic analysis of *Calvatia gigantea*, the first sequenced species of *Lycoperdaceae*

SHI Yanhua LIN Weiping\* HOU Dianhai\*

School of Bioscience and Technology, Weifang Medical University, Weifang, Shandong 261053, China

**Abstract:** [Background] The puffball mushroom is commonly used traditional medicinal edible fungus, which has multiple pharmacological effects, including antimicrobial, anti-inflammatory, anti-tussive, anti-cancer and suppressing cell growth effects. So far, most of the researches on puffball mushrooms have

**Foundation items:** Natural Science Foundation of Shandong Province (ZR2020QC014); Scientific Research Projects of Colleges and Universities in Shandong Province (J18KA176)

\*Corresponding authors: Tel: 86-536-8462495

E-mail: LIN Weiping: linwpwf@163.com; HOU Dianhai: houdh@wfmc.edu.cn

**Received:** 22-12-2020; **Accepted:** 15-02-2021; **Published online:** 16-03-2021

基金项目：山东省自然科学基金青年项目(ZR2020QC014)；山东省高校科研计划项目(J18KA176)

\*通信作者：Tel: 0536-8462495

E-mail: 林维平: linwpwf@163.com; 侯典海: houdh@wfmc.edu.cn

收稿日期: 2020-12-22; 接受日期: 2021-02-15; 网络首发日期: 2021-03-16

been focused on extraction and pharmacological effect of their bioactive agents. Nevertheless, little is known about the genomic information of puffball mushrooms. **[Objective]** To acquire genomic information of puffball mushrooms and mine protein-encoding gene of potential application in industry. **[Methods]** A combination of Illumina and PacBio sequencing technologies was used for genome sequencing, and analysis of genome structure and gene annotation were performed against multiple database, including GO. The sequenced puffball mushroom was subjected to ITS identification, while CAZymes and secondary metabolite biosynthetic gene clusters were identified using dbCAN and antiSMASH, respectively. **[Results]** The acquired puffball genome was 46.04 Mb in size, encoding 11 903 proteins including GO-annotated genes. Based on the genomic information, the puffball species was identified as *Calvatia gigantea* using ITS (Internal Transcribed Spacer). A total of 25 genes related to terpene and ergosterol biosynthetic pathway were analyzed, while 317 CAZymes and 18 secondary metabolite biosynthetic gene clusters were identified. In addition, an  $\alpha$ -amylase from CAZymes was characterized as a potential secretory, tannin-resistant  $\alpha$ -amylase. **[Conclusion]** This study supplied the first sequenced puffball mushroom, and found a potential tannin-resistant  $\alpha$ -amylase and multiple new secondary metabolite biosynthetic gene clusters. This study will supply important basic genetic information for genetic evolution and functional researches of puffball mushrooms, their further application in food industry, and development of secondary metabolites of pharmaceutical value.

**Keywords:** *Calvatia*, genome, third-generation sequencing, secondary metabolite biosynthetic gene clusters

马勃是一类真菌的通称, 大多属担子菌门马勃科(*Lycoperdaceae*), 也有部分为地星科(*Geastraceae*)<sup>[1]</sup>。马勃是著名的传统药用真菌, 具有多种药理作用, 包括抗菌消炎、止血止咳、降血糖、抗肿瘤和抑制细胞增殖作用等<sup>[2-4]</sup>。马勃可食用, 其化学成分相对复杂, 主要含有小分子化合物, 如氨基酸、甾类、萜类和醌类物质, 以及大分子化合物, 包括多糖、蛋白质和多肽等有益活性成分<sup>[5]</sup>。Lam 等从龟裂马勃(*Calvatia caelate*)中分离出一种泛素样活性多肽 CULP, 该多肽含有 76 个氨基酸残基, 可抑制人乳腺癌 MDA-MB-231 细胞增殖<sup>[6]</sup>。接着 Ng 等从龟裂马勃的新鲜子实体中分离出活性蛋白 Calcaelin, 该蛋白具有抑制肿瘤细胞有丝分裂和增殖生长的能力, 同时发现其在鼠脾细胞中具有抗细胞分裂原活性<sup>[7]</sup>。最近 Lee 等从拟秃马勃(*Calvatia nipponica*)中分离出 N,N-二甲基邻氨基苯甲酸和(7Z,10Z)-7,10-十八碳二烯酸甲酯 2 种主要化合物, 可能用于治疗男性功能障碍, 具有滋补功效<sup>[8]</sup>。

大秃马勃, 属于马勃科(*Lycoperdaceae*)的秃马勃属(*Calvatia*)<sup>[9]</sup>。与其他马勃真菌类似, 其富含氨基酸, 至少包括 8 种必需氨基酸和 10 种非必需氨

基酸<sup>[3,10]</sup>, 并含有多种复杂的活性成分, 包括具有抗氧化作用的多酚类和黄酮类物质<sup>[5]</sup>。大秃马勃子实体的甲醇提取物能够有效降低患糖尿病小鼠的血糖浓度, 可作为潜在的抗糖尿病治疗因子<sup>[11]</sup>, 而且体外实验显示甲醇提取物也能够抑制人肺癌 A549 细胞的有丝分裂并诱导细胞凋亡<sup>[2]</sup>。研究发现大秃马勃多糖(*Calvatia gigantea* Polysaccharide, CGP)对乳腺癌细胞和宫颈癌细胞的增殖有一定的抑制作用<sup>[12]</sup>。早期研究发现大秃马勃产可降解儿茶素的酶<sup>[13]</sup>, 预示着大秃马勃可能在降解毒性酚类及多酚物质以及聚合丹宁类物质中具有应用价值<sup>[13-14]</sup>。大秃马勃也产生抗单宁酸  $\alpha$ -淀粉酶, 而且在植物碳水化合物上生长, 随着含有的毒性单宁类化合物含量提升, 抗单宁酸  $\alpha$ -淀粉酶产量也随之上升, 并呈现很好的动力学特征<sup>[15-16]</sup>。大秃马勃还能够产脂肪酶, 其单位体积的高酶活力产量可比拟脂肪酶高产微生物<sup>[17]</sup>。马勃可以食用这一特点, 也表明大秃马勃的脂肪酶未来可适用于食品工业<sup>[17]</sup>。然而目前尚无马勃基因组测序, 需要基因组遗传信息促进后续明确的基因和蛋白功能研究。

在本研究中, 我们结合 Illumina 二代测序和

PacBio 三代测序, 获得首个马勃真菌基因组序列。基于基因组信息, 我们通过 rDNA ITS 鉴定将该马勃真菌归类为大秃马勃(*Calvatia gigantea*), 并进一步对基因组进行了注释以及碳水化合物活性酶和次级代谢物基因簇的鉴定。本研究提供了首个马勃真菌基因组数据, 并为后续马勃真菌生物学研究、活性化合物挖掘和医疗及食品、工业应用等提供基础的遗传信息资源。

## 1 材料与方法

### 1.1 材料

#### 1.1.1 菌种来源

本研究鉴定为大秃马勃(*Calvatia gigantea*)的菌株, 购自中国普通微生物菌种保藏管理中心(CGMCC5.752), 该菌在库中标注的名称为皮树丝马勃(*Mycenastrum corium*)。

#### 1.1.2 培养基

固体 PDA 培养基和液体 PDA 培养基参照文献[18]进行配制。

#### 1.1.3 主要试剂和仪器

DNeasy Blood & Tissue Kit, Qiagen 公司。全温振荡培养箱, 上海知楚仪器有限公司; NanoDrop 分光光度仪, Thermo 公司。

### 1.2 方法

#### 1.2.1 菌株培养和基因组提取

菌种接种于固体 PDA 培养基上, 28 °C 培养 14 d 后转接到液体 PDA 培养基中 28 °C、120 r/min 培养 14 d, 4 °C、10 000×g 离心 10 min 收集菌体后液氮冷冻处理。将真菌用研钵磨碎后采用 Sanghai-Marooof 等的方法<sup>[19]</sup>提取 DNA, 接着参照试剂盒使用说明书, 用 DNeasy Blood & Tissue Kit 对 DNA 继续进行洗涤和重提。DNA 浓度和质量用 NanoDrop 分光光度仪测后用于基因组测序<sup>[20]</sup>。

#### 1.2.2 基因组测序和组装

基因组测序和组装使用浙江安诺优达生物科技有限公司的 Illumina HiSeq X Ten 和 PacBio Sequel I 高通量测序平台。所得的原始数据经过过滤得到高质量的 DNA 序列数据后依照平台基因组

装流程进行处理。简略来讲, 对 PacBio Sequel I 测序生成的三代测序数据, 分别使用 Canu<sup>[21]</sup>, SMARTdenovo<sup>[22]</sup>和 WTDBG<sup>[23]</sup>软件进行装配, 选出 WTDBG 软件组装结果进行后续处理。接着使用 PacBio 的 Arrow 软件<sup>[24]</sup>进行自我单个碱基校正。然后使用 Illumina HiSeq X Ten 测序生成的二代测序数据, 运用 Pilon 软件<sup>[25]</sup>对三代测序组装的序列进行校正, 最终得到校正后的基因组。基因组组装的准确性和完整性使用 BUSCO 软件(V3.0.1)和 Basidiomycota\_odb9 数据库<sup>[26]</sup>进行评估。

#### 1.2.3 基因组注释

基因组注释利用基因预测软件 Augustus<sup>[27]</sup>、Gene Mark<sup>[28]</sup>、Glimmer HMM<sup>[29]</sup>和 SNAP<sup>[30]</sup>分别进行蛋白质编码基因的从头预测, 然后使用 EVIDENCEModeler (EVM)<sup>[31]</sup>将所有预测的基因整合成一个非冗余的完整基因集。然后使用 BLAST<sup>[32]</sup>将集成基因基于同源性与各种常见的功能数据库进行比较。所使用的常见的功能数据库包括 eggNOG<sup>[33]</sup>、GO<sup>[34]</sup>、KEGG<sup>[35]</sup>、PFAM<sup>[36]</sup>和 SwissProt<sup>[37]</sup>、NT (<https://www.ncbi.nlm.nih.gov/nucleotide/>)、NR (<ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr.gz>)。

另外, 对于基因组中非编码 RNA 的预测, 使用 tRNAscan-SE 软件<sup>[38]</sup>预测转运 RNA (tRNA), 同时, 将基因组在已知非编码 RNA 库 Rfam 数据库<sup>[39]</sup>进行比对, 以预测核糖体 RNA (rRNA)、微 RNA (miRNA)和小核仁 RNA (snoRNA)。对于重复序列的预测, 先使用 RepeatMasker 和 RepeatProteinmasker 软件<sup>[40]</sup>将基因组序列在数据库 RepBase<sup>[41]</sup>进行同源比对, 生成基于数据库搜索的重复序列库。然后使用 RepeatModeler 软件(<http://www.repeatmasker.org/RepeatModeler/>)进行重复序列的从头预测, 生成 *De Novo* 重复序列库。再用 RepeatMasker 软件将基因组与 *De Novo* 重复序列库进行比对。最后使用 EVM 软件对所有预测的结果进行整合。对于转座子的串联重复序列的预测, 从头预测还使用 Tandem Repeats Finder (TRF)软件在基因序列中进行搜索<sup>[42]</sup>。

### 1.2.4 rDNA ITS 真菌鉴定

根据基因组非编码 RNA 中的 rRNA 预测结果,提取相邻 18S rRNA 基因和 28S rRNA 基因之间的内源转录间隔区(Internally Transcribed Spacer, ITS)序列(包含 5.8S rRNA 基因)。从多个重复的 ITS 序列中选取 1 个作为本实验菌株的 ITS 代表序列(该序列位于 Contig 10 的 55 507–56 984 位)(GenBank 序列号为 MW672586)用于 rDNA ITS 鉴定分析。从 UNITE 数据库<sup>[43]</sup>中搜索马勃科(*Lycoperdaceae*)秃马勃属(*Calvatia*)中物种的 rDNA ITS 序列进行下载,搜索参数设置如下:“Threshold”为“1.5”;“include”为“All SH-s”。同样,皮树丝马勃(*Mycenastrum corium*) rDNA ITS 序列也在 UNITE 数据库中下载得到,并用于后续分析。鳞柄灰锤菌(*Tulostoma squamosum*)的 rDNA ITS 序列(GenBank 登录号为 DQ415732)作为外群(Outgroup)<sup>[9]</sup>。所有 rDNA ITS 使用 MAFFT 的 FFT-NS-i 算法进行序列比对<sup>[44]</sup>,然后对比对序列进行手动整理,删除比对序列中所对应的通用引物<sup>[45]</sup> ITS1 (5'-TCCGTAGGTGAACCTGCGG-3')序列与其左侧序列,以及反向引物 ITS4 (5'-TCCTCCGCTTATTGATATGC-3')序列与其右侧序列。接着用软件 IQ-TREE 2<sup>[46-47]</sup>通过 ModelFinder<sup>[48]</sup>选择出最优的 DNA 进化模型,然后使用软件 IQ-TREE 1.6.1 依据该进化模型构建最大似然树,并用 FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>)查看和手动调整进化树。

### 1.2.5 碳水化合物活性酶和次级代谢物基因簇的鉴定

基因组中所含有的碳水化合物活性酶(CAZymes)使用带有 HMMER、DIAMOND<sup>[49]</sup>和 Hotpep 搜索方法的 dbCAN2 在线预测软件<sup>[50]</sup>进行鉴定。次级代谢产物相关的生物合成基因簇使用 antiSMASH 5.0<sup>[51]</sup>进行预测:将基因组序列数据和注释文件(gff3 格式)上传,选择的检测严谨度“Detection Strictness”为“Relaxed”,“Extra Features”为“All on”。

### 1.2.6 生物信息学分析

蛋白序列通过 NCBI BLASTp<sup>[32]</sup>进行手动序列比对,并通过 NCBI Conserved Domain Search Service (CD Search)<sup>[52]</sup>进行功能域搜索和分析。蛋白

分泌信号肽使用 SignalP 5.0<sup>[53]</sup>进行预测,N-糖基化位点和 Epsilon 氨基糖基化的赖氨酸位点分别使用 NetNGlyc 1.0 (<http://www.cbs.dtu.dk/services/NetNGlyc/>)和 NetGlycate 1.0<sup>[54]</sup>进行预测,O-(alpha)-糖基化位点和 C-甘露糖糖基化位点则使用 DictyOGlyc 1.1<sup>[55]</sup>和 NetCGlyc 1.0<sup>[56]</sup>进行预测。

## 2 结果与分析

### 2.1 大秃马勃基因组序列的一般特征

结合 5 984 Mb Illumina HiSeq X Ten 数据和 16 431 Mb PacBio Sequel I 数据,组装出大秃马勃的基因组(GenBank 登录号 JACAAK000000000),其总长度为 46.04 Mb,GC 含量为 47.50%,包括 89 个重叠群(Contigs) (N50 大小为 3.04 Mb)。基因组的 BUSCO 评估值为 95.5% (其中 Complete BUSCOs 为 91.8%),说明基因组组装和注释的准确性和完整性好。我们从基因组中共预测出 11 903 个蛋白编码基因,其中基因平均长度为 1 860.32 bp。基因的蛋白编码序列的平均长度为 1 328.64 bp,平均每个基因有 5.49 外显子(平均 241.90 bp)和 5.46 内含子(平均 119.34 bp)。此外,我们还预测出 254 个 tRNA 基因、61 个 rRNA 基因、10 个 snoRNA 基因。重复序列的总长度为 15.23 Mb (占基因组的 33.07%),其中串联重复序列占基因组长度的 0.17%,而转座子序列总大小为 15.15 Mb (占基因组的 32.90%)。在这些转座子序列中,DNA 转座子、长端重复序列(Long Terminal Repeats, LTRs)和非 LTR 转座子序列分别占基因组的 0.19%、20.59%和 2.88%。

### 2.2 大秃马勃的 ITS 鉴定

所有 51 条 ITS 经序列比对和手动修剪后,生成含 747 个位点的序列文件。IQ-TREE 选出了最优的 DNA 进化模型为“TIM2+F+R3”,使用该模型,运用 IQ-TREE 构建了最大似然树。本研究中的 CGMCC5.752 和大秃马勃(ITS 的 GenBank 登录号为 DQ112623)归在同一分支,进化距离非常近(分支长度接近为 0),而并未与栓皮马勃(*Mycenastrum corium*)聚在一起。因此本研究将其修订为大秃马勃(*Calvatia gigantea*),菌株编号为 5.752 (图 1)。

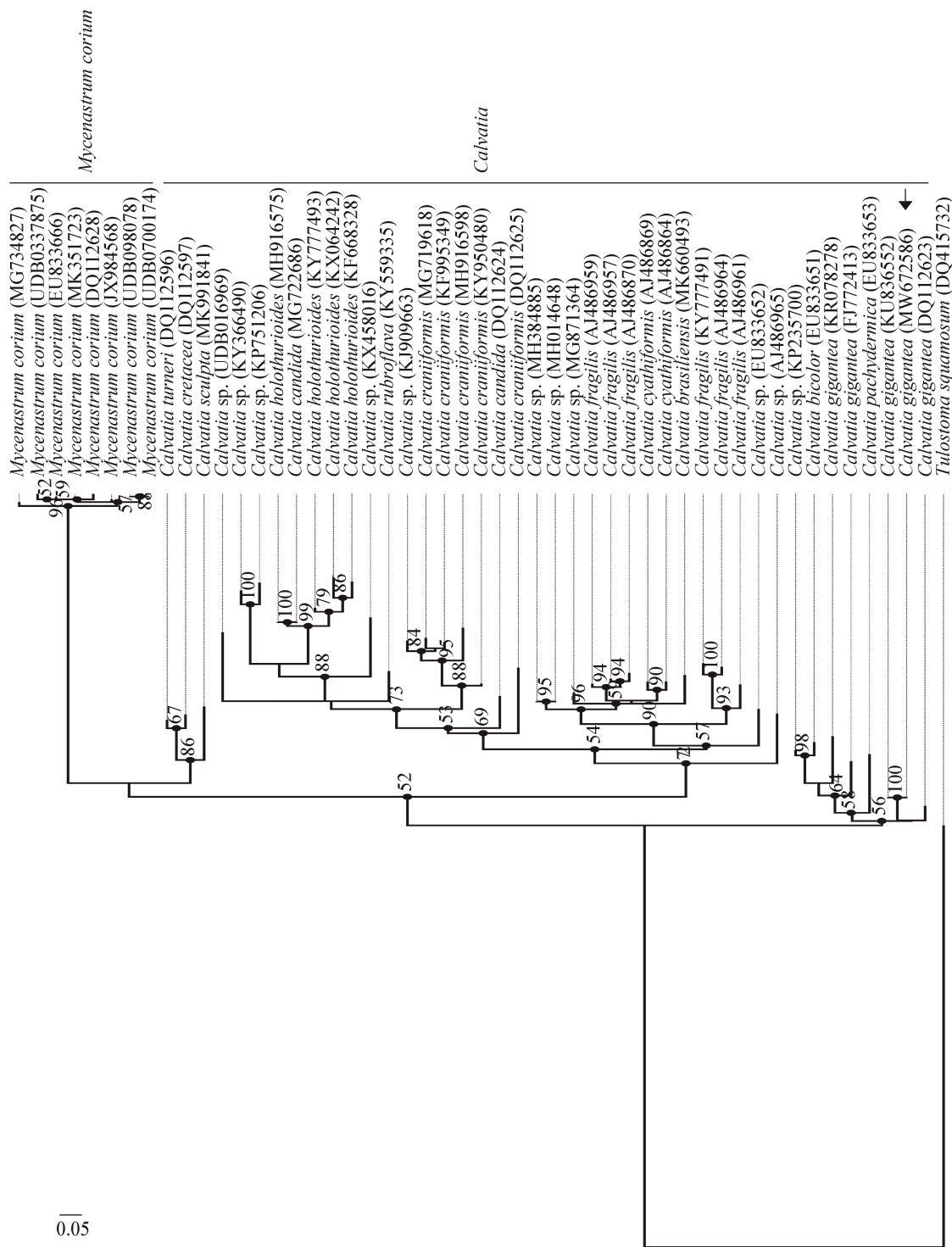


图 1 基于 rDNA ITS 序列构建的秃马勃属系统发育树

Figure 1 Phylogenetic tree on the rDNA ITS sequence of the genus *Calvatia* species

注: 括号中为菌株 GenBank 序列号; 黑实心圆代表 Bootstrap 值大于 50 的节点; 标尺代表支长 0.05; 箭头所指为本项目的大秃马勃菌株  
Note: The GenBank accession number of each strain was indicated in brackets; The node with bootstrap value  $\geq 50$  was indicated as black solid circle; The strain of *Calvatia gigantea* used in this study were indicated as black arrow

### 2.3 大秃马勃 11 903 个蛋白编码基因的功能注释

通过与包括 SwissProt、GO、KEGG、NT、NR、PFAM 等在内的多个数据库进行基于同源性搜索和比对, 共有 10 444 个蛋白编码基因得到了注释, 占总蛋白编码基因总数(11 903)的 87.74%, 其中 6 634 个得到了 GO (Gene Ontology)功能注释, 占总蛋白编

码基因总数的 55.73%。根据 GO 归类, 在 Cellular Component 类别中, 蛋白大部分归类在 Cell Part、Organelle 和 Organelle Part 条目中, 在 Biological Process 类别中, 蛋白大部分归类在 Cellular Process、Metabolic Process 和 Single-Organism Process 条目中, 而在 Molecular Function 类别中, 蛋白大部分归类在 Binding 和 Catalytic 条目中(图 2)。尽管如此,

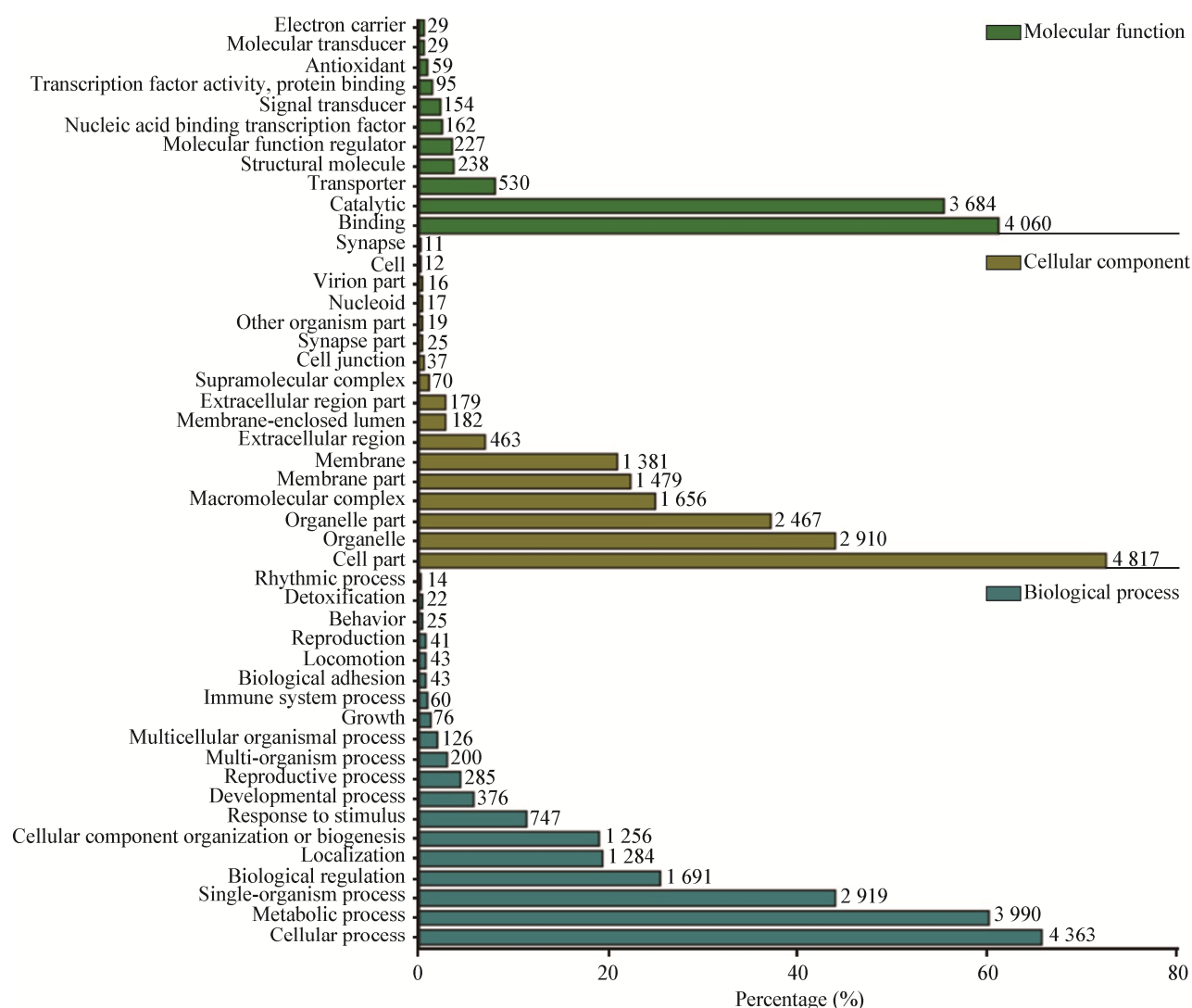


图 2 大秃马勃 5.752 GO 功能注释

Figure 2 GO annotation of the genome of *Calvatia gigantea* 5.752

注: 基因数量标记在各百分比条的右侧; 百分比条的长度代表对应 GO 类别中基因数量占 GO 注释基因总数量(6 634 个)的百分比。基因数小于等于 10 的 GO 分类条目未在图中显示

Note: The gene number was noted at the right of each bar; The length of each bar represents the percentage of genes in each GO class (the percentage of total GO-annotated genes was set as 100%); Each GO class with number of genes  $\leq 10$  was not shown

在预测的蛋白质编码基因中, 仍有 5 269 个基因(占总基因数的 44.27%)仍未得到功能注释, 而且, 其中仍有 649 个基因没有在使用数据库中找到与之同源的序列, 因而本研究的基因数据集不仅丰富了基因数据库资源, 也为进一步研究马勃基因功能, 挖掘新的代谢途径和新的具有药用价值的代谢物提供基础的遗传信息资源。

## 2.4 萜类和麦角甾醇合成途径相关基因的鉴定

麦角甾醇又称麦角固醇, 对细胞的生理功能起到重要的作用, 并具有抗炎、降脂、防癌、免疫调节等功效<sup>[57]</sup>。其作为真菌细胞膜中的重要组成成分, 还可作为其质量控制的重要检测方法<sup>[58]</sup>。萜类化合物是一类结构丰富、生物活性多样的次级代谢产物, 在治疗各种疾病中有着非常重要的作用。研

究表明, 马勃中含有较多的萜类化合物<sup>[4]</sup>。麦角甾醇代谢途径研究的已非常明确, 而且与萜类骨架合成途径重叠以及倍半萜和三萜的骨架合成途径重叠(图 3)。真菌主要通过甲羟戊酸途径(Mevalonate Pathway, MVA 途径)以乙酰辅酶 A 为原料合成异戊烯焦磷酸和二甲基烯丙基焦磷酸 2 个产物。这 2 种物质是所有萜类化合物的前体物质。然后 2 分子异戊烯焦磷酸和 1 分子二甲基烯丙基焦磷酸, 在法尼基焦磷酸合酶催化下形成法尼基焦磷酸, 而后在鲨烯合成酶和鲨烯环氧化酶作用下, 依序生成鲨烯和环氧鲨烯(图 3)。法尼基焦磷酸、鲨烯和环氧鲨烯作为骨架, 可用于各种倍半萜和三萜的合成(图 3)。相较于萜类骨架合成途径, 从环氧鲨烯到麦角甾醇的合成途径是麦角甾醇所独有的(图 3)。针对麦

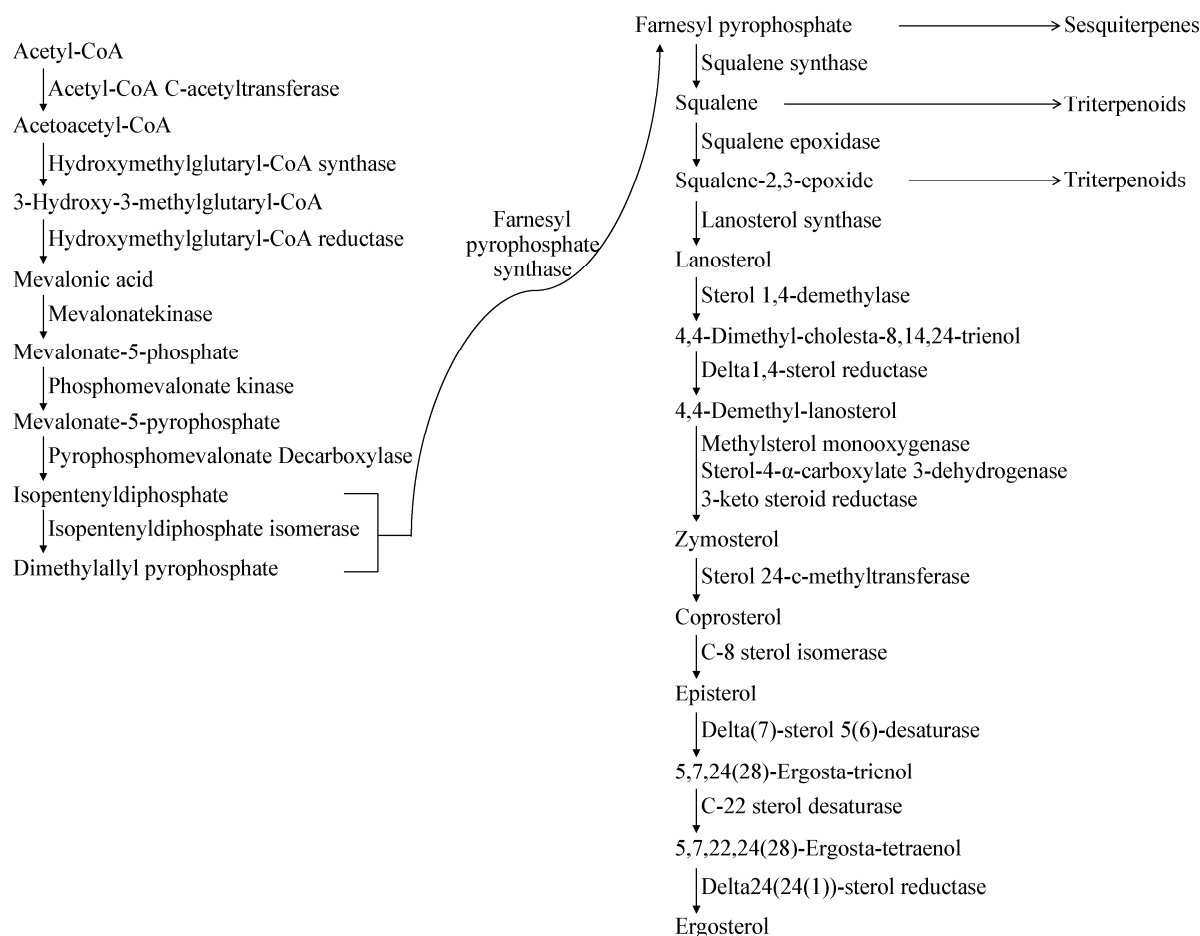


图 3 萜类和麦角甾醇合成途径概览

Figure 3 Overview of terpene and ergosterol biosynthetic pathway

角甾醇整个合成途径所涉及的 21 个酶(表 1), 大秃马勃基因组注释共发现 25 个基因, 其中萜类骨架合成相关基因 10 个, 倍半萜和三萜骨架合成相 关基因 2 个, 麦角甾醇合成相关基因 15 个(包括 倍半萜和三萜合成相关基因 2 个) (表 1)。这些酶 都在细胞质内参与代谢途径, 信号肽预测显示这

表 1 大秃马勃 5.752 基因组中预测的萜类和麦角甾醇合成途径相关基因

Table 1 Predicted genes related to terpene and ergosterol biosynthetic pathway from *Calvatia gigantea* 5.752 genome

序号 No.	酶 Enzyme	位置 <sup>#</sup> Location <sup>#</sup>	外显子/内含子 Exon/Intron	氨基酸 Amino acid	酶号 No. of enzyme
1	乙酰辅酶 A 酰基转移酶 <sup>a</sup> Acetyl-CoA C-acetyltransferase <sup>a</sup>	4: 2 056 050–2 057 545 11: 1 113 517–1 114 955	6/5 6/5	414 381	[EC:2.3.1.9]
2	羟甲基戊二酰辅酶 A 合酶 <sup>a</sup> Hydroxymethylglutaryl-CoA synthase <sup>a</sup>	5: 1 650 395–1 654 135	10/9	696	[EC:2.3.3.10]
3	羟甲基戊二酰辅酶 A 还原酶 <sup>a</sup> Hydroxymethylglutaryl-CoA reductase (NADPH) <sup>a</sup>	11: 1 311 975–1 317 376	5/4	1 358 <sup>TM</sup>	[EC:1.1.1.34]
4	甲羟戊酸激酶 <sup>a</sup> Mevalonate kinase <sup>a</sup>	9: 2 097 657–2 099 053	6/5	339	[EC:2.7.1.36]
5	磷酸甲羟戊酸激酶 <sup>a</sup> Phosphomevalonate kinase <sup>a</sup>	1: 2 162 798–2 164 751	8/7	505	[EC:2.7.4.2]
6	焦磷酸甲羟戊酸脱羧酶 <sup>a</sup> Pyrophosphomevalonate decarboxylase <sup>a</sup>	1: 661 538–663 109	4/3	309	[EC:4.1.1.33]
7	异戊烯基二磷酸异构酶 <sup>a</sup> Isopentenylidiphosphate isomerase <sup>a</sup>	3: 2 972 439–2 973 364	4/3	251	[EC:5.3.3.2]
8	法尼基焦磷酸合酶 <sup>a</sup> Farnesyl pyrophosphate synthase <sup>a</sup>	50: 1 713–2 675 50: 9 447–9 979	3/2 2/1	287 163	[EC:2.5.1.1;2.5.1.10]
9	鲨烯合酶 <sup>b,c</sup> Squalene synthase <sup>b,c</sup>	5: 2 732 722–2 734 624	6/5	488 <sup>TM</sup>	[EC:2.5.1.21]
10	鲨烯环氧化物 <sup>b,c</sup> Squalene epoxidase <sup>b,c</sup>	5: 2 775 389–2 777 081	4/3	507	[EC:1.14.99.7]
11	羊毛甾醇合酶 <sup>c</sup> Lanosterol synthase <sup>c</sup>	5: 883 323–886 108	12/11	729	[EC:5.4.99.7]
12	甾醇 1,4 脱甲基酶 <sup>c</sup> Sterol 1,4-demethylase <sup>c</sup>	9: 974 919–976 947 10: 1 153 547–1 157 168	6/5 7/6	437 <sup>TM</sup> 659 <sup>TM</sup>	[EC:1.14.13.70]
13	1,4 甾醇还原酶 <sup>c</sup> Delta1,4-sterol reductase <sup>c</sup>	9: 1 816 986–1 818 575	6/5	424 <sup>TM</sup>	[EC:1.3.1.70]
14	甲基甾醇单氧化酶 <sup>c</sup> Methylsterol monooxygenase <sup>c</sup>	14: 1 221 205–1 222 448	5/4	337 <sup>TM</sup>	[EC:1.14.13.72]
15	甾醇-4- $\alpha$ -羧化-3-脱氢酶 <sup>c</sup> Sterol-4- $\alpha$ -carboxylate 3-dehydrogenase <sup>c</sup>	7: 1 656 891–1 658 422 11: 806 672–808 337	4/3 2/1	457 529 <sup>TM</sup>	[EC:1.1.1.170]
16	3-类固醇还原酶 <sup>c</sup> 3-keto steroid reductase <sup>c</sup>	4: 1 083 642–1 085 063	5/4	400	[EC:1.1.1.270]
17	甾醇 24-c-甲基转移酶 <sup>c</sup> Sterol 24-c-methyltransferase <sup>c</sup>	10: 1 715 140–1 716 420	5/4	349	[EC:2.1.1.41]
18	C-8 固醇异构酶 <sup>c</sup> C-8 sterol isomerase <sup>c</sup>	14: 1 219 789–1 220 286	1/0	165	[EC:5.3.3.5]
19	$\Delta$ (7)-固醇 5(6)-去饱和酶 <sup>c</sup> Delta(7)-sterol 5(6)-desaturase <sup>c</sup>	9: 1 887 737–1 888 970	5/4	337 <sup>TM</sup>	[EC:1.14.19.20]
20	C-22-甾醇去饱和酶 <sup>c</sup> C-22 sterol desaturase <sup>c</sup>	3: 2 225 003–2 226 808	6/5	510	[EC:1.14.14.-]
21	$\Delta$ 24[24(1)]-甾醇还原酶 <sup>c</sup> Delta24(24(1))-sterol reductase <sup>c</sup>	5: 1 789 072–1 790 978	9/8	486 <sup>TM</sup>	[EC:1.3.1.71]

注: <sup>a</sup>: 萜类骨架合成相关基因; <sup>b</sup>: 倍半萜和三萜合成相关基因; <sup>c</sup>: 麦角甾醇合成相关基因; #: 冒号前面数字表示重叠群的序号, 后面数字代表基因起始-终止位置; TM: 跨膜区

Note: <sup>a</sup>: Predicted genes related to terpenoid backbone biosynthesis; <sup>b</sup>: Predicted genes related to sesquiterpenoid and triterpenoid biosynthesis; <sup>c</sup>: Predicted genes related to ergosterol biosynthesis; #: The number before the colon indicates the contig number while the number after the colon indicates the start-end site of the corresponding gene; TM: Transmembrane domain



些酶并无信号肽。跨膜区预测发现,共有 8 种酶(共 9 个基因)属于跨膜蛋白(表 1), 其中  $\Delta 24[24(1)]$ -甾醇还原酶含有 9 个跨膜区, 1,4 甾醇还原酶含有 7 个跨膜区, 羟甲基戊二酰辅酶 A 还原酶含有 5 个跨膜区,  $\Delta(7)$ -固醇 5(6)-去饱和酶和甾醇 1,4 脱甲基酶(位于 Contig 9)分别含有 3 和 2 个跨膜区, 鲨烯合酶、甾醇 1,4 脱甲基酶(位于 Contig 10)、甲基甾醇单氧化酶和甾醇-4- $\alpha$ -羧化-3-脱氢酶则均含有 1 个跨膜区。

2.5 大秃马勃基因组中碳水化合物酶的鉴定

依据 dbCAN2 的建议, 选择至少由 2 种搜索方法鉴定出来的蛋白作为碳水化合物酶。本研究共鉴定出 317 个碳水化合物酶, 其中 194 个由 3 种搜索方法鉴定出来。根据 CAZy 数据库中的碳水化合物酶分类, 大秃马勃碳水化合物酶可划分为 6 大类 97 个超家族(表 2), 其中糖苷水解酶(Glycoside Hydrolases, GHs)和辅助酶类(Auxiliary Activities,

AAs)的数量明显大于其他类别, 这也与马勃的腐生生活方式一致。在所鉴定到的 GHs 中, 大秃马勃基因组中共发现 8 个  $\alpha$ -淀粉酶家族成员, 即 GH13 家族成员<sup>[59]</sup>。综合手动 BLAST、CD Search 和 SignalP 信号肽分析, 发现 GH13 家族仅有 1 个蛋白具有淀粉结合功能域(Starch-Binding Domain, SBD), 即  $\alpha$ -淀粉酶碳水化合物结合模块(Carbohydrate-Binding Module, CBM), 而且具有信号肽(1-15 aa) (图 4)。该蛋白大小为 558 个氨基酸, 其编码基因位于基因组的 Contig 3 正链 241 405-243 830 位, 含有 12 个外显子和 11 个内含子。糖基化预测结果显示该蛋白无 N-糖基化修饰位点, 2 个潜在的 Epsilon 氨基糖基化的赖氨酸位点(第 2 和第 543 位赖氨酸), 1 个 O-( $\alpha$ )-糖基化位点(第 32 位丝氨酸)和 1 个 C-甘露糖基化位点(第 416 位色氨酸) (图 4)。根据前期研究<sup>[15-16]</sup>, 推测该蛋白具有抗丹宁  $\alpha$ -淀粉酶活性, 可考虑后续对其进行克隆表达和酶活性特征研究分析。

表 2 大秃马勃 5.752 碳水化合物酶(CAZymes)信息总览  
Table 2 Overview of the CAZymes of *Calvatia gigantea* 5.752

分类 Category	蛋白数量 Numbers of protein	超家族数量 Numbers of superfamily
碳水化合物结合模块类 Carbohydrate-binding modules (CBMs)	2	1
糖苷水解酶类 Glycoside hydrolases (GHs)	169	50
糖基转移酶类 Glycosyl transferases (GTs)	48	22
多糖裂解酶类 Polysaccharide lyases (PLs)	9	7
碳水化合物酯酶类 Carbohydrate esterases (CEs)	20	8
辅助酶类 Auxiliary activities (AAs)	69	9

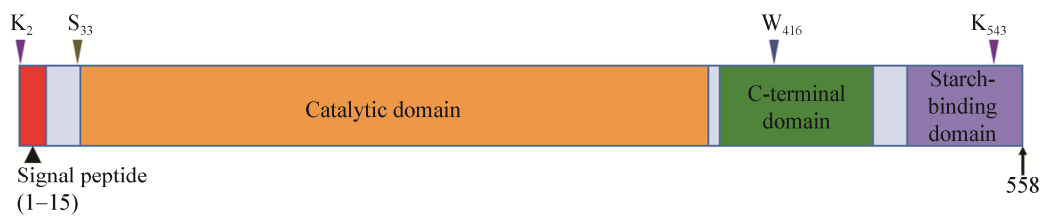


图 4 潜在抗丹宁  $\alpha$ -淀粉酶的结构功能域和糖基化位点预测结果

Figure 4 Predicted function domains and glycosylation sites of the potential tannin-resistant  $\alpha$ -amylase

Note: K<sub>2</sub>, K<sub>543</sub>: Glycation of  $\epsilon$  amino group of lysine; S<sub>33</sub>: O-( $\alpha$ )-GlcNAc glycosylation site; W<sub>416</sub>: C-mannosylation site

2.6 大秃马勃 18 个次级代谢物生物合成基因簇的鉴定

使用 antiSMASH 5.2.0, 本研究在大秃马勃基因组的 10 个重叠群(Contigs)中鉴定了 18 个次级代谢物基因簇, 其中包括 9 个萜类(Terpene)基因簇、5 个非核糖体肽合成酶样(NRPS-like)基因簇、2 个 I 型聚酮合酶(T1PKS)基因簇、1 个吲哚(Indole)基因簇和 1 个铁载体(Siderophore)基因簇(表 3)。所有萜类基因簇中都有萜类合成酶, 在 I 型聚酮合酶和非核糖体肽合成酶样基因簇中分别都有编码 PKS 和 NRPS 的基因。这些基因簇和已知的 antiSMASH 数据库中的基因簇相似性不高(antiSMASH 5.2.0 未给出相似度数值)。

3 讨论

本研究综合二代和三代测序数据, 经过组装, 首次提供了一株马勃真菌的基因组信息, 该基因组

完整性较高, 预测和注释了 10 444 个蛋白编码基因, 其中有 649 个新蛋白编码基因, 为今后了解马勃真菌生物学提供了重要的遗传信息基础。但基因组仍有较大部分(44.27%)蛋白未得到功能注释, 后期需要系统的功能基因组学研究。本研究基于基因组数据中的 rDNA ITS 序列, 将该株马勃真菌鉴定为大秃马勃。基因组测序将为真菌鉴定和进化提供更丰富的信息并促进更为准确的分析<sup>[60]</sup>, 但目前大秃马勃为首次测序的马勃真菌, 后续研究仍需丰富的马勃真菌基因组测序研究。

本研究鉴定出 317 个碳水化合物酶, 为了解马勃真菌腐生生活方式提供了信息, 结合前人研究, 鉴定出一个潜在的可分泌胞外的抗丹宁  $\alpha$ -淀粉酶, 为拓展可食用马勃真菌资源在食品工业上的应用提供了基因信息<sup>[15]</sup>, 但有价值的碳水化合物酶仍有待开发和应用。另外, 本研究鉴定出 18 个次级代

表 3 大秃马勃 5.752 基因组中 18 个次级代谢物合成基因簇总览

Table 3 Summary of the 18 secondary metabolite biosynthetic gene clusters of the *Calvatia gigantea* 5.752

序号 No.	基因组中的位置 Location in genome			次级代谢物生物合成基因簇类型 Type of secondary metabolite biosynthetic gene cluster
	重叠群序号 No. of contig	起始位置 Start	终止位置 End	
1	2	789 186	810 624	萜类 Terpene
2	3	253 643	275 108	吲哚 Indole
3	4	2 046 382	2 094 084	I 型聚酮合酶 T1PKS
4	5	109 182	128 607	萜类 Terpene
5	5	2 507 366	2 553 112	非核糖体肽合成酶样 NRPS-like
6	5	2 722 722	2 744 624	萜类 Terpene
7	6	85 416	106 750	萜类 Terpene
8	6	1 274 541	1 318 780	非核糖体肽合成酶样 NRPS-like
9	7	210 687	254 192	非核糖体肽合成酶样 NRPS-like
10	7	1 232 923	1 254 141	萜类 Terpene
11	9	579 022	599 840	萜类 Terpene
12	9	1 339 149	1 383 158	非核糖体肽合成酶样 NRPS-like
13	9	1 839 653	1 855 402	铁载体 Siderophore
14	10	1 151 654	1 196 330	I 型聚酮合酶 T1PKS
15	10	1 449 476	1 471 235	萜类 Terpene
16	11	1 300 674	1 345 205	非核糖体肽合成酶样 NRPS-like
17	14	305 930	325 122	萜类 Terpene
18	14	2 382 491	2 403 494	萜类 Terpene

谢物生物合成基因簇, 包含萜类、非核糖体肽和聚酮化合物合成基因簇。这些次级代谢物结构复杂多样<sup>[61-62]</sup>, 往往具有重要的生理、生物活性, 如抑菌、降胆固醇、免疫抑制和抗肿瘤等作用<sup>[63]</sup>。虽然次级代谢物中的各种化合物及其衍生物的多样性不断被揭示<sup>[64]</sup>, 但次级代谢物生物合成过程复杂, 人们对其知之甚少, 仍需要大量研究<sup>[61,65]</sup>。尽管如此, 本研究发现的基因簇丰富了次级代谢物生物合成基因资源<sup>[64]</sup>, 也为未来发掘具有重要医用价值的次级代谢物提供信息基础<sup>[61]</sup>。

我们期待, 本研究将在首次提供的马勃真菌基因组基础上, 加快马勃真菌的基因组测序研究, 并进一步促进对马勃遗传进化、基因功能和生物学过程的研究, 以及其在食品工业应用与药用价值开发研究。

## REFERENCES

- [1] Zhao HZ, Xu YY, Fu XY, Fan L. The progress of food and medical on values of puff-balls[J]. Microbiology China, 2007, 34(2): 367-369 (in Chinese)  
赵会珍, 胥艳艳, 付晓燕, 范黎. 马勃的食药价值及其研究进展[J]. 微生物学通报, 2007, 34(2): 367-369
- [2] Eroğlu C, Seçme M, Atmaca P, Kaygusuz O, Gezer K, Bağcı G, Dodurga Y. Extract of *Calvatia gigantea* inhibits proliferation of A549 human lung cancer cells[J]. Cytotechnology, 2016, 68(5): 2075-2081
- [3] Pan X, Zhang J, Zhu TH, Liang HP. Analysis of puffballs' phylogenetic relationships from *Eucalyptus grandis* forest[J]. Scientia Silvae Sinicae, 2011, 47(5): 68-75 (in Chinese)  
潘欣, 张健, 朱天辉, 梁洪萍. 巨桉林下马勃分子系统发育的关系[J]. 林业科学, 2011, 47(5): 68-75
- [4] Sui X, Xie Y, Yang F, Fu D, Li P, Huang B. The research progress of medicinal fungi *Lasiosphaera* seu *Calvatia*[J]. Heilongjiang Science, 2016, 7(21): 12-13,24 (in Chinese)  
隋新, 谢洋, 杨帆, 付东, 李鹏, 黄波. 药用真菌马勃的研究进展[J]. 黑龙江科学, 2016, 7(21): 12-13,24
- [5] Kivrak I, Kivrak S, Harmandar M. Bioactive compounds, chemical composition, and medicinal value of the giant puffball, *Calvatia gigantea* (higher basidiomycetes), from Turkey[J]. International Journal of Medicinal Mushrooms, 2016, 18(2): 97-107
- [6] Lam YW, Ng TB, Wang HX. Antiproliferative and antimitogenic activities in a peptide from puffball mushroom *Calvatia caelata*[J]. Biochemical and Biophysical Research Communications, 2001, 289(3): 744-749
- [7] Ng TB, Lam YW, Wang HX. Calcaelin, a new protein with translation-inhibiting, antiproliferative and antimitogenic activities from the mosaic puffball mushroom *Calvatia caelata*[J]. Planta Medica, 2003, 69(3): 212-217
- [8] Lee S, Kim MJ, Lee BS, Ryoo R, Kim HK, Kim KH. Cumulative effects of constituents from the mushroom *Calvatia nipponica* on the contractility of penile corpus cavernosum smooth muscle[J]. Mycobiology, 2020, 48(2): 153-156
- [9] Larsson E, Jeppson M. Phylogenetic relationships among species and genera of *Lycoperdaceae* based on ITS and LSU sequence data from north European taxa[J]. Mycological Research, 2008, 112(1): 4-22
- [10] Kivrak İ, Kivrak Ş, Harmandar M. Free amino acid profiling in the giant puffball mushroom (*Calvatia gigantea*) using UPLC-MS/MS[J]. Food Chemistry, 2014, 158: 88-92
- [11] Ogbale OO, Nkumah AO, Linus AU, Falade MO. Molecular identification, *in vivo* and *in vitro* activities of *Calvatia gigantea* (macro-fungus) as an antidiabetic agent[J]. Mycology, 2019, 10(3): 166-173
- [12] Zhao YS, Wang JP, Song AR, Wang NN, Han ZJ, Sun XB. Polysaccharide extraction from *Calvatia gigantea* and its *in vitro* anti-tumor activity[J]. Chinese Journal of Modern Applied Pharmacy, 2012, 29(7): 574-578 (in Chinese)  
赵友生, 王进平, 宋爱荣, 王宁宁, 韩志建, 孙晓波. 马勃多糖提取及体外抗肿瘤研究[J]. 中国现代应用药学, 2012, 29(7): 574-578
- [13] Galiotou-Panayotou M, Rodis P, Macris BJ, Stathakos D. Purification of a novel enzyme involved in catechin degradation by *Calvatia gigantea*[J]. Applied Microbiology and Biotechnology, 1988, 28(6): 543-545
- [14] Galiotou-Panayotou M, Macris BJ. Degradation of condensed tannins by *Calvatia gigantea*[J]. Applied Microbiology and Biotechnology, 1986, 23(6): 502-506
- [15] Komninos J, Kekos D, Macris BJ, Galiotou-Panayotou M. Tannin-resistant  $\alpha$ -amylase from *Calvatia gigantea*[J]. Biotechnology and Bioengineering, 1988, 32(7): 939-941
- [16] Kekos D, Macris BJ. Production and characterization of amylase from *Calvatia gigantea*[J]. Applied and Environmental Microbiology, 1983, 45(3): 935-941
- [17] Christakopoulos P, Tzia C, Kekos D, Macris BJ. Production and characterization of extracellular lipase from *Calvatia gigantea*[J]. Applied Microbiology and Biotechnology, 1992, 38(2): 194-197
- [18] Chen YT, Hou DH, Yang X, Wang FY, Zhao YX, Gao YY, Lin WP, Sun TY. Effects of illumination and pH on the growth and antioxidant activities of *Pycnoporus sanguineus*[J]. Journal of Microbiology, 2019, 39(3): 29-34 (in Chinese)  
陈艳通, 侯典海, 杨晓, 王福悦, 赵亚烜, 高媛媛, 林维

- 平, 孙同毅. 光照和 pH 对血红密孔菌生长及抗氧化活性的影响[J]. 微生物学杂志, 2019, 39(3): 29-34
- [19] Saghai-Maroo MA, Soliman KM, Jorgensen RA, Allard RW. Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics[J]. Proceedings of the National Academy of Sciences of the United States of America, 1984, 81(24): 8014-8018
- [20] Lin WP, Jia GT, Sun HY, Sun TY, Hou DH. Genome sequence of the fungus *Pycnoporus sanguineus*, which produces cinnabarinic acid and pH- and thermo-stable laccases[J]. Gene, 2020, 742: 144586
- [21] Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation[J]. Genome Research, 2017, 27(5): 722-736
- [22] Istace B, Friedrich A, D'Agata L, Faye S, Payen E, Beluche O, Caradec C, Davidas S, Cruaud C, Liti G, et al. *De novo* assembly and population genomic survey of natural yeast isolates with the Oxford Nanopore MinION sequencer[J]. GigaScience, 2017, 6(2): giw018
- [23] Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2[J]. Nature Methods, 2020, 17(2): 155-158
- [24] Chaisson MJ, Tesler G. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory[J]. BMC Bioinformatics, 2012, 13(1): 238
- [25] Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng QD, Wortman J, Young SK, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement[J]. PLoS One, 2014, 9(11): e112963
- [26] Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs[J]. Bioinformatics, 2015, 31(19): 3210-3212
- [27] Stanke M, Steinkamp R, Waack S, Morgenstern B. AUGUSTUS: a web server for gene finding in eukaryotes[J]. Nucleic Acids Research, 2004, 32(suppl\_2): W309-W312
- [28] Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training[J]. Genome Research, 2008, 18(12): 1979-1990
- [29] Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders[J]. Bioinformatics, 2004, 20(16): 2878-2879
- [30] Korf I. Gene finding in novel genomes[J]. BMC Bioinformatics, 2004, 5(1): 59
- [31] Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments[J]. Genome Biology, 2008, 9(1): R7
- [32] McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools[J]. Nucleic Acids Research, 2004, 32(Suppl 2): W20-W25
- [33] Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences[J]. Nucleic Acids Research, 2016, 44(D1): D286-D293
- [34] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Michael Cherry J, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene Ontology: tool for the unification of biology[J]. Nature Genetics, 2000, 25(1): 25-29
- [35] Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets[J]. Nucleic Acids Research, 2012, 40(D1): D109-D114
- [36] El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, et al. The Pfam protein families database in 2019[J]. Nucleic Acids Research, 2019, 47(D1): D427-D432
- [37] Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003[J]. Nucleic Acids Research, 2003, 31(1): 365-370
- [38] Chan PP, Lowe TM. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences[M]. Gene Prediction, 2019
- [39] Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. Rfam: annotating non-coding RNAs in complete genomes[J]. Nucleic Acids Research, 2005, 33(Suppl 1): D121-D124
- [40] Tarailo-Graovac M, Chen NS. Using RepeatMasker to identify repetitive elements in genomic sequences[J]. Current Protocols in Bioinformatics, 2009, Chapter 4: Unit 4.10
- [41] Bao WD, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes[J]. Mobile DNA, 2015, 6(1): 11
- [42] Benson G. Tandem repeats finder: a program to analyze DNA sequences[J]. Nucleic Acids Research, 1999, 27(2): 573-580
- [43] Nilsson RH, Larsson KH, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, Schigel D, Kennedy P, Picard K, Glöckner FO, Tedersoo L, et al. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications[J]. Nucleic Acids Research, 2019, 47(D1): D259-D264
- [44] Katoh K, Kuma KI, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment[J]. Nucleic Acids Research, 2005, 33(2): 511-518
- [45] Innis MA, Garfield DH, Sninsky JJ, White TJ. PCR Protocols: A Guide to Methods and Applications[M]. Pittsburgh, Academic Press, 1990: 315-322

- [46] Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era[J]. *Molecular Biology and Evolution*, 2020, 37(5): 1530-1534
- [47] Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies[J]. *Molecular Biology and Evolution*, 2015, 32(1): 268-274
- [48] Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates[J]. *Nature Methods*, 2017, 14(6): 587-589
- [49] Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND[J]. *Nature Methods*, 2015, 12(1): 59-60
- [50] Zhang H, Yohe T, Huang L, Entwistle S, Wu PZ, Yang ZL, Busk PK, Xu Y, Yin YB. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation[J]. *Nucleic Acids Research*, 2018, 46(W1): W95-W101
- [51] Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline[J]. *Nucleic Acids Research*, 2019, 47(W1): W81-W87
- [52] Lu SN, Wang JY, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, et al. CDD/SPARCLE: the conserved domain database in 2020[J]. *Nucleic Acids Research*, 2020, 48(D1): D265-D268
- [53] Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, Von Heijne G, Nielsen H. SignalP 5.0 improves signal peptide predictions using deep neural networks[J]. *Nature Biotechnology*, 2019, 37(4): 420-423
- [54] Johansen MB, Kierner L, Brunak S. Analysis and prediction of mammalian protein glycation[J]. *Glycobiology*, 2006, 16(9): 844-853
- [55] Gupta R, Jung E, Gooley AA, Williams KL, Brunak S, Hansen J. Scanning the available *Dictyostelium discoideum* proteome for O-linked GlcNAc glycosylation sites using neural networks[J]. *Glycobiology*, 1999, 9(10): 1009-1022
- [56] Julenius K. NetCGlyc 1.0: prediction of mammalian C-mannosylation sites[J]. *Glycobiology*, 2007, 17(8): 868-876
- [57] Chen H, Chen DQ, Li QF, Li PF, Chen H, Zhao YY. Research progress on pharmacology, pharmacokinetics and determination of ergosta-4, 6, 8(14), 22-tetraen-3-one[J]. *China Journal of Chinese Materia Medica*, 2014, 39(20): 3905-3909 (in Chinese)  
陈哈, 陈丹倩, 李全福, 李鹏飞, 陈华, 赵英永. 麦角甾酮的药理活性、药代动力学及含量测定研究进展[J]. *中国中药杂志*, 2014, 39(20): 3905-3909
- [58] Xu WY, Li A, Hu D, Zhang ZC, Wang SM. Research on pharmacognosics of medicinal fungi *Lycoperdon wrightii*[J]. *Edible Fungi of China*, 2018, 37(4): 47-51 (in Chinese)  
徐维茵, 李昂, 胡德, 张智超, 王淑敏. 药用真菌白刺马勃的生药学研究[J]. *中国食用菌*, 2018, 37(4): 47-51
- [59] Stam MR, Danchin EGJ, Rancurel C, Coutinho PM, Henrissat B. Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of  $\alpha$ -amylase-related proteins[J]. *Protein Engineering, Design and Selection*, 2006, 19(12): 555-562
- [60] Hibbett D. The invisible dimension of fungal diversity[J]. *Science*, 2016, 351(6278): 1150-1151
- [61] Keller NP, Turner G, Bennett JW. Fungal secondary metabolism—from biochemistry to genomics[J]. *Nature Reviews Microbiology*, 2005, 3(12): 937-947
- [62] Anke T, Schöffler A. *Physiology and Genetics: Selected Basic and Applied Aspects*[M]. Heidelberg, Springer, 2018: 231-275
- [63] Martín JF, García-Estrada C, Zeilinger S. *Biosynthesis and Molecular Genetics of Fungal Secondary Metabolites*[M]. New York: Springer New York, 2014
- [64] Hautbergue T, Jamin EL, Debrauwer L, Puel O, Oswald IP. From genomics to metabolomics, moving toward an integrated strategy for the discovery of fungal secondary metabolites[J]. *Natural Product Reports*, 2018, 35(2): 147-173
- [65] Schueffler A, Anke T. Fungal natural products in research and development[J]. *Natural Product Reports*, 2014, 31(10): 1425-1448