

研究报告



一株稀有海洋真菌 *Stachybotrys longispora* FG216 的 *De Novo* 测序及基因组学分析

曹玉陵 唐思敏 张添 包斌 吴文惠*

上海海洋大学食品学院海洋生物制药实验室 上海 201306

摘要:【背景】长孢葡萄穗霉菌(*Stachybotrys longispora*) FG216 是一株稀有海洋真菌, 其次生代谢产物 FGFC1 具有纤溶活性。进行 *S. longispora* FG216 的基因组序列分析, 将充实和促进海洋微生物功能基因和次生代谢产物合成生物学的基础研究和应用研究。【目的】解析 *S. longispora* FG216 的基因组序列, 分析基因组生物功能和同源相似性关系, 分析次生代谢产物纤溶活性化合物 FGFC1 的相关基因。【方法】基于 Illumina HiSeq 高通量测序平台对 *S. longispora* FG216 菌株进行 *De Novo* 测序, 使用 SSPACE、Augustus 等软件进行组装、编码基因预测、基因功能注释、物种共线性分析以及预测 FGFC1 次生代谢产物合成基因簇。【结果】*S. longispora* FG216 的基因组测序总长度为 45 622 830 bp, 共得到 605 个 Scaffold, GC 含量为 51.31%, 注释预测得到 13 329 个编码基因和 169 个非编码 RNA。基因组测序数据提交至国家微生物科学数据中心(编号为 NMDC60016264), 其中 13 053、8 422、8 460、7 714 和 2 847 个基因分别能够在 NR、KEGG、KOG、GO 和 CAZy 数据库匹配到注释信息。比较基因组学分析发现, *Stachybotrys* 具有保守性, 核心基因占基因家族总数目的 71.44%, *S. longispora* FG216 与 *S. chlorohalonata* IBT 40285 的相似性最高; 同时, 预测得到 101 个次生代谢产物合成基因簇, 其中 18 个基因簇与已知的化合物相匹配。通过 antiSMASH 预测, Cluster 57 是编码合成 FGFC1 母核结构异吡啶酮的基因簇, 与 *S. chlorohalonata* IBT 40285 中的基因簇相似度为 40%。【结论】海洋稀有真菌 *S. longispora* FG216 的基因组信息已上传至国家微生物科学数据中心公开使用, 为 *Stachybotrys* 种属的研究提供了重要的参考意义, 同时发现了 *S. longispora* FG216 次生代谢产物纤溶活性化合物 FGFC1 母核部分编码基因是 Cluster 57。

关键词: 长孢葡萄穗霉菌, *De Novo* 测序, 基因功能注释, 次生代谢, 比较基因组学

De novo sequencing and genomics analysis of a rare marine fungus *Stachybotrys longispora* FG216

CAO Yuling TANG Simin ZHANG Tian BAO Bin WU Wenhui*

Marine Biopharmaceutical Laboratory, School of Food Science, Shanghai Ocean University, Shanghai 201306, China

Abstract: [Background] *Stachybotrys longispora* FG216 is a rare marine fungus, and its secondary

Foundation items: Shanghai Innovation Action Project (1749074222500); National Natural Science Foundation of China (81502955, 81750110548)

*Corresponding author: E-mail: whwu@shou.edu.cn

Received: 22-02-2021; **Accepted:** 01-04-2021; **Published online:** 17-05-2021

基金项目: 上海市创新行动计划(1749074222500); 国家自然科学基金(81502955, 81750110548)

*通信作者: E-mail: whwu@shou.edu.cn

收稿日期: 2021-02-22; **接受日期:** 2021-04-01; **网络首发日期:** 2021-05-17

metabolite FGFC1 has fibrinolytic activity. The genome sequence analysis of *S. longispora* FG216 will enrich and promote the basic research and applied research of synthetic biology of marine microbial functional genes and secondary metabolites. **[Objective]** To detected the genome sequence of *S. longispora* FG216, analyze the biological function and homologous similarity relation of the genome, and reveal the related secondary metabolite genes of the fibrinolytic active compound FGFC1. **[Methods]** Based on Illumina HiSeq high-throughput sequencing platform, *S. longispora* FG216 strain was *de novo* sequenced. SSPACE, Augustus and other software were used to perform assembly, encoding gene prediction, gene function annotation, species collinear analysis and prediction of FGFC1 secondary metabolite synthesis gene cluster. **[Results]** The total sequence length of *S. longispora* FG216 genome was 45 622 830 bp, and a total of 605 scaffolds were obtained, with a GC content of 51.31%. Annotations predicted 13 329 coding genes and 169 non-coding RNAs. The genome sequencing data submitted to the National Microbiology Science Data Center numbered NMDC60016264, of which 13 053, 8 422, 8 460, 7 714 and 2 847 genes can be matched with the annotation information in NR, KEGG, KOG, GO and CAZy databases respectively. According to comparative genomics analysis, *Stachybotrys* is conservative, with core genes accounting for 71.44% of the total gene families. *S. longispora* FG216 and *S. chlorohalonata* IBT 40285 had the highest similarity. At the same time, 101 secondary metabolite synthesis gene clusters were predicted, among them 18 gene clusters matched the known compounds. Under the antiSMASH predicts that Cluster 57 is a gene cluster encoding the synthesis of the parent nuclear structure of FGFC1 isoidindolinone, and the similarity of the gene cluster with *S. chlorohalonata* IBT 40285 is 40%. **[Conclusion]** The genome information of the rare marine fungus *S. longispora* FG216 has been uploaded to the National Microbial Science Data Center for public use, which provides important reference for the study of *Stachybotrys* species. Cluster 57 was also found to be the partial encoding gene of the parent nucleus of the fibrinolytic active compound FGFC1, a secondary metabolite of *S. longispora* FG216.

Keywords: *Stachybotrys longispora*, *de novo* sequencing, gene function annotation, secondary metabolism, comparative genomics

葡萄穗霉属(*Stachybotrys*)是丝状真菌,隶属于葡萄穗霉科(*Stachybotryaceae*),其下包含 75 个种属,多出现在富含纤维素的潮湿环境中,通过降解腐烂植物中的纤维素补充养分^[1]。

从 *Stachybotrys* 的次生代谢产物中发现了许多结构新颖且具有显著生物活性的化合物。例如:从海绵中分离出的真菌 *Stachybotrys chartarum* 代谢产生的新型异吡啶酮型生物碱显著抑制 HIV-1 的毒性^[2];海百合内生真菌 *Stachybotrys chartarum* 952 的次生代谢化合物被确认具有降脂和降胆固醇的生物活性^[3];一株 *Stachybotrys chartarum* MF347 代谢产生的新型化合物能够选择性抑制耐甲氧西林的金黄色葡萄球菌的生长^[4]。随着对 *Stachybotrys* 研究的深入,更多次生代谢化合物的生物活性被解析,特别是生物活性优良且结构新颖的化合物有可能成为重要的先导化合物。

截至目前, *Stachybotrys* 的基因组学研究甚少,NCBI 仅存 4 种 *S. chartarum* 和 1 种 *Stachybotrys chlorohalonata* 共 5 条基因组记录^[5]。研究发现 *Stachybotrys longispora* FG216 代谢产生显著纤溶活性的次生代谢产物 FGFC1,显示出优良的海洋溶栓药物成药性^[6]。针对海洋稀有真菌 *S. longispora* 的基因组学研究,通过 *De Novo* 测序并进行基因预测和功能注释,分析次生代谢产物合成基因簇,基于比较基因组学建立海洋稀有真菌 *S. longispora* 的种属关系。

1 材料与方法

1.1 供试菌株

S. longispora FG216 分离自舟山离岸 100 m 的海水中,保藏在上海海洋大学海洋生物制药实验室,在中国典型培养物保藏中心的保藏号为 CCTCCM2012272。

1.2 主要试剂和仪器及培养基

PDA 培养基, 青岛高科技工业园海博生物技术有限公司; 酵母提取物、细菌学蛋白胨, Oxoid 公司; 无水乙醇, 国药集团化学试剂有限公司; 异丙醇、 β -巯基乙醇, 阿拉丁试剂(上海)有限公司; Ezup 柱式真菌基因组 DNA 抽提试剂盒, 生工生物工程(上海)股份有限公司。

恒温摇床, 太仓市华美生化仪器厂; 电泳仪, 北京六一生物科技有限公司; NanoDrop 2000 超微量分光光度计, 赛默飞世尔科技(中国)有限公司。

培养基的配制参考文献[7] (g/L): 葡萄糖 35.0, 可溶性淀粉 10.0, 脱脂大豆粉 20.0, 蛋白胨 5.0, 牛肉膏蛋白胨 5.0, 酵母提取物 3.0, 氯化钠 2.0, 磷酸氢二钾 0.5, 硫酸镁 0.05。调节 pH 为 5.8, 1×10^5 Pa 灭菌 20 min。

1.3 培养条件

取出冻藏的 *S. longispora* FG216 菌种于 37 °C 水浴锅预热, 挑取菌液接种到灭菌的 PDA 培养基中, 28 °C 恒温培养 3 d 长出菌丝体, 使用接种环挑取长势较好的菌落接种至灭菌培养基中, 28 °C、180 r/min 恒温培养 5 d, 最后 10 000 r/min 离心 5 min 收集菌体, 并使用 PBS 重悬洗涤菌体 2 次, 以避免培养基对实验造成误差。

1.4 DNA 提取及质量检测

称取收集的菌体约 100 mg 放入研钵中, 加入液氮快速研磨成粉末, 然后按照 Ezup 柱式真菌基因组 DNA 抽提试剂盒的操作说明提取基因组并溶于 50 μ L TE 缓冲液中, 取 5 μ L 用于琼脂糖凝胶电泳检测基因组纯度, 使用 NanoDrop 2000 超微量分光光度计检测基因组浓度。

1.5 基因组测序和组装分析

经质量检测合格的基因组 DNA 交付给苏州金唯智生物科技有限公司构建测序文库。首先将大片段 DNA 打断约为 500 bp 大小的片段, 将打断形

成的粘性末端修复为平末端, 通过 3'端加碱基“A”, 使得 DNA 片段能与 3'端带有“T”碱基的接头连接, 电泳回收目的片段连接产物。参考 VAHTS™ Universal DNA Library Prep Kit for Illumina® V3 试剂盒说明书对两端带有接头的 DNA 片段进行 PCR 扩增, 最后用检验合格的文库进行 Cluster 制备和测序。使用二代测序数据质量统计软件 Cutadapt (V1.9.1)对测序原始数据去除接头和低质量序列等, 得到后续信息分析用的 Clean Data。

基于优化后的数据, 先使用 Velvet (Version 1.2.10)软件进行 k-mer 分析构建 De Bruijn 图, 拼接出分段的 Contig 序列。接着应用 SSPACE (Version 3.0)软件将文库测序得到的 Reads 比对到 Contig 序列上, 操作 GapFiller (Version 1-10)软件将所有文库测序得到的 Reads 比对到 Scaffold 序列上。最终得到未知碱基 N 比例较低、序列长度较长的 Scaffold 序列。

1.6 编码基因和非编码 RNA 预测

Augustus (Version 3.3)软件可以有效预测编码基因的起始位点和编码区, 提高了对高 GC 序列预测的准确性, 在保证预测准确性的前提下, 降低了预测假阳性。非编码 RNA (Non-Coding RNA)是从基因组上转录而来, 并不翻译成蛋白, 在 RNA 水平上就能行使各自的生物学功能, 包括 rRNA、tRNA、snRNA、snoRNA 和 microRNA 等多种已知功能的 RNA, 还包括一些未知功能的 RNA。使用 Rfam (Version 12.0)数据库可以进行序列比对, 分析基因组中的非编码 RNA。

1.7 基因功能注释

基因功能注释是将预测得到的编码基因的蛋白序列与各数据库中包含的蛋白质进行序列比对。因此, 基于预测得到的编码基因的蛋白序列, 使用 BLAST (Version 2.2.31+)等软件与 NR、KEGG、GO 等数据库中的蛋白序列进行比对, 序列比对的 *E*-Value 设为 $1e-5$, 序列比对长度应不小于蛋白长度的 60%, 选取最优的匹配结果作为该

基因的注释结果,同时对预测出的基因进行功能注释分析。

基于 Diamond (Version 0.8.15)软件使用 National Center for Biotechnology Information (NCBI)的 BLAST 功能对基因组进行 NR 注释;使用 BLAST (Version 2.2.28+)软件通过与 Kyoto Encyclopedia of Genes and Genomes (KEGG)数据库比对进行 KEGG 注释;Blast2go (Version 2.5)用于 Gene Ontology (GO)数据库比对进行 GO 注释;Hmmscan (Version 3.1b2)软件则用于基因组的 KOG 注释;对比 Carbohydrate-Active Enzymes Database (CAZy)数据库进行 CAZy 注释。

1.8 基因组比较分析

基于 OrthoVenn2 (<https://orthovenn2.bioinfotoolkits.net>)网站分析 *Stachybotrys* 的直系同源基因。参数设置如下:*E*-Value 为 $1e-2$ 、Inflation Value 为 1.5,用来分析核心基因、兼有基因和独有基因。*Stachybotrys* 物种的基因组信息从 NCBI (<https://www.ncbi.nlm.nih.gov/>)下载。

1.9 次生代谢产物合成基因簇预测分析

antiSMASH (<https://fungismash.secondarymetabolites.org/>)是用于细菌和真菌抗生素及次生代谢产物合成基因簇预测的专业软件,运用 antiSMASH 软件 Detection Strictness 设置为 Relaxed, Extra Features 选择 KnownClusterBlast、ActiveSiteFinder、SubClusterBlast 和 RREFinder 对 *S. longispora* FG216 中次生代谢产物合成基因簇及其可能合成的化合物核心结构进行预测分析。

2 结果与分析

2.1 基因组 DNA 质量

S. longispora FG216 基因组 DNA 经 0.7%琼脂糖核酸电泳检测(图 1),在紫外线下显影出现单一条带,可以看出基因组大小在 23 kb 左右,有轻微的降解,没有 RNA 或者蛋白质的污染, NanoDrop 2000 检测核酸浓度为 55.9 ng/ μ L,说明提取的基因组 DNA 质量良好,可用于进一步的建库分析。

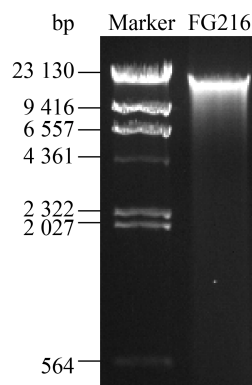


图 1 *S. longispora* FG216 提取的 DNA 电泳图

Figure 1 Electrophoresis of DNA extracted from *S. longispora* FG216

Note: Marker: λ /Hind III DNA Marker

2.2 基因组组装结果和基因组组分分析

对测序结果原始图像数据利用软件 Bcl2fastq 进行图像碱基识别,结果共检测出 Reads 的数量为 15 437 842 个,碱基的数量是 2 315 676 300 bp,相对应 Reads 的平均长度为 150 bp,Phred 数值大于 20 和 30 的碱基占总体碱基的百分比分别为 97.34%和 92.81%,其 GC 含量为 50.44%。使用二代测序数据质量统计软件 Cutadapt 过滤后数据 Reads 数量为 15 407 082 个,占原始数据的 99.80%;剩余碱基的数量是 2 300 824 378 bp,占总碱基的 99.36%,结果表明该基因组的测序结果可靠,数据精准有效、误差较小。

经过组装分析基因组序列共得到 605 条序列,总碱基长度达到 45 622 830 bp,其中最短的序列长度为 181 bp,最长的序列长度为 1 144 105 bp, N50 的长度为 295 293 bp, GC 含量为 51.31%, Ns 为 0%验证了基因组测序结果准确,基因组组装结果已提交至国家微生物科学数据中心,登录号为 NMDC60016264。如图 2 所示,测序覆盖平均深度在该基因组的 GC 值附近会出现一个峰值,而在该值两端偏离该值越远;散点图呈现出类似泊松分布的形状,表明分析测序数据并不存在 GC 偏向性,同时 Mapping Ratio 为 97.39%,基因组覆盖度为 100%,平均覆盖深度为 48.87,这为基因功能分析提供了良好的基础条件。

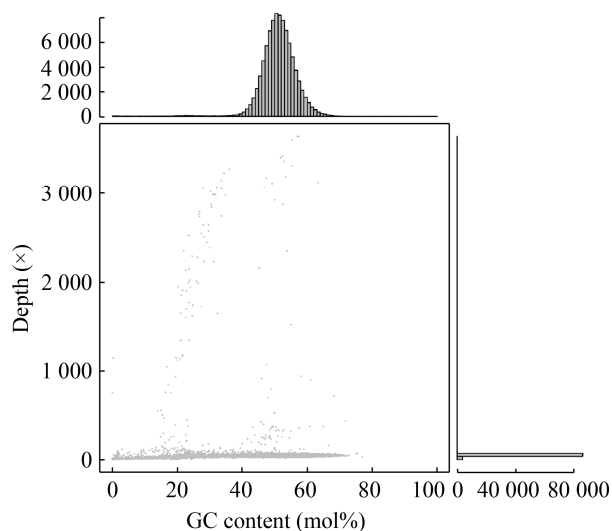


图2 *S. longispora* FG216 基因组 GC 含量与覆盖深度关联分析统计图

Figure 2 Correlation analysis statistical chart of GC content and coverage depth of *S. longispora* FG216 genome

基因组完成组装后, 使用软件 Augustus 对其编码基因进行预测, 共预测出 13 329 条编码基因, 编码基因的平均长度为 1 594.42 bp, 编码基因的总长度为 21 252 085 bp, 占基因组组装总碱基数目的 46.58%, N50 为 1 845 bp, GC 碱基比例为 54.72%, Ns 值为 0%。非编码 RNA 是不编码蛋白质的 RNA, 不仅包括 rRNA、tRNA、snRNA、snoRNA 和 microRNA 等多种已知功能的 RNA, 还包括一些未知功能的 RNA。该基因组共检测到 169 条 ncRNA, 包括 46 条 rRNA、86 条 tRNA 和 37 条其他非编码 RNA。基因组重复序列分析显示共出现 DNA 转座子 26 个, 其他散在分布于基因组中的重复序列 1 684 个、简单重复序列 7 083 个以及低复杂度序列 771 个。

2.3 基因功能注释

2.3.1 NR 数据库注释

基因功能注释的主要方法是将预测得到的编码基因的蛋白序列使用 BLAST 等比对软件与各数据库中包含的蛋白质进行序列比对, 选取最优的匹配结果作为该基因的注释结果。NR 数据库

(Non-Redundant Protein Database) 是一个非冗余的蛋白数据库, 经过比对, 13 053 个基因在 NR 数据库上得到注释, 占编码基因总数的 97.93%。

2.3.2 KEGG 代谢途径分析

KEGG 数据库是系统分析基因功能和基因组信息的数据库。基因功能划分为 6 大类, 包括细胞过程 (Cellular Processes)、环境信息处理 (Environmental Information Processing)、遗传信息处理 (Genetic Information Processing)、人类疾病 (Human Diseases)、新陈代谢 (Metabolism) 和生物体系统 (Organismal Systems)。 *S. longispora* FG216 基因组能够对应到 KEGG Pathway 的 8 422 个基因, 富集在 379 代谢通路中, 其中涉及基因数目最多的代谢通路为氨基酸代谢途径 (ko01230) 226 个基因, 氨基苯甲酸酯降解途径 (ko00627) 有 217 个基因, 碳代谢途径 (ko01200) 有 212 个基因, 嘌呤代谢途径 (ko00230) 有 193 个基因, 甘氨酸、丝氨酸和苏氨酸的代谢途径 (ko00260) 有 186 个基因, 如图 3 所示。统计表明, 在新陈代谢 (Metabolism) 这一大类中, 各种代谢均注释到较多的基因数量, 表明 *S. longispora* FG216 菌株具有代谢功能种类较多、能参与多种物质代谢途径转化的特点。

2.3.3 KOG 数据库注释

S. longispora FG216 属于真核生物, 因此使用子库 KOG 进行注释 (图 4) 可以推测该序列的功能。其共有 8 460 条基因序列匹配到相应的功能, 数目较多的功能类别分别是: 一般功能预测 (General Function Prediction Only) 1 541 个基因, 次生代谢产物生物合成、运输和分解代谢 (Secondary Metabolites Biosynthesis, Transport and Catabolism) 704 个, 翻译后修饰、蛋白质周转、蛋白伴侣 (Posttranslational Modification, Protein Turnover, Chaperones) 591 个, 碳水化合物的运输和代谢 (Carbohydrate Transport and Metabolism) 587 个, 表明 *S. longispora* FG216 菌株中参与该类反应过程的基因数较多, 具有丰富次生代谢反应为主的功能型蛋白。

图3 *S. longispora* FG216 基因组 KEGG 分析统计Figure 3 KEGG analysis of *S. longispora* FG216 genome

2.3.4 GO 数据库注释

GO 数据库定义了三大亚类, 基因组注释到基因种类和数量最多的是分子功能 (Molecular Function), 其次是生物途径 (Biological Process) 和细胞组分 (Cellular Component) (图5)。共有 7 714 个基因成功获得 GO 功能注释, 基因数目最多的类别分别是: 催化活性 (Catalytic Activity) 4 156 个、结合 (Binding) 3 762 个和代谢过程 (Metabolic Process) 3 319 个。

2.3.5 CAZy 数据库注释

CAZy 数据库将 2 847 个基因匹配到相关酶, 根据图 6 可以看出, 分类包括: 糖苷水解酶 (Glycoside Hydrolases, GH) 893 个, 糖基结合模块 (Carbohydrate-Binding Modules, CBM) 762 个, 糖基转移酶 (Glycosyl Transferases, GT) 710 个, 辅助功能的酶 (Auxiliary Activities, AA) 289 个, 糖脂酶 (Carbohydrate Esterases, CE) 121 个, 多糖裂合酶 (Polysaccharide Lyases, PL) 72 个。

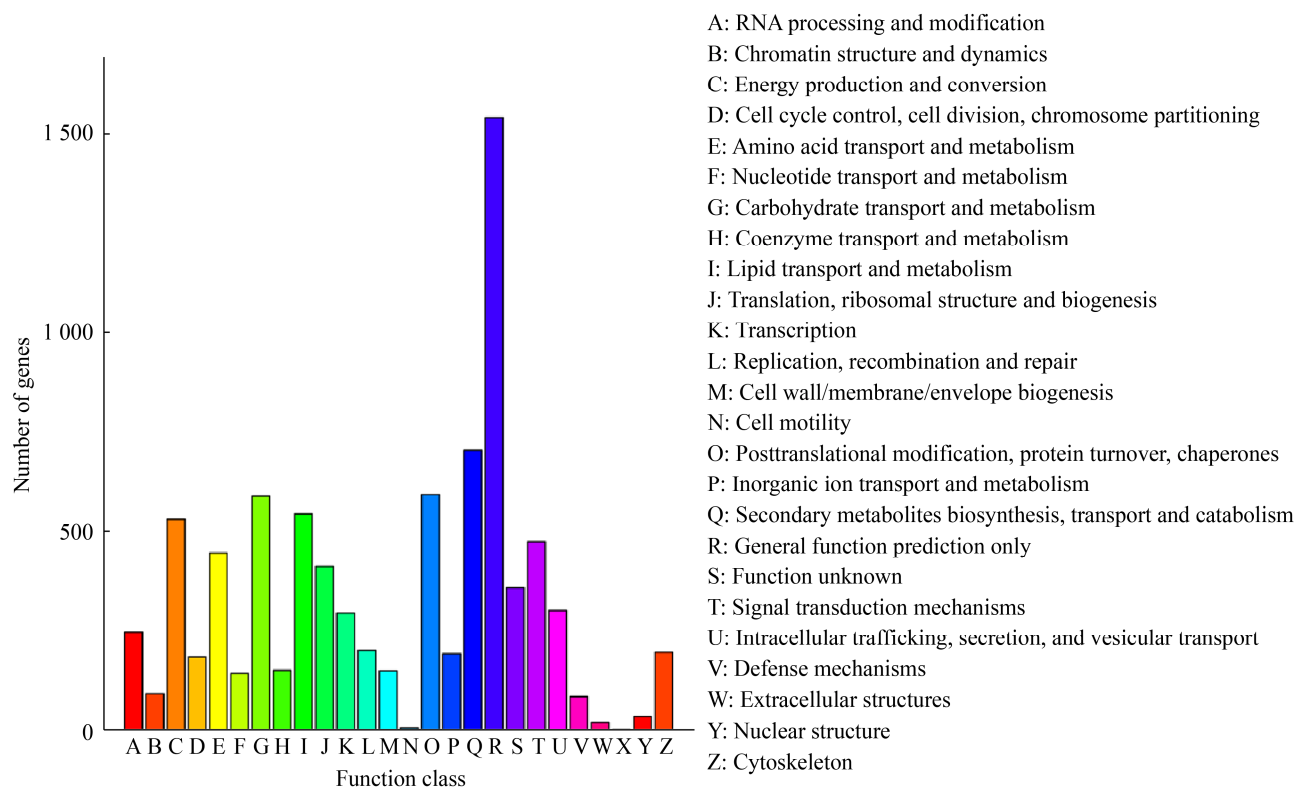


图 4 *S. longispora* FG216 基因组 KOG 分析统计
Figure 4 KOG analysis of *S. longispora* FG216 genome

2.4 比较基因组学分析

Stachybotrys, 属于致病菌, 可以产生多种对生物体有害的毒素。根据 NCBI 数据库信息展示, 对于葡萄穗霉属的基因组研究甚少, 因此检索选取引用已知基因组序列的 5 种菌 *S. chartarum*、*S. chartarum* IBT 40288、*S. chartarum* IBT 40293、*S. chartarum* IBT 7711 和 *S. chlorohalonata* IBT 40285 作为对照, 分析其基因组相关性。如表 1 所示, *S. longispora* FG216 具有最少的 Scaffolds (605) 和最多的 CDS 以及最长的碱基长度 45.6 Mb, 推测是由于近几年测序技术的快速发展, 使得基因组测序的结果更加精准和准确。

对 5 种菌株的基因序列使用 OrthoVenn2 进行差异比较, 结果如图 7 所示, 7 325 个基因家族是 5 种葡萄穗霉属共有的, 所占基因家族总数目的比例为 71.44%, 是该种属的核心基因。独有基因包括 *S. longispora* FG216 的 448 个基因家族、

S. chartarum IBT 40288 的 14 个基因家族、*S. chlorohalonata* IBT 40285 的 13 个基因家族、*S. chartarum* IBT 7711 的 3 个基因家族和 *S. chartarum* IBT 40293 的 3 个基因家族。*S. longispora* FG216 所含有的基因家族数量最少为 8 862 个, 独有基因在 5 种葡萄穗霉属中是最多的。通过成对基因组间重叠簇的热图可以看出, *S. longispora* FG216 与 *S. chlorohalonata* IBT 40285 基因组的相似性最高。

在 *Stachybotrys* 的核心基因中, 通过表 2 的 GO 富集分析可以看出, 基因功能与蛋白转运 (Protein Transport) 相关的数目最多, 为 97 个; 其次是 DNA 模板转录 (Transcription, DNA-Templated) 87 个, 氧化还原酶活性 (Oxidoreductase Activity) 84 个, 同时发现 4 个与黄曲霉毒素生物合成过程 (Aflatoxin Biosynthetic Process) 相关的功能基因。对于 *S. longispora* FG216 的独有基因, GO 分析显示功能基因数目最多的是碳水化合物运输

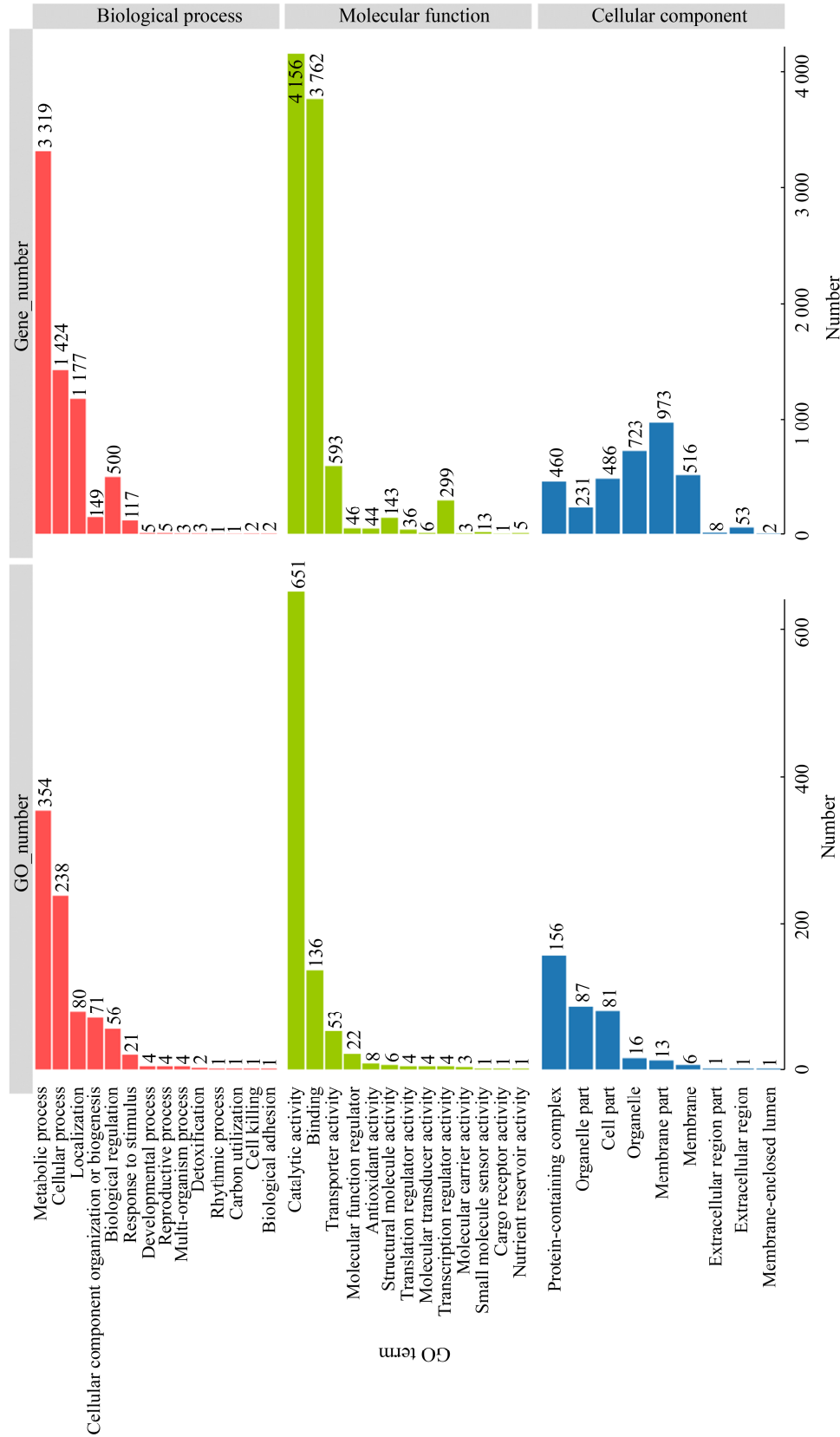


图 5 *S. longispora* FG216 基因组 GO 分析统计
Figure 5 GO analysis of *S. longispora* FG216 genome

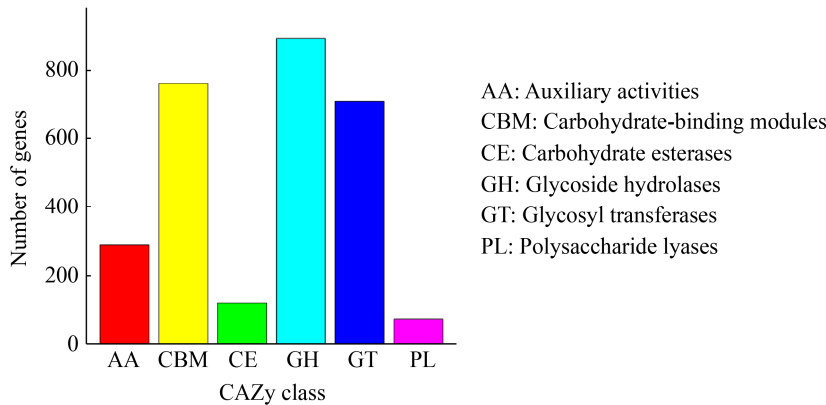


图 6 *S. longispora* FG216 基因组 CAZy 分析统计
Figure 6 CAZy analysis of *S. longispora* FG216 genome

表 1 *S. longispora* FG216 与其他 *Stachybotrys* 基因组比较
Table 1 Comparison of *S. longispora* FG216 genomes with other *Stachybotrys*

Items	<i>S. longispora</i> FG216	<i>S. chartarum</i>	<i>S. chartarum</i> IBT 40288	<i>S. chartarum</i> IBT 40293	<i>S. chartarum</i> IBT 7711	<i>S. chlorohalonata</i> IBT 40285
Scaffolds	605	1 434	2 351	2 342	2 290	2 802
Total length (Mb)	45.6	41.01	36.01	36.48	36.88	34.39
CDS	13 329	—	11 368	11 453	11 530	10 706
Largest scaffold (Mb)	1.144	4.955	0.433	0.762	0.861	0.484
N50 (bp)	295 293	1 219 284	126 356	200 047	176 820	114 117
GC (%)	51.31	51.2	53.4	53.2	53.1	53.2

注: —: 未找到相关 CDS。 *S. chartarum*、 *S. chartarum* IBT 40288、 *S. chartarum* IBT 40293、 *S. chartarum* IBT 7711 和 *S. chlorohalonata* IBT 40285 的登录号分别为 GCA_001021365.1、 GCA_000732765.1、 GCA_000732565.1、 GCA_000730325.1 和 GCA_000732775.1
Note: —: No related CDS found. *S. chartarum* GenBank accession No: GCA_001021365.1; *S. chartarum* IBT 40288 GenBank accession No: GCA_000732765.1; *S. chartarum* IBT 40293 GenBank accession No: GCA_000732565.1; *S. chartarum* IBT 7711 GenBank accession No: GCA_000730325.1; *S. chlorohalonata* IBT 40285 GenBank accession No: GCA_000732775.1

(Carbohydrate Transport) 和 多 糖 分 解 过 程 (Polysaccharide Catabolic Process), 它们均为 9 个, 其中发现有 4 个吲哚乙酰胺水解酶活性 (Indoleacetamide Hydrolase Activity) 相关的功能基因 (表 3)。 *S. longispora* FG216 和 *S. chlorohalonata* IBT 40285 具有最多的兼有基因 103 个, 基因功能富集在氧化还原酶活性 [Oxidoreductase Activity, Acting on NAD(P)H, Oxygen as Acceptor] 和丝氨酸型内肽酶活性 (Serine-Type Endopeptidase Activity)。

2.5 次生代谢产物合成基因簇预测分析结果

antiSMASH 软件预测 *S. longispora* FG216 基因组共得到 101 个基因簇, 其中, 编码聚酮合酶 (Polyketide Synthase, PKS) 和非核糖体多肽合成酶

(Non-Ribosomal Peptide Synthase, NRPS) 占基因簇的主要部分, T1 PKS 有 53 个, NRPS-Like 有 22 个, NRPS 有 19 个, 参与萜烯合成 (Terpene) 有 12 个, 吲哚合成 (Indole) 有 9 个。 *S. longispora* FG216 基因组编码的蜂蜜曲菌素 ((-)-Mellein) 基因簇 Cluster 1 与 *Parastagonospora nodorum* 的生物合成基因簇相似性是 100%; 6-甲基水杨酸 (6-Methylsalicylic Acid) 基因簇 Cluster 15 与 *Aspergillus terreus* 的生物合成基因簇相似性是 100%; 二甲基辅酶 (Dimethylcoprogen) 基因簇 Cluster 55 与 *Alternaria alternata* 的生物合成基因簇相似性是 100%; AbT1 的基因簇 Cluster 73 与 *Aureobasidium pullulans* 的生物合成基因簇相似性是 100%。

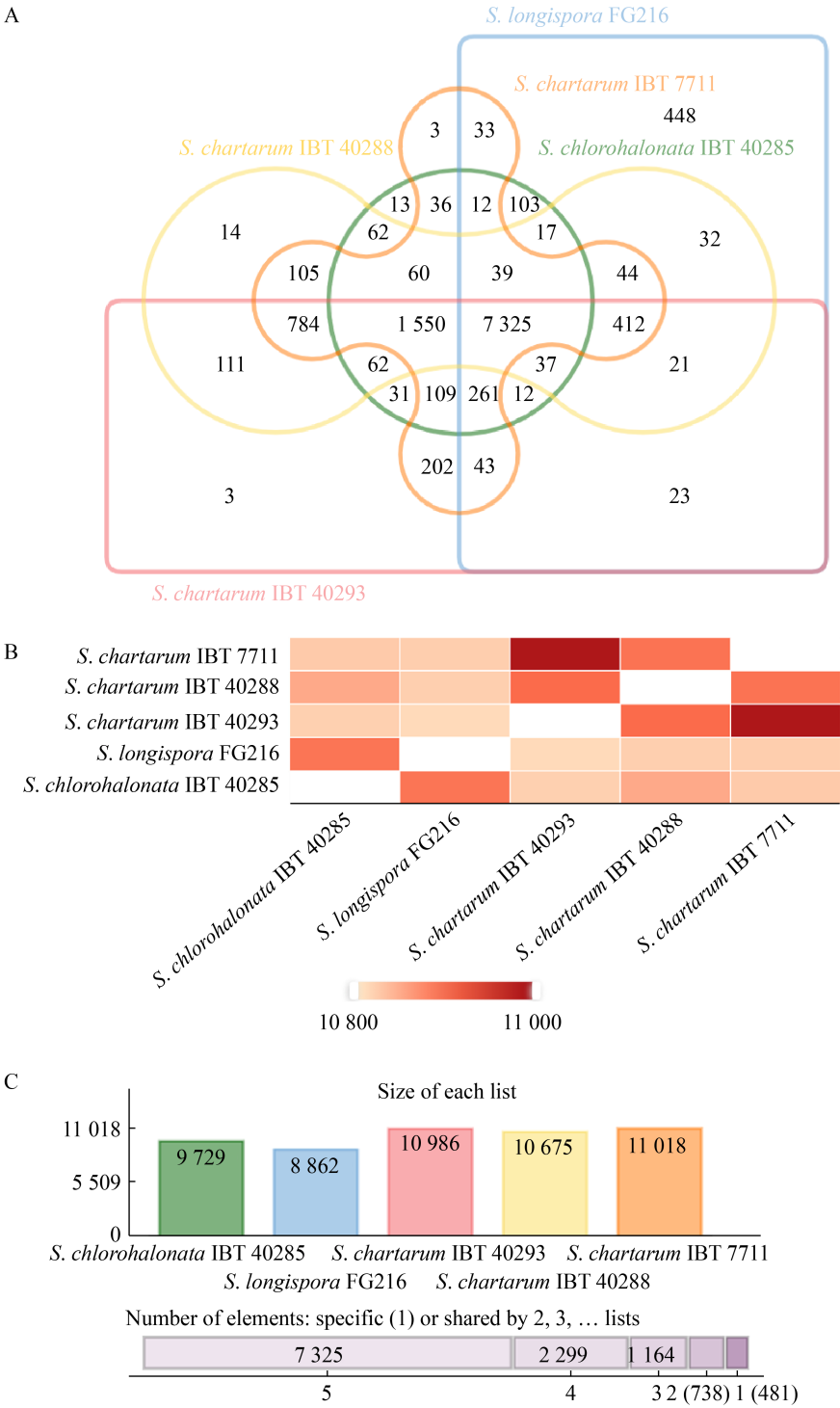


图 7 *S. longispora* FG216 基因组信息的差异比较

Figure 7 Differential comparison of genomic information of *S. longispora* FG216

注：A：5 种 *Stachybotrys* 核心基因、独有基因、兼有基因韦恩图分析；B：成对基因家族间重叠簇的热图分析；C：5 种 *Stachybotrys* 基因家族数目分析

Note: A: Analysis of core genes, unique genes and concurrent genes of five *Stachybotrys*; B: Heat map analysis of overlapping gene clusters between pairs of gene families; C: Number analysis of five *Stachybotrys* gene families

表 2 *Stachybotrys* 核心基因 GO 分析
Table 2 GO analysis of *Stachybotrys* core gene

GO ID	Count	Name	Namespace
GO:0016491	84	Oxidoreductase activity	Molecular function
GO:0016705	32	Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	Molecular function
GO:0015031	97	Protein transport	Biological process
GO:0016746	5	Transferase activity, transferring acyl groups	Molecular function
GO:0006886	29	Intracellular protein transport	Biological process
GO:0045493	24	Xylan catabolic process	Biological process
GO:0000398	28	Mrna splicing, via spliceosome	Biological process
GO:0045122	4	Aflatoxin biosynthetic process	Biological process
GO:0006351	87	Transcription, DNA-templated	Biological process
GO:0016125	6	Sterol metabolic process	Biological process
GO:0006364	74	rRNA processing	Biological process
GO:0004497	16	Monooxygenase activity	Molecular function

表 3 *S. longispora* FG216 独有基因 GO 分析
Table 3 *S. longispora* FG216 unique gene GO analysis

GO ID	Count	Name	Namespace
GO:0008643	9	Carbohydrate transport	Biological process
GO:0043864	4	Indoleacetamide hydrolase activity	Molecular function
GO:0035442	4	Dipeptide transmembrane transport	Biological process
GO:0043312	3	Neutrophil degranulation	Biological process
GO:0019477	4	L-lysine catabolic process	Biological process
GO:0009058	4	Biosynthetic process	Biological process
GO:0006807	3	Nitrogen compound metabolic process	Biological process
GO:0006847	2	Plasma membrane acetate transport	Biological process
GO:0000272	9	Polysaccharide catabolic process	Biological process

根据次生代谢产物合成基因簇预测结果显示,发现一条基因簇 Cluster 57 涉及异吲哚啉酮(Isoindolinone)的生物合成(图 8),与 *S. chlorohalonata* IBT 40285 中对应合成基因簇的相似度有 40%,但是其 2 条核心生物合成基因 NRPS-Like、PKS 在 *S. longispora* FG216 中是存在的,而异吲哚则是 FGFC1 的核心结构,是生物合成中的关键部分(图 9)。如表 4 所示, *S. longispora* FG216 和 *S. chlorohalonata* IBT 40285 中相对应基因的 BLAST 比较分析发现,共有的 4 个基因的覆盖度都超过 90%、相似度超过 60%,可信度较高,9 个相邻的基因组成长度为 49 082 bp 的完整基因簇,

推测为纤溶活性化合物 FGFC1 的生物合成基因簇。根据基因簇和 FGFC1 的化学结构式推测出在 *S. longispora* FG216 中的生物合成途径,红色方框内的生物合成步骤是确定的,合成的冬青生菌素 B (Illicolin B)是 FGFC1 的前体化合物,而接下来的合成过程需要继续探索(图 10)。

预测得到的结果表明, *S. longispora* FG216 具有合成多种次生代谢产物的潜力和基因元件,具体能否合成这些次生代谢产物还与其培养条件等因素有关,由于 *S. longispora* FG216 的次生代谢产物相关文献报道较少,未能就预测的其他次生代谢产物与其分离得到的已知次生代谢产物对应上。

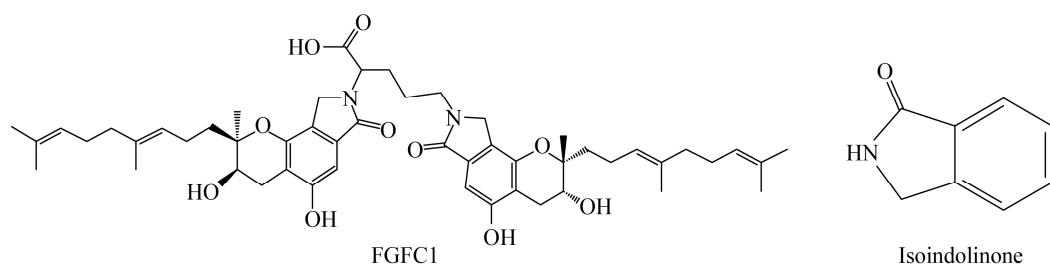


图 8 FGFC1 和异吲哚啉酮的化学结构

Figure 8 Chemical structure of FGFC1 and isindolinone

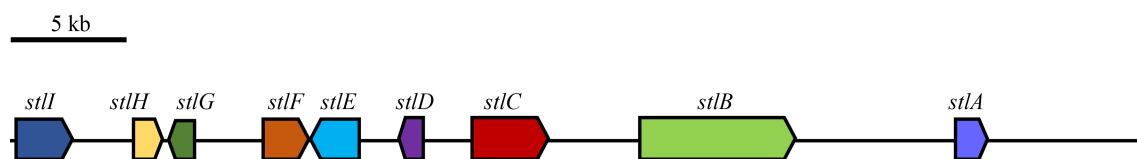
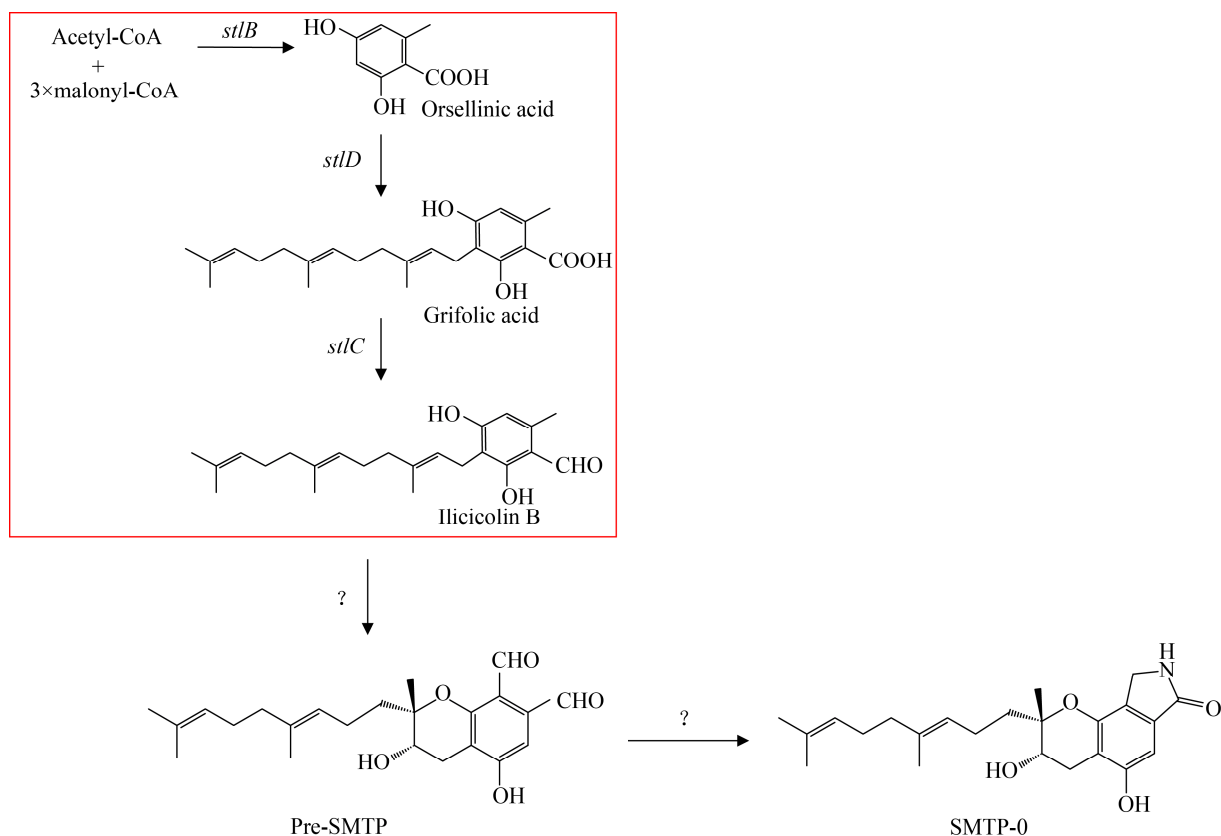
图 9 *S. longispora* FG216 基因组中异吲哚啉酮的生物合成基因簇Figure 9 Isoindolinone biosynthetic gene cluster in *S. longispora* FG216 genome

图 10 推测的 FGFC1 生物合成途径

Figure 10 The presumed FGFC1 biosynthetic pathway

表 4 *S. longispora* FG216 与 *S. chlorohalonata* IBT 40285 异吲哚啉酮生物合成基因簇的比较
Table 4 Comparison of isoindolinone biosynthetic gene clusters between *S. longispora* FG216 and *S. chlorohalonata* IBT 40285

Gene	Locus tag	<i>S. chlorohalonata</i> IBT 40285 homologue	Coverage (%)	Protein identity (%)	Putative function
<i>stII</i>	g7747	—	—	—	Protein RTA1 [<i>Trichoderma asperellum</i>]
<i>stIH</i>	g7748	—	—	—	Long-chain specific acyl-CoA dehydrogenase-like protein [<i>Acremonium chrysogenum</i> ATCC 11550]
<i>stIG</i>	g7749	S40285_07609	100	62.22	Glycoside hydrolase [<i>Massariosphaeria phaeospora</i>]
<i>stIF</i>	g7750	—	—	—	Rhamnogalacturonate lyase B [<i>Aspergillus bombycis</i>]
<i>stIE</i>	g7751	—	—	—	Siderophore iron transporter mirB [<i>Verticillium dahliae</i> VdLs.17]
<i>stID</i>	g7752	S40285_10521	94	75.86	Prenyltransferase [<i>Stachybotrys bisbyi</i>]
<i>stIC</i>	g7753	S40285_07611	99	73.77	NRPS-like enzyme [<i>Stachybotrys bisbyi</i>]
<i>stIB</i>	g7754	S40285_07612	99	76.09	Polyketide esynthase [<i>Stachybotrys bisbyi</i>]
<i>stIA</i>	g7755	—	—	—	Polyketide synthase [<i>Aspergillus tubingensis</i>]

注：—：不存在相似基因
Note: —: There is no similar gene

3 讨论与结论

海洋微生物因为其独特的高压、高盐、低氧、低光照等异于陆地的生存环境，使其在进化过程中可以产出具有结构独特、生理活性等特征的一系列次生代谢产物，成为研究治疗各种疾病的热点，对人类的健康起到关键性作用^[8]。海洋真菌是天然产物的重要生产者，根据 2010–2013 年的海洋微生物中天然产物的调查显示，63%的天然产物来自于海洋真菌^[9]。然而在 1981–2014 年间，被批准的 1 211 种小分子治疗药物中有将近 65%是与天然产物产生直接或间接的联系^[10]。从海洋真菌中分离得到的化合物结构类型主要包括聚酮类、生物碱类、肽类、萜类、甾体类、内酯类等，许多化合物结构新颖，而且具有多种药理学活性比如抗癌、抗炎、抑菌、抗糖尿病等，有非常重要的研究价值^[11]。

Stachybotrys 作为一种产毒素菌株引起社会的广泛关注，其容易引起呼吸系统疾病，对人体健康产生负面影响。随着研究的不断深入，*Stachybotrys* 的相关习性逐渐被了解，根据其生物特性先后从土壤^[12]、腐烂纤维布^[13]、海绵^[14]、红树林^[15]、珊瑚^[16]等分离出多种葡萄穗霉属真菌，然而大多

数以黑葡萄穗霉菌为主。最近几年，常见研究报道 *Stachybotrys* 分离出有着特殊结构的天然产物具有相关的生物活性，其中 *S. longispora* FG216 经过探索，发现可以稳定地产出具有纤溶活性的小分子化合物，对于治疗血栓性疾病有着重要的意义^[17]。随着天然产物的研究增多，新型化合物的发现进入瓶颈期，分析其原因在于海洋微生物难以分离培养、化合物只在特定自然环境下产出、发现的化合物重复率较高、微生物产物单一等，极大地限制了新型天然产物的发展^[18]。近年来，基因测序技术的飞速发展以及分析生物信息学和遗传学等相关技术的成熟，为天然产物的探索提供了新的研究方法。通过对可培养的海洋微生物进行基因组挖掘，从中发现、筛选并鉴定出潜在的新型天然产物相关的沉默基因簇，极大地推动了单一微生物的生产能力，促进其代谢产物多样化，减少生物资源浪费。

在本次研究中，首次利用 *De Novo* 技术对从海水中分离得到的 *S. longispora* FG216 菌株进行基因组测序，总共得到长度为 45.6 Mb 的基因组序列，13 329 个编码基因占基因组组装总碱基数目的 46.58%，相较于 *Stachybotrys* 其他已知的 5 个物

种的基因组数据而言, 基因组测序结果较好, 检测到的基因数目最多。KEGG、GO、KOG 数据库注释结果显示编码基因主要富集在包括氨基酸代谢、碳水化合物代谢及异源物质降解和代谢等代谢功能区, 行使催化活性、结合和转运等分子功能, 有利于适应环境。除了维持和促进真菌的生长, 虽然 *S. longispora* FG216 中对应的萜类化合物和聚酮类化合物代谢相关的编码基因相对较少, 但是验证了其有产出相应新型化合物的潜力。从代谢途径来看, *S. longispora* FG216 仍然有许多次生代谢产物有待探索, 有重要的研究价值。通过基因组学分析, 验证 *Stachybotrys* 作为致病菌, 从 KEGG 注释中可以看出其具有多种致病性, 可能引起各种感染性、癌症、神经退行性疾病等重大人体疾病, 对公共安全有很大的影响。在 *Stachybotrys* 研究的过程中, 研究人员应注意防护, 佩戴防护措施, 谨慎操作。

S. longispora FG216 基因组与 *S. chlorohalonata* IBT 40285、*S. chartarum* IBT 40288、*S. chartarum* IBT 40293、*S. chartarum* IBT 7711 作比较, 通过物种同源性分析表明, *Stachybotrys* 是高度保守的物种, 具有较高比例的核心基因, 可预测生物功能相似, 根据核心基因的生物功能预测有产出黄曲霉毒素的能力, 而且与 *S. chlorohalonata* IBT 40285 有最高的生物相似性, 推测可产出具有相似结构的天然产物。*S. longispora* FG216 有关于吲哚乙酰胺水解酶活性的独有基因, 可能参与其代谢产物 FGFC1 的合成。antiSMASH 软件共预测出 101 个基因簇和 18 个化合物合成基因簇。基因簇的类型主要有 NRPS、T1PKS、T3PKS 和 NRPS-Like, 还包括其他的萜烯、吲哚和磷酸盐的合成基因簇。然而从 *S. longispora* FG216 提取到的纤溶活性化合物 FGFC1 在化学结构上属于呋喃并异吲哚酮类衍生物, 也是杂倍半萜烯化合物的一种, 从预测出的基因簇中匹配到一种相关的化合物异吲哚啉酮的合成基因簇 Cluster 57, 主要由 NRPS-Like 和 T1PKS 类型的基因簇构成, 整体

相似度虽然只有 40%, 但是 2 条核心基因的存在极大地增加了 FGFC1 的合成由 Cluster 57 基因簇完成的可能性, 进而对 FGFC1 的生物合成途径进行了预测, 对于接下来从生物信息学和分子水平上去解析 FGFC1 的生物合成途径打下了坚实的基础。

本研究首次报道了 *S. longispora* FG216 的基因组信息, 为深入了解葡萄穗霉菌的次生代谢合成途径提供了参考信息, 为解决葡萄穗霉属间的进化、分类及鉴定等研究提供了参考数据, 对以后 *S. longispora* FG216 的相关研究具有重要意义。

致谢: 感谢上海海洋大学吴文惠教授对本研究的指导。

REFERENCES

- [1] Lombard L, Houbaken J, Decock C, Samson RA, Meijer M, Réblová M, Groenewald JZ, Crous PW. Generic hyper-diversity in *Stachybotriaceae*[J]. *Persoonia*, 2016, 36: 156-246
- [2] Li Y, Liu D, Cen S, Proksch P, Lin WH. Isoindolinone-type alkaloids from the sponge-derived fungus *Stachybotrys chartarum*[J]. *Tetrahedron*, 2014, 70(39): 7010-7015
- [3] Li YM, Tong SC, Li TS. Adaptive fuzzy output-feedback control for output constrained nonlinear systems in the presence of input saturation[J]. *Fuzzy Sets and Systems*, 2014, 248: 138-155
- [4] Wu B, Oesker V, Wiese J, Malien S, Schmaljohann R, Imhoff JF. Spirocyclic drimanes from the marine fungus *Stachybotrys* sp. strain MF347[J]. *Marine Drugs*, 2014, 12(4): 1924-1938
- [5] Semeiks J, Borek D, Otwinowski Z, Grishin NV. Comparative genome sequencing reveals chemotype-specific gene clusters in the toxigenic black mold *Stachybotrys*[J]. *BMC Genomics*, 2014, 15: 590
- [6] Fu SQ, Yan T, Wu WH, Zhu QG, Guo RH, Chen SQ, Bao B. Fibrinolysis promoting activity of a novel metabolic product of marine microbe-FGFC1[J]. *Pharmaceutical Care and Research*, 2015, 15(2): 99-102 (in Chinese)
傅诗情, 严婷, 吴文惠, 朱全刚, 郭锐华, 陈山乔, 包斌. 海洋微生物代谢产物 FGFC1 的纤溶促进作用[J]. *药学服务与研究*, 2015, 15(2): 99-102
- [7] Qian SY, Yang J, Zhang BS, Jiang SN, Wu WH, Guo RH, Bao CL, Bao B. Research of biosynthesis of fibrinolytic compounds from marine *Stachybotrys longispora*[J]. *Chinese Journal of Marine Drugs*, 2019, 38(3): 52-58 (in Chinese)
钱士云, 杨晶, 张本帅, 姜胜男, 吴文惠, 郭锐华, 包春

- 玲, 包斌. 海洋长孢葡萄穗霉菌生物合成纤溶活性化合物的研究[J]. 中国海洋药物, 2019, 38(3): 52-58
- [8] Zhang HT. Current situation of marine drugs research[J]. Journal of Guangdong Medical University, 2020, 38(3): 251-254 (in Chinese)
张海涛. 海洋药物的研究现状[J]. 广东医科大学学报, 2020, 38(3): 251-254
- [9] Zhao CY, Zhu TH, Zhu WM. New marine natural products of microbial origin from 2010 to 2013[J]. Chinese Journal of Organic Chemistry, 2013, 33(6): 1195-1234 (in Chinese)
赵成英, 朱统汉, 朱伟明. 2010-2013 之海洋微生物新天然产物[J]. 有机化学, 2013, 33(6): 1195-1234
- [10] Newman DJ, Cragg GM. Natural products as sources of new drugs from 1981 to 2014[J]. Journal of Natural Products, 2016, 79(3): 629-661
- [11] Rateb ME, Ebel R. Secondary metabolites of fungi from marine habitats[J]. Natural Product Reports, 2011, 28(2): 290-344
- [12] Zhai NP, Li GY, Xu C, Li TT, Geng YH, Zhang M. Identification of three species of *Stachybotryaceae* in the soil of mountain slope in Henan[J]. Chinese Journal of Soil Science, 2019, 50(4): 878-884 (in Chinese)
翟妮平, 李光宇, 徐超, 李田田, 耿月华, 张猛. 河南山坡土壤中三种葡萄穗霉科真菌鉴定[J]. 土壤通报, 2019, 50(4): 878-884
- [13] Picart P, Pastor FIJ, Orejas M. Transcriptional analysis of the lichenase-like gene *cel12A* of the filamentous fungus *Stachybotrys atra* BP-A and its relevance for lignocellulose depolymerization[J]. International Microbiology, 2021, 24(2): 197-205
- [14] Liu D, Li Y, Guo XC, Ji W, Lin WH. Chartarlactams Q-T, dimeric phenylspirodrimanones with antibacterial and antiviral activities[J]. Chemistry & Biodiversity, 2020, 17(6): e2000170
- [15] Gan Q, Lin CY, Lu CJ, Chang YM, Che Q, Zhang GJ, Zhu TJ, Gu QQ, Wu ZQ, Li MY, et al. Staprexanthones, xanthone-type stimulators of pancreatic β -cell proliferation from a mangrove endophytic fungus[J]. Journal of Natural Products, 2020, 83(10): 2996-3003
- [16] Yang B, Long JY, Pang XY, Lin XP, Liao SR, Wang JF, Zhou XF, Li YQ, Liu YH. Structurally diverse polyketides and phenylspirodrimanones from the soft coral-associated fungus *Stachybotrys chartarum* SCSIO41201[J]. The Journal of Antibiotics, 2021, 74(3): 190-198
- [17] Hong ST, Wu WH, Zhou Y, Yan T, Zhou Y, Bao B. Effect of indolone compound from marine microbial sources on conformation characteristics of fibrinolytic activity factors[J]. Chinese Journal of Marine Drugs, 2015, 34(1): 59-66 (in Chinese)
洪少彤, 吴文惠, 周喻, 严婷, 周寅, 包斌. 海洋微生物来源吲哚酮纤溶化合物影响纤溶因子构象特性的研究[J]. 中国海洋药物, 2015, 34(1): 59-66
- [18] Molinski TF, Dalisay DS, Lievens SL, Saludes JP. Drug development from marine natural products[J]. Nature Reviews Drug Discovery, 2009, 8(1): 69-85