

黑茶益生菌种“金花”菌基因组测序及分析

袁超¹, 袁丹丹², 王淼², 刘飞¹, 朱希强^{1, 2*}, 凌沛学¹

(1. 山东省药学院, 山东省生物药物重点实验室, 山东 济南 250101; 2. 湖南惟楚福瑞达生物科技有限公司, 湖南 益阳 413000)

摘要:目的 提取“金花”菌基因组并测序分析。方法 从黑茶中分离得到一株冠突散囊菌, 命名为YKY807, 提取其基因组, 利用三代测序技术和Pacbio RS 测序平台进行测序和序列分析。结果 序列全长为27 701 366 kb, 其GC含量为49.87%。总基因数量为8405, 总基因序列长度为12 750 105 kb, 编码基因占基因组的41.5%, 编码基因总长度为11 398 657 kb, 编码基因平均长度为1356.1 kb。结论 已获得冠突散囊菌基因组序列信息, 序列已提交至GenBank数据库, 登录号为MAQV00000000。

关键词: 冠突散囊菌; 基因组测序; 序列分析

中图分类号: Q78 文献标识码: A 文章编号: 1672-979X (2016) 04-0259-04

Genome Sequencing and Sequence Analysis of Dark Tea Probiotics *Eurotium cristatum*

YUAN Chao¹, YUAN Dan-dan², WANG Miao², LIU Fei¹, ZHU Xi-qiang^{1,2}, LING Pei-xue¹

(1. Shandong Provincial Key Laboratory of Biopharmaceuticals, Shandong Academy of Pharmaceutical Sciences, Jinan 250101, China; 2. Hunan Weichu Freda Bio-Technology Co., Ltd., Yiyang 413000, China)

Abstract: Objective To obtain the whole genome sequence data of *Eurotium cristatum*. **Methods** A strain of *Eurotium cristatum* was isolated from dark tea, named *Eurotium cristatum* YKY807. The genome was extracted, then the genome sequencing and sequence analysis was performed using the third generation sequencing technology and the Pacbio RS platform. **Results** The total genome sequence length of strain *Eurotium cristatum* YKY807 is 27 701 366 kb, with an average GC content of 49.87%. The total gene number is 8405, the total gene length is 12 750 105 kb, and the coding sequence (CDS) accounts for 41.15% of the genome; the total CDS length is 11 398 657 kb, and the average CDS length is 1356.1 kb. **Conclusion** The whole genome sequence data of *Eurotium cristatum* has been obtained. The complete genome sequence of YKY807 has been submitted to GenBank (the accession number is MAQV00000000).

Key Words: *Eurotium cristatum*; genome sequencing; sequence analysis

“金花”菌, 学名冠突散囊菌 (*Eurotium cristatum*), 属散囊菌目发菌科散囊菌属, 是湖南安化黑茶所独有的益生菌种^[1]。大量的研究和临床数据表明, 冠突散囊菌能有效提升人体免疫

收稿日期: 2016-05-25

作者简介: 袁超 (1985-), 男, 山东潍坊人, 硕士, 工程师, 研究方向: 生物技术 Tel: 0531-81213080

E-mail: yuanchao_l@sina.com

*通讯作者: 朱希强 (1967-), 男, 山东济南人, 博士, 研究员, 研究方向: 生物技术 Tel: 0531-81213080

E-mail: xistrong@sina.com

能力,具有抗癌、抗氧化、抗衰老等作用^[2]。此外,消脂减肥功效远超其他菌种^[3]。

为更好地研究冠突散囊菌的作用机制,我们开展了冠突散囊菌基因组的研究。从黑茶中分离得到目的菌株,经鉴定为冠突散囊菌,命名为YKY807 (*Eurotium cristatum* YKY807),经纯种培养后,提取基因组并测序分析,为寻找其功能基因奠定基础。

1 仪器与材料

1.1 仪器

RS232G紫外分光光度计 (Eppendorf);
DYY-6C电泳仪 (北京六一)。

1.2 材料

查氏培养基^[4] (自制); E5038植物/真菌DNA分离试剂盒 (Sigma)。

1.3 菌种

冠突散囊菌YKY807,分离自湖南惟楚福瑞达生物科技有限公司所产黑茶。

2 方法

2.1 基因组提取

2.1.1 将冠突散囊菌孢子液 (10^8 个/mL)按2%接种量接种到查氏培养基中,28℃摇床培养20 h。收集菌丝体,洗涤。

2.1.2 采用Sigma E5038植物/真菌DNA分离试剂盒提取基因组,提取步骤按试剂盒说明书进行。

2.1.3 用琼脂糖凝胶电泳检测DNA样品纯度^[5]。

2.1.4 用紫外分光光度计检测DNA样品浓度。

2.2 基因组测序及拼接

采用全基因组鸟枪法 (whole genome shotgun, WGS)策略,利用Illumina Miseq+Pacbio RS 平台进行测序^[6]。采用FastQC对数据进行质量控制,使用A5.2.0软件完成拼接。

2.3 基因组完整性与连续性评估

真核生物中存在一类蛋白质,该类蛋白质的氨基酸序列在大部分真核生物中比较保守,将基因组拼装后的序列与这些蛋白质进行比较,检测

这类蛋白质序列的完整性,从而达到间接评估基因组拼装完整性与连续性的目的^[7]。本实验采用CEGMA (core eukaryotic genes mapping approach, version 2.5)来评估基因组拼装的完整性和连续性。

2.4 蛋白质编码基因预测

2.4.1 采用CEGMA获取真核生物中保守的基因序列,以此基因集作为Augustus (version 3.03)基因模型训练集,构建冠突散囊菌YKY807基因组的基因预测模型。

2.4.2 对Augustus第一轮预测的基因预测结果与Swiss-Prot数据库进行比对,选取保守序列作为Augustus基因预测的训练集,以构建的基因预测模型获得基因预测结果;同时以此序列集作为训练集,分别构建glimmerHMM (version 3.0.1)和SNAP (version 2006-07-28)的基因预测模型,并获得相应的基因预测结果。

2.4.3 采用Exonerate (version 2.2.0)软件,利用近缘物种的蛋白质序列进行同源预测。

2.4.4 采用EVIDENCEModeler (version r2012-06-25)软件对上述3种预测结果进行整合^[8]。

3 结果与分析

3.1 基因组提取

3.1.1 用琼脂糖凝胶电泳对提取的DNA纯度进行分析,琼脂糖凝胶电泳结果见图1。

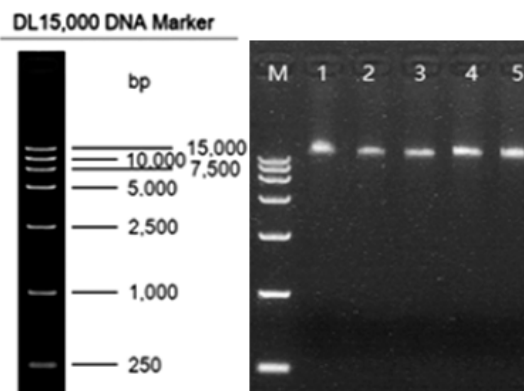


图1 冠突散囊菌YKY807琼脂糖凝胶电泳图

由图1可见,提取的DNA电泳检测条带单一,无降解,无蛋白质污染,DNA纯度符合建库要求。

3.1.2 利用紫外分光光度计测定提取的DNA浓度,结果见表1。

表1 冠突散囊菌YKY807紫外分光光度计定量结果

泳道	样品名称	浓度/ng·μL ⁻¹	体积/μL	总量/μg
1	YKY807	109.24	50	5.46
2	YKY807	64.10	50	3.21
3	YKY807	120.05	50	6.00
4	YKY807	92.47	50	4.62
5	YKY807	81.89	50	4.09

由表1可见,提取的DNA样品浓度最小为64.10 ng/μL,最大为120.05 ng/μL,均大于20 ng/μL,DNA浓度符合建库要求。

综上,用Sigma E5038植物/真菌DNA分离试剂盒提取的冠突散囊菌YKY807基因组DNA符合建库要求。

3.2 基因组测序及拼接

将基因组测序数据利用A5 2.0软件进行拼接,拼接结果见表2。

表2 基因组拼装数据统计

项目	数据
总序列数量	11
总序列长度/bp	27 701 366
GC百分含量/%	49.87
不确定碱基数量	0
最短序列长度/bp	582 486
最长序列长度/bp	3 958 284
N50/bp	2 970 423

注: N50为将所有序列长度从长到短排列将序列长度按照该顺序依次相加,当相加的长度达序列总长度的50%时,最后一条序列的长度。

由表2可见,使用A5 2.0软件拼接后,共得到11个序列,其中最短序列582 486 kb,最长序列3 958 284 kb;序列全长为27 701 366 kb,其GC含量为49.87%。

3.3 基因组完整性与连续性评估

基因组拼装的完整性和连续性评估结果见表3。

表3 基因组拼装完整性和连续性评估

项目	序列个数	基因组中检测到的保守蛋白序列的数量占总保守蛋白数量的百分比/%
YKY807基因组保守蛋白序列	237	95.56
YKY807基因组完整保守蛋白序列	230	92.74
CEGMA定义保守蛋白序列	248	-

由表3可见,在基因组中,CEGMA检测到了其中237个保守蛋白序列,占CEGMA定义总保守蛋白质序列数量的95.56%;在这237个保守蛋白序列中,230个蛋白质序列为完整序列,占CEGMA定义总保守蛋白序列数量的92.74%,占YKY807基因组总保守蛋白质序列数量的97.05%。CEGMA的检测结果显示基因组拼接结果有较好的基因组连续性和完整性。

3.4 蛋白质编码基因预测

采用EVIDENCEModeler(version r2012-06-25)软件对预测结果进行整合,基因预测统计结果见表4。

表4 基因预测结果统计

项目	预测结果
总基因数量	8405
总基因序列长度	12 750 105
总基因长度占基因组百分比	46.03
平均基因长度	1516.9
编码基因总长度	11 398 657
编码基因占基因组的比例	41.15
编码基因的平均长度	1356.1

由表4可见,YKY807总基因数量为8405,总基因序列长度为12 750 105 kb,编码基因占基因组的41.15%,编码基因总长度为11 398 657 kb,编码基因平均长度为1356.1 kb。

4 结论

我们通过菌种分离纯化获得了冠突散囊菌纯种并提取其基因组。利用国内外最先进的三代测序技术和Pacbio RS 测序平台获得了整个基因组信息。序列已提交至GenBank数据库, 登录号为MAQV01000000。本工作对寻找冠突散囊菌保健功能基因、阐释黑茶的保健功效及提升黑茶的质量、促进该产业的发展有重要意义。

参考文献

- [1] 叶聿程. 茯砖茶中冠突散囊菌的研究进展[J]. 生物技术世界, 2014, (9): 64.
- [2] 刘石泉, 赵运林, 胡治远, 等. 冠突散囊菌基因组DNA提取方法比较研究[J]. 茶叶, 2011, 37(4): 225-230.
- [3] 刘子音, 许爱清, 李宗军, 等. 茯砖茶中冠突散囊菌及其代谢产物研究进展[J]. 茶叶通讯, 2010, 37(1): 23-26.
- [4] 姚茂君, 黄群, 陈林杰, 等. 冠突散囊菌的分离及液态发酵特性[J]. 食品与发酵工业, 2007, 33(6): 28-31.
- [5] 李尔汉, 杨慧林, 王筱兰, 等. 肌苷生产菌枯草芽孢杆菌 ATCC 13952 的全基因组测序及序列分析[J]. 微生物学报, 2015, 55(12): 1560-1567.
- [6] Geng W T, Cao M F, Song C J, *et al.* Complete genome sequence of *Bacillus amyloliquefaciens* LL3, which exhibits glutamic acid-independent production of poly- γ - glutamic acid[J]. *J Bacteriol*, 2011, 193 (13): 3393-3394.
- [7] 梁振山, 曾令兵, 万腊根. 基因组测序技术及其临床应用进展[J]. 实验与检验医学, 2016, 34(1): 48-51.
- [8] Boeckmann B, Bairoch A, Apweiler R, *et al.* The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003[J]. *Nucleic Acids Res*, 2003, 31(1): 365-370.

《食品与药品》文稿中有关插图的要求

1 图的分类

线条图 常用的有: 简易函数图、流程图、色谱图等

照片图 本刊采用黑白照片图

2 基本要求

同一内容不应既用文字或表又用图, 类同事物或现象无须重复插图。

插图中所选用的名词术语应与正文中一致。不应出现正文中没有或与正文表述内容不相关的文字、数字和符号。

同一文章中各插图的大小、字号应统一。插图应排列在文中第一次提到它的文字段后面。对于图的说明和分析列于图后。

图题应简洁明确。图题与图序之间空1个字距, 图题用黑体小四。分图可用ABC或abc表示。注释和说明语应置于图题之上, 较图题小一号字。

标目通常由量名称和相应的单位组成, 一般用比的形式, 如: t/d , t/min , t/h , 酶活性/ $U \cdot mg^{-1}$, A_{260nm} , 抑制率/%等。