

酱香大曲中链霉菌菌株FBKL4.005全基因组测序及序列分析

李豆南^{1,2}, 王晓丹¹, 罗小叶¹, 黄 魏¹, 邱树毅^{1,2,*}

(1. 贵州大学 贵州省发酵工程与生物制药重点实验室, 酿酒与食品工程学院, 贵州 贵阳 550025;

2. 贵州大学生命科学学院, 贵州 贵阳 550025)

摘 要: 链霉菌FBKL4.005是本课题组前期从酱香大曲中分离得到的1株耐高温菌株, 能够在37~55℃的范围内生长代谢, 为了能够更深入地探究该菌株在酿酒过程中的代谢机理和功能性作用, 完成大曲极端微生物开发应用, 对于菌株的序列信息进行解析是最为全面高效的方式。本研究采用Illumina HiSeq 2000测序平台对菌株FBKL4.005进行全基因组测序, 共得到67 771 429个reads, 2.47 G的数据量, 通过对测序数据进行拼接, 得到69个contigs, 菌株整个基因组大小为9 454 406 bp, GC含量为73.03%, 经过滤处理后得到7.33 Mb的有效数据量及7 946个注释基因, 并通过GO、COG、KEGG、NR、TCDB数据库的比对分析, 完成菌株基因组预测与基本功能的注释, 获得的基因注释信息为今后深入开展酱香大曲中耐高温特性链霉菌类群代谢途径及机理的研究提供了理论支持。

关键词: 酱香型大曲; 链霉菌属; 全基因组测序; 基因功能注释

Whole Genome Sequencing and Sequence Analysis of *Streptomyces* sp. FBKL4.005 from Moutai-Flavor *Daqu*

LI Dounan^{1,2}, WANG Xiaodan¹, LUO Xiaoye¹, HUANG Wei¹, QIU Shuyi^{1,2,*}

(1. Guizhou Provincial Key Laboratory of Fermentation Engineering and Biological Pharmacy,

School of Liquor and Food Engineering, Guizhou University, Guiyang 550025, China;

2. College of Life Sciences, Guizhou University, Guiyang 550025, China)

Abstract: *Streptomyces* sp. FBKL4.005 was a heat-resistant strain isolated from Moutai-flavor *Daqu*, which could grow and metabolize at 37–55 °C. In order to get insights into the metabolic mechanism and functional role of the strain in liquor-making for the exploitation and application of extremophiles in *Daqu*, whole genome sequencing of strain FBKL4.005 was performed on the Illumina HiSeq 2000 sequencing platform. A total of 67 771 429 reads and 2.47 G data were obtained. By *de novo* sequence splicing, 69 contigs were also acquired. The complete genomic size of stain was 9 454 406 bp with a GC content of 73.03%. After filtration treatment, 7.33 Mb effective data and 7 946 annotated genes were acquired. We conducted a comparative analysis among the GO, COG, KEGG, NR and TCDB database for genomic prediction and functional annotation. The acquirement of gene annotation information can provide an important foundation for further research on the metabolic pathways and mechanism of heat resistant *Streptomyces* sp. in Moutai-flavor *Daqu*.

Keywords: Moutai-flavored *Daqu*; *Streptomyces* sp.; whole genome sequencing; gene functional annotation

DOI:10.7506/spkx1002-6630-201814031

中图分类号: Q939.13

文献标志码: A

文章编号: 1002-6630 (2018) 14-0206-07

引文格式:

李豆南, 王晓丹, 罗小叶, 等. 酱香大曲中链霉菌菌株FBKL4.005全基因组测序及序列分析[J]. 食品科学, 2018, 39(14): 206-212. DOI:10.7506/spkx1002-6630-201814031. <http://www.spkx.net.cn>

LI Dounan, WANG Xiaodan, LUO Xiaoye, et al. Whole genome sequencing and sequence analysis of *Streptomyces* sp. FBKL4.005 from moutai-flavor *Daqu*[J]. Food Science, 2018, 39(14): 206-212. (in Chinese with English abstract) DOI:10.7506/spkx1002-6630-201814031. <http://www.spkx.net.cn>

收稿日期: 2017-05-30

基金项目: 贵州省工业攻关项目(黔科合GZ字[2014]3012); 贵州省重大专项(黔科合重大专项字GZ[2013]6009号)

第一作者简介: 李豆南(1991—), 男, 博士研究生, 研究方向为发酵工程。E-mail: 530034619@qq.com

*通信作者简介: 邱树毅(1963—), 男, 教授, 博士, 研究方向为酶工程、发酵工程、食品生物技术、白酒酿造。

E-mail: syqiu@gzu.edu.cn

链霉菌 (*Streptomyces*) 从分类学上归属于放线菌目, 是最为高等的放线菌类群, 具有发育良好的分枝菌丝, 主要分布于土壤环境中。研究表明, 该类群主要是一类革兰氏阳性、化能异养型好氧菌^[1], 该菌属下的菌株多数能够合成结构复杂的次生代谢产物 (如抗生素等), 具有良好的生物活性, 是一类具有很大应用潜力的微生物^[2]。近年来, 在酱香型、浓香型白酒酿造过程中也陆续分离得到该菌属下的多种菌株, 并对其代谢特性进行了一定的探究, 确认该菌属在酿造体系中存在重要的生物调控作用^[3-4]。

全基因组测序是针对未知基因组序列的物种进行个体全部基因组测序, 具有测序覆盖面广、准确性极高的技术优势^[5]。近年来随着该技术的快速发展, 以第2代测序技术为主体的全基因组测序技术已成为了一种高效检测手段, 广泛应用于生物、医学等领域。在微生物学方面, 以低错误率、低成本化和测序通量高为特征的罗氏454、Illumina HiSeq 2000测序平台已演变为基本的二代全基因组测序技术, 成为了探究未知微生物类群多样性、生物学特性、代谢功能机制的重要手段^[6-7], 已经在土壤、肠道环境中微生物类群的分析鉴定以及工程菌株的开发中发挥了重要的作用^[8-10]。针对一些极端环境中的微生物而言, 由于其难以分离、可培养性低的特点, 传统的分析方法已不适用, 这时全基因组测序技术的优势便展现出来^[11], 其中, 在诸如油田中嗜热采油芽孢杆菌 (*Geobacillus thermodenitrificans*)^[12]、新西兰地热区域中嗜酸甲烷氧化细菌^[13-14]、水稻根际土壤中的斯氏假单胞菌 (*Pseudomonas stutzeri*)^[15]、蜡状芽孢杆菌^[16]、深海沉积物中耐压耐冷的希瓦氏菌属^[17]、两极地区永久冻土中的产甲烷菌类群^[18]、氯消毒饮用水中的抗性变形菌类群^[19]以及高酸性金属矿中的硫杆菌、酸微菌群落^[20]等极端环境下微生物的研究中均有该技术的运用, 为微生物防治石油污染、特殊代谢途径的发现和极端环境微生物多样性等方面的研究起到了巨大的推动作用。

本研究的开展建立在课题组王晓丹等^[21]之前的研究之上, 其利用罗氏454 FLX+高通量测序平台对贵州3种酱香型高温大曲中细菌的群落结构多样性进行了深入研究, 发现在属水平上链霉菌属为大曲中的优势细菌属之一, 含量均占到大曲总生物量的1%以上, 暗示该类群可能具有重要的功能性作用; 基于这一结果, 本课题组通过分离方法的设计, 成功从大曲中得到1株具有耐高温特性的链霉菌株FBKL4.005, 并以该菌株为材料进行全基因组测序分析, 从基因功能注释的角度快速剖析菌株的代谢特征, 为日后酱香大曲中该类群功能性研究打下基础。

1 材料与方法

1.1 材料与试剂

链霉菌菌株 (实验室编号: FBKL4.005) 筛选分离自茅台某酒厂高温酱香大曲; 细菌基因组DNA提取试剂盒 美国Biomiga公司; 制霉菌素、新生霉素、琼脂糖、Tris、EDTA二钠 北京索莱宝科技有限公司; 其他化学试剂均为国产分析纯。

1.2 仪器与设备

Thermo高速冷冻离心机 美国贝克曼库尔特有限公司; 核酸凝胶电泳仪、凝胶成像仪 德国耶拿分析仪器股份公司; 微型旋涡混合仪、Flex cycler多功能聚合酶链式反应 (polymerase chain reaction, PCR) 仪 美国Bio-Rad公司; BXM-30R立式灭菌锅 上海博讯实业有限公司医疗设备厂; SW-CJ-1D净化工作台 苏州净化设备有限公司。

1.3 方法

1.3.1 培养基的选择与配方

菌株纯化、斜面保藏培养基: ISP₂培养基。

液体种子扩培培养基: 葡萄糖10 g/L、酵母膏4 g/L、蛋白胨4 g/L、酵母浸粉4 g/L、K₂HPO₄ 4 g/L、KH₂PO₄ 2 g/L、MgSO₄ · 7H₂O 0.5 g/L, pH 7.2~7.4。

1.3.2 菌株总DNA的提取

从实验室低温 (4℃) 保藏的ISP₂斜面培养基上刮取适量待测菌株, 接种于100 mL液体种子扩培培养基中, 于45℃培养24 h后, 采用细菌基因组DNA提取试剂盒提取基因组DNA, 操作步骤参照试剂盒说明书进行, 所得基因组利用0.8%的琼脂糖电泳后在凝胶成像仪上对提取效果进行检测并用于基因组测序。

1.3.3 全基因组测序数据处理及分析

基因组DNA提取后采用Illumina HiSeq 2000测序平台完成全基因组测序, 上机前先利用HiSeq平台完成建库, 使用Agilent 2100对文库的插入片段进行检测, 并运用实时定量PCR技术对文库的有效浓度进行准确定量, 保证文库质量。将测序得到的原始数据进行过滤处理, 分别去除质量值不大于38的低质量碱基、N碱基达到10 bp、与连接物之间overlap超过15 bp的reads, 去除样品宿主以及重复污染, 得到有效数据。

经过预处理后得到的有效数据, 使用SOAP denovo组装软件^[22-23]进行序列拼接组装, 根据序列的双端信息, 确定出contig排列, 取长度大于500 bp的序列, 根据不同的K-mer进行组装得到初步组装结果, 然后采用krskgf、gapclose等软件对初步组装结果进行优化和补洞, 将contig连接成基因组序列, 得到最终组装结果, 并与GO (Gene Ontology)、COG (Cluster of Orthologous Groups of proteins)、KEGG (Kyoto Encyclopedia of Genes and

Genomes)、NR (Non-Redundant Protein Database)、TCDB (Transporter Classification Database) 数据库进行比对分析, 从而获得菌株基因组功能注释信息。

2 结果与分析

2.1 菌株基因组的组装

利用SOAP denovo组装软件^[22-23]对测序得到的序列双端的信息数据进行拼接(原始序列67 771 429个reads, 读长150 bp, 文库插入片段大小6 000 bp), 共获得7个scaffolds和69个contigs, 共9 441 858 bp的序列信息, 其中用于判断基因组拼接效果优劣的N50大小为352 516 bp, N90大小为116 315 bp; 经过优化与补洞处理后, 获得菌株整个基因组大小为9 454 406 bp, GC含量为73.03%, 基因平均长度为976 bp, 其中编码区基因总数为8 316个, 占整个基因组序列大小86.01%。由菌株基因GC含量及测序深度分布图可知(图1), 经过拼接组装后的GC含量主要分布在60%~80%之间, 序列深度多分布在70~200×的范围; Unigene长度分布集中在300~1 100 bp及大于2 000 bp范围内(图2), 其中400~500 bp长度的基因分布最多, 达到669个, 占总数的8.00%左右。

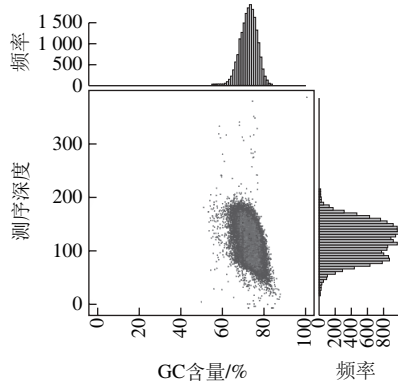


图1 菌株GC含量及测序深度分布图

Fig. 1 GC Content and sequencing depth distribution

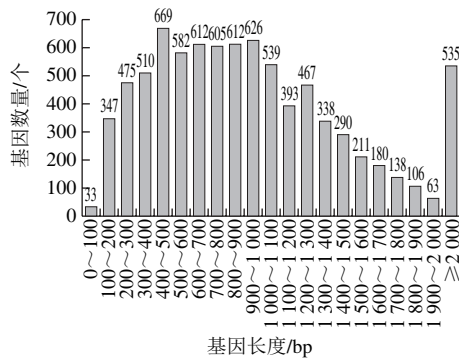


图2 菌株Unigene长度分布

Fig. 2 Unigene length distribution

2.2 菌株基因组的功能注释

将菌株测序数据与GO (Released November, 2013)、COG (Released September, 2015)、KEGG (Released April, 2016)、NR (Released April, 2016)、TCDB (Released April, 2016) 功能数据库进行BLAST比对, 同时使用antiSMASH (version 2.0.2) 对基因组上已知次级代谢基因簇进行预测, 对上述结果完成过滤 (BLASTP, $evalue \leq 1e^{-5}$), 对于每一条序列的BLAST结果, 均选取得分最高的比对结果 ($identity \geq 0\%$, $coverage \geq 40\%$) 进行注释。发现菌株基因组共有7 946个基因得到成功比对注释, 占基因总数的95.38%, 共有258个基因在所有数据库中得到注释, 占基因总数的3.10%; 其中, 在NR、GO、COG数据库中得到功能注释的基因较多, 分别为7 903、5 358个和5 720个, 占基因总数的94.86%、64.31%及68.66%, 而在TCDB数据库中得到注释的基因最少, 仅有451个, 占基因总数的5.41%, 具体的菌株基因组功能注释分布情况见图3。

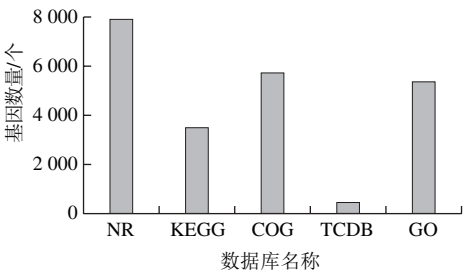


图3 菌株基因功能注释分布情况

Fig. 3 Database distribution of gene functional annotation from the strain

2.2.1 NR数据库基因注释分析

表1 菌株基因组NR数据库比对分析结果
Table 1 Results of BLAST with NR database

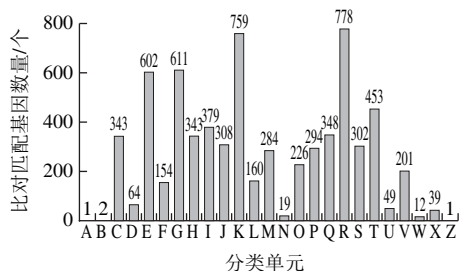
物种分类等级	注释到的物种	基因数量	百分比/%
放线菌目	链霉菌属 (<i>Streptomyces</i> sp.)	4 544	57.50
	吸水链霉菌 (<i>S. hygroscopicus</i>)	1 171	14.82
	熏衣草灰链霉菌 (<i>S. lavenduligriseus</i>)	342	4.33
	产色链霉菌 (<i>S. achromogenes</i>)	293	3.71
	玫瑰产色链霉菌 (<i>S. roseochromogenes</i>)	229	2.90
	<i>S. durhamensis</i>	173	2.19
	淡青链霉菌 (<i>S. glaucescens</i>)	50	0.63
	<i>S. leeuwenhoekii</i>	45	0.57
	山丘链霉菌 (<i>S. collinus</i>)	44	0.56
	<i>S. olindensis</i>	39	0.49
	龟裂链霉菌 (<i>S. rimosus</i>)	35	0.44
	阿氟曼链霉菌 (<i>S. scabiei</i>)	31	0.39
	单霉素链霉菌 (<i>S. monomycinii</i>)	30	0.38
	番薯链霉菌 (<i>S. ipomoeae</i>)	28	0.35
	绿色产色链霉菌 (<i>S. viridochromogenes</i>)	28	0.35
	肿瘤链霉菌 (<i>S. turgidiscabies</i>)	25	0.32
	卡特利链霉菌 (<i>S. cattleya</i>)	24	0.30

续表1

物种分类等级	注释到的物种	基因数量	百分比/%
	<i>S. aureocirculatus</i>	19	0.24
	白色链霉菌 (<i>S. albus</i>)	18	0.23
	库兹涅尔氏菌属 (<i>Kutzneria</i> sp.)	18	0.23
	<i>S. fulvoviolaceus</i>	17	0.22
	灰浅红链霉菌 (<i>S. griseorubens</i>)	17	0.22
	拒霉素链霉菌 (<i>S. resistomycificus</i>)	16	0.20
	淡黄色链霉菌 (<i>S. flaveolus</i>)	16	0.20
	棒状链霉菌 (<i>S. clavuligerus</i>)	15	0.19
	阿维链霉菌 (<i>S. avermitilis</i>)	15	0.19
	双色链霉菌 (<i>S. bicolor</i>)	15	0.19
	冰城链霉菌 (<i>S. bingchengensis</i>)	15	0.19
非放线菌目	其他	591	7.48

NR是一个非冗余的蛋白质数据库^[24], 通过BLAST软件, 将菌株FBKL4.005的基因序列翻译为相应的氨基酸序列, 并与NR数据库进行比对, 得到最终的注释结果。菌株一共有7 903个基因被注释到, 占基因总数的94.86%, 其中菌株FBKL4.005拼接基因组在属水平上与现有的链霉菌属 (*Streptomyces* sp.) 基因组核酸序列相似性在57.50%左右, 总体相似度最高; 而在种水平上则与吸水链霉菌 (*S. hygroscopicus*) 具有14.82%最高相似性。

2.2.2 COG数据库基因注释分析



A. RNA加工与修饰; B.染色质结构与动力学; C.能量产生与转化; D.细胞周期调控、细胞分裂, 染色体分隔; E.氨基酸转运与代谢; F.核苷酸转运和代谢; G.碳水化合物转运和代谢; H.辅酶转运与代谢; I.脂质转运与代谢; J.翻译, 核糖体结构和生物合成; K.转录; L.复制, 重组和修复; M.细胞壁/细胞膜的生成; N.细胞运动; O.翻译后修饰, 蛋白质转换, 伴侣; P.无机离子转运与代谢; Q.次级代谢产物的合成、转运和代谢; R.一般功能预测; S.未知功能; T.信号转导机制; U.细胞内的转运、分泌和囊泡运输; V.防御机制; W.胞外结构; X.噬菌体原、转座子遗传; Z.细胞骨架。

图4 菌株基因组COG数据库比对分析结果

Fig. 4 COG functional classification of maitake Unigenes

COG是由NCBI创建并维护的蛋白数据库, 根据细菌、藻类和真核生物完整基因组的编码蛋白系统进化关系分类构建而成^[25]。利用BLAST软件, 将菌株FBKL4.005的翻译得到的氨基酸序列, 与COG数据库进行比对, 完成相应基因的功能注释。从功能注释分析结果可以看出 (图4), 菌株FBKL4.005基因功能注释结果总共可以分为25类, 其中具有一般功能预测的注释结果最为丰富, 共778个, 占注释基因总数的13.60%, 其次

是与物质代谢密切相关的759个转录注释结果, 占注释基因总数的13.27%, 而与碳水化合物转运和代谢、氨基酸转运与代谢、信号转导机制等功能相关的基因也得到较多的注释结果, 分别为611、602个和453个, 此外还发现了302个功能未知的基因, 有待今后进一步研究。

2.2.3 GO数据库注释分析

表2 菌株基因组GO功能分类

Table 2 Gene ontology classification of the strain

GO数据库功能分类	基因功能描述	基因数量/个
细胞学组件	细胞	1 514
	细胞组分	1 514
	胞外区	21
	胞外区组分	21
	大分子复合物	207
	膜封闭内腔	15
	细胞器	164
	细胞器组分	63
	病毒	36
	病毒组分	36
	生物附着	18
	生物调节	993
	细胞组分的组织与生物合成	140
	细胞过程	2 677
	凋亡	4
生物学途径	发育过程	17
	定位建立	777
	免疫系统过程	3
	定位	785
	运动	17
	代谢过程	3 161
	多细胞有机体过程	74
	多细胞生物过程	9
	生物过程负向调控	50
	氮源利用	2
	生物过程正向调控	7
	生物过程调控	979
	繁殖	46
	生殖过程	45
分子功能	应激反应	475
	信号	300
	病毒繁殖	36
	抗氧化活性	12
	连接	2 697
	催化活性	2 987
	通道调节活性	1
	电子载体活性	2
	酶调节活性	14
	分子转导活性	224
	核酸结合转录因子活性	537
	蛋白结合转录因子活性	202
	结构分子活性	77
	转运活性	291

GO是1988年由基因本体联合会创立基因本体论数据库, 能够通过细胞学组件、生物学途径、分子功能3大分

支数据库对物种基因组进行分类和准确描述^[26]。将菌株FBKL4.005翻译氨基酸序列与GO数据库进行比对,对GO的不同分类比对结果进行统计,得到菌株功能基因含量的分布情况(图5)。菌株共有5 358个基因在GO数据库中被注释到,在3大类功能注释数据下又可分为43种功能注释结果(表2),其中的细胞学组件类有10个分支,共3 591个基因注释结果,菌株基因组与细胞、细胞组分功能组表现出最高相关性,各有1 514个;生物学途径类注释存在22个分支,共10 615个,其中菌株基因组与代谢过程、细胞过程、生物调节、生物过程调控等功能组相关性较高,分别为3 161、2 677、993个和979个;分子功能类有11个分支共7 044个相关性注释结果,其中菌株基因组与催化活性、连接功能组相关性最高,分别为2 987个和2 697个。

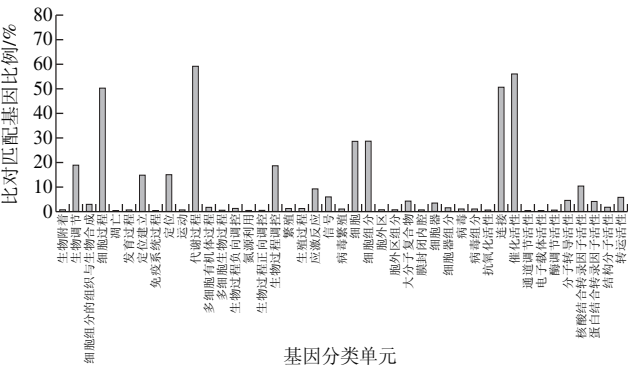


图5 菌株的功能基因含量分布情况

Fig. 5 Functional gene distribution of strain FBKL4.005

2.2.4 KEGG数据库注释分析

KEGG建立于1995年,是全面分析基因表达产物在细胞中的代谢途径及功能作用的重要参考,其中,最核心的便是KEGG PATHWAY数据库;通过数据库的比对查阅,可以快速方便地确定发挥某类功能相关的所有注释基因^[27-28]。本研究将菌株的氨基酸序列与KEGG数据库进行比对,对KEGG不同分类比对结果完成统计,发现菌株基因组共有3 492个基因得到注释,占菌株基因总数的43.95%,碳水化合物、氨基酸以及脂肪代谢为菌株基因组最主要涉及的几种代谢通路,分别有355、345个和155个基因注释结果。

进一步通过KEGG PATHWAY数据库分析,确定菌株有179个物质代谢通路得到注释,其中嘌呤代谢(93个)、糖酵解与糖异生作用(91个)、丁酸甲酯代谢(82个)、氨基苯甲酸酯降解代谢(79个)、脂肪酸代谢(78个)、精氨酸/脯氨酸代谢(76个)以及氨基糖/核苷酸糖代谢(75个)等通路与菌株基因组相比拥有较高的相关度(表3)。此外,通过比对结果的汇总还发现338个与白酒特征风味物质产生通路、187个与土腥味

物质代谢通路、87个与环境污染降解通路、74个与大分子糖类物质降解通路以及57个与链霉素和新霉素等抗生素产生通路相关的基因;因此根据这些信息,初步推测该菌株在白酒酿造体系中可能与特征风味产生有一定的关系,同时可能还具有一定的抗生素代谢功能和大分子物质降解能力,为今后该菌株代谢功能的探究打下了基础。

表3 菌株基因组KEGG数据库主要代谢通路分析
Table 3 Main metabolic pathways of the strain from KEGG database

代谢图编码	注释代谢通路名称	基因数量/个(≥40)
map00230	嘌呤代谢	93
map00010	糖酵解-糖异生途径	91
map00650	丁酸甲酯代谢	82
map00627	氨基苯甲酸酯降解	79
map00071	脂肪酸代谢	78
map00330	精氨酸和脯氨酸代谢	76
map00520	氨基糖、核苷酸糖代谢	75
map00620	丙酮酸代谢	73
map00280	缬氨酸、亮氨酸和异亮氨酸代谢	72
map00680	甲烷代谢	72
map00640	丙酮代谢	69
map00903	柠檬烯、蒎烯降解	69
map00240	嘧啶代谢	67
map00190	氧化磷酸化作用	66
map00260	甘氨酸、丝氨酸和苏氨酸代谢	65
map00362	苯甲酸降解	60
map00380	色氨酸代谢	60
map00500	淀粉、蔗糖代谢	60
map00310	赖氨酸降解	54
map00363	酚类降解	53
map00350	酪氨酸代谢	52
map00626	萘的降解	50
map02020	双组分系统	50
map00720	原核生物碳固定	49
map00250	丙氨酸、天冬氨酸和谷氨酸	48
map00051	果糖和甘露糖代谢	46
map00630	乙醛酸及脱羧产物代谢	46
map00052	半乳糖代谢	44
map00910	氮源代谢	44
map00270	半胱氨酸和蛋氨酸代谢	43
map00410	丙氨酸代谢	43
map00625	氯化烷烃的降解	43
map00624	多环芳烃降解	41
map00061	脂肪酸生物合成	40
map00281	香叶醇降解	40
map00860	叶绿素代谢	40

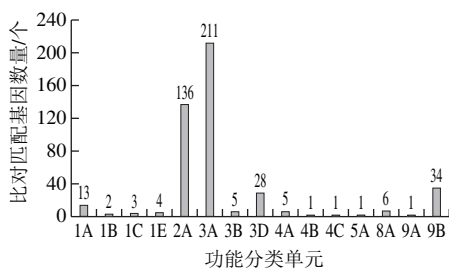
2.2.5 TCDB数据库注释分析

TCDB是转运蛋白分类数据库,包括离子通道的分类系统,该数据库提供了TC编号、描述信息和超过600个转运蛋白家族的实例数据;而转移系统则以5个级别进行分类,每个级别都对应于TC编号里的一个编号,用以表示特定类型的转运蛋白质^[29]。研究中使用BLAST软件,将菌株的氨基酸序列与TCDB数据库进行比对,分别得到

TCDB一级分类统计结果与二级分类统计结果。其中从一级分类统计结果来看(表4), 菌株基因组涉及到初级主动运输以及电化学势驱动转运体功能的基因最多, 分别为244个与136个, 初步表明了菌株FBKL4.005可能多以主动运输、电化学转运的方式分泌和吸收各种物质; 而从二级分类统计结果来看(图6), 与单向传递体、协同转运子、逆向转运子转运以及P-P磷酸化驱动转运蛋白相关的基因最多, 分别为211个和136个, 为菌株主要的转运蛋白系统。

表4 菌株TCDB数据库一级分类统计结果
Table 4 First level classification of the strain by TCDB database

TCDB一级分类编号	功能分类说明	基因数量/个
1	通道与孔道	22
2	电化学势驱动转运体	136
3	初级主动运输	244
4	基团移位	7
5	跨膜电子载体	1
8	运输辅助因素	6
9	不完全特征性运输系统	35



1A. α 型通道; 1B. β -桶型孔蛋白; 1C.成孔毒素(蛋白质和多肽); 1E.穴蛋白; 2A.转运(单向传递体、协同转运子、逆向转运子); 3A. P-P磷酸化驱动转运蛋白; 3B.脱羧反应驱动转运蛋白; 3D.氧化还原驱动蛋白; 4A.磷酸转移驱动转运蛋白; 4B.烟酰胺核糖核苷酸吸收转运蛋白; 4C.酰基辅酶A连接酶耦合转运蛋白; 5A.跨膜双电子转移载体; 8A.辅助转移蛋白; 9A.未知的转运蛋白的生化机制; 9B.假定转运蛋白。

图6 菌株TCDB数据库二级分类统计结果

Fig. 6 Second level classification of the strain by TCDB database

2.2.6 菌株次级代谢基因簇分析

次级代谢产物是微生物在一定的生长时期, 以初级代谢产物为前体合成的一系列大分子生长非必需物质; 通过次级代谢基因簇注释分析可以看出(表5), 菌株基因组有关萜烯类物质合成以及非核糖体肽合成酶的基因簇最多, 各有5个, 此外还发现了与芳香类化合物合成有关的基因簇, 其中萜烯类物质是酱香型白酒生产中链霉菌属的主要代谢产物, 与酿造过程中的土霉味产生密切相关, 同时在药香型以及浓香型白酒中也均有该类物质少量检出^[30-32], 是白酒中与风味形成有关的物质之一; 而芳香类化合物则同样存在于各种香型的白酒中, 同样是构成白酒风味的重要成分^[33]。

表5 菌株次级代谢基因簇分析

Table 5 Secondary metabolic gene clusters of the strain

Cluster编号	功能分类说明
2/6/8/14/29	萜烯类物质的合成(发酵风味相关)
10/12/15/19/33	非核糖体肽合成酶
4/9/11/32	细菌素产生
7/13/17	嗜铁素合成
21/22	四氢嘧啶
16/31	丁内酯类物质的合成
5/30/34	多功能酶复合物合成
18/27	黑色素产生
20/24	芳香类化合物的合成(发酵风味相关)
1	氨基香豆素的合成
23	查尔酮类化合物合成
25	芳香类、多功能酶复合物的合成
3/26/28	其他

3 讨论与结论

本研究采用Illumina HiSeq第2代高通量测序技术对分离自酱香大曲的具有耐高温特性的链霉菌菌株FBKL4.005进行全基因组测序, 经序列拼接组装后, 确定菌株基因组大小为9 454 406 bp, 是一种GC含量高达73.03%的微生物类群, 同时将整合基因组数据与GO、COG、KEGG、NR、TCDB等几大基本数据库的进行比对分析, 完成菌株基因组各方面功能的注释及数据统计工作, 从分子生物学的角度探究了该菌株的生物学特性以及代谢功能机制。

总体而言, 菌株FBKL4.005在基因上与链霉菌属尤其是该属下的吸水链霉菌具有最高的相似性, 从功能预测的角度来看, COG数据库的比对结果显示转录、碳水化合物、氨基酸代谢是菌株主要的功能预测结果; 通过GO数据库的比对, 细胞、细胞组分、代谢过程、催化活性等注释结果呈现出较高的相关性; 利用KEGG数据库中的代谢通路分析手段, 确定了菌株主要代谢通路的组成, 发现多个与白酒特征风味、土腥味、糖类物质降解、链霉素和新霉素产生等代谢通路相关的基因; 此外, 对菌株转运蛋白系统及次级代谢产物基因簇进行了探究, 从结果来看转运子转运、磷酸化驱动转运蛋白系统主导的主动运输、电化学转运为菌株运输物质的主要方式, 同时发现与白酒风味、土腥味物质代谢相关的基因结构。而针对类似菌株, 荆新云^[34]也完成了全测序分析, 获得了全长8 047 771 bp的基因组、7 570个编码基因以及23个次级代谢产物基因簇, 同时发现与其耐热性相关普遍胁迫蛋白基因的拷贝数比一般链霉菌高。与耐高温链霉菌4F相比, 菌株FBKL4.005基因组更大、次级代谢基因簇较多, 且发现了高温链霉菌4F所没有的风味代谢相关基因, 但其耐热性基因却未能有效证实。

研究针对酱香大曲中具有耐高温特性的链霉菌菌株,采用全基因组测序技术对其生物学特性和代谢功能机制进行探究,得到大量基因组学信息,为今后深入开展酱香大曲中耐热性链霉菌代谢功能特征及相关机制的研究提供了重要的数据参考。

参考文献:

- [1] CHATER K F, SOR B, JOON L K, et al. The complex extracellular biology of *Streptomyces*[J]. FEMS Microbiology Reviews, 2010, 34(2): 171-198. DOI:10.1111/j.1574-6976.2009.00206.x.
- [2] 傅俊曾, 冀兰鑫, 谢春峰. 链霉菌属微生物来源的生物活性天然产物[J]. 药物评价研究, 2010, 33(2): 153-158. DOI:10.7501/j.issn.0253-6376.
- [3] 尤小龙. 酱香白酒酿造过程放线菌代谢活性成分研究[D]. 贵阳: 贵州大学, 2016: 20-47.
- [4] 王涛, 游玲, 赵东, 等. 浓香型白酒酿造相关放线菌挥发性产物分析[J]. 食品科学, 2012, 33(14): 184-187.
- [5] 丁秀芹. 微生物基因组研究进展[J]. 科技资讯, 2009(5): 11-12. DOI:10.3969/j.issn.1672-3791.2009.05.009.
- [6] BARTRAM A, LYNCH M, STEARNS J, et al. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads[J]. Applied & Environmental Microbiology, 2011, 77(11): 3846-3852. DOI:10.1128/AEM.05896-11.
- [7] LUO C, TSEMENTZI D, KYRPIDES N, et al. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample[J]. PLoS ONE, 2012, 7(2): e30087. DOI:10.1371/journal.pone.0030087.
- [8] LI X, DING C, ZHANG T, et al. Fungal pathogen accumulation at the expense of plant-beneficial fungi as a consequence of consecutive peanut monoculturing[J]. Soil Biology and Biochemistry, 2014, 72: 11-18. DOI:10.1016/j.soilbio.2014.01.019.
- [9] YU Z, YANG J, ZHOU J, et al. Water stratification affects the microeukaryotic community in a subtropical deep reservoir[J]. Journal of Eukaryotic Microbiology, 2014, 61(2): 126-133. DOI:10.1111/jeu.12090.
- [10] 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. DOI:10.13343/j.cnki.wsxb.20150155.
- [11] 孙欣, 高莹, 杨云锋. 环境微生物的宏基因组学研究新进展[J]. 生物多样性, 2013, 21(4): 393-400. DOI:10.3724/SP.J.1003.2013.08050.
- [12] FENG L, WANG W, CHENG J. Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir[J]. Proceedings of the National Academy of Sciences of the United States of America, 2007, 104(13): 5602-5607. DOI:10.1073/pnas.0609650104.
- [13] DUNFIELD P, YURYEV A, SENIN P. Methane oxidation by an extremely acidophilic bacterium of the phylum Verrucomicrobia[J]. Nature, 2007, 450: 879-882. DOI:10.1038/nature06411.
- [14] HOU S, MAKAROVA K, SAW J. Complete genome sequence of the extremely acidophilic methanotroph isolate V4, *Methylacidiphilum infernorum*, a representative of the bacterial phylum Verrucomicrobia[J]. Biology Direct, 2008, 3(1): 1-25. DOI:10.1186/1745-6150-3-26.
- [15] YAN Y, YANG J, DOU Y. Nitrogen fixation island and rhizosphere competence traits in the genome of root-associated *Pseudomonas stutzeri* A1501[J]. Proceedings of the National Academy of Sciences of the United States of America, 2008, 105(21): 7564-7569. DOI:10.1073/pnas.0801093105.
- [16] 喻婵, 王琰, 徐承晨, 等. 蜡状芽孢杆菌菌群代谢途径的分析和比较[J]. 遗传, 2011, 33(10): 1057-1066. DOI:10.3724/SP.J.1005.2011.01057.
- [17] WANG F, WANG J, JIAN H. Environmental adaptation: genomic analysis of the piezotolerant and psychrotolerant deep-sea iron reducing bacterium *Shewanella piezotolerans* WP3[J]. Journal of Biotechnology, 2008, 136(4): e1937. DOI:10.1016/j.jbiotec.2008.07.1275.
- [18] MACKELPRANG R, WALDROP M, DEANDELIS K, et al. Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw[J]. Nature, 2011, 480: 368-371. DOI:10.1038/nature10576.
- [19] SHI P, JIA S, ZHANG X, et al. Metagenomic insights into chlorination effects on microbial antibiotic resistance in drinking water[J]. Water Research, 2013, 47(1): 111-120. DOI:10.1016/j.watres.2012.09.046.
- [20] KUANG J, HUANG L, CHEN L, et al. Contemporary environmental variation determines microbial diversity patterns in acid mine drainage[J]. ISME Journal, 2013, 7(5): 1038-1050. DOI:10.1038/ismej.2012.139.
- [21] 王晓丹, 班世栋, 周鸿翔, 等. 贵州省遵义地区3个酱香型大曲细菌群落的比较分析[J]. 食品科学, 2016, 37(7): 110-116. DOI:10.7506/spkx1002-6630-201607021.
- [22] LI R, ZHU H, RUAN J, et al. *De novo* assembly of human genomes with massively parallel short read sequencing[J]. Genome Research, 2010, 20(2): 265-272. DOI:10.1101/gr.097261.109.
- [23] LI R, LI Y, KRISTENSEN K, et al. SOAP: short oligonucleotide alignment program[J]. Bioinformatics, 2008, 24(5): 713-714. DOI:10.1093/bioinformatics/btn025.
- [24] LI W, JAROSZEWSKI L, GODZIK A. Tolerating some redundancy significantly speeds up clustering of large protein databases[J]. Bioinformatics, 2002, 18(1): 77-82. DOI:10.1093/bioinformatics/18.1.77.
- [25] TATUSOV R, FEDOROVA N, JACKSON J, et al. The COG database: an updated version includes eukaryotes[J]. BMC Bioinformatics, 2003, 4(1): 41. DOI:10.1186/1471-2105-4-41.
- [26] ASHUBUENER M, BALL C, BLAKE J, et al. Gene Ontology: tool for the unification of biology[J]. Nature Genetics, 2000, 25(1): 25-29. DOI:10.1038/75556.
- [27] KANEHISA M, GOTO S, KAWASHIMA S, et al. The KEGG resource for deciphering the genome[J]. Nucleic Acids Research, 2004, 32(Suppl 1): D277-D280. DOI:10.1093/nar/gkh063.
- [28] KANEHISA M, GOTO S, HATTORI M, et al. From genomics to chemical genomics: new developments in KEGG[J]. Nucleic Acids Research, 2006, 34(Suppl 1): D354-D357. DOI:10.1093/nar/gkj102.
- [29] MILTON S, VAMSEE S, DORJEE G, et al. The transporter classification database[J]. Nucleic Acids Research, 2014, 42: D251. DOI:10.1093/nar/gkt1097.
- [30] 杨帆, 陈良强, 林琳, 等. 一株放线菌的分离鉴定及其代谢产物研究[J]. 酿酒科技, 2014(9): 42-44. DOI:10.13746/j.njkj.2014.0174.
- [31] 范文来, 胡光源, 徐岩. 顶空固相微萃取-气相色谱-质谱法测定药香型白酒中萜烯类化合物[J]. 食品科学, 2012, 33(14): 110-116.
- [32] 徐占成, 陈勇, 莫凯, 等. 剑南春酒中萜烯类生理活性物质的分析研究[J]. 食品与发酵科技, 2009, 45(5): 1-2. DOI:10.3969/j.issn.1674-506X.2009.05-001.
- [33] 沈怡方. 白酒生产技术全书[M]. 北京: 中国轻工业出版社, 2015: 45-53.
- [34] 荆新云. 链霉菌4F基因组测序、蛋白质组分析及基因组工程[D]. 北京: 中国科学院大学, 2013: 104-108.