

1株来源于茅台酒酿造过程宛氏拟青霉MTDF-01的全基因组测序及分析

王和玉¹, 刘延峰², 张巧玲¹, 堵国成², 杨帆¹, 李江华^{2,*}

(1.贵州茅台酒股份有限公司, 贵州 仁怀 564501; 2.江南大学 工业生物技术教育部重点实验室, 江苏 无锡 214122)

摘要:为深入了解宛氏拟青霉(*Paecilomyces variotii*)在酱香型白酒生产过程中代谢机理,为今后进一步挖掘及调控*P. variotii*在白酒酿造过程中代谢提供必要生物信息学基础。通过PacBio RS II测序平台对*P. variotii* MTDF-01进行全基因组测序,并且对测序数据进行基因组组装、基因预测和功能注释、碳水化合物基因预测、次级代谢产物合成基因簇预测,以及已与报道全基因组序列的分离于土壤中的耐甲醛*P. variotii* No.5进行比较基因组分析和共线性分析。基因组拼接得到19个基因组骨架和19个重叠群,总长度为30 833 540 bp, GC平均含量为47.46%。基因组中预测到8 815个基因、5 044个简单重复序列,221个tRNA,49个rRNA。总共注释基因8 662个,其中淀粉和纤维素水解酶基因分别有60个和165个,次级代谢产物合成基因簇23个。*P. variotii* MTDF-01与*P. variotii* No.5基因组存在翻转、异位等基因组重排,两者基因组共存在16 414处单核苷酸碱基突变,其中有525个碱基缺失,335个碱基插入,15 554个单碱基替换。本研究利用PacBio RS II测序平台获得了*P. variotii* MTDF-01的全基因组信息,并且分析*P. variotii* MTDF-01的潜在功能,为深入了解*P. variotii*在白酒生产过程中的代谢机理提供了遗传信息基础,对今后酱香型白酒中风味和健康物质的调控研究具有重要意义。

关键词:茅台酒;拟青霉;全基因组测序;基因注释

Whole-Genome Sequencing and Analysis of *Paecilomyces variotii* MTDF-01, Isolated during Maotai-Flavor Liquor Brewing

WANG Heyu¹, LIU Yanfeng², ZHANG Qiaoling¹, DU Guocheng², YANG Fan¹, LI Jianghua^{2,*}

(1. Kweichow Moutai Distillery Co. Ltd., Renhuai 564501, China;

2. Key Laboratory of Industrial Biotechnology, Ministry of Education, Jiangnan University, Wuxi 214122, China)

Abstract: This study aimed to gain insight into the metabolic mechanism of *Paecilomyces variotii* in the production of Maotai-flavor liquor for the purpose of providing an essential bioinformatic basis for further understanding and regulating the metabolism of this strain. The whole-genome sequencing of *P. variotii* MTDF-01, a strain isolated during the brewing of Maotai-flavor liquor, was carried out on the PacBio RS II platform. The sequencing reads were assembled and annotated via gene prediction and functional annotation. In the whole genome, 19 scaffolds and 19 contigs were obtained. The total length was 30 833 540 bp and the average GC content was 47.46%. A total of 8 815 genes, 5 044 simple sequence repeats (SSR), 221 tRNA and 49 rRNAs were predicted. The number of annotated genes was 8 662. The genome contained 60 starch hydrolase-coding genes and 165 cellulose hydrolase-coding genes, as well as 23 secondary metabolite biosynthetic gene clusters. The genome rearrangement of MTDF-01 was different from that of *P. variotii* No.5, a formaldehyde-resistant strain isolated from soil. Synteny analysis indicated that genome rearrangement and translocation occurred among these genomes. A total of 16 414 single nucleotide base mutations were found in the genomes of the two strains, including 525 base deletions, 335 base insertions and 15 554 single base substitutions. The high-quality whole-genome sequence and potential function of *P. variotii* MTDF-01 obtained in this study are of great significance for future studies on the regulation of the flavor and health-benefiting substances of Maotai-flavor liquor.

Keywords: Moutai-flavor liquor; *Paecilomyces variotii*; whole-genome sequencing; gene annotation

DOI:10.7506/spkx1002-6630-20181008-036

中图分类号: TS261.1

文献标志码: A

文章编号: 1002-6630(2019)24-0185-08

收稿日期: 2018-10-08

第一作者简介: 王和玉(1973—)(ORCID: 0000-0002-0907-6189),男,工程师,学士,研究方向为白酒酿造微生物。

E-mail: zql911@126.com

*通信作者简介: 李江华(1968—)(ORCID: 0000-0002-3089-8961),男,教授,博士,研究方向为发酵工程和酶工程。

E-mail: lijiahua@jiangnan.edu.cn

引文格式:

王和玉, 刘延峰, 张巧玲, 等. 1 株来源于茅台酒酿造过程宛氏拟青霉MTDF-01的全基因组测序及分析[J]. 食品科学, 2019, 40(24): 185-192. DOI:10.7506/spkx1002-6630-20181008-036. <http://www.spkx.net.cn>

WANG Heyu, LIU Yanfeng, ZHANG Qiaoling, et al. Whole-genome sequencing and analysis of *Paecilomyces variotii* MTDF-01, isolated during Maotai-flavor liquor brewing[J]. Food Science, 2019, 40(24): 185-192. (in Chinese with English abstract) DOI:10.7506/spkx1002-6630-20181008-036. <http://www.spkx.net.cn>

白酒发酵过程中复杂的微生物种群是决定白酒风味特征的重要因素^[1-3]。霉菌能够分泌多种酶类分解原料为发酵过程微生物提供代谢底物, 同时其自身可生成各种代谢产物, 对白酒风味形成有重要贡献^[4-6]。在白酒生产中主要的功能霉菌包括毛霉、曲霉等。在酱香型白酒的酿造过程制曲、堆积发酵和窖池发酵中, 宛氏拟青霉(*Paecilomyces variotii*)是优势霉菌之一^[7], 研究表明该菌具有高糖化酶活力, 能够降解原料中的淀粉^[8], 并且在制曲工艺、堆积工艺以及兼性条件下分别进行纯种固态发酵, 能代谢生成酸、醇、醛、酮、酯及芳香族化合物等多种物质, 发酵产生多种风味物质和风味前体物质, 包括苯乙醇、苯乙酸等, 是酱香型白酒中众多呈香呈味物质中的组成部分^[9]。同时*P. variotii*也能够分泌降解糖醛的酶, 可有效降低酒醅中的糖醛^[10]。

目前, 在白酒酿造行业, 针对酿造微生物功能研究已从传统的发酵分析手段转向代谢组学结合宏基因组、转录组进行全方面的解析^[11-14], 迄今为止, 白酒酿造微生物全基因组测序报道有地衣芽孢杆菌(*Bacillus licheniformis*) GMCC3963^[15]、华根霉(*Rhizopus chinensis*) CTCCM201021^[16]、耐高温放线菌(*Thermoactinomyces daqus*) H-18^[17]以及窖泥中分离得到的丁酸梭菌(*Clostridium butyricum*) JKY6D110^[18]等。*P. variotii*作为酱香型白酒酿造过程的优势微生物之一, 在制曲及制酒阶段活动活跃, 目前对其研究主要在于产酶特性及代谢产物分析^[8], *P. variotii*全基因组分析鲜见报道, 对*P. variotii*发酵代谢机制及生理生化等微生物特性缺乏足够的认识和了解, 限制了深入开发和利用该微生物。

本研究通过PacBio RS II测序平台, 对茅台酒酿造过程中分离得到的*P. variotii* MTDF-01进行全基因组测序, 根据测序获得的数据进行基因组分析、基因功能注释和比较基因组学分析, 为今后进一步深入了解*P. variotii*在酱香型白酒生产过程中的代谢机理以及为调控菌株代谢能力提供重要的生物信息学基础。

1 材料与方法

1.1 材料与试剂

菌株MTDF-01分离自茅台酒酿造过程。

真菌基因组DNA提取试剂盒 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

ACB-401超净工作台 新加坡Esco科技有限公司; spx-250B-z生化培养箱 上海福马实验设备有限公司; MaxQ 6000轨道摇床 赛默飞世尔科技公司; 5804R台式高速冷冻离心机 德国Eppendorf公司; GelDoc XR凝胶成像系统 美国Bio-Rad公司。

1.3 方法

1.3.1 菌体培养与收集

菌体培养: 拟青霉划线PDA琼脂斜面培养基37 °C培养72 h后通过无菌生理盐水洗下孢子, 制备成孢子悬浮液。

菌体摇瓶培养: 吸取450 μL孢子悬浮液(孢子浓度数量级10⁷)接种至20 mL/250 mL的PDA液体培养基中, 摇床培养30 °C、转速200 r/min。将培养好的菌体通过过滤收集。

1.3.2 菌株总DNA提取

采用真菌基因组DNA提取试剂盒提取基因组DNA, 操作步骤参照试剂盒说明书。提取得到的基因组通过0.8%琼脂糖凝胶电泳检测。

1.3.3 全基因组测序组装

本次测序委托武汉生物技术研究完成。样本质检合格后, 用Covarisg-TUBE对基因组DNA随机打断, 进行文库构建。利用磁珠富集、纯化大片段DNA, 对片段化的DNA进行损伤修复、末端修复; 在DNA片段两端连接茎环状测序接头, 并利用外切酶去除连接失败的片段, 纯化后, 采用Agilent 2100 Bioanalyzer HighSensitivity Kit进行文库质量检测。采用第3代测序仪PacBioRS II对DNA进行非扩增长片段测序, 测序完成后去除adapter序列及截掉reads部分区域的低质量碱基, 采用HGAP^[19]流程进行组装, 挑选长reads作为种子序列, 其他的较短reads通过BLASR软件^[20]比对到种子序列, 进行校正。最终以校正后的高质量种子reads采用Celera assembler组装软件^[21]用OLC组装算法进行组装。利用Quiver软件^[22]进行组装结果的优化和校正。在Quiver校正结果的基础上, 去除低覆盖度的重叠群及冗余的重叠群。

1.3.4 基因预测与注释

通过EVM(EvidenceModeler)软件^[23]整合Augustus

软件^[24]、SNAP软件^[25]、GeneWise软件^[26]基因预测结果获得基因结构预测,将获得的基因提交COG (cluster of orthologous groups of proteins)^[27]、GO (gene ontology)^[28]、KEGG (kyoto encyclopedia of genes and genomes)^[29-30]、NR (non-redundant protein database)、Swiss-Prot数据库^[31]、TrEMBL数据库^[31]进行比对,获得功能注释信息。

1.3.5 物种进化树构建及分析

挑选菌株进行系统发育树构建,运用MEGA 7.0软件 (<https://www.megasoftware.net/>) Neighbor-Joining法进行分析,以18S rRNA基因进行比较分析构建系统进化树。

1.3.6 CAZy糖酶分析

通过dbCAN HMMs 3.0在线工具^[32] (<http://csbl.bmb.uga.edu/dbCAN/index.php>) 将*P. variotii* MTDF-01的CDS序列与碳水化合物活性酶 (CAZy) 数据库 (<http://www.cazy.org/>) 进行比对分析。比对时dbCAN的阈值参数设置如下:如果比对的序列长度大于80 aa, 阈值 $<1 \times 10^{-5}$, 否则阈值 $>1 \times 10^{-3}$ 。

1.3.7 次级代谢产物合成基因簇分析

利用antiSMASH 3.0^[33] (<https://fungismash.secondarymetabolites.org>) 对MTDF-01菌株中次生代谢物合成基因组簇进行预测。

1.3.8 基因组共线性分析

利用Mummer程序 (<https://sourceforge.net/projects/mummer/postdownload>, 版本3.23) 对MTDF-01及NCBI中已测序的菌株*P. variotii* No.5进行分析比较,并用MUMmerplot将结果可视化。菌株*P. variotii* No.5序列从NCBI网站下载 (ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/497/085/GCA_000497085.1_PVAR5_assembly01/)。

2 结果与分析

2.1 MTDF-01全基因组基本特征及初步分析

PacBio RS II是首个商业化应用的第3代测序平台,其采用特有的单分子实时技术使得测序读长较第2代测序技术显著增加(半数测序读长大于20 kb),同时减弱GC性^[19]。通过PacBio RS II测序平台对菌株MTDF-01进行全基因组测序,获得菌株MTDF-01全基因组数据。测序数据过滤后总数据量为2.58 G,平均测序读长为14.64 kb,最长的测序读长达到46.94 kb,测序深度为83.81×,DNA靶序列平均长度为7 823 bp。对原始数据进行组装后得到参考序列由19个重叠群组成,形成19个基因组骨架,总长度为30 833 540 bp,其中最重叠群为7 588 950 bp,GC平均含量为47.46%。MTDF-01基因组测序数据提交至NCBI, GenBank登录号为SAMN09987769。

2.2 系统进化树构建

为进一步确认MTDF-01的种属关系,本研究以18S rRNA基因进行比较分析,构建系统发育树(图1)。结果表明MTDF-01与*P. variotii* CBS 102.74、*P. variotii* CBS 101075聚为一支,可以确定为*P. variotii*,同时与*P. variotii* No.5也聚为一支,两者物种亲缘关系较近。

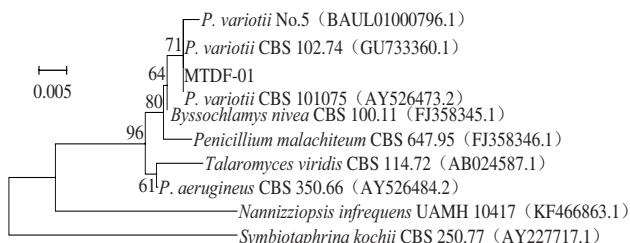


图1 通过MEGA Neighbor-Joining法构建的真菌分子系统发育树

Fig. 1 Molecular phylogenetic tree of fungi constructed by Neighbor-Joining method with MEGA software

2.3 基因结构预测

基于ab initio预测软件(Augustus、SNAP、GeneMark-ES)及同源比对预测两种方法对*P. variotii* MTDF-01测序拼接片段进行基因预测与编码区分析,共预测出基因8 815个,基因总长度为13 530 750 bp,平均基因长度为1 740 bp,平均编码序列长度为478 bp。预测基因序列采用多个数据库(KOG\GO\KEGG\NR\Swiss-Prot\TrEMBL)进行比对,获得相应功能注释信息,最终注释基因8 662个,详细统计数据见表1。在基因组序列上共找到5 044个简单重复序列,分别采用同源比对和de novo预测的方式对基因组的重复序列进行注释,结果显示该基因组中含有3.80%的重复序列。在基因序列上共找到tRNA数量为221个, rRNA为49个。

表1 基因功能注释统计
Table 1 Gene function annotation statistics of MTDF-01

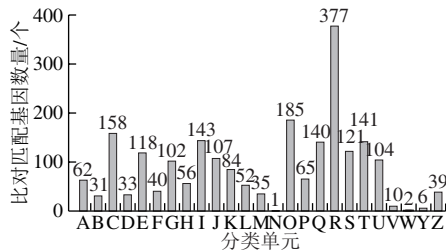
类型	注释数量	占预测基因比例/%
KOG	1 972	22.37
GO	5 490	62.28
KEGG	1 614	18.31
NR	8 657	98.21
UniProt	6 339	71.91
TrEMBL	8 536	96.83
预测基因总数	8 815	100

2.4 基因功能分类

2.4.1 COG聚类分析

NCBI创建并维护的蛋白数据库COG是根据细菌、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。真核生物的序列功能可通过真核基因组数据库KOG进行比对及预测。对*P. variotii* MTDF-01基因功能进行初步分析,选择利用COG进行基因注释与功能分

类,共注释基因1 972 个,占总蛋白数的22.37%,可分为25 个功能组,3 大类分别为细胞过程及信号、信息储存及加工、代谢过程,分别注释基因数为556、336 个和822 个。从图2可看出,分类为R(普通功能预测)的基因数量最多为377 个,其余基因序列主要集中在O(蛋白翻译后修饰)、C(能量产生与转换)、I(脂质转运和代谢)及Q(次级代谢物合成、转运和代谢)等分类单元。



A. RNA加工和修饰; B.染色体结构和动力; C.能源生产与转换; D.细胞周期控制, 细胞分裂, 染色体分配; E.氨基酸转运和代谢; F.核苷酸转运和代谢; G.碳水化合物转运和代谢; H.辅酶转运和代谢; I.脂质转运和代谢; J.翻译、核糖体结构和生物转化; K.转录; L.复制、重组和修复; M.细胞壁/膜/包膜生物发生; N.细胞运动; O.蛋白翻译后修饰, 蛋白转换, 分子伴侣; P.无机离子转运和代谢; Q.次级代谢物合成、转运和分解代谢; R.普通功能预测; S.功能未知; T.信号转导机制; U.细胞内运输, 分泌和小泡运输; V.防御机制; W.细胞外结构; Y.核结构; Z.骨架。

图2 MTDF-01预测蛋白COG功能分类

Fig. 2 COG cluster analysis of MTDF-01 proteins

2.4.2 GO聚类分析

基于基因本体数据库GO功能分类, 该数据库按照细胞学组件、分子功能、生物学过程对蛋白进行分类注释。本研究将MTDF-01菌株进行蛋白质功能分析, 共预测得到5 490 个蛋白, 注释统计结果如表2所示。可看出菌株的GO功能聚类在细胞、细胞组分、连接、催化活性、细胞过程、代谢过程等条目的基因数占据优势, 其中细胞和细胞组分各有1 187 个基因, 分子功能中的连接和催化活性功能分别有2 929 个和2 922 个, 生物学过程中的细胞过程和代谢过程分别注释到2 077 个及2 914 个基因, 表明在菌株的蛋白功能主要为细胞组成、代谢和酶催化。

表2 MTDF-01预测蛋白GO分类

Table 2 Gene ontology classification of MTDF-01 proteins

功能分类	功能描述	基因数量/个
细胞学组件	细胞	1 187
	细胞组分	1 187
	细胞器组分	318
	细胞器	826
	细胞膜组分	556
	细胞膜组分	824
	蛋白质复合物	477
	超分子复合物	7
	膜封闭腔	83
	胞外区	5
	类核	1
	细胞连接	1

续表2

功能分类	功能描述	基因数量/个
分子功能	催化活性	2 922
	连接	2 929
	信号转导活性	23
	转运活性	304
	抗氧化活性	27
	分子功能调节剂	42
	结构分子活性	126
	营养储存活性	6
	转录调节活性	226
	分子载体活性	1
	定位	727
	刺激应答	278
	代谢过程	2 914
	细胞过程	2 077
生物学过程	生物调节	498
	生物过程调节	472
	生物过程负调控	22
	信号	123
	细胞组分组织或生物发生	202
	多菌种过程	3
	细胞聚集	1
	生物相	1
	消毒	3
	ATP分解代谢过程	5
	发育过程	4
	生物过程的正向调节	16
	繁殖	5
	生殖过程	5
	免疫系统过程	1
	运动	1
	生物黏附	1
	GTP分解代谢过程	35
	细胞增殖	1
	碳利用	1

2.4.3 KEGG代谢途径预测分析

本研究通过将*P. variotii* MTDF-01与KEGG数据库进行比对分析, 共注释得到物质代谢通路274 个。MTDF-01中预测代谢通路(图3)在新陈代谢、遗传信息加工、细胞过程3 个大类中涉及基因较多, 遗传信息加工(转录、翻译、蛋白质的折叠和加工等)共有566 个基因, 细胞过程(转运代谢、细胞运动、细胞生长与死亡细胞通讯等)共有382 个基因, 而涉及代谢的这类基因最多, 行使的功能较多, 既有细胞本身合成氨基酸的基因, 也有氨基酸代谢、膜蛋白基因以及能量系统基因等, 共有1 506 个。在代谢中涉及基因最多的通路主要有嘌呤代谢途径(ko00230)、氧化磷酸化途径(ko00190)、嘌呤代谢途径(ko00240)、糖酵解途径(ko00010)、淀粉与蔗糖代谢途径(ko00500)、精氨酸与脯氨酸代谢途径(ko00330)。

通过比对结果发现, MTDF-01中涉及氨基酸代谢的代谢途径有20 条共462 个基因, 包含了苯丙氨酸、色氨

酸和酪氨酸等芳香族氨基酸代谢途径,芳香族氨基酸可为苯乙醇、苯乙酸等风味物质代谢提供前体物质,但缺失精氨酸合成途径的相关基因。进一步对白酒特征风味物质相关通路分析发现MTDF-01具有苯乙醇、苯乙酸、亚麻油酸、苯甲酸、肉桂酸乙酯、乙酸、乳酸等风味物质的代谢途径。其中苯乙醇和苯乙酸可通过苯丙氨酸、酪氨酸和色氨酸生物合成途径将苯丙氨酸在氧化氢-过氧化物酶作用下生成2-苯乙酰胺,进一步在酰胺酶作用下生成苯乙酸,苯乙酸可在乙醛脱氢酶和芳基乙醇脱氢酶作用下生成苯乙醇。苯丙氨酸也可通过芳香族L-氨基酸脱羧酶作用生成苯乙胺,苯乙胺在芳基乙醇脱氢酶作用下生成苯乙醇。在*P. variotii* MTDF-01中有多条通路可合成乙酸,该菌株具有乙醇脱氢酶和乙醛脱氢酶相关基因能够将乙醇氧化成乙酸,还可丙酮酸代谢途径、乙醛酸和二羧酸盐代谢途径代谢生成乙酸。注释结果表明*P. variotii* MTDF-01有亚麻油酸、油酸、花生四烯酸、二十二碳六烯酸等多种有机酸合成的关键酶。有文献报道,*P. variotii*具有降解糠醛的功能^[10],但在注释结果中并未发现糠醛降解途径。根据预测信息可推测该菌株在白酒发酵过程中可能对白酒风味代谢具有重要贡献。

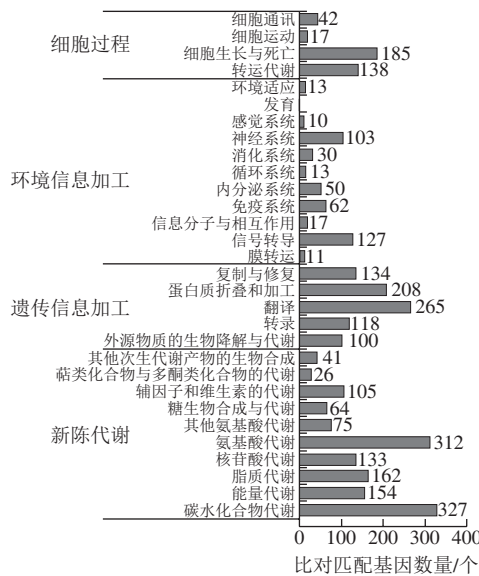


图3 *P. variotii* MTDF-01 KEGG功能分类

Fig. 3 KEGG cluster analysis of *P. variotii* MTDF-01 proteins

2.4.4 CAZy糖酶分析

在白酒酿造过程中,淀粉酶能够将原料中的淀粉降解生成葡萄糖等利于微生物利用的还原糖,为发酵进程提供营养底物,促进酵母等微生物生长代谢。原料中的纤维成分较多,纤维素酶能够对纤维素进行降解,释放淀粉,有利于糖化酶的作用,同时纤维素酶能够将原料淀粉中3%左右的纤维素和半纤维素转化成可发酵性糖,有利于酵母的利用^[34],此外,纤维素酶也会影响白酒的风味^[35]。

将基因组序列与CAZy数据库进行比对,在*P. variotii* MTDF-01的基因组中共有907个基因编码的蛋白质结构域属CAZy家族,包括425个糖苷水解酶(glycoside hydrolases, GHs)、307个糖苷转移酶、75个碳水化合物酯酶、5个多糖裂解酶、213个碳水化合物结合组件(carbohydrate-binding modules, CBMs),其中194个基因含有多种组分结构域,表明MTDF-01含有丰富的碳水化合物代谢酶。CBMs作为一些GHs所含有的非催化活性结构的结构域,使得GHs能靠近底物,以增加GHs锚定底物的能力,从而影响GHs的酶催化活性。通过预测MTDF-01的基因组中共有108个GHs基因含有CBMs结构域。

淀粉降解的酶主要为 α -淀粉酶、 α -葡萄糖苷酶(糖化酶)等,纤维素酶主要为 $\alpha\beta$ -葡聚糖酶、 β -葡萄糖苷酶,木聚糖是半纤维素中最主要的碳水化合物,因此木聚糖酶也成为了半纤维素降解中的主要酶类。为进一步探索*P. variotii* MTDF-01对淀粉及纤维素的降解能力,本研究对在*P. variotii* MTDF-01基因组中淀粉/糖原、纤维素等降解酶相关酶进行统计,见表3。结果表明在*P. variotii* MTDF-01基因组中含有多个淀粉及纤维素水解酶基因,分别为60个及165个,可推测*P. variotii* MTDF-01具有降解淀粉和纤维素的潜力。在白酒生产过程中,纤维素酶来源广泛,大曲中的细菌、真菌和放线菌都有产纤维素酶的能力。在曲霉^[36]、*Paenibacillus* sp.、*Acinetobacter* sp.^[37]等微生物中均有较强的纤维素降解力,但对*P. variotii*纤维素降解方面研究较少,*P. variotii*纤维素降解能力仍需进一步进行实验验证。

表3 *P. variotii* MTDF-01淀粉及纤维素降解酶CAZy预测结果

Table 3 Predicted starch and cellulose degrading-enzymes of *P. variotii* MTDF-01 from CAZy database

酶	作用底物	GH家族	基因数量/个
α -淀粉酶	淀粉	GH1	2
		GH13_1	3
		GH13_5	1
		GH4	9
		GH13	21
α -葡萄糖苷酶	淀粉或糖原	GH15	5
		GH31	16
		GH63	1
		GH133	2
		GH5	23
纤维素酶	纤维素	GH6	8
		GH51	2
		GH12	1
		GH1	2
		GH3	16
β -葡萄糖苷酶	纤维素	GH5_9	6
		GH17	18
		GH12	1
		GH16	18
		GH43	65
木聚糖酶	木聚糖	GH51	2

2.4.5 次级代谢产物合成基因簇分析

表 4 MTDF-01次级代谢产物合成基因簇
Table 4 Secondary metabolic gene clusters of MTDF-01

基因簇类型	起始位置	终止位置	最相似的已知基因簇	相似性/%
萜烯类	1 491 356	1 512 630	—	
萜烯类	1 518 311	1 540 201	—	
萜烯类	2 388 359	2 409 990	—	
萜烯类	5 272 735	5 295 215	—	
萜烯类	5 532 588	5 553 976	—	
<i>Tlpks</i>	6 261 255	6 309 223	—	
<i>Tlpks</i>	7 171 116	7 219 328	—	
<i>Tlpks</i>	2 192 107	2 238 203	—	
<i>Tlpks</i>	4 469 790	4 517 240	Depudecin	33
<i>Tlpks</i>	2 579 770	2 614 871	—	
<i>Tlpks</i>	474 429	518 376	—	
<i>Tlpks</i>	1 344 826	1 386 321	Aflatox	8
<i>Tlpks-Nrps</i>	519 870	571 911	—	
<i>Tlpks-Nrps</i>	2 513 077	2 565 191	—	
<i>Nrps</i>	3 610 441	3 661 454	—	
<i>Nrps</i>	667 526	711 974	—	
<i>Nrps</i>	400 849	455 107	—	
<i>Nrps</i>	2 244 192	2 302 863	—	
<i>Nrps</i>	771 510	820 882	—	
<i>Nrps</i>	1 123 207	1 177 593	—	
Other	3 704 030	3 745 548	—	
Other	2 197 753	2 242 793	—	
Other	1 528 501	1 571 827	—	

注：—未知功能基因簇。

聚酮合酶基因合成酶途径、非核糖体多肽合成酶途径与聚酮合酶基因合成酶-非核糖体多肽合成酶混合代谢是大多数真菌毒素合成中的关键代谢途径。同时在真菌的萜类化合物代谢过程中，也能产生多种真菌毒素^[16]。用anti-SMASH对基因组进行次级代谢产物合成分析预测，基因组中共预测得到23个次级代谢产物基因簇，主要为聚酮合酶（polyketide synthase, PKS）基因簇、非核糖体多肽合成酶（non-ribosomal peptide synthase, NRPS）基因簇和萜烯类，具体预测结果如表4所示。在预测到的基因簇中，*pks*和*Nrps*基因簇共15个，占预测总基因簇的65.21%，*Tlpks*有7个，*Nrps*有6个，*Tlpks-Nrps*有2个，萜烯类合成途径基因簇有5个。但所有基因簇与已知次级代谢产物进行BLAST比对发现，*P. variotii* MTDF-01的基因组中PKS途径、NRPS途径与PKS-NRPS混合代谢途径、萜烯类途径中的基因簇并未匹配及注释到相似性高的基因信息，相似度最高的仅为33%。因此可认为*P. variotii* MTDF-01中不存在PKS、NRPS、PKS-NRPS混合代谢途径、萜烯类等途径合成代谢能力，*P. variotii* MTDF-01不具备合成该类物质的能力。

通过KEGG对萜类化合物合成途径分析时也发现，在*P. variotii* MTDF-01中注释到参与萜类化合物代谢途径

的基因仅有26个，其中萜类化合物合成途径注释到18个基因，主要为萜类物质骨架结构的真菌甲羟戊酸合成途径中注释到相对完整的代谢途径，共16个基因，但并未注释到特异性后修饰相关蛋白序列，因此，根据基因组预测信息可表明*P. variotii* MTDF-01不具备真菌毒素合成能力，是白酒酿造过程的安全菌株。

2.5 与*P. variotii* No.5比较基因组学分析

2.5.1 *P. variotii* No.5全基因组概况

目前已报道的关于*P. variotii*的全基因组测序的菌株为*P. variotii* No.5（NBRC109023）。*P. variotii* No.5是1株分离于土壤中的耐甲醛的真菌，能够在20 min内降解2%的甲醛^[38]。将*P. variotii* MTDF-01与*P. variotii* No.5菌株进行比较基因组学研究，以研究不同环境*P. variotii*基因组信息的差异，为研究不同来源*P. variotii*基因组框架、组成及蛋白差异提供理论基础及研究方向。两个菌株基本信息统计结果如表5所示。

表 5 *P. variotii* MTDF-01与*P. variotii* No.5基本特征比较分析
Table 5 Comparative analysis of general features of *P. variotii* MTDF-01 and *P. variotii* No.5

菌株名称	<i>P. variotii</i> MTDF-01	<i>P. variotii</i> No.5
基因组大小/bp	30 833 540	29 762 401
G+C含量/%	47.6	48.6
蛋白编码序列数量	8 815	8 877
重叠群数量	19	1 053

2.5.2 共线性分析

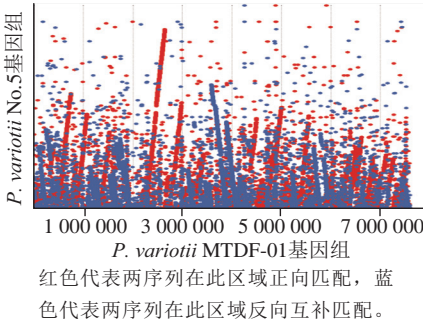


图 4 *P. variotii* MTDF-01与*P. variotii* No.5基因组序列线性结构比较
Fig. 4 Global alignment between *P. variotii* MTDF-01 and No.5

通过比较*P. variotii* MTDF-01与*P. variotii* No.5两个基因组的同源性，绘制共线性关系点阵图（图4）以观测结构性的差异。结果表明*P. variotii* MTDF-01与*P. variotii* No.5在基因组水平上存在较大差异。从图4可看出，两者基因组中有多条大片段能够比对上，两株菌基因组存在一定相似性，但是存在大量倒位、重排等现象。对单核苷酸多态性进行分析发现*P. variotii* MTDF-01与*P. variotii* No.5两个基因组共存在16 414处单核苷酸碱基突变，其中有525个碱基缺失，335个碱基插入，15 554个单碱基替换。

进一步对蛋白编码序列进行分析发现：*P. variotii* MTDF-01共预测CDS序列8 815个，*P. variotii* No.5共预测CDS序列8 877个，总匹配蛋白数（相似度大于50%）为7 743个，其中仅有26个基因完全匹配上，其余7 717个基因存在氨基酸层面的突变、插入或缺失，相似度在90%以上的基因有5 803个，占总匹配蛋白数的74.94%。两者在氨基酸水平上既有相似性又有各自的独特性。通过KEGG注释结果表明No.5存在大量环境信息处理相关基因，主要涉及信号转导。由于*P. variotii* MTDF-01和*P. variotii* No.5两株菌株来源环境差异较大，*P. variotii* No.5来源于土壤环境，而*P. variotii* MTDF-01在白酒酿造环境中经过高温高酸等环境胁迫条件长期驯化，因此这也可能导致两株菌在基因组及功能的差异。白酒酿造环境的长期驯化使*P. variotii* MTDF-01显示出独特性。

3 讨论与结论

Oka等^[38]将从土壤中分离得到的*P. variotii* No.5进行全基因组测序，并根据其具有降解甲醛的功能，对参与甲醛代谢的相关途径进行了分析，发现了7个相关蛋白。Radwan等^[39]从喷气燃料中分离得到*Paecilomyces* sp. BYSS01，预测发现共有334种真菌酶参与碳水化合物代谢，并且能够参与芳香烃和正烷烃的降解。但在白酒酿造过程中关于*P. variotii*的基因组研究鲜见报道。本研究通过对白酒酿造过程中分离得到*P. variotii* MTDF-01进行全基因组测序，拼接得到基因组总长度为30 833 540 bp，GC平均含量为47.46%。通过COG、GO、KEGG等数据库比较分析发现MTDF-01菌株基因主要涉及遗传信息加工、细胞转运及代谢过程，在代谢过程中主要参与碳水化合物、氨基酸、核苷酸、能量代谢等代谢通路，这为进一步了解*P. variotii*代谢机理提供了理论基础。基于KEGG和全基因组测序数据代谢网络注释发现，*P. variotii* MTDF-01具有苯乙醇、苯乙酸、苯甲酸、肉桂酸乙酯、乙酸、乳酸等代谢途径并具有合成多种有机酸的关键酶，此外，通过碳水化合物预测和次级代谢产物分析表明*P. variotii* MTDF-01含有多个淀粉和纤维素水解酶，具有降解淀粉和纤维素的潜力，不具备真菌毒素合成能力，是白酒酿造过程的安全菌株。对*P. variotii* MTDF-01与*P. variotii* No.5基因组进行比较基因组学分析发现，两者存在显著差异，存在翻转及异位等基因组重排现象，*P. variotii* MTDF-01具有独特性。

本研究利用全基因组测序对白酒酿造过程中分离得到的*P. variotii* MTDF-01进行分析，在*P. variotii* MTDF-01

中发现了参与酿酒原料代谢及风味物质合成代谢的基因及代谢通路，为解析*P. variotii*的酿造功能提供了新思路，从基因组水平对*P. variotii*代谢途径的挖掘，为深入了解*P. variotii*在白酒生产过程中的代谢机理提供了参考信息及研究方向，对以后*P. variotii*的相关研究具有重要意义。

参考文献：

- [1] 任聪, 杜海, 徐岩. 中国传统发酵食品微生物组研究进展[J]. 微生物学报, 2017, 57(6): 885-898. DOI:10.13343/j.cnki.wsxb.20170102.
- [2] XU Y Q, SUN B G, FAN G S, et al. The brewing process and microbial diversity of strong flavour Chinese spirits: a review[J]. Journal of the Institute of Brewing, 2017, 123(1): 5-12. DOI:10.1002/jib.404.
- [3] JIN G Y, ZHU Y, XU Y. Mystery behind Chinese liquor fermentation[J]. Trends in Food Science & Technology, 2017, 63: 18-28. DOI:10.1016/j.tifs.2017.02.016.
- [4] WU Q, CHEN B, XU Y. Regulating yeast flavor metabolism by controlling saccharification reaction rate in simultaneous saccharification and fermentation of Chinese Maotai-flavor liquor[J]. International Journal of Food Microbiology, 2015, 200: 39-46. DOI:10.1016/j.ijfoodmicro.2015.01.012.
- [5] 曾婷婷, 章肇敏, 吴生文. 根霉对特香型白酒风味的影响研究[J]. 中国酿造, 2012, 31(7): 143-145.
- [6] HUANG Y H, YI Z L, JIN Y L, et al. New microbial resource: microbial diversity, function and dynamics in Chinese liquor starter[J]. Scientific Reports, 2017, 7(1): 1-14. DOI:10.1038/s41598-017-14968-8.
- [7] 陈笔. 酱香型白酒酿造过程中霉菌群落结构以及霉菌与酵母相互作用的研究[D]. 无锡: 江南大学, 2014, 13-16.
- [8] CHEN B, WU Q, XU Y. Filamentous fungal diversity and community structure associated with the solid state fermentation of Chinese Maotai-flavor liquor[J]. International Journal of Food Microbiology, 2014, 179(6): 80-84. DOI:10.1016/j.ijfoodmicro.2014.03.011.
- [9] 杨帆, 林琳, 王和玉, 等. 拟青霉固态发酵代谢产物分析[J]. 酿酒科技, 2011(7): 41-43; 46. DOI:10.13746/j.njkj.2011.07.026.
- [10] 沈海月, 范文来, 徐岩. 白酒生产中高效降解糠醛微生物的筛选与鉴定[J]. 食品工业科技, 2010, 31(3): 207-209. DOI:10.13386/j.issn1002-0306.2010.03.073.
- [11] 彭素琴, 吴群, 徐岩, 等. 产酱香地衣芽孢杆菌CGMCC3963耐受特征及基于转录组学的耐受机制分析[J]. 微生物学通报, 2014, 41(12): 2395-2403. DOI:10.13344/j.microbiol.china.140172.
- [12] SONG Z W, DU H, ZHANG Y, et al. Unraveling core functional microbiota in traditional solid-state fermentation by high-throughput amplicons and metatranscriptomics sequencing[J]. Frontiers in Microbiology, 2017, 8: 1-14. DOI:10.3389/fmicb.2017.01294.
- [13] LIU J, WU Q, WANG P, et al. Synergistic effect in core microbiota associated with sulfur metabolism in spontaneous Chinese liquor fermentation[J]. Applied & Environmental Microbiology, 2017, 83(24): AEM.01475-17. DOI:10.1128/AEM.01475-17.
- [14] HUANG Y H, YI Z L, JIN Y L, et al. Metatranscriptomics reveals the functions and enzyme profiles of the microbial community in Chinese nong-flavor liquor starter[J]. Frontiers in Microbiology, 2017, 8: 1747. DOI:10.3389/fmicb.2017.01747.
- [15] WU Q, PENG S, YU Y, et al. Genome sequence of *Bacillus licheniformis* CGMCC3963, a stress-resistant strain isolated in a Chinese traditional solid-state liquor-making process[J]. Genome Announc, 2013, 1(1). DOI:10.1128/genomeA.00060-12.

- [16] 吴荣, 王栋, 徐岩, 等. 基于全基因组数据华根霉产真菌毒素分析[J]. 微生物学通报, 2013, 40(8): 1437-1447. DOI:10.13344/j.microbiol.china.2013.08.004.
- [17] SU Y, XU Y Q, XIN C H, et al. Genome sequence of *Thermoactinomyces daqu* H-18, a novel thermophilic species isolated from high-temperature *Daqu*[J]. Genome Announc, 2014, 3(1): 1394-1400. DOI:10.1128/genomeA.01394-14.
- [18] LI C R, WANG Y S, XIE G P, et al. Complete genome sequence of *Clostridium butyricum* JKY6D1 isolated from the pit mud of a Chinese flavor liquor-making factory[J]. Journal of Biotechnology, 2016, 220: 23-24. DOI:10.1016/j.jbiotec.2016.01.003.
- [19] NAKANO K, SHIROMA A, SHIMOJI M, et al. Advantages of genome sequencing by long-read sequencer using SMRT technology in medical area[J]. Human Cell, 2017, 30(3): 149-161. DOI:10.1007/s13577-017-0168-8.
- [20] CHAISSON M J, 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. DOI:10.1186/1471-2105-13-238.
- [21] MYERS E W, SUTTON G G, DELCHER A L, et al. A whole-genome assembly of *Drosophila*[J]. Science, 2000, 287: 2196-2204. DOI:10.1126/science.287.5461.2196.
- [22] CHIN C, ALEXANDER D H, MARKS P, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data[J]. Nature Methods, 2013, 10(6): 563-569. DOI:10.1038/nmeth.2474.
- [23] HAAS B J, SALZBERG S L, ZHU W, et al. Automated eukaryotic gene structure annotation using Evidence Modeler and the program to assemble spliced alignments[J]. Genome Biology, 2008, 9(1): R7. DOI:10.1186/gb-2008-9-1-r7.
- [24] STANKE M, KELLER O, GUNDUZ I, et al. AUGUSTUS: ab initio prediction of alternative transcripts[J]. Nucleic Acids Research, 2006, 34(Suppl 2): W435-W439. DOI:10.1093/nar/gkl200.
- [25] BROMBERG Y, ROST B. SNAP: predict effect of non-synonymous polymorphisms on function[J]. Nucleic Acids Research, 2007, 35(11): 3823-3835. DOI:10.1093/nar/gkm238.
- [26] BIRNEY E, CLAMP M, DURBIN R. GeneWise and GenomeWise[J]. Genome Research, 2004, 14(5): 988-995. DOI:10.1101/gr.1865504.
- [27] TATUSOV R L, NATALE D A, GARKAVTSEV I V, et al. The COG database: new developments in phylogenetic classification of proteins from complete genomes[J]. Nucleic Acids Research, 2001, 29(1): 22-28. DOI:10.1093/nar/29.1.22.
- [28] ASHBURNER M, BALL C A, BLAKE J A, et al. Gene Ontology: tool for the unification of biology[J]. Nature Genetics, 2000, 25(1): 25-29. DOI:10.1038/75556.
- [29] 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.
- [30] 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.
- [31] BOUTET E, LIEBERHERR D, TOGNOLLI M, et al. UniProtKB/Swiss-Prot, the manually annotated section of the UniProt knowledge base: how to use the entry view[J]. Methods in Molecular Biology, 2016, 1374: 23-54. DOI:10.1007/978-1-4939-3167-5_2.
- [32] YIN Y, MAO X, YANG J C, et al. dbCAN: a web resource for automated carbohydrate-active enzyme annotation[J]. Nucleic Acids Research, 2012, 40: 445-451. DOI:10.1093/nar/gks479.
- [33] WEBER T, BLIN K, DUDELA S, et al. antiSMASH 3.0: a comprehensive resource for the genome mining of biosynthetic gene clusters[J]. Nucleic Acids Research, 2015, 43: 237-243. DOI:10.1093/nar/gkv437.
- [34] 姜淑荣. 浅谈纤维素酶在酒类生产中的应用[J]. 中国酿造, 2008, 27(9): 12-15.
- [35] 李旭晖, 吴生文, 张志刚. 纤维素酶对大曲酒风味物质影响的探讨[J]. 中国酿造, 2011, 30(6): 80-83.
- [36] 钟霞, 赵盈盈, 张健, 等. 一株产纤维素酶曲霉菌白酒糟发酵培养基的优化[J]. 中国酿造, 2012, 31(3): 94-96.
- [37] 刘茂柯, 唐玉明, 熊洪, 等. 白酒酒酯纤维素降解菌的多样性分析及分离筛选[J]. 食品与发酵工业, 2018, 44(4): 35-41.
- [38] OKA T, EKINO K, FUKUDA K, et al. Draft genome sequence of the formaldehyde-resistant fungus *Byssoschlamys spectabilis* No.5 (Anamorph *Paecilomyces variotii* No.5) (NBRC109023)[J]. Genome Announcements, 2014, 2(1). DOI:10.1128/genomeA.01162-13.
- [39] RADWAN O, GUNASEKERA T S, RUIZ O N. Draft genome sequence of *Byssoschlamys* sp. isolate BYSS01, a filamentous fungus adapted to the fuel environment[J]. Genome Announcements, 2018, 6(10): e00164-18. DOI:10.1128/genomeA.00164-18.