

基因组功能注释分析米曲霉 ZA189 优势酿造性能的遗传基础

侯莎^{1,2}, 吴昌正², 潘力¹, 王斌¹

(1. 华南理工大学生物科学与工程学院, 广东广州 510006) (2. 广东海天创新技术有限公司, 广东佛山 528000)

摘要: 米曲霉 ZA189 是以米曲霉沪酿 3.042 为出发菌株经传统诱变获得的, 具有中性蛋白酶、淀粉酶、谷氨酰胺酶活力高的特性。为深入研究米曲霉 ZA189 的遗传背景, 评估其作为酱油酿造菌株的性能, 本研究利用新一代纳米孔测序技术完成了米曲霉 ZA189 基因组的测序、组装和功能注释。获得的基因组全长 36.89 Mb, 包含 16 条 Scaffold, 测序深度达到 132.05×。基因注释表明, 米曲霉 ZA189 基因组包含 263 个 tRNA 基因和 72 个 rRNA 基因, 这是米曲霉高效表达蛋白的遗传基础。KOG 和 KEGG 注释表明米曲霉 ZA189 具有强大的能量合成、蛋白质合成及次级代谢产物合成能力, 这是其在发酵工业广泛应用并产生多种风味功能物质的遗传基础。碳水化合物活性酶 CAZy 注释和蛋白酶注释表明, 米曲霉 ZA189 基因组中包含多达 330 个糖苷水解酶 (GH) 和 30 个蛋白酶基因, 这是米曲霉 ZA189 作为酱油酿造菌株降解大豆蛋白和碳水化合物的关键性能指标。通过本研究对酱油酿造菌株米曲霉 ZA189 的基因组有了更加深入的理解, 通过对蛋白表达系统基因、次级代谢基因、碳水化合物活性酶及蛋白酶的功能注释, 为指导其在酱油酿造中的应用提供了理论支持。

关键词: 米曲霉 ZA189; 基因组; 纳米孔测序技术; 碳水化合物活性酶 CAZy; 蛋白酶

文章编号: 1673-9078(2021)06-63-71

DOI: 10.13982/j.mfst.1673-9078.2021.6.0673

Genome Functional Annotation Analysis of the Genetic Basis for the Superior Brewing Performance of *Aspergillus oryzae* ZA189

HOU Sha^{1,2}, WU Chang-zheng², PAN Li¹, WANG Bin¹

(1. School of Biology and Biological Engineering, South China University of Technology, Guangzhou 510006, China)

(2. Guangdong Haitian Innovation TECH Co. Ltd., Foshan 528500, China)

Abstract: *Aspergillus oryzae* ZA189 was obtained via traditional mutagenesis using *A. oryzae* Huniang 3042 as the starting strain, and possessed high activities of neutral protease, amylase and glutaminase. To further investigate the genetic background of *A. oryzae* ZA189 and evaluate its performance as the strain for soy sauce brewing, the sequencing, assembly and functional annotation of the genome of *A. oryzae* ZA189 was achieved using the new generation nanopore sequencing technology in this study. The length of the obtained genome was 36.89 Mb, having 16 Scaffolds and sequencing depth of 132.05×. Gene annotation showed that the *A. oryzae* ZA189 genome contained 263 tRNA genes and 72 rRNA genes, which was the genetic basis for high-efficiency *A. oryzae* protein expression. KOG and KEGG annotations illustrated that *A. oryzae* ZA189 possessed strong capabilities of energy synthesis, protein synthesis and secondary metabolite synthesis, which formed the genetic foundation for its wide application in the fermentation industry and the production of multiple flavor functional substances. Annotation of carbohydrate-active enzymes (CAZys) and proteases showed that *A. oryzae* ZA189 genome contained as many as 330 glycoside hydrolase

引文格式:

侯莎,吴昌正,潘力,等.基因组功能注释分析米曲霉 ZA189 优势酿造性能的遗传基础[J].现代食品科技,2021,37(6):63-71

HOU Sha, WU Chang-zheng, PAN Li, et al. Genome functional annotation analysis of the genetic basis for the superior brewing performance of *Aspergillus oryzae* ZA189 [J]. Modern Food Science and Technology, 2021, 37(6): 63-71

收稿日期: 2020-07-17

基金项目: 广东省重点研发计划项目 (2020B020226007); 广州市科技计划项目 (202002030207); 广东省调味食品生物发酵先进技术企业重点实验室开放基金项目 (2017B030302002); 中央高校基本科研业务费项目 (2019MS092)

作者简介: 侯莎 (1986-), 女, 博士后, 研究方向: 微生物学与发酵工程

通讯作者: 王斌 (1981-), 男, 博士, 副教授, 研究方向: 工业生物技术及发酵工程

(GH) genes and 30 protease genes, which is the key performance indicator of *A. oryzae* ZA189 as a soy sauce brewing strain in degrading soybean proteins and carbohydrates. Through this research, a more in-depth understanding of the genome of the soy sauce brewing strain, *A. oryzae* ZA189. Functional annotation of protein expression system genes, secondary metabolism genes, and CAZys and proteases provides a theoretical support for guiding its application in soy sauce brewing.

Key words: *A. oryzae* ZA189; genome; nanopore sequencing technology; carbohydrate-active enzymes (CAZys); protease

目前我国酱油生产绝大多数采用的是上海市酿造科学研究所 1958 年研发获得的米曲霉沪酿 3.042, 其初始菌株来源于福建省永春县酱园, 中科院编号为 AS3.863。米曲霉沪酿 3.042 是以 AS3.863 为出发菌经过紫外诱变获得的。该菌具有分子孢子量大、繁殖速度快、抗杂菌生长能力强、易制曲、培养粗放、蛋白酶活力较高及综合酶系分泌丰富等特点。根据统计结果显示, 目前我国 90% 的酱油企业仍然沿用此菌种。第一株完成全基因组测序的米曲霉菌株是 RIB40, 由日本学者于 2005 年完成, 基因组大小为 37 Mb, 共 8 条染色体, 包含 12074 个基因^[1]。截至目前共有 10 株米曲霉的基因组被测序^[2]。沪酿 3.042 的基因组测序工作, 由天津科技大学曹小红课题组在 2012 年完成, 基因组全长 36.6 Mb, 注释基因 11639 个, 与 RIB40 比较发现沪酿 3.042 基因组中参与转运系统、酯类形成和氨基酸代谢等功能的数百个基因存在差异, 这些差异是两种菌株发酵风味差异的遗传基础^[3]。

Oxford Nanopore Technologies 技术 (ONT 测序) 是新一代基于纳米孔的单分子实时电信号测序技术^[4], 目前主要包括三种平台 (MinION、GridIONx5、PromethION), 其测序原理是相同的: DNA/RNA 链在马达蛋白的带领下与镶嵌在生物膜上的纳米孔蛋白结合并解螺旋, 在生物膜两侧电压差的作用下 DNA/RNA 链以一定的速率通过纳米孔通道蛋白, 由于 DNA/RNA 碱基化学性质的差异, 当其通过纳米孔通道时将引起电化学信号的变化^[5], 通过对电信号的检测及转换, 可获得相应碱基类型的信息, 从而完成测序^[6]。其特点是测序读长长 (超过 150 kb), 测序速度快, 测序数据实时监控。基于该方法已实现多种微生物基因组的测序工作, 发表相关文章 300 余篇。

米曲霉 ZA189 菌株是以沪酿 3.042 为出发菌株, 采用常压室温等离子体 (Atmospheric Room Temperature Plasma, ARTP) 诱变获得的突变菌株。与出发菌株相比, ZA189 制曲获得的成曲中, 中性蛋白酶、淀粉酶及谷氨酰胺酶活力均有所提高, 发酵原油中氨基酸态氮、全氮、还原糖和谷氨酸均有所提高, 而且更适用于酱油现代化圆盘工艺和大罐生产, 所得原油味道更浓郁, 色泽更鲜明, 具备酱香浓郁, 鲜甜突出, 滋味醇厚持久。

本研究利用 ONT 测序技术对米曲霉 ZA189 进行全基因组测序, 对其基因组序列进行组装、注释, 并进行功能注释和分析, 在此基础上分析该菌的胞外水解活力、安全性等特征, 为指导其在酱油酿造中的应用提供数据支持。

1 材料与方法

1.1 菌株及培养基

菌株: ZA189, 由沪酿 3.042 经诱变获得。

培养基 DPY: 2% (W/V) 葡萄糖, 1% (W/V) 蛋白胨, 酵母粉和 KH_2PO_4 各 0.50% (W/V), 0.05% (W/V) $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 2% 琼脂 (W/V)、pH 5.0。

1.2 试剂及仪器设备

1.2.1 试剂

苯酚:氯仿:异戊醇 (25:24:1) (索莱宝); Tris-base (sigma); SDS (sigma); 琼脂糖 (sigma); EDTA (国药试剂); SQK-LSK109 连接试剂盒 (Oxford Nanopore Technologies)。

溶菌缓冲液: Tris-HCl 50 mM, EDTA 50 mM, SDS 3% (W/V), pH=8.0, 高温高压灭菌后保存使用。

1.2.2 主要仪器设备

生化培养箱, 上海一恒科学仪器有限公司; 超净工作台, 苏州净化设备有限公司; 小型离心机, 德国 Eppendorf 公司; 低温冷冻离心机, 德国 Eppendorf 公司; 水平电泳系统, 美国 BIO-RAD 公司; 凝胶成像系统, 美国 BIO-RAD 公司等。

1.3 实验方法

1.3.1 米曲霉培养

取冰箱冻存的 ZA189 孢子 (约 $10^7/\text{mL}$) 500 μL 到加有玻璃纸隔离的 DPY 平板上, 30 $^\circ\text{C}$ 恒温培养 24 h 至长出丰富菌丝。

1.3.2 基因组 DNA 的提取

用灭菌的药匙刮取菌丝, 将菌丝转移至研钵 (提前液氮预冷) 中, 液氮研磨成粉末 (需研磨充分)。称取适量粉末 (3 g) 加入到 65 $^\circ\text{C}$ 预热的溶菌缓冲液 (1.5 g/10 mL) 中。65 $^\circ\text{C}$ 孵育 0.5 h, 每隔 5 min 用 5 mL

去尖枪头轻轻混匀。加入等体积的 DNA 提取液, 轻轻颠倒混匀, 12000×g, 4 °C 离心 10 min。取上清(切勿吸到中间层白色沉淀), 加入等体积氯仿, 轻轻颠倒混匀, 4 °C 12000×g 离心 10 min。取上清, 加入 2.5 倍体积预冷的无水乙醇, -20 °C 沉淀 DNA 2 h 左右。12000×g, 4 °C 离心 15 min, 去除上清, 用 70% (V/V) 无水乙醇洗涤沉淀 2 次。适量水溶解 DNA, 加入 1% 体积的 RNA 酶处理 30 min, 通过凝胶电泳检测提取的基因组质量。

1.3.3 基因组 DNA 质量检测、文库构建及测序

实验流程按照 Oxford Nanopore Technologies (ONT) 公司提供的标准 protocol 执行 (<https://nanoporetech.com/>), 包括样品质量检测、文库构建、文库质量检测和文库测序等流程, 文库构建包括如下步骤: (1) 提取高质量基因组 DNA, 利用 Nanodrop、0.35% 琼脂糖凝胶电泳进行纯度、浓度和完整性质检; (2) BluePippin 全自动核酸回收系统回收大片段 DNA; (3) 文库构建及测序 (SQK-LSK109 连接试剂盒): ① 大片段 DNA 的获得: 提取高质量的 DNA, 样本质检合格后, 对基因组 DNA 进行随机打断; 利用磁珠富集、纯化大片段 DNA, 将大片段进行切胶回收; ② 片段修复: 对筛选好的大片段 DNA 进行损伤修复、末端修复及 3' 末端加 A, 并对反应产物进行纯化; ③ 连接反应: 片段修复产物进行测序相关接头的连接及纯化, 得到最终的上机测序文库; ④ 定量检测: 用 Qubit 精确地对建好的 DNA 文库进行定量检测; ⑤ 文库上机: 取一定量的 DNA 文库, 与上机相关试剂混合后加入 Flow cell 中, 利用 PromethION 测序仪进行实时单分子测序, 获得原始测序数据。

1.3.4 基因组测序数据的组装

测序获得的原始序列, 需要进行质控、组装、注释等, 才能获得生物学相关信息, 分析生物学功能。

(1) 原始数据质控: 测序原始数据格式为包含所有原始测序信号的二进制 fast5 格式, 通过 MinKNOW 软件包中的 Albacore 软件进行 base calling 后会将 fast5 格式数据转换为 fastq 格式, 经进一步过滤接头、低质量及短片段 (长度 < 2000 bp) 的 reads 后, 得到总的数据集。

(2) 基因组组装: 使用 Canu v1.5^[7] 软件对过滤后 subreads 进行纠错, 然后使用 NECAT 软件对纠错后的 subreads 进行组装, 最后采用 Pilon v1.22^[8] 软件利用二代数据进一步对组装基因组进行纠错, 得到最终准确度更高的基因组。

1.3.5 数据处理

1.3.5.1 基因预测及功能注释

(1) 基因预测: 蛋白编码基因预测主要采用从头预测、基于同源蛋白预测, 然后对 2 种预测结果进行整合。使用 Genscan v1.0、Augustus v2.4、GlimmerHMM v3.0.4、GeneID v1.4、SNAP (version2006-07-28) 进行从头预测; 使用 GeMoMa v1.3.1 进行基于同源蛋白的预测。最后利用 EVM v1.1.1 整合上述 2 种方法得到的预测结果, 并用 PASA v2.0.2 进行修饰。使用软件 tRNAscan-SE v1.3.1 预测基因组中的 tRNA, 使用软件 Infernal v1.1 基于 Rfam v12.0 数据库预测基因组中的 rRNA。

(2) 基因功能注释: 利用预测得到的基因序列与 KOG、KEGG、Swiss-Prot (2015_01)、TrEMBL、Nr 等功能数据库做 BLAST (v2.2.29) 比对, 得到基因功能注释结果并进行功能富集分析。利用软件 hmmer v3.0 基于 Pfam (27.0) 数据库进行 Pfam 功能注释。利用软件 hmmer 基于碳水化合物相关酶数据库 (CAZy^[9]) (<http://www.cazy.org/>) 进行碳水化合物酶类基因的功能注释。

1.3.5.2 胞外分泌蛋白及蛋白酶分析

(1) 信号肽预测: 使用软件 SignalP 4.0 对所有的预测到的基因的蛋白序列进行分析, 找出含有信号肽的蛋白。

(2) 跨膜蛋白预测: 使用软件 tmhmm v2.0 对所有预测基因的蛋白序列进行分析, 找出含有跨膜螺旋的蛋白, 即为跨膜蛋白。

(3) 分泌蛋白预测: 在上述预测的含有信号肽的蛋白中去掉含有跨膜螺旋的蛋白, 剩余的蛋白即为分泌蛋白。

(4) 蛋白酶分析: 根据基因功能注释的结果, 从分泌蛋白中筛选出蛋白酶基因, 对其进行功能聚类分析。

2 结果与讨论

2.1 米曲霉 ZA189 测序数据质控与基因组组装

Nanopore 测序的下机数据的原始数据经格式转换、过滤接头、过滤低质量及短片段后, 得到 4,872,627,875 bp 高质量测序数据集, 测序深度为 132.05×。将 reads 按照不同梯度长度分布统计 (图 1), 其中 10 kb~20 kb 和 20 kb~30 kb 之间分布的 reads 最多, 分别为 39.43% 和 25.12%。

利用过滤后的高质量测序数据, 使用 NECAT 软

件进行基因组组装,最后采用 Pilon 软件利用二代数据进一步对组装基因组进行纠错,得到高准确度的基因组,组装得到完整基因组总长 36.89 Mb, Scaffold 数目为 16 条(米曲霉为 8 条染色体), scaffold N50 长度为 2.37 Mb。而出发菌株沪酿 3.042 测序全长 36.6 Mb^[3],米曲霉 ZA189 基因组相比出发菌株增加的序列(0.29 Mb)可能是序列插入或者片段重复导致的。

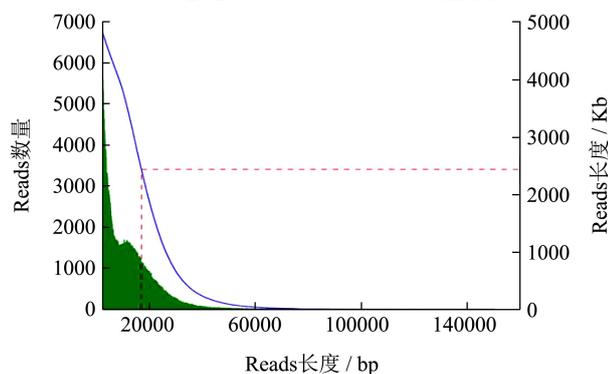


图1 米曲霉 ZA189 基因组测序 reads 长度分布统计

Fig.1 Length distribution of sequencing reads in the genome of *A. oryzae* ZA189

注: Reads 数量: 绿色柱状图; Reads 长度: 对应于蓝色实线; 红色虚线表示 reads 的 N50 长度。Reads N50: 把 reads 从长到短进行排列,然后相加,当恰好加到 1 M 的 50%,也就是 500 k 的时候,那一条 reads 的长度叫做 Reads N50。

2.2 米曲霉 ZA189 基因组中的基因注释

表 1 米曲霉 ZA189 蛋白编码基因的预测

Table 1 Prediction of protein-coding genes in *A. oryzae* ZA189

基因预测方法	预测软件	同源物种	基因数目
从头预测	Augustus		10,322
	Genscan		8,786
	GeneID		14,230
	GlimmerHMM		12,635
	SNAP		12,671
同源蛋白预测		<i>Aspergillus fumigatus</i>	9,595
	GeMoMa	<i>Aspergillus nidulans</i>	9,357
		<i>Aspergillus niger</i>	10,412
整合	EVM		12,468

表 2 基因信息统计

Table 2 Statistics of gene information

基因序列总长度/bp	基因平均长度/bp	外显子总长度/bp	内含子总长度/bp
24,760,823	1,985.95	22,337,256	2,423,567

蛋白编码基因预测主要采用从头预测和基于同源蛋白预测,然后对 2 种预测结果进行整合,共预测得

到 12,468 个米曲霉基因(表 1)。ZA189 基因组中编码蛋白基因的长度为 24,760,823 bp,平均长度为 1,986 bp,占基因组长度 67.10%,其中外显子区总长度占 60.54%,内含子区长度占 6.57%(表 2)。出发菌株沪酿 3.042 则预测到 11379 个基因,外显子区总长度为 44%^[3],基因数量和外显子长度比例的差异,一方面是测序技术和分析手段的进步引起的,另一方面也体现了 ZA189 优势酿造性能的分子基础。

非编码 RNA 即不编码蛋白质的 RNA,包括 microRNA、rRNA 和 tRNA 等多种已知功能的 RNA,针对非编码 RNA 的结构特点,采用不同的策略预测不同的非编码 RNA。在米曲霉 ZA189 基因组中,共预测得到 263 个 tRNA、72 个 rRNA。而出发菌株沪酿 3.042 则预测到 243 个 tRNA^[3],丰富的蛋白质翻译相关 RNA 基因的存在,表明 ZA189 具有大量表达蛋白质的遗传基础。

2.3 米曲霉 ZA189 基因组的功能注释

ZA189 相比出发菌株沪酿 3.042 酶活力更高,根据项目组前期研究,ZA189 应用于酱油发酵中(表 3 所示),制曲阶段孢子数和原料消耗率显著低于出发菌株沪酿 3.042 ($p < 0.05$),中性蛋白酶活力和淀粉酶活力分别达到 3 210.21 U/g 和 480.29 U/g,谷氨酰胺酶活力达到 3.48 U/g,较出发菌株提高 38%;发酵原油的氨基酸态氮含量为 1.16 g/100 mL、全氮含量为 1.94 g/100 mL、还原糖含量为 11.05 g/100 mL、谷氨酸含量为 13.69 g/kg,均高于出发菌株沪酿 3.042^[10]。

对米曲霉 ZA189 基因组注释的基因进行 KOG 功能注释(图 2),其中以下功能类群富集了大量功能基因: C-能量生产与转换(Energy production and conversion)、E-氨基酸的运输与代谢(Amino acid transport and metabolism)、G-糖运输与代谢(Carbohydrate transport and metabolism)、I-脂肪运输与代谢(Lipid transport and metabolism)、O-翻译后修饰(Posttranslational modification, protein turnover, chaperones)、Q-次级代谢(Secondary metabolisms biosynthesis, transport and catabolism)、T-信号传导(Signal transduction mechanisms)、U-胞内囊泡转运及分泌(Intracellular trafficking, secretion, and vesicular transport)。类群 C、E、G、I、O、U 富集大量的功能基因,表明米曲霉具有强大的能量合成能力、蛋白质合成及转运的能力,这是其在发酵工业广泛应用的遗传基础^[11]。类群 Q 富集大量基因,表明米曲霉具有较强的次级代谢产物合成能力,这为米曲霉产生多种多样的风味功能物质提供了可能。

表3 ZA189 发酵性能分析^[10]

Table 3 Fermentation feature analysis of ZA189

成曲指标					
菌种名称	孢子数/($\times 10^8$ CFU/g)	原料消耗率/%	中性蛋白酶活力/(U/g)	淀粉酶活力/(U/g)	谷氨酰胺酶活力/(U/g)
As3.951	2.54 \pm 0.03	6.81 \pm 0.92	2 568.00 \pm 193.32	417.30 \pm 12.12	2.54 \pm 0.11
ZA189	2.17 \pm 0.07	6.02 \pm 0.55	3 210.21 \pm 370.20	480.29 \pm 52.01	3.48 \pm 0.23

原油指标					
菌种名称	氨基酸态氮/(g/100 mL)	全氮/(g/100 mL)	全氮收得率/%	还原糖/(g/100 mL)	谷氨酸/(g/kg)
As3.951	1.09 \pm 0.12	1.71 \pm 0.32	84.04 \pm 7.21	8.62 \pm 2.14	11.13 \pm 1.03
ZA189	1.16 \pm 0.09	1.94 \pm 0.29	91.97 \pm 2.27	11.05 \pm 2.10	13.69 \pm 1.49

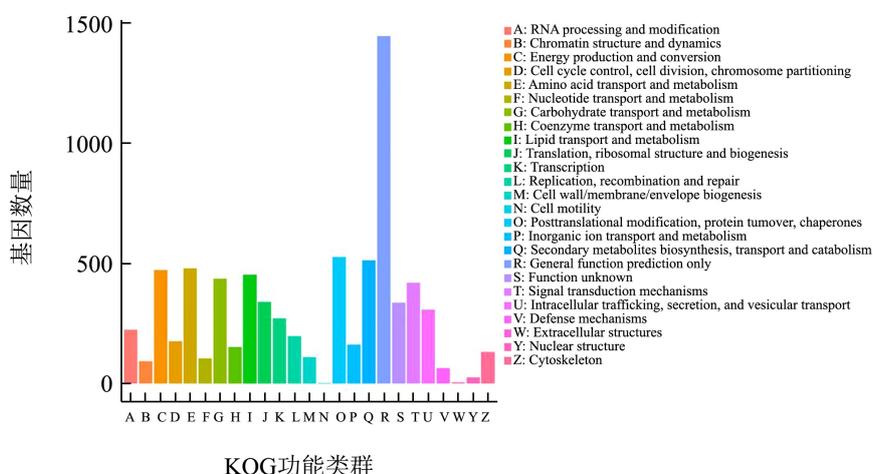


图2 米曲霉 ZA189 基因的 KOG 功能注释分析

Fig.2 KOG functional annotation of genes in *A. oryzae* ZA189 genome

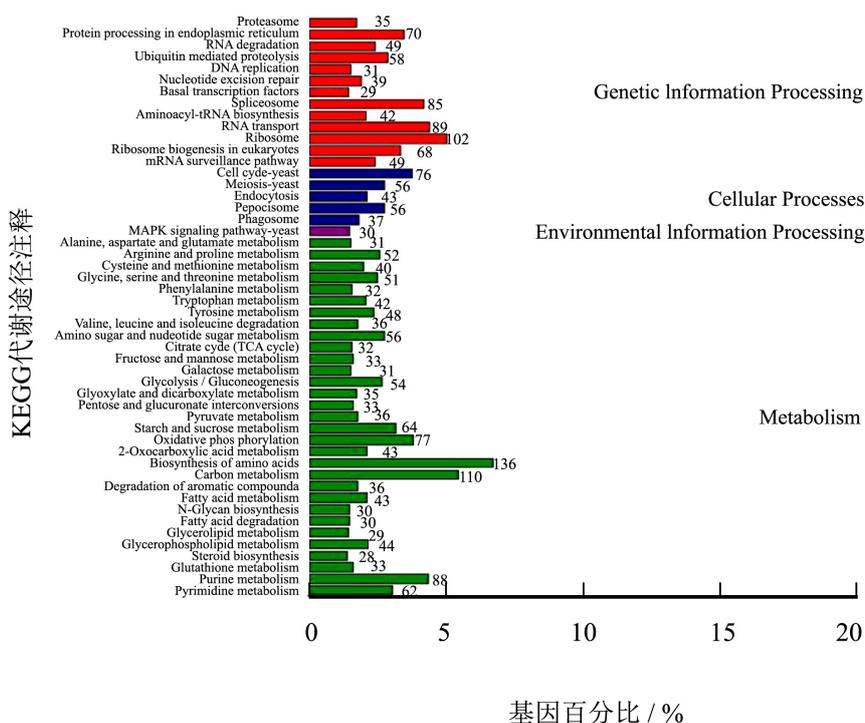


图3 米曲霉 ZA189 基因的 KEGG 代谢途径注释

Fig.3 KEGG pathway analysis of annotated genes in the genome of *A. oryzae* ZA189

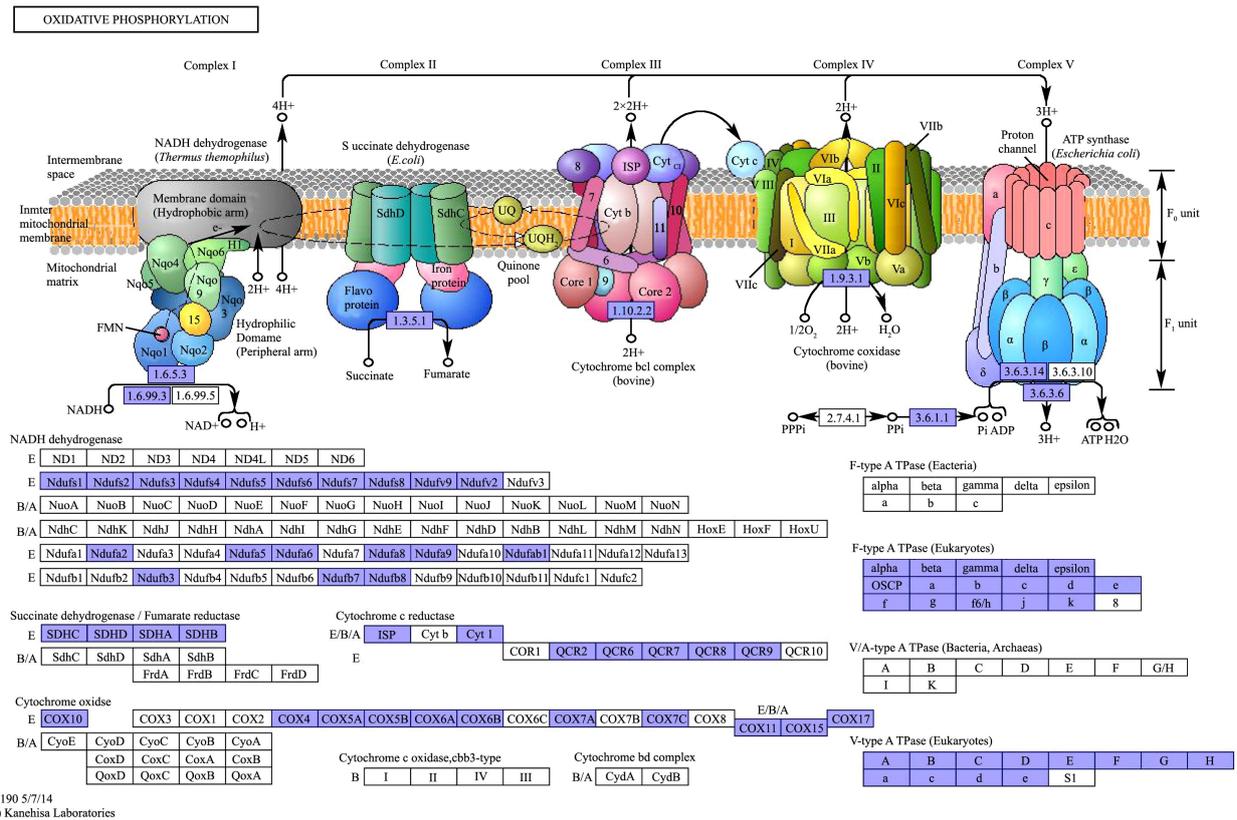


图 4 氧化磷酸化途径上富集的米曲霉 ZA189 基因

Fig.4 Annotated genes of *A. oryzae* ZA189 enriched in oxidative phosphorylation

注：紫色方块表示米曲霉 ZA189 基因组中有对应的氧化磷酸化基因的富集。

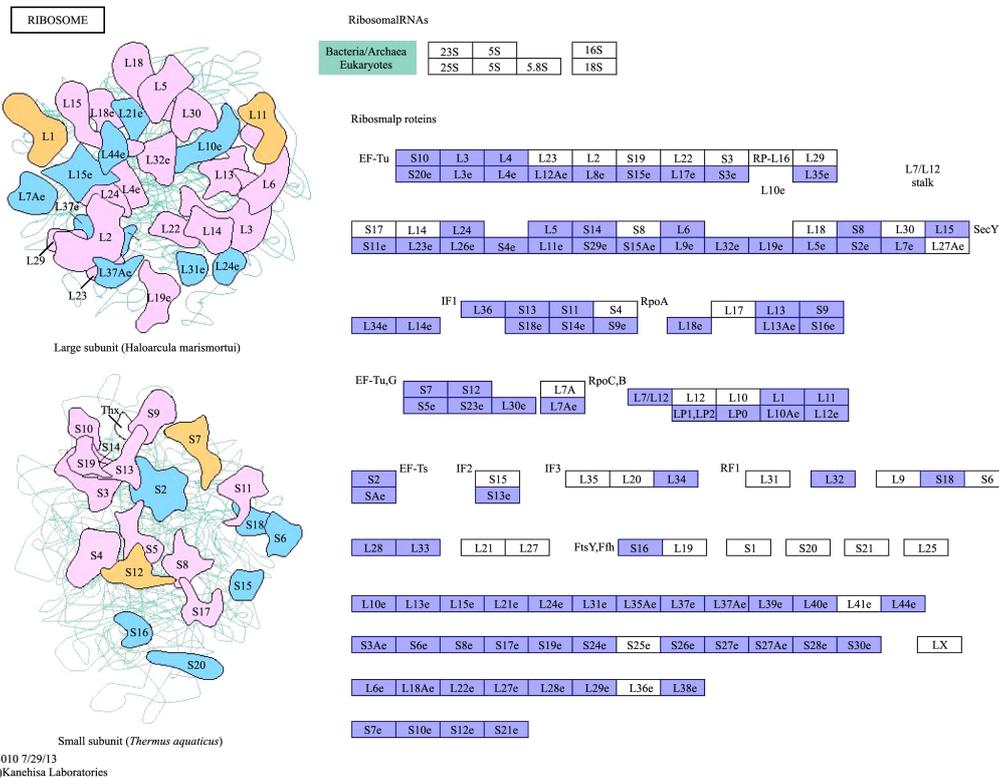


图 5 核糖体上富集的米曲霉 ZA189 基因

Fig.5 Annotated genes of *A. oryzae* ZA189 enriched in ribosome

注：紫色方块表示米曲霉 ZA189 基因组中有对应的核糖体相关基因的富集。

表 4 米曲霉 ZA189 基因组中的蛋白酶

Table 4 The proteases annotated in the genome of *A. oryzae* ZA189

米曲霉 ZA189 基因	蛋白酶注释
EVM0008810.1	proteinase [<i>Aspergillus oryzae</i> RIB40]
EVM0001047.1	vacuolar protease A [<i>Aspergillus oryzae</i> RIB40]
EVM0006563.1	peptidase [<i>Aspergillus oryzae</i> RIB40]
EVM0005139.1	aspartic-type endopeptidase ctsD [<i>Aspergillus oryzae</i> RIB40]
EVM0005811.1	dipeptidyl peptidase 4 [<i>Aspergillus oryzae</i> RIB40]
EVM0003410.1	alkaline serine protease AorO, putative [<i>Aspergillus oryzae</i> 3.042]
EVM0012091.1	aspergillopepsin A-like aspartic endopeptidase [<i>Aspergillus oryzae</i> RIB40]
EVM0011875.1	tripeptidyl-peptidase sed2 [<i>Aspergillus oryzae</i> RIB40]
EVM0007102.1	leucine aminopeptidase 1 [<i>Aspergillus oryzae</i> RIB40]
EVM0010937.1	aspartic-type endopeptidase [<i>Aspergillus oryzae</i> RIB40]
EVM0006663.1	elastinolytic metalloproteinase Mep [<i>Aspergillus flavus</i> NRRL3357]
EVM0006944.1	carboxypeptidase S1 [<i>Aspergillus oryzae</i> RIB40]
EVM0004260.1	neutral protease 2 precursor [<i>Aspergillus oryzae</i> RIB40]
EVM0004981.1	carboxypeptidase S1 [<i>Aspergillus oryzae</i> RIB40]
EVM0002600.1	aspartic-type endopeptidase opsB [<i>Aspergillus oryzae</i> RIB40]
EVM0000761.1	metalloproteinase [<i>Aspergillus oryzae</i> RIB40]
EVM0010518.1	X-Pro dipeptidyl-peptidase (S15 family) protein [<i>Aspergillus oryzae</i> RIB40]
EVM0010058.1	extracellular elastinolytic metalloproteinase, putative [<i>Aspergillus oryzae</i> 3.042]
EVM0008970.1	metalloproteinase [<i>Aspergillus oryzae</i> RIB40]
EVM0000267.1	alkaline D-peptidase, putative [<i>Aspergillus flavus</i> NRRL3357]
EVM0008297.1	subtilisin-like serine protease pepC [<i>Aspergillus oryzae</i> RIB40]
EVM0010379.2	peptidase S41 family protein [<i>Aspergillus oryzae</i> RIB40]
EVM0004697.1	neutral protease II [<i>Aspergillus oryzae</i>]
EVM0004205.1	serine-type carboxypeptidase F [<i>Aspergillus oryzae</i> RIB40]
EVM0002638.1	aspartic-type endopeptidase, putative [<i>Aspergillus oryzae</i> 3.042]
EVM0003663.1	serine peptidase, putative [<i>Aspergillus flavus</i> NRRL3357]
EVM0006011.1	aspartyl protease, putative [<i>Aspergillus oryzae</i> 3.042]
EVM0000010.1	carboxypeptidase CpyA/Prcl, putative [<i>Aspergillus flavus</i> NRRL3357]
EVM0007930.1	carboxypeptidase C [<i>Aspergillus oryzae</i> 3.042]
EVM0004940.1	carboxypeptidase S1 [<i>Aspergillus oryzae</i> RIB40]

对米曲霉 ZA189 基因组注释的基因进行 KEGG 途径分析 (图 3), 以下途径富集了大量功能基因: 内质网中的蛋白处理 (Protein processing in endoplasmic reticulum)、剪接体 (spliceosome)、RNA 转运 (RNA transport)、核糖体 (Ribosome)、氧化磷酸化 (Oxidative phosphorylation)、氨基酸生物合成 (Biosynthesis of amino acids)、碳代谢 (Carbon metabolism)、嘌呤代谢 (Purine metabolism)。图 4、图 5 反映了氧化磷酸化、核糖体上的基因富集情况。这与 KOG 功能富集分析的结果是一致的, 说明米曲霉在能量代谢、蛋白质合成、碳代谢 (与生长相关) 方面具有遗传优势。此前关于米曲霉酿造模式菌株 RIB40 转录组研究的结

果表明, 与能量代谢、蛋白质合成等相关的基因在固体发酵条件下具有高的转录表达水平^[12], 也说明米曲霉在基因组上与能量代谢、蛋白质合成相关优势也体现在实际的发酵应用中。

2.4 米曲霉 ZA189 的碳水化合物活性酶

碳水化合物活性酶 (Carbohydrate-Active enZymes, CAZy), 主要包含与糖苷键降解、修饰及生成相关的酶类家族。碳水化合物活性酶数据库, 主要包含 5 大分类: 糖苷水解酶 (Glycoside Hydrolases, GHs)、糖基转移酶 (Glycosyl Transferases, GTs)、多糖裂解酶 (Polysaccharide Lyases, PLs)、碳水化合物

酯酶 (Carbohydrate Esterases, CEs)、辅助活性酶 (Auxiliary Activities, AAs)。此外, 该数据库还包含与碳水化合物结合相关的酶 (Carbohydrate-Binding Modules, CBMs)。将米曲霉 ZA189 基因组预测基因蛋白序列, 利用软件 hmmer 基于碳水化合物相关酶数据库 (CAZy) 进行碳水化合物酶类基因的功能注释 (图 6), 结果显示米曲霉 ZA189 基因组中包含多达 330 个糖苷水解酶 (GH) 基因, 占 CAZy 酶总数的 42.69%, 这为米曲霉糖苷水解酶的开发利用提供了基础数据, 这也与 ZA189 在酱油成曲阶段高淀粉酶活特性密切相关。例如, Gomi K 等从米曲霉中制备了分泌表达的糖苷水解酶 GH3, 并对其进行了酶学性质研究^[13], Tang, C.D. 等人的研究表明米曲霉的糖苷水解酶可以用于魔芋粉的水解以及作为饲料添加剂^[14]。

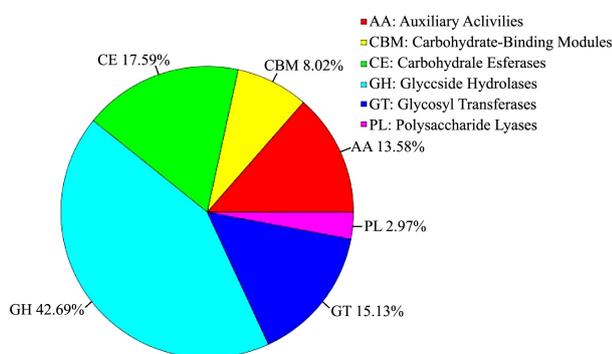


图 6 米曲霉 ZA189 基因组的碳水化合物活性酶 (CAZys)

Fig.6 Carbohydrate-Active enZymes (CAZys) annotated in *A. oryzae* ZA189 genome

2.5 米曲霉 ZA189 基因组中的蛋白酶的功能分析

利用 SignalP 4.0 软件预测米曲霉 ZA189 基因组注释的基因, 共获得 1224 个具有信号肽的基因, 利用 tmhmm v2.0 软件预测跨膜蛋白, 在米曲霉 ZA189 基因组中共找到 2668 个跨膜蛋白基因。从具有信号肽的 1224 个基因中去掉包含的跨膜蛋白基因, 剩余的基因即为分泌蛋白基因, 共 888 个。根据 Nr 基因功能注释的结果, 确定了分泌蛋白中的 30 个蛋白酶基因 (表 4), 包括 2 个中性蛋白酶、2 个碱性蛋白酶, 还有氨肽酶、羧肽酶、内肽酶等。若以活性中心来划分, 包括 2 个金属蛋白酶、4 个丝氨酸蛋白酶、5 个天门冬氨酸蛋白酶四大类。在沪酿 3.042 和采用 N+ 离子注入方式诱变获得的菌株 100-8 中, 分析到 25 种蛋白酶基因^[15]。因此, 可看出 ZA189 菌株具有更多蛋白酶基因, 使其具有高水平分泌表达大量蛋白酶的潜力, 这是由其制备的成曲中蛋白酶活力提升的原因, 也是其

作为酱油酿造菌株降解大豆蛋白的关键性能指标^[16]。

3 结论

本研究利用新一代纳米孔测序技术完成了米曲霉 ZA189 基因组的测序、组装和功能注释。获得的基因组全长 36.89 Mb, 包含 16 条 Scaffold, 测序深度达到 132.05 \times , 所得基因组在完整性和质量上达到米曲霉模式菌株 RIB40 的基因组质量水平。基因注释表明, 米曲霉 ZA189 基因组包含丰富的 tRNA 和 rRNA 基因, 这是米曲霉高效表达蛋白的遗传基础。KOG 和 KEGG 注释表明米曲霉 ZA189 具有强大的能量合成、蛋白质合成及次级代谢产物合成能力, 这是其在发酵工业广泛应用并产生多种风味功能物质的遗传基础。碳水化合物活性酶 CAZy 注释和蛋白酶注释表明, 米曲霉 ZA189 基因组中包含多达 330 个糖苷水解酶 (GH) 和 30 个蛋白酶基因, 这是米曲霉 ZA189 作为酱油酿造菌株高蛋白酶活、淀粉酶活和谷氨酰胺酶活的分子基础。通过本研究对酱油酿造菌株米曲霉 ZA189 的基因组及其功能有了更加深入的理解, 为指导其在酱油酿造中的应用提供了理论支持。

参考文献

- [1] Machida M, Asai K, Sano M, et al. Genome sequencing and analysis of *Aspergillus oryzae* [J]. Nature, 2005, 438(7071): 1157-1161
- [2] 吴琴琴, 孙敏, 陈雨, 等. 米曲霉功能基因组研究策略和进展 [J]. 生物技术通报, 2019, 35(8): 186-192
WU Qin-qin, SUN Min, CHEN Yu, et al. Strategies and advances in functional genomics of *Aspergillus oryzae* [J]. Biotechnology Bulletin, 2019, 35(8): 186-192
- [3] Zhao G, Yao Y, Qi W, et al. Draft genome sequence of *Aspergillus oryzae* strain 3.042 [J]. Eukaryotic Cell, 2012, 11(9): 1178
- [4] Deamer D, Akeson M, Branton D. Three decades of nanopore sequencing [J]. Nature Biotechnology, 2016, 34(5): 518-524
- [5] Magi A, Semeraro R, Mingrino A, et al. Nanopore sequencing data analysis: state of the art, applications and challenges [J]. Briefings in Bioinformatics, 2018, 19(6): 1256-1272
- [6] Jain M, Olsen H E, Paten B, et al. The oxford nanopore MinION: delivery of nanopore sequencing to the genomics community [J]. Genome Biology, 2016, 17(1): 239
- [7] Koren S, Walenz B P, Berlin K, et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting

- and repeat separation [J]. *Genome Research*, 2017, 27(5): 722-736
- [8] Walker B J, Abeel T, Shea T, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement [J]. *PloS One*, 2014, 9(11): e112963
- [9] Cantarel B L, Coutinho P M, Rancurel C, et al. The carbohydrate-active enzymes database (CAZy): an expert resource for glycomics [J]. *Nucleic Acids Research*, 2009, 37(Database issue): D233-238
- [10] 童星,彭勃.一株高酶活力米曲霉菌株的选育及其在酱油生产中的应用[J].*中国酿造*,2018,37(11):51-55
- TONG Xing, PENG Bo. Breeding of *Aspergillus oryzae* with high enzyme activity and its application in soy sauce production [J]. *China Brewing*, 2018,37(11): 51-55
- [11] Kang C, Yu X W, Xu Y. Cloning and expression of a novel prolyl endopeptidase from *Aspergillus oryzae* and its application in beer stabilization [J]. *Journal of Industrial Microbiology & Biotechnology*, 2015, 42(2): 263-272
- [12] Wang B, Guo G, Wang C, et al. Survey of the transcriptome of *Aspergillus oryzae* via massively parallel mRNA sequencing [J]. *Nucleic Acids Research*, 2010, 38(15): 5075-5087
- [13] Kudo K, Watanabe A, Ujiie S, et al. Purification and enzymatic characterization of secretory glycoside hydrolase family 3 (GH3) aryl beta-glucosidases screened from *Aspergillus oryzae* genome [J]. *Journal of Bioscience and Bioengineering*, 2015, 120(6): 614-623
- [14] Tang C D, Shi H L, Tang Q H, et al. Genome mining and motif truncation of glycoside hydrolase family 5 endo-beta-1,4-mannanase encoded by *Aspergillus oryzae* RIB40 for potential konjac flour hydrolysis or feed additive [J]. *Enzyme and Microbial Technology*, 2016, 93-94: 99-104
- [15] Zhao G, Yao Y, Hou L, et al. Comparison of the genomes and transcriptomes associated with the different protease secretions of *Aspergillus oryzae* 100-8 and 3.042 [J]. *Biotechnol Lett*, 2014, 36(10): 2053-2058
- [16] 黄艳,王静,高庭,等.一株高产蛋白酶的米曲霉在酱油酿造中的发酵性能对比研究[J].*中国调味品*,2018,43(12):80-82, 88
- HUANG Yan, WANG Jing, GAO Ting, et al. A comparative study on fermentation performance of *Aspergillus oryzae* with high protease activity in soy sauce fermentation [J]. *China Condiment*, 2018, 43(12): 80-82, 88

(上接第 230 页)

- [29] Tao Y-S, Jin G-J, Y-H Xu, et al. Wine aroma response to different participation of selected *Hanseniaspora uvarum* in mixed fermentation with *Saccharomyces cerevisiae* [J]. *Food Research International*, 2018, 108: 119-127
- [30] Duan C-Q, Zhang X-K, Lan Y-B, et al. Changes in monosaccharides, organic acids and amino acids during Cabernet Sauvignon wine ageing based on a simultaneous analysis using gas chromatography-mass spectrometry [J]. *Journal of the Science of Food and Agriculture*, 2018, 98(1): 104-112