

高产聚苹果酸黑色素短梗霉CGMCC18996全基因组组装注释及关键蛋白分析

王舸楠^{1,2,3}, 李佳谦^{1,2,3}, 李雨桐^{1,2}, 陈世伟¹, 王淑贤¹, 赵廷彬⁴, 贾士儒^{1,2,3}, 乔长晟^{1,2,3,4,*}

(1.天津科技大学生物工程学院, 天津 300457;

2.工业发酵微生物教育部重点实验室暨天津市工业微生物重点实验室, 天津 300457;

3.天津市微生物代谢与发酵过程控制技术工程中心, 天津 300457; 4.天津慧智百川生物工程有限公司, 天津 300457)

摘要: 通过PacBio Sequel II及Illumina NovaSeq 6000测序平台对高产聚苹果酸 (polymalic acid, PMLA) 产黑色素短梗霉 (*Aureobasidium melanogenum*) 基因组进行测序, 对测序得到的下机文件进行组装, 并结合转录组数据进行基因结构注释。结果表明, 产黑色素短梗霉基因组中共有6 202个基因, 主要参与碳水化合物转运及代谢、氨基酸转运代谢、转录后修饰、RNA加工及修饰等生物活动。功能注释结果显示基因组中有大部分基因与过氧化物体有关, 进而对菌株进行透射电镜拍摄, 发现菌体中存在圆形的类过氧化物体 (乙醛酸体) 结构, 提示菌体可通过乙醛酸循环途径生成苹果酸。最后, 对可能与PMLA合成有关的磷酸烯醇式丙酮酸羧化酶激酶和苹果酸合成酶进行了蛋白质结构预测, 发现这两种蛋白可能具备合成苹果酸的能力。本研究结果可为产黑色素短梗霉菌株的PMLA代谢提供一定的参考, 同时, 组装的基因组文件已上传至相关数据库, 为后续的菌株开发利用提供基础。

关键词: 聚苹果酸; 产黑色素短梗霉; 基因组测序; 基因组组装; 蛋白质预测

Genome Assembly and Annotation of a High-Polymalic Acid (PMLA) Producing Strain *Aureobasidium melanogenum* CGMCC18996 and Analysis of Its Key Proteins Related to PMLA Synthesis

WANG Genan^{1,2,3}, LI Jiaqian^{1,2,3}, LI Yutong^{1,2}, CHEN Shiwei¹, WANG Shuxian¹, ZHAO Tingbin⁴, JIA Shiru^{1,2,3}, QIAO Changsheng^{1,2,3,4,*}

(1. College of Bioengineering, Tianjin University of Science and Technology, Tianjin 300457, China; 2. Key Laboratory of Industrial Fermentation Microbiology, Ministry of Education, Tianjin Key Laboratory of Industrial Microbiology, Tianjin 300457, China;

3. Tianjin Engineering Center of Microbial Metabolism and Fermentation Process Control, Tianjin 300457, China;

4. Tianjin Huizhi Baichuan Bioengineering Co. Ltd., Tianjin 300457, China)

Abstract: In this study, we applied PacBio Sequel II and Illumina NovaSeq 6000 sequencing platforms to sequence the genome of a high-polymalic acid (PMLA)-producing strain, *Aureobasidium melanogenum* CGMCC18996, and used different assemblers to obtain a high-quality genome assembly, which was then annotated using transcriptomic data. The results indicated a total of 6 202 genes were found in the *A. melanogenum* genome, mainly involved in carbohydrate transport and metabolism, amino acid transport and metabolism, post-translational modification, RNA processing and modification. Meanwhile, functional annotation revealed that most genes in the genome were related to peroxisome in the strain. Transmission electron microscopy (TEM) indicated the existence of a circular peroxisome-like (glyoxysome) structure in the cells, demonstrating the ability to malic acid through the glyoxylate cycle. Finally, we predicted the protein structures of two enzymes related to PMLA biosynthesis, phosphoenolpyruvate carboxykinase (PCKA) and malate synthase (MASY). It was found that the enzymes could have the ability to synthesize malic acid. This study could provide a reference for

收稿日期: 2022-09-06

基金项目: 天津市科委基金项目 (21YDTPJC00140); 宁夏回族自治区重点研发计划项目 (2022CMG02006; 2021BEG03011); 工业发酵微生物教育部重点实验室暨天津市工业微生物重点实验室开放课题 (2020KF006)

第一作者简介: 王舸楠 (1994—) (ORCID: 0000-0001-9501-869X), 男, 博士研究生, 研究方向为轻工技术与工程。

E-mail: 758104598@qq.com

*通信作者简介: 乔长晟 (1969—) (ORCID: 0000-0003-2058-3867), 男, 教授, 博士, 研究方向为生物工程。

E-mail: qiaochangsheng@tust.edu.cn

metabolism regulation in *A. melanogenum* for improved PMLA production, and the assembled genome has been uploaded to the database, which could provide the basis for the future development and utilization of *A. melanogenum* CGMCC18996.

Keywords: polymeric acid; *Aureobasidium melanogenum*; genome sequencing; genome annotation; protein prediction

DOI:10.7506/spkx1002-6630-20220906-059

中图分类号: Q93

文献标志码: A

文章编号: 1002-6630(2023)16-0213-07

引文格式:

王舸楠, 李佳谦, 李雨桐, 等. 高产聚苹果酸黑色素短梗霉CGMCC18996全基因组组装注释及关键蛋白分析[J]. 食品科学, 2023, 44(16): 213-219. DOI:10.7506/spkx1002-6630-20220906-059. <http://www.spkx.net.cn>

WANG Genan, LI Jiaqian, LI Yutong, et al. Genome assembly and annotation of a high-polymeric acid (PMLA) producing strain *Aureobasidium melanogenum* CGMCC18996 and analysis of its key proteins related to PMLA synthesis[J]. Food Science, 2023, 44(16): 213-219. (in Chinese with English abstract) DOI:10.7506/spkx1002-6630-20220906-059. <http://www.spkx.net.cn>

聚苹果酸 (polymeric acid, PMLA) 是以苹果酸为唯一单体的均聚高分子聚合物, 属于聚酯类聚合物, 具有高生物相容性、高水溶性、生物可吸收性、化学可衍生性、可降解性和无免疫原性等多种优良性能^[1]。PMLA分子的侧链带有可修饰性较强的羧基, 一些分子能够通过官能团反应引入其聚合链上, 同时PMLA水解产生的苹果酸是理想的调酸剂, 这使PMLA在生物医药、食品和生物材料领域具有潜在的应用前景^[2-6]。

研究表明产黑色素短梗霉 (*Aureobasidium melanogenum*) 是一种具有较强产PMLA能力的类酵母真菌^[4], 但目前对其PMLA生物合成研究并不透彻, 对通路中关键酶的研究并不清晰。产量还有很大的提升空间^[7]。同时, 该菌基因组信息依旧较少, 限制了该菌株的开发利用。因此, 对该菌种进行基因组测序及组装可为改造菌种, 提高产量提供理论依据。

目前, 主流的基因测序分为两种, 单分子测序 (如PacBio三代测序平台) 和高通量测序 (如二代Illumina平台), 其中单分子测序具有读长长组装结果好的特点, 但其测序成本与错误率较高; 高通量测序采用双端测序策略, 结果兼备质量高、价格低的特点, 但其测序长度较短, 组装会产生较多的片段 (contigs)。因此采用三代测序结果组装再通过二代测序结果进行纠错 (基因组polish), 能够很大程度降低contigs数目同时保持基因组准确率, 在完成基因组组装后, 可通过转录组测序结果对基因组进行结构注释, 以同时获得高质量基因组注释结果^[8]。

本研究通过PacBio Sequel II及Illumina NovaSeq 6000测序平台对高产PMLA产黑色素短梗霉基因组进行测序, 通过不同组装软件对测序的下机文件进行组装及优化; 结合转录组数据对组装结果进行基因结构注释。之后对基因组注释结果进行不同数据库的功能注释, 同时分析PMLA合成关键蛋白。通过这种方法, 期望得到适合于基因组分析以及后续分子生物学实验的高质量基因组, 为产黑色素短梗霉的开发利用提供一定生物信息学参考, 并同时为其他类似物种的基因组组装提供思路。

1 材料与方法

1.1 材料与试剂

菌株黑色素短梗霉, 已保存至中国普通微生物菌种保藏管理中心 (保藏号: CGMCC18996); 基因组提取试剂盒 湖南艾科瑞生物工程有限公司; 乙腈 (色谱纯) 北京鼎国昌盛生物技术有限公司; 磷酸二氢钾、磷酸氢二钾、磷酸 (均为色谱纯) 天津市大茂化学试剂厂。

1.2 仪器与设备

UC7超薄切片仪 德国Leica公司; LC-10高效液相色谱仪 日本岛津公司; ZWYR-D2403恒温培养振荡器上海智城分析仪器制造有限公司; LRH-250A生化培养箱 泰宏医疗器械有限公司; NanoDrop One超微量分光光度计 美国赛默飞世尔科技公司。

1.3 方法

1.3.1 基因组提取与测序

本实验的基因组提取、建库以及测序委托诺禾致源科技股份有限公司 (北京) 完成, 实验流程均按照诺禾致源公司测序方法进行 (https://www.novogene.com/novo/sd_cpjs_6.html, https://www.novogene.com/novo/ed_cpjs_2.html)。

1.3.2 数据处理及生物信息学分析

使用曙光云计算服务器 (曙光信息产业股份有限公司) 进行生物信息学分析。最终得到相关的测序下机文件以及组装的基因组文件已上传至国家生物信息中心 (项目号: PRJCA011444)。

1.3.2.1 原始下机文件获取及质控

通过Illumina NovaSeq 6000平台获取高通量测序产生的fastq基因组测序文件, 通过PacBio Sequel II测序平台获取单分子测序产生的bam基因组测序文件, 同时验证高通量测序和单分子测序文件的md5值与平台提供是否一致。

1.3.2.2 基因组组装及评估

使用SPAdes-3.15.5^[9]组装软件对二代测序clean reads进行组装, 为了得到最优 N_{50} 及 L_{50} 值, 将k-mer大

小设定为77、87、97、107、117及127进行组装。并通过QUAST^[10]软件对组装结果进行评估。三代测序通过Flye^[11]、Canu^[12]、NextDenovo (<https://github.com/Nextomics/NextDenovo>)以及MECAT2^[13]组装软件进行组装,通过quickmerge^[14]将最优结构与其他组装结果进行融合,之后对得到的融合后结果进行去重复处理进一步降低contigs数目。

1.3.2.3 基因组注释

使用BRAKER2^[15]软件进行基因组注释,首先通过HISAT2将转录组序列比对到基因组上,生成bam文件,再将生成的bam文件与基因组文件作为输入参数使用BRAKER2进行注释。将获得的氨基酸序列通过InterProScan^[16]与eggNOG-mapper^[17]进行注释得到UniProtKB AC/ID(基因名),将这些基因名进行京都基因与基因组百科全书(Kyoto Encyclopedia of Genes and Genomes, KEGG) (<https://www.kegg.jp/kegg/>)、基因本体论(Gene Ontology, GO)^[18]及直系同源集(Clusters of Orthologous Groups, COG)^[19]数据库注释,并通过antiSMASH^[20]对次级代谢产物合成基因簇进行预测。

1.3.3 菌体形态

透射电镜拍摄方法如下,将电子显微镜固定剂(2.5%戊二醛溶于磷酸缓冲液)添加到经离心分离的菌株中,重悬菌体,使用0.1 mol/L磷酸漂洗液漂洗3次,1%锇酸固定液固定1 h;之后使用0.1 mol/L磷酸漂洗液漂洗3次,丙酮脱水3次。使用纯EMBed 812树脂(90529-77-4)包埋,移入65 °C烘箱中聚合48 h;使用超薄切片机制至60~80 nm薄片,用2%乙酸铀饱和醇溶液和柠檬酸铅染色15 min。最后,使用JEM1200电镜进行观察拍摄,得到放大4 000倍的产黑色素短梗霉CGMCC18996图像。

1.3.4 蛋白质结构预测

蛋白质结构预测使用Alphafold2-2.2.0^[21],将基因组注释的氨基酸序列作为输入参数,运行模式参数选择monomer,其余均选用默认选项进行蛋白质结构预测,最终在生成文件中选择最优结果(即rank0结果)。之后将蛋白预测结果与蛋白质数据库(Protein Data Bank, PDB)^[22]中已上传的晶体结构进行比对(磷酸烯醇式丙酮酸羧化酶激酶(phosphoenolpyruvate carboxykinase, PCKA): 1YLH^[23];苹果酸合成酶(malate synthase, MASY): 3CUZ^[24])。

2 结果与分析

2.1 通过二代测序数据进行基因组预组装

基于Illumina NovaSeq 6000测序平台对产黑色素短梗霉菌基因组进行测序,共得到35.85 Gb×2的双端测序结果。且该测序平台得到的raw reads具有较高的质量评分($Q>35$)。之后,选取质控后clean reads进行后续的组装实验。

不同k-mer长度对产黑色素短梗霉clean reads进行组装结果见表1,组装结果显示,产黑色素短梗霉菌基因组大

小约为44 Mb。其中,在k-mer长度为127时得到 N_{50} 值为908 694, L_{50} 值为14,基因组覆盖度为121.2×。基于该结果可以判断该菌基因组长度理论值在44 Mb左右;且随着k-mer长度的增加, N_{50} 值增大, L_{50} 值减小;若继续增大k-mer值会进一步优化组装参数。但由于SPAdes软件进行组装的k-mer值最大为127,这可能是考虑到继续加大k-mer值对测序深度要求较高,从而提高测序成本。

表1 k-mer长度对组装结果的影响

Table 1 Effect of k-mer size on assembly results

k-mer 长度	Scaffolds数 (>500 bp)	N_{50}	L_{50}	GC 含量/%	测序深度	基因组 长度/Mb
87	373	343 039	36	50.05	323.3×	44.52
97	254	411 413	33	50.06	266.0×	44.59
107	184	672 496	22	50.06	209.9×	44.64
117	161	806 284	17	50.06	154.8×	44.65
127	150	908 694	14	50.06	121.2×	44.66

2.2 三代测序组装结果及基因组结构注释

基于PacBio平台的单分子测序共产生46 Gb大小的bam基因组测序文件,转换为fasta文件后,根据二代组装结果设置基因组大小为44 Mb进行组装。各组装软件组装结果如表2所示,其中,选用Canu组装的最优结果通过quickmerge软件与其他组装结果进行融合,并结合二代测序文件进行基因组纠正(polish),在删除重复contigs后的最终组装结果见表3。

结合转录组测序文件进行基因组结构注释共注释出15 684个基因,并找到基因编码区与氨基酸预测区,因结合转录组测序文件进行结构注释,这些基因可能包含有可变剪切和重复注释的结构,会增加注释出的基因数目。因此,进行功能注释和基因名称注释后需删除重复的基因名;最终,共获得6 202个基因注释结果。

表2 三代组装结果

Table 2 Results of third-generation assembly

组装软件	contigs数	N_{50}	L_{50}	GC含量/%	基因组长度/Mb
Flye	344	254 950	32	50.01	38
Nextdenovo	89	549 604	12	49.96	25
Canu	42	2 190 588	9	50.12	46
MECAT2	193	420 943	20	50.11	36
融合结果	35	2 204 220	8	50.09	42

表3 产黑色素短梗霉菌基因组组装结果

Table 3 Results of genome assembly of *A. melanogenum*

名称	contigs数	N_{50}	L_{50}	GC含量/%	基因组长度/Mb
产黑色素短梗霉	26	2 204 220	8	50.09	42

2.3 GO、KEGG、COG以及antiSMASH次级代谢注释结果

对注释出的6 202个基因进行GO、KEGG与COG数据库注释,结果如图1所示。其中,COG注释结果显示大部分基因与碳水化合物转运及代谢、氨基酸转运代谢、转录后修饰、RNA加工及修饰有关。KEGG注释结果显示大部分基因所处代谢通路与核糖体、过氧化物体、RNA转运有关。GO注释结果显示大部分基因与RNA、过

氧化物体以及线粒体有关。最终, antiSMASH次级代谢物预测结果共发现4个非核糖体肽合成酶(nonribosomal peptide synthetase, NRPS)基因簇、5个一类聚酮酶(polyketide synthetase, pks)基因簇、3个β-内酯类合成基因簇以及7个萜类合成基因簇, 其中1个一类pks基因簇和黑色素合成有关(相关性100%), 1个一类pks基因簇与黑麦酮酸类化合物合成有关(相关性18%)。

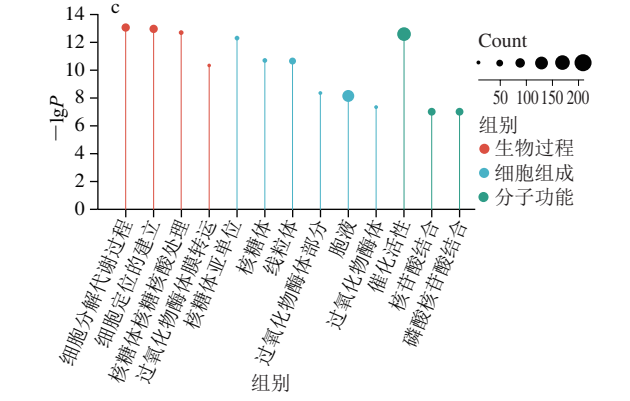
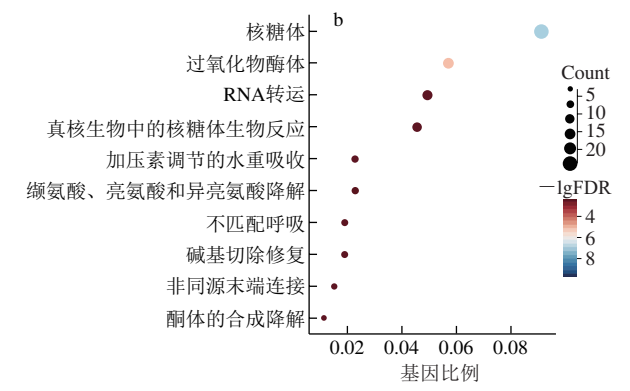
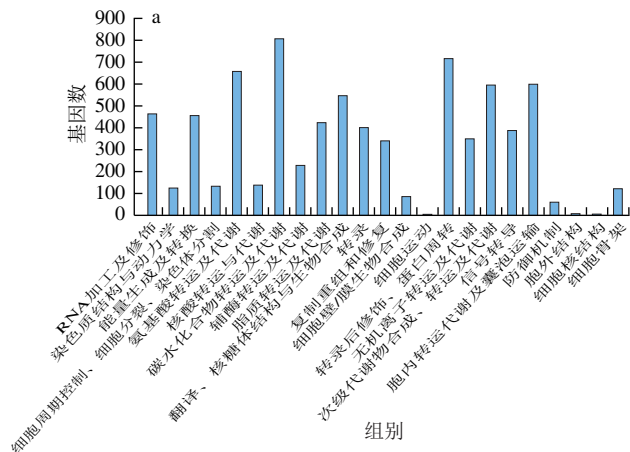


图1 COG (a)、KEGG (b)、GO (c) 数据库注释结果

2.4 产黑色素短梗霉菌体透射电镜结果

如图2所示, 在低产组中菌株具有较大细胞核(N)以及周围存在的具有双层膜结构线粒体(M), 在高产组中出现了类似乙醛酸循环体的圆形结构, 提示黑色素短梗霉中可能存在有乙醛酸循环途径。

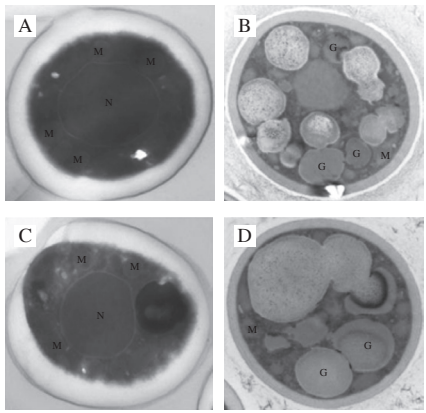


图2 产黑色素短梗霉PMLA低产组(a、c)及PMLA高产组(b、d)透射电镜结果

Fig. 2 TEM results of low- and high-PMLA producing *A. melanogenum*

2.5 PMLA合成相关基因结构预测

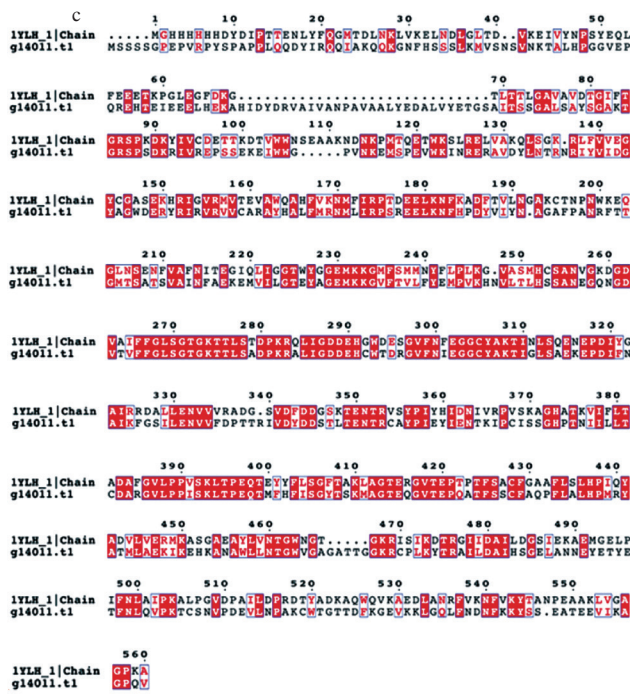
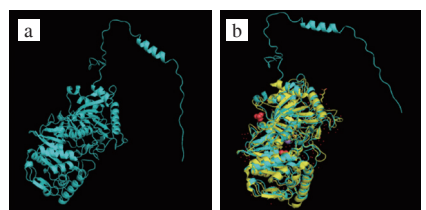


图3 AlphaFold对PCKA预测结果(a)及与PDB中已上传晶体结构的对比(b)和序列对比结果(c)

Fig. 3 Results of PCKA prediction by AlphaFold software (a), crystal structure of PCKA versus that in PDB database (b), and amino acid alignment results (c)

在对基因结构与基因名注释后，找到PCKA、MASY的蛋白质序列，其晶体结构预测结果如图3、4所示，其中，PCKA、MASY与PDB上传晶体结构（PCKA: 1YLH^[23]，MASY: 3CUZ^[24]）进行比对，结果显示预测的蛋白结构与PDB数据库中序列比对结果基本一致，氨基酸比对结果中序列与其晶体结构中小分子配体的结合位点也具有较高的一致性。同时蛋白的保守作用位点如PCKA中86R、140V、146G、287G、288D、289D，MASY中440C、275C、276G、277R、278W在比对结果中一致。

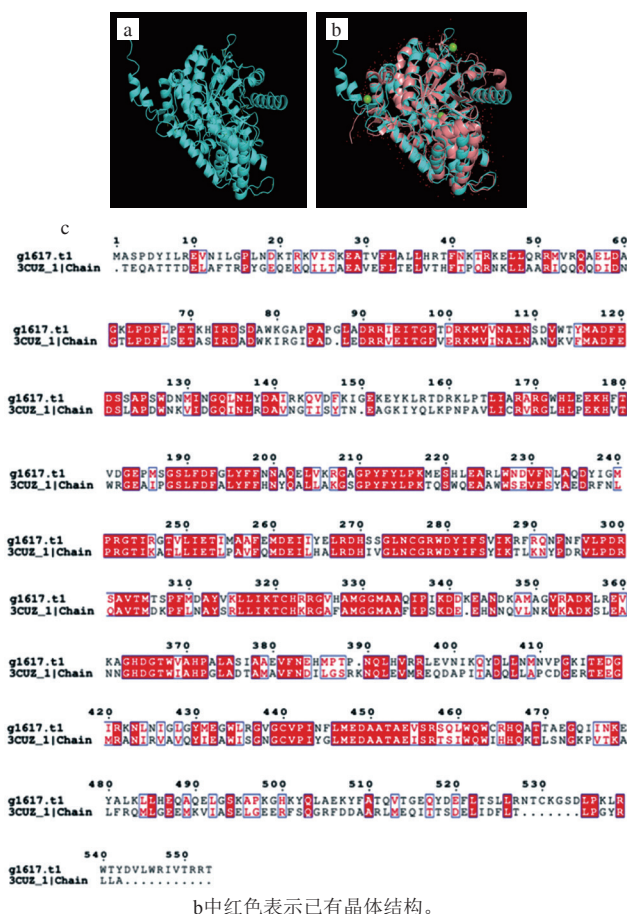


图4 Alphafold对MAS Y预测结果(a)及与PDB中已上传晶体结构的比对(b)和序列对比结果(c)

Fig. 4 Results of MASY prediction by Alphafold software (a), crystal structure of MASY versus that in PDB database (b), and amino acid alignment results (c)

3 讨论

结合基因组及转录组测序结果可以组装并注释出质量较高的基因组。本研究通过三代测序(Pacbio sequel II平台)、二代测序(Illumina NovaSeq 6000)平台对产黑色素短梗霉基因组进行测序，因三代组装需要预估基

因组大小，因此，首先通过二代数据进行基因组预组装共得到44 Mbp基因组大小。之后考察了不同组装软件对产黑色素短梗霉基因组的组装效果，在一般的默认选项下，Canu软件得到了较好的组装效果。对该组装结果通过二代数据修正并去重后得到包含26个contigs、 N_{50} 为2 204 220、GC值为50.09%、大小为42 Mb的较高质量基因组组装结果。通过转录组数据对该组装结果进行结构注释，共找到6 202个基因。产黑色素短梗霉属于出芽短梗霉的亚种，其在不同环境中也会呈现出酵母状与菌丝体状的不同形态^[25]。将基因组组装与注释结果与酵母和丝状真菌的模式生物基因组进行比较；其中酿酒酵母(*Saccharomyces cerevisiae*)基因组大小为12.15 Mb，编码6 016个蛋白，GC含量为38.15%^[26]；构巢曲霉基因组大小为30.30 Mb，编码10 008个蛋白，GC含量为50.10%^[27]。因此，产黑色素短梗霉基因组组成更偏向构巢曲霉，且本实验室先前通过无参转录组注释发现很多与构巢曲霉同源的基因^[28]。因构巢曲霉是丝状真菌的模式生物，该结果提示产黑色素短梗霉可能同样适用构巢曲霉的分子转化方法。

本研究通过Illumina NovaSeq 6000平台对产黑色素短梗霉进行了大约35 GB下机文件大小的高通量测序，该测序量的测序深度较大，理论测序深度为1 000×。通过二代组装软件进行组装后，最终组装出 N_{50} 值为908 694的基因组，该结果仅达到三代测序组装结果 N_{50} 的41.4%，因此，考虑到组装结果以及建库与测序成本^[29]，三代测序进行组装辅以二代测序进行基因组修正的测序方法更具有性价比。

通过COG、KEGG与GO数据库注释结果表明，产黑色素短梗霉中基因表达主要位于核糖体、线粒体以及过氧化物体等细胞器中，其中，有研究表明线粒体中的三羧酸循环、乙醛酸循环体中的乙醛酸循环以及细胞质中的还原性三羧酸途径与PMLA的生物合成有关^[7,30-31]。功能注释结果中出现了大量与过氧化物体有关的基因，而能够产生苹果酸的乙醛酸循环体也属于一类过氧化物体。因此，该结果说明在产黑色素短梗霉中可能存在乙醛酸循环体。实验室先前研究发现乙醛酸体中的MASY可通过乙醛酸途径生成苹果酸，相比于线粒体中存在的苹果酸/天冬氨酸穿梭体系，乙醛酸循环体中苹果酸可能可以直接通过单层乙醛酸体膜进入细胞质中而被非核糖体肽合成酶(non-ribosomal peptide synthase, NRPS)聚合合成PMLA^[28]。透射电镜结果也显示，在加入CaCO₃的高产PMLA组中，菌体内部出现了较多的圆形细胞器，提示菌体中存在乙醛酸循环体的可能。antiSMASH注释结果表明在产黑色素短梗霉中存在4个NRPS基因簇，该结果也同样反映出NRPS蛋白聚合苹果酸形成PMLA的可能性。同时，注释出的与黑色素合成的pks基因簇可解释该菌种在生长过程中逐渐变黑的现象。

实验室先前研究表明^[28], *PCKA*、*MASY*和*NRPS*基因在PMLA高产组中发生了较大幅度的上调。因此,通过对基因组注释得到的*PCKA*和*MASY*的编码区进行了蛋白结构预测,并将预测结果与PDB上已有的晶体结构进行比对。比对结果表明,蛋白预测结果与已有晶体结构基本吻合,且保守结构域相似,提示预测蛋白与其他菌株中已存在蛋白可能具有相似的功能。*PCKA*是糖异生途径的限速酶,一般情况下催化草酰乙酸转化为磷酸烯醇式丙酮酸,但有研究表明*PCKA*在真菌细胞质中可反向催化磷酸烯醇式丙酮酸生成草酰乙酸^[23]。而细胞质中的草酰乙酸会进一步转化为PMLA的前体物质苹果酸。*MASY*可催化乙醛酸循环中的第二部反应,将乙醛酸转化为苹果酸^[24],而乙醛酸体的单层膜结构可能使苹果酸被动运输进入细胞质,进而被细胞质中存在的*NRPS*蛋白聚合成为PMLA^[28]。本实验没有对*NRPS*蛋白预测的原因是因为其氨基酸序列较长(5 000 aa左右),软件无法预测,其次是在PDB上还没有找到近似的蛋白晶体结构,比对较为困难。

本研究对产黑色素短梗霉进行二代、三代基因组测序并组装得到最优组装结果,后通过转录组数据对基因组进行结构注释,将得到的结构注释结果进行不同数据库的功能注释。同时对产黑色素短梗霉进行了透射电镜拍摄,电镜结果提示菌株中存在乙醛酸循环体结构,该结构可能与PMLA合成有关,最终,实验通过对PMLA合成相关的*PCKA*和*MASY*蛋白进行结构预测并与PDB上已有电镜结果进行比对。发现产黑色素短梗霉中这两种蛋白与PDB上已有晶体结构基本一致,且这两种蛋白功能最终都与苹果酸代谢有关。本实验结果可为产黑色短梗霉菌株的PMLA代谢提供一定的参考,相关的测序下机文件以及组装的基因组文件已上传至国家生物信息中心(PRJCA011444),为后续的菌株开发利用提供基础。

参考文献:

- [1] ISRAEL L L, BRAUBACH O, GALSTYAN A, et al. A combination of tri-leucine and angiopep-2 drives a polyanionic polymalic acid nanodrug platform across the blood-brain barrier[J]. *ACS Nano*, 2019, 13: 1253-1271. DOI:10.1021/acsnano.8b06437.
- [2] CHI Z, LIU G L, LIU C G, et al. Poly(beta-L-malic acid) (PMLA) from *Aureobasidium* spp. and its current proceedings[J]. *Applied Microbiology and Biotechnology*, 2016, 100: 3841-3851. DOI:10.1007/s00253-016-7404-0.
- [3] 张亚南, 王洁, 张猛, 等. 直接缩聚法载药纳米胶束的合成及pH控释研究[J]. *化工新型材料*, 2015(43): 158-160.
- [4] WANG G N, SHI B Y, ZHANG P, et al. Effects of corn steep liquor on beta-poly(L-malic acid) production in *Aureobasidium melanogenum*[J]. *AMB Express*, 2020, 10: 211. DOI:10.1186/s13568-020-01147-8.
- [5] ZOU X, CHENG C, FENG J, et al. Biosynthesis of poly malic acid in fermentation: advances and prospects for industrial application[J]. *Critical Reviews in Biotechnology*, 2019, 39: 408-421. DOI:10.1080/07388551.2019.1571008.
- [6] ZHOU Q, YANG T H, QIAO Y B, et al. Preparation of poly(beta-L-malic acid)-based charge-conversional nanoconjugates for tumor-specific uptake and cellular delivery[J]. *International Journal of Nanomedicine*, 2015, 10: 1941-1952. DOI:10.2147/ijn.S78547.
- [7] CAO W F, LOU J Q, QI B K, et al. beta-Poly(L-malic acid) production by fed-batch culture of *Aureobasidium pullulans* ipe-1 with mixed sugars[J]. *Engineering in Life Sciences*, 2014, 14: 180-189. DOI:10.1002/elsc.201200189.
- [8] LI W, LI K, ZHANG Q J, et al. Improved hybrid *de novo* genome assembly and annotation of African wild rice, *Oryza longistaminata*, from Illumina and PacBio sequencing reads[J]. *The Plant Genome*, 2020, 13: e20001. DOI:10.1002/tpg2.20001.
- [9] ANDREY P, DMITRY A, DMITRY M et al. Using SPAdes *De Novo* assembler[J]. *Current Protocols in Bioinformatics*, 2020, 70: e102. DOI:10.1002/cpbi.102.
- [10] ALEXEY G, VLADISLAV S, NIKOLAY V, et al. QUAST: quality assessment tool for genome assemblies[J]. *Bioinformatics*, 2013, 29: 1072-1075. DOI:10.1093/bioinformatics/btt086.
- [11] MIKHAIL K, DEREK M B, BAHAR B, et al. metaFlye: scalable long-read metagenome assembly using repeat graphs[J]. *Nature Methods*, 2020, 17: 1103-1110. DOI:10.1038/s41592-020-00971-x.
- [12] SETGEY N, BRIAN W, ARANG R, et al. HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads[J]. *Genome Research*, 2020, 30: 263566.120. DOI:10.1101/gr.263566.120.
- [13] XIAO C L, CHEN Y, XIE S Q, et al. MECAT: fast mapping, error correction, and *de novo* assembly for single-molecule sequencing reads[J]. *Nature Methods*, 2017, 14: 1072-1074. DOI:10.1038/nmeth.4432.
- [14] CHAKRABORTY M, BALDWIN-BROWN J G, LONG A D, et al. Contiguous and accurate *de novo* assembly of metazoan genomes with modest long read coverage[J]. *Nucleic Acids Research*, 2016, 44: e147. DOI:10.1093/nar/gkw654.
- [15] TOMAS B, KATHARINA J H, ALEXANDRE L, et al. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database[J]. *NAR Genomics and Bioinformatics*, 2021, 3. DOI:10.1093/nargab/lqaa108.
- [16] PHILIP J, DAVID B, CHANG H, et al. InterProScan 5: genome-scale protein function classification[J]. *Bioinformatics*, 2014, 30: 1236-1240. DOI:10.1093/bioinformatics/btu031.
- [17] CARLOS P C, ANA H P, IVICA L, et al. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale[J]. *Molecular Biology and Evolution*, 2021, 38: 5825-5829. DOI:10.1093/molbev/msab293.
- [18] CARBON S, DOUGLASS E, GOOD B M, et al. The Gene Ontology resource: enriching a gold mine[J]. *Nucleic Acids Research*, 2020, 49: D325-D334. DOI:10.1093/nar/gkaa1113.
- [19] JAMINE H C, DAMIAN S, DAVIDE H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5 090 organisms and 2 502 viruses[J]. *Nucleic Acids Research*, 2018, 47: D309-D314. DOI:10.1093/nar/gky1085.
- [20] KAI B, SIMOIN S, ALEXANDER M K, et al. antiSMASH 6.0: improving cluster detection and comparison capabilities[J]. *Nucleic Acids Research*, 2021, 49: W29-W35. DOI:10.1093/nar/gkab335.
- [21] ROSHAN R, JASON L, ROBERT V, et al. MSA transformer[J]. *bioRxiv*, 2021, 43: 8-58. DOI:10.1101/2021.02.12.430858.
- [22] HELEON M, BERMAN, J W, FENG Z K, et al. The Protein Data Bank[J]. *Nucleic Acids Research*, 2000, 28: 235-242. DOI:10.1093/nar/28.1.235.

- [23] LEDUC Y A, PRASAD L, LAIVENIEKS M, et al. Structure of PEP carboxykinase from the succinate-producing *Actinobacillus succinogenes*: a new conserved active-site motif[J]. Acta Crystallographica Section D, 2005, 61: 903-912. DOI:10.1107/S0907444905008723.
- [24] LOHMAN J R, OLSON A C, REMINGTON S J. Atomic resolution structures of *Escherichia coli* and *Bacillus anthracis* malate synthase A: comparison with isoform G and implications for structure-based drug discovery[J]. Protein Science, 2008, 17: 1935-1945. DOI:10.1110/ps.036269.108.
- [25] GOSTINČAR C, OHM R A, KOGEJ T, et al. Genome sequencing of four *Aureobasidium pullulans* varieties: biotechnological potential, stress tolerance, and description of new species[J]. BMC Genomics, 2014, 15: 549.
- [26] ENGEL S R, WONG E D, NASH R S, et al. New data and collaborations at the *Saccharomyces* Genome Database: updated reference genome, alleles, and the alliance of genome resources[J]. Genetics, 2022, 220. DOI:10.1093/genetics/iyab224.
- [27] WORTMAN J R, GILSENNAN J M, GILSENNAN V J, et al. The 2008 update of the *Aspergillus nidulans* genome annotation: a community effort[J]. Fungal Genetics and Biology, 2009, 46(1): S2-13. DOI:10.1016/j.fgb.2008.12.003.
- [28] WANG G, YIN H, ZHAO T, et al. Full-length *de novo* transcriptome assembly of *Aureobasidium melanogenum* CGMCC18996 to analyze the beta-poly(*L*-malic acid) biosynthesis pathway under the CaCO₃ addition[J]. Food Science and Human Wellness, 2023, 12(4): 1248-1256.
- [29] PHILIP A, SATHEESH N, TANSY P, et al. Revolutionising public health reference microbiology using whole genome sequencing: *Salmonella* as an exemplar[J]. bioRxiv, 2015, 3: 3225. DOI:10.1101/033225.
- [30] LEE B S, HOLLER E. β -Poly(*L*-malate) production by non-growing microplasmidia of *Physarum polycephalum*: effects of metabolic intermediates and inhibitors[J]. FEMS Microbiology Letters, 2000, 193: 69-74. DOI:10.1016/S0378-1097(00)00457-2.
- [31] LEE B S, MAURER T, LALBITZER H R, et al. β -Poly(*L*-malate) production by *Physarum polycephalum*-C¹³ nuclear magnetic resonance studies[J]. Applied Microbiology and Biotechnology, 1999, 52: 415-420. DOI:10.1007/s002530051540.