



## 聚苹果酸生产菌出芽短梗霉 CCTCC M2012223 的全基因组测序及序列分析

王永康<sup>1</sup>, 宋晓丹<sup>1</sup>, 李晓荣<sup>1</sup>, 杨尚天<sup>2</sup>, 邹祥<sup>1\*</sup>

<sup>1</sup>西南大学药学院, 重庆药物过程与质量控制工程技术中心, 重庆 400715

<sup>2</sup>William G. Lowrie Department of Chemical and Biomolecular Engineering, The Ohio State University, Columbus, OH 43210, USA

**摘要:**【目的】解析出芽短梗霉 CCTCC M2012223 的基因组序列信息, 分析其代谢产物聚苹果酸、黑色素、普鲁兰多糖合成相关基因, 为深入研究遗传多样性和代谢工程改造提供序列背景信息。【方法】使用 Illumina HiSeq 高通量测序平台对出芽短梗霉 CCTCC M2012223 菌株进行全基因组测序, 并对测序数据进行序列拼接, 基因预测与功能注释, COG/GO 聚类分析, 比较基因组学分析等。下载其他 5 株出芽短梗霉基因组序列, 比较分析 6 株菌的种内同源基因、全基因组进化以及代谢产物合成相关基因。【结果】出芽短梗霉 CCTCC M2012223 基因组序列全长 30756831 bp, GC 含量 47.49%, 编码 9452 个基因。比较基因组分析表明出芽短梗霉 CCTCC M2012223 的基因组组装长度最长, 6 株菌的同源基因数达到 7092 个, 普鲁兰多糖和聚苹果酸合成相关基因的蛋白序列有很高的保守性。出芽短梗霉 CCTCC M2012223 和 *Aureobasidium pullulans* var. *melanogenum* 亲缘关系最近, 而这 2 株菌的黑色素合成相关基因的蛋白序列有一些插入和突变。【结论】本研究解析了出芽短梗霉 CCTCC M2012223 的基因组序列信息, 获得黑色素、普鲁兰多糖和聚苹果酸合成相关基因, 为后续的代谢机制解析和改造提供相关依据。

**关键词:** 出芽短梗霉, 全基因组测序, 比较基因组分析, 代谢机制, 聚苹果酸

出芽短梗霉(*Aureobasidium pullulans*)是一种极具工业应用价值的真菌, 分类上属于半知菌门(Deuteromycophyta)短梗霉属(*Aureodacidium*), 通常大多数出芽短梗霉菌会产生黑色素, 俗称黑酵

母真菌。由于出芽短梗霉的遗传不稳定性, 目前已分离并鉴定出多个变种<sup>[1]</sup>。出芽短梗霉能代谢产生聚苹果酸、普鲁兰糖、酶制剂、黑色素、短梗霉毒素等多种产物<sup>[2-3]</sup>, 其中聚苹果酸是其代谢产生

基金项目: 国家自然科学基金(31571816); 国家“863计划”(2015AA021005, 2014AA021205); 重庆市社会事业与民生保障专项(cstc2016shmszx80075); 重庆市知识产权专项基金(CQIPO2015224)

\*通信作者。Tel: +86-23-68251225; Fax: +86-23-68251048; E-mail: zhx1030@swu.edu.cn

收稿日期: 2016-05-15; 修回日期: 2016-06-21; 网络出版日期: 2016-07-01

的一种聚酯型聚合物,以 L-苹果酸为唯一单体聚合,具有良好的水溶性、生物可降解性、易修饰性和无免疫原性等特点,在药物载体、组织工程、食品包装等领域具有广阔的应用前景。

本课题组前期筛选到 1 株高产聚苹果酸的出芽短梗霉菌株 CCTCC M2012223,已对其遗传转化、聚合途径、发酵调控等方面开展了系统研究<sup>[4-7]</sup>,并建立了基于聚苹果酸水解制备 L-苹果酸的新工艺技术<sup>[8]</sup>。尽管出芽短梗霉 CCTCC M2012223 不分泌黑色素,但在发酵过程中也存在少量副产物普鲁兰多糖和浅色色素形成,影响聚苹果酸产物得率和产品外观。因此,为深入了解聚苹果酸合成代谢规律,有必要对菌株的遗传背景有更加深入的了解,本研究报道了聚苹果酸生产菌出芽短梗霉 CCTCC M2012223 的全基因组序列,并与已报道 5 株出芽短梗霉基因组进行比较分析,对代谢产物聚苹果酸、普鲁兰多糖及黑色素合成的相关基因进行功能分析,为深入研究菌株的遗传多样性及代谢工程改造研究提供背景信息。

## 1 材料和方法

### 1.1 菌株培养与基因组 DNA 的提取

产聚苹果酸菌株出芽短梗霉 CCTCC M2012223 由西南大学药学院发酵工程研究室筛选保藏。菌株在 PDA 斜面上活化 2 d,挑取单菌落于培养基(g/L):葡萄糖 60.0, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> 2.0, KH<sub>2</sub>PO<sub>4</sub> 0.1, MgSO<sub>4</sub> 0.1, ZnSO<sub>4</sub> 0.1, KCl 0.5, 于 25 °C 培养 24 h,使用 OMEGA 基因组 DNA 提取试剂盒按照操作步骤提取基因组 DNA, TBS-380 荧光计(Turner BioSystems Inc., Sunnyvale, CA)测定 DNA 浓度,采用高质量的 DNA 样品( $OD_{260/280}=1.8-2.0$ , 总量>6 μg)构建片段文库。

### 1.2 文库构建和 Illumina HiSeq 测序

将纯化的基因组 DNA 打断成小片段,使用 T4 DNA 聚合酶将其补平成平末端,随后在 5'端加上磷酸基团,并在 3'末端加 A 碱基,最后在片段两端连接接头。经琼脂糖凝胶电泳纯化后,PCR 扩增以富集文库。构建好的双端文库于 Illumina HiSeq 测序平台完成全基因组测序。

### 1.3 原始测序数据质量剪切、序列组装与注释

Illumina HiSeq 测序得到的原始图像数据经过 Illumina Base Calling 软件 CASAVA v1.8.2 ([http://support.illumina.com/sequencing/sequencing\\_software/casava.ilmmn](http://support.illumina.com/sequencing/sequencing_software/casava.ilmmn))转化为原始测序数据(Raw sequencing data),使用 SeqPrep (<https://github.com/jstjohn/SeqPrep>)软件确定包含接头和引物的读段,使用 Sickle (<https://github.com/najoshi/sickle>)软件进行质量剪切,得到高质量序列数据(Clean data)。使用 SOAPdenovo v2.04 (<http://soap.genomics.org.cn/>)拼接软件对 Clean data 进行多个 K-mer 参数的拼接,得到最优的组装结果,随后运用 GapCloser v1.12 软件(<http://soap.genomics.org.cn/>)对组装结果进行局部内洞填充和碱基校正。采用软件 RepeatMasker version 4.0.3 来识别、分类基因组序列的重复元件,并用“N”屏蔽基因组中的低复杂度重复序列(low-complexity sequences)和散在重复序列(interspersed repeats)。

利用 AUGUSTUS version 2.5.5 (<http://bioinf.uni-greifswald.de/augustus/>)软件对 repeat-masked 的基因组进行基因预测,将预测基因的蛋白序列分别与 NCBI-NR 蛋白质数据库 SwissProt (<http://uniprot.org>), KEGG (<http://www.genome.jp/kegg/>)<sup>[9]</sup>, COG (<http://www.ncbi.nlm.nih.gov/COG/>)<sup>[10]</sup>, GO (<http://geneontology.org/>), SignalP (<http://www.cbs.dtu.dk/services/SignalP/>),

Transmembrane domain (<http://www.cbs.dtu.dk/services/TMHMM/>)和 CAZy 数据库比对以完成蛋白功能注释。分别使用 tRNAscan-SE v1.23 (<http://lowelab.ucsc.edu/tRNAscan-SE/>)<sup>[11]</sup>软件和 RNAmmer v1.2 (<http://www.cbs.dtu.dk/services/RNAmmer/>)<sup>[12]</sup>软件预测 tRNA 和 rRNA。

#### 1.4 比较基因组学分析

从 GenBank 数据库下载 4 株出芽短梗霉变种基因组序列及注释信息, 分别是 *A. pullulans* var. *pullulans* (EXF-150, CBS 100280) (GenBank: AYEO000000000)、*A. pullulans* var. *melanogenum* (EXF-3378, CBS 110374) (GenBank: AYEN000000000), *A. pullulans* var. *subglaciale* (EXF-2481, CBS 123387) (GenBank: AYYB000000000)、*A. pullulans* var. *namibiae* (EXF-3398, CBS 147.97) (GenBank: AYEM000000000)<sup>[1]</sup>; 从 <http://genome.fungalgenomics.ca/> 下载 *A. pullulans* ATCC 62921 的基因组数据。将出芽短梗霉 CCTCC M2012223 的基因组序列与 5 株菌进行比较分析, 统计其基本特征, 使用 ORTHOMCL (<http://www.orthomcl.org/orthomcl/>) 对 6 株菌的基因进行聚类, 使用 single-copy orthologs 鉴定单拷贝直系同源, 使用 MUSCLE (<http://www.ebi.ac.uk/Tools/msa/muscle/>) 分别对单拷贝直系同源进行比对, 然后进行 concatenated alignment 使用 MEGA (<https://mega.nz/>) 软件进行系统进化分析。

#### 1.5 代谢产物相关基因分析

将出芽短梗霉 CCTCC M2012223 预测基因的蛋白序列分别与已报道的出芽短梗霉黑色素、普鲁兰多糖、聚苹果酸合成相关基因进行同源性分析得到相应的蛋白序列, 同样的方法得到其他 5 株出芽短梗霉的相应蛋白, 然后对不同菌株的蛋白进行比对和同源性分析。

## 2 结果和分析

### 2.1 基因组组装与注释

原始测序数据结果表明, 出芽短梗霉 CCTCC M2012223 基因组中有 28995154 个 pair-end reads, 读长 223 bp, 总碱基数 6466639670, 平均插入片段长度 400 bp, 平均测序深度 210.25 $\times$ 。经过序列拼接得到 393 个 contigs, 其中 262 个大于 1000 bp, contig N50 为 342222 bp, 再经局部内洞填充、碱基校正, 得到出芽短梗霉 CCTCC M2012223 的完整基因组序列, 全长 30756831 bp, GC 含量 47.49%, 预测其编码 9452 个基因, 基因总长度 14465974 bp, 编码基因占基因组百分比为 47.0%, 平均长度 1530 bp, 基因组图见图 1。

将出芽短梗霉 CCTCC M2012223 预测基因的蛋白序列与 string v9.0 (<http://string-db.org/>) 数据库进行 BLASTp 比对, 可以获得基因所对应的 COG (clusters of orthologous groups of proteins) 注释结果, 并根据 COG 注释结果对蛋白进行功能归类。如图 2 所示, 其中参与糖类转运和代谢(G, carbohydrate transport and metabolism), 翻译、核糖体结构和合成(J, translation, ribosomal structure and biogenesis), 氨基酸转运和代谢(E, amino acid transport and metabolism), 能量产生和传递(C, energy production and conversion)以及翻译后修饰、蛋白质周转、分子伴侣(O, posttranslational modification, protein turnover, chaperones)的蛋白占较大比例。

将预测基因的蛋白序列与 GO 数据库进行 BLASTp 比对获得 GO 注释信息, 使用软件 blast2go (<http://www.blast2go.com/b2ghome>) 进行 GO 功能聚类分析。GO 注释包括 3 方面的内容: 生物学过程(biological processes), 细胞组分

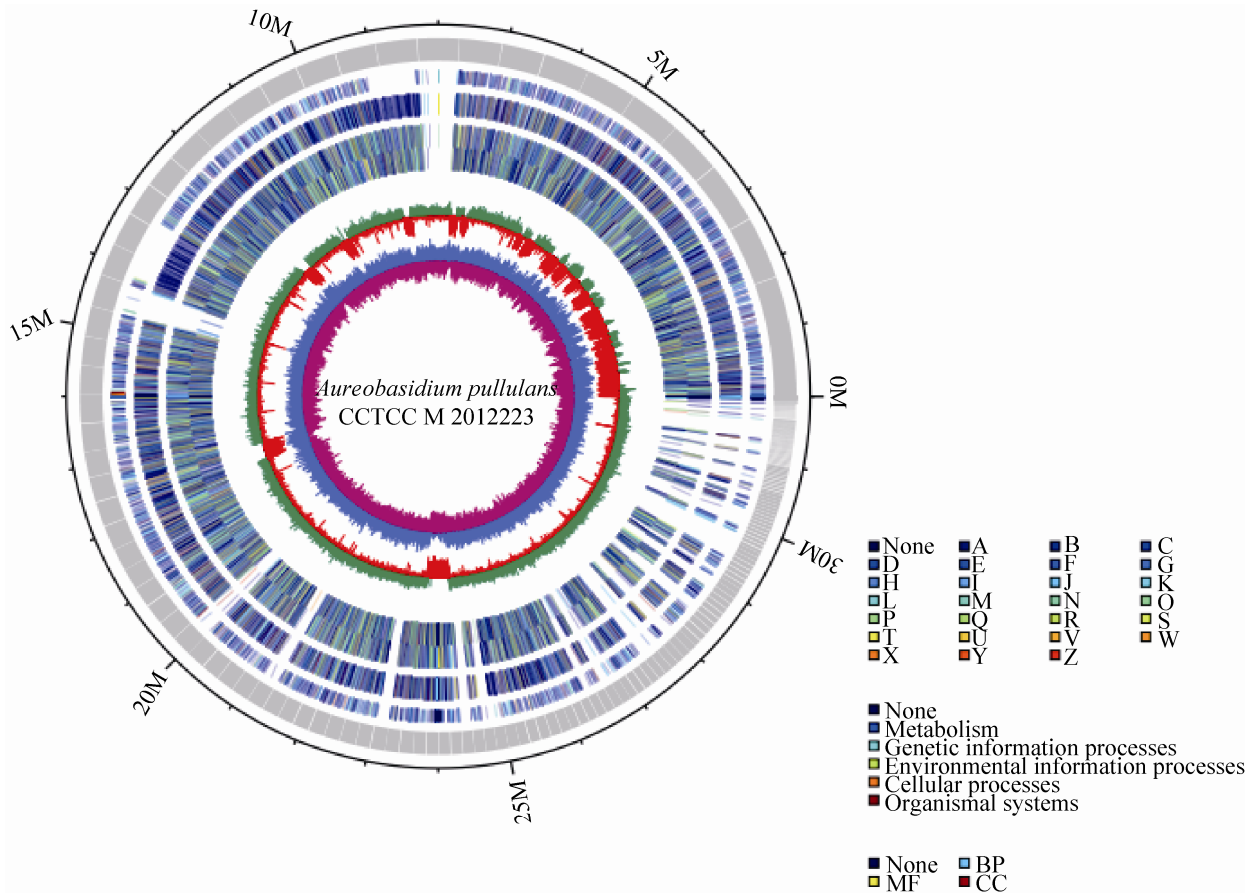


图 1. 出芽短梗霉 CCTCC M2012223 基因组图

Figure 1. Map of the genome of *A. pullulans* CCTCC M2012223. A: RNA processing and modification; B: chromatin structure and dynamics; C: energy production and conversion; D: cell cycle control, cell division, chromosome partitioning; E: amino acid transport and metabolism; F: nucleotide transport and metabolism; G: carbohydrate transport and metabolism; H: coenzyme transport and metabolism; I: lipid transport and metabolism; J: translation, ribosomal structure and biogenesis; K: transcription; L: replication, recombination and repair; M: cell wall/membrane/envelope biogenesis; N: cell motility; O: posttranslational modification, protein turnover, chaperones; P: inorganic ion transport and metabolism; Q: secondary metabolites biosynthesis, transport and catabolism; R: general function prediction only; S: function unknown; T: signal transduction mechanisms; U: intracellular trafficking, secretion, and vesicular transport; V: defense mechanisms; W: extracellular structures; X: mobilome: prophages, transposons; Y: nuclear structure; Z: cytoskeleton.

(cellular components) 和分子功能 (molecular functions)。如图 3 所示, 在生物学过程方面, 主要与代谢过程 (metabolic process)、细胞内过程 (cellular process)、单组织过程 (single organism process)、定位建立 (establishment of localization)

和定位 (localization) 有关; 在细胞组分方面, 主要与细胞部分 (cell part)、细胞 (cell)、膜 (membrane) 和细胞器 (organelle) 有关; 在分子功能方面, 主要与催化活性 (catalytic activity)、结合 (binding) 和转运活性 (transporter activity) 有关。

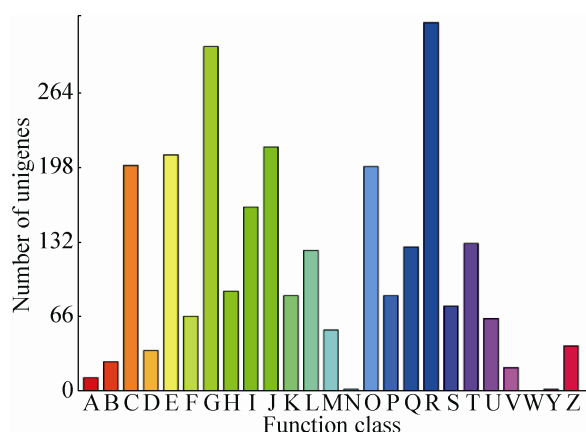


图 2. 出芽短梗霉 CCTCC M2012223 的 COG 功能归类

Figure 2. Gene distribution based on COG classification of *A. pullulans* CCTCC M2012223. A: RNA processing and modification; B: chromatin structure and dynamics; C: energy production and conversion; D: cell cycle control, cell division, chromosome partitioning; E: amino acid transport and metabolism; F: nucleotide transport and metabolism; G: carbohydrate transport and metabolism; H: coenzyme transport and metabolism; I: lipid transport and metabolism; J: translation, ribosome structure and biogenesis; K: transcription; L: replication, recombination and repair; M: cell wall/membrane/envelope biogenesis; N: cell motility; O: posttranslational modification, protein turnover, chaperones; P: inorganic ion transport and metabolism; Q: secondary metabolites biosynthesis, transport and catabolism; R: general function prediction only; S: function unknown; T: signal transduction mechanisms; U: intracellular trafficking, secretion, and vesicular transport; V: defense mechanisms; W: extracellular structures; Y: nuclear structure; Z: cytoskeleton.

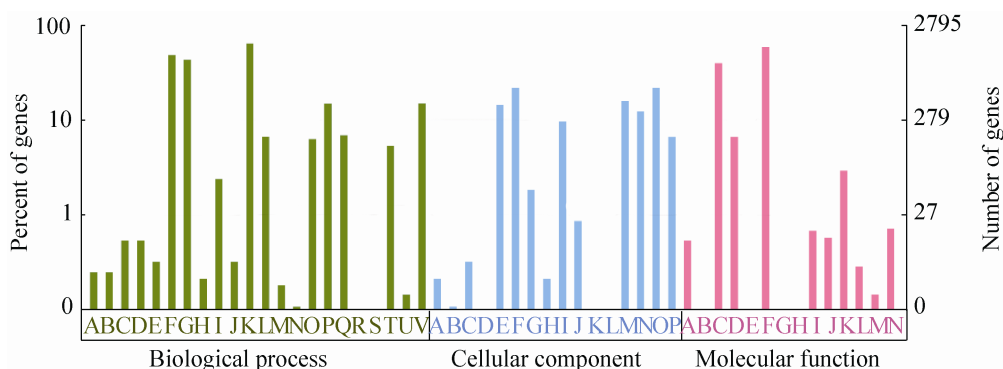


图 3. 出芽短梗霉 CCTCC M2012223 的 GO 功能聚类

Figure 3. Gene distribution based on Gene Ontology of *A. pullulans* CCTCC M2012223. Biological process: A: reproduction; B: multi-organism process; C: developmental process; D: positive regulation of biological process; E: multicellular organismal process; F: cellular process; G: single-organism process; H: growth; I: signaling; J: negative regulation of biological process; K: metabolic process; L: biological regulation; M: reproductive process; N: immune system process; O: regulation of biological process; P: establishment of localization; Q: cellular component organization or biogenesis; R: rhythmic process; S: cell killing; T: response to stimulus; U: locomotion; V: localization. Cell components: A: virion part; B: extracellular matrix; C: extracellular region part; D: collagen trimer; E: organelle; F: cell part; G: membrane-enclosed lumen; H: virion; I: membrane part; J: extracellular region; K: extracellular matrix part; L: nucleoid; M: membrane; N: macromolecular complex; O: cell; P: organelle part. Molecular function: A: electron carrier activity; B: protein binding transcription factor activity; C: binding; D: transporter activity; E: protein tag; F: catalytic activity; G: metallochaperone activity; H: nutrient reservoir activity; I: enzyme regulator activity; J: molecular transducer activity; K: structural molecule activity; L: nucleic acid binding transcription factor activity; M: receptor activity; N: antioxidant activity.

## 2.2 比较基因组学分析

**2.2.1 6株出芽短梗霉基因组的基本特征:**将6株出芽短梗霉的基因组组装和注释信息进行比较,其基本特征列于表1。出芽短梗霉 CCTCC M2012223 的基因组组装长度最长达到 30.75 Mb, 包含的 contigs 和 scaffolds 数量也最多;出芽短梗霉 CCTCC M2012223、*A. pullulans* var. *melanogenum* 和 *A. pullulans* ATCC 62921 的基因组 GC 含量低于 50.0%, 其他 3 株菌的 GC 含量高于 50.0%;从基因组注释结果来看,除出芽短梗霉 CCTCC M2012223 外,其他 5 株菌的注释基因数量均大于

10000, 其中 *A. pullulans* var. *pullulans* 的注释基因数量最多达到 11866 个,这可能与其高渗耐受力有关<sup>[1]</sup>;6株出芽短梗霉菌株的基因外显子 GC 含量在 51.79%–52.99%之间,而出芽短梗霉 CCTCC M2012223 的内含子数量最少,平均每 2.28 个基因之间有 1 个内含子。

采用 OrthoMCL 软件对 6 株出芽短梗霉菌株的氨基酸(或核苷酸)序列进行比对,选取一定阈值进行相似性聚类得到种内同源基因,如图 4 所示,6 株菌的同源基因数为 7062 个,分别占其总基因数的 74.7%、59.6%、65.4%、68.8%、66.7%和 63.5%,

表 1. 出芽短梗霉 CCTCC M2012223 与其他株的基因组组装和注释统计结果

Table 1. Genome assembly and annotation statistics for *A. pullulans* CCTCC M2012223 and other varieties

Statistic	Value per <i>A. pullulans</i> variety <sup>a</sup>					
	ApC <sup>b</sup>	ApP <sup>b</sup>	ApS <sup>b</sup>	ApN <sup>b</sup>	ApM <sup>b</sup>	ApA <sup>b</sup>
Assembly statistics						
Assembly length/Mb	30.75	29.62	25.80	25.43	26.20	28.10
Contig length total/Mb	30.75	29.59	25.79	25.43	26.20	–
Number of contigs	393	209	84	55	174	–
Contig N50	28	11	12	10	12	–
Contig L50/kb	342.22	779.84	805.54	1053.18	652.37	–
Number of scaffold	323	186	75	47	150	49
Scaffold N50	26	10	11	9	10	–
GC content/%	47.49	50.02	50.78	51.12	49.85	49.90
Gene statistics						
Number of genes	9452	11866	10809	10266	10594	11129
Protein length (amino acids, median)	421	369	377	380	372	–
Exon length (bp, median)	306	369	363	288	294	–
Gene length (bp, median)	1197	1399	1424	1266	1232	–
Intron length (bp, median)	55	57	57	57	57	–
Transcript length (bp, median)	1389	1297	1325	1140	1116	–
GC content of exon/%	51.79	52.52	52.53	52.99	51.87	52.66
Intron genes number	6883	8407	7819	7632	7797	–
Introns pre gapped gene number	2.28	2	2	2	2	–

a: part data are cited from Gostinčar et al<sup>[1]</sup>. (2014); b: ApC, *A. pullulans* CCTCC M2012223; ApP: *A. pullulans* var. *pullulans*; ApS: *A. pullulans* var. *subglaciale*; ApN: *A. pullulans* var. *namibiae*; ApM: *A. pullulans* var. *melanogenum*; ApA: *A. pullulans* ATCC 62921.

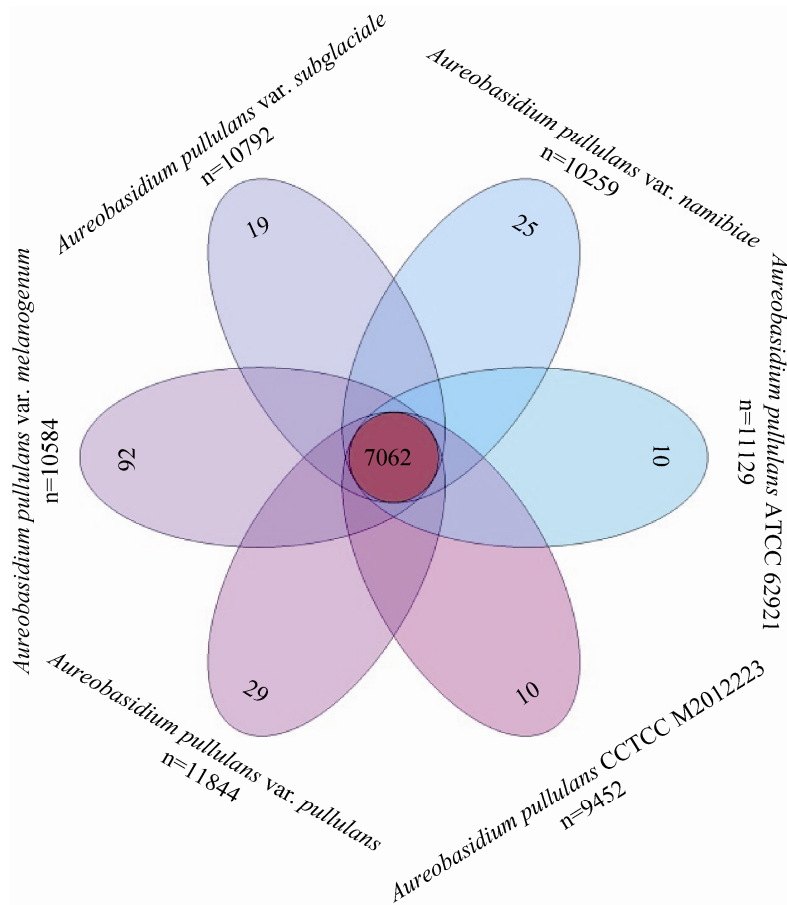


图 4. 6 株出芽短梗霉同源基因比较分析

Figure 4. Venn's diagram of *A. pullulans* varieties homologous genes.

而其特有基因数分别为 10、29、19、25、92 和 10 个, 说明这 6 株出芽短梗霉菌株具有较高的同源性, 同时在进化上也存在一定的差异。

在同源基因分析的基础上, 选取 6 株出芽短梗霉都含有且为单一拷贝的同源基因(避免旁系

同源蛋白的干扰)进行多序列比对, 构建全基因组进化树, 如图 5 所示。表明出芽短梗霉 CCTCC M2012223 在进化上与 *A. pullulans* var. *melanogenum* 的亲缘关系最近, 与 *A. pullulans* var. *pullulans* 和 *A. pullulans* ATCC 62921 的亲缘关系最远。

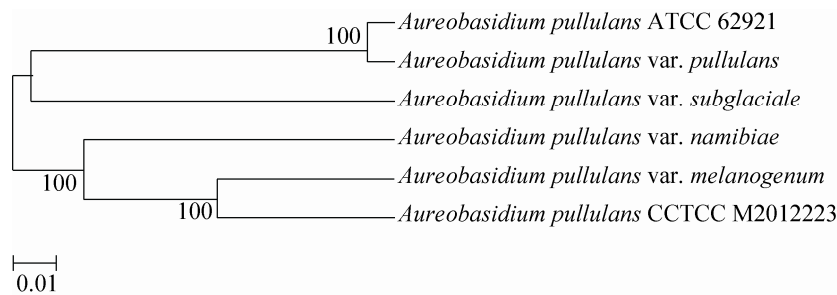


图 5. 6 株出芽短梗霉全基因组进化树分析

Figure 5. Genome-wide phylogenetic tree of *A. pullulans* varieties. The number indicates the bootstrap value.

**2.2.2 黑色素代谢相关基因分析：**出芽短梗霉能产生 1,8-二羟基萘(DHN)黑色素,对于大多数真菌来说,黑色素在多种压力条件下对细胞有保护作用<sup>[13]</sup>。微生物主要通过酪氨酸酶,漆酶,儿茶酚酶和聚酮合酶途径合成黑色素<sup>[14]</sup>,而座囊菌纲主要以乙酰辅酶 A 或丙二酰辅酶 A 作为前体通过聚酮合酶途径合成 DHN-黑色素, Gostinčar 等<sup>[1]</sup>据此鉴定了出芽短梗霉中黑色素合成途径中的相关蛋白,包括乙酰辅酶 A (acetyl-coenzyme A)、丙二酰辅酶 A 酰基载体蛋白转酰基酶(malonyl coenzyme A acyl carrier protein transacylase, MCAT)、四羟基萘/三羟基萘还原酶相似蛋白(tetrahydroxynaphthalene/trihydroxynaphthalene reductase-like protein)、萘酚还原酶(hydroxynaphthalene reductase)、以及小柱孢酮脱水酶相似蛋白(scytalone dehydratase like protein)。 *A. pullulans* var. *melanogenum* 是能分泌黑色素的一个变种<sup>[15]</sup>,与之相比,出芽短梗霉 CCTCC M2012223 不产生黑色素,通常只少量分泌浅色色素,因此,将 2 株菌黑色素合成途径相关蛋白进行了比对分析。如表 2 所示,除萘酚还原酶外,其余 3 个蛋白的氨基酸序列均发生突变,

其中 2 株菌的乙酰辅酶 A、丙二酰辅酶 A 酰基载体蛋白转酰基酶(MCAT)和小柱孢酮脱水酶相似蛋白的序列一致性分别为 83.49%、79.44%和 97.73%,表明出芽短梗霉 CCTCC M2012223 黑色素合成途径中的乙酰辅酶 A 和 MCAT 可能功能上表现出差异,导致黑色素合成可能受到限制。

**2.2.3 普鲁兰多糖合成相关基因分析：**普鲁兰多糖是由麦芽三糖为单体通过  $\alpha$ -1,6 键连接形成的多聚糖<sup>[16]</sup>。其生物合成途径中涉及尿苷二磷酸葡萄糖焦磷酸化酶(uridien diphosphoglucose pyrophosphorylase, UDPG pyrophosphorylase)  $\alpha$ -葡萄糖变位酶( $\alpha$ -phosphoglucose mutase, PPM)、普鲁兰糖合成酶(pullulan synthetase, PUL)等关键酶<sup>[17-18]</sup>。将出芽短梗霉 CCTCC M2012223 中普鲁兰多糖合成相关酶蛋白序列进行同源性分析(图 6)。结果表明 6 株菌的 PPM、UDPG-pyrophosphorylase、PUL 序列一致性分别达到 98.77%、99.22%和 85.81%,且出芽短梗霉 CCTCC M2012223 和 *A. pullulans* var. *melanogenum* 的相应蛋白序列一致性最高。

表 2. 出芽短梗霉 CCTCC M2012223 与 *A. pullulans* var. *melanogenum* 黑色素合成途径相关蛋白比对分析  
Table 2. Comparative analysis of proteins involved in melanin pathway between *A. pullulans* CCTCC M2012223 and *A. pullulans* var. *melanogenum*

Proteins	Function	Sequence Identity/%	Mutant form
Acetyl-coenzyme A	The common intermediate of catabolism, a precursor of melanin synthesis	83.49	99 amino acids insertion mutations in C-terminal, 7 amino acid mutations
Malonyl coenzyme A acyl carrier protein transacylase (MCAT)	Transfer malonic acid groups of malonyl coenzyme A to the mercaptan of acyl carrier protein (ACP)	79.44	More than 50 amino acids mutations, 9 amino acids insertion mutations
Hydroxynaphthalene reductase	Catalytic 1,3,6,8-tetrahydroxynaphthalene to generate scytalone	100	No amino acid mutation
Scytalone dehydratase like protein	Catalytic scytalone to generate trihydroxynaphthalene	1,3,8-97.73	3 amino acid mutations: Ser→Ala, Val→Ile, Lys→Gln



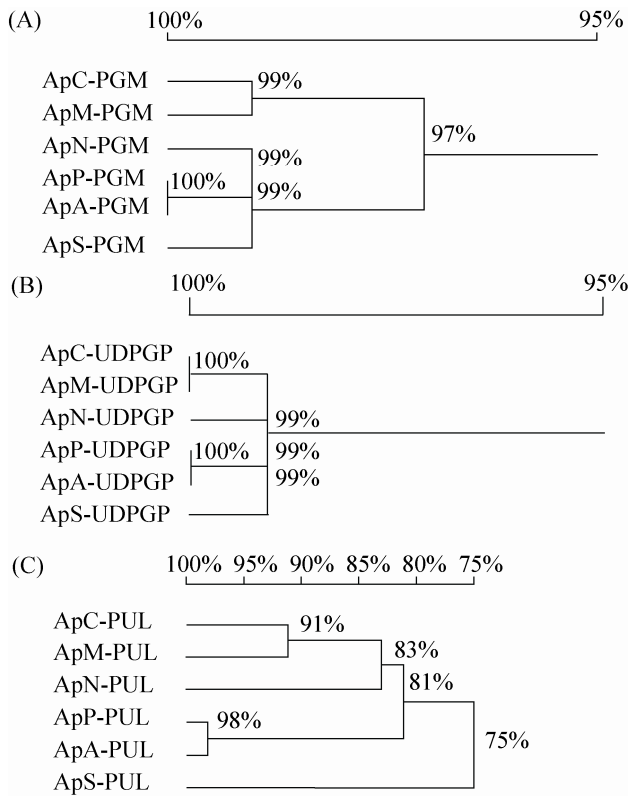


图 6. 6 株出芽短梗霉普鲁兰多糖合成相关基因的蛋白序列同源性分析

Figure 6. Homology analysis of proteins sequences involved in pullulan biosynthesis in *A. pullulans* varieties. A:  $\alpha$ -phosphoglucose mutase (PGM); B: UDPG-pyrophosphorylase (UDPGP); C: pullulan synthetase (PUL). The percentage indicates the sequence identity.

**2.2.4 聚苹果酸聚合相关基因分析：**聚苹果酸 (Polymalic acid, PMA)是以 L-苹果酸单体聚合而成的聚酯型聚合物<sup>[19]</sup>, 聚苹果酸的前体 L-苹果酸的来源主要有 3 条途径: TCA 循环、乙醛酸循环和胞质丙酮酸羧化途径<sup>[20]</sup>, 而 L-苹果酸聚合生成聚苹果酸的聚合途径可能涉及苹果酰辅酶 A 连接酶(malate-CoA ligase, MCL)和聚苹果酸合成酶 (polymalic acid synthase, PAS)的参与<sup>[21]</sup>。本课题组前期已从出芽短梗霉 CCTCC M2012223 中克隆

得到 MCL 基因<sup>[5]</sup>, 通过蛋白质序列比对得到其他 5 株出芽短梗霉的 MCL 蛋白, 进行同源性分析, 序列一致性达到 97.80%, 且出芽短梗霉 CCTCC M2012223 和 *A. pullulans* var. *melanogenum* 的 MCL 蛋白序列一致性最高(图 7); 而 PAS 的相关序列信息尚未见报道。

### 3 讨论

出芽短梗霉是一种遗传多样性真菌, 代谢产物丰富。本研究通过全基因组测序获得了出芽短梗霉 CCTCC M2012223 的基因组序列和注释信息, 结果表明其基因组序列全长 30756831 bp, GC 含量 47.49%, 编码 9452 个基因, 与另外已测序完成的 5 株出芽短梗霉菌株进行比较基因组分析, 发现其基因组的基本特征与其他菌株相似度高, 6 株菌的同源基因数达到 7092 个, 同时也存在各自不同数量的特有基因。全基因组进化分析表明出芽短梗霉 CCTCC M2012223 与 *A. pullulans* var. *melanogenum* 的亲缘关系最近, 与 *A. pullulans* var. *pullulans* 和 *A. pullulans* ATCC 62921 的亲缘关系

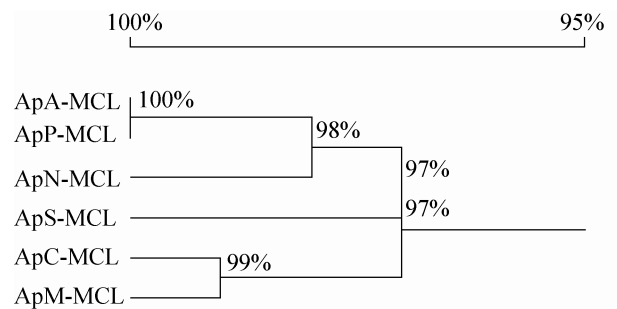


图 7. 6 株出芽短梗霉 MCL 基因的蛋白序列同源性分析

Figure 7. Homology analysis of MCL proteins in *A. pullulans* varieties. The percentage indicates the sequence identity.

最远，也表明不同菌株在进化过程中存在遗传不稳定性。

出芽短梗霉 CCTCC M2012223 是课题组筛选的 1 株聚苹果酸高产菌株，已在合作企业投入生产，但在发酵过程中也存在少量副产物普鲁兰多糖和浅色素，影响产物得率和质量外观。对该菌株的聚苹果酸、普鲁兰多糖和黑色素合成相关蛋白与其余 5 株已测序菌株进行蛋白序列比对发现，6 株菌的聚苹果酸和普鲁兰多糖合成相关蛋白表现出较高的保守性，而出芽短梗霉 CCTCC M2012223 与 *A. pullulans* var. *melanogenum* 尽管亲缘关系最近，但黑色素合成相关基因及表型表现出差异。黑色素是大多数出芽短梗霉代谢产生的一种色素，对细胞起到保护作用，然而发酵过程产生黑色素通常会造成本代谢产物下游分离提取困难，因此选育不产黑色素的出发菌株具有潜在的工业应用价值。*A. pullulans* var. *melanogenum* 是出芽短梗霉分泌黑色素的一个变种<sup>[15]</sup>，尽管出芽短梗霉 CCTCC M2012223 与之亲缘关系最近，但黑色素合成途径中除萘酚还原酶外，乙酰辅酶 A、丙二酰辅酶 A 酰基载体蛋白转酰基酶等蛋白氨基酸序列存在一些插入和突变，可能导致其色素合成代谢产生变化。

出芽短梗霉 CCTCC M2012223 基因组测序的完成，将为其菌株代谢工程改造及调控机制研究提供了大量的生物学信息。对聚苹果酸代谢产物而言，通过基因破坏、CRISPR/Cas9 基因编辑等技术对副产物普鲁兰多糖和黑色素的去除，以及聚苹果酸合成功能基因挖掘则成为可能，本课题组也正在开展相关的研究，并取得了一定进展。随着出芽短梗霉功能基因组学研究的不断深入，

将会获得更多优良性状的基因工程菌株，推进聚苹果酸及苹果酸发酵的产业化进程。

## 参 考 文 献

- [1] Gostinčar C, Ohm RA, Kogej T, Sonjak S, Turk M, Zajc J, Zalar P, Grube M, Sun H, Han J, Sharma A, Chiniquy J, Ngan CY, Lipzen A, Barry K, Grigoriev IV, Gunde-Cimerman N. Genome sequencing of four *Aureobasidium pullulans* varieties: biotechnological potential, stress tolerance, and description of new species. *BMC Genomics*, 2014, 15(1): 549.
- [2] Li Y, Chi Z, Wang GY, Wang ZP, Liu GL, Lee CF, Ma ZC, Chi ZM. Taxonomy of *Aureobasidium* spp. and biosynthesis and regulation of their extracellular polymers. *Critical Reviews in Microbiology*, 2015, 41(2): 228–237.
- [3] Chi ZM, Wang F, Chi Z, Yue LX, Liu GL, Zhang T. Bioproducts from *Aureobasidium pullulans*, a biotechnologically important yeast. *Applied Microbiology and Biotechnology*, 2009, 82(5): 793–804.
- [4] Zou X, Yang J, Tian X, Guo MJ, Li ZH, Li YZ. Production of polymalic acid and malic acid from xylose and corncob hydrolysate by a novel *Aureobasidium pullulans* YJ 6-11 strain. *Process Biochemistry*, 2016, 51(1): 16–23.
- [5] Wu XY, Zhou F, Tu GW, Zou X. Gene cloning, expression and characterization of malate-CoA ligase in the polymerization pathway of polymalic acid from *Aureobasidium pullulans*. *Acta Microbiologica Sinica*, 2014, 54(8): 919–925. (in Chinese)  
吴小燕, 周峰, 涂光伟, 邹祥. 聚苹果酸聚合途径中苹果酰辅酶 A 连接酶基因的克隆、表达及酶学性质. *微生物学报*, 2014, 54(8): 919–925.
- [6] Zou X, Wang YK, Tu GW, Zan ZQ, Wu XY. Adaptation and transcriptome analysis of *Aureobasidium pullulans* in corncob hydrolysate for increased inhibitor tolerance to malic acid production. *PLoS One*, 2015, 10(3): e0121416.
- [7] Tu GW, Wang YK, Feng J, Li XR, Guo MJ, Zou X. *Agrobacterium tumefaciens*-mediated transformation of *Aureobasidium pullulans* and high-efficient screening for

- polymalic acid producing strain. *Chinese Journal Biotechnology*, 2015, 31(7): 1063–1072. (in Chinese)
- 涂光伟, 王永康, 冯骏, 李晓荣, 郭美锦, 邹祥. 农杆菌介导的出芽短梗霉遗传转化及高效筛选聚苹果酸高产菌株. *生物工程学报*, 2015, 31(7): 1063–1072.
- [8] Zou X, Zhou YP, Yang ST. Production of polymalic acid and malic acid by *Aureobasidium pullulans* fermentation and acid hydrolysis. *Biotechnology and Bioengineering*, 2013, 110(8): 2105–2113.
- [9] Kanehisa M. Molecular network analysis of diseases and drugs in KEGG//Mamitsuka H, DeLisi C, Kanehisa M. Data Mining for Systems Biology. New York: Humana Press, 2013: 263–275.
- [10] Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, 2000, 28(1): 33–36.
- [11] Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, 1997, 25(5): 955–964.
- [12] Lagesen K, Hallin P, Rødland EA, Stærfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, 2007, 35(9): 3100–3108.
- [13] Gostinčar C, Muggia L, Grube M. Polyextremotolerant black fungi: oligotrophism, adaptive potential, and a link to lichen symbioses. *Frontiers in Microbiology*, 2012, 3: 390.
- [14] Taborda CP, da Silva MB, Nosanchuk JD, Travassos LR. Melanin as a virulence factor of *Paracoccidioides brasiliensis* and other dimorphic pathogenic fungi: a minireview. *Mycopathologia*, 2008, 165(4/5): 331–339.
- [15] Zalar P, Gostinčar C, de Hoog GS, Uršič V, Sudhadham M, Gunde-Cimerman N. Redefinition of *Aureobasidium pullulans* and its varieties. *Studies in Mycology*, 2008, 61: 21–38.
- [16] Cheng KC, Demirci A, Catchmark JM. Pullulan: biosynthesis, production, and applications. *Applied Microbiology and Biotechnology*, 2011, 92(1): 29–44.
- [17] Catley BJ, McDowell W. Lipid-linked saccharides formed during pullulan biosynthesis in *Aureobasidium pullulans*. *Carbohydrate Research*, 1982, 103(1): 65–75.
- [18] Ma ZC, Chi Z, Geng Q, Zhang F, Chi ZM. Disruption of the pullulan synthetase gene in siderophore-producing *Aureobasidium pullulans* enhances siderophore production and simplifies siderophore extraction. *Process Biochemistry*, 2012, 47(12): 1807–1812.
- [19] Ouchi T, Fujino A. Synthesis of poly( $\alpha$ -malic acid) and its hydrolysis behavior *in vitro*. *Die Makromolekulare Chemie*, 1989, 190(7): 1523–1530.
- [20] Zelle RM, de Hulster E, van Winden WA, de Waard P, Dijkema C, Winkler AA, Geertman JMA, van Dijken JP, Pronk JT, van Maris AJA. Malic acid production by *Saccharomyces cerevisiae*: engineering of pyruvate carboxylation, oxaloacetate reduction, and malate export. *Applied and Environmental Microbiology*, 2008, 74(9): 2766–2777.
- [21] Willibald B, Bildl W, Lee BS, Holler E. Is  $\beta$ -poly(L-malate) synthesis catalysed by a combination of  $\beta$ -L-malyl-AMP-ligase and  $\beta$ -poly(L-malate) polymerase? *European Journal of Biochemistry*, 1999, 265(3): 1085–1090.

# Complete genome sequencing of polymalic acid-producing strain *Aureobasidium pullulans* CCTCC M2012223

Yongkang Wang<sup>1</sup>, Xiaodan Song<sup>1</sup>, Xiaorong Li<sup>1</sup>, Shang-tian Yang<sup>2</sup>, Xiang Zou<sup>1\*</sup>

<sup>1</sup> Chongqing Engineering Research Center for Pharmaceutical Process and Quality Control, College of Pharmaceutical Sciences, Southwest University, Chongqing 400715, China

<sup>2</sup> William G. Lowrie Department of Chemical and Biomolecular Engineering, The Ohio State University, Columbus, OH 43210, USA

**Abstract:** [Objective] To explore the genome sequence of *Aureobasidium pullulans* CCTCC M2012223, analyze the key genes related to the biosynthesis of important metabolites, and provide genetic background for metabolic engineering. [Methods] Complete genome of *A. pullulans* CCTCC M2012223 was sequenced by Illumina HiSeq high throughput sequencing platform. Then, fragment assembly, gene prediction, functional annotation, and GO/COG cluster were analyzed in comparison with those of other five *A. pullulans* varieties. [Results] The complete genome sequence of *A. pullulans* CCTCC M2012223 was 30756831 bp with an average GC content of 47.49%, and 9452 genes were successfully predicted. Genome-wide analysis showed that *A. pullulans* CCTCC M2012223 had the biggest genome assembly size. Protein sequences involved in the pullulan and polymalic acid pathway were highly conservative in all of six *A. pullulans* varieties. Although both *A. pullulans* CCTCC M2012223 and *A. pullulans* var. *melanogenum* have a close affinity, some point mutation and inserts were occurred in protein sequences involved in melanin biosynthesis. [Conclusion] Genome information of *A. pullulans* CCTCC M2012223 was annotated and genes involved in melanin, pullulan and polymalic acid pathway were compared, which would provide a theoretical basis for genetic modification of metabolic pathway in *A. pullulans*.

**Keywords:** *Aureobasidium pullulans*, genome-wide comparative analysis, complete genome sequencing, metabolic mechanism, polymalic acid

(本文责编: 李磊)

---

Supported by the National Natural Science Foundation of China (31571816), by the National High Technology Research and Development Program of China (863 Program) (2015AA021005, 2014AA021205), by the Chongqing Social and People's Livelihood Guarantee Special Program (cstc2016shmszx80075), and by the Chongqing Intellectual Property Special Foundation Program (CQIPO2015224)

\*Corresponding author. Tel: +86-23-68251225; Fax: +86-23-68251048; E-mail: zhx1030@swu.edu.cn

Received: 15 May 2016; Revised: 21 June 2016; Published online: 1 July 2016