

# *Ruminiclostridium papyrosolvans* DSM2782的全基因组测序及生物功能挖掘

尤梦成<sup>1,2</sup>, 邱子立<sup>3</sup>, 许成钢<sup>2</sup>, 任振兴<sup>3\*</sup>

(1. 山西大学 生物技术研究所, 化学生物学与分子工程教育部重点实验室, 山西 太原 030006;

2. 浙江农林大学 动物科技学院·动物医学院, 浙江 杭州 311300;

3. 山西大学 应用化学研究所, 山西 太原 030006)

**摘要:** *Ruminiclostridium papyrosolvans* (溶纸梭菌, *R. papyrosolvans*) DSM2782 作为嗜中温产纤维小体厌氧梭菌中进化程度最高的纤维素降解梭菌, 由于其基因组信息缺失会影响进一步的实验探究。为了能够有效地挖掘 *R. papyrosolvans* DSM2782 的生物学功能, 并探究其是否具有合成有活性的次级代谢产物的潜力。采用二代 Illumina 联合三代 PacBio 测序技术, 对 *R. papyrosolvans* DSM2782 进行全基因组测序和功能分析, 并与同类型的嗜中温产纤维小体梭菌进行比较基因组学分析, 最后对其不同碳源下的发酵液采用液相色谱-质谱联用技术 (Liquid Chromatography-Mass Spectrometry, LC-MS) 进行非靶向代谢检测。研究结果表明 *R. papyrosolvans* DSM2782 基因组序列全长为 5 027 861 bp, GC 含量为 37.1%。将完整的基因组序列提交至美国国家生物技术信息中心 (National Center for Biotechnology Information, NCBI) 获得 GenBank 登录号为 CP119677.1。通过原核生物基因组注释工具 (Prokaryotic Genome Annotation Pipeline, PGAP) 共注释了 4 274 个蛋白编码基因、4 个 ncRNA、24 个 rRNA 以及 62 个 tRNA。为进一步对 *R. papyrosolvans* DSM2782 基因组的功能进行分析, 使用直系同源簇 (Cluster of Orthologous Group, COG)、基因本体 (Gene Ontology, GO)、京都基因与基因组百科全书 (Kyoto Encyclopedia of Genes and Genomes, KEGG) 数据库对其编码基因序列进行注释, 分别有 3 503、1 301、2 009 个编码基因被注释。通过 dbCAN3 数据库在 *R. papyrosolvans* DSM2782 基因组中检测到 231 个不同的碳水化合物活性水解酶。通过与其他梭菌属基因组进行 ANI 比较, 结果显示 *R. papyrosolvans* DSM2782 与 *R. sp.* BNL1100、*R. cellulolyticum* H10 以及 *R. papyrosolvans* C7 具有较高的同源性。同时对 *R. papyrosolvans* DSM2782 与 *R. papyrosolvans* C7 进行共线性分析, 结果表明 *R. papyrosolvans* DSM2782 与 *R. papyrosolvans* C7 之间的大部分区域的序列高度同源。利用生物合成基因预测软件 antiSMASH 预测 *R. papyrosolvans* DSM2782 基因组中存在 23 个生物合成基因簇。通过非靶向代谢组学分析, 发现 *R. papyrosolvans* DSM2782 可能产生 4-羟基苯甲酸、尼泊金丙酯、丁酸、四氢嘧啶、去甲哈尔满、左旋多巴等物质, 这也表明 *R. papyrosolvans* DSM2782 能够产生具有活性的次级代谢产物。*R. papyrosolvans* DSM2782 全基因组测序以及代谢组学结果表明其具有降解木质纤维素的功能, 同时也能够产生多种具有活性的次级代谢产物。本研究为后续深入开展 *R. papyrosolvans* DSM2782 生物功能研究提供了丰富的基因组信息和基因资源。

**关键词:** 溶纸梭菌; 全基因组测序; 碳水化合物活性水解酶; 生物合成基因簇

中图分类号: Q93

文献标志码: A

文章编号: 0253-2395(2025)02-0416-13

收稿日期: 2024-08-31; 接受日期: 2024-12-30

基金项目: 国家自然科学基金 (32070045; 32170053)

作者简介: 尤梦成 (1995—), 男, 河南信阳人, 博士, 研究方向为梭菌的遗传代谢机制。E-mail: 1351382072@qq.com

\* 通信作者: 任振兴 (REN Zhenxing), E-mail: zxren@sxu.edu.cn

引文格式: 尤梦成, 邱子立, 许成钢, 等. *Ruminiclostridium papyrosolvans* DSM2782 的全基因组测序及生物功能挖掘 [J]. 山西大学学报 (自然科学版), 2025, 48(2): 416-428. DOI: 10.13451/j.sxu.ns.2024178.

## Whole Genome Sequencing Analysis and Function Mining of *Ruminiclostridium papyrosolvens* DSM2782

YOU Mengcheng<sup>1,2</sup>, QIU Zili<sup>3</sup>, XU Chenggang<sup>2</sup>, REN Zhenxing<sup>3\*</sup>

(1. Key Laboratory of Chemical Biology and Molecular Engineering of Ministry of Education, Institute of Biotechnology, Taiyuan 030006, China;

2. College of Animal Science and Technology & College of Veterinary Medicine, Zhejiang Agriculture and Forestry University, Hangzhou 311300, China;

3. Institute of Applied Chemistry, Shanxi University, Taiyuan 030006, China)

**Abstract:** *Ruminiclostridium papyrosolvens* (*R. papyrosolvens*) DSM2782 is the most evolved cellulose-degrading clostridia among the mesophilic cellulolytic anaerobic bacteria, however, the lack of genomic information will affect further experimental research. In order to effectively explore the biological functions and the potential to synthesize active secondary metabolites of *R. papyrosolvens* DSM2782, the whole genome of *R. papyrosolvens* DSM2782 was sequenced and analyzed by combination of the second-generation Illumina and the third-generation PacBio sequencing platform, and comparative genomic analysis was performed with the same type of mesophilic cellulolytic clostridium. Finally, the fermentation broth was subjected to liquid chromatography–mass spectrometry (LC-MS) non-targeted metabolic detection. The results showed that the complete genome of *R. papyrosolvens* DSM2782 is a single circular 5 027 861-bp chromosome with 37.1% GC content. The genome sequence was submitted to National Center for Biotechnology Information (NCBI) to obtain the GenBank accession number CP119677.1. In total, 4 274 protein-coding genes, 4 ncRNA, 24 rRNA, and 62 tRNA were annotated by prokaryotic genome annotation pipeline (PGAP). Furthermore, to further analyze the function of *R. papyrosolvens* DSM2782 genome, the whole genome sequence was gene-annotated using Cluster of Orthologous Group (COG), Gene Ontology (GO), and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases, and 3 503, 1 301, and 2 009 coding genes were annotated, respectively, 231 different carbohydrate-active hydrolases were detected in *R. papyrosolvens* DSM2782 using the dbCAN3 database. Compared with other clostridia genomes by ANI, the results showed that *R. papyrosolvens* DSM2782 shared high homology with *R. sp.* BNL1100, *R. cellulolyticum* H10, and *R. papyrosolvens* C7. At the same time, collinearity analysis showed that the sequences of most regions between *R. papyrosolvens* DSM2782 and *R. papyrosolvens* C7 were highly homologous. Moreover, the biosynthetic gene prediction software antiSMASH was used to predict the presence of 23 biosynthetic gene clusters in the genome of *R. papyrosolvens* DSM2782. Through non-targeted metabolism analysis in different carbon sources, it was found that *R. papyrosolvens* DSM2782 may produce substance such as 4-Hydroxybenzoic acid, propylparaben, butyric acid, Ectoine, norharman, and 3,4-Dihydroxyphenylalanine, which also indicates that *R. papyrosolvens* DSM2782 can produce active secondary metabolism. The whole gene sequencing and metabolomics results of *R. papyrosolvens* DSM2782 showed that it has the function of degrading lignocellulose and also produce a variety of active secondary metabolites. This research provides genomic information and gene resources for the subsequent in-depth development of the biocatalytic functions of *R. papyrosolvens* DSM2782.

**Key words:** *Ruminiclostridium papyrosolvens*; whole genome sequencing; CAZymes; biosynthetic gene cluster

### 0 引言

作为世界上最丰富的可持续性的自然资源,木质纤维素是化石能源的一种有前途的替代品,可以缓解化石燃料面临的资源枯竭问题<sup>[1]</sup>。但是木质纤维素结构复杂,难以降解,从而限制了其开发利用。嗜中温厌氧纤维素降解梭菌,如 *Ruminiclostridium papyrosolvens* (*R. papyrosolvens*) DSM2782<sup>[2-3]</sup>,能够分泌高效降解纤维素的纤维素酶复合体-纤维小体( cellulose-

some)<sup>[4-7]</sup>。因此,阐明这类嗜中温厌氧纤维素降解梭菌通过纤维小体降解纤维素的机制,对于通过这类纤维素降解细菌对木质纤维素进行有效降解并生产清洁能源和具有附加价值的化学品具有重要的研究意义。

另一方面,这类嗜中温厌氧纤维素降解梭菌相对于其他致病梭菌和产溶剂梭菌,含有更多的生物合成基因簇。我们通过基因组分析比较发现纤维素降解梭菌基因组中含有编码次级代谢产物生物合成基因簇的数量远多于致病梭菌和

产溶剂梭菌。研究人员陆续从非致病性的厌氧梭菌中发现了许多具有临床价值的天然产物,例如从 *R. cellulolyticum* H10 中分离出来的聚合硫酰胺化合物-Closthioamide<sup>[8-9]</sup>,是人类首次从厌氧纤维素降解梭菌中分离出来的具有抗耐甲氧西林的金黄色葡萄球菌以及耐万古霉素的肠球菌的天然活性产物<sup>[10]</sup>。这一研究成果揭开了从厌氧纤维素降解梭菌中发现天然产物的序幕。

本研究利用二代 Illumina 联合三代 PacBio 测序技术对 *R. papyrosolvans* DSM2782 进行全基因组测序。通过 COG (Cluster of Orthologous Group)、GO (Gene Ontology)、KEGG (Kyoto Encyclopedia of Genes and Genomes)、CAZyme 以及 anti-SMASH 等数据库对 *R. papyrosolvans* DSM2782 进行基因功能注释和生物合成基因簇的预测分析,并通过非靶向代谢组学探究 *R. papyrosolvans* DSM2782 是否能够产生具有活性的次级代谢产物。以此为深入开展该类型的具有产纤维小体的木质纤维素降解特性的厌氧梭菌产生具有活性的天然产物的研究奠定基础。

## 1 材料和方法

### 1.1 实验材料

#### 1.1.1 实验菌种培养

本研究所用的 *R. papyrosolvans* DSM2782 来自于本实验室挑取的 *R. papyrosolvans* DSM2782<sup>[3]</sup> 单克隆菌株。先通过平板划线法从培养 *R. papyrosolvans* DSM2782 的 GS-2 固体厌氧培养基<sup>[11]</sup> 平板上挑取 *R. papyrosolvans* DSM2782 的单克隆菌落,接种至 5 mL 的 GS-2 液体厌氧培养基中,35 °C 培养 1 d,取 1.5 mL 菌液接种于 50 mL 的 GS-2 液体厌氧培养基中,35 °C 培养 1 d,收集菌体并放置 -80 °C 超低温冰箱备用。

#### 1.1.2 仪器和设备

厌氧培养箱,龙跃仪器设备有限公司;立式压力蒸汽灭菌器,江阴滨江医疗设备有限公司;电热恒温培养箱,上海精宏实验设备有限公司;超净工作台,上海博讯医疗生物仪器股份有限公司;移液枪,艾本德国际贸易有限公司;pH 计,奥豪斯国际贸易有限公司;超微量紫外分光光度计,英国柏楛有限公司;高速台式

冷冻离心机,艾本德国际贸易有限公司等。

### 1.2 实验方法

#### 1.2.1 GS-2 厌氧培养基配制

无水磷酸二氢钾 1.5 g,无水磷酸氢二钾 3.8 g,尿素 2.1 g,*L*-半胱氨酸 1.0 g,MOPs 10.0 g,酵母提取物 6.0 g,二水合柠檬酸三钠 3.0 g,六水氯化镁 1.0 g,无水氯化钙 0.11 g,七水硫酸亚铁 0.001 34 g,0.1% 刃天青 500 μL,加入蒸馏水溶解,定容至 1 000 mL,并调整 pH 为 7.4(固体培养基另需要加入琼脂 15 g)。通过厌氧培养箱对培养基溶液进行除氧操作,待除氧完成后进行各种规格的液体培养基的分装,最后通过高压灭菌锅在 121 °C 下对 GS-2 厌氧培养基进行 20 min 高压蒸汽灭菌,待液体厌氧培养基冷却备用。

#### 1.2.2 *R. papyrosolvans* DSM2782 的 DNA 提取

通过 FastPure Bacteria DNA Isolation MiniKit 试剂盒 (Vazyme) 提取收集好的 *R. papyrosolvans* DSM2782 单克隆菌株的基因组 DNA。使用超微量分光光度计检测提取的 DNA 的纯度和浓度。DNA 样品检测合格后,通过干冰将提取的 *R. papyrosolvans* DSM2782 基因组 DNA 以及收集好的 *R. papyrosolvans* DSM2782 的菌体运输至派森诺生物科技有限公司 (Personalbio) 进行全基因组测序。

#### 1.2.3 *R. papyrosolvans* DSM2782 基因组测序、质控与组装

使用上海派森诺科技股份有限公司的 Illumina NovaSeq 和 PacBio Sequel 测序平台对 *R. papyrosolvans* DSM2782 进行全基因组测序,使用 AdapterRemoval<sup>[12]</sup> 以及 SOAPec<sup>[13]</sup> 对二代数据进行清洗。同时使用 Canu 软件<sup>[14]</sup> 对由 PacBio 获得的下机数据进行拼接,得到 contig 序列。将清洗后的二代高质量测序数据使用 Pilon<sup>[15]</sup> 软件对通过 PacBio 平台获取的 contigs 结果进行校正,最后使用 Unicycler<sup>[16]</sup> 软件拼接得到完整的基因组序列。

#### 1.2.4 基因组序列分析

将完整的 *R. papyrosolvans* DSM2782 基因组序列上传至 NCBI (National Center for Biotechnology Information) 数据库,通过 NCBI 数据库自带的注释工具 PGAP (Prokaryotic Genome Annota-

tion Pipeline)<sup>[17]</sup>对 *R. papyrosolvens* DSM2782 完整的基因组序列进行注释。使用 COG 数据库<sup>[18]</sup>对 *R. papyrosolvens* DSM2782 基因组序列进行同源蛋白簇功能注释;使用 dbCAN3 数据库<sup>[19]</sup>对 *R. papyrosolvens* DSM2782 基因组序列进行注释;使用 FastANI 软件<sup>[20]</sup>对 *R. papyrosolvens* DSM2782 的基因组序列以及其他类型的梭菌基因组序列进行对比分析,探究 *R. papyrosolvens* DSM2782 与参考菌株之间的同源关系。利用 MCScanX<sup>[21]</sup>对 *R. papyrosolvens* DSM2782 与 *R. papyrosolvens* C7 进行共线性分析,通过 pfam 数据库<sup>[22]</sup>预测 cohesin 和 dockerin 结构域,使用 Easyfig<sup>[23]</sup>对 *R. papyrosolvens* DSM2782 与 *R. papyrosolvens* C7 含有的纤维小体基因簇 *cip-cel* 以及 *xyl-doc* 进行同源性比对分析。使用 antiSMASH<sup>[24]</sup>在线预测 *R. papyrosolvens* DSM2782 中的生物合成基因簇,并统计 *R. papyrosolvens* DSM2782 基因组中存在的生物合成基因簇。最后,将在不同碳源下培养的 *R. papyrosolvens* DSM2782 的发酵液通过 Vanquish (Thermo Fisher Scientific)超高效液相色谱仪进行检测,通过 Waters ACQUITY UPLC BEH Amide (2.1 mm × 100 mm, 1.7 μm) 色谱柱以及 Phenomenex Kinetex C18 (2.1 mm × 100 mm 2.6 μm) 色谱柱对目标化合物进行色谱分离,进样体积均为 2 μL。使用 Xcalibur 软件控制 Orbitrap Exploris 120 质谱仪进行一级、二级质谱数据采集。使用 ProteoWizard 软件将质谱数据转成 mzXML 格式,与 BiotreeDB 数据库进行对比,得到代谢物的鉴定结果和相对定量值。从而探究不同的碳源是否能够诱导 *R. papyrosolvens* DSM2782 合成具有生物活性的次级代谢产物。

## 2 结果

### 2.1 *R. papyrosolvens* DSM2782 基因组组装结果和基因组分析

利用 Illumina NovaSeq 二代测序技术和 PacBio Sequel 三代测序技术相结合的测序方式对 *R. papyrosolvens* DSM2782 进行测序(表 1),通过 Illumina NovaSeq 二代测序技术共得到 6 893 190 条读长(reads),总计 1 033 978 500 bp。通过 PacBio Sequel 三代测序技术共得到 196 472

条 reads,总计 1 947 591 954 bp, Reads N90 为 7 187 bp, reads N50 为 10 890 bp。将测序得到的原始数据经过过滤分析过滤掉 0.75% 的低质量 reads 之后,最终得到高质量的碱基 1 016 433 614 bp。基于 Illumina 平台的二代测序结果和 PacBio 平台的三代测序结果对 *R. papyrosolvens* DSM2782 的 Clean Data 进行组装,最终得到一条全长为 5 027 861 bp 的环状基因组序列,GC 含量为 37.1% (图 1)。将所得的 *R. papyrosolvens* DSM2782 完整的基因组序列上传至 NCBI 数据库,获得 GenBank 登录号为 CP119677.1。基因注释结果显示 *R. papyrosolvens* DSM2782 的蛋白编码基因数量为 4 274 个,此外还注释 *R. papyrosolvens* DSM2782 含有 4 个 ncRNA、24 个 rRNA 和 62 个 tRNA。

### 2.2 COG 数据库注释结果分析

在 *R. papyrosolvens* DSM2782 基因组中,一共有 3 503 个基因在 COG 数据库中被注释(图 2)。结果显示在 *R. papyrosolvens* DSM2782 中共存在 4 大类、22 小类功能基因。其中未知功能(S)最多,占总数的 19.13%。其他功能主要集中在转录(K, 9.53%)、碳水化合物的运输和代谢(G, 7.83%)、氨基酸运输和代谢(E, 6.66%)、能量产生和转换(C, 6.31%)、复制、重组和修饰(L, 6.29%)、无机离子运输和代谢(P, 5.29%)、信号转导机制(T, 5.21%)、细胞壁/膜/包膜生物发生(M, 5.15%)、翻译、核糖体结构和生物发生(J, 5.05%)、辅酶运输和代谢(H, 4.48%)等方面。特别地,在 *R. papyrosolvens* DSM2782 基因组中,注释为转录功能的基因数量有 353 个,说明 *R. papyrosolvens* DSM2782 能够较好地适应外界环境的变化。此外,碳水化合物的运输和代谢功能基因有 290 个,反映其具有高效降解木质纤维素的能力。同时还发现 *R. papyrosolvens* DSM2782 在细胞壁、细胞膜的生物反应方面有 191 个基因得到注释,暗示 *R. papyrosolvens* DSM2782 具有生成生物膜的能力。

### 2.3 GO 注释结果分析

通过 GO 数据库对 *R. papyrosolvens* DSM2782 全基因组进行注释(图 3)。结果显示在生物学过程中,共有 1 134 个基因被注释到,translation

**表1** *R. papyrosolvans* DSM2782基因组组装信息  
**Table 1** The assembled information of the genome of *R. papyrosolvans* DSM2782

Sequence platform	Illumina NovaSeq
Total reads number	6 893 190
Total sequence length/bp	1 033 978 500
N rate/%	0.0034
GC Content/%	37.16
Q20 rate/%	97.59
Q30 rate/%	92.73
Average reads number	150
Sequence platform	PacBio Sequel
Total reads number	196 472
Total sequence length/bp	1 947 591 954
N20/bp	13 919
N20 number	21 515
N50/bp	10 890
N50 number	69 392
N90/bp	7 187
N90 number	155 010
N number	0
N rate/%	0
GC Content/%	38.42
High quantity reads number	6 841 542
High quantity sequence length/bp	1 016 433 614
GC_Content	37.2
Q20 rate/%	97.97
Q30 rate/%	93.27
Total length/bp	5 027 861
GC Content/%	37.1
ORF number	4 274
ncRNA number	4
rRNA number	24
tRNA number	62

以及 carbohydrate metabolic process 拥有最多的基因注释数量,分别为 57、50 个,表明 *R. papyrosolvans* DSM2782 自身的复杂性以及碳代谢的多样性。在细胞组分中,有 346 个基因被注释到,membrane 以及 cytoplasm 占主导地位,分别为 77、50 个,表明其在物质的结合、转运和催化上都有较强的能力。在分子功能中,有 1 485 个基因被注释到,这个数值在该大类甚至在整個注释结果中数量都是最多的,其中 ATP binding、DNA binding 以及 hydrolase activity, hydrolyzing O-glycosyl compounds 的数量分别为 92、72 和 60 个,显示出该菌的蛋白功能主要表现在细胞组成、遗传信息调控和能量代谢等方面,

反映出 *R. papyrosolvans* DSM2782 的生物活跃性。

#### 2.4 KEGG 注释结果分析

通过 KEGG 数据库对 *R. papyrosolvans* DSM2782 全基因组序列进行注释,共发现 7 大类功能基因(图 4)。在注释的代谢途径中,最主要的途径是氨基酸代谢、蔗糖和淀粉代谢以及丙酮酸代谢等,其中,蔗糖和淀粉代谢途径中的内切葡聚糖酶能够将纤维素降解为纤维二糖,这表明 *R. papyrosolvans* DSM2782 具有降解木质纤维素的潜力。在细胞过程中,关于 Transporters 的基因注释数量最多,有 365 个,Two-component system 为 189 个,ABC transporters 为 162 个,同时 Ribosome 也有 108 个基因注释数量,表明 *R. papyrosolvans* DSM2782 能够利用多种类型的 ABC 物质转运系统来适应环境的变化。另外,*R. papyrosolvans* DSM2782 在 Enzymes with EC numbers 方面为 105 个,说明该菌能够充分适应环境的变化并且同时具有较强的碳水化合物代谢和转运能力。另外,有 19 个基因被注释到聚酮化合物合成酶基因途径,推测这些基因和 *R. papyrosolvans* DSM2782 能够产生生物活性物质有关。

#### 2.5 CAZy 注释结果分析

碳水化合物活性水解酶能够参与各种碳水化合物的分解代谢,*R. papyrosolvans* DSM2782 的 KEGG 代谢通路分类结果显示注释到碳水化合物代谢通路下的功能基因数目最多。碳水化合物活性水解酶数据库(CAZy)是关于能够合成或分解复杂碳水化合物的数据库,主要包括糖苷水解酶(Glycoside Hydrolases, GHs)、糖基转移酶(Glycosyltransferases, GTs)、多糖裂解酶(Polysaccharide Lyases, PLs)、碳水化合物酯酶(Carbohydrate Esterases, CEs)、辅助活性酶类(Auxiliary Activities, AAs)和碳水化合物结合模块(Carbohydrate binding modules, CBMs)这 6 类能催化碳水化合物降解、修饰以及生物合成的相关酶系家族。结果表明在 *R. papyrosolvans* DSM2782 的基因组中检测到 231 个不同的碳水化合物活性水解酶(图 5),包括 56 个 CBM 家族,23 个 CE 家族,111 个 GH 家族,37 个 GT 家族,3 个 PL 家族以及 1 个 AA 家族。其中,GH

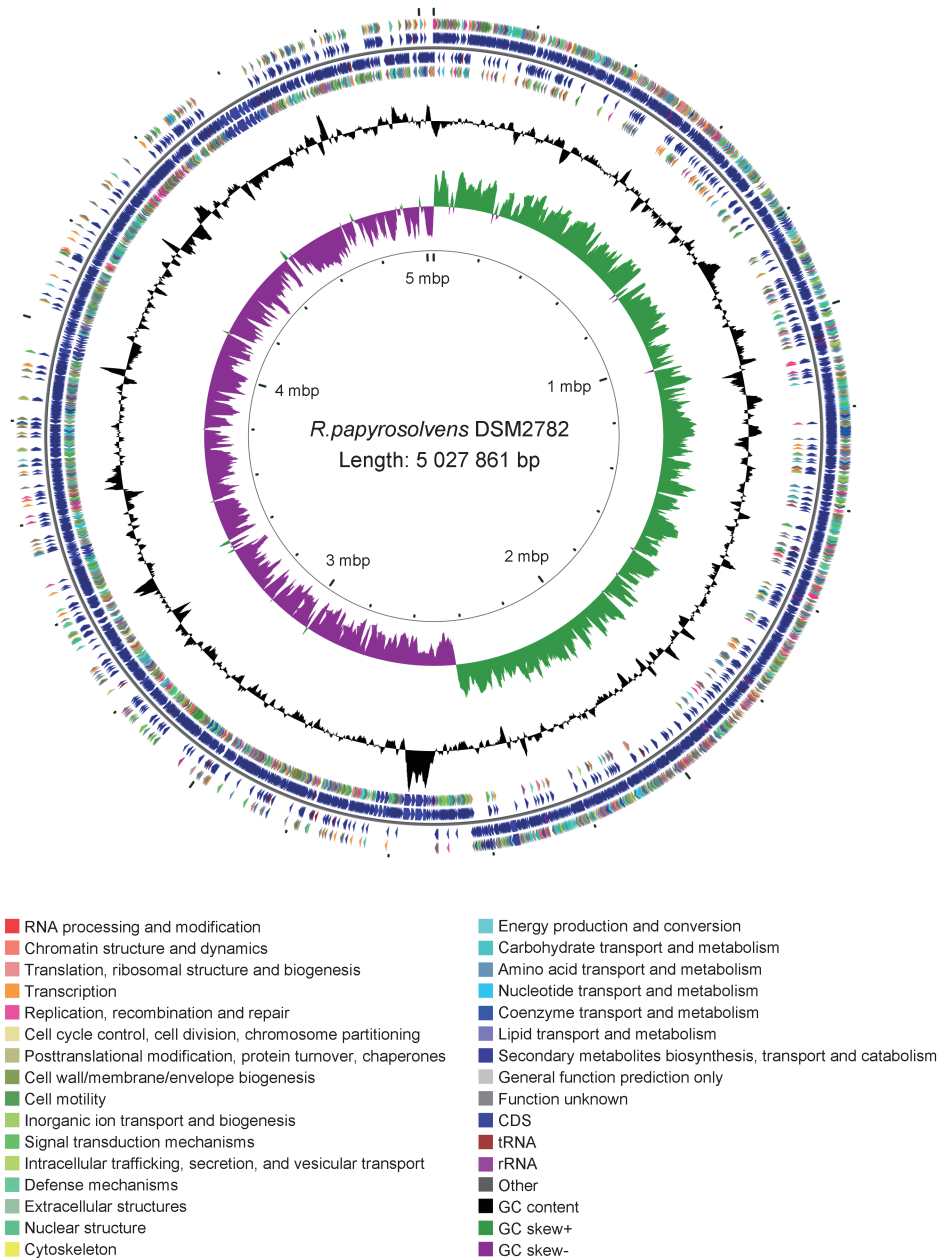


图1 *R. papyrosolvens* DSM2782 基因组圈图

从内到外,第一圈代表基因组大小;第二圈代表GC Skew;第三圈代表GC含量;第四圈和第七圈代表每个CDS所属于的COG分类;第五圈和第六圈代表CDS、tRNA及rRNA在基因组上的位置。

Fig. 1 Genome circle map of the *R. papyrosolvens* DSM2782

From the inside to the outside of the genomic map image, the first circle represents genome size; The second circle represents GC Skew; The third circle represents the GC content; The fourth and seventh circles represent the COG classification of the CDS; The fifth and sixth circles represent the positions of CDS, tRNAs and rRNAs.

家族数量和种类最为丰富,其中GH9、GH43、GH5、GH10的数量分别为14、13、9、6。CBM家族中含有数量最多的种类为CBM6、CBM3、CBM9,数量分别为17、7、5。CE家族中数量最多的分别为CE4、CE1,数量分别为8、6。此外,在GH家族、CBM家族、CE家族以及PL家

族中,分别含有51、32、7、3个纤维小体基因,相应地在GH家族、CBM家族、CE家族以及AA家族中,分别含有60、24、16、1个游离酶基因。这些结果表明 *R. papyrosolvens* DSM2782 拥有种类和数量丰富的纤维小体基因以及游离酶基因,能够高效调控不同种类的碳水化合物活

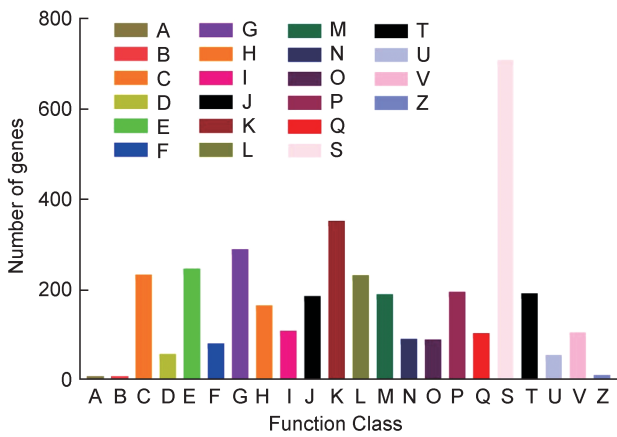


图2 *R. papyrosolvans* DSM2782基因组COG功能注释分类  
 Fig. 2 Functional classification of COG of the *R. papyrosolvans* DSM2782

性水解酶的表达以此从环境中获取营养来维持自身的生命活动。

2.6 梭菌属同源性分析

本研究对11个不同种类的梭菌属进行ANI分析,结果显示(图6)。*R. cellubioparum* DSM1351与*R. termitidis* CT1112以及*R. hungatei* DSM14427之间同源关系最为相近,ANI值为96.5%以及77.5%。*C. acetobutylicum* ATCC824与*C. cellulovorans* 743B之间同源关系最为相近,ANI值为74.5%。*R. papyrosolvans* DSM2782与*R. cellulolyticum* H10、*R. sp.* BNL1100、*R. josui* JCM17888以及*R. papyrosolvans* C7之间同源关系最为相近,ANI值分别为83.4%、88.7%、85.1以及88.4%。*R. herbifermentans* MA18与*R. sufflavum* DSM19573之间同源关系最为相近,ANI

值为77.6%。这些结果表明*R. papyrosolvans* DSM2782与*R. cellulolyticum* H10、*R. sp.* BNL1100、*R. josui* JCM17888以及*R. papyrosolvans* C7的同源性较高,同时也表明了*Ruminiclostridium*与*Clostridium*之间的种间差异性。

对*R. papyrosolvans* DSM2782与*R. papyrosolvans* C7进行基因组比较分析探究*R. papyrosolvans*不同菌株之间的差异性。结果表明*R. papyrosolvans* DSM2782与*R. papyrosolvans* C7之间的大部分区域的序列高度同源(图7(a))。由于*R. papyrosolvans* C7基因组信息严重缺失,这也可能导致ANI分析结果中*R. papyrosolvans* C7与*R. papyrosolvans* DSM2782的同源性略低于*R. papyrosolvans* DSM2782与*R. sp.* BNL1100之间的同源性(图6)。对*R. papyrosolvans* DSM2782与*R. papyrosolvans* C7含有的纤维小体基因簇进行同源性分析,发现纤维素降解基因簇*cip-cel*之间的基因同源性很高。但由于*R. papyrosolvans* C7中降解半纤维素的基因簇*xyl-doc*的信息缺失严重,导致无法准确得出*R. papyrosolvans* DSM2782与*R. papyrosolvans* C7所持有的纤维小体基因簇之间的整体同源性(图7(b))。尽管如此,*R. papyrosolvans* DSM2782与*R. papyrosolvans* C7之间的同源性仍然高于其他类型的梭菌。

2.7 *R. papyrosolvans* DSM2782的次级代谢产物合成基因簇分析结果

使用 antiSMASH 对 *R. papyrosolvans*

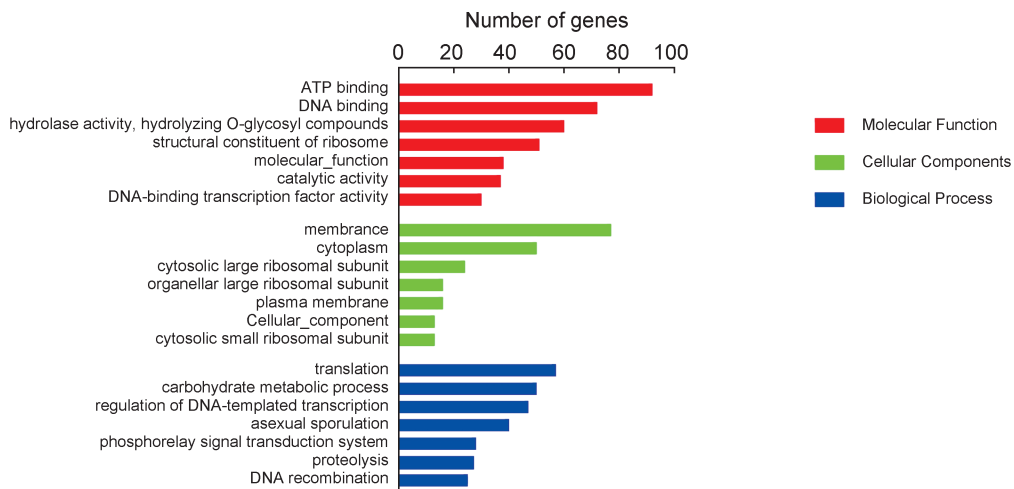


图3 *R. papyrosolvans* DSM2782基因组GO功能注释  
 Fig. 3 GO functional annotation of the *R. papyrosolvans* DSM2782

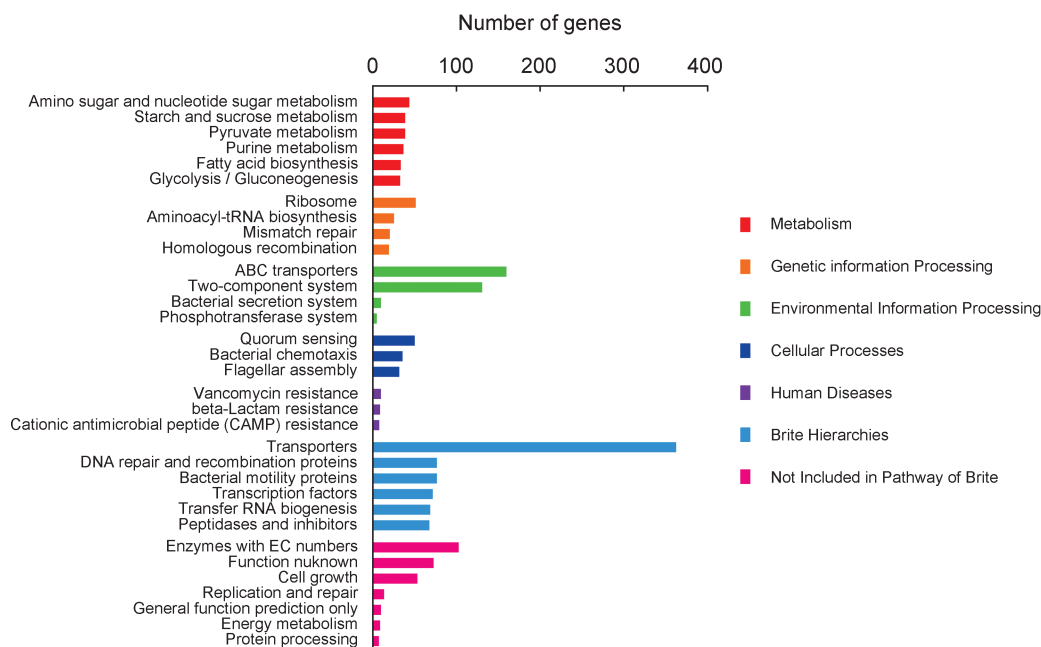


图4 *R. papyrosolvans* DSM2782 基因组 KEGG 代谢途径注释

Fig. 4 KEGG pathway annotation of the *R. papyrosolvans* DSM2782

DSM2782 是否含有生物合成基因簇进行预测, 结果表明 *R. papyrosolvans* DSM2782 基因组中含有 23 个生物合成基因簇, 这些生物合成基因簇类型大多为编码非核糖体肽合成酶 (NRPS)、核糖体合成和翻译修饰后的多肽 (RiPP)<sup>[25]</sup> 以及聚酮化合物合成酶 (PKS) (表 2)。其中, 数量最多的生物合成基因簇编码非核糖体肽合成酶, 剩下的大多编码核糖体合成和翻译修饰后的多肽以及聚酮化合物合成酶等。在 *R. papyrosolvans* DSM2782 基因组含有的生物合成基因簇中, P0092\_RS14160-P0092\_RS14255 与合成 Rc-AIP1 的基因簇相似性为 83%<sup>[26]</sup>, P0092\_RS11420-P0092\_RS11760 与合成 macrobrevin 的基

因簇相似性为 66%<sup>[27]</sup>。此外, *R. papyrosolvans* DSM2782 含有的大多数生物合成基因簇与 anti-SMASH 数据库中收录的生物合成基因簇的相似度总体较低, 暗示 *R. papyrosolvans* DSM2782 具有合成不同于现有生物活性化合物结构的潜力。

### 2.8 不同碳源培养下的 *R. papyrosolvans* DSM2782 代谢产物结果分析

使用非靶向代谢技术检测在以纤维二糖以及秸秆为碳源培养下的 *R. papyrosolvans* DSM2782 产生的次级代谢产物, 在 *R. papyrosolvans* DSM2782 中能检测到具有生物活性的物质分别有 4-羟基苯甲酸、尼泊金丙酯、丁酸、阿魏

表3 *R. papyrosolvans* DSM2782 发酵液中鉴定的生物活性物质

Table 3 Identification of bioactive substances in the fermentation broth of the *R. papyrosolvans* DSM2782

序号	中文名称	英文名称	MS2 分值	化合物种类
1	4-羟基苯甲酸	4-Hydroxybenzoic acid	4	苯环型化合物
2	壬酸	Pelargonic acid	3.98	脂质和类脂分子
3	尼泊金丙酯	Propyl paraben	3.98	苯环型化合物
4	乳酸	Lactate	3.95	有机酸及其衍生物
5	丁酸	Butanoic acid	3.94	脂质和类脂分子
6	阿魏酸	Ferulate	3.87	苯丙素类和聚酮类
7	四氢嘧啶	Ectoine	3.84	有机酸及其衍生物
8	硫胺素	Thiamine	3.8	有机杂环化合物
9	去甲哈尔满	beta-Carboline	3.66	有机杂环化合物
10	左旋多巴	3,4-Dihydroxyphenylalanine	3.66	有机酸及其衍生物

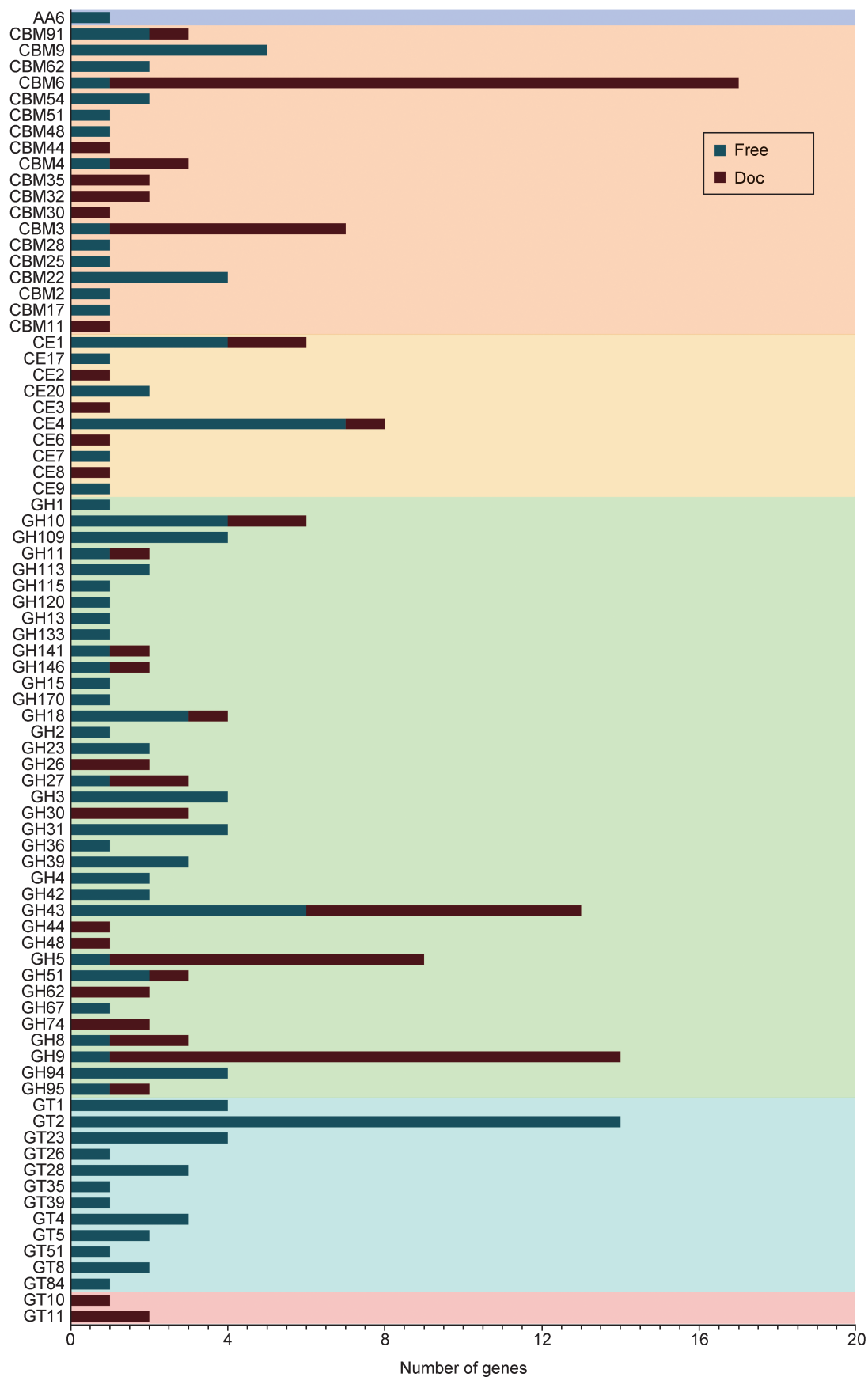


图5 *R. papyrosolvens* DSM2782 CAZyme 注释分类

Fig. 5 CAZyme annotation and classification of the *R. papyrosolvens* DSM2782

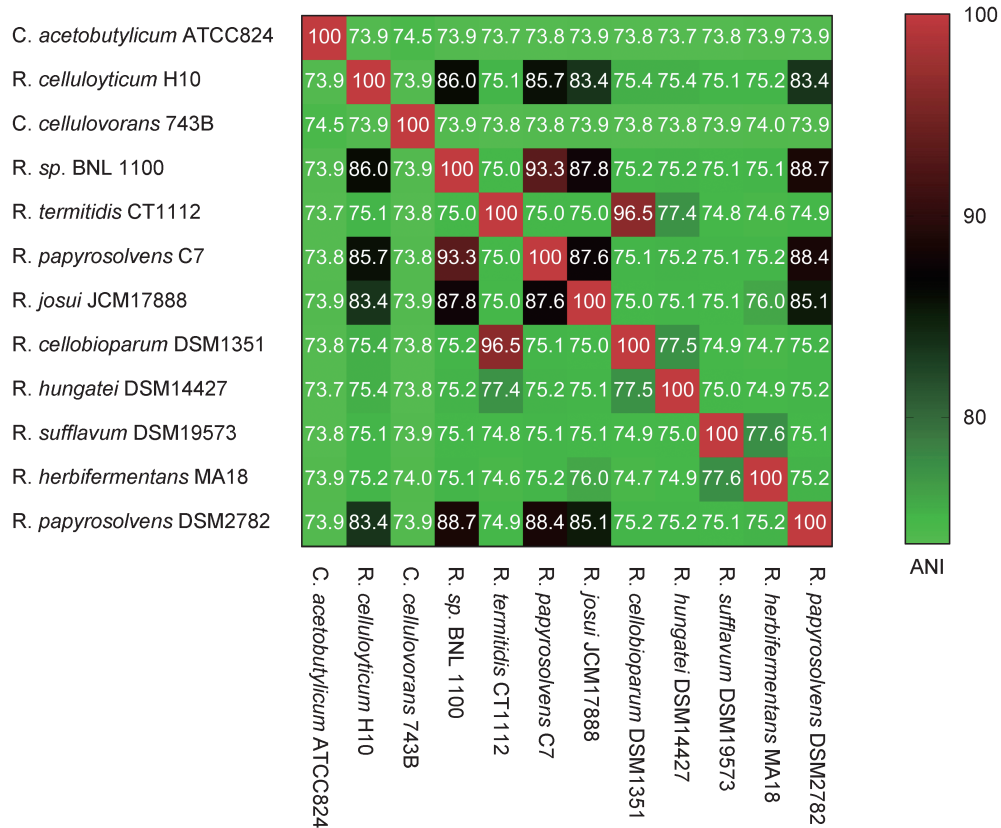


图6 梭菌属 ANI 热图

Fig. 6 ANI heatmap of the clostridium

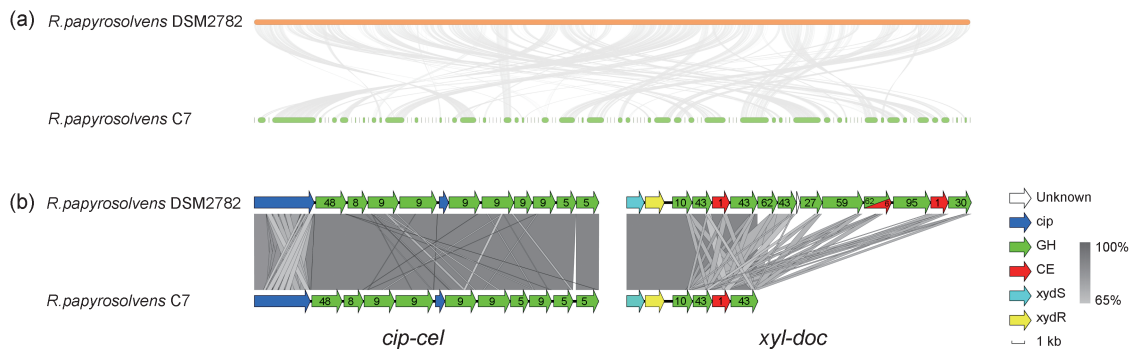


图7 *R. papyrosolvens* DSM2782与*R. papyrosolvens* C7同源性分析

(a) *R. papyrosolvens* DSM2782与*R. papyrosolvens* C7基因组共线性分析; (b) *R. papyrosolvens* DSM2782与*R. papyrosolvens* C7纤维小体基因簇同源性分析。

Fig. 7 Homology analysis between *R. papyrosolvens* DSM2782 and *R. papyrosolvens* C7

(a) Collinearity analysis of the *R. papyrosolvens* DSM2782 and *R. papyrosolvens* C7; (b) Homology analysis of the *cip-cel* and *xyl-doc* gene cluster between *R. papyrosolvens* DSM2782 and *R. papyrosolvens* C7.

酸、四氢嘧啶、去甲哈尔满以及左旋多巴等具有不同结构的次级代谢产物(表3)。研究表明这些次级代谢产物均具有一定的生物活性。例如,去甲哈尔满属于有机杂环化合物<sup>[28]</sup>,是一种很有前景的抗癌光敏剂,也能够用作抗生素佐剂,具有增强常规抗生素的功效。丁酸属于脂

质和类脂分子化合物<sup>[29]</sup>,对于宿主的健康起着非常重要的作用。阿魏酸属于苯丙素类和聚酮类化合物<sup>[30]</sup>,是*R. papyrosolvens* DSM2782降解秸秆时产生的代谢产物,能够起到缓解血管痉挛等作用。这些结果表明*R. papyrosolvens* DSM2782能够产生各种具有不同结构的并且具

表2 *R. papyrosolvens* DSM2782 生物合成基因簇预测结果Table 2 Prediction of gene clusters of the *R. papyrosolvens* DSM2782 for synthesis of secondary metabolites

Region ID	Type	Gene Start	Gene End	Most similar known cluster	Similarity/%
1.1	betalactone	P0092_RS00595	P0092_RS00680	—	—
1.2	ranthipeptide	P0092_RS01875	P0092_RS01970	—	—
1.3	ranthipeptide	P0092_RS03520	P0092_RS03605	—	—
1.4	NRPS, transAT-PKS-like	P0092_RS03710	P0092_RS03865	—	—
1.5	thioamide-NRP	P0092_RS04000	P0092_RS04245	—	—
1.6	NRPS-like	P0092_RS07265	P0092_RS07460	octapeptin C4	11
1.7	ranthipeptide	P0092_RS07975	P0092_RS08085	—	—
1.8	cyclic-lactone-autoinducer	P0092_RS08555	P0092_RS08680	—	—
1.9	cyclic-lactone-autoinducer	P0092_RS09050	P0092_RS09180	—	—
1.10	RiPP-like, transAT-PKS, NRPS-like, NRP-me-tallophore, NRPS	P0092_RS10780	P0092_RS11245	ulbactin F/ulbactin G	42
1.11	PKS-like, transAT-PKS, NRPS	P0092_RS11420	P0092_RS11760	macrobrevin	66
1.12	cyclic-lactone-autoinducer	P0092_RS14160	P0092_RS14255	Rc-AIP1	83
1.13	RiPP-like	P0092_RS15365	P0092_RS15410	—	—
1.14	NRPS-like, NRPS, T1PKS	P0092_RS15670	P0092_RS15825	—	—
1.15	ranthipeptide	P0092_RS15980	P0092_RS16060	—	—
1.16	NRPS-like	P0092_RS17025	P0092_RS17260	locillomycin/locillomycin B/locillomycin C	14
1.17	cyclic-lactone-autoinducer, transAT-PKS, NRPS	P0092_RS17495	P0092_RS17825	xantholipin	4
1.18	NRPS	P0092_RS18370	P0092_RS18530	—	—
1.19	thiopeptide, LAP	P0092_RS18760	P0092_RS18900	clostridiolysin S	22
1.20	cyclic-lactone-autoinducer	P0092_RS19350	P0092_RS19510	—	—
1.21	T1PKS, transAT-PKS-like	P0092_RS20100	P0092_RS20275	—	—
1.22	betalactone	P0092_RS21610	P0092_RS21745	—	—
1.23	cyclic-lactone-autoinducer	P0092_RS21785	P0092_RS21860	—	—

注:—表示未预测到相似的基因簇。

有不同活性的次级代谢产物。这些结果将进一步为后续研究 *R. papyrosolvens* DSM2782 产生具有活性的次级代谢产物提供更多的参考信息。

### 3 讨论与结论

本研究通过基因组二代和三代测序技术对 *R. papyrosolvens* DSM2782 进行全基因组测序, 获得了 *R. papyrosolvens* DSM2782 的完整基因组序列。通过一系列生物信息学分析, 结果表明 *R. papyrosolvens* DSM2782 在木质纤维素降解方面具有很高的应用潜力。此外, 通过 antiS-MASH 预测结果表明, 在 *R. papyrosolvens* DSM2782 基因组内还含有数量相当丰富的生物合成基因簇, 暗示其具有能够合成有应用价值的生物活性化合物的潜力。

通过全基因组测序与分析, 表明在 *R. papy-*

*rosolvens* DSM2782 基因组中含有数量相当丰富的纤维小体基因以及基因簇, 例如纤维小体基因簇主要为 *cip-cel* 以及 *xyl-doc*, 数量较多的纤维小体酶为 GH43、GH5、GH9 等家族。其中, GH43 的种类和功能较为复杂, GH5 和 GH9 则主要是纤维素酶, 这表明 *R. papyrosolvens* DSM2782 能够高效降解纤维素, 同时也能够降解其他种类的复杂多糖, 例如半纤维素、木聚糖等。联合基因组学结合蛋白组学以及转录组学等技术, 我们能够探究在复杂多糖下 *R. papyrosolvens* DSM2782 基因组中不同种类的纤维小体基因的表达变化, 有望为构建高效降解木质纤维素的纤维小体的人工合成与应用提供新的纤维小体酶基因。

随着基因组测序技术的飞速发展, 极大地推动了对嗜中温纤维素降解梭菌的基因组研究

进展。通过结合基因组挖掘、一株多种化合物 (One Strain Many Compounds, OSMAC) 策略<sup>[31]</sup> 以及代谢组学等研究方法, 能够不断地扩大嗜中温纤维素降解梭菌资源库, 并在嗜中温纤维素降解梭菌中发现更多的生物合成基因簇以及结构新颖的次级代谢产物。通过基因编辑技术改造 *R. papyrosolvens* DSM2782, 使其在降解木质纤维素的同时也产生具有活性的次级代谢产物。这样不仅能够避免因为木质纤维素难以降解造成资源的浪费, 而且还能够进一步丰富微生物天然产物资源库及合成途径。这将极大地扩展嗜中温纤维素降解梭菌的应用范围。因此, 对于通过 OSMAC 策略是否能够诱导 *R. papyrosolvens* DSM2782 产生具有生物活性的化合物还有待进一步实验探究。此外还需要进行实验探究如何激活在实验室条件下 *R. papyrosolvens* DSM2782 基因组中沉默的生物合成基因簇, 从而开发 *R. papyrosolvens* DSM2782 产生具有活性的次级代谢产物的潜力。

综上所述, 本研究利用全基因组测序技术首次揭示了 *R. papyrosolvens* DSM2782 作为嗜中温纤维素降解菌株以及其本身含有数量和种类丰富的生物合成基因簇的特点, 表明其作为嗜中温厌氧纤维素降解梭菌不仅能够应用于纤维素降解, 而且还具有产生有活性的天然产物的潜力的双重研究提供了科学的理论依据。

#### 参考文献:

- [1] RAJ T, CHANDRASEKHAR K, NARESH KUMAR A, et al. Recent Advances in Commercial Biorefineries for Lignocellulosic Ethanol Production: Current Status, Challenges and Future Perspectives[J]. *Bioresour Technol*, 2022, **344**(Pt B): 126292. DOI: 10.1016/j.biortech.2021.126292.
- [2] HEMME C L, MOUTTAKI H, LEE Y J, et al. Sequencing of Multiple Clostridial Genomes Related to Biomass Conversion and Biofuel Production[J]. *J Bacteriol*, 2010, **192**(24): 6494-6496. DOI:10.1128/JB.01064-10.
- [3] YOU M C, ZHAO Q Y, LIU Y S, et al. Insights into Lignocellulose Degradation: Comparative Genomics of Anaerobic and Cellulolytic *Ruminiclostridium*-type Species [J]. *Front Microbiol*, 2023, **14**: 1288286. DOI:10.3389/fmicb.2023.1288286.
- [4] BÉGUIN P, LEMAIRE M. The Cellulosome: An Exocellular, Multiprotein Complex Specialized in Cellulose Degradation[J]. *Crit Rev Biochem Mol Biol*, 1996, **31**(3): 201-236. DOI:10.3109/10409239609106584.
- [5] DOI R H, TAMARU Y. The *Clostridium cellulovorans* Cellulosome: An Enzyme Complex with Plant Cell Wall Degrading Activity[J]. *Chem Rec*, 2001, **1**(1): 24-32. DOI: 10.1002/1528-0691(2001)1: 124:: AID-TCR5>3.0.CO;2-W.
- [6] ARTZI L, BAYER E A, MORAÏS S. Cellulosomes: Bacterial Nanomachines for Dismantling Plant Polysaccharides[J]. *Nat Rev Microbiol*, 2017, **15**(2): 83-95. DOI: 10.1038/nrmicro.2016.164.
- [7] FIEROBE H P, PAGÈS S, BÉLAÏCH A, et al. Cellulosome from *Clostridium cellulolyticum*: Molecular Study of the Dockerin/Cohesin Interaction[J]. *Biochemistry*, 1999, **38**(39): 12822-12832. DOI:10.1021/bi9911740.
- [8] LINCKE T, BEHNKEN S, ISHIDA K, et al. Closthioamide: An Unprecedented Polythioamide Antibiotic from the Strictly Anaerobic Bacterium *Clostridium cellulolyticum*[J]. *Angew Chem Int Ed*, 2010, **49**(11): 2011-2013. DOI:10.1002/anie.200906114.
- [9] LETZEL A C, PIDOT S J, HERTWECK C. A Genomic Approach to the Cryptic Secondary Metabolome of the Anaerobic World[J]. *Nat Prod Rep*, 2013, **30**(3): 392-428. DOI:10.1039/c2np20103h.
- [10] MIARI V F, SOLANKI P, HLEBA Y, et al. In Vitro Susceptibility to Closthioamide among Clinical and Reference Strains of *Neisseria gonorrhoeae*[J]. *Antimicrob Agents Chemother*, 2017, **61**(10): e00929-17. DOI: 10.1128/AAC.00929-17.
- [11] ZOU X, REN Z X, WANG N, et al. Function Analysis of 5'-UTR of the Cellulosomal *xyl*-*doc* cluster in *Clostridium papyrosolvens*[J]. *Biotechnol Biofuels*, 2018, **11**: 43. DOI:10.1186/s13068-018-1040-0.
- [12] SCHUBERT M, LINDGREEN S, ORLANDO L. AdapterRemoval V2: Rapid Adapter Trimming, Identification, and Read Merging[J]. *BMC Res Notes*, 2016, **9**: 88. DOI:10.1186/s13104-016-1900-2.
- [13] LUO R B, LIU B H, XIE Y L, et al. SOAPdenovo2: An Empirically Improved Memory-efficient Short-read de Novo Assembler[J]. *Gigascience*, 2012, **1**(1): 18. DOI: 10.1186/2047-217X-1-18.
- [14] 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 Res*, 2017, **27**(5): 722-736. DOI: 10.1101/gr.215087.116.
- [15] WALKER B J, ABEEL T, SHEA T, et al. Pilon: An Inte-

- grated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement[J]. *PLoS One*, 2014, **9**(11): e112963. DOI: 10.1371/journal.pone.0112963.
- [16] WICK R R, JUDD L M, GORRIE C L, *et al.* Unicycler: Resolving Bacterial Genome Assemblies from Short and Long Sequencing Reads[J]. *PLoS Comput Biol*, 2017, **13**(6): e1005595. DOI:10.1371/journal.pcbi.1005595.
- [17] TATUSOVA T, DICUCCIO M, BADRETDIN A, *et al.* NCBI Prokaryotic Genome Annotation Pipeline[J]. *Nucleic Acids Res*, 2016, **44**(14): 6614–6624. DOI: 10.1093/nar/gkw569.
- [18] TATUSOV R L, GALPERIN M Y, NATALE D A, *et al.* The COG Database: A Tool for Genome-scale Analysis of Protein Functions and Evolution[J]. *Nucleic Acids Res*, 2000, **28**(1): 33–36. DOI:10.1093/nar/28.1.33.
- [19] ZHENG J F, GE Q W, YAN Y C, *et al.* dbCAN3: Automated Carbohydrate-active Enzyme and Substrate Annotation[J]. *Nucleic Acids Res*, 2023, **51**(W1): W115–W121. DOI:10.1093/nar/gkad328.
- [20] JAIN C, RODRIGUEZ-R L M, PHILLIPPY A M, *et al.* High Throughput ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries[J]. *Nat Commun*, 2018, **9**(1): 5114. DOI: 10.1038/s41467-018-07641-9.
- [21] WANG Y P, TANG H B, DEBARRY J D, *et al.* MCS-canX: A Toolkit for Detection and Evolutionary Analysis of Gene Synteny and Collinearity[J]. *Nucleic Acids Res*, 2012, **40**(7): e49. DOI:10.1093/nar/gkr1293.
- [22] MISTRY J, CHUGURANSKY S, WILLIAMS L, *et al.* Pfam: The Protein Families Database in 2021[J]. *Nucleic Acids Res*, 2021, **49**(D1): D412–D419. DOI: 10.1093/nar/gkaa913.
- [23] SULLIVAN M J, PETTY N K, BEATSON S A. Easyfig: A Genome Comparison Visualizer[J]. *Bioinformatics*, 2011, **27**(7): 1009–1010. DOI:10.1093/bioinformatics/btr039.
- [24] BLIN K, SHAW S, AUGUSTIJN H E, *et al.* AntiSMASH 7.0: New and Improved Predictions for Detection, Regulation, Chemical Structures and Visualisation [J]. *Nucleic Acids Res*, 2023, **51**(W1): W46–W50. DOI: 10.1093/nar/gkad344.
- [25] HUDSON G A, MITCHELL D A. RiPP Antibiotics: Biosynthesis and Engineering Potential[J]. *Curr Opin Microbiol*, 2018, **45**: 61–69. DOI:10.1016/j.mib.2018.02.010.
- [26] MOLLOY E M, DELL M, HÄNSCH V G, *et al.* Enzyme-primed Native Chemical Ligation Produces Autoinducing Cyclopeptides in Clostridia[J]. *Angew Chem Int Ed*, 2021, **60**(19): 10670–10679. DOI: 10.1002/anie.202016378.
- [27] CHAKRABORTY K, KIZHAKKEKALAM V K, JOY M. Polyketide-derived Macrobrevins from Marine Macroalga-associated *Bacillus amyloliquefaciens* as Promising Antibacterial Agents Against Pathogens Causing Nosocomial Infections[J]. *Phytochemistry*, 2022, **193**: 112983. DOI: 10.1016/j.phytochem.2021.112983.
- [28] KUSHWAHA P, KUMAR V, SAHA B. Current Development of  $\beta$ -carboline Derived Potential Antimalarial Scaffolds[J]. *Eur J Med Chem*, 2023, **252**: 115247. DOI: 10.1016/j.ejmech.2023.115247.
- [29] VAN IMMERSEEL F, DUCATELLE R, DE VOS M, *et al.* Butyric Acid-producing Anaerobic Bacteria as a Novel Probiotic Treatment Approach for Inflammatory Bowel Disease[J]. *J Med Microbiol*, 2010, **59**(Pt 2): 141–143. DOI:10.1099/jmm.0.017541-0.
- [30] STOMPOR-GORAŁY M, MACHACZKA M. Recent Advances in Biological Activity, New Formulations and Prodrugs of Ferulic Acid[J]. *Int J Mol Sci*, 2021, **22**(23): 12889. DOI:10.3390/ijms222312889.
- [31] BODE H B, BETHE B, HÖFS R, *et al.* Big Effects from Small Changes: Possible Ways to Explore Nature's Chemical Diversity[J]. *Chembiochem*, 2002, **3**(7): 619–627. DOI:10.1002/1439-7633(20020703)3:7<619::AID-CBIC619>3.0.CO;2-9.