



产 *S*-雌马酚梭菌 C1 三代全基因组测序及功能基因筛选鉴定

胡云霏¹, 阳婷¹, 王春霖¹, 冯然¹, 陈华海¹, 李降龙², 尹业师^{1*}

1 湖南科技学院化学与生物工程学院, 湘南优势植物资源综合利用湖南省重点实验室, 湖南 永州 425199

2 湖南海路生物科技有限公司, 湖南 常德 415400

胡云霏, 阳婷, 王春霖, 冯然, 陈华海, 李降龙, 尹业师. 产 *S*-雌马酚梭菌 C1 三代全基因组测序及功能基因筛选鉴定. 微生物学报, 2022, 62(1): 333–345.

Hu Yunfei, Yang Ting, Wang Chunlin, Feng Ran, Chen Huahai, Li Xianglong, Yin Yesi. Whole genome sequencing and functional gene screening of *S*-equol-producing *Clostridium* sp. C1. *Acta Microbiologica Sinica*, 2022, 62(1): 333–345.

摘要:【目的】挖掘产 *S*-雌马酚梭菌 C1 转化大豆苷元产生 *S*-雌马酚的功能基因, 为梭菌 C1 的 *S*-雌马酚转化机制研究提供参考, 并为利用合成生物学方法生产 *S*-雌马酚提供新基因资源。

【方法】利用 GridION 测序平台, 对梭菌 C1 进行第三代全基因组测序、基因组组装和功能注释等分析, 从 C1 菌全基因组中筛选和鉴定参与 *S*-雌马酚生物转化的功能基因。【结果】C1 全基因组大小为 3 035 113 bp, 预测编码 3 166 个基因, 包含 53 个 tRNA、15 个 rRNA、4 个 ncRNA 和 1 个基因岛。通过生物信息学分析, 发现 C1-07020 基因编码蛋白与已报道的 *Lactococcus* sp. 20-92 大豆苷元还原酶具有 44.8% 的氨基酸序列相似性和相同的 3 个功能保守结构域, 体外蛋白功能验证表明, C1-07020 具有大豆苷元还原酶功能。此外, C1 菌中没有发现与已知产 *S*-雌马酚菌相似的功能基因簇或大豆苷元还原酶以外的其他功能基因。【结论】在 C1 中鉴定到一个新的产 *S*-雌马酚功能基因, 并发现了 C1 可能具备特殊的产 *S*-雌马酚机制, 实验所获基础数据可为进一步挖掘产 *S*-雌马酚新功能基因、了解 *S*-雌马酚的生成机制及体外产 *S*-雌马酚基因资源的应用提供参考。

关键词: 产 *S*-雌马酚菌; 全基因组; 大豆苷元还原酶; 二氢大豆苷元; 功能基因

基金项目: 中央引导地方科技发展专项资金(2019XF5067); 湖南省自然科学基金(2021JJ40220)

Supported by the Project for Guiding Local Science and Technology Development by the Central Government (2019XF5067) and by the Natural Science Foundation of Hunan, China (2021JJ40220)

*Corresponding author. E-mail: yinyeshi@126.com

Received: 26 April 2021; Revised: 27 July 2021; Published online: 4 November 2021

Whole genome sequencing and functional gene screening of *S*-equol-producing *Clostridium* sp. C1

HU Yunfei¹, YANG Ting¹, WANG Chunlin¹, FENG Ran¹, CHEN Huahai¹, LI Xianglong², YIN Yesi^{1*}

¹ Key Laboratory of Comprehensive Utilization of Advantage Plants Resources in Hunan South, College of

Chemistry and Bioengineering, Hunan University of Science and Engineering, Yongzhou 425199, Hunan, China

² Hailu Biotech Co., Ltd, Changde 415400, Hunan, China

Abstract: [Objective] This paper aims to investigate genes related to the production of *S*-equol from daidzein in *Clostridium* sp. C1, which is expected to serve as a reference for exploring the *S*-equol transformation mechanism in C1 and discover genes for the synthesis of *S*-equol by synthetic biology techniques. [Methods] Through the third-generation sequencing (GridION), assembly, and function annotation of the whole genome of C1, genes related to the biotransformation of *S*-equol were screened and identified. [Results] The genome was 3 035 113 bp, with 3 166 protein-coding genes, 53 tRNA genes, 15 rRNA genes, 4 ncRNA genes, and 1 genomic island. Bioinformatics analysis revealed that C1-07020 protein shared 44.8% amino acids and 3 conserved domains with daidzein reductase of *Lactococcus* sp. 20-92. *In vitro* verification showed that C1-07020 protein had similar functions to daidzein reductase. In addition, no other *S*-equol-producing gene clusters or other functional genes were identified from C1, which suggests that C1 may have a different metabolic mechanism from other *S*-equol-producing bacteria. [Conclusion] One *S*-equol-producing gene was identified from C1 and C1 might have a unique *S*-equol synthesis mechanism. The results of this study can serve as a reference for further exploring *S*-equol-producing genes, the production mechanism, and *in vitro* utilization of the genes.

Keywords: *S*-equol-producing bacteria; whole genome; daidzein reductase; dihydrodaidzein; functional genes

作为一种类雌激素活性最高、对人体没有毒副作用的大豆异黄酮代谢物，*S*-雌马酚在防治更年期综合征、雌激素相关癌症、心脑血管疾病、肥胖和调节抗氧化应激、炎症等方面发挥重要作用^[1-5]。豆制品中不含 *S*-雌马酚，人体本身也不能合成 *S*-雌马酚，现已明确，肠道产 *S*-雌马酚菌的存在是个体具备产 *S*-雌马酚能力的必要条件^[6]，且不同人群产 *S*-雌马酚能力差异较大^[3]。目前已报道了多种产 *S*-雌马酚菌，但研究较为清楚的菌株只有 *Lactococcus* sp. 20-92、*Slackia isoflavoniconvertens* sp. HE8、

Slackia sp. NATTS、*Eggerthella* sp. YY7918 和 *Adlercreutzia equolifaciens* sp. DSM19450^{T[7]}，而且对于产 *S*-雌马酚菌的种类、丰度与人是否具备产 *S*-雌马酚能力及豆制品摄入之间的关系仍不清楚。体外补充 *S*-雌马酚是满足非产雌马酚者需求的最佳方式之一。利用产 *S*-雌马酚基因，采用合成生物学的方式进行体外发酵生产，是实现 *S*-雌马酚量产的有效途径。

已有研究表明，转化大豆苷元(daidzein, DZN)合成 *S*-雌马酚的过程需要大豆苷元还原酶(daidzein reductase, DZNR)、外消旋酶

(dihydrodaidzein racemase, DDRC)、二氢大豆苷元还原酶(dihydrodaidzein reductase, DHDR)和四氢大豆苷元还原酶(tetrahydrodaidzein reductase, THDR)的参与,它们分别将 DZN 转化为 R-二氢大豆苷元(R-dihydrodaidzein, R-DHD)、S-DHD、四氢大豆苷元(tetrahydrodaidzein, THD)和 S-雌马酚(图 1)^[8-10]。根据以上参与 S-雌马酚代谢转化的 4 个功能基因序列,本实验室前期采用兼容表达载体,将这 4 个基因同时转化到大肠杆菌中,构建了可将 DZN 转化为 S-雌马酚的基因工程菌,在体外实现了 S-雌马酚的发酵合成,但 S-雌马酚发酵产量仍有待提升^[7,11]。由于已掌握的产雌马酚菌种资源有限,加大对产雌马酚菌种资源的挖掘、深入了解产 S-雌马酚作用机制是突破这一瓶颈,实现 S-雌马酚量产工艺的重要途径。

梭菌 C1 是一株从小鸡盲肠内容物中分离获得的革兰氏阳性丝状厌氧菌^[6],该菌可在厌氧条件下将 DZN 转化为 S-雌马酚。为深入了解 C1 菌中参与 S-雌马酚合成的功能基因及其代谢通路信息,本研究采用第三代全基因组测序技术,获得 C1 菌的全基因组信息,然后根据已

报道的产 S-雌马酚功能基因信息,筛选和鉴定 C1 菌中潜在的参与 S-雌马酚生物合成的功能基因,为了解 C1 菌产 S-雌马酚机制以及功能基因资源的体外应用提供参考。

1 材料与方法

1.1 材料

1.1.1 菌种和质粒

梭菌 C1、pETDute-1 质粒保存于本实验室,大肠杆菌 BL21 (DE3) (北京全式金)。

1.1.2 主要仪器与试剂

厌氧工作站(英国 Electrotek);台式冷冻浓缩仪(美国 LABCONCO);高效液相色谱仪(日本 Shimadzu);SunFire™ C₁₈ 检测柱(5 μm, 4.6 mm×250 mm 柱);微量分光光度计(德国 Norllu);蛋白电泳与转膜系统(美国 Thermo);多功能酶标仪(美国 BioTek)。

Gifu anaerobic medium (GAM, 青岛海博莱);DZN、DHD (上海大赛璐);细菌 DNA 提取试剂盒(德国 Qiagen);琼脂糖 Ni 亲和层析柱(美国 Thermo);BCA 蛋白定量试剂盒、NAD(P)H (北京索莱宝)。

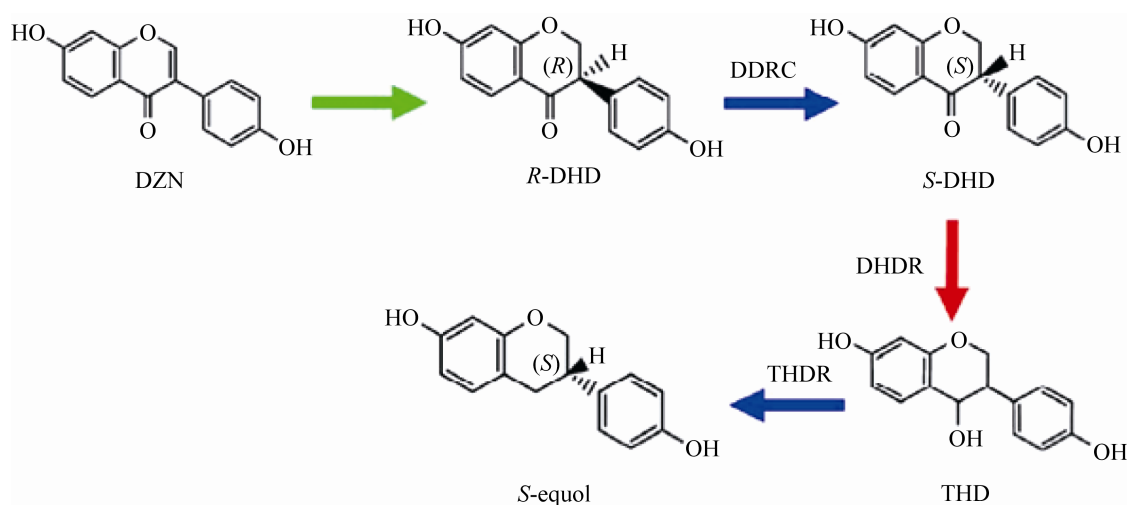


图 1 S-雌马酚的催化合成途径示意图

Figure 1 Biosynthesis pathway of S-equiol.

1.2 梭菌 C1 的培养与基因组 DNA 提取

取保存在 -80°C 的梭菌 C1 接种到 10 mL GAM 培养基中,在 37°C 厌氧工作站培养 36 h,然后以 5%的比例传代接种到 100 mL 培养基中,继续培养 36 h。将培养的菌液在 10 000 r/min 条件下离心 5 min,收集的菌体通过 DNA 提取试剂盒进行 DNA 提取,并分别通过 0.75%琼脂糖凝胶电泳和微量分光光度计测定提取 DNA 的浓度、纯度和降解程度。

1.3 全基因组测序

本次测序委托武汉未来组生物科技有限公司完成。将检测合格的 DNA 样品进行凝胶回收和 DNA 文库构建,利用 GridION 测序仪进行单分子实时测序。

1.4 基因组组装和结构分析

对测序样品依次进行第二代和第三代测序,获得的数据首先利用 canu 软件^[12]进行初步组装,然后根据第二代测序的数据,通过 pilon 软件^[13]对第三代测序数据进行矫正、冗余剔除和基因组成环处理,再通过 circlator 软件^[14]将成环序列的原点移动到基因组的复制起始位点,获得最终的基因组序列。

C1 菌编码基因用 prodial 软件^[15]进行预测;tRNA 基因用 tRNAscan-SE 软件进行预测;rRNA 基因用 RNAmmer 软件^[16]进行预测,并在 NCBI 中通过 16S rRNA 序列比对分析,获取 C1 菌的分类信息;其他 ncRNA 用 infernal 软件搜索 Rfam 数据库进行预测^[17];CRISPR 用 minced 软件^[18]进行预测,基因岛用 Islander 软件^[19]进行预测。在 NCBI 中通过 BLASTn 分析,将 C1 菌基因组序列与质粒数据库进行比对(<https://ftp.ncbi.nlm.nih.gov/refseq/release/plasmid/>),预测潜在的质粒信息。

1.5 基因组功能注释

根据 C1 菌基因组编码蛋白序列信息,通过

BLAST 软件在以下数据库中对 C1 的基因组进行功能注释: TIGRFAMs^[20]、Rfam、Pfam^[21]、GO^[22]、KEGG^[23]、refseq^[24]、COG^[25],然后根据注释结果进行分类统计分析。

1.6 产 S-雌马酚功能基因筛选分析

根据已知主要产 S-雌马酚细菌 *Lactococcus* sp. 20-92 (Lac 20-92, GenBank: AB558141.1) 和 *Eggerthella* sp. YY7918 (YY7918, GenBank: AP012211.1)的 4 种产 S-雌马酚功能基因信息: DZNR、DDRC、DHDR 和 THDR,利用 NCBI 的 Local BLAST 工具(<https://blast.ncbi.nlm.nih.gov/Blast.cgi>)对 C1 菌的全基因组序列进行序列相似性分析,筛选 C1 基因组中可能参与 S-雌马酚生物合成的功能基因;然后将筛选基因在 NCBI 的 conserved domain database (CDD)中进行功能保守区域分析(<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>),并利用 Bio2Rxn (<http://design.rxnfinder.org/bio2rxn/>)和 Motif Scan (https://myhits.sib.swiss/cgi-bin/motif_scan)工具做酶促反应预测分析,获取候选基因可能具备的酶功能特性。

1.7 C1-07020 基因的克隆表达和鉴定

通过原核基因克隆表达技术,将 C1-07020 基因克隆到 pETDuet-1 质粒(fp: CCCGCGGCC GCATGAAAAACAAATATTACCCTCA, rp: CC GCCTAGGTTAGATTTGTCTGGCTGCTAT, 下划线为酶切位点),然后转化到大肠杆菌 BL21 (DE3)中,构建 pETDuet-1 C1-07020 BL21 (DE3)工程菌。该工程菌在含有 100 $\mu\text{g/mL}$ 氨苄青霉素 LB 选择培养基中, 37°C 、180 r/min 条件下孵育,当细菌 OD_{600} 为 0.6 时,加入终浓度为 0.25 mmol/L 的 IPTG,然后在 16°C 、120 r/min 条件下诱导表达 16 h;对诱导表达所得的菌体进行超声波破碎,并通过 SDS-PAGE 测定重组蛋白的表达情况;同时,取 1 mL 诱导表达后的菌液,加入

20 $\mu\text{g/mL}$ 的 DZN (100 mg/mL , DMSO 溶解), 在 37 $^{\circ}\text{C}$ 发酵 4 h, 然后通过以下方法检测发酵产物的生成情况。

在发酵后的样品中加入 0.7 mL 的乙酸乙酯, 剧烈振荡混匀, 静置 10 min, 然后在 5 000 r/min 条件下离心 5 min, 转移萃取上清液到新的离心管中, 再按同样的方式对样品萃取 2 次, 将总共 3 次的萃取液进行冷冻离心浓缩, 获得的冻干样品加入 0.2 mL 甲醇溶解, 最后通过 HPLC 检测 DZN 的转化情况, 检测条件为: 10%乙腈和 0.1%乙酸为流动相 A, 90%乙腈和 0.1%乙酸为流动相 B, 在流速 1 mL/min、70%流动相 A 和 30%流动相 B 的条件下洗脱 15 min, 检测温度 30 $^{\circ}\text{C}$, 检测波长 254 nm 和 275 nm, 进样量 10 μL 。

1.8 C1-07020 重组蛋白的功能检测

按 1.7 的方法表达 C1-07020 重组蛋白, 然后通过琼脂糖 Ni 亲和层析柱进行纯化, 并通过 SDS-PAGE 和 BCA 试剂盒测定纯化蛋白的纯度和浓度。以纯化的 C1-07020 重组蛋白为催化酶, 按以下反应体系测定重组蛋白的大豆胍元还原酶活性(1 mL 体系): 20 $\mu\text{g/mL}$ 的 C1-07020 重组蛋白, 10 $\mu\text{g/mL}$ 的 DZN, 500 $\mu\text{g/mL}$ 的 NADPH 或 500 $\mu\text{g/mL}$ 的 NADH, 0.5 mg/mL 的亚硫酸氢钠, 0.3 mg/mL 的二硫苏糖醇, 0.2 mg/mL 的苯甲基磺酰氟, 加入 100 mmol/L pH 7.0 磷酸钾作为缓冲液, 在 37 $^{\circ}\text{C}$ 厌氧工作站孵育 4 h。孵育完成后的样品按照实验 1.7 的 HPLC 方法检测 DZN 和 DHD。

2 结果与分析

2.1 梭菌 C1 全基因组概况

第三代全基因组测序结果显示, C1 菌基因组由 1 个 contig 组成, 大小为 3 035 113 bp, 平均(G+C)含量为 31.66%, 通过预测共得到 3 166 个完整的编码区(CDS), 占整个基因组的 89.51%。

进一步的基因组结构分析发现, C1 菌基因组中可能编码 53 个 tRNA、15 个 rRNA、4 个 ncRNA 和 1 个基因岛, 但没有预测到 CRISPR 和质粒序列, 所得到的核基因组见图 2。细菌分类结果显示, C1 菌与 *Clostridium* sp. TM-40 和 *Clostridium* sp. AUH-JLC39 的 16S rRNA 相似性分别为 99.93%和 99.64%。梭菌 C1 的全基因组数据已提交至 GenBank, 登录号为 CP073631。

2.2 梭菌 C1 基因组注释分析

利用 TIGRFAMs、Rfam、Pfam、GO、KEGG、refseq 和 COG 数据库, 对 C1 菌全基因组中 3 166 个 CDS 进行蛋白功能注释, 统计结果见表 1。其中, 在所有数据库中都出现的功能基因数量为 559 个, 占比 17.66%, 而至少在一个数据库当中出现的功能基因数量达到 3 048 个, 占比为 96.27%。进一步对 COG 和 KEGG 数据库中注释的功能基因进行分析, 以获取 C1 菌中与 *S*-雌马酚合成相关的次级代谢产物合成通路的基因信息。

2.2.1 COG 功能注释结果

将 C1 菌注释基因的氨基酸序列在 COG 数据库中做比对分析。总共有 1 665 个 CDS 得到注释(图 3), 按照 24 种不同的功能属性进行分类, 其中与转录(transcription)以及翻译、核糖体结构和生物形成(translation, ribosomal structure and biogenesis)相关的预测基因数量最多, 分别为 214 和 210 个, 占注释基因总数的 12.85%和 12.61%。其次是与碳水化合物转运和代谢(carbohydrate transport and metabolism)以及氨基酸转运与代谢(amino acid transport and metabolism), 分别为 156 和 136 个, 占注释基因总数的 9.37%和 8.17%, 其他功能属性的注释基因数量比例相对较少。此外, 在所有注释基因中, 与次级代谢物生物合成、运输和分解代谢相关的注释基因有 20 个, 这其中可能包含参与 *S*-雌马酚生物合成的功能基因。

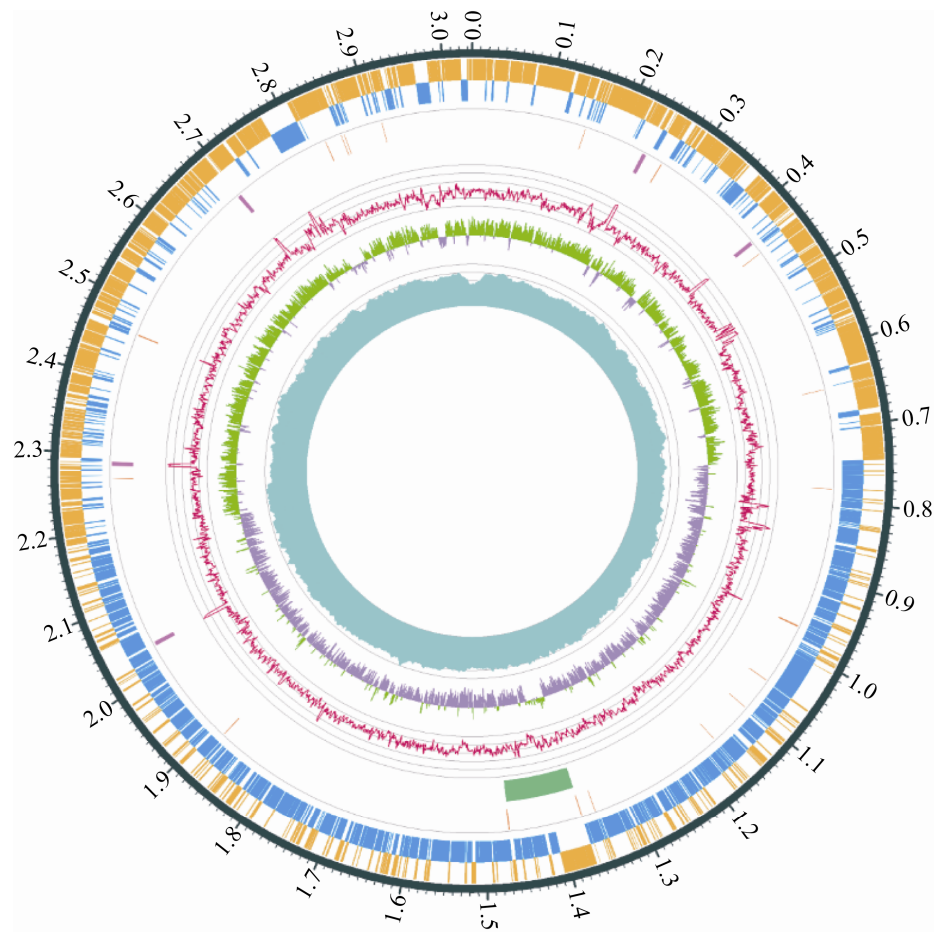


图 2 梭菌 C1 核基因组圈图
Figure 2 Nuclear genome of *Clostridium* sp. C1. From the outside to the inside: coding genes (plus strand), coding genes (minus strand), tRNA (orange) and rRNA (blue), gene island (green), GC content, GC-skew, sequencing depth.

表 1 梭菌 C1 基因组功能蛋白注释统计
Table 1 Functional protein annotation statistics of *Clostridium* sp. C1 genome

Database	Number of annotated functional proteins	Proportion/%
COG	1 665	52.59
KEGG	1 324	41.82
GO	1 636	51.67
Refseq	3 039	95.99
Pfam	2 395	75.65
TIGRFAMs	757	23.91
All the databases	559	17.66
At least one database	3 048	96.27
Total	3 166	100.00

2.2.2 KEGG 功能注释结果

C1 菌全基因组的 KEGG 数据库分析结果显示，总共有 1 324 个基因获得功能注释，根据代谢通路相关信息可分为 6 大类(图 4)。在这些功能注释基因中，主要归属于代谢(metabolism)、遗传信息处理(genetic information processing)和环境信息处理(environmental information processing)相关通路，其中，又以代谢相关的注释基因数量最多。进一步分析发现，与次级代谢生物合成通路相关的基因有 19 个，参与外来物生物降解与代谢通路的注释基因有 17 个，它们可能与 C1 菌的 S-雌马酚生成代谢通路相关。

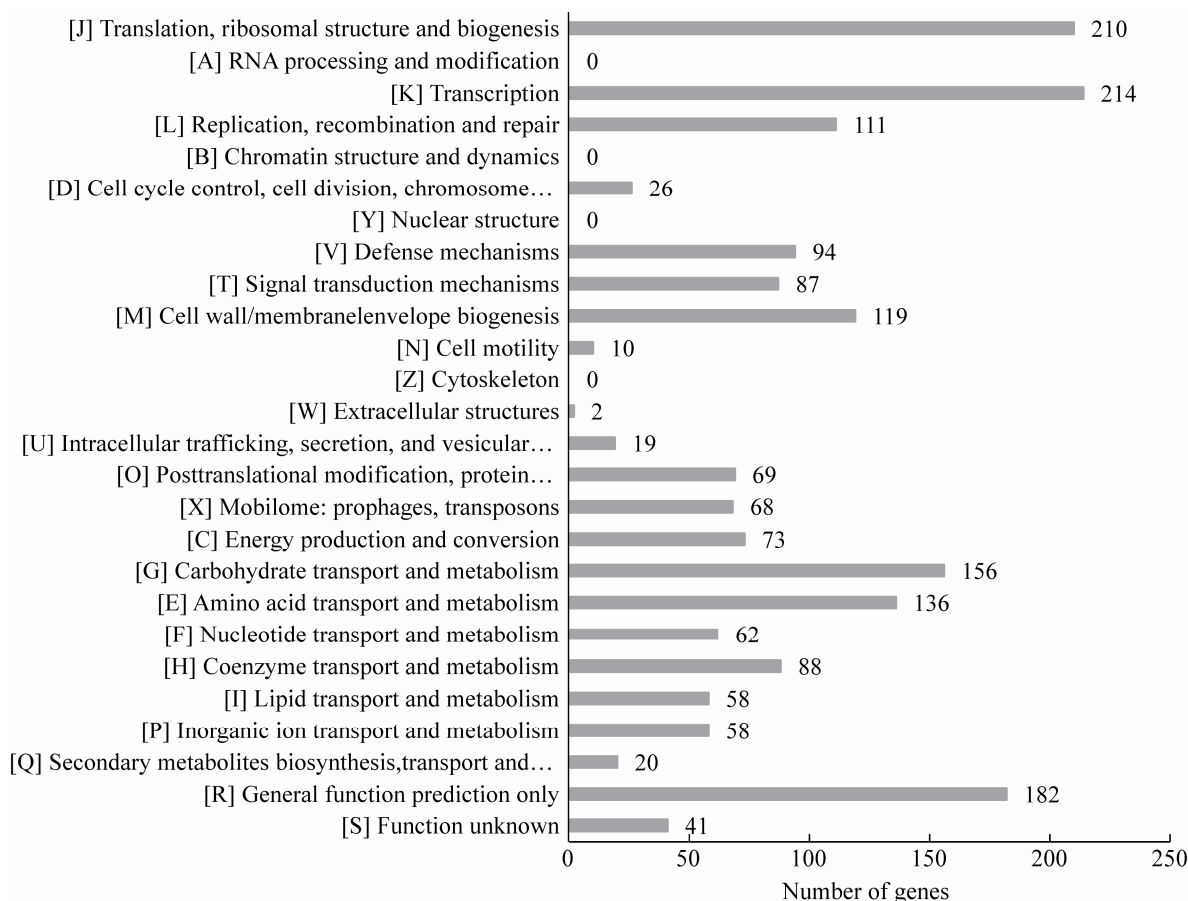


图3 C1 菌基因组编码蛋白 COG 功能分类统计

Figure 3 Classification and statistics of COG function of C1 genome coding proteins.

2.3 产 *S*-雌马酚功能基因筛选

利用已知产 *S*-雌马酚细菌 Lac 20-92 和 YY7918 的 4 个功能基因序列信息(DZNR、DDRC、THDR 和 THDR), 对梭菌 C1 的所有 CDS 进行氨基酸序列相似性分析和潜在功能区域分析, 并结合 COG 和 KEGG 的基因注释结果, 从 C1 菌中获取产 *S*-雌马酚功能基因信息。结果显示, 从 C1 中筛选到一个与 Lac 20-92 的 DZNR 相似性最高的基因——C1-07020 (相似性为 44.8%)。进一步的保守结构域分析发现, C1-07020 中具有与 DZNR 类似的旧黄色酶家族黄素单核苷酸结合区域(OYE-like FMN binding domain)、[4Fe-4S]簇基序(4Fe-4S cluster motif)和辅酶结合基序(coenzyme binding motifs)保守

结构域(图 5)。同时, 酶功能预测分析显示, C1-07020 具有氧化还原酶活性, 是一种 NAD(P)H 依赖型脱氢酶, 在 NAD(P)H 存在条件下, 具有促进氢原子转移的还原酶功能, 这些预测的功能特点都与 DZNR 的大豆苜蓿还原酶功能相似。另一方面, 在 C1 菌基因组中并没有发现与已知 DDRC、DHDR 和 THDR 相似的基因, 也未能通过保守结构域或酶功能分析发现与这 3 个功能基因相似的信息。

2.4 C1-07020 基因功能验证

为了验证 C1-07020 是否具有 DZNR 的大豆苜蓿还原酶功能, 通过原核基因克隆表达技术, 将 C1-07020 基因构建到 pETDuet-1 质粒中, 并转化到大肠杆菌 BL21 (DE3)中进行表达和纯化,

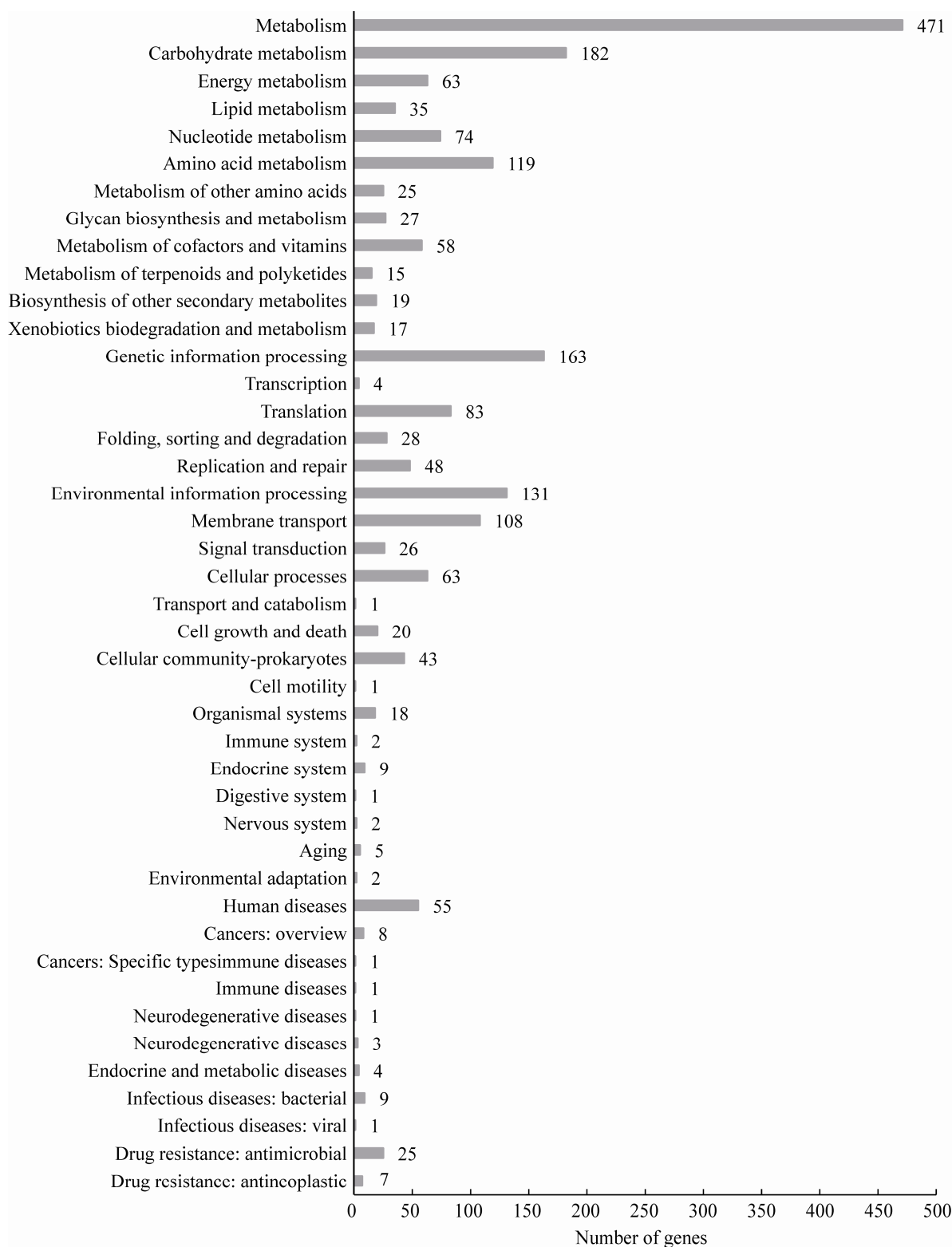


图 4 C1 菌基因组编码蛋白 KEGG 功能分类统计

Figure 4 Functional classification and statistics of KEGG annotated genes encoded by C1 genome.

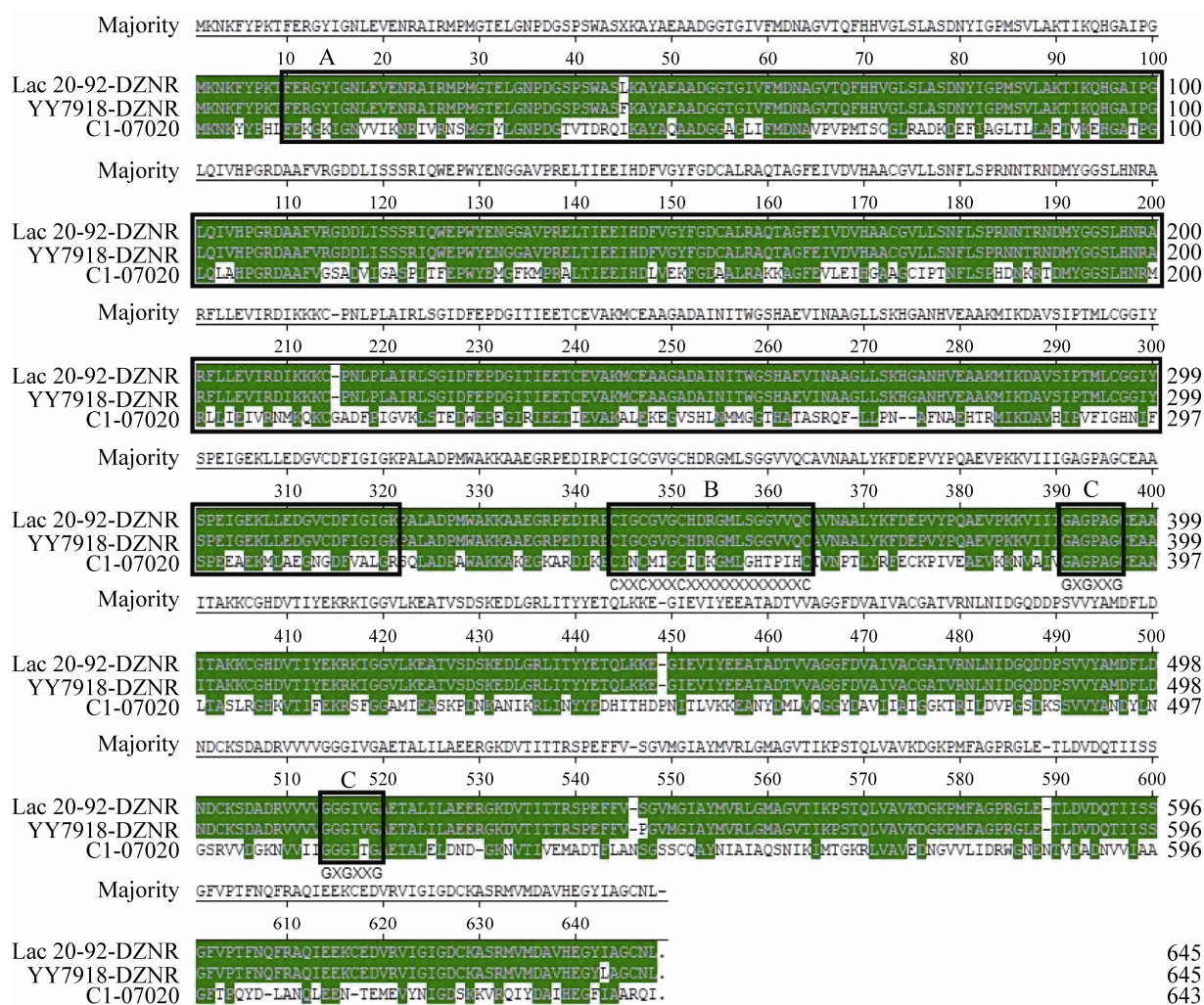


图 5 C1-07020 与 DZNR 保守结构域分析

Figure 5 Conserved domains analysis of C1-07020 and DZNR. The same amino acid residues are represented in green; A: OYE-like FMN binding domain; B: 4Fe-4S cluster motif; C: coenzyme binding motifs; the common amino acid residues of 4Fe-4S cluster motif and coenzyme binding motifs are below box B and C, respectively.

最终获得可溶性表达的 C1-07020 重组蛋白，浓度可达 1.05 mg/mL。以 DZN 为底物，测定 C1-07020 重组蛋白的酶促反应，HPLC 检测结果表明，在辅酶 NADH 的存在以及严格厌氧环境下，反应液中有 DHD 的生成，说明 C1-07020 重组蛋白具备大豆昔元还原酶的功能，可将 DZN 转化为 DHD。并且发现，NADPH 不能代替 NADH 满足 C1-07020 的酶活性要求(图 6)。

2.5 C1 菌产 S-雌马酚特性分析

已报道的产 S-雌马酚菌都具有 S-雌马酚合成代谢基因簇的特点。首先对 C1-07020 可能所处的代谢通路进行分析，但在 COG 和 KEGG 和 GO 等数据库中都未发现与 C1-07020 相关的代谢通路。在确定 C1-07020 是产 S-雌马酚功能基因后，对 C1-07020 基因编码区上下游 30 个基因作进一步的分析(图 7)。保守结构域、酶功能

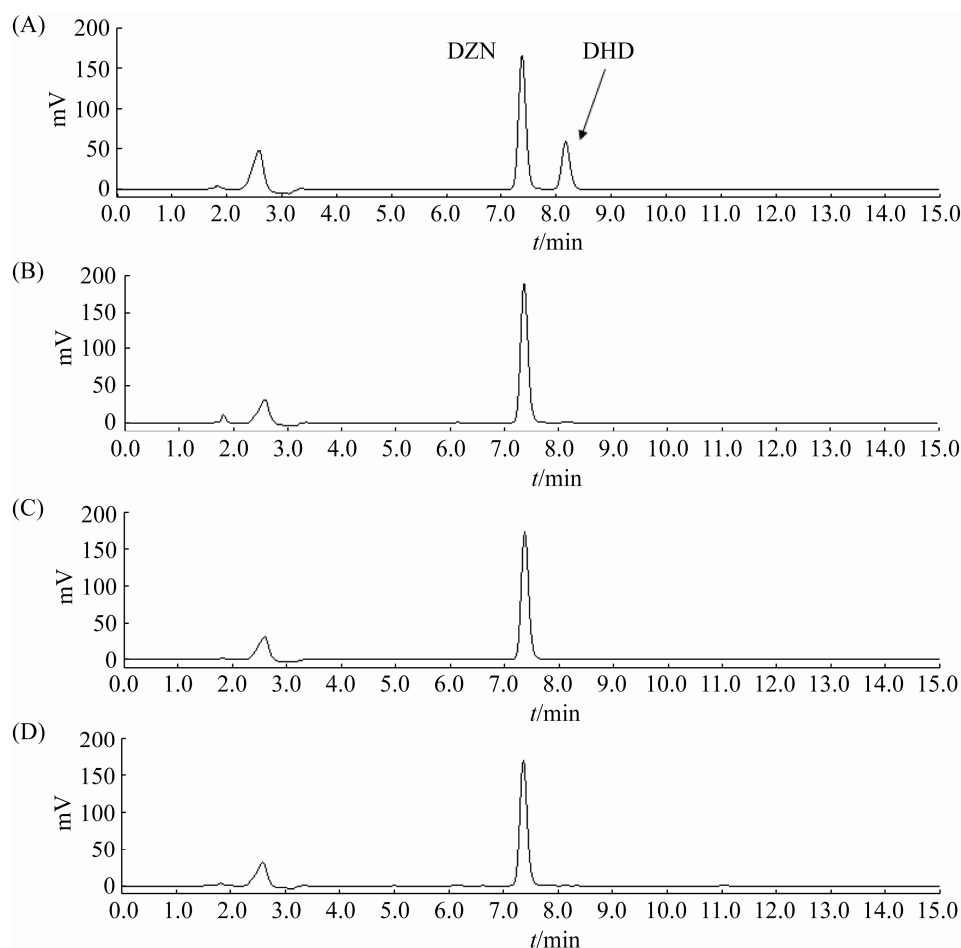


图 6 C1-07020 重组蛋白的大豆苷元还原酶功能检测

Figure 6 Detection of daidzein reductase function of C1-07020 recombinant protein. A: reaction solution with C1-07020 recombinant protein, NADH under anaerobic condition; B: reaction solution with C1-07020 recombinant protein, NADPH under anaerobic condition; C: reaction solution with C1-07020 recombinant protein, NADH under aerobic condition; D: reaction solution without C1-07020 recombinant protein.

和 motif scan 比对分析结果显示, 在这些基因中, 并没有与 DDRC、DHDR 和 DHDR 相似的基因, 也没有发现可能类似的基因功能特性, COG、KEGG 和 GO 等数据库中也未发现与此相关的代谢通路信息, 因此, C1 菌中产 *S*-雌马酚基因可能并不是以基因簇形式存在。

3 讨论

为了解 C1 菌的产 *S*-雌马酚特性, 本研究利用第三代测序技术, 对 C1 菌的全基因组进行

测序, 获取细菌的全基因组数据, 并根据已报道参与 *S*-雌马酚合成的功能基因信息, 通过生物信息学方法在 C1 菌中筛选到一个潜在的 DZNR 基因——C1-07020。体外功能验证实验表明, C1-07020 具有 DZNR 的功能, 可将 DZN 转化为 DHD (图 6), 这是 *S*-雌马酚生物合成的第一步酶促反应。

进一步将 C1-07020 编码蛋白序列与已报道主要产 *S*-雌马酚菌的 DZNR 进行对比分析, 发现 C1-07020 具有与 DZNR 相似的 3 个保守结

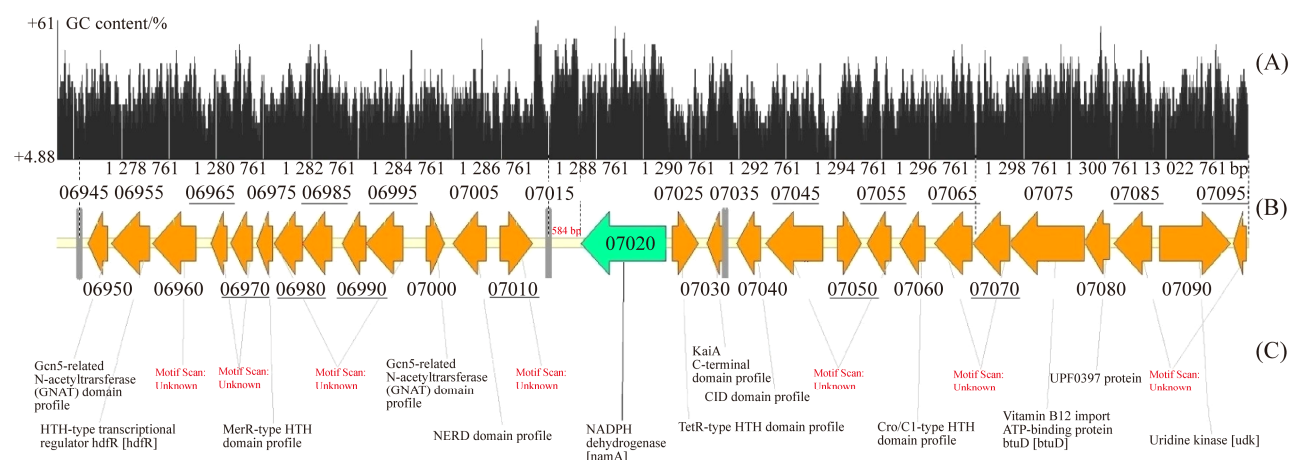


图 7 梭菌 C1 可能的产 *S*-雌马酚基因簇分析

Figure 7 Analysis of the possible *S*-equol gene cluster in *Clostridium* C1. A: GC content of potential *S*-equol-producing gene cluster; B: sequence information of upstream and downstream genes of C1-07020 predicted by whole genome sequencing; C: functional prediction analysis of each CDS.

构域(图 5), 属于旧黄色酶(OYE)家族蛋白^[26]。C1-07020 编码蛋白的 N 端包含一个 OYE 样 FMN 结合域、一个 4Fe-4S 铁硫簇基序(CXXCX3CX12C)和 2 个辅酶结合基序(GXGXG)。氢原子转移在多种生物氧化还原反应中起关键作用, 在已知的 OYE 家族蛋白中, FMN 是底物的结合区, 并可作为电子转移的载体; 4Fe-4S 铁硫簇基序和辅酶基序则参与辅酶的结合以及电子的转移, 从而促进氢原子的转移^[27]。DZN 转化为 DHD 正是一种氢原子转移的生物还原反应, C1-07020 体外功能验证结果表明, 只有在 NADH 和严格的厌氧条件下, 才能实现 DHD 的转化(图 6), NADH 可能为酶促反应提供还原力, 但 NADPH 无此功能, 厌氧条件也与 DZNR 的氧敏感相似^[9], 具有较为典型的 OYE 家族蛋白特性。

另一方面, C1-07020 与 *Lactococcus* sp. 20-92 DZNR 的氨基酸序列相似性并不高(44.8%), 表明不同的 DZNRs 可能具有不同的酶促特性, 例如, C1-07020 是 NADH 依赖型还原酶(图 6A-B), 而 Lac 20-92 DZNR 则属于

NADPH 依赖型还原酶^[9]。产 *S*-雌马酚细菌能高效转化投入发酵液的 DZN 为 DHD, 但无论是已知的 DZNRs 还是 C1-07020, 重组蛋白的体外酶活性都不高, 即使提高 NADH 浓度和保持严格的厌氧环境也不能有效提高酶活性(数据未给出), 推测 DZN 还原酶活性可能还受到其他辅助因子的影响。

在确定了 C1-07020 是 C1 菌的产 *S*-雌马酚功能基因后, 进一步在 C1 菌的全基因组中筛选其他产 *S*-雌马酚功能基因, 但无论是功能基因注释、次级代谢通路预测还是基因簇分析^[9,28-29], 都未能找到类似 DDRC、DHDR 或 THDR 的功能基因。对 C1-07020 基因上下游区域进行分析(可能的基因簇, 图 7), 并对部分未知功能的基因进行了体外功能验证(数据未给出), 也并没有发现参与 *S*-雌马酚生物合成的其他功能基因。因此, C1 菌中可能没有产 *S*-雌马酚基因簇, 并可能存在不同于 DDRC、DHDR 或 THDR 的产 *S*-雌马酚功能基因。

本研究从梭菌 C1 的全基因组中筛选和鉴定到一个具有 DZNR 功能的产 *S*-雌马酚功能基

因 C1-07020, 并猜测 C1 菌可能具有新的产 *S*-雌马酚功能基因和不同于已知产 *S*-雌马酚菌的代谢机制。后续研究获得新类基因可为体外 *S*-雌马酚的生物发酵生产提供基因资源, 所掌握的 C1 菌基础信息可为进一步挖掘 *S*-雌马酚合成新基因、了解 *S*-雌马酚的生成机制提供重要参考。

参考文献

- [1] Mayo B, Vázquez L, Flórez AB. Equol: a bacterial metabolite from the daidzein isoflavone and its presumed beneficial health effects. *Nutrients*, 2019, 11(9): E2231.
- [2] Daily JW, Ko BS, Ryuk J, Liu ML, Zhang WJ, Park S. Equol decreases hot flashes in postmenopausal women: a systematic review and meta-analysis of randomized clinical trials. *Journal of Medicinal Food*, 2019, 22(2): 127–139.
- [3] Matsumoto T, Kojima M, Takayanagi K, Taguchi K, Kobayashi T. Role of *S*-equol, indoxyl sulfate, and trimethylamine N-oxide on vascular function. *American Journal of Hypertension*, 2020, 33(9): 793–803.
- [4] Fatima A, Khan MS, Ahmad MW. Therapeutic potential of equol: a comprehensive review. *Current Pharmaceutical Design*, 2020, 26(45): 5837–5843.
- [5] Feng XL, Ho SC, Zhan XX, Zuo LSY, Mo XF, Zhang X, Abulimiti A, Huang CY, Zhang CX. Serum isoflavones and lignans and odds of breast cancer in pre- and postmenopausal Chinese women. *Menopause*, 2021, 28(4): 413–422.
- [6] 胡云霏, 陈华海, 尹业师. 雌马酚产生细菌及其雌马酚合成代谢机制. *微生物学报*, 2019, 59(8): 1452–1462.
Hu YF, Chen HH, Yin YS. Research progress in equol-producing bacteria and their metabolism. *Acta Microbiologica Sinica*, 2019, 59(8): 1452–1462. (in Chinese)
- [7] Li BJ. Advances in exploring equol production and application. *Journal of Food Processing and Preservation*, 2019, 43(11): e14205.
- [8] Shimada Y, Takahashi M, Miyazawa N, Abiru Y, Uchiyama S, Hishigaki H. Identification of a novel dihydrodaidzein racemase essential for biosynthesis of equol from daidzein in *Lactococcus* sp. strain 20-92. *Applied and Environmental Microbiology*, 2012, 78(14): 4902–4907.
- [9] Schröder C, Matthies A, Engst W, Blaut M, Braune A. Identification and expression of genes involved in the conversion of daidzein and genistein by the equol-forming bacterium *Slackia isoflavoniconvertens*. *Applied and Environmental Microbiology*, 2013, 79(11): 3494–3502.
- [10] Lee PG, Lee UJ, Song H, Choi KY, Kim BG. Recent advances in the microbial hydroxylation and reduction of soy isoflavones. *FEMS Microbiology Letters*, 2018, 365(19): fny195.
- [11] Li HL, Mao SM, Chen HH, Zhu LY, Liu W, Wang X, Yin YS. To construct an engineered (*S*)-equol resistant *E. coli* for *in vitro* (*S*)-equol production. *Frontiers in Microbiology*, 2018, 9: 1182. DOI: 10.3389/fmicb.2018.01182.
- [12] Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Research*, 2017, 27(5): 722–736.
- [13] Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng QD, Wortman J, Young SK, Earl AM. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*, 2014, 9(11): e112963.
- [14] Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biology*, 2015, 16: 294.
- [15] Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 2010, 11: 119.
- [16] 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.
- [17] Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, Bateman A, Finn RD, Petrov AI. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Research*, 2018, 46(D1): D335–D342.
- [18] Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, Hugenholtz P. CRISPR recognition tool (CRT): a tool for automatic detection of clustered

- regularly interspaced palindromic repeats. *BMC Bioinformatics*, 2007, 8: 209.
- [19] Hudson CM, Lau BY, Williams KP. Islander: a database of precisely mapped genomic islands in tRNA and tmRNA genes. *Nucleic Acids Research*, 2015, 43(D1): D48–D53.
- [20] Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E. TIGRFAMs and genome properties in 2013. *Nucleic Acids Research*, 2012, 41(D1): D387–D395.
- [21] Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, 2016, 44(D1): D279–D285.
- [22] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. *Nature Genetics*, 2000, 25(1): 25–29.
- [23] Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research*, 2014, 42(D1): D199–D205.
- [24] O’Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao YM, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li WJ, Maglott D, Masterson P, McGarvey KM, Murphy MR, O’Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun HZ, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, Pruitt KD. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, 2016, 44(D1): D733–D745.
- [25] Galperin MY, Makarova KS, Wolf YI, Koonin EV. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Research*, 2015, 43(D1): D261–D269.
- [26] Willistein M, Bechtel DF, Müller CS, Demmer U, Heimann L, Kayastha K, Schünemann V, Pierik AJ, Ullmann GM, Ermler U, Boll M. Low potential enzymatic hydride transfer via highly cooperative and inversely functionalized flavin cofactors. *Nature Communications*, 2019, 10: 2074.
- [27] Ernst C, Kayastha K, Koch T, Venceslau SS, Pereira IAC, Demmer U, Ermler U, Dahl C. Structural and spectroscopic characterization of a HdrA-like subunit from *Hyphomicrobium denitrificans*. *The FEBS Journal*, 2021, 288(5): 1664–1678.
- [28] Flórez, Vázquez, Rodríguez, Redruello, Mayo. Transcriptional regulation of the equol biosynthesis gene cluster in *Adlercreutzia equolifaciens* DSM19450T. *Nutrients*, 2019, 11(5): 993.
- [29] Kawada Y, Yokoyama S, Yanase E, Niwa T, Suzuki T. The production of *S*-equol from daidzein is associated with a cluster of three genes in *Eggerthella* sp. YY7918. *Bioscience of Microbiota, Food and Health*, 2016, 35(3): 113–121.

(本文责编 李磊)