

# 嗜盐四联球菌 CGMCC 3792 全基因组测序和胞外多糖基因簇分析

张敏, 金焱, 黄钧, 周荣清, 迟原龙, 吴重德\*  
(四川大学轻工科学与工程学院 成都 610065)

**摘要** 目的:嗜盐四联球菌为一种乳酸菌,其产生的高分子聚合物胞外多糖具有重要的应用价值。为探究嗜盐四联球菌产胞外多糖的分子机制,对该菌进行全基因组测序,并对其基因组和胞外多糖基因簇进行分析。方法:采用纳米孔三代测序技术对嗜盐四联球菌 CGMCC 3792 的全基因组进行测序,采用生物信息学方法对其进行基因预测和功能注释,并深入挖掘合成胞外多糖的相关基因,在基因水平上分析嗜盐四联球菌 CGMCC 3792 合成胞外多糖的分子机制。结果:嗜盐四联球菌 CGMCC 3792 基因组全长 2 573 722 bp,编码 2 633 个基因。对胞外多糖生物合成途径进行分析,发现其可利用甘露糖、果糖、麦芽糖、葡萄糖和纤维二糖合成胞外多糖。通过基因组分析发现,该基因组含有 1 个 15.4 kb 的胞外多糖基因簇,其中包括起调控作用的基因 *epsA* 和 *epsB*,决定链长的基因 *epsC* 和 *epsD*,以及 2 个糖基转移酶基因和编码翻转酶的基因 *wzx*。结论:本研究在基因水平上揭示了嗜盐四联球菌 CGMCC 3792 产胞外多糖的分子机制,为其工业化应用奠定了基础。

**关键词** 嗜盐四联球菌; 纳米孔测序; 胞外多糖; 基因簇

文章编号 1009-7848(2025)01-0314-08 DOI: 10.16429/j.1009-7848.2025.01.030

乳酸菌(Lactic acid bacteria, LAB)是世界公认的安全菌株(GARS),广泛应用于发酵食品中。LAB 产生的胞外多糖(Exopolysaccharide, EPS)因独特的理化性质与生物活性而受到越来越多学者的青睐,例如抗氧化、抗菌、抗癌、免疫调节等活性<sup>[1]</sup>。EPS 作为一种天然高分子聚合物,具有可再生、生物安全、生物降解和生物相容性等特点<sup>[2]</sup>,与合成化合物相比具有竞争优势。嗜盐四联球菌(*Tetragenococcus halophilus*, *T. halophilus*)是一种中度嗜盐乳酸菌,能在含 25%氯化钠的条件下生长,广泛存在于高盐发酵食品中,如酱油和鱼露<sup>[3]</sup>,在食品发酵过程中产生风味物质,改善发酵食品风味<sup>[4]</sup>。此外,前期的研究表明,嗜盐四联球菌产生的胞外多糖具有抗氧化、抗冷冻以及冻干保护作用<sup>[5]</sup>。本文利用牛津纳米孔三代测序技术对嗜盐四联球菌 CGMCC 3792 进行全基因组测序,通过基因组分析阐明该菌株的 EPS 合成途径,并分析其 EPS 基因簇,以期对 EPS 的发酵生产奠定理论基础。

收稿日期: 2024-01-12

基金项目: 国家自然科学基金项目(32272280,31871787)

第一作者: 张敏,女,博士生

通信作者: 吴重德 E-mail: cdwu@scu.edu.cn

## 1 材料与方法

### 1.1 试验菌株

本研究所用菌株嗜盐四联球菌 CGMCC 3792 (*Tetragenococcus halophilus* CGMCC 3792) 分离自酱油醪,经生理生化和 16S rDNA 序列分析鉴定,保藏于中国普通微生物菌种保藏管理中心(CGMCC)。

### 1.2 主要试剂、仪器和设备

MRS 肉汤培养基,英国 OXOID 公司;基因组 DNA 提取试剂盒 NucleoBond® HMW DNA kit,德国 MN NucleoBond 公司;Qubit™ dsDNA HS Assay Kit,美国 ThermoFisher 公司;Agencourt AMPure XP Beads,美国 Beckman 公司;其余试剂均为国产分析纯级。

GL-20G-II 型高速冷冻离心机,上海安亨科学仪器厂;SW-CJ-2FD 型超净工作台,苏州净化设备有限公司;JJ500 型分析天平,常熟双杰测试仪器厂;S220 型 Covaris 超声波 DNA 破碎仪,美国 Covaris 公司;WH-3 型微型漩涡混合仪,上海沪西分析仪器厂有限公司;Q33226 型 Thermo Qubit 4.0 荧光定量仪,美国 ThermoFisher 公司。

### 1.3 菌株培养与基因组 DNA 的提取

将-80℃甘油管保藏的嗜盐四联球菌 CGMCC 3792 接种于 5 mL 添加 6%氯化钠的 MRS 肉汤培养基,30℃静置培养 24 h,以 5%接种量接种于

100 mL 新鲜的添加 6%氯化钠的 MRS 肉汤培养基,连续培养 3 代。培养液以 8 000 r/min 在 4 ℃离心 5 min,并用无菌水清洗 2 次,采用细菌基因组 DNA 提取试剂盒 (NucleoBond® HMW DNA kit) 提取菌株基因组 DNA。

#### 1.4 文库构建与测序

利用 0.7%琼脂糖凝胶检测所提取 DNA 的完整性,采用 Qubit™ dsDNA HS Assay Kit 对 DNA 浓度精确定量。将得到的 DNA 用 Covaris 超声波 DNA 破碎仪破碎,制备成约 500 bp 的片段,补平末端并添加碱基 A, 连接接头序列, 通过 Agencourt AMPure XP Beads 纯化产物, 在 MinION Flow Cell (Oxford Nanopore Technologies,ONT) 上进行单分子纳米孔 DNA 测序。

#### 1.5 基因组注释

采用 Prokka 软件<sup>[6]</sup>和 NCBI NR (<http://ncbi.nlm.nih.gov/>) 数据库对基因组进行基因预测和注释。结合 RAST (Rapid annotation using subsystem technology)Server -RAST Annotation Server (基于子系统技术的快速注释服务器,<http://rast.nmpdr.org/>) 网上在线工具、COG (<http://www.ncbi.nlm.nih.gov/COG/>)、基因本体论 (Gene ontology,

GO,<http://www.geneontology.org/>)和京都基因与基因组百科全书 (Kyoto encyclopedia of genes and genomes,KEGG,<http://www.kegg.jp/>)<sup>[7]</sup>等数据库完成基因的预测及功能注释。采用碳水化合物活性酶 (Carbohydrate active enzymes,CAZy,<http://www.cazy.org/>)数据库<sup>[8]</sup>对基因及蛋白序列进行比对,获得对应碳水化合物活性酶注释信息。

## 2 结果与分析

### 2.1 嗜盐四联球菌 CGMCC 3792 的基因组特征

基于三代测序技术对嗜盐四联球菌 CGMCC 3792 进行全基因组测序,基因组圈图如图 1 所示,从外向内,第 1 圈表示基因组大小刻度(kb),第 2 圈表示 GC 含量,第 3 圈表示测序深度,第 4、5 圈分别表示正、负链基因元件 (CDS, rRNA, tRNA),第 6、7 圈分别表示正、负链的 CDS,不同颜色表示不同的 COG 功能分类。嗜盐四联球菌 CGMCC 3792 基因组全长为 2 573 722 bp, 平均 GC 含量为 36.09%, 包含 62 个 tRNA,15 个 rRNA。基因组共编码 2 633 个基因,基因总长度为 2 239 700 bp,占总基因组长度的 87.02%。

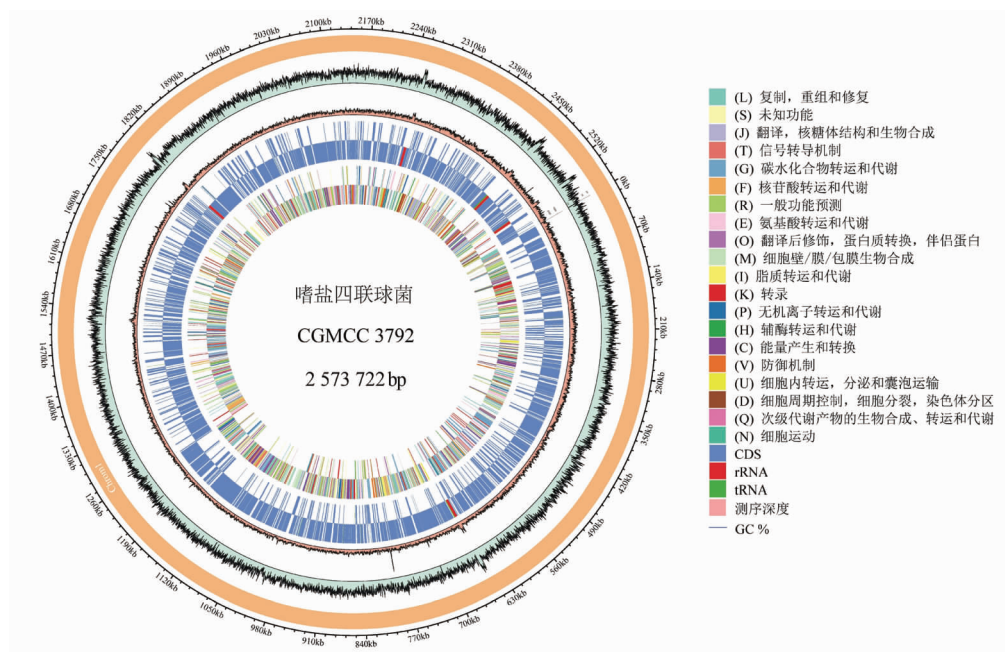


图 1 嗜盐四联球菌 CGMCC 3792 的基因组圈图

Fig.1 Circular genome map of *T. halophilus* CGMCC 3792

## 2.2 嗜盐四联球菌 CGMCC 3792 基因组功能注释

对嗜盐四联球菌 CGMCC 3792 基因组分别进行 COG、GO 和 KEGG 数据库比对注释。如图 2 所示,通过 COG 基因功能预测分析,在所有的 2 633 个基因中,有 1 945 个基因分别注释到 20

个 COG 类别中。除去功能未知的 182 个基因外,占比较大的是 R(一般功能预测)、G(碳水化合物转运和代谢)、L(复制、重组和修复)、K(转录)、J(翻译、核糖体结构和生物合成)以及 E(氨基酸转运和代谢),占比分别为 13.78%,12.08%,8.53%,8.07%,7.71%,7.61%。

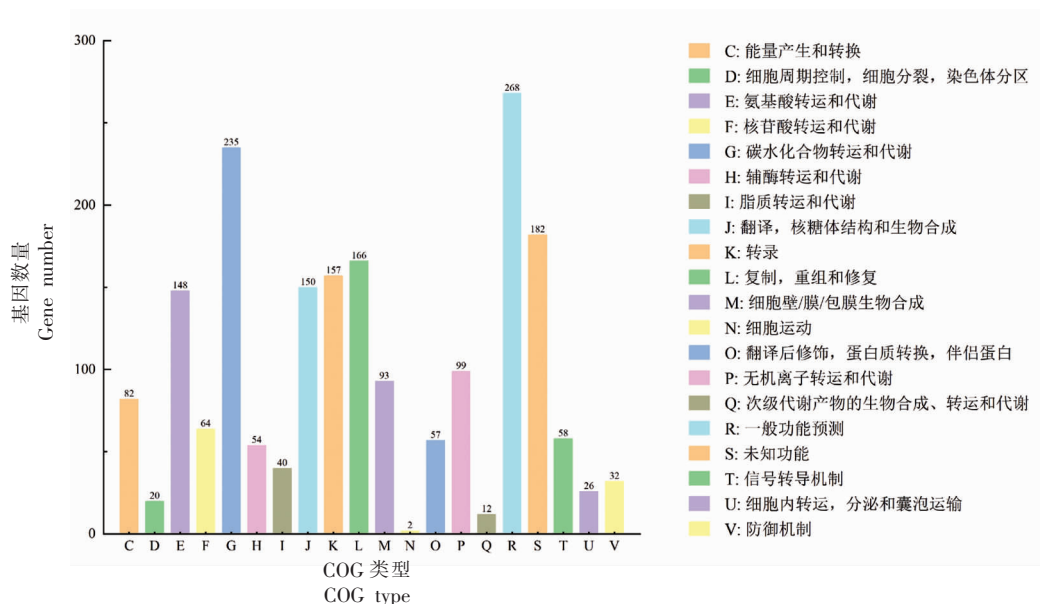


图 2 嗜盐四联球菌 CGMCC 3792 的 COG 功能分类图

Fig.2 COG functional classification map of *T. halophilus* CGMCC 3792

将嗜盐四联球菌 CGMCC 3792 基因组数据与 GO 数据库比对,结果如图 3 所示。共有 1 933 个基因分别注释到生物过程(Biological process)、细胞组分(Cellular component)和分子功能(Molecular function)3 个大类中,进一步可细分为 48 个二级功能类别。在生物过程大类的 22 个二级类别中,代谢过程(Metabolic process)和细胞过程(Cellular process)占比最高,分别有 1 286,1 255 个基因;此外,被注释到单生物过程(Single-organism process)、定位活性(Establishment of localization)、定位(Localization)、刺激反应(Response to stimulus)、生物调节(Biological regulation)、生物过程调控(Regulation of biological process)和细胞组织部分或生物合成(Cellular component organization or biogenesis)的基因数量也比较多。在细胞组分大类的 14 个二级分类中,细胞(Cell)和细胞部分(Cell part)占比最高,均有 993 个基因,其次是膜(Membrane)、膜部分(Membrane part)、高分

子复合物(Macromolecular complex)和细胞器(Organelle)。而在分子功能大类中有 12 个二级分类,占比最高的是催化活性(Catalytic activity),有 1 300 个基因,其次是结合活性(Binding)和转运活性(Transporter activity)。

如图 4 所示,共有 1 948 个基因注释到 KEGG 代谢通路中,共注释到 6 大类别,分别为细胞过程(Cellular processes)、环境信息处理(Environmental information processing)、遗传信息处理(Genetic information processing)、人类疾病(Human diseases)、新陈代谢(Metabolism)和生物体系统(Organismal systems)。其中大部分基因被注释到新陈代谢(Metabolism)中,在二级分类中,注释到碳水化合物代谢(Carbohydrate metabolism)的基因有 401 个,占比最高,而膜运输(Membrane transport)有 209 个基因,氨基酸代谢(Amino acid metabolism)有 189 个基因。

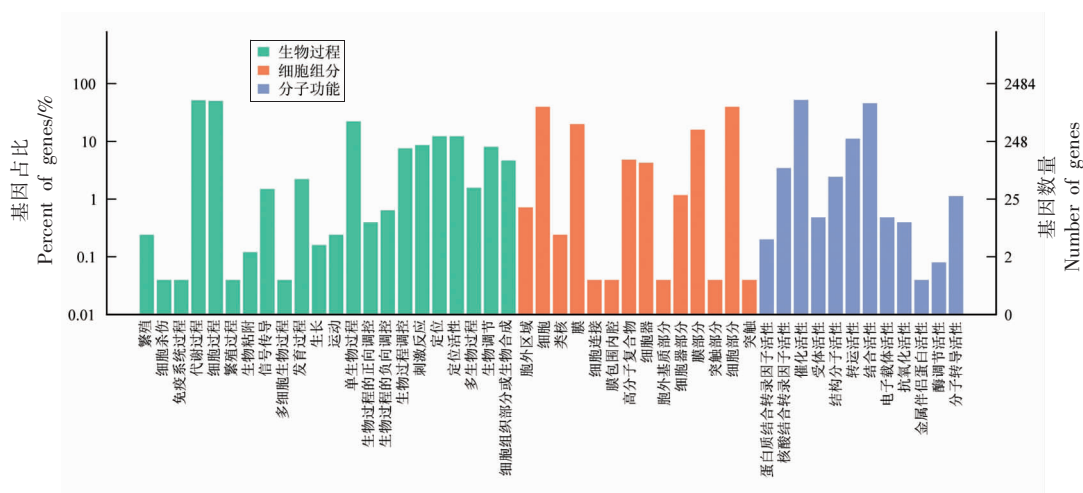


图 3 嗜盐四联球菌 CGMCC 3792 的 GO 功能分类图

Fig.3 GO functional classification map of *T. halophilus* CGMCC 3792

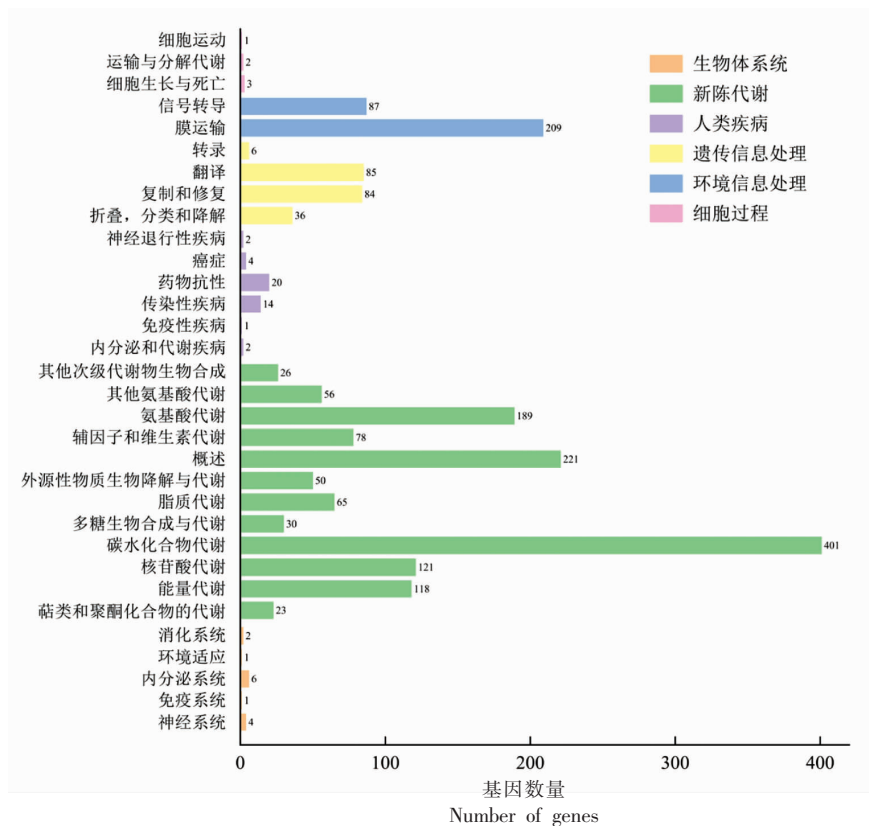


图 4 嗜盐四联球菌 CGMCC 3792 的 KEGG 途径分析

Fig.4 KEGG pathway analysis of *T. halophilus* CGMCC 3792

### 2.3 碳水化合物活性酶分析

通过 CAZy 分析了嗜盐四联球菌 CGMCC 3792 的碳水化合物活性酶功能基因, 共有 98 个基因注释到 5 大类别 (1 个基因可能注释到多个类别), 如图 5 所示, 分别为糖苷水解酶(Glycoside

hydrolases, GHs)、糖基转移酶 (Glycosyl transferases, GTs)、碳水化合物酯酶(Carbohydrate esterases, CEs)、辅助酶类 (Auxiliary activities, AAs)和碳水化合物结合模块(Carbohydrate-binding modules, CBMs)。其中, 注释到 GHs 家族蛋白



的基因最多,有45个,其次是CEs家族和GTs家族,分别为24,18个。GHs家族具有水解复杂碳水化合物化合物的能力,能代谢糖并形成各种糖核苷酸,被认为是参与碳水化合物代谢的关键酶<sup>[9]</sup>。在GHs大类中,占比较多的是GH13家族和GH1家族,分别有9个和8个基因,而GH13家族和GH1家族分别隶属于 $\alpha$ -葡萄糖苷酶和 $\beta$ -葡萄糖苷酶大类<sup>[10-11]</sup>。GTs家族是糖基转移酶家族,催化糖基从

糖基供体转移到糖基受体,形成EPS重复单元,是EPS合成过程中至关重要的酶类<sup>[9]</sup>。在GTs大类中,GT2、GT27、GT4、GT5和GT81家族均注释到4个基因。通过CAZy分析发现,嗜盐四联球菌CGMCC 3792含有丰富的糖苷水解酶和糖基转移酶,有助于细胞利用各类碳水化合物进行生长代谢和EPS的生物合成。

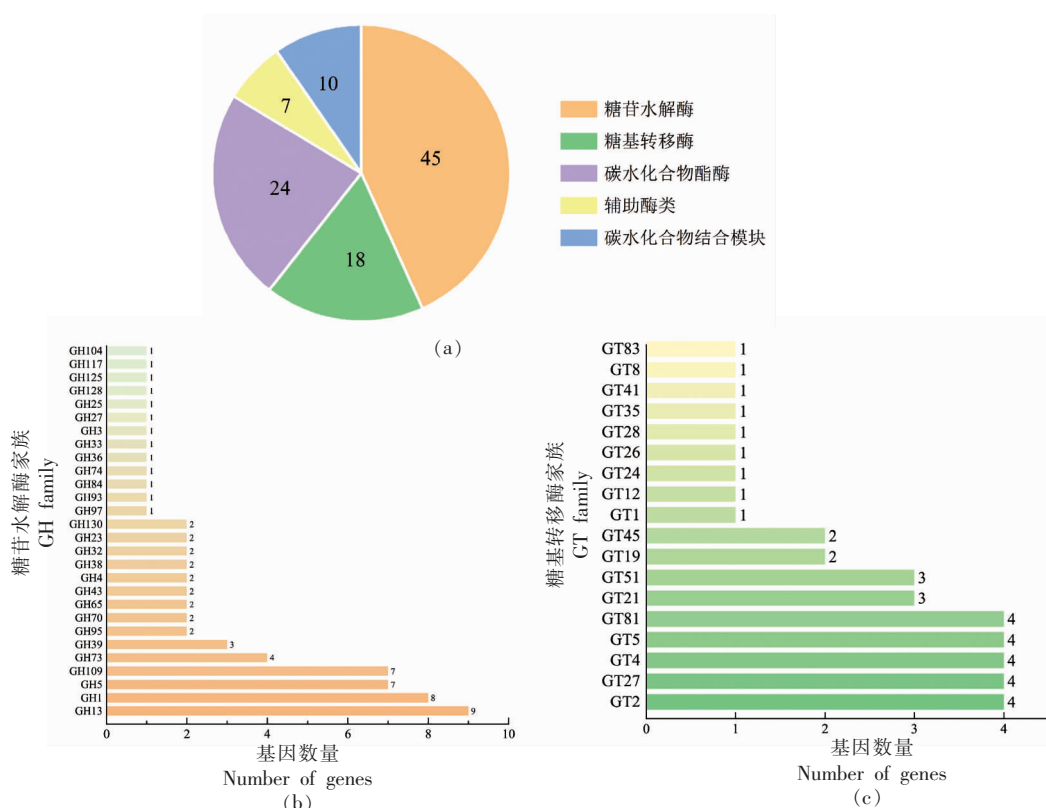


图5 CAZy在嗜盐四联球菌CGMCC 3792中的分布(a)、糖苷水解酶(b)和糖基转移酶(c)亚家族分类的基因个数分布

Fig.5 Distribution of CAZy in *T. halophilus* CGMCC 3792 (a), distribution of genes in GH (b) and GT (c) subfamilies

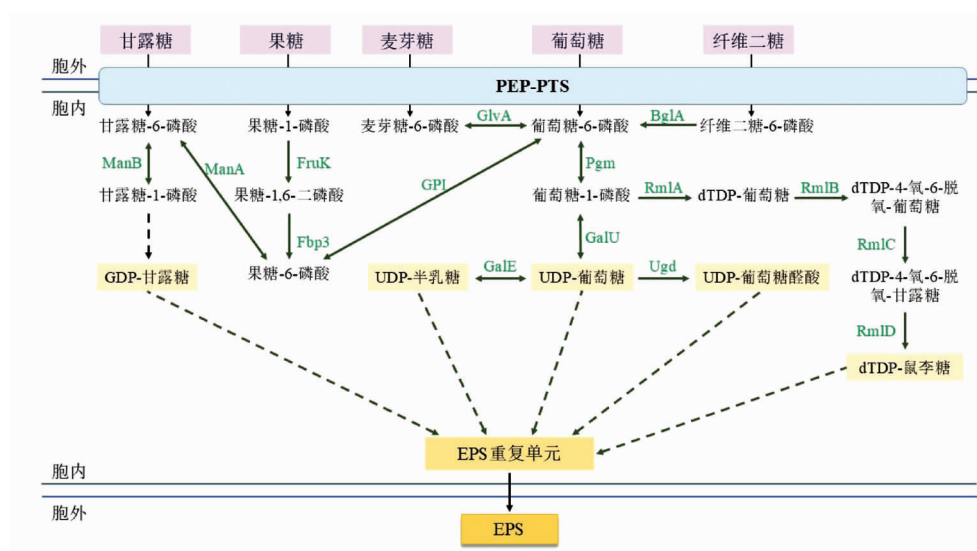
## 2.4 嗜盐四联球菌CGMCC 3792 EPS生物合成途径分析

乳酸菌胞外多糖独特的理化性质和生物活性使其拥有重要的应用价值,因此,阐明胞外多糖的生物合成途径对提高胞外多糖产量和加强其应用具有重要意义。磷酸烯醇式丙酮酸-磷酸转移酶系统(Phosphoenolpyruvate - phosphotransferase system, PEP - PTS)是广泛存在于乳酸菌和其它细菌中的一种糖转运系统,催化碳水化合物的转运

并伴随磷酸化。如图6所示,经生物信息学分析发现,嗜盐四联球菌CGMCC 3792可以利用甘露糖、果糖、麦芽糖、葡萄糖和纤维二糖作为碳源进行细胞生长和EPS合成。首先,嗜盐四联球菌CGMCC 3792利用PEP-PTS转运系统将甘露糖、果糖、麦芽糖、葡萄糖以及纤维二糖转运至细胞内,并进行磷酸化处理。其中,葡萄糖-6-磷酸作为重要的中间产物,可以与麦芽糖-6-磷酸、果糖-6-磷酸以及葡萄糖-1-磷酸分别在麦芽糖-6-磷酸葡

糖苷酶(GlvA)、葡萄糖-6-磷酸异构酶(GPI)和葡萄糖磷酸变位酶(Pgm)的催化下相互转化。而果糖-6-磷酸还可以在甘露糖-6-磷酸异构酶(ManA)的催化下与甘露糖-6-磷酸相互转化。葡萄糖-1-磷酸转化成 UDP-葡萄糖后,又可以在 UDP-葡萄糖-4-差向异构酶(GalE)和 UDP 葡萄糖 6-脱氢酶(Ugd)作用下,生成 UDP-半乳糖和 UDP-葡萄糖醛酸。另外,甘露糖转化为甘露糖-6-磷酸后,在磷酸甘露糖变位酶(ManB)的作用下,转化为甘

露糖-1-磷酸,随后转化为 GDP-甘露糖为 EPS 重复单元的合成提供前体。dTDP-鼠李糖从葡萄糖-1-磷酸经过葡萄糖-1-磷酸胸腺嘧啶转移酶(RmlA)、dTDP-葡萄糖 4,6-脱氢酶(RmlB)等一系列酶催化后生成。最后,GDP-甘露糖、UDP-半乳糖、UDP-葡萄糖、UDP-葡萄糖醛酸和 dTDP-鼠李糖在糖基转移酶的作用下,转移到脂质载体形成 EPS 重复单元,由翻转酶和跨膜蛋白输送到细胞外<sup>[12]</sup>,从而完成 EPS 的生物合成。



注:ManB:磷酸甘露糖变位酶;ManA:甘露糖-6-磷酸异构酶;FruK:1-磷酸果糖激酶;Fbp3:果糖-1,6-二磷酸酶;GPI:葡萄糖-6-磷酸异构酶;GlvA:麦芽糖-6-磷酸葡糖苷酶;Pgm:葡萄糖磷酸变位酶;GalU:UTP-葡萄糖-1-磷酸尿苷酰转移酶;GalE:UDP-葡萄糖-4-差向异构酶;Ugd:UDP 葡萄糖 6-脱氢酶;BglA:6-磷酸-β-葡糖苷酶;RmlA:葡萄糖-1-磷酸胸腺嘧啶转移酶;RmlB:dTDP-葡萄糖 4,6-脱氢酶;RmlC:dTDP-4-脱氢鼠李糖 3,5-异构酶;RmlD:dTDP-4-脱氢鼠李糖还原酶。

图 6 嗜盐四联球菌 CGMCC 3792 的 EPS 生物合成途径

Fig.6 EPS biosynthetic pathway of *T. halophilus* CGMCC 3792

## 2.5 嗜盐四联球菌 CGMCC 3792 EPS 基因簇分析

EPS 的产生通常都是由基因簇介导。通过基因组分析发现,嗜盐四联球菌 CGMCC 3792 染色体上存在一条完整的 EPS 基因簇,如图 7 所示,包含 15 个基因,全长 15.4 kb。其中,基因 *epsA* 和 *epsB* 分别编码 LytR 家族的转录调节因子和酪氨酸蛋白磷酸酶,负责调控 EPS 的产生<sup>[12-13]</sup>。基因 *epsA* 在 EPS 的生物合成中起着至关重要的作用,Dertli 等<sup>[14]</sup>研究表明,敲除 *epsA* 基因完全消除了 EPS 的产生。基因 *epsC* 和 *epsD* 分别编码链长决

定蛋白和酪氨酸蛋白激酶,链长决定蛋白通过控制重复单元的添加数量来控制多糖链的长度,对 EPS 的链长起决定作用<sup>[12,15]</sup>。有 2 个编码糖基转移酶的基因,分别是 *GTF1* 和 *GTF2*,主要负责将单糖转移至多糖链上,从而完成 EPS 重复单元的合成,在 EPS 的单糖组成中起关键作用<sup>[16-17]</sup>。基因 *wzx* 编码翻转酶(Flippase),识别重复单元并将其翻转过细胞膜<sup>[15]</sup>。此外,还有 1 个磷酸转移酶基因 *pglC*,4 个编码多糖生物合成蛋白的基因 *pglF*、*wbpA*、*rmlB* 和 *ugd*,以及 2 个假定蛋白基因。

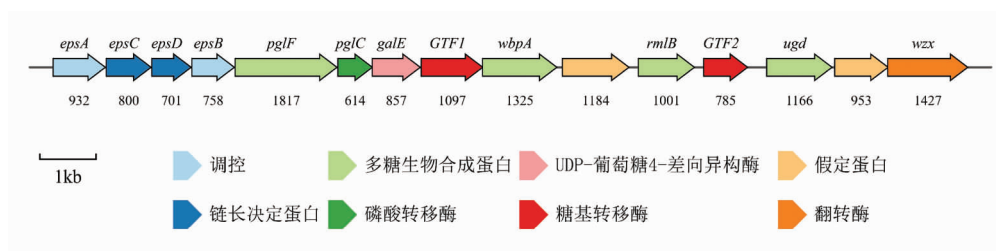


图7 嗜盐四联球菌 CGMCC 3792 的 EPS 基因簇

Fig.7 EPS gene cluster of *T. halophilus* CGMCC 3792

### 3 结论

采用纳米孔三代测序技术对嗜盐四联球菌 CGMCC 3792 进行了全基因组测序,对基因组上的基因进行了 COG、GO、KEGG 和 CAZy 注释。CAZy 分析表明,嗜盐四联球菌 CGMCC 3792 具有丰富的糖苷水解酶和糖基转移酶。同时,分析了嗜盐四联球菌 CGMCC 3792 的 EPS 合成途径和 EPS 基因簇,结果表明,嗜盐四联球菌 CGMCC 3792 可以利用甘露糖、果糖、麦芽糖、葡萄糖和纤维二糖等糖类进行细胞的生长代谢和 EPS 合成。本研究的相关结果将为嗜盐四联球菌 CGMCC 3792 生产 EPS 提供理论依据,同时为该菌株的工业化应用提供研究基础。

### 参 考 文 献

- [1] ZHOU Y, CUI Y H, QU X J. Exopolysaccharides of lactic acid bacteria: Structure, bioactivity and associations: A review [J]. Carbohydrate Polymers, 2019, 207: 317–332.
- [2] GUERREIRO B M, FREITAS F, LIMA J C, et al. Demonstration of the cryoprotective properties of the fucose-containing polysaccharide FucoPol[J]. Carbohydrate Polymers, 2020, 245: 116500.
- [3] UDOMSIL N, RODTONG S, TANASUPAWAT S, et al. Proteinase-producing halophilic lactic acid bacteria isolated from fish sauce fermentation and their ability to produce volatile compounds[J]. International Journal of Food Microbiology, 2010, 141 (3): 186–194.
- [4] KUDA T, IZAWA Y, ISHII S, et al. Suppressive effect of *Tetragenococcus halophilus*, isolated from fish-*nukazuke*, on histamine accumulation in salted and fermented fish[J]. Food Chemistry, 2012, 130 (3): 569–574.
- [5] ZHANG M, ZENG S Y, HAO L Y, et al. Structural characterization and bioactivity of novel exopolysaccharides produced by *Tetragenococcus halophilus* [J]. Food Research International, 2022, 155: 111083.
- [6] SEEMANN T. Prokka: Rapid prokaryotic genome annotation[J]. Bioinformatics, 2014, 30(14): 2068–2069.
- [7] KANEHISA M, GOTO S. KEGG: kyoto encyclopedia of genes and genomes [J]. Nucleic Acids Research, 2000, 28(1): 27–30.
- [8] LOMBARD V, GOLACONDA RAMULU H, DRULA E, et al. The carbohydrate-active enzymes database (CAZy) in 2013[J]. Nucleic Acids Research, 2014, 42(D1): D490–D495.
- [9] 代安娜尔, 乔少婷, 孙思霖, 等. 嗜热链球菌 I-MAU20246 胞外多糖基因簇及其表达分析[J]. 微生物学报, 2023, 63(1): 403–418.
- [10] DAI A N E, QIAO S T, SUN S L, et al. Identification and expression profiling of exopolysaccharide gene cluster in *Streptococcus thermophilus* I-MAU20246[J]. Acta Microbiologica Sinica, 2023, 63 (1): 403–418.
- [11] MENG X F, GANGOITI J, BAI Y X, et al. Structure-function relationships of family GH70 glucanase and 4, 6- $\alpha$ -glucanotransferase enzymes, and their evolutionary relationships with family GH13 enzymes[J]. Cellular and Molecular Life Sciences, 2016, 73: 2681–2706.
- [12] 范琴, 杨金茹, 唐湘华, 等. 粪便微生物宏基因组来源 GH1  $\beta$ -葡萄糖苷酶的重组表达及酶学性质[J]. 微生物学通报, 2021, 48(12): 4581–4599.
- [13] FAN Q, YANG J R, TANG X H, et al. Purification and characterization of GH1  $\beta$ -glucosidase from fecal microbes metagenome[J]. Microbiology China,

- 2021, 48(12): 4581–4599.
- [12] 乔少婷, 代安娜尔, 解敏, 等. *Streptococcus thermophilus* IMAU20551 胞外多糖基因簇及其表达分析[J]. 食品科学, 2022, 43(22): 9.  
QIAO S T, DAI A N E, XIE M, et al. *Streptococcus thermophilus* IMAU20551 eps gene cluster and its expression analysis[J]. Food Science, 2022, 43(22): 9.
- [13] 田佳乐, 刘洋, 李嘉雯, 等. 产胞外多糖嗜热链球菌 *eps* 基因簇的表达及生物信息学分析[J]. 中国食品学报, 2021, 21(11): 244–251.  
TIAN J L, LIU Y, LI J W, et al. Expression of eps gene cluster in *Streptococcus thermophilus* exopolysaccharide and bioinformatics analysis[J]. Journal of Chinese Institute of Food Science and Technology, 2021, 21(11): 244–251.
- [14] DERTLI E, MAYER M J, COLQUHOUN I J, et al. *EpsA* is an essential gene in exopolysaccharide production in *Lactobacillus johnsonii* FI9785[J]. Microbial Biotechnology, 2016, 9(4): 496–501.
- [15] 孔令慧, 赵林森, 夏永军, 等. 嗜热链球菌胞外多糖生物合成的研究进展[J]. 食品安全质量检测学报, 2019, 10(2): 284–290.  
KONG L H, ZHAO L S, XIA Y J, et al. Research advance in the exopolysaccharide biosynthesis of *Streptococcus thermophilus* [J]. Journal of Food Safety and Quality, 2019, 10(2): 284–290.
- [16] 李柏良, 赵莉, 王成凤, 等. 嗜热链球菌 KLDS3.1012 胞外多糖合成途径的基因组学及表型特征分析[J]. 食品科学, 2019, 40(6): 136–142.  
LI B L, ZHAO L, WANG C F, et al. Genomic and phenotypic characterization of the biosynthesis pathway of exopolysaccharides in *Streptococcus thermophilus* KLDS3.1012 [J]. Food Science, 2019, 40(6): 136–142.
- [17] ZHANG Y, DAI X F, JIN H N, et al. The effect of optimized carbon source on the synthesis and composition of exopolysaccharides produced by *Lactobacillus paracasei* [J]. Journal of Dairy Science, 2021, 104(4): 4023–4032.

### Whole Genome Sequencing and Exopolysaccharide Gene Cluster Analysis of *Tetragenococcus halophilus* CGMCC 3792

ZHANG Min, JIN Yao, HUANG Jun, ZHOU Rongqing, CHI Yuanlong, WU Chongde\*  
(College of Biomass Science and Engineering, Sichuan University, Chengdu 610065)

**Abstract** Objective: *Tetragenococcus halophilus* is a kind of lactic acid bacteria, the high polymer exopolysaccharide produced by it has important application value. In order to explore the molecular mechanism of exopolysaccharide produced by *Tetragenococcus halophilus*, the whole genome of the strain was sequenced and analyzed, which provided a theoretical basis for the production and application of exopolysaccharide from *Tetragenococcus halophilus*. Methods: The whole genome of *Tetragenococcus halophilus* CGMCC 3792 was sequenced by nanopore third-generation sequencing technology, and the gene prediction and functional annotation were performed by bioinformatics method. In addition, the genes related to the biosynthesis of exopolysaccharide were further explored, and the molecular mechanism of exopolysaccharide synthesis by *Tetragenococcus halophilus* CGMCC 3792 was analyzed deeply at the gene level. Results: The total length of the genome of *Tetragenococcus halophilus* CGMCC 3792 was 2 573 722 bp, encoding 2 633 genes. The biosynthesis pathway of exopolysaccharide was analyzed, and results showed that exopolysaccharide could be synthesized from mannose, fructose, maltose, glucose and cellobiose. Through genomic analysis, it was found that the genome contained a 15.4 kb exopolysaccharide gene cluster, including the regulating genes *epsA* and *epsB*, the chain length determining genes *epsC* and *epsD*, and two glycosyltransferase genes as well as the flippase encoding gene *wzx*. Conclusion: This study revealed the molecular mechanism of exopolysaccharide production by *Tetragenococcus halophilus* CGMCC 3792 at the gene level, which laid a foundation for its industrial application.

**Keywords** *Tetragenococcus halophilus*; nanopore sequencing; exopolysaccharide; gene cluster