



利用木质纤维素类生物质产微生物絮凝剂 *Pseudomonas boreopolis* GO2 的全基因组测序及 比较基因组学分析

冯嘉茵^{1,2}, 何继堃^{1,2}, 孙嘉怡^{1,2}, 符学志^{1,2}, 袁佳骛^{1,2}, 郭海朋^{1,2*}

1 宁波大学 农产品质量安全危害因子与风险防控国家重点实验室, 浙江 宁波 315211

2 宁波大学海洋学院, 浙江 宁波 315211

冯嘉茵, 何继堃, 孙嘉怡, 符学志, 袁佳骛, 郭海朋. 利用木质纤维素类生物质产微生物絮凝剂 *Pseudomonas boreopolis* GO2 的全基因组测序及比较基因组学分析[J]. 微生物学报, 2023, 63(2): 786-804.

FENG Jiayin, HE Jikun, SUN Jiayi, FU Xuezhi, YUAN Jiaao, GUO Haipeng. Genome sequencing and comparative genome analysis of *Pseudomonas boreopolis* GO2, a strain producing biofloculant with lignocellulosic biomass[J]. Acta Microbiologica Sinica, 2023, 63(2): 786-804.

摘要:【目的】*Pseudomonas boreopolis* GO2 可以利用木质纤维素类生物质为唯一碳源发酵产微生物絮凝剂。解析菌株 GO2 的全基因组特征可为利用木质纤维素类生物质定向合成多糖型微生物絮凝剂提供分子基础。【方法】利用 Illumina NovaSeq 测序平台对菌株 GO2 进行测序, 用 SMRT 等软件进行基因组组装、系统发育分析、基因预测和功能注释, 并与 4 株近缘模式株进行了比较基因组分析。【结果】菌株 GO2 基因组大小为 4 498 896 bp, GC 含量为 69.5%, 共编码 3 906 个基因。菌株 GO2 与 *Pseudomonas boreopolis* JCM 13306 的 16S rRNA 基因相似性、平均核苷酸一致性 (average nucleotide identity, ANI)、DNA-DNA 杂交 (DNA-DNA hybridization, DDH) 值最高, 分别为 99.93%、98.36% 和 88.00%, 将菌株 GO2 命名为 *Pseudomonas boreopolis* GO2。比较基因组学分析发现, GO2 与 4 个近缘模式菌株共有 2 348 个直系同源核心基因, 主要参与碳水化合物代谢、氨基酸代谢以及能量代谢等过程, 而 GO2 菌株特有的 307 个基因主要与转录、复制、修复、细胞壁/膜/包膜的生物发生相关。菌株 GO2 基因组中含有 226 个碳水化合物活性酶 (carbohydrate-active enzymes, CAZymes) 基因, 占总基因数的 5.79%, 包括 82 个与植物细胞壁降解相关的糖苷水解酶 (glycoside hydrolases, GHs) 家族基因和由糖基转移酶 (glycosyltransferases, GTs) 家族基因形成的 3 个多

资助项目: 国家自然科学基金(21801143); 宁波市公益计划(2019C10077)

This work was supported by the National Natural Science Foundation of China (21801143) and the Ningbo Public Welfare Science and Technology Project (2019C10077).

*Corresponding author. Tel/Fax: +86-574-87600164, E-mail: guohaipeng@nbu.edu.cn

Received: 2022-06-28; Accepted: 2022-09-07; Published online: 2022-09-23

糖合成基因簇。【结论】根据菌株 GO2 基因组序列的比对分析, 将其命名为 *Pseudomonas boreopolis* GO2, 其基因组中包含丰富的植物细胞壁降解酶基因和 3 个多糖合成基因簇, 这些基因可能在菌株 GO2 直接利用木质纤维素类生物质合成微生物絮凝剂过程中发挥着重要作用。

关键词: *Pseudomonas boreopolis* GO2; 微生物絮凝剂; 比较基因组; CAZy 家族基因; 基因簇

Genome sequencing and comparative genome analysis of *Pseudomonas boreopolis* GO2, a strain producing bioflocculant with lignocellulosic biomass

FENG Jiayin^{1,2}, HE Jikun^{1,2}, SUN Jiayi^{1,2}, FU Xuezhi^{1,2}, YUAN Jiaao^{1,2}, GUO Haipeng^{1,2*}

1 State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-Products, Ningbo University, Ningbo 315211, Zhejiang, China

2 School of Marine Science, Ningbo University, Ningbo 315211, Zhejiang, China

Abstract: [Objective] *Pseudomonas boreopolis* GO2 synthesizes bioflocculant with lignocellulosic biomass. In this study, the whole genome sequencing and comparative genome analysis of GO2 were performed to explore its genome characteristics and analyze genes related to lignocellulose degradation and polysaccharide bioflocculant synthesis. **[Methods]** Illumina NovaSeq was used for genome sequencing, and SMRT and other software for genome assembly, phylogenetic analysis, and gene prediction and annotation. Moreover, the genome was respectively compared with that of four similar model strains. **[Results]** The genome of GO2 was 4 498 896 bp, with guanine-cytosine (GC) content of 69.5% and 3 906 coding genes. It showed the highest 16S rRNA gene similarity (99.93%), DNA-DNA hybridization (DDH) (88.00%), and average nucleotide identity (ANI) (98.36%) to *P. boreopolis* JCM 13306. Therefore, this strain was named as *P. boreopolis* GO2. GO2 shared 2 348 core orthologous genes with the four model strains, and these genes were mainly involved in the pathways of carbohydrate metabolism, amino acid metabolism, and energy metabolism. The 307 specific genes of GO2 were mainly involved in the pathways of transcription, replication, repair, and cell wall/membrane biogenesis. GO2 contained 226 carbohydrate-active enzymes (CAZymes) genes, accounting for 5.79% of the total number of genes. Among the CAZymes genes, 82 belonged to the glycoside hydrolases (GHs) family which was related to plant cell wall degradation and 3 clusters of polysaccharide synthesis genes were formed by glycosyltransferases (GTs) family genes. **[Conclusion]** Based on the comparative analysis of genome sequences, GO2 is named *Pseudomonas boreopolis* GO2, and its genome contains rich plant cell wall degrading enzyme genes and three polysaccharide synthesis gene clusters, which may play an important role in the synthesis of bioflocculants by with lignocellulosic biomass in GO2.

Keywords: *Pseudomonas boreopolis* GO2; bioflocculant; comparative genome analysis; CAZy family genes; gene cluster

微生物絮凝剂(microbial bioflocculant, MBF)是一种由微生物分泌产生的具有絮凝活性的天然高分子物质^[1], 主要成分包括多糖、蛋白质以及核酸等天然有机高分子化合物^[2]。与传统絮凝剂相比, MBF 具有絮凝效能高、安全无毒、易降解等优点, 可用于废水处理、饮用水净化和微藻收获等, 被认为是最有潜力替代传统无机和有机高分子絮凝剂的环保型絮凝剂^[3-7]。然而由于其生产成本低、产量低的问题, 限制了 MBF 的工业化应用。碳源是微生物发酵过程中除水之外需求最大的营养物, 其对细菌的生长和代谢具有巨大的影响。目前产絮菌合成 MBF 的最佳底物大多为价格较高的葡萄糖、木糖、甘露糖、半乳糖或蔗糖等单糖和寡糖化合物, 这在一定程度上增加了它的生产成本。

木质纤维素类生物质来源广泛, 价格低廉, 是地球上微生物最丰富的碳源, 由纤维素、半纤维素和木质素组成^[8-9]。在自然情况下, 木质纤维素类生物质中的纤维素和半纤维素不能被大多数微生物直接利用^[10-11], 但某些微生物可以通过分泌包括纤维素酶、半纤维素酶和木质素酶等一系列活性酶将纤维素和半纤维素降解成小分子糖后吸收利用^[12-13], 这些酶类均属于碳水化合物活性酶(carbohydrate-active enzymes, CAZymes)。CAZymes 种类和活性跟细菌基因组中 CAZy 基因特性密切相关, 通常细菌基因组中 CAZy 基因数占总基因数的 1%–3%, 而木质纤维素降解细菌含有丰富的 CAZy 基因, 占比为 3%–5%, 如木质纤维素降解菌 *Bacillus velezensis* LC1^[14], *Pseudomonas stutzeri* MP4687^[15] 和 *Chitinophaga* sp. CB10^[16]。按照功能可以将 CAZymes 分为糖苷水解酶(glycoside hydrolases, GHs)、糖基转移酶(glycosyltransferases, GTs)、多糖裂合酶(polysaccharide lyases, PLs)、碳水化合物酯酶(carbohydrate esterases, CEs)、碳水化

合物结合模块(carbohydrate-binding modules, CBMs)和具有辅助活性的酶(auxiliary activities, AAs)等六大类。其中 GHs 参与木质纤维素类生物质中纤维素和半纤维素的降解, AAs 参与木质素的降解。

通过微生物来降解木质纤维素类生物质, 同时生产 MBF, 将有可能极大地降低它的生产成本^[17]。我国学者 Liu 等^[7-8]发现可以直接转化木质纤维素生物质合成絮凝剂的菌株 *Bacillus agaradhaerens* C9 和 *Cellulosimicrobium cellulans* L804, 并用该絮凝剂收获微藻。本研究前期筛选到一株可以直接利用木质纤维素类生物质作为碳源转化生成多糖型 MBF 的菌株 GO2, 可直接利用玉米秸秆为发酵碳源生产 MBF, 其发酵液的絮凝活力可达 95%, 发酵液中纤维素酶和木聚糖酶活性对其絮凝剂活力的贡献率高达 50%以上^[18]。进一步研究发现, GO2 菌株产的多糖型 MBF 在不同温度和酸碱度条件下仍保留较强的活力, 在微藻回收^[18]和重金属移除^[19]方面表现出了巨大的应用前景。为了进一步揭示 GO2 菌株本身的遗传特性, 深入了解并挖掘其生物学功能和基因资源, 本研究结合基因组测序和生物信息学方法, 分析菌株 GO2 基因组中与木质纤维素降解和多糖 MBF 合成相关的功能基因; 采用比较基因组学分析, 比较菌株 GO2 与近缘模式菌株的基因组异同, 了解 GO2 的基因组特征, 解析 GO2 基因组中生物质降解及多糖型絮凝活性物质的合成和调控基因, 以期为后期利用木质纤维素类生物质为碳源产微生物絮凝剂工程菌的创制及工业化应用提供理论基础。

1 材料与amp;方法

1.1 菌株和培养条件

本实验室前期筛选得到的产絮菌 *Pseudomonas* sp. GO2 为革兰氏阴性杆状细菌, 该

菌株可以直接发酵各种生物质产多糖型 MBF^[18]。该菌株保存于实验室-80 °C超低温冰箱,为获得其发酵液,将该菌株接种于 Luria-Bertani (LB) 液体培养基中,于 30 °C、200 r/min 转速下振荡培养 24 h。取 2 mL 细菌培养液,10 000×g 离心 1 min,收集细菌用于基因组 DNA 提取。

1.2 基因组提取、测序

菌株 GO2 基因组 DNA 的提取按照细菌基因组 DNA 提取试剂盒说明书进行。用 Qubit[®] 2.0 荧光仪 (Invitrogen, Thermo Scientific) 对提取的 DNA 进行定量,用安捷伦生物分析仪 (Agilent, Santa Clara) 检测 DNA 质量。采用 Illumina NovaSeq PE150 测序平台完成 GO2 的全基因组测序 (北京诺禾生物信息技术有限公司)。

1.3 基因组组装与注释

测序获得的原始数据去除低质量序列,使用 SMRT Link v5.1.0 (www.pacb.com/support/software-downloads) 进行基因组组装。将组装好的序列分别于 GO (gene ontology)、KEGG (Kyoto encyclopedia of genes and genomes)、COG (clusters of orthologous groups)、NR (non-redundant protein database)、TCDB (transporter classification database) 和 Swiss-Prot 等 7 个数据库进行基因功能注释。对以上 7 个数据库进行全基因组 Blast (e 值小于 10^{-5} , 最小比对长度百分比大于 40%)。分泌蛋白用 Signal P 数据库注释,CAZymes 用碳水化合物活性酶数据库^[20]注释。

1.4 系统发育及同源分析

利用 EzBioCloud (<https://www.ezbiocloud.net/>) 中 TrueBac[™] ID system 对 GO2 基因组序列进行全基因组序列比对,挑选与 GO2 基因组序列相似度最高的 14 株模式菌株,分别为 *Pseudomonas boreopolis* JCM 13306 (GCA_014654535.1)、*Pseudomonas knackmussii* B13 (GCA_000689415.1)、*Pseudomonas aeruginosa*

DSM 50071 (GCA_001045685.1)、*Pseudomonas furukawaii* KF707 (GCA_002355475.1)、*Pseudomonas stutzeri* ATCC 17588 (GCA_004359495.1)、*Pseudomonas guangdongensis* CCTCCAB (GCA_900105885.1)、*Pseudomonas rhodesiae* LMG 17764 (GCA_013285305.1)、*Pseudomonas balearica* DSM 6083 (GCA_000818015.1)、*Xanthomonas maliensis* M97 (GCA_000401255.1)、*Xanthomonas campestris* ATCC 33913 (GCA_000007145.1)、*Xanthomonas vesicatoria* ATCC 35937 (GCA_001908725.1)、*Xanthomonas hyacinthi* CFBP 1156 (GCA_002939895.1)、*Xanthomonas pisi* DSM 18956 (GCA_001010415.1) 和 *Xanthomonas oryzae* ATCC 35933 (GCA_000482445.1)。下载以上菌株的全基因组序列,提取基因组中的 16S rRNA 基因全长序列,使用 Mega X 软件^[21]基于最大似然法 (1 000 bootstrap) 构建系统发育树。使用 TYGS 平台 (<https://tygs.dsmz.de/>) 基于基因组序列构建全基因组系统发育树^[22],并通过 iTOL (<https://itol.embl.de/>) 对进化树进行可视化与美化。使用 JSpeciesWS (<https://jspecies.ribohost.com/jspeciesws/>) 和 GGDC (<https://ggdc.dsmz.de/>) 平台计算以上各基因组之间的平均核苷酸一致性 (average nucleotide identity, ANI)^[23] 和 DNA-DNA 杂交值 (DNA-DNA hybridization, DDH)^[24]。

1.5 比较基因组学分析

根据菌株 GO2 的系统发育及同源分析,菌株 GO2 与模式菌株 *P. boreopolis* JCM 13306 同源,除了该模式菌株,其与 *Xanthomonas* 属细菌同源性高于 *Pseudomonas* 属的其他细菌,因此,菌株 GO2 及其模式菌株应被划分为 *Xanthomonas* 属。但在本研究中,我们遵循其模式菌株命名仍将 GO2 命名为 *P. boreopolis* GO2。为了进一步分析菌株 GO2 基因组特性,我们从 14 个模式菌株中挑选了 4 个与菌株 GO2

基因组序列相似度最高的菌株进行比较基因组学分析。这 4 个模式菌株分别为 *P. boreopolis* JCM 13306、*X. maliensis* M97、*X. campestris* ATCC 33913 和 *X. vesicatoria* ATCC 35937。全基因组的共线性分析使用 Mauve 软件^[25]。核心基因组和泛基因组分析基于从 Orthofinder^[26]输出的序列中提取的 5 株菌的核心基因及 GO2 的特异性基因，分别用 COG 和 KEGG 数据库进行功能注释。

2 结果与分析

2.1 菌株 GO2 的全基因组测序、组装及注释

使用 Illumina NovaSeq PE150 平台对 GO2

基因组进行测序。组装获得大小为 4 498 896 bp 的基因组，由一个没有质粒的环状染色体组成(图 1)。基因组的平均 GC 含量为 69.5%，包含 3 906 个预测的 CDS，长度为 3 890 595 bp，占全长的 86.48% (表 1)。已将 GO2 的基因组序列提交到 NCBI 的 GenBank 数据库中，登录号为 CP097162。

GO2 的全基因组序列共有 3 906 个蛋白编码基因，占基因组总长的 86.48%。菌株 *P. boreopolis* GO2 共有 2 614 个基因具有 GO 注释功能，占整个序列的 66.92%，其中生物学过程分布的序列最多，占 51.63%，分子功能占 29.12%，细胞组分占 19.23%。生物学过程中的

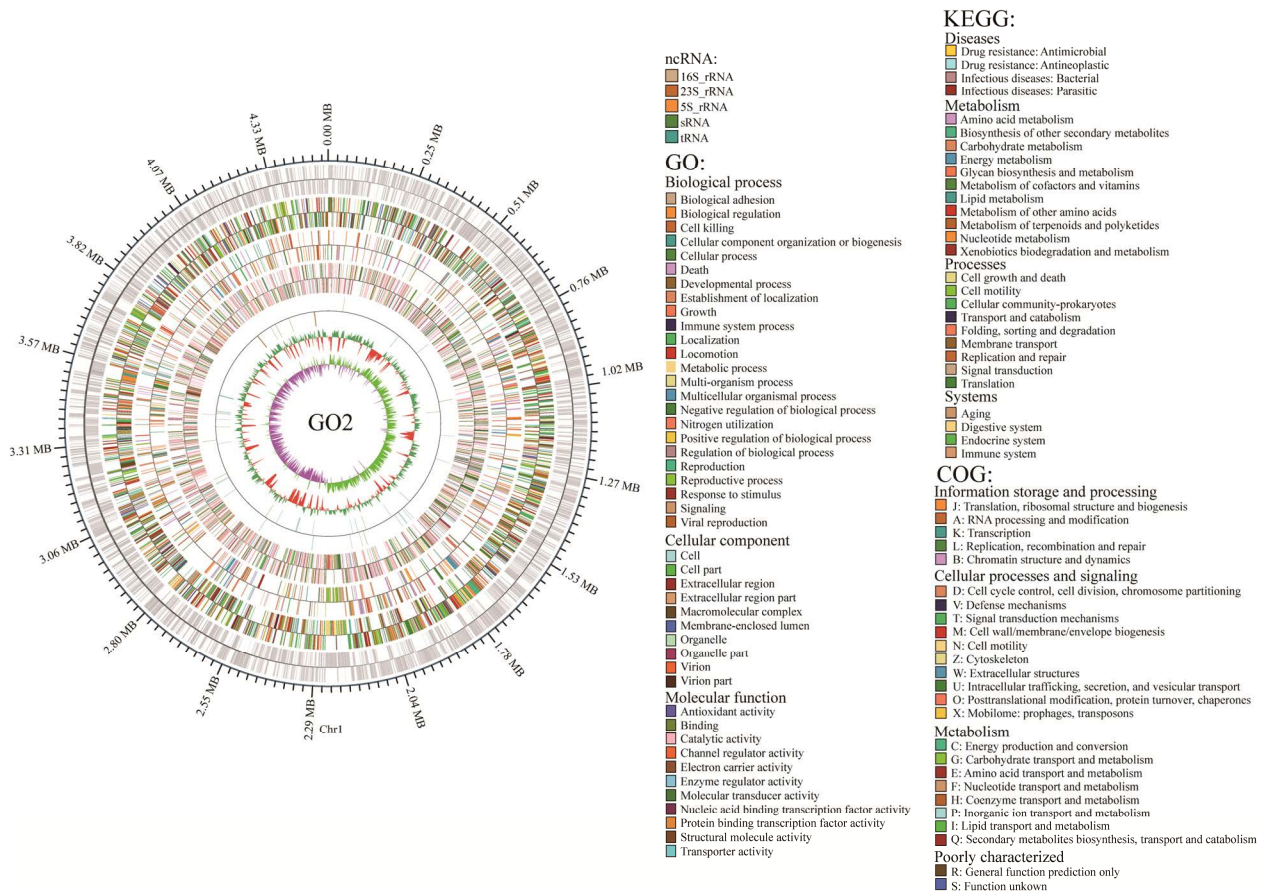


图 1 *Pseudomonas boreopolis* GO2 的表观修饰分布圈图

Figure 1 Circular representation of *Pseudomonas boreopolis* GO2. Based on tRNA and rRNA, GO, KEGG, and COG results, circular diagrams of chromosomes are drawn. The concentric circles show (reading outwards) GC skew, GC content, tRNA and rRNA, GO, KEGG, COG, genes, DNA coordinates.

表 1 *Pseudomonas boreopolis* GO2 基因组基本特征

Table 1 Genome features of *Pseudomonas boreopolis* GO2

Genome features	Value
Genome size (bp)	4 498 896
G+C content (%)	69.5
Gene number	3 906
Gene length (bp)	3 890 595
% of genome	86.48
5S rRNA genes	2
16S rRNA genes	2
23S rRNA genes	2
tRNA genes	55
Other ncRNA genes	12

代谢过程 (metabolic process) 和细胞内过程 (cellular process)、分子功能层次下催化活性 (catalytic activity) 和结合 (binding) 单元在菌株 GO2 基因组中分布的基因较多(图 2)。

GO2 的全基因组中有 3 158 个基因具有 COG 注释功能, 占全基因序列的 80.85%。GO2 基因组具体 COG 功能分类如图 3A 所示。菌株 *P. boreopolis* GO2 基因组中参与一般功能预测 (general function prediction, R) 的蛋白质最多, 共有 324 个, 其次为转录(transcription, K)、氨基酸转运与代谢(amino acid transport and metabolism, E)、细胞壁/膜/包膜的生物发生(cell wall/membrane/

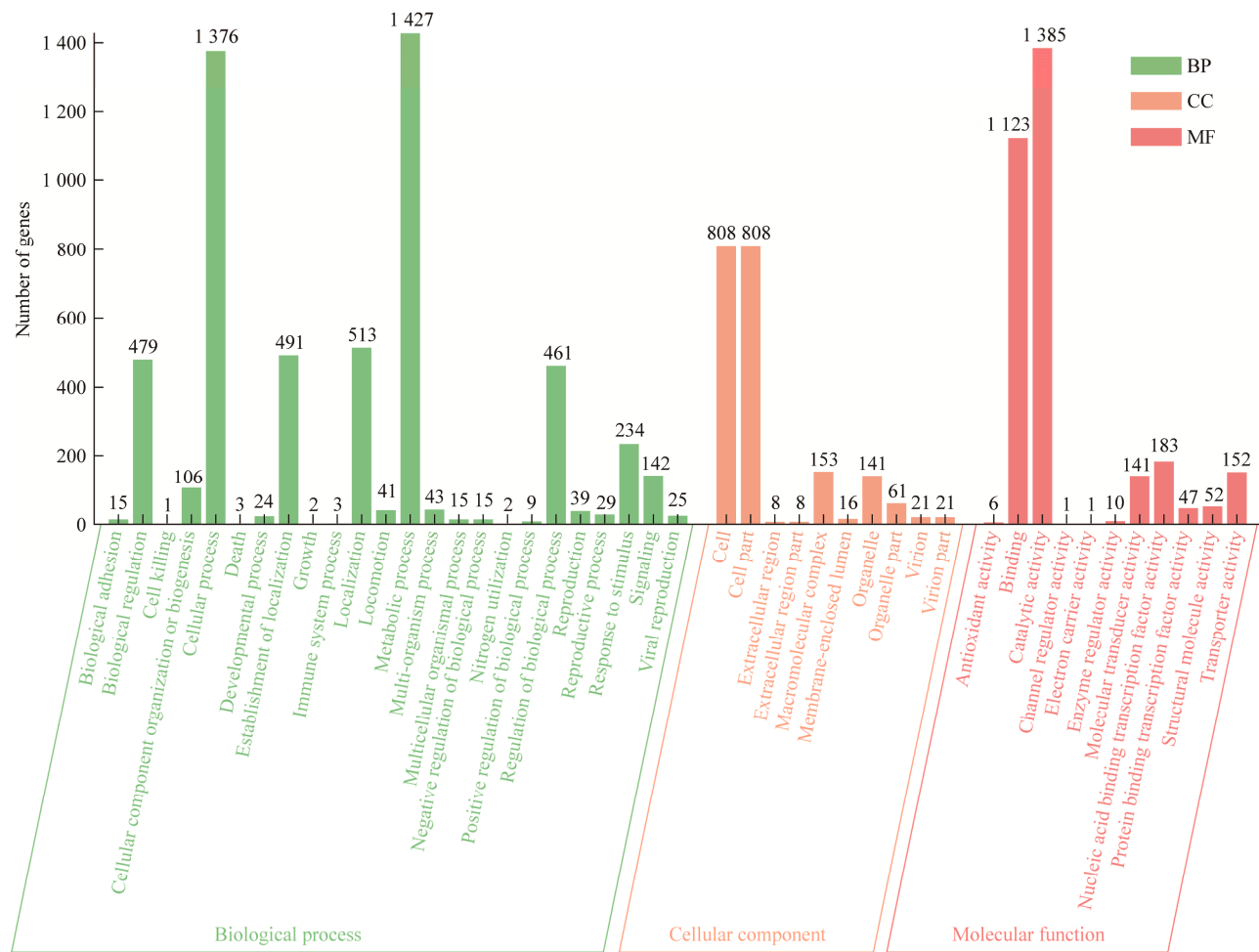


图 2 *Pseudomonas boreopolis* GO2 全基因组的 GO 功能分类

Figure 2 Gene annotation and functional classification in GO database of *Pseudomonas boreopolis* GO2.

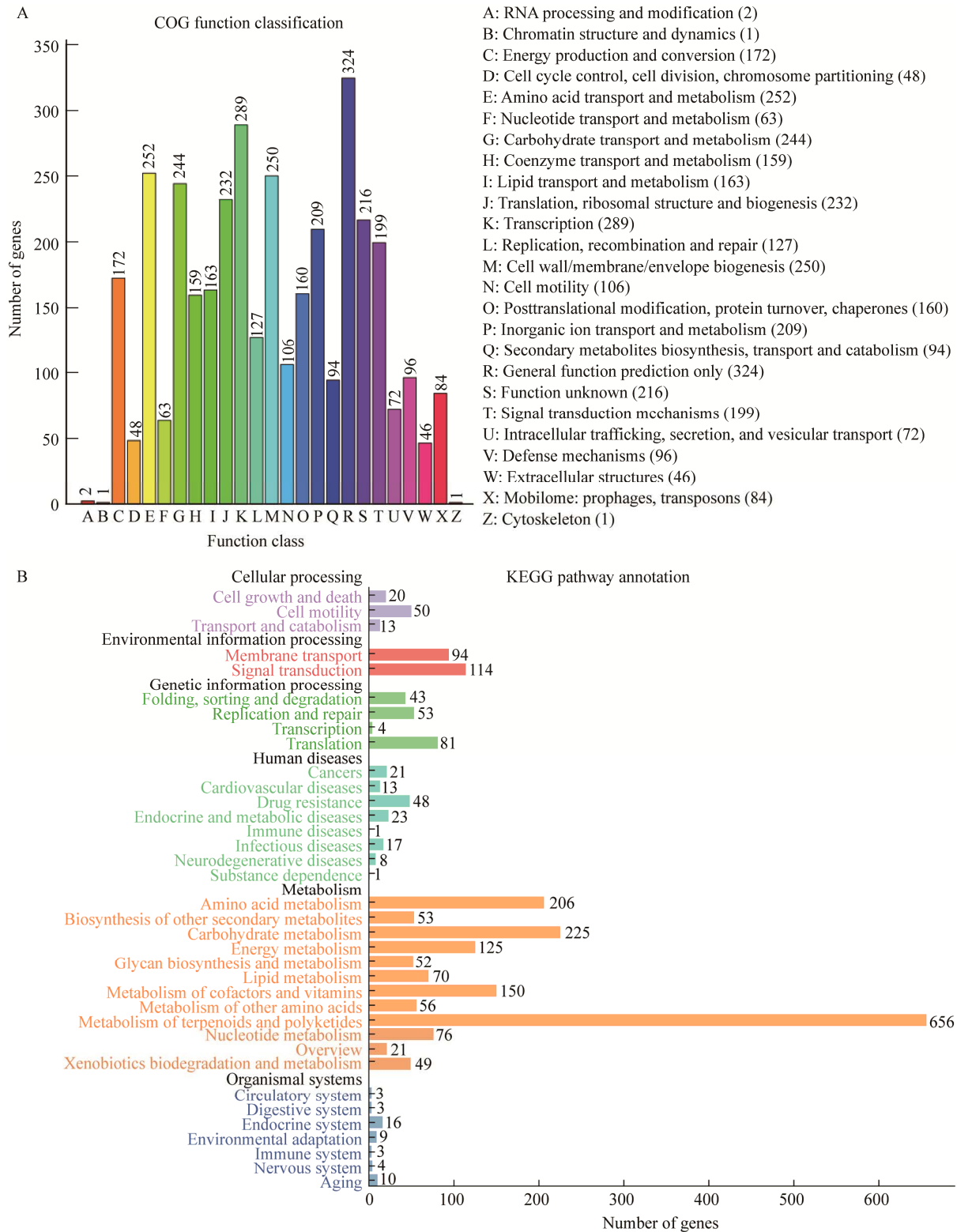


图 3 *Pseudomonas boreopolis* GO2 全基因组的 COG 和 KEGG 功能分类

Figure 3 Gene distribution based on COG (A) and KEGG (B) metabolic pathways of *Pseudomonas boreopolis* GO2.

envelope biogenesis, M)、糖类转运和代谢(carbohydrate transport and metabolism, G)的基因数量分别为 289、252、250 和 244 个。

GO2 基因组 KEGG 分类如图 3B 所示, 代谢通路中属于萜类和聚酮类代谢(metabolism of terpenoids and polyketides)、糖代谢(carbohydrate metabolism)和氨基酸代谢(amino acid metabolism)的通路被注释到较多的基因, 分别为 665、225 和 198 个, 丰富的物质代谢途径使 *P. boreopolis* GO2 可以利用各种复杂的化合物以满足自身代谢需求。在 225 个参与糖代谢的基因中, 有 23 个基因参与半乳糖的代谢, 18 个基因参与果糖和甘露糖的代谢, 30 个基因参与淀粉和蔗糖的代谢。另外, 菌株 GO2 中含有较多参与多糖和粘多糖的生物合成和降解的基因, 这些基因可能与 GO2 能降解木质素物质有关, 说明 GO2 能利用多种多糖作为碳源。

2.2 菌株 GO2 的系统发育树分析

基于 16S rRNA 基因序列比对结果, 菌株 GO2 与模式株 *P. boreopolis* JCM 13306 的 16S rRNA 基因相似度最高为 99.93%, 表明 GO2 与

P. boreopolis JCM 13306 为同一物种。此外, 菌株 GO2 及其模式菌株 *P. boreopolis* JCM 13306 与 *Xanthomonas* 属细菌聚在一支, 而与 *Pseudomonas* 属其他细菌明显分离(图 4)。基于全基因组序列的系统发育树分析也发现菌株 GO2 及其模式株 *P. boreopolis* JCM 13306 与 *Xanthomonas* 属的进化距离较近(图 5), 因此菌株 GO2 及其模式株应属于 *Xanthomonas* 属而不是 *Pseudomonas* 属。

为了验证以上推断, 基于全基因组比对计算 14 个菌株基因组的 ANI 值(图 6)和 DDH 值(表 2)。结果表明, 菌株 GO2 与 *P. boreopolis* JCM 13306 的 ANI 值和 DDH 值高达 98.6%和 88.0%, 进一步表明 GO2 与 *P. boreopolis* JCM 13306 为同一物种, 将其命名为 *P. boreopolis* GO2。不同 *Xanthomonas* 属模式菌株的 ANI 值为 78.1%–87.0%, 而不同 *Pseudomonas* 属模式菌株的 ANI 值为 74.0%–80.8%。本研究中菌株 GO2 和 *P. boreopolis* JCM 13306 与 *Xanthomonas* 属模式菌株的 ANI 值为 78.7%–81.0%, 而与 *Pseudomonas* 属模式菌株的 ANI 值仅为 68.0%–70.0% (图 6), 表明模式株 *P. boreopolis*

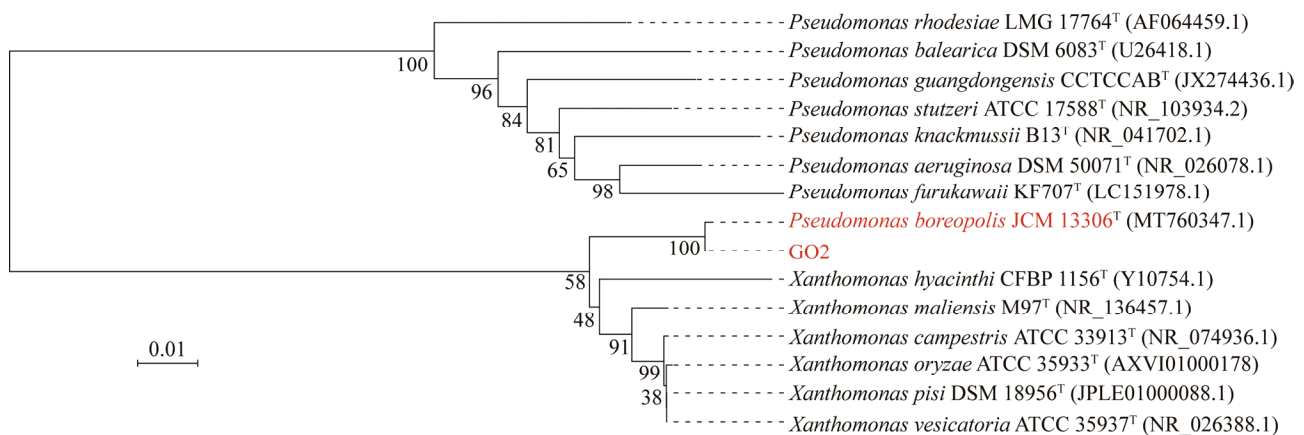


图 4 基于最大似然法构建的 GO2 16S rRNA 基因的系统发育树

Figure 4 The phylogenetic tree of GO2 16S rRNA gene based on the maximum likelihood method. Red font represents GO2 and its type strain, the serial number in parentheses represents the accession number of the sequence, the distance scale represents the variance of the sequence, and the number on the branch represents the bootstrap value.

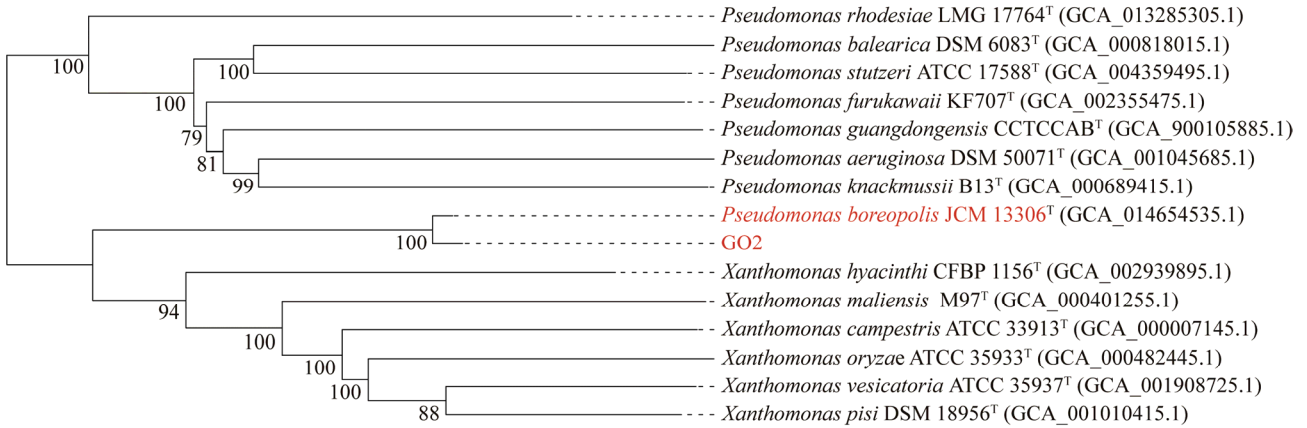


图 5 基于 GO2 全基因组序列的系统发育树

Figure 5 The phylogenetic tree based on the complete sequences of GO2 genome. Red font represents GO2 and its type strain, the serial number in parentheses represents the accession number of the genome, and the number on the branch represents the bootstrap value.

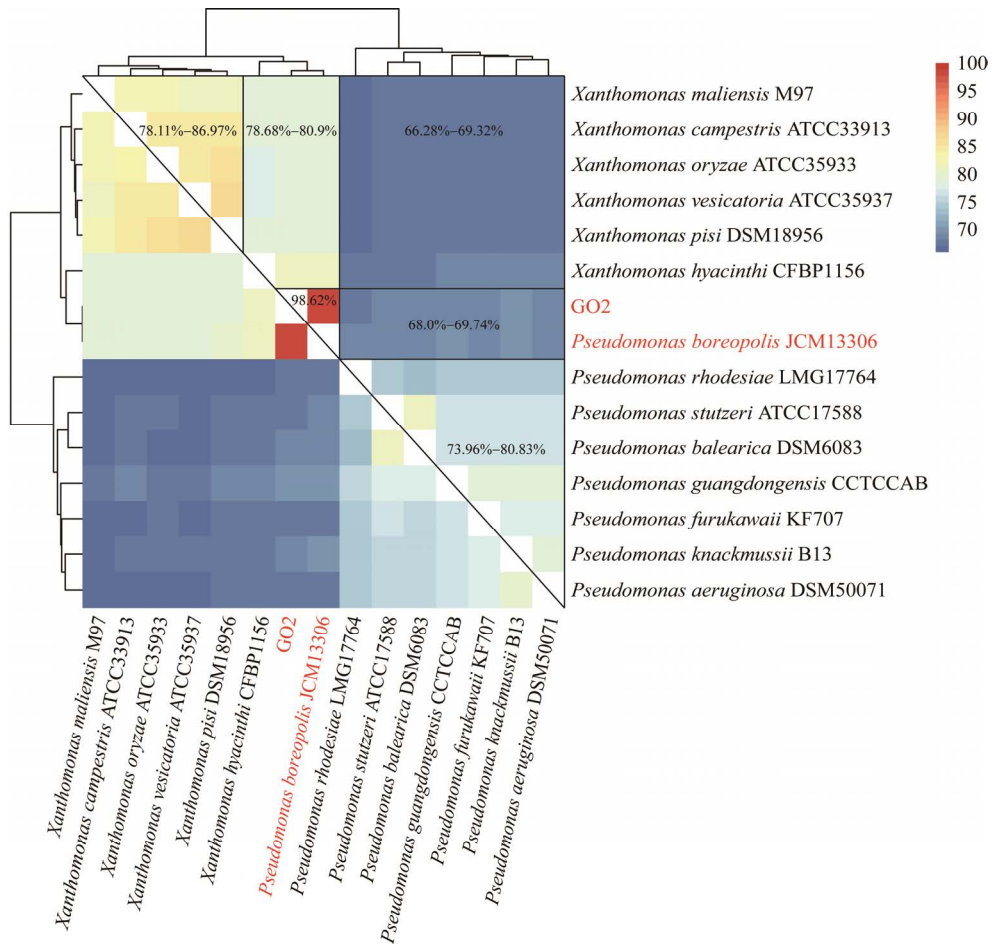


图 6 菌株 GO2 和其他相似模式菌株的 ANI

Figure 6 ANI values between GO2 and other similar type strains.

表 2 菌株 GO2 和其他相似模式菌株的 DDH
Table 2 DDH values between GO2 and other similar type strains

Strain names	<i>Pseudomonas boreopolis</i> GO2 (%)
<i>Pseudomonas boreopolis</i> JCM 13306	88.0
<i>Xanthomonas maliensis</i> M97	21.5
<i>Xanthomonas campestris</i> ATCC 33913	24.1
<i>Xanthomonas oryzae</i> ATCC 35933	21.3
<i>Xanthomonas vesicatoria</i> ATCC 35937	22.0
<i>Xanthomonas pisi</i> DSM 18956	21.5
<i>Xanthomonas hyacinthi</i> CFBP 1156	27.9
<i>Pseudomonas rhodesiae</i> LMG 17764	12.9
<i>Pseudomonas stutzeri</i> ATCC 17588	13.2
<i>Pseudomonas balearica</i> DSM 6083	13.3
<i>Pseudomonas guangdongensis</i> CCTCCAB	13.4
<i>Pseudomonas furukawaii</i> KF707	13.1
<i>Pseudomonas knackmussii</i> B13	13.4
<i>Pseudomonas aeruginosa</i> DSM 50071	13.1

JCM 13306 应当属于 *Xanthomonas* 属。此外, 菌株 GO2 与 *Xanthomonas* 属模式菌株的 DDH 值为 21.3%–27.9%, 而与 *Pseudomonas* 属模式菌株的 DDH 值仅为 12.9%–13.4% (表 2)。以上结果表明, 菌株 GO2 应为 *Xanthomonas* 属细菌而不是 *Pseudomonas* 属。

2.3 比较基因组学分析

2.3.1 全基因组序列比较分析

系统发育分析结果表明, *P. boreopolis* GO2 与同种模式菌株 *P. boreopolis* JCM 13306 的进化关系最近, 且 *P. boreopolis* GO2 与 *Xanthomonas*

属模式菌株的亲缘关系相较于 *Pseudomonas* 属更近, 研究表明 *Xanthomonas* 属的菌株具有产多糖微生物絮凝剂的能力^[27-29]。因此, 选取 GO2 的同种模式菌 *P. boreopolis* JCM 13306 以及亲缘关系最近的 *Xanthomonas* 属模式菌株 *X. maliensis* M97、*X. campestris* ATCC 33913 和 *X. vesicatoria* ATCC 35937 与 *P. boreopolis* GO2 进行全基因组比较分析。

菌株 GO2 与参考菌株的基因组基本特征如表 3 所示, 与菌株 GO2 和 *P. boreopolis* JCM 13306 相比, *X. maliensis* M97、*X. campestris* ATCC 33913 和 *X. vesicatoria* ATCC 35937 的基因组较大, 而 *P. boreopolis* JCM 13306 与 GO2 的 GC 含量等方面十分相近, 均在 69% 左右。通过对 5 个菌株的基因组特征比较, 发现菌株 GO2 和 *P. boreopolis* JCM 13306 在基因组组成、大小和各项特征均比较相似。

2.3.2 基因组共线性分析

P. boreopolis GO2 与 *P. boreopolis* JCM 13306 之间共线性较好(图 7), 但 GO2 与参考菌株基因组之间都存在或多或少的插入、缺失、易位和倒位等基因重排现象, 这可能归结于它们为适应不同生存环境而产生了基因变化。

2.3.3 泛基因和核心基因组分析

对 5 株菌的全基因组进行同源基因聚类分析(图 8A), 获得 4 345 个直系同源基因, 其中核心基因 2 348 个, 占总直系同源基因的 54.04%,

表 3 *Pseudomonas boreopolis* GO2 及近缘 4 株菌的全基因组序列基本特征

Table 3 General genomic features of *Pseudomonas boreopolis* GO2 and other four similar type strains

General features	<i>P. boreopolis</i> GO2	<i>P. boreopolis</i> JCM 13306	<i>X. maliensis</i> M97	<i>X. campestris</i> ATCC 33913	<i>X. vesicatoria</i> ATCC 35937
Genome size (Mb)	4.49	4.33	4.96	5.07	5.30
GC content (%)	69.5	68.9	67.0	65.1	64.1
Gene number	3 906	3 833	4 895	4 294	4 332
rRNA	6	3	1	6	6
tRNA	55	55	23	55	55

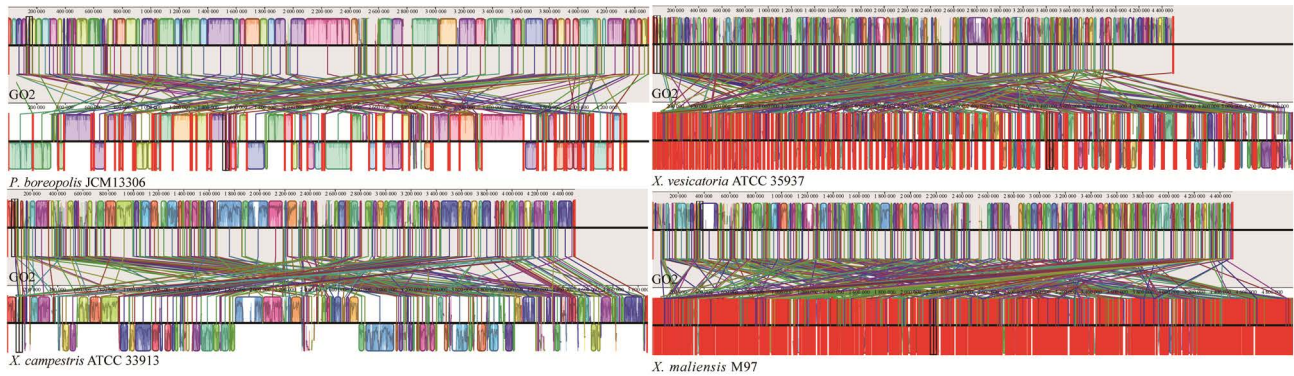


图7 *Pseudomonas boreopolis* GO2与近缘4株模式菌 *P. boreopolis* JCM 13306、*Xanthomonas maliensis* M97、*X. campestris* ATCC 33913 和 *X. vesicatoria* ATCC 35937 的共线性分析

Figure 7 Synteny blocks between *Pseudomonas boreopolis* GO2 and other four similar type strains *P. boreopolis* JCM 13306, *Xanthomonas maliensis* M97, *X. campestris* ATCC 33913 and *X. vesicatoria* ATCC 35937.

另外 1 997 个为附属基因。同时,各菌株存在的特有基因数分别为 231 个(*P. boreopolis* JCM 13306)、884 个(*X. maliensis* M97)、490 个(*X. campestris* ATCC 33913)、816 个(*X. vesicatoria* ATCC 35937)和 307 个(*P. boreopolis* GO2)。

共有的 2 248 个核心基因注释到 COG 的 20 个已知功能分类中(图 8B),其中与代谢相关的比例最高(39.1%)。核心基因主要富集在[E]氨基酸转运和代谢、[J]转录与核糖体结构的生物发生、[M]细胞壁/膜/包膜的生物发生中。GO2 特有基因中主要分布在转录、复制、修复、以及细胞壁/膜/包膜的生物发生(图 8B)。KEGG 分析显示(图 8C),与新陈代谢相关的基因在核心基因组中所占比例最大。在这些基因中,大多数与碳水化合物代谢、氨基酸代谢以及能量代谢有关。

2.4 碳水化合物活性酶比较分析

基因组中的 CAZy 家族基因与 CAZymes 的数量、功能和活性直接相关。菌株 GO2 的基因组中共鉴定出 226 个(占总基因数的 5.79%) CAZy 家族基因(图 9),主要包括 159 个多糖降解相关基因(115 个 GHs、32 个 CBMs、8 个 CEs、

3 个 PLs、1 个 AAs)和 67 个多糖合成相关基因,明显高于近缘菌株 *P. boreopolis* JCM 13306 (189 个, 4.93%)、*X. maliensis* M97 (167 个, 3.41%)、*X. campestris* ATCC 33913 (160 个, 3.73%)和 *X. vesicatoria* ATCC 35937 (169 个, 3.90%)的 CAZy 基因数量。菌株 GO2 基因组中 159 个多糖降解相关基因有 82 个基因参与植物细胞壁的降解,根据其功能的不同,可进一步细分为降解纤维素、半纤维素、果胶和木质素的相关基因(表 4)。包括 GH3、GH5、GH6、GH9、GH12、GH31、GH94 和 GH144 等家族的 17 个基因注释为编码降解纤维素的葡聚糖酶、葡萄糖苷酶以及纤维素二糖磷酸化酶;共有 56 个基因参与半纤维素的降解,包括 28 个编码木聚糖酶的相关基因,15 个编码甘露糖苷酶的相关基因和 4 个 CBM13,13 个降解半乳糖相关基因和 3 个降解阿拉伯糖相关基因;PL1、PL10、GH28、GH95、GH115、CE8 和 CE12 等家族的 10 个基因参与果胶酶的编码;AA6 参与木质素的降解。CBM13 可能增强了 GO2 对木聚糖、葡甘露糖和一些辅酶的亲合能力。除了参与植物细胞壁降解的 82 基因外,其余 77 个多糖降解基因

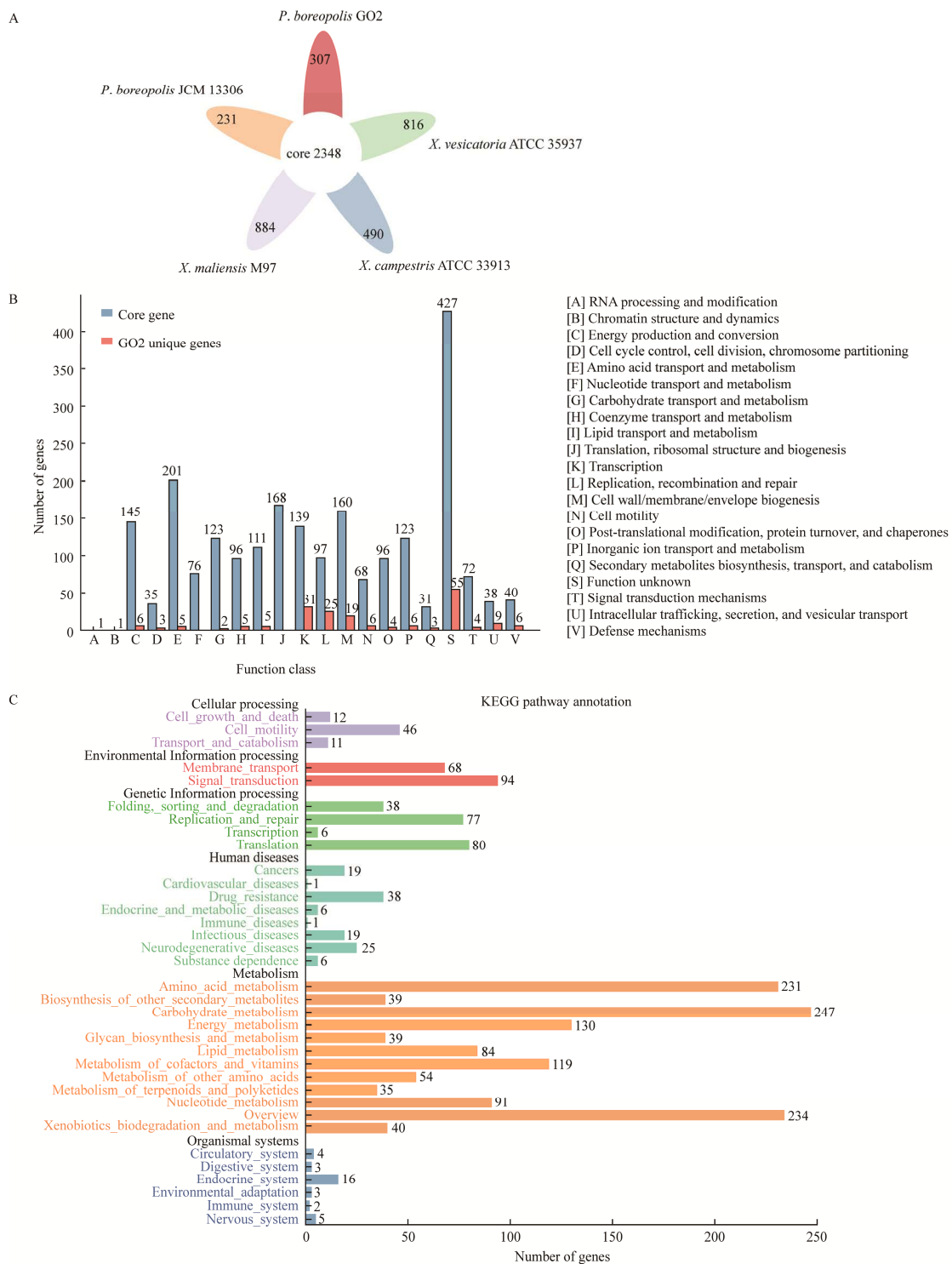


图 8 *Pseudomonas boreopolis* GO2 与近缘 4 株菌的泛基因组分析

Figure 8 Analysis of Pan-genome among *Pseudomonas boreopolis* GO2 and other four similar type strains. A: The numbers of orthologous gene families and unique genes. B: The number of core orthologous gene and GO2-specific 307 genes assigned in COG categories. C: The number of core orthologous gene assigned in KEGG categories.

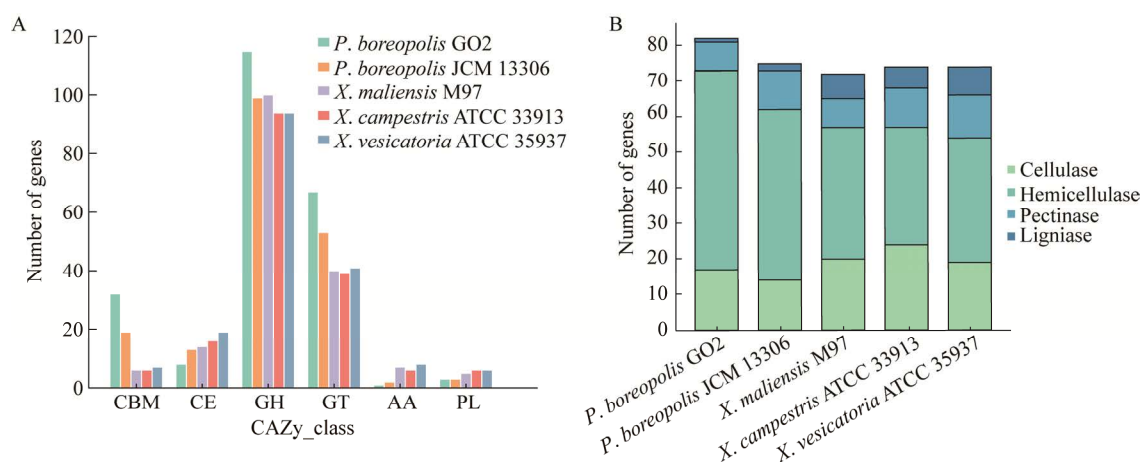


图9 *Pseudomonas boreopolis* GO2 与近缘 4 株菌的碳水化合物酶和植物细胞壁降解酶比较

Figure 9 Comparison of CAZymes and plant cell wall degrading enzymes among *Pseudomonas boreopolis* GO2 and other four similar type strains.

表4 *Pseudomonas boreopolis* GO2 基因组中的植物细胞壁降解酶基因统计

Table 4 Statistics of predicted plant cell wall degradation CAZymes in genome of *Pseudomonas boreopolis* GO2

CAZy family	Known activity	Number	CAZy family	Known activity	Number
Cellulases					
GH1	β -glucosidase	1	GH12	Endoglucanase	1
GH3	Endoglucanase	7	GH31	Glucosidase	1
GH5	Endoglucanase	1	GH94	Cellobiose phosphorylase	1
GH6	Endoglucanase	1	GH144	Endoglucanase	2
GH9	Endoglucanase	1	CBM6	Binding to cellulose	1
Hemicellulases					
Xylose enzymes					
GH10	Endo-1,4- β -xylanase	1	CE1	Acetyl xylan esterase	1
GH11	Endo-1,4- β -xylanase	2	CE4	Acetyl xylan esterase	2
GH43	β -xylosidase	16	CBM13	Binding to xylan	4
GH67	α -glucuronidase	2			
Mannose enzymes					
GH2	β -mannosidase	7	GH92	α -mannosidase	3
GH47	α -mannosidase	1	GH125	Exo- α -1,6-mannosidase	1
Galactose enzymes					
GH27	α -galactosidase	1	GH53	Exo- β -1,4-galactanase	2
GH35	β -galactosidase	3	CBM32	Binding to galactose	3
GH36	α -galactosidase	2	CBM51	Binding to galactose	2
Arabinose enzymes					
GH51	α -L-arabinofuranosidase	1	GH127	α -L-arabinofuranosidase	1
GH121	β -L-arabinofuranosidase	1			
Pectinases					
PL1	Pectate lyase	2	GH115	Xylan- α -1,2-glucuronidase	1
PL10	Pectate lyase	1	CE8	Pectin methylesterase	1
GH28	Polygalacturonase	1	CE12	Pectin methylesterase	2
Ligninases					
AA6	1,4-benzoquinone reductase	1			

主要与淀粉酶、溶菌酶和几丁质酶有关。基因组中丰富的多糖降解酶基因表明菌株 GO2 能够更好地水解生物物质。此外, 菌株 GO2 基因组中共有 61 个 GTs, 这也高于其他近缘 4 株菌, 主要被预测为纤维素合酶、 β -葡萄糖醛酸转移酶和 β -N-乙酰甘露糖醛酸转移酶, 可能参与 UDP-鼠李糖、UDP-葡萄糖和 GDP-甘露糖等核苷酸糖前体的合成。

2.5 多糖型微生物絮凝剂合成基因簇分析

细菌胞外多糖(exopolysaccharide, EPS)的合成是一个多步反应过程, 需要多种特殊功能的蛋白质参与核苷酸糖前体的合成、重复单元的装配和聚合转运, 编码这些蛋白质的相关基因通常排列在一些基因簇。通过对 4 株菌株基因组中多糖合成基因簇进行分析发现, 4 株菌株均有一个黄原胶合成基因簇(*Gum*)和一个鼠李糖合成基因簇(*Rml*), 但是它们在基因组中的位置、排列方向和基因数量不同(图 10)。菌株 GO2 和 *X. campestris* ATCC 33913 的 *Gum* 基因簇分布于基因组中后部分, 分别位于 1 801 589–1 812 993 bp 和 2 899 664–2 917 444 bp 区间, *P. boreopolis* JCM13306 和 *X. vesicatoria* ATCC 35937 分布于基因组的前半部分, 分别位于 55 515–70 940 bp 和 2–12 716 bp 区间。4 株菌株中, GO2 和 *P. boreopolis* JCM13306 的 *Gum* 基因簇高度同源(98.9%–99.87%)且排列方向相同, 与 *X. campestris* ATCC 33913 和 *X. vesicatoria* ATCC 35937 同源性较低(50.30%–84.44%)且排列方向相反。菌株 GO2 的 *Gum* 基因簇包括 *GumD*、*GumE*、*GumF*、*GumH*、*GumI*、*GumJ*、*GumK*、*GumL* 和 *GumM* (GO2_GM001578–GO2_GM001587), 除 *X. vesicatoria* ATCC 35937 不含有 *GumD* 外, 这些基因均存在于其他 2 株菌株的 *Gum* 基因簇中, 但与其他 3 株菌株相比, GO2 的 *Gum* 基因簇缺少 *GumB*、*GumC*、*GumG*、

GumN 和 *GumP*, 其中 *GumG* 和 *GumP* 为 *Xanthomonas* 属的 2 株模式菌株所特有。

RmlABCD 编码蛋白参与鼠李糖前体 L-TDP-鼠李糖的生物合成过程, 本研究中的 4 株菌株均具有 *RmlABCD* 基因, 其基因大小相似且串联排列方式一致(图 10)。菌株 GO2 的 *RmlA* (GO2_GM000771) 编码蛋白与 *P. boreopolis* JCM13306 和 *X. campestris* ATCC 33913 编码的葡萄糖-1-磷酸胸苷转移酶具有较高的同源性, 分别为 98.98%和 89.46%; *RmlB* (GO2_GM000770) 编码蛋白与 *P. boreopolis* JCM 13306 中 dTDP-葡萄糖-4,6-脱水酶具有 99.72%的同源性; *RmlC* 和 *RmlD* (GO2_GM000772 和 GO2_GM000773) 与 *P. boreopolis* JCM13306 中 dTDP-4-脱氢鼠李糖 3,5-表异构酶和 dTDP-鼠李糖合酶的编码基因完全相同(100%)。同时, 菌株 GO2 的鼠李糖合成基因簇中的 CBM5 和 CBM50 为该菌株所特有, 可能对同基因簇上的其他鼠李糖合成基因的功能起辅助作用^[30]。此外, 相比于其他模式菌株, 菌株 GO2 还含有一个长度为 8.64 kb 的多糖合成基因簇, 其基因排列方向倒序, 包括 2 个 GH0 (未分类)、2 个 GT0 (未分类)、2 个 GT4 和 1 个 GT2 基因, 这些基因的大小相似, 约为 1 000 bp。

3 讨论与结论

本研究对可利用木质纤维素产微生物絮凝剂的细菌菌株 GO2 进行了全基因组测序, 其基因组由 1 个大小为 4 498 896 bp 的染色体组成, GC 含量为 69.5%, 3 906 个 CDS 序列翻译后蛋白序列在 GO、COG、KEGG 数据库中分别有 2 614、3 158、3 736 个蛋白序列注释出了功能, 3 个数据库中注释结果具有很高的一致性, 预测的 CDS 蛋白被注释最多的单元与代谢相关。通过 16S rRNA 基因全长序列比对发现, 菌株

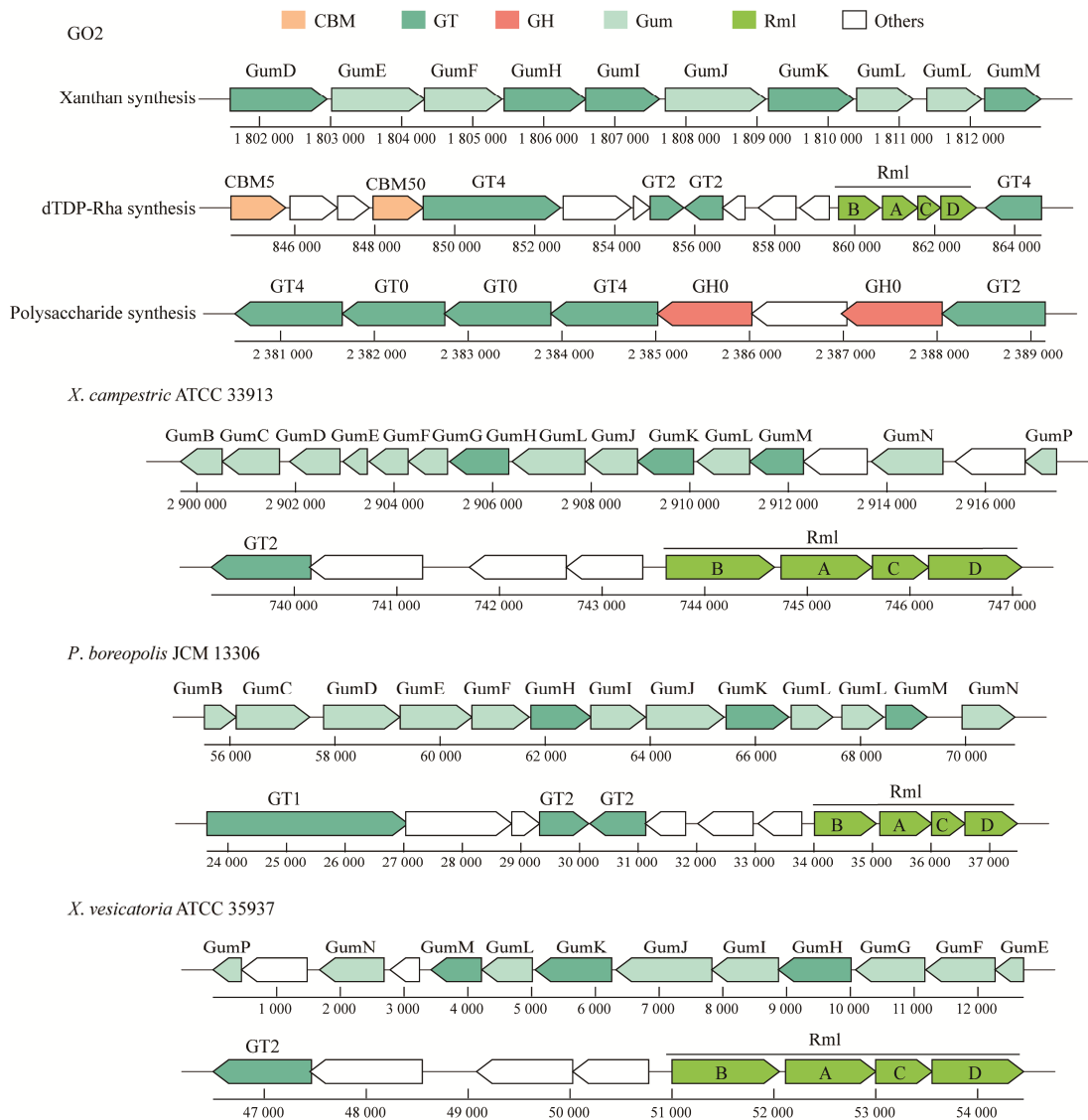


图 10 *Pseudomonas boreopolis* GO2 与近缘模式菌的胞外多糖生物合成基因簇

Figure 10 The exopolysaccharide biosynthesis gene clusters of *Pseudomonas boreopolis* GO2 and other similar type strains.

GO2 与模式菌株 *P. boreopolis* JCM 13306 具有最高的相似度为 99.93%，在 16S rRNA 和全基因组系统发育树中聚类最近。通过 ANI 比对也发现菌株 GO2 与 *P. boreopolis* JCM 13306 具有最高的 ANI 值，为 98.36%，高于同种分类阈值 ANI \geq 95%^[31]；但菌株 GO2 和 *P. boreopolis* JCM 13306 与 *Xanthomonas* 属模式菌的 ANI 为 78.7%–81.0%，介于 *Xanthomonas* 属细菌的 ANI

值之间，高于 *Pseudomonas* 属其他模式菌之间的 ANI 值。研究表明细菌同属分类阈值为 ANI \geq 73.98%^[32]，因此，菌株 GO2 和 *P. boreopolis* JCM 13306 划分为 *Xanthomonas* 属更合理。此外，DDH 比对也发现，菌株 GO2 与 *Xanthomonas* 属模式菌的 DDH 值高于 *Pseudomonas* 属的其他模式菌，进一步表明菌株 GO2 应被划分到 *Xanthomonas* 属。全基因组共线性分析显示菌

株 GO2 与 *P. boreopolis* JCM 13306 的共线性关系较好, 说明菌株 GO2 与模式菌株可能在株水平上有较小差异, 为了保持其模式菌株的命名相同, 我们将菌株 GO2 命名为 *P. boreopolis* GO2^[33]。此外, 菌株 GO2 与 *P. boreopolis* JCM 13306 基因组之间存在倒位和易位等染色体变异事件, 这可能是由于它们的生长环境不同的原因^[34], 如在 2 227 673–2 409 609 bp 处观察到大小为 17 kb 的片段在 *P. boreopolis* GO2 基因组中发生了倒位现象, 该片段含有 *tkl*、*gap* 和 *mod* 等基因涉及到磷酸戊糖途径、糖酵解过程和物质跨膜转运。

碳水化合物活性酶(CAZymes)家族包括糖苷水解酶(GHs)、糖基转移酶(GTs)、多糖裂解酶(PLs)、碳水化合物酯酶(CEs)、辅助活性酶(AAs)和碳水化合物结构域(CBMs), 参与各种寡糖、多糖和糖复合物等碳水化合物的代谢过程。GHs、CEs、PLs 和 AAs 主要与多糖化合物的降解有关, CBMs 可与其他家族的酶结合提高其催化活性, GTs 主要负责二糖、寡糖和多糖前体核苷酸糖的生物合成。微生物降解植物细胞壁的反应过程需要多种碳水化合物酶的参与, 研究表明细菌基因组中 CAZymes 通常占基因含量的 1%–3%^[35], 本研究中, 菌株 GO2 基因组中的 CAZymes 占比为 5.79%, 而其他 4 株近缘菌为 3.41%–4.93%, 表明菌株 GO2 及 4 株近缘菌均具有较好地木质纤维素降解能力。进一步比较碳水化合物活性酶基因的数量和功能发现, 相比于 4 株近缘菌, 菌株 GO2 含有更多的植物细胞壁降解酶的相关基因, 尤其是纤维素和半纤维素降解酶基因, 研究表明秸秆主要成分为纤维素(32%–47%)、半纤维素(19%–27%)和木质素(5%–24%)^[36], 因此, 菌株 GO2 对木质纤维素的强降解力, 可能归因于其较高的纤维素酶和半纤维素酶编码基因数量。GH43 家族

是降解生物质的重要碳水化合物酶家族, 并且在许多植物细胞壁降解微生物的基因组中扩张^[37], 其具有的 β -木糖苷酶、内切 α -L-阿拉伯糖酶和 1,3- β -半乳糖苷酶活性对半纤维素的降解起着关键作用^[38]。本研究中, 菌株 GO2 基因组中含有的 GH43 家族基因(16 个), 明显多于其他 4 株菌, 这说明菌株 GO2 具有较强的半纤维素降解能力。AAs 家族基因与木质素降解相关, 经注释后发现菌株 GO2 基因组中含有 AA6 家族基因, 主要编码 1,4-苯醌类还原酶, 研究表明醌类化合物是木质素降解的重要中间体, 苯醌类氧化还原酶参与醌类化合物的氧化还原循环^[39]。另外研究也发现 *Pseudomonas* 属细菌主要通过分泌 AA6 和 AA1 家族相关酶来参与木质素的降解^[40], 因此该家族基因可能在菌株 GO2 降解木质素过程中发挥关键作用。此外, 菌株 GO2 基因组中还含有较多 α -淀粉酶基因, 表明菌株 GO2 能够较好地利用植物细胞中的淀粉类多糖。

GTs 将核苷酸糖转移到脂质载体上, 负责核苷酸糖前体的装配^[41]。本研究中, 菌株 GO2 及其他 4 株近缘菌株基因组中含有丰富的 GTs 基因, 且与其他 4 株菌相比, 菌株 GO2 含有的 GTs 最多, 如 GO2_GM002823、GO2_GM002824、GO2_GM002825 和 GO2_GM002826 等基因参与多种单糖如葡萄糖、半乳糖、甘露糖和鼠李糖等的核苷酸糖前体的生物合成和转运。尽管细菌胞外多糖的结构各异, 但其生物合成途径相似, 总结不同多糖的合成代谢机制, 可发现胞外多糖生物合成途径的主要分为 4 种: (1) Wzx/Wzy 依赖型途径; (2) ABC 转运体依赖型途径; (3) 合酶依赖型途径; (4) 利用单一蔗糖酶胞外合成^[42]。黄原胶等杂多糖结构具有高度多样化, 通常由 Wzx/Wzy 依赖型途径合成, 且通过该途径合成多糖的菌株多糖合成基因簇中

含有翻转酶(Wzx)和聚合酶(Wzy)基因^[42]。研究报道 *Xanthomonas* 属细菌具有产生黄原胶的能力^[27-29]。本研究中,菌株 GO2 产生的多糖型微生物絮凝剂为杂多糖^[43],其 *Gum* 基因簇中 *GumE* 和 *GumJ* 分别编码 Wzx/Wzy 依赖型途径中的聚合酶(Wzy)和翻转酶(Wzx),因此,推测菌株 GO2 中含有与黄原胶合成类似途径——Wzx/Wzy 依赖型途径。黄原胶生物合成 *Gum* 基因簇中, *GumD* (GT26)编码蛋白可将磷酸葡萄糖转移到脂质载体上, *GumM* (GT26)、*GumH* (GT4)、*GumI* (GT94)和 *GumK* (GT70)编码蛋白参与葡萄糖、甘露糖和葡萄糖醛酸的糖基转移,组成由葡萄糖-甘露糖-葡萄糖醛酸组成的五糖完整重复单元, *GumF* 和 *GumL* 负责甘露糖末端的乙酰化和丙酮化, *GumE* 将重复单元聚合后由 *GumJ* 编码的外排蛋白转移到细胞周质。而与其他近缘模式菌相比,菌株 GO2 的黄原胶合成基因簇缺少黄原胶合成最终步骤中的链长决定簇蛋白(*GumC*)和胞外多糖合成外排蛋白(*GumB*),这可能是造成菌株 GO2 不能合成分泌完整的黄原胶,只能产生黄原胶的单糖组分即葡萄糖、甘露糖和葡萄糖醛酸的原因。鼠李糖合成 *Rml* 基因簇中,磷酸葡萄糖由 *RmlA* 编码的糖基转移酶合成 dDTP-葡萄糖,随后在 *RmlBCD* 编码蛋白的作用下生成 dDTP-鼠李糖,同时,该基因簇中含有 CBM 家族基因,表明菌株 GO2 的鼠李糖合成能力可能强于其他模式菌株。此外,菌株 GO2 基因组中存在一个多糖合成基因簇为该菌株所特有,包含 2 个 GH0、2 个 GT0、2 个 GT4 和 1 个 GT2,该基因簇可能与微生物絮凝剂其他组分如半乳糖、木糖和果糖的生物合成密切相关,进一步表明黄原胶不是菌株 GO2 多糖合成的主要途径,其基因组中可能还存在其他多糖生物合成途径。这与我们前期研究发现菌株 GO2 产生的多糖型微生物絮凝

剂成分包括甘露糖、葡萄糖、葡萄糖醛酸、鼠李糖、半乳糖、木糖以及果糖^[43]的结果一致。因此,这些多糖合成基因簇中的基因及 GTs 可能是影响微生物絮凝剂产量的关键基因。

根据微生物絮凝剂的组成和 GO2 的基因组特性,推测 GO2 以木质纤维素为碳源合成多糖型生物絮凝剂的过程包括以下几步:首先,生物物质的细胞壁成分如纤维素、半纤维素和果胶被 GO2 分泌的 CAZymes 水解为葡萄糖、木糖、半乳糖和甘露糖等单糖,随后单糖通过黄原胶和鼠李糖等胞外多糖生物合成途径合成 UDP-葡萄糖、UDP-葡萄糖醛酸和 UDP-鼠李糖等核苷酸糖前体,并在多种 GTs 的作用下装配到脂质载体上形成重复单元,最后这些重复单元聚合在一起形成微生物絮凝剂被分泌到细胞外。因此,菌株 GO2 基因组中 CAZymes 基因如 GHs 和 GTs 等可能是其利用木质纤维素类生物质产微生物絮凝剂的关键基因,这为今后通过利用基因工程技术构建高产微生物絮凝剂的工程菌株奠定了分子生物学基础。

参考文献

- [1] SALEHIZADEH H, YAN N. Recent advances in extracellular biopolymer flocculants[J]. *Biotechnology Advances*, 2014, 32(8): 1506-1522.
- [2] SHAHADAT M, TENG TT, RAFATULLAH M, SHAIKH ZA, SREEKRISHNAN TR, ALI SW. Bacterial bioflocculants: a review of recent advances and perspectives[J]. *Chemical Engineering Journal*, 2017, 328: 1139-1152.
- [3] AYANGBENRO AS, BABALOLA OO, AREMU OS. Bioflocculant production and heavy metal sorption by metal resistant bacterial isolates from gold mining soil[J]. *Chemosphere*, 2019, 231: 113-120.
- [4] BEN REBAH F, MNIF W, SIDDEEG SM. Microbial flocculants as an alternative to synthetic polymers for wastewater treatment: a review[J]. *Symmetry*, 2018, 10(11): 556.
- [5] KURNIAWAN SB, ABDULLAH SRS, IMRON MF, SAID NSM, ISMAIL N, HASAN HA, OTHMAN AR,

- PURWANTI IF. Challenges and opportunities of biocoagulant/bioflocculant application for drinking water and wastewater treatment and its potential for sludge recovery[J]. International Journal of Environmental Research and Public Health, 2020, 17(24): E9312.
- [6] LI Z, ZHONG S, LEI HY, CHEN RW, YU Q, LI HL. Production of a novel bioflocculant by *Bacillus licheniformis* X14 and its application to low temperature drinking water treatment[J]. Bioresource Technology, 2009, 100(14): 3650-3656.
- [7] LIU W, ZHAO C, JIANG J, LU Q, HAO Y, WANG L, LIU C. Bioflocculant production from untreated corn stover using *Cellulosimicrobium cellulans* L804 isolate and its application to harvesting microalgae[J]. Biotechnology for Biofuels, 2015, 8: 170.
- [8] LIU C, HAO Y, JIANG J, LIU W. Valorization of untreated rice bran towards bioflocculant using a lignocellulose-degrading strain and its use in microalgal biomass harvest[J]. Biotechnology for Biofuels, 2017, 10: 90.
- [9] CHEN SL, SUN S, ZHONG CY, WANG TF, ZHANG Y, ZHOU JG. Bioconversion of lignocellulose and simultaneous production of cellulase, ligninase and bioflocculants by *Alcaligenes faecalis*-X3[J]. Process Biochemistry, 2020, 90: 58-65.
- [10] WANG L, MA F, LEE DJ, WANG AJ, REN NQ. Bioflocculants from hydrolysates of corn stover using isolated strain *Ochrobactium ciceri* W2[J]. Bioresource Technology, 2013, 145: 259-263.
- [11] LIU W, HAO Y, JIANG J, ZHU A, ZHU J, DONG Z. Production of a bioflocculant from *Pseudomonas veronii* L918 using the hydrolyzate of peanut hull and its application in the treatment of ash-flushing wastewater generated from coal fired power plant[J]. Bioresource Technology, 2016, 218: 318-325.
- [12] LOMBARD V, GOLACONDA RAMULU H, DRULA E, COUTINHO PM, HENRISSAT B. The carbohydrate-active enzymes database (CAZy) in 2013[J]. Nucleic Acids Research, 2014, 42(Database issue): D490-D495.
- [13] NGUYEN STC, FREUND HL, KASANJIAN J, BERLEMONT R. Function, distribution, and annotation of characterized cellulases, xylanases, and chitinases from CAZy[J]. Applied Microbiology and Biotechnology, 2018, 102(4): 1629-1637.
- [14] LI YQ, LEI L, ZHENG L, XIAO XM, TANG H, LUO CB. Genome sequencing of gut symbiotic *Bacillus velezensis* LC1 for bioethanol production from bamboo shoots[J]. Biotechnology for Biofuels, 2020, 13: 34.
- [15] PATEL M, PATEL HM, VOHRA N, DAVE S. Complete genome sequencing and comparative genome characterization of the lignocellulosic biomass degrading bacterium *Pseudomonas stutzeri* MP4687 from cattle rumen[J]. Biotechnology Reports, 2020, 28: e00530.
- [16] FUNNICELLI MIG, PINHEIRO DG, GOMES-PEPE ES, de CARVALHO LAL, CAMPANHARO JC, FERNANDES CC, KISHI LT, CARARETO ALVES LM, de MACEDO LEMOS EG. Metagenome-assembled genome of a *Chitinophaga* sp. and its potential in plant biomass degradation, as well of affiliated *Pandoraea* and *Labrys* species[J]. World Journal of Microbiology and Biotechnology, 2021, 37(9): 162.
- [17] SIDDEEG SM, TAHOON MA, REBAH FB. Agro-industrial waste materials and wastewater as growth media for microbial bioflocculants production: a review[J]. Materials Research Express, 2020, 7(1): 12001.
- [18] GUO HP, HONG CT, ZHENG BS, LU F, JIANG DA, QIN WS. Bioflocculants' production in a biomass-degrading bacterium using untreated corn stover as carbon source and use of bioflocculants for microalgae harvest[J]. Biotechnology for Biofuels, 2017, 10: 306.
- [19] FENG JY, XU YJ, DING JH, HE JK, SHEN YH, LU GM, QIN WS, GUO HP. Optimal production of bioflocculant from *Pseudomonas* sp. GO2 and its removal characteristics of heavy metals[J]. Journal of Biotechnology, 2022, 344: 50-56.
- [20] ANTHONY L, ELODIE D, VINCENT L, COUTINHO PEDRO M, BERNARD H. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes[J]. Biotechnology for Biofuels, 2013, 6(1): 41.
- [21] SUDHIR K, GLEN S, MICHAEL L, CHRISTINA K, KOICHIRO T. MEGA X: molecular evolutionary genetics analysis across computing platforms[J]. Molecular Biology and Evolution, 2018, 35(6): 1547-1549.
- [22] MEIER-KOLTHOFF JP, GÖKER M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy[J]. Nature Communications, 2019, 10(1): 1-10.
- [23] IMCHANG L, YEONG OK, SANG-CHEOL P, JONGSIK C. OrthoANI: an improved algorithm and software for calculating average nucleotide identity[J]. International Journal of Systematic and Evolutionary Microbiology, 2016, 66(2): 1100-1103.
- [24] JOHAN G, KONSTANTINIDIS KONSTANTINOS T, KLAPPENBACH JOEL A, TOM C, PETER V, TIEDJE

- JAMES M. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2007, 57(Pt 1): 81-91.
- [25] DARLING AC, MAU B, BLATTNER FR, PERNA NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements[J]. *Genome Research*, 2004, 14(7): 1394-1403.
- [26] EMMS DAVID M, STEVEN K. OrthoFinder: phylogenetic orthology inference for comparative genomics[J]. *Genome Biology*, 2019, 20(1): 238.
- [27] VORHÖLTER FJ, SCHNEIKER S, GOESMANN A, KRAUSE L, BEKEL T, KAISER O, LINKE B, PATSCHKOWSKI T, RÜCKERT C, SCHMID J, SIDHU VK, SIEBER V, TAUCH A, WATT SA, WEISSHAAR B, BECKER A, NIEHAUS K, PÜHLER A. The genome of *Xanthomonas campestris* pv. *campestris* B100 and its use for the reconstruction of metabolic pathways involved in xanthan biosynthesis[J]. *Journal of Biotechnology*, 2008, 134(1/2): 33-45.
- [28] LU GT, MA ZF, HU JR, TANG DJ, HE YQ, FENG JX, TANG JL. A novel locus involved in extracellular polysaccharide production and virulence of *Xanthomonas campestris* pathovar *campestris*[J]. *Microbiology: Reading, England*, 2007, 153(Pt 3): 737-746.
- [29] DEMIRCI AS, PALABIYIK I, ALTAN DD, APAYDIN D, GUMUS T. Yield and rheological properties of exopolysaccharide from a local isolate: *Xanthomonas axonopodis* pv. *vesicatoria*[J]. *Electronic Journal of Biotechnology*, 2017, 30: 18-23.
- [30] BORASTON ALISDAIR B, BOLAM DAVID N, GILBERT HARRY J, DAVIES GIDEON J. Carbohydrate-binding modules: fine-tuning polysaccharide recognition[J]. *The Biochemical Journal*, 2004, 382(Pt 3): 769-781.
- [31] CHIRAG J, RODRIGUEZ-R LUIS M, PHILLIPPY ADAM M, KONSTANTINIDIS KONSTANTINOS T, SRINIVAS A. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries[J]. *Nature Communications*, 2018, 9(1): 5114.
- [32] BARCO RA, GARRITY GM, SCOTT JJ, AMEND JP, NEALSON KH, EMERSON D. A genus definition for bacteria and archaea based on a standard genome relatedness index[J]. *mBio*, 2020, 11(1): e02475-19.
- [33] DOOLITTLE WF, PAPKE RT. Genomics and the bacterial species problem[J]. *Genome Biology*, 2006, 7(9): 116.
- [34] CONOVER DO, DUFFY TA, HICE LA. The covariance between genetic and environmental influences across ecological gradients[J]. *Annals of the New York Academy of Sciences*, 2009, 1168(1): 100-129.
- [35] HUANG L, ZHANG H, WU PZ, ENTWISTLE S, LI XQ, YOHE T, YI HD, YANG ZL, YIN YB. dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation[J]. *Nucleic Acids Research*, 2017, 46(D1): D516-D521.
- [36] ZHAO Y, DAMGAARD A, CHRISTENSEN TH. Bioethanol from corn stover-a review and technical assessment of alternative biotechnologies[J]. *Progress in Energy and Combustion Science*, 2018, 67: 275-291.
- [37] KOHLER A, KUO AL, NAGY LG, MORIN E, BARRY KW, BUSCOT F, CANBÄCK B, CHOI C, CICHOCKI N, CLUM A, COLPAERT J, COPELAND A, COSTA MD, DORÉ J, FLOUDAS D, GAY G, GIRLANDA M, HENRISSAT B, HERRMANN S, HESS J, et al. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists[J]. *Nature Genetics*, 2015, 47(4): 410-415.
- [38] KEITH M, NICOLAS L, VINCENT L, BERNARD H. Dividing the large glycoside hydrolase family 43 into subfamilies: a motivation for detailed enzyme characterization[J]. *Applied and Environmental Microbiology*, 2016, 82(6): 1686-1692.
- [39] SILVA JP, TICONA ARP, HAMANN PRV, QUIRINO BF, NORONHA EF. Deconstruction of lignin: from enzymes to microorganisms[J]. *Molecules: Basel, Switzerland*, 2021, 26(8): 2299.
- [40] XU C, SU X, WANG J, ZHANG F, SHEN G, YUAN Y, YAN L, TANG H, SONG F, WANG W. Characteristics and functional bacteria in a microbial consortium for rice straw lignin-degrading[J]. *Bioresource Technology*, 2021, 331: 125066.
- [41] THIBODEAUX CJ, MELANÇON CE, LIU HW. Natural-product sugar biosynthesis and enzymatic glycodiversification[J]. *Angewandte Chemie-International Edition*, 2008, 47(51): 9814-9859.
- [42] JOCHEN S, VOLKER S, BERND R. Bacterial exopolysaccharides: biosynthesis pathways and engineering strategies[J]. *Frontiers in Microbiology*, 2015, 6: 496.
- [43] GUO HP, FU XZ, CHEN YF, FENG JY, QI ZY, YAN MC, ZHENG BS, QIN WS, SHAO QS. Production of polysaccharide bioflocculants and gene co-expression network analysis in a biomass-degrading bacterium, *Pseudomonas* sp. GO2[J]. *Renewable Energy*, 2022, 188: 997-1007.