



研究报告

副干酪乳杆菌 PC-01 全基因组测序及不同副干酪乳杆菌菌株比较基因组学分析

吕瑞瑞¹ 李伟程¹ 康小红² 张和平^{*1}

1 内蒙古农业大学 乳品生物技术与工程教育部重点实验室 农业农村部奶制品加工重点实验室 内蒙古乳品生物技术与工程重点实验室 内蒙古 呼和浩特 010018

2 内蒙古蒙牛乳业(集团)股份有限公司 内蒙古 呼和浩特 011500

摘要:【背景】副干酪乳杆菌(*Lactobacillus paracasei*)作为乳酸菌中重要菌种之一,常被认为是优良益生菌开发的潜在资源。【目的】以 *L. paracasei* PC-01 和 *L. paracasei* Zhang 为例,分析不同 *L. paracasei* 的基因组差异和遗传背景,为菌株的鉴定和开发奠定基础。【方法】采用 PacBio SMRT 三代测序技术对 *L. paracasei* PC-01 进行全基因组测序,结合 2 株 *L. paracasei* 模式菌株和公开的 36 株全基因组数据,通过比较基因组学方法揭示 39 株 *L. paracasei* 菌株之间的差异。【结果】*L. paracasei* PC-01 基因组不包含质粒,染色体大小为 2 829 251 bp,GC 含量为 46.64%; *L. paracasei* Zhang 包含一个质粒基因组大小为 2 898 456 bp,GC 含量为 46.51%; 不同 *L. paracasei* 菌株基因组大小、质粒数及 GC 含量均存在一定差异。*L. paracasei* 群体为开放式基因组,基因组具有高度多样性。基于核心基因构建系统发育树对于 *L. paracasei* 种内区分效果最好, *L. paracasei* PC-01 和 *L. paracasei* Zhang 处于不同的进化分支。约占 *L. paracasei* PC-01 基因组长度 91% 的片段与 *L. paracasei* Zhang 基因组匹配率大于 90%, *L. paracasei* PC-01 注释到 10 个与转录调节鼠李糖代谢功能相关的特有基因,如 *aguA*、*clpB* 等; *L. paracasei* Zhang 包含 *mshA*、*ltrA* 等特有基因,与糖基转移等功能相关。【结论】通过比较基因组学揭示 *L. paracasei* PC-01 和 *L. paracasei* Zhang 的遗传信息,解析了不同 *L. paracasei* 菌株之间的差异,为 *L. paracasei* 的鉴定和应用奠定了遗传学基础。

关键词: 副干酪乳杆菌, 全基因组测序, 基因组图谱, 比较基因组学分析

Foundation items: Science and Technology Major Projects of Inner Mongolia Autonomous Region (zdx2018018); Inner Mongolia Science and Technology Project (201802063); China Agriculture Research System of MOF and MARA

***Corresponding author:** E-mail: hepingdd@vip.sina.com

Received: 11-11-2020; **Accepted:** 24-12-2020; **Published online:** 07-04-2021

基金项目: 内蒙古自治区科技重大专项(zdx2018018); 内蒙古自治区科技计划(201802063); 财政部和农业农村部: 国家现代农业产业技术体系

***通信作者:** E-mail: hepingdd@vip.sina.com

收稿日期: 2020-11-11; **接受日期:** 2020-12-24; **网络首发日期:** 2021-04-07

Whole genome sequencing of *Lactobacillus paracasei* PC-01 and comparative genomics analysis about *Lactobacillus paracasei* strains

LÜ Ruirui¹ LI Weicheng¹ KANG Xiaohong² ZHANG Heping^{*1}

¹ Inner Mongolia Agricultural University, Key Laboratory of Dairy Biotechnology and Engineering, Ministry of Education; Key Laboratory of Dairy Products Processing, Ministry of Agriculture and Rural Affairs; Inner Mongolia Key Laboratory of Dairy Biotechnology and Engineering, Hohhot, Inner Mongolia 010018, China

² Inner Mongolia Meng Niu Dairy Industry (Group) Company Limited, Hohhot, Inner Mongolia 011500, China

Abstract: [Background] *Lactobacillus paracasei*, as one of the important strains in lactic acid bacteria, is often considered as a potential resource for the development of probiotics. [Objective] *L. paracasei* PC-01 and *L. paracasei* Zhang were taken as examples to analyze the genomic differences and genetic backgrounds of different *L. paracasei* strains, in order to laying a foundation for the identification and development of strains. [Methods] The whole genome sequencing of *L. paracasei* PC-01 was sequenced by PacBio SMRT third-generation sequencing technology, combining with 2 *L. paracasei* model strains and the whole genome data of 36 strains which had published. Through the comparative genomics methods to revealed the differences of 39 *L. paracasei* strains. [Results] The genome of *L. paracasei* PC-01 has one chromosome, with a size of 2 829 251 bp and it's GC content was 46.64%; *L. paracasei* Zhang contained a plasmid and the genome size was 2 898 456 bp and it's GC content was 46.51%. There were some differences of *L. paracasei* strains about genome size, the number of plasmid and GC content. *L. paracasei* is an open genome with a high genome diversity. Phylogenetic tree constructed by core genes has the best effect on species differentiation of *L. paracasei*. What's more, *L. paracasei* PC-01 and *L. paracasei* Zhang were in different evolutionary branches. *L. paracasei* PC-01 and Zhang had a fragment matching rate is 91% of their genomes, and *L. paracasei* PC-01 has 10 specific genes related to transcriptional regulation and rhamnoglycan metabolism, such as *aguA* and *clpB*. *L. paracasei* Zhang contains *mshA*, *ltrA* specific genes were related to glycosyl transfer and other functions. [Conclusion] Through the comparative genomics revealed the genetic information of *L. paracasei* PC-01 and Zhang and found the differences of *L. paracasei* strains, which laid the genetic foundation for the identification and application of *L. paracasei* in the future.

Keywords: *Lactobacillus paracasei*, whole genome sequencing, genome atlas, comparative genomics analysis

随着人们对健康的关注, 益生菌制剂由于其安全性高、促进宿主健康以及增强免疫力等特性引起广泛关注^[1]。副干酪乳杆菌(*Lactobacillus paracasei*)作为乳杆菌属中重要菌种之一, 属于革兰氏阳性菌异型发酵乳酸菌, 广泛分布于肠道、乳制品及青贮饲料等生态位, 其具有抑制肠道有害菌、降血压等益生特性^[2]。*L. paracasei* 能产生黏附素从而黏附到宿主肠道上皮细胞, 从而维持肠道平衡发挥益生功能^[3]。因此, *L. paracasei* 菌株及其代谢产物被广泛应用于医疗、食品及发酵工业等领域, 具有较高的经济价值。

随着高通量测序技术和分析手段的不断发展, 越来越多菌株的基因组完成了测序, 可进一步揭示菌株基因组异同。比较基因组学是在全基因组测序的基础上, 对 2 条及以上的 DNA 链进行分析, 比较物种基因组间的相似性和差异性, 以了解不同物种间基因功能和进化关系^[4]。Sun 等^[5]对乳杆菌属的 213 株菌株及亚种的模式菌株基因组通过比较基因组学讨论了不同乳杆菌模式菌株的群体遗传和干酪乳杆菌分类学问题。Tamara 等^[6]通过对 34 株不同的 *L. paracasei* 基因组测序并进行比较基因组学分析, 研究 *L. paracasei* 进化和生态

位适应之间的关系。*L. paracasei* 作为益生菌资源开发的菌株之一, 对于不同 *L. paracasei* 菌株基因组差异及遗传多样性的研究将成为重点, 通过比较基因组学方法了解其基因组信息及遗传信息, 阐明不同 *L. paracasei* 基因组之间的差异, 对于其在工业发酵、医疗及食品生产的应用至关重要。

本研究团队前期从西藏牦牛发酵乳中分离得到一株 *L. paracasei* PC-01^[7], 实验表明该菌益生特性和安全性评价均良好; 孟和毕力格等从酸马奶中分离得到 240 株乳酸菌^[8], 筛选得到的新菌株 *L. paracasei* Zhang 具有降胆固醇^[9]、免疫调节、抗氧化等益生作用^[10]。本研究对 *L. paracasei* PC-01 全基因组测序, 并结合 NCBI RefSeq 数据库下载的 38 株 *L. paracasei*, 通过比较基因组学分析从基因组水平探究不同菌株之间的遗传异质性, 比较 *L. paracasei* PC-01 和 *L. paracasei* Zhang 的差异, 以期对 *L. paracasei* 的开发和应用奠定遗传学基础。

1 材料与方法

1.1 菌株

L. paracasei 菌株来自内蒙古农业大学乳酸菌菌种资源库(Lactic Acid Bacteria Collection Center, LABCC), 菌株编号为 IMAU60230, *L. paracasei* PC-01 (IMAU60230)是 2007 年分离自西藏拉萨地区酸牦牛奶中的菌株。

截至 8 月 12 日, NCBI (National Center for Biotechnology Information, <https://www.ncbi.nlm.nih.gov/>) RefSeq 数据库包含 38 株 *L. paracasei* 基因组完成图已全部下载。

1.2 主要试剂和仪器

Wizard[®] Genomic DNA Purification Kit, Promega 公司; MRS 培养基, 广州环凯微生物科技有限公司; SMRT 建库试剂盒(DNA Template Prep Kit 1.0)、测序试剂(DNA/Polymerase Binding Kit P6 V2 和 DNA Sequencing Bundle 4.0), Pacific Biosciences 公司。

电热恒温培养箱, 上海一恒科技有限公司;

台式高速离心机, Eppendorf 公司; 超微量紫外分光光度计, NanoDrop 公司; Qubit 2.0 荧光计, Life Technologies 公司; PacBio SMRT RS II 测序平台, Pacific Biosciences 公司。

1.3 菌株培养与基因组 DNA 的提取、测序

菌株 *L. paracasei* PC-01 的 DNA 提取具体操作步骤如下: (1) 将该菌株接种到 5 mL 的 MRS 液体培养基^[11] (62 g/L)中, 37 °C 无氧条件下培养 24 h; (2) 以 2%的接种量到 50 mL 的 MRS 液体培养基中, 在 37 °C 的厌氧工作站扩培 24 h; (3) 扩培至 3 代菌液 3 000×g 离心 10 min 收集菌泥; (4) 将收集的菌泥用 PBS 缓冲液清洗 2 遍, 准备进行全基因组 DNA 的提取。具体的全基因组 DNA 提取参考钟智等^[12]的方法。

提取后的 DNA 具体处理如下: (1) 使用超微量紫外分光光度计测定浓度, 用 0.6%的琼脂糖凝胶电泳检测 DNA 片段的纯度和完整度; (2) 将符合要求的 DNA 保存至 2 mL 离心管中, 加入 0.46×磁珠进行吸附, 直至磁珠全部吸附到磁力架上; (3) 将废液吸出, 用 75%乙醇对吸附有 DNA 的离心管快速洗涤 2 遍, 将废液全部弃掉; (4) 用回溶液将纯化后的 DNA 从磁珠上回溶, 用 Qubit 检测纯化后 DNA 的浓度, 计算 DNA 质量(回溶后 DNA 体积与浓度相乘), 要求 DNA 质量大于 7 μg 且体积小于 150 μL。将符合要求的高质量 DNA 按照 PacBio SMRT 全基因组 DNA 建库流程建立 10 kb 文库, 建库后根据 PacBio SMRT RS II 测序平台上机流程进行全基因组测序。一株菌使用一个 PacBio SMRT Cell, 平均可测得高于 150×的原始数据。

1.4 基因组测序和组装

将测序所得的原始数据进行质量评估, 去掉低质量 Reads 和接头, 通过 SMRT[®] Portal (V2.7) 中 RS_HGAP_Assembly.3 软件对 *L. paracasei* PC-01 序列进行质控和菌株基因组组装。使用 Circlator (V1.5.5)^[13]对组装完成的基因组进行环化。

1.5 比较基因组学分析

1.5.1 平均核酸一致性 (Average Nucleotide Identity, ANI)和总核酸一致性(Total Nucleotide Identity, TNI)值的计算

研究中用自制 Perl 脚本计算 *L. paracasei* 的 ANI 和 TNI 值, 计算方法参考 Goris 等^[14]和 Chen 等^[15]方法。

1.5.2 全基因组圈图和功能基因组分析

将组装完成的 *L. paracasei* PC-01 全基因组序列上传至 RAST 网站 (Repaid Annotation using Subsystem Technology; <http://rast.nmpdr.org/rast.cgi>) 进行注释, 下载其注释文件。注释文件及核酸序列上传 CGView Server (http://stothard.afns.ualberta.ca/cgview_server/) 进行圈图绘制。*L. paracasei* PC-01 基因组基本信息采用自制 Perl 脚本完成。

1.5.3 构建核心基因集、附属基因集

核心基因基于 Prokka^[16]软件对菌株基因组进行基因预测后, 采用 Roary^[17]软件识别核心基因集(Core Gene)和附属基因集(Accessory Gene), 其中以编码蛋白氨基酸相似性大于 95%的原则识别核心基因, 核心基因代表所有菌株共有基因; 仅在 1 个菌株中出现的基因称为特有基因。

1.5.4 系统发育树和共线性分析

为探究 *L. paracasei* 的系统发育关系, 利用 Roary 得到的核心基因, Bao 等^[18]前期使用 10 种看家基因对干酪乳杆菌(*Lactobacillus casei*)进行多位点序列分析, 本文从中选取 *murE* 和 *uvrC* 进行基因串联, 使用 RNAmmer^[19]提取 16S rRNA 基因序列, 使用自制脚本选取 Score 最高值进行合并, 利用邻接(Neighbor-Joining, NJ)法构建 3 种系统发育树。使用 BRIG 软件^[20]对 *L. paracasei* PC-01 和 *L. paracasei* Zhang 的基因组序列进行差异分析。

1.5.5 数据作图

使用 R (V3.5.0)软件中的“pheatmap”程序包绘制热图。

2 结果与分析

2.1 副干酪乳杆菌基因组信息

将样品 *L. paracasei* PC-01 测序数据评估组装, 其基因组序列已提交至 GeneBank 数据库, 序列号为 PRJNA671823。*L. paracasei* PC-01 的基因组圈图如图 1 所示, 并统计 *L. paracasei* 所有菌株的基因组信息(表 1), 将 *L. paracasei* Zhang 作为参考菌株对 *L. paracasei* PC-01 进行分析。*L. paracasei* PC-01 组装后仅为 1 条染色体, 没有质粒, 基因组大小为 2 829 251 bp, GC 含量为 46.64%, 包含 2 756 个蛋白质编码区(Coding Sequence, CDS); *L. paracasei* Zhang 组装结果统计显示 1 条染色体和 1 个质粒, 基因组大小为 2 898 456 bp, GC 含量为 46.45%, 包括 2 705 个 CDS; 不同菌株基因组大小、质粒数、GC 含量等特征均各有差异。张俊等^[2]研究发现, *L. paracasei* 的 GC 含量在 46.2%–46.6%范围, 可编码蛋白 2 800–3 100 个左右, 这与本文统计结果基本一致。

2.2 副干酪乳杆菌 ANI 和 TNI 分析

ANI 常用于细菌基因组的鉴定, 是鉴定细菌基因组种内关系的黄金方法^[21], 通过基因组序列比对判断是否为同一种或者亚种; TNI 是计算核

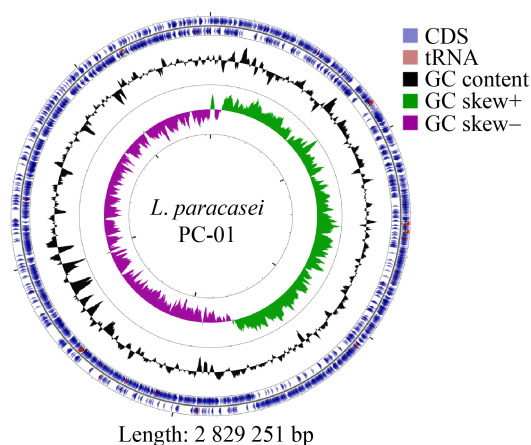


图 1 *L. paracasei* PC-01 全基因组圈图
Figure 1 The whole genome circle-map of *L. paracasei* PC-01

表 1 副干酪乳杆菌基因组基本信息

Table 1 The genome information of *L. paracasei* strains

Strains	Accession No.	Scaffold number	Genome size	Max length	Min length	GC content (%)
<i>L. paracasei</i> PC-01	—	1	2 829 251	2 829 327	2 829 327	46.64
<i>L. paracasei</i> Zhang	GCA 000019245.3	2	2 898 456	2 861 969	36 487	46.43
ATCC25302 ^T	GCA 000159495.1	125	3 003 415	290 533	825	46.48
DSM20258 ^T	GCA 001436485.1	389	2 383 240	43 276	202	46.42
ATCC 334	GCA 000014525.1	2	2 924 325	2 895 264	29 061	46.58
BL23	GCA 000026485.1	1	3 079 196	3 079 196	3 079 196	46.34
362.5013889	GCA 000155515.2	3	3 025 352	2 939 026	24 207	46.30
BD-II	GCA 000194765.1	2	3 127 288	3 069 926	57 362	46.29
LC2W	GCA 000194785.1	2	3 077 434	3 039 042	38 392	46.31
W56	GCA 000318035.1	2	3 132 096	3 075 780	56 316	46.29
LOCK919	GCA 000418515.1	2	3 143 369	3 113 601	29 768	46.22
N1115	GCA 000582665.1	5	3 064 279	2 938 059	3 833	46.45
JCM 8130	GCA 000829035.1	3	3 017 804	2 995 875	10 775	46.55
CAUH35	GCA 001191565.1	5	2 973 347	2 770 411	31 497	46.31
L9	GCA 001244395.1	1	3 076 437	3 076 437	3 076 437	46.35
KL1	GCA 001514415.1	1	2 918 888	2 918 888	2 918 888	46.60
IIA	GCA 002079285.1	4	3 246 143	3 055 892	45 033	46.25
TK1501	GCA 002257625.1	1	2 942 538	2 942 538	2 942 538	46.54
FAM18149	GCA 002442835.1	6	2 969 707	2 710 941	28 557	46.31
TMW 1.1434	GCA 002813615.1	2	3 170 114	3 063 698	106 416	46.37
HD1.7	GCA 002865565.1	1	3 039 280	3 039 280	3 039 280	46.42
HDS-01	GCA 002902825.1	1	3 038 767	3 038 767	3 038 767	46.42
EG9	GCA 003177075.1	4	3 074 406	2 927 257	12 035	46.43
Lpc10	GCA 003199005.1	1	3 052 122	3 052 122	3 052 122	46.31
LC355	GCA 003268715.1	3	3 143 531	3 044 973	38 999	46.38
ZFM54	GCA 003627255.1	3	3 048 677	3 015 887	13 282	46.34
Feb-12	GCA 003957435.1	2	3 096 558	3 022 095	74 463	46.32
IJH-SONE68	GCA 003966835.1	3	3 181 204	3 084 917	45 267	46.39
SRCM103299	GCA 004141835.1	3	3 187 447	3 081 420	38 400	46.38
CBA3611	GCA 007292115.1	2	3 102 530	3 035 735	66 795	46.30
NJ	GCA 007637635.1	1	3 083 412	3 083 412	3 083 412	46.40
10266	GCA 008329845.1	2	3 044 908	3 012 260	32 648	46.33
TCS	GCA 008807095.1	2	3 144 483	3 077 701	66 782	46.28
IBB3423	GCA 009739485.1	3	3 240 583	3 183 386	5 986	46.30
TD 062	GCA 009834405.1	7	2 867 519	2 667 255	8 093	46.43
CACC 566	GCA 009931715.1	3	3 240 202	3 123 521	51 921	46.14
347-16	GCA 012955485.1	3	3 219 033	3 102 350	51 922	46.14
Lp02	GCA 013307125.1	3	3 092 262	2 991 176	47 041	46.33
MGYG-HGUT-02388	GCA 902386635.1	1	3 076 437	3 076 437	3 076 437	46.35

注: ^T: 模式菌株; —: 无登录号Note: ^T: The model strain; —: No accession No.

酸之间匹配程度,二者均可用于菌株的区分。基于 38 株 *L. paracasei* 全基因组完成图和 *L. paracasei* PC-01 进行 ANI (图 2A) 和 TNI (图 2B) 分析。结果显示,39 株 *L. paracasei* 的 ANI 值全部大于 98.09%, TNI 值全部大于 77.28%; *L. paracasei* PC-01 和 *L. paracasei* Zhang 之间 ANI 值为 98.37%, TNI 值为 80.74%。

2.3 核心基因和泛基因组

结果显示(图 3), *L. paracasei* 群体为开放式基因组,随着菌株基因组数量不断增加,泛基因个数呈现增加的趋势,而核心基因数量趋于稳定状态。李晓姝等^[22]发现 *L. paracasei* 的基因组呈开放

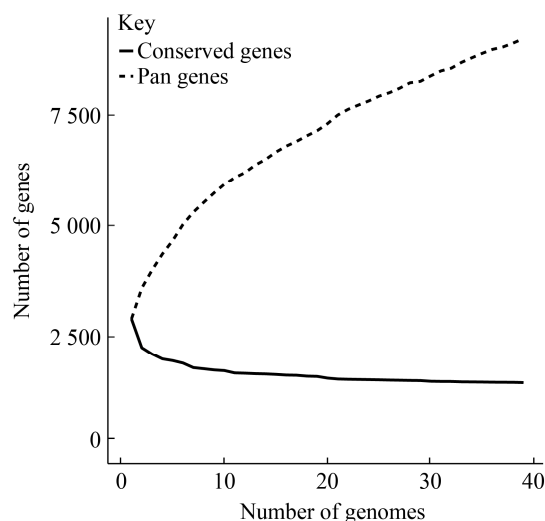


图 3 核心基因和泛基因组分析

Figure 3 The analyze of core-pan genome

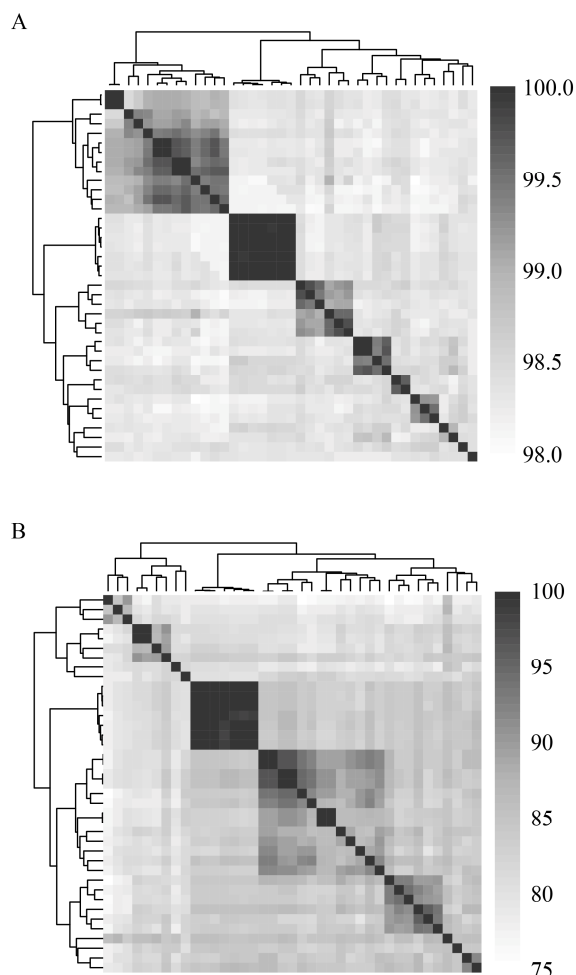


图 2 平均核酸一致性和总核酸一致性

Figure 2 Average nucleotide identity and total nucleotide identity

型,导致这一结果的原因可能是由于 *L. paracasei* 分布广泛,可与外界各种遗传物质发生交换。

为比较 *L. paracasei* 不同菌株之间的差异,通过 Prokka 和 Roary 软件进行分析注释,以氨基酸一致性大于 95% 为标准,统计核心基因和泛基因的数量。39 株 *L. paracasei* 的泛基因组包括了 9 226 个基因,其中核心基因 1 458 个,特有基因 3 003 个。王彦杰^[23]对 78 株粪肠球菌进行比较基因组分析,发现其核心基因数量占有所有基因的 47.2%。然而,本研究中 *L. paracasei* 核心基因占泛基因数量的 15.8%,表明大部分 *L. paracasei* 是易变的,其基因组多样性较高。

2.4 基于核心基因、16S rRNA 基因、看家基因的系统发育树

系统发育树可将同一群体内不同个体的遗传距离与进化关系可视化,是研究群体结构与物种进化的前提^[24]。本研究将 *L. paracasei* subsp. *tolerans* DSM20258^T 作为外群,以 16S rRNA、*murE-uvrC* 串联基因和核心基因序列通过邻接法构建系统发育树,比较不同 *L. paracasei* 在不同方法下的区分情况,以了解 *L. paracasei* PC-01 和 *L. paracasei* Zhang 之间的进化关系。

2.4.1 基于 16S rRNA 基因序列构建的系统发育树

细菌 16S rRNA 基因序列被称为细菌鉴定的黄金标准, 在细菌学和分类学方面常被用于细菌的检测和鉴定^[25]。本研究基于 39 株 *L. paracasei* 的 16S rRNA 基因序列构建系统发育树(图 4), 结果显示, 各 *L. paracasei* 的 16S rRNA 基因序列差异较小, *L. paracasei* 的各菌株并未明显区分开。前期研究表明 16S rRNA 基因因其序列多拷贝及高度保守性等特点, 导致该方法在区分一些近缘种或者亚种中存在一定缺陷^[26]。

2.4.2 基于 *murE-uvrC* 串联基因序列构建的系统发育树

基于 *murE-uvrC* 串联基因序列构建的系统发育树显示(图 5), *L. paracasei* Zhang 与 *L. paracasei* subsp. *paracasei* ATCC25302^T 聚为一簇, *L. paracasei* PC-01 亲缘距离稍远, 总体而言区分效果并不明显。这可能是由于干酪乳杆菌及其近缘物种之间关系复杂且亲缘关系较近, 虽比 16S rRNA 基因序列区分效果明显, 但仍不理想。

2.4.3 基于核心基因序列构建的系统发育树

基于 1 458 个核心基因序列构建系统发育树(图 6), 结果显示 39 株 *L. paracasei* 主要分为 3 个分支, 集中在一个分支表明其亲缘关系更近。*L. paracasei* subsp. *tolerans* DSM20258^T 为代表的分支由 4 株菌组成; *L. paracasei* PC-01 与 *L. paracasei* Zhang 从核心基因序列构建的系统发育树可知二者存在一定差异, *L. paracasei* PC-01 与 ATCC25302^T 遗传距离更近, 且各个菌株区分更明显。

通过不同方法构建系统发育树发现, 基于不同 *L. paracasei* 菌株的核心基因构建的系统发育树对菌株的区分和鉴定效果最好, 干酪乳杆菌及其近缘菌种间的 16S rRNA 基因序列相似度大于 99%^[27], 也解释了本研究结果中 *L. paracasei* 菌株的 16S rRNA 基因序列区分能力最差的原因。因此不同菌株核心基因构建的系统发育树可作为更准确快速地区分 *L. paracasei* 种及亚种的可行方法,

为益生菌种的筛选和鉴定奠定一定的理论基础。

2.5 基于 BRIG 对副干酪乳杆菌 PC-01 和 Zhang 的差异分析

不同菌株基因组的可视化有利于确定密切相关菌株基因型差异。Blast Ring Image Generator (BRIG) 可生成显示多个原核生物基因组比较的图像, 是可视化基因组比较数据的通用方法^[20]。本文以 *L. paracasei* Zhang 染色体为中心参考序列, 比较 *L. paracasei* PC-01 与其基因组的一致性和差异。结果显示(图 7), 二者基因组之间有一定差异, 匹配度最高为 100%, 最低为 61.30%; 约有 2 567 583 bp 基因组片段匹配度大于 90%, 占总片段长度的 91%; 分析发现 11 631 bp 基因片段匹配度小于 70%, 约占总片段的 0.4%。

2.6 *L. paracasei* PC-01 和 *L. paracasei* Zhang 的基因比较

将 Roary 注释结果进行统计, 发现 *L. paracasei* PC-01 特有基因 169 个, 其中注释到功能的有 10 个基因, 分别为 *aguA*、*araE*、*chbR*、*clpB*、*pglH*、*rhaA*、*rhaB*、*rhaD*、*rhaM*、*yicI*, 与转录调节、鼠李糖代谢相关; *L. paracasei* Zhang 包含 51 个特有基因, 其中注释到功能的有 4 个, 分别为 *epsE*、*yybR*、*mshA*、*ltrA*, 与糖基转移、转录调节等代谢功能相关。*L. paracasei* PC-01 基因注释到的代谢通路多集中于糖代谢, 这与 *L. paracasei* 和鼠李糖乳杆菌在系统发育和表型上相似, 以及具有相同的肽聚糖类型这一研究结果相符合^[2,28]。

乳酸菌对胃肠道的耐酸、耐胆盐特性及黏附能力是其定殖人体发挥益生功效的基础条件。人体胆盐浓度由于其高渗透压会对菌体细胞产生一定伤害^[29]。部分乳酸菌可编码胆盐水解酶(Bile Salt Hydrolase, BSH)降低胆盐对自身的损伤。Zhang 等^[30]分析 *L. paracasei* Zhang 全基因组发现 BSH 的编码基因; *L. paracasei* PC-01 的胆盐耐受性较好, 通过全基因组分析在该菌基因组注释到 BSH 同源基因, 其良好的耐胆盐特

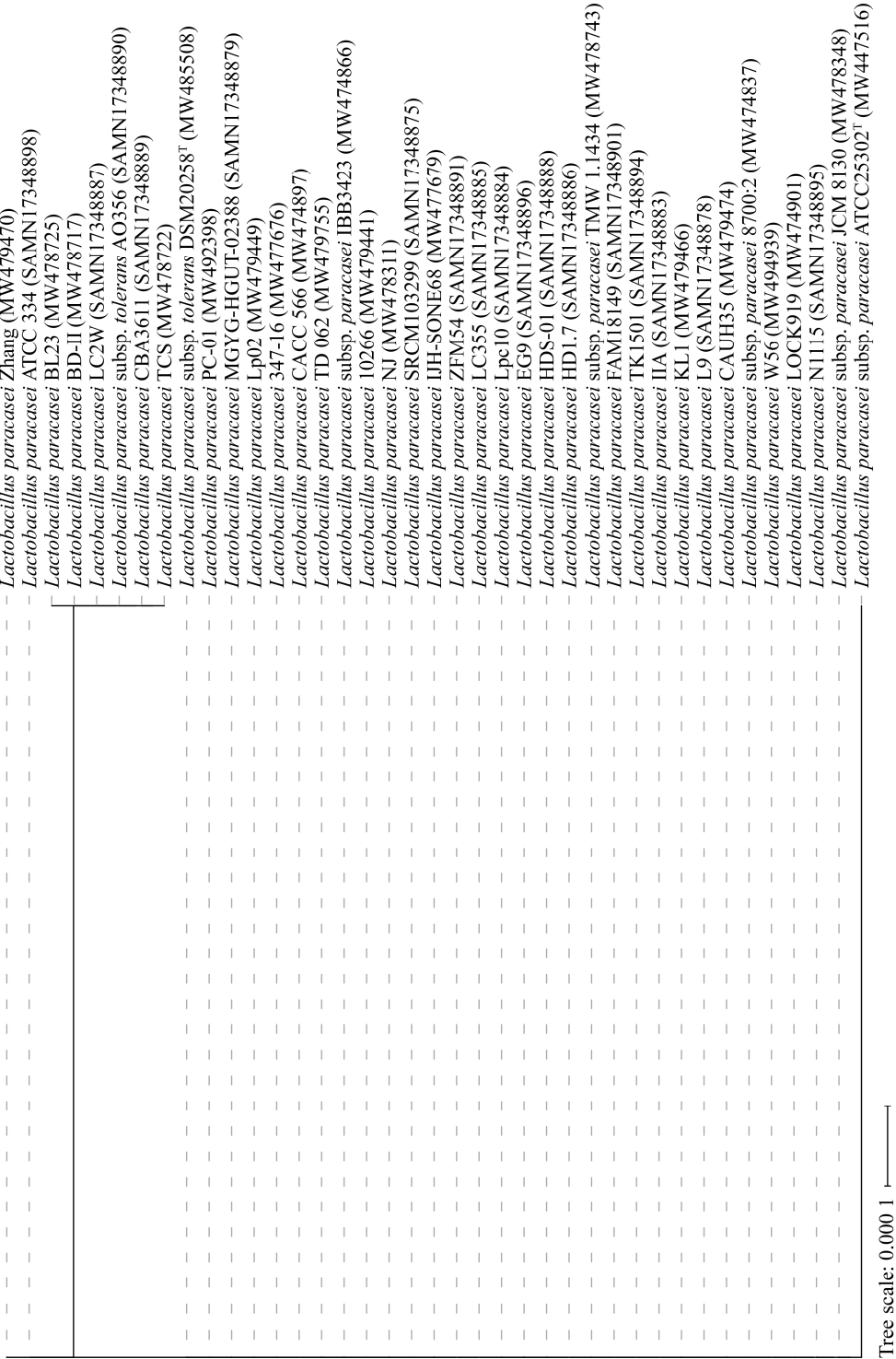


图 4 基于 16S rRNA 基因序列构建的系统发育树
Figure 4 Phylogenetic tree based on 16S rRNA gene sequences
注: 比例尺显示水平线的长度, 代表碱基替换数

Note: The scale bar represents the horizontal branch lengths and the number of substitutions per nucleotide position

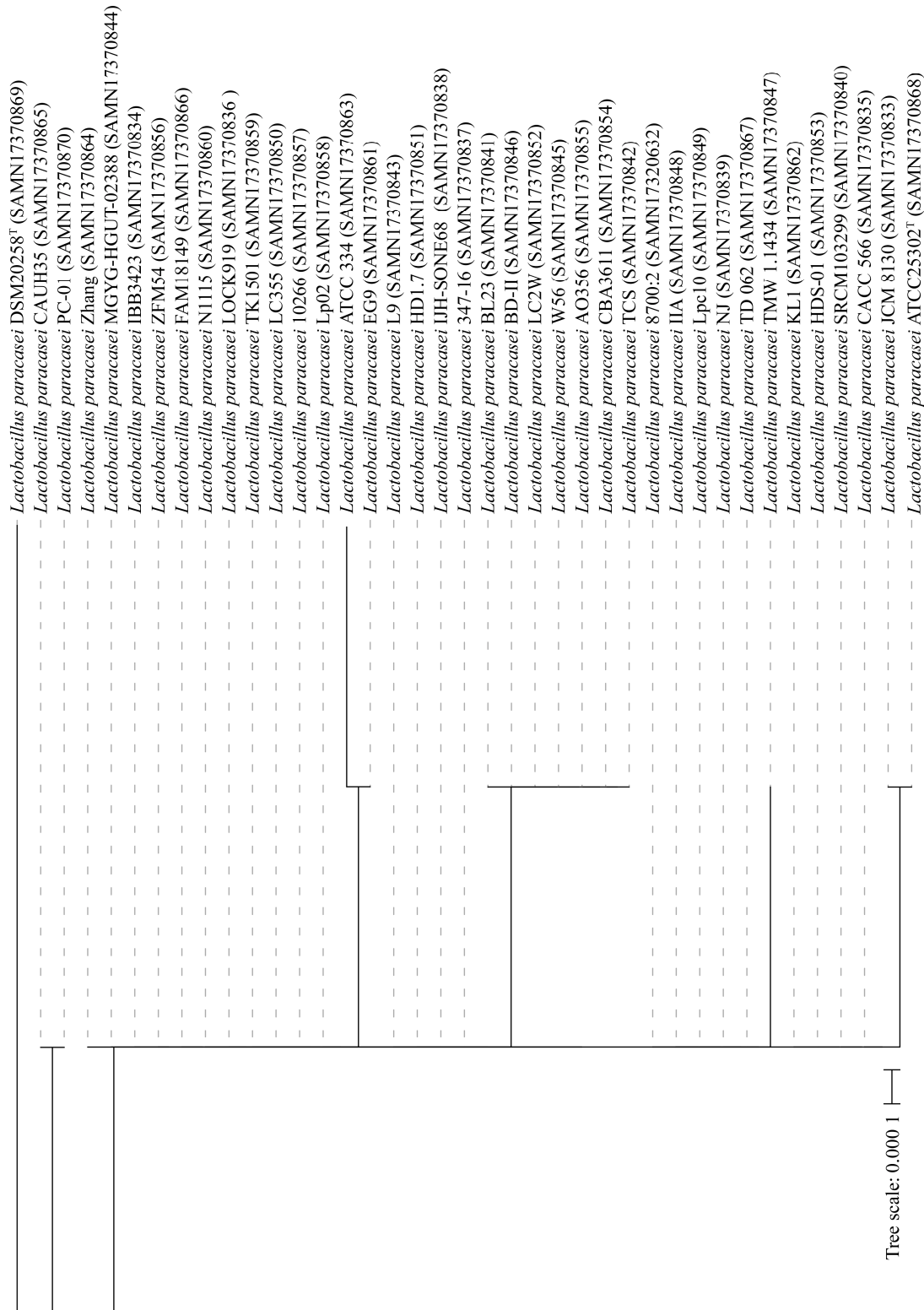


图 5 基于 *murE-uvrC* 串联序列构建的系统发育树
Figure 5 Phylogenetic tree based on *murE-uvrC* gene sequences

注: 比例尺显示水平线的长度, 代表碱基替换数

Note: The scale bar represents the horizontal branch lengths and the number of substitutions per nucleotide position

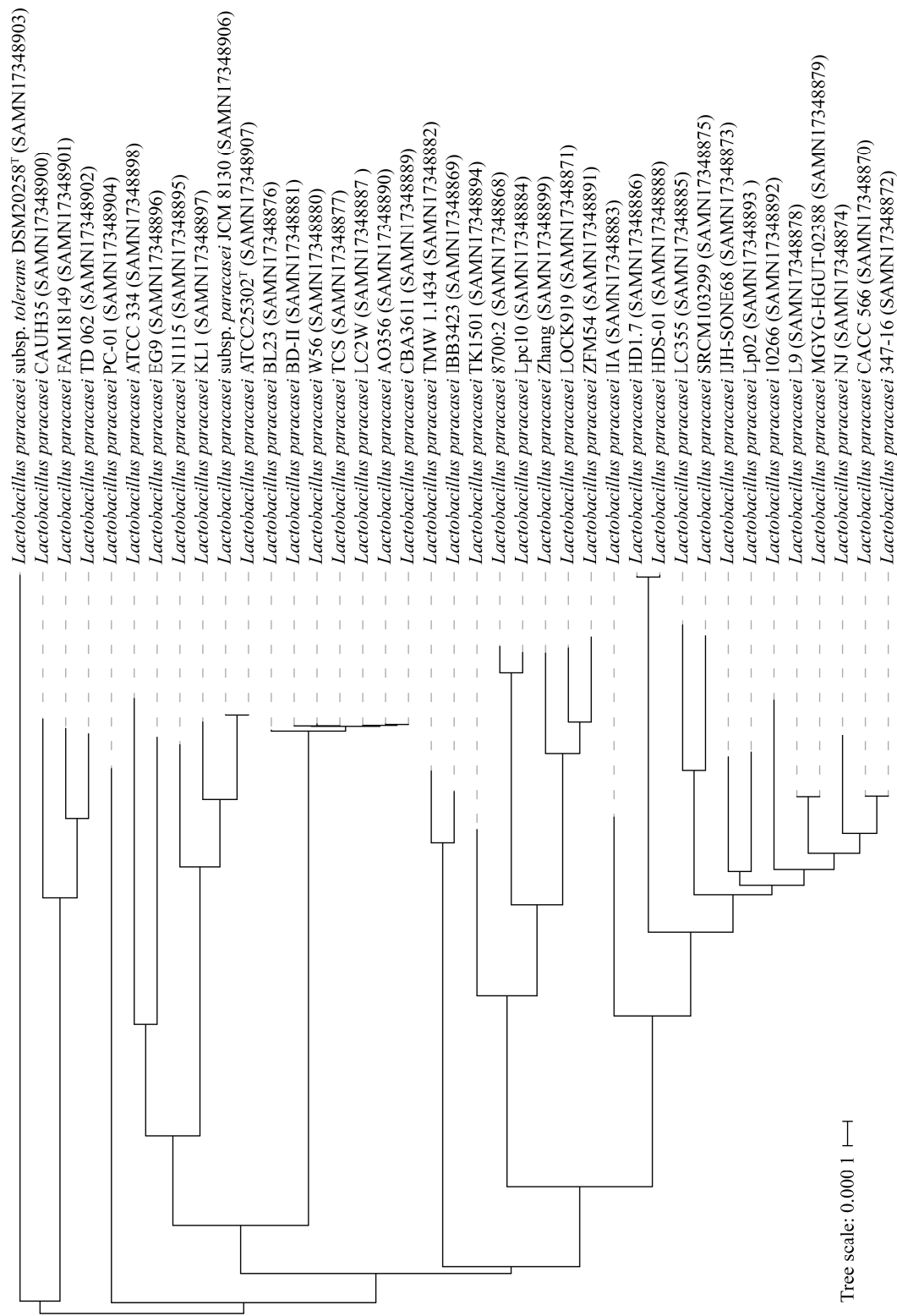


图 6 基于核心基因序列的系统发育树

Figure 6 Phylogenetic tree based on core gene sequences

注：比例尺显示水平线的长度，代表碱基替换数

Note: The scale bar represents the horizontal branch lengths and the number of substitutions per nucleotide position

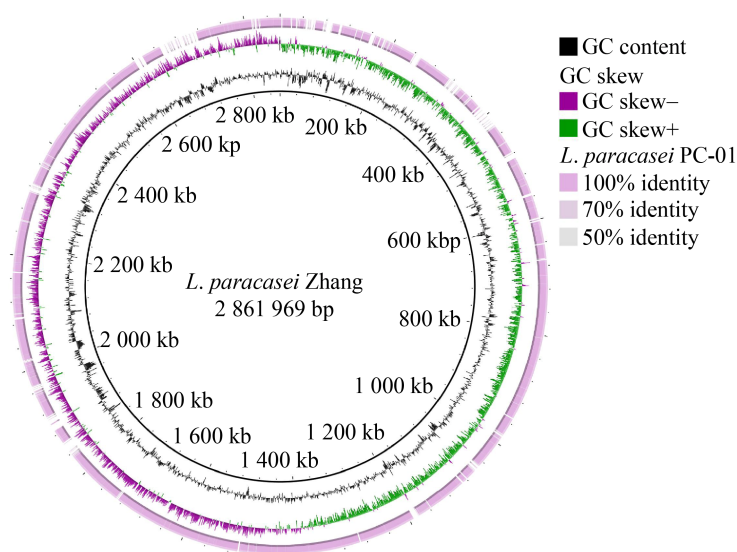


图 7 以副干酪乳杆菌 Zhang 为参考基因组的 BRIG 分析

Figure 7 BRIG analysis based on *L. paracasei* Zhang as reference genome

性可能与该基因有关。Alayande 等^[31]对嗜酸乳杆菌 PNW3 基因组测序组装后, 注释到 39 个具有重要益生特性相关蛋白, 分别与乳酸、生物活性肽及黏附分子的生成有关。在 *L. paracasei* PC-01 和 Zhang 也发现 6 个与该研究相同的编码蛋白(表 2), 主要与乳酸、生物活性肽、黏附分子和抗应力分子的产生, 以及改善宿主的代谢有关。前期研究报道 *L. paracasei* Zhang 具有抗氧化、降血脂血糖及免疫调节等益生功能^[9-10], Wang 等^[32]发现便秘人群摄入 *L. paracasei* Zhang 益生菌发酵乳后, 通过调节炎症、肠道微生物和代谢途径改善了便秘情况。从基因组水平注释到 *L. paracasei* PC-01 具有益生相关的蛋白, 表明 *L. paracasei* PC-01 是一株具有潜在益生特性的菌株, 但其益生功能需进一步探究。

3 讨论

L. paracasei Zhang 是 2004 年从内蒙古自治区锡林郭勒盟草原传统发酵酸马奶分离得到的一株具有益生特性的乳酸菌^[8]。目前许多研究证实, *L. paracasei* Zhang 具有耐酸、耐胆盐等特性, 可在人体肠道定殖, 对于宿主免疫调节、降低胆固

醇等方面均有一定作用^[33], 该菌已在生产中有广泛推广和应用。本实验室 2007 年从西藏自治区拉萨传统牦牛发酵乳中分离得到的 *L. paracasei* PC-01 也具有良好益生特性。*L. paracasei* 及其近缘物种之间亲缘关系复杂, 其分类学地位一直存在争议难以确定^[34]。本研究旨在基因组水平上通过比较基因组学方法揭示 2 株具有益生特性的 *L. paracasei* 之间的差异。结果显示 *L. paracasei* PC-01 基因组大小为 2 829 251 bp, 没有质粒; *L. paracasei* Zhang 基因组大小为 2 898 456 bp, 包含 1 个质粒; 二者基因组 GC 含量及 CDS 数量各不相同; *L. paracasei* PC-01 和 *L. paracasei* Zhang 之间 ANI 值为 98.37%, TNI 值为 80.74%。通过对 39 株 *L. paracasei* 的基因组分析发现, *L. paracasei* 属于开放式基因组, 不同 *L. paracasei* 的基因组信息各有不同, 而且它们之间的核酸一致性存在差异。上述结果表明 *L. paracasei* 基因组具有高度遗传多样性。

此外, 本研究先后通过基于 16S rRNA、看家基因 *murE-uvrC* 串联及核心基因分别构建种内系统发育树, 由于干酪乳杆菌及其近缘物种 16S rRNA

表 2 *L. paracasei* PC-01 和 *L. paracasei* Zhang 具有重要益生作用的编码蛋白
Table 2 Protein-encoding sequences putative for important probiotic functions of *L. paracasei* PC-01 and *L. paracasei* Zhang

Important probiotic features	Associated functional proteins/peptides
Production of lactic acids	L-lactate dehydrogenase
Production of bioactive peptides	S-ribosylhomocysteine lyase
Production of adhesion molecules	ATP synthase epsilon chain
Production of stress resistance molecules	ATP-binding subunit C ₁ pE
Improving host metabolism	Poly (glycerol-phosphate) alpha-glucosyltransferase (EC 2.4.1.52) PTS system, cellobiose-specific IIB component

基因序列相似性高达 99%，因此仅通过 16S rRNA 基因序列和表型结合难以将干酪乳杆菌群准确区分^[35]，前期研究表明看家基因 *murE-uvrC* 串联序列比 16S rRNA 基因序列对干酪乳杆菌的区分效果好^[26]，周江林等^[36]认为看家基因区分洋葱伯克霍尔德复合群菌株是存在异常的。本研究中 16S rRNA 基因序列和看家基因 *murE-uvrC* 串联序列均不能在菌株水平有效区分 *L. paracasei* 菌株，但基于 39 株 *L. paracasei* 的 1 458 个核心基因所构建的系统发育树，可明显区分 *L. paracasei* PC-01 和 *L. paracasei* Zhang 在不同的进化分支，并可在菌株水平上对 39 株菌株进行有效区分。从系统发育结构上看，*L. paracasei* PC-01 与 *L. paracasei* subsp. *tolerans* DSM20258^T 遗传距离更近。表明 2 株益生菌菌株具有不同的遗传背景，基于核心基因的系统发育树可以有效在菌株水平对不同益生菌菌株进行区分。

为进一步了解 2 株益生菌基因组的差异，本研究以 *L. paracasei* Zhang 为参考菌株，使用 BRIG 软件进行分析发现 *L. paracasei* PC-01 大约 91%基因组片段与 *L. paracasei* Zhang 基因组匹配率大于 90%，占 *L. paracasei* PC-01 基因组 0.4%的片段与 *L. paracasei* Zhang 的匹配率小于 70%。此外，通过 Roary 对注释结果统计 *L. paracasei* PC-01 和 *L. paracasei* Zhang 各自的特有基因，*L. paracasei* PC-01 注释到 169 个特有基因，其中 159 个是假定蛋白，注释到 10 个与转录调节、鼠李糖代谢、N-乙酰半乳糖氨基转移酶合成代谢等功能相关的

特有基因，如 *aguA*、*araE*、*rhaD*，其中与鼠李糖代谢的特有基因包含 4 个，因此推测 *L. paracasei* PC-01 可能有较强的鼠李糖代谢或合成能力；*L. paracasei* Zhang 注释到的特有基因数量相对较少，共包含 51 个特有基因，其中有 4 个注释到功能的基因，包括 *epsE*、*yybR* 等基因，与糖基转移、转录调节等代谢功能相关。进一步分析发现 *L. paracasei* PC-01 和 *L. paracasei* Zhang 基因组上存在 BSH 基因，这与二者良好的耐胆盐特性相符合。上述结果表明，通过比较基因组的方法可揭示具有益生特性的 *L. paracasei* PC-01 和 *L. paracasei* Zhang 在基因组基本信息和特有基因等方面的差异，以为后续益生菌株的开发奠定遗传学基础。

4 结论

随着生物测序技术不断进步及生物信息分析的更新，越来越多的 *L. paracasei* 菌株基因组完成了测序。本研究通过 PacBio SMRT 三代测序技术对 *L. paracasei* 基因组进行测序和组装，结合比较基因组学方法发现 *L. paracasei* PC-01 和 *L. paracasei* Zhang 基因组大小、质粒数和二者基因组匹配度均存在差异。不同 *L. paracasei* 基因组长度、质粒数及 GC 含量各不相同。核心基因构建的系统发育树对 *L. paracasei* 种内区分效果最好，*L. paracasei* PC-01 和 *L. paracasei* Zhang 处于不同的进化分支。与 *L. paracasei* Zhang 相比，有 0.4%的 *L. paracasei* PC-01 基因组片段匹配率小于 70%，*L. paracasei* PC-01 注释到 10 个特有基因，

如 *aguA*、*clpB* 等, 与转录调节、鼠李糖代谢等功能相关; *L. paracasei* Zhang 包含 *mshA*、*ltrA* 等特有基因, 发现与糖基转移等功能相关。通过比较基因组学揭示 *L. paracasei* 不同菌株遗传信息, 发现不同 *L. paracasei* 菌株之间的差异, 为不同 *L. paracasei* 益生菌菌株的区分鉴定和应用奠定了遗传学基础。

REFERENCES

- [1] Yang ZZ, Hou BC, Xie XN, Guo BH. Adhesion analysis of *Lactobacillus paracasei* BS 08 based on genomics[J]. China Dairy Industry, 2020, 48(8): 10-14 (in Chinese)
杨郑州, 侯保朝, 谢晓娜, 郭炳豪. 基于基因组学对副干酪乳杆菌 BS 08 黏附作用的分析[J]. 中国乳品工业, 2020, 48(8): 10-14
- [2] Zhang J, Zhao BT, Yang FM. Review on the probiotic characteristics and application of *Lactobacillus paracasei*[J]. Packaging and Food Machinery, 2019, 37(4): 47-52 (in Chinese)
张俊, 赵保堂, 杨富民. 副干酪乳杆菌的益生特性及其应用研究进展[J]. 包装与食品机械, 2019, 37(4): 47-52
- [3] Yang X, Guo WH. Function and safety evaluation of probiotics[J]. China Food Safety Magazine, 2018(12): 47-48 (in Chinese)
杨秀, 郭文慧. 益生菌的功能及安全性评价[J]. 食品安全导刊, 2018(12): 47-48
- [4] Wei LP, Liu YY, Dubchak I, Shon J, Park J. Comparative genomics approaches to study organism similarities and differences[J]. Journal of Biomedical Informatics, 2002, 35(2): 142-150
- [5] Sun ZH, Harris HMB, McCann A, Guo CY, Argimón S, Zhang WY, Yang XW, Jeffery IB, Cooney JC, Kagawa TF, et al. Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera[J]. Nature Communications, 2015, 6: 8322
- [6] Tamara S, Michiel W, Justyna P, Christian C, Sylvain B, Jos B, Johan ETHV, Roland JC. *Lactobacillus paracasei* comparative genomics: towards species pan-genome definition and exploitation of diversity[J]. PLoS One, 2013, 8(7): e68731
- [7] Bao QH, Liu WJ, Yu J, Wang WH, Qing MJ, Chen X, Wang F, Zhang JC, Zhang WY, Qiao JM, et al. Isolation and identification of cultivable lactic acid bacteria in traditional yak milk products of Gansu Province in China[J]. The Journal of General and Applied Microbiology, 2012, 58(2): 95-105
- [8] Meng H, Wu RN, Wang LP, Yang XJ, Xu J, Dong Y, Sun ZH, Zhang HP. Isolation and identification of *Lactobacillus* from koumiss collected in Inner Mongolia and People's Republic of Mongolia[J]. China Dairy Industry, 2004, 32(11): 6-11 (in Chinese)
孟和毕力格, 乌日娜, 王立平, 杨续金, 徐杰, 董莹, 孙志宏, 张和平. 不同地区酸马奶中乳杆菌的分离及其生物学特性的研究[J]. 中国乳品工业, 2004, 32(11): 6-11
- [9] Tuoya, Suyalema, Zhang HP. Effect of *Lb.casei* Zhang on koumiss on the production of cytokines in sera of mouse[J]. Food Science, 2006, 27(11): 488-491 (in Chinese)
托娅, 苏雅勒玛, 张和平. 乳杆菌 *Lb.casei* Zhang 对小鼠血清中细胞因子水平的影响[J]. 食品科学, 2006, 27(11): 488-491
- [10] Zhang JC, Guo Z, Sun ZH, Chen W, Zhang HP. The effects of probiotic on human intestinal microbiota: taking *Lactobacillus casei* Zhang as example[J]. Journal of Chinese Institute of Food Science and Technology, 2011, 11(9): 58-68 (in Chinese)
张家超, 郭壮, 孙志宏, 陈卫, 张和平. 益生菌对肠道菌群的影响: 以 *Lactobacillus casei* Zhang 研究为例[J]. 中国食品学报, 2011, 11(9): 58-68
- [11] Liu WJ, Chen MX, Duo LN, Wang JC, Guo S, Sun HT, Menghe B, Zhang HP. Characterization of potentially probiotic lactic acid bacteria and bifidobacteria isolated from human colostrum[J]. Journal of Dairy Science, 2020, 103(5): 4013-4025
- [12] Zhong Z, Sun TS, Chen YF. Genomic insights into the high exopolysaccharides-producing bacterium *Streptococcus thermophilus* ND-07[J]. China Dairy Industry, 2018, 46(4): 9-11, 21 (in Chinese)
钟智, 孙天松, 陈永福. 基因组分析揭示 *Streptococcus thermophilus* ND-07 富产胞外多糖分子机制[J]. 中国乳品工业, 2018, 46(4): 9-11, 21
- [13] Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. Circlator: automated circularization of genome assemblies using long sequencing reads[J]. Genome Biology, 2015, 16(1): 294
- [14] Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 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
- [15] Chen JP, Yang XW, Chen JW, Cen Z, Cui YJ. SISP: a fast species identification system for prokaryotes based on total nucleotide identity of whole genome sequences[EB/OL]. 2015. DOI: 10.11979/idxm.201561008
- [16] Seemann T. Prokka: rapid prokaryotic genome annotation[J]. Bioinformatics, 2014, 30(14): 2068-2069
- [17] Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. Roary: rapid large-scale prokaryote Pan genome analysis[J]. Bioinformatics, 2015, 31(22): 3691-3693
- [18] Bao QH, Song YQ, Xu HY, Yu J, Zhang WY, Menghe B,

- Zhang HP, Sun ZH. Multilocus sequence typing of *Lactobacillus casei* isolates from naturally fermented foods in China and Mongolia[J]. Journal of Dairy Science, 2016, 99(7): 5202-5213
- [19] Lagesen K, Hallin P, Rødland EA, Stærfeldt HH, Rognes T, Ussery DW. RNaMMer: consistent and rapid annotation of ribosomal RNA genes[J]. Nucleic Acids Research, 2007, 35(9): 3100-3108
- [20] Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons[J]. BMC Genomics, 2011, 12(1): 402
- [21] Chirag J, Luis MRR, Adam MP, Konstantinos TK, Srinivas A. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries[J]. Nature Communications, 2018, 9: 5114
- [22] Li XS, Yin RM, Mao BY, Cui SM, Zhao JX. Genetic diversity and antibiotic resistance of *Lactobacillus paracasei*[J]. Food and Fermentation Industries, 2019, 45(14): 1-8 (in Chinese)
李晓姝, 殷瑞敏, 毛丙永, 崔树茂, 赵建新. 副干酪乳杆菌的基因多样性及其抗生素耐受性分析[J]. 食品与发酵工业, 2019, 45(14): 1-8
- [23] Wang YJ. Comparative genomic analysis of *Enterococcus faecalis* isolated from dairy products and other environments[D]. Hohhot: Master's Thesis of Inner Mongolia Agricultural University, 2018 (in Chinese)
王彦杰. 粪肠球菌乳源分离株与其它环境分离株基因组比较分析[D]. 呼和浩特: 内蒙古农业大学硕士学位论文, 2018
- [24] Song YQ. Study on population genetics and functional genomics of *Lactobacillus delbrueckii* subsp. *bulgaricus*[D]. Hohhot: Doctoral Dissertation of Inner Mongolia Agricultural University, 2018 (in Chinese)
宋宇琴. 德氏乳杆菌保加利亚亚种的群体遗传学和功能基因组学研究[D]. 呼和浩特: 内蒙古农业大学博士学位论文, 2018
- [25] Tang YW, Ellis NM, Hopkins MK, Smith DH, Dodge DE, Persing DH. Comparison of phenotypic and genotypic techniques for identification of unusual aerobic pathogenic gram-negative bacilli[J]. Journal of Clinical Microbiology, 1998, 36(12): 3674-3679
- [26] Duolana, Song YQ, Li WC, Li J, Xu HY, Sun ZH, Zhang HP. Phylogenetic analysis of *Lactobacillus casei* and closely related species/subspecies based on sequences of housekeeping genes[J]. Microbiology China, 2018, 45(12): 2751-2761 (in Chinese)
多拉娜, 宋宇琴, 李伟程, 李晶, 徐海燕, 孙志宏, 张和平. 干酪乳杆菌的近缘种及亚种部分看家基因的系统发育分析[J]. 微生物学通报, 2018, 45(12): 2751-2761
- [27] Mori K, Yamazaki K, Ishiyama T, Katsumata M, Kobayashi K, Kawai Y, Inoue N, Shinano H. Comparative sequence analyses of the genes coding for 16S rRNA of *Lactobacillus casei*-related taxa[J]. International Journal of Systematic Bacteriology, 1997, 47(1): 54-57
- [28] Salvetti E, Torriani S, Felis GE. The genus *Lactobacillus*: a taxonomic update[J]. Probiotics and Antimicrobial Proteins, 2012, 4(4): 217-226
- [29] Hu AH, Ao XL, Chen C, Pu B, Chen AJ, Jiang HX. Research progress on mechanism of lactic acid bacteria acid and bile salt resistance[J]. Science and Technology of Food Industry, 2015, 36(8): 380-383, 389 (in Chinese)
胡爱华, 敖晓琳, 陈岑, 蒲彪, 陈安均, 姜欢笑. 乳酸菌耐酸耐胆盐机制的研究进展[J]. 食品工业科技, 2015, 36(8): 380-383, 389
- [30] Zhang WY, Yu DL, Sun ZH, Wu RN, Chen X, Chen W, Meng H, Hu SN, Zhang HP. Complete genome sequence of *Lactobacillus casei* Zhang, a new probiotic strain isolated from traditional homemade koumiss in Inner Mongolia, China[J]. Journal of Bacteriology, 2010, 192(19): 5268-5269
- [31] Alayande KA, Aiyegoro OA, Ateba CN. Distribution of important probiotic genes and identification of the biogenic amines produced by *Lactobacillus acidophilus* PNW₃[J]. Foods, 2020, 9(12): 1840
- [32] Wang JC, Bai XY, Peng CT, Yu ZJ, Li BH, Zhang WY, Sun ZH, Zhang HP. Fermented milk containing *Lactobacillus casei* Zhang and *Bifidobacterium animalis* ssp. *lactis* V9 alleviated constipation symptoms through regulation of intestinal microbiota, inflammation, and metabolic pathways[J]. Journal of Dairy Science, 2020, 103(12): 11025-11038
- [33] Guo Z, Wang JC, Yan LY, Chen W, Liu XM, Zhang HP. *In vitro* comparison of probiotic properties of *Lactobacillus casei* Zhang, a potential new probiotic, with selected probiotic strains[J]. LWT - Food Science and Technology, 2009, 42(10): 1640-1646
- [34] Li XS. Differences in genomics and physiological characteristics of *Lactobacillus paracasei* from different sources[D]. Wuxi: Master's Thesis of Jiangnan University, 2019 (in Chinese)
李晓姝. 不同来源的副干酪乳杆菌的基因组及生理特性差异研究[D]. 无锡: 江南大学硕士学位论文, 2019
- [35] Zhang Z, Yu XJ, Cao YH, Yao S. Taxonomic changes of *Lactobacillus casei* and its closely related species[J]. Food and Fermentation Industries, 2020, 46(16): 243-248 (in Chinese)
张哲, 于学健, 曹艳花, 姚粟. 干酪乳杆菌 *Lactobacillus casei* 及相近菌种的分类变迁[J]. 食品与发酵工业, 2020, 46(16): 243-248
- [36] Zhou JL, Kong N, Zhang Q, Hu MD, Zhou J, Yue JJ, Ren HG, Jin Y, Liang L. Genome-based correction of some strains of *Burkholderia cepacia* complex[J]. Letters in Biotechnology, 2019, 30(6): 733-739 (in Chinese)
周江林, 孔娜, 张琪, 胡明达, 周静, 岳俊杰, 任洪广, 靳远, 梁龙. 洋葱伯克霍尔德菌复合群部分菌株基于基因组的物种水平重鉴定[J]. 生物技术通讯, 2019, 30(6): 733-739