



食品用乳酸杆菌和双歧杆菌泛基因组多态性及菌株间功能基因差异分析

陈楠^{1,2,3}, 郭亚琼^{2,4}, 律娜^{1,2}, 李凤琴⁵, 陈历俊^{6,7}, 向雪松^{1,8}, 王欣^{1,9}, 刘烈刚^{1,10}, 王红伟¹, 智发朝^{1,11}, 何方^{1,12}, 陈卫¹³, 杨瑞馥^{1,14}, 朱宝利^{1,2,15,16,17*}

1. 中国营养学会益生菌益生元与健康分会, 北京 100052;
2. 中国科学院微生物研究所, 北京 100101;
3. 微源数智(北京)医疗科技有限公司, 北京 100068;
4. 中国科学院大学, 北京 100049;
5. 国家食品安全风险评估中心, 国家卫生健康委员会食品安全风险评估重点实验室, 北京 100021;
6. 北京三元食品股份有限公司, 国家母婴乳品健康工程技术研究中心, 北京 100163;
7. 北京市乳品工程技术研究中心, 母乳研究技术创新中心, 北京 100163;
8. 中国疾病预防控制中心营养与健康所, 北京 100050;
9. 浙江省农业科学院食品科学研究所, 杭州 310021;
10. 华中科技大学同济医学院公共卫生学院, 武汉 430030;
11. 南方医科大学南方医院消化科, 广东省南方消化病研究所广东省胃肠病重点实验室, 广州 510515;
12. 四川大学华西公共卫生学院/四川大学华西第四医院, 成都 610041;
13. 江南大学食品学院, 无锡 214122;
14. 军事科学院军事医学研究院微生物流行病学研究所, 北京 100071;
15. 济南微生态生物医学省实验室, 济南 250117;
16. 西南医科大学基础医学院, 泸州 646000;
17. 国家人类基因组北方研究中心/北京诺赛基因组研究中心有限公司, 北京 100176

* 联系人, E-mail: zhubaoli@im.ac.cn

收稿日期: 2023-03-17; 接受日期: 2023-05-06; 网络版发表日期: 2023-05-16

国家重点研发计划(批准号: 2021YFA1301002)、济南微生态生物医学省实验室科研项目(批准号: JNL-2022013B)和国家自然科学基金(批准号: 32170068)资助

摘要 益生菌是一类有益于人体健康的微生物, 主要包括细菌和真菌. 益生菌的鉴定一直是学界亟需解决的技术问题, 传统的分类学方法只能做种属水平的鉴定, 而基于基因组学分析可鉴定到种或株水平. 本研究对公共数据库中记载的我国《可用于食品的菌种名单》中的16种乳酸杆菌和5种双歧杆菌的基因组数据进行了全基因组比对分析, 并应用核心基因平均核苷酸相似性(core genes average nucleotide identity, cANI)进行了物种的重新分类和鉴定方法探索. 结果显示, 乳酸杆菌和双歧杆菌的cANI值具有种属特异性, 其中乳酸杆菌的cANI值在93.6%~99.6%之间, 双歧杆菌的cANI值在94.9%~98.1%之间, 同时发现25株乳酸杆菌和1株两歧双歧杆菌有分类学错误. 对247株乳酸杆菌和113株双歧杆菌基因组完成图的比对分析结果显示, 乳酸杆菌和双歧杆菌菌株间的差异基因

引用格式: 陈楠, 郭亚琼, 律娜, 等. 食品用乳酸杆菌和双歧杆菌泛基因组多态性及菌株间功能基因差异分析. 中国科学: 生命科学, 2023, 53: 711–724
Chen N, Guo Y Q, Lyu N, et al. Identification, safety and functional prediction of food-grade lactobacilli and *Bifidobacterium* based on whole genome sequences (in Chinese). Sci Sin Vitae, 2023, 53: 711–724, doi: 10.1360/SSV-2023-0043

数目最多可达436个, 菌株间的cANI相似值最小可达1个单核苷酸多态性(single nucleotide polymorphism, SNP). 针对乳酸杆菌和双歧杆菌的毒力基因和耐药基因分析显示, 在目前用于食品生产的菌株中均未发现毒力相关基因, 同时发现有20%以上的菌株携带不同种类的耐药基因. 结论: 乳酸杆菌和双歧杆菌不同种的cANI值差异较大, 有种属特异性, 可用于益生菌的种属鉴定; 同一物种不同菌株间的基因和SNP数目差异可用于菌株鉴定.

关键词 益生菌, 泛基因组, cANI, 菌株差异基因, SNP, 毒力相关基因, 耐药基因

益生菌是指在摄入(或给与)一定量的基础上, 能给宿主带来健康益处的活微生物^[1,2], 其功能为改善人体肠道微生态, 调节人体免疫功能^[3]. 截至2022年7月, 在全球范围内使用的益生菌种类涵盖36个属84个种, 其中细菌29个属74个种, 真菌7个属10个种^[4]. 细菌类的益生菌主要为乳酸杆菌(*Lactobacilli*)和双歧杆菌(*Bifidobacterium*), 包括38种乳酸杆菌和5种双歧杆菌, 其中只有一部分经过益生功能验证的才能称为益生菌. 中国国家卫生健康委员会公布的《可用于食品的菌种名单》(2022年8月更新版)规定了可用于食品的16个属33个种的细菌(包括16种乳酸杆菌和5种双歧杆菌)和1种酵母菌属的真菌. 中国国家药典委员会公布的《中华人民共和国药典》(2020年版)规定了可用于药品的细菌菌种(株), 分属于5个属11个种(包括2种乳酸杆菌和3种双歧杆菌)^[5]. 尽管各个国家或地区有比较严格的管理制度和规范, 但目前益生菌的生产和使用仍存在三个方面的问题, 即益生菌的分类鉴定、益生菌的功能评价和安全性评估.

目前益生菌的菌种鉴定主要有三种方法, 其一为传统的分类学方法, 主要依赖于细菌的表型特征和生理生化特性. 这种方法一般只能鉴定到科水平, 很难准确鉴定到属或种水平^[6-8]. 因此, 自20世纪70年代以来, 基于分子生物学的分子分型和鉴定方法逐渐占据主导地位; 但是由于分子分型和鉴定方法是基于细菌16S rRNA基因序列的差异(同种相似度98.7%以上), 或者基于两个细菌基因组DNA的杂交系数CoT值的大小(同种杂交系数在70%以上)^[7,9], 因此这种方法一般只能鉴定到属或种水平^[10-12]. 基于基因组学的分类学鉴定方法起始于2010年之后^[13,14], 是目前细菌种水平鉴定的较为可靠的方法, 也是未来细菌种水平鉴定的主要方法. 近年来, 为了尽量减少基因水平转移的影响, 研究者以单拷贝核心基因(single copy core genes)的平均核苷酸相似性(average nucleotide identity, ANI)

或平均氨基酸相似性(average amino acid identity, AAI)作为细菌物种分类鉴定的判断标准^[15,16], 以进行更准确的系统发育分析或物种分类鉴定. 另一方面, 基于微生物基因组学的分析方法还可以校正传统菌种鉴定方法和分子分型方法准确性不足导致的错误, 例如, 乳酸杆菌早期将261个种均归属为一个乳杆菌属(*Lactobacillus*), 但是从分类学角度看, 乳杆菌属下不同的细菌种群在基因组、形态学和生态学多样化方面超过了一般细菌种属的多样性程度. 对原乳杆菌属的2443菌株基因组进行大规模测序分析发现, 该菌属可重新分类为24个属^[15,16]. 同样的问题也存在于芽孢杆菌属(*Bacillus*): 在大规模分析了原芽孢杆菌属(科)303个种的基因组之后, 将它们重新分类为18个属^[17]. 目前, 国际分类权威组织在LPSN(List of Prokaryotic names with Standing in Nomenclature)中已经接受了上述物种的分类学地位变化和新的分类学名称. 中国也对相关食品用的菌种名单进行了更新, 其中涉及到的物种有9种乳酸杆菌、2种双歧杆菌、1种芽孢杆菌、1种丙酸杆菌和1种葡萄球菌.

在不同物种的益生菌功能鉴定方面, 目前国内外主要采用体外实验、动物试验和人体实验等方式鉴定和评估益生菌菌株的功能, 还没有采用基因组学的分析评估方法, 主要原因为同一物种不同菌株的基因组所携带的功能基因差异大, 涉及的问题主要是细菌物种的泛基因组(pangenome)概念. 泛基因组最早由美国科学家提出, 其定义为同一物种细菌中不同菌株基因总数的总和^[18]. 泛基因组可以分为核心基因组(core genome)和非必需基因组(dispensable genome), 其中核心基因组是指同一物种中95%以上的菌株基因组共有的基因, 是此物种生存必需的基因. 因此, 同一物种益生菌的功能基因是不同的, 例如, 鼠李糖乳酪杆菌(*Lactocaseibacillus rhamnosus*)的功能基因簇*spaCBA-srtC1*, 其编码的菌毛与肠道定植、调节人体免疫水

平、抵抗致病菌等多种功能密切相关, 仅有27%的鼠李糖乳酪杆菌菌株含有该功能基因簇^[19]; 植物乳植杆菌(*Lactiplantibacillus plantarum*)的*ltaS*, *tagG*等12个磷壁酸相关的功能基因, 与宿主的免疫调节、抗炎等功能密切相关, 在599个基因组中仅有两株含有完整的12个基因的功能基因通路^[20]。

益生菌的安全性是益生菌评价的另一重要问题, 涉及的内容主要为菌株的耐药性和潜在的致病性。目前, 国内外进行益生菌安全性评估的技术和方法主要有基于基因组学的毒力相关基因分析和耐药基因分析、利用高效液相色谱检测潜在毒性物质、用最小抑菌浓度(minimum inhibitory concentration, MIC)进行耐药性检测和动物及人体临床试验等^[21~23], 后两种是传统的安全性评估方法。基因组分析能够获得益生菌菌株的毒力相关基因、耐药基因等功能基因的信息, 一般作为菌株安全性评估的第一步, 后续再结合传统的评估技术和方法, 以获得益生菌的耐药性和潜在的致病性信息。细菌毒力相关基因和耐药基因的多样性比较复杂, 目前国际上尚未形成统一的标准。

本研究对我国可用于食品的16种乳酸杆菌和5种双歧杆菌进行泛基因组分析, 利用cANI对分类不准确的菌株分类进行了校正, 并且进一步采用不同物种菌株基因组的完成图进行比较分析, 在基因组水平研究同种不同菌株之间的基因差异性, 初步探究基因组学技术用于食品工业用菌株水平鉴定的可能性, 结合对毒力相关基因、耐药基因的预测分析, 为益生菌安全评估提供科学依据。

1 材料与方法

1.1 数据收集

本研究自美国国家生物技术信息中心(National Center for Biotechnology Information, NCBI)的GenBank数据库(<https://ncbi.nlm.nih.gov/genbank>)下载了16种乳酸杆菌和5种双歧杆菌的基因组数据(截止到2021年5月20日), 共3818个基因组。上述物种均为我国《可用于食品的菌种名单》中的物种(亚种按照种水平进行分析), 其中部分乳酸杆菌和双歧杆菌菌株为食品生产用菌株。本文中乳酸杆菌种名使用分类更新后的中文名称, 更新前后中文、拉丁文名称对照关系详见网络版附表1。

1.2 基因组质控、注释及泛基因组分析

使用CheckM软件的默认参数(estimated completeness>50%, estimated contamination<10%)对基因组的完整性、污染度等指标进行评估, 保留下的3780个质量基因组使用PROKKA v.1.13软件进行注释。使用PROKKA gff作为输入文件, 利用软件Roary v.3.12.0进行泛基因组分析。乳酸杆菌每个物种的基因组数量如下: 嗜酸乳杆菌(*L. acidophilus*)58个, 卷曲乳杆菌(*L. crispatus*)156个, 德氏乳杆菌(*L. delbrueckii*)103个, 格式乳杆菌(*L. gasseri*)55个, 瑞士乳杆菌(*L. helveticus*)144个, 约氏乳杆菌(*L. johnsonii*)60个, 马乳酒样乳杆菌(*L. kefirifaciens*)9个, 干酪乳酪杆菌(*L. casei*)18个, 副干酪乳酪杆菌(*L. paracasei*)242个, 鼠李糖乳酪杆菌226个, 发酵粘液乳杆菌(*Limosilactobacillus fermentum*)96个, 罗伊氏粘液乳杆菌(*L. reuteri*)208个, 植物乳植杆菌599个, 唾液联合乳杆菌(*Ligilactobacillus salivarius*)193个, 弯曲广布乳杆菌(*Latilactobacillus curvatus*)24个, 清酒广布乳杆菌(*L. sakei*)56个。双歧杆菌每个物种的基因组数量如下: 青春双歧杆菌(*B. adolescentis*)325个, 动物双歧杆菌(*B. animalis*)91个, 两歧双歧杆菌(*B. bifidum*)123个, 短双歧杆菌(*B. breve*)123个, 长双歧杆菌(*B. longum*)871个。

1.3 核心基因平均核苷酸相似性(cANI)分析

对每一个物种内的所有菌株进行Roary分析, 得到每个物种的单拷贝核心基因, 使用软件FastANI计算每一个物种内所有菌株的cANI, 使用Perl脚本将分析结果整理成矩阵后, 使用R语言的ggplot包进行作图。

1.4 系统发育树的构建

利用核心基因序列构建系统发育树, 分析乳酸杆菌和双歧杆菌同一物种不同菌株的系统发育关系。使用MATTF软件对核心基因进行比对, 再使用FastTree软件采用最大似然法构建系统发育树。

1.5 基因组比对分析

为进一步研究乳酸杆菌和双歧杆菌菌株水平的功能基因差异, 本实验针对鼠李糖乳酪杆菌、副干酪乳酪杆菌、罗伊氏粘液乳杆菌、植物乳植杆菌等4种乳酸杆菌和短双歧杆菌、两歧双歧杆菌、长双歧杆菌等

3种双歧杆菌, 选取基因组为完成图级别的菌株, 进行基因组比对分析. 基于1.2节中的基因组注释和泛基因组分析结果, 对同一物种中的菌株进行两两比较, 得到菌株差异基因, 使用Perl脚本将菌株差异基因数量整理成矩阵后, 使用R语言的pheatmap包进行作图.

1.6 毒力相关基因和耐药基因分析

使用DIAMOND软件, 把每个基因组的氨基酸序列与毒力因子数据库(The Virulence Factor Database, VFDB)进行比对, 进行毒力相关基因的分析. 筛选蛋白质序列相似度 $\geq 95\%$, 覆盖度 $\geq 95\%$ 的比对结果作为阳性结果.

使用综合抗生素耐药性数据库(The Comprehensive Antibiotic Resistance Database, CARD)提供的耐药基因序列及耐药基因识别工具RGI v5.2.0, 按照默认参数对基因组进行耐药基因的分析^[24]. 可移动耐药基因使用软件ResFinder v4.0, 按照默认参数进行分析^[25]. 筛选蛋白质序列相似度 $\geq 95\%$, 覆盖度 $\geq 95\%$ 的比对结果作为阳性结果.

2 结果

2.1 食品用乳酸杆菌和双歧杆菌的泛基因组特征分析

本研究对重新分类的6个属的16种乳酸杆菌和双歧杆菌属的5种双歧杆菌进行了大规模基因组比对分析, 对3818个基因组以完整度 $> 50\%$, 污染度 $< 10\%$ 为标准进行质控, 获得了3780个高质量基因组, 其中乳酸杆菌为2247个基因组, 双歧杆菌为1533个基因组. 为了深入地分析不同益生菌菌株的功能基因, 本研究利用Roary软件分析了每一个物种的泛基因组特征, 结果见表1. 乳酸杆菌中, 基因组数量最多的种为植物乳植杆菌, 有599个的基因组, 泛基因组总基因数目为32034个, 核心基因组为1240个基因, 最大的基因组有3638个基因, 最小的基因组2635个基因; 基因组数量最少的为马乳酒样乳杆菌, 有9个基因组, 泛基因组总基因数目为4202个, 核心基因组为1504个基因, 最大的基因组有2961个基因, 最小的基因组有2091个基因. 双歧杆菌属中基因组数量最多的种为长双歧杆菌, 有871个基因组, 泛基因组总基因数目为25863个, 核心基因组为807个基因, 最大的基因组有2589个基因, 最小的基因

组有1359个基因; 基因组数量最少的为动物双歧杆菌, 有91个基因组, 泛基因组总基因数目为4352个, 核心基因组为1105个基因, 最大的基因组有1928个基因, 最小的基因组有1537个基因. 为了更直观展示乳酸杆菌和双歧杆菌的泛基因组特征, 绘制了每个物种的泛基因组和核心基因组的特征曲线(结果见网络版附图1), 发现各个物种的泛基因组基因数量随着基因组数量的增加而持续增加, 核心基因组基因数量则随着基因组数量的增加而减少, 具备开放型泛基因组的典型特征. 上述结果表明, 可用于食品的16种乳酸杆菌和5种双歧杆菌均为开放型泛基因组, 这与前期针对长双歧杆菌^[26]、植物乳植杆菌^[27]、唾液联合乳杆菌^[28]等包含益生菌菌株的菌种, 进行的基因组学研究结果相一致. 开放型泛基因组表明菌种具有较高的基因多样性, 与益生菌的益生功能菌株特异性密切相关.

2.2 乳酸杆菌中不同物种的核心基因组比对分析

乳酸杆菌是一类利用碳水化合物在代谢过程中发酵产生乳酸的革兰氏阳性、无芽孢且厌氧的杆状细菌, 被广泛应用于维护人体健康. 乳酸杆菌菌株原属于同一个乳杆菌属, 包含261个种, 随着基因组学分类学方法的发展, 国际上将乳酸杆菌重新进行了分类, 在校正了原有的乳杆菌属的同时, 规定了23个新属^[15,16]. 中国可用于食品的乳酸杆菌共有16个菌种, 其中分类变更的涉及9个种, 分别为干酪乳酪杆菌、副干酪乳酪杆菌、鼠李糖乳酪杆菌、发酵粘液乳杆菌、罗伊氏粘液乳杆菌、植物乳植杆菌、唾液联合乳杆菌、弯曲广布乳杆菌和清酒广布乳杆菌(详见表1和网络版附表1).

以基因组为基础进行分类学研究时, 一般以95%的基因组ANI作为细菌种水平分类鉴定的阈值^[14,29-32]. 为了减少水平基因转移的影响, 目前, 多数研究者以cANI作为细菌物种分类鉴定的判断标准^[15,16], 以进行更准确的系统发育分析或物种分类鉴定. 本研究采用cANI的方法分析了16种乳酸杆菌种内菌株的序列相似性, 基于cANI值的箱线图分别如图1A所示. 乳酸杆菌的16个种中, 种内最小cANI值高于95%的种有瑞士乳杆菌、卷曲乳杆菌、德氏乳杆菌、约氏乳杆菌、马乳酒样乳杆菌、副干酪乳酪杆菌、鼠李糖乳酪杆菌、发酵粘液乳杆菌和唾液联合乳杆菌, 其中最高的为瑞士乳杆菌, 种内最小cANI值为98.4%; 格氏乳杆菌和清酒广布乳杆菌的种内最小cANI值略小于95%, 分别为

表 1 《可用于食品的菌种名单》中的16种乳酸杆菌和5种双歧杆菌的泛基因组**Table 1** Genes distributed across 16 lactobacilli and 5 *Bifidobacterium* species pangenome in the List of Microorganisms Used in Food

中文翻译 ^{a)}	拉丁文名称	基因组数量 ^{b)}	泛基因组	核心基因组	非必需基因组
乳杆菌属	<i>Lactobacillus</i>				
嗜酸乳杆菌	<i>Lactobacillus acidophilus</i>	58	2822	1681	1141
卷曲乳杆菌	<i>Lactobacillus crispatus</i>	156	16836	753	16083
德氏乳杆菌	<i>Lactobacillus delbrueckii</i>	103	11137	734	10403
格式乳杆菌	<i>Lactobacillus gasseri</i>	55	7610	750	6860
瑞士乳杆菌	<i>Lactobacillus helveticus</i>	144	13589	858	12731
约氏乳杆菌	<i>Lactobacillus johnsonii</i>	60	9853	951	8902
马乳酒样乳杆菌	<i>Lactobacillus kefiranofaciens</i>	9	4202	1504	2698
乳酪杆菌属	<i>Lactocaseibacillus</i>				
干酪乳酪杆菌	<i>Lactocaseibacillus casei</i>	18	6207	1515	4692
副干酪乳酪杆菌	<i>Lactocaseibacillus paracasei</i>	242	23868	1027	22841
鼠李糖乳酪杆菌	<i>Lactocaseibacillus rhamnosus</i>	226	19261	1025	18236
粘液乳杆菌属	<i>Limosilactobacillus</i>				
发酵粘液乳杆菌	<i>Limosilactobacillus fermentum</i>	96	12612	742	11870
罗伊氏粘液乳杆菌	<i>Limosilactobacillus reuteri</i>	208	23089	858	22231
乳植杆菌属	<i>Lactiplantibacillus</i>				
植物乳植杆菌	<i>Lactiplantibacillus plantarum</i>	599	32034	1240	30794
联合乳杆菌属	<i>Ligilactobacillus</i>				
唾液联合乳杆菌	<i>Ligilactobacillus salivarius</i>	193	13776	1003	12773
广布乳杆菌属	<i>Latilactobacillus</i>				
弯曲广布乳杆菌	<i>Latilactobacillus curvatus</i>	24	6609	1182	5427
清酒广布乳杆菌	<i>Latilactobacillus sakei</i>	56	6285	1280	5005
双歧杆菌属	<i>Bifidobacterium</i>				
青春双歧杆菌	<i>Bifidobacterium adolescentis</i>	325	12803	1170	11633
动物双歧杆菌	<i>Bifidobacterium animalis</i>	91	4352	1105	3247
两歧双歧杆菌	<i>Bifidobacterium bifidum</i>	123	8405	1134	7271
短双歧杆菌	<i>Bifidobacterium breve</i>	123	11150	1142	10008
长双歧杆菌	<i>Bifidobacterium longum</i>	871	25863	807	25056

a): 中文属名参照国家卫生健康委公布的《可用于食品的菌种名单》更新名称; 与旧名称对应关系见网络版附表1. b): 经过物种校正之后的基因组数量

94.8%和93.6%; 嗜酸乳杆菌、弯曲广布乳杆菌、植物乳植杆菌、罗伊氏粘液乳杆菌和干酪乳酪杆菌的种内最小cANI值分别仅为82.77%, 87.99% 86.96%, 70.10%和70.45%, 远低于种水平鉴定的95%阈值, 表明在经过属水平重新分类后, 乳酸杆菌的菌株水平分类仍存在一定问题, 主要原因是NCBI是一个初级的基因组数据库, 特别是在微生物基因组数据方面, 存在许多物种分类学的错误, 干酪乳酪杆菌与副干酪乳酪杆菌在已有

研究中也表现出存在错误分类的情况^[29]. 对嗜酸乳杆菌、弯曲广布乳杆菌、植物乳植杆菌、罗伊氏粘液乳杆菌和干酪乳酪杆菌等5个种内的25个具有异常cANI值的菌株基因组, 进行“分类学检查”(NCBI Taxonomy Check)验证, 结果如网络版附表2所示, 嗜酸乳杆菌的2株基因组分别属于卷曲乳杆菌和瑞士乳杆菌; 弯曲广布乳杆菌的2株基因组与*L. fragifolii*最相似; 植物乳植杆菌的4株基因组分别与副植物乳植杆菌和戊糖乳杆

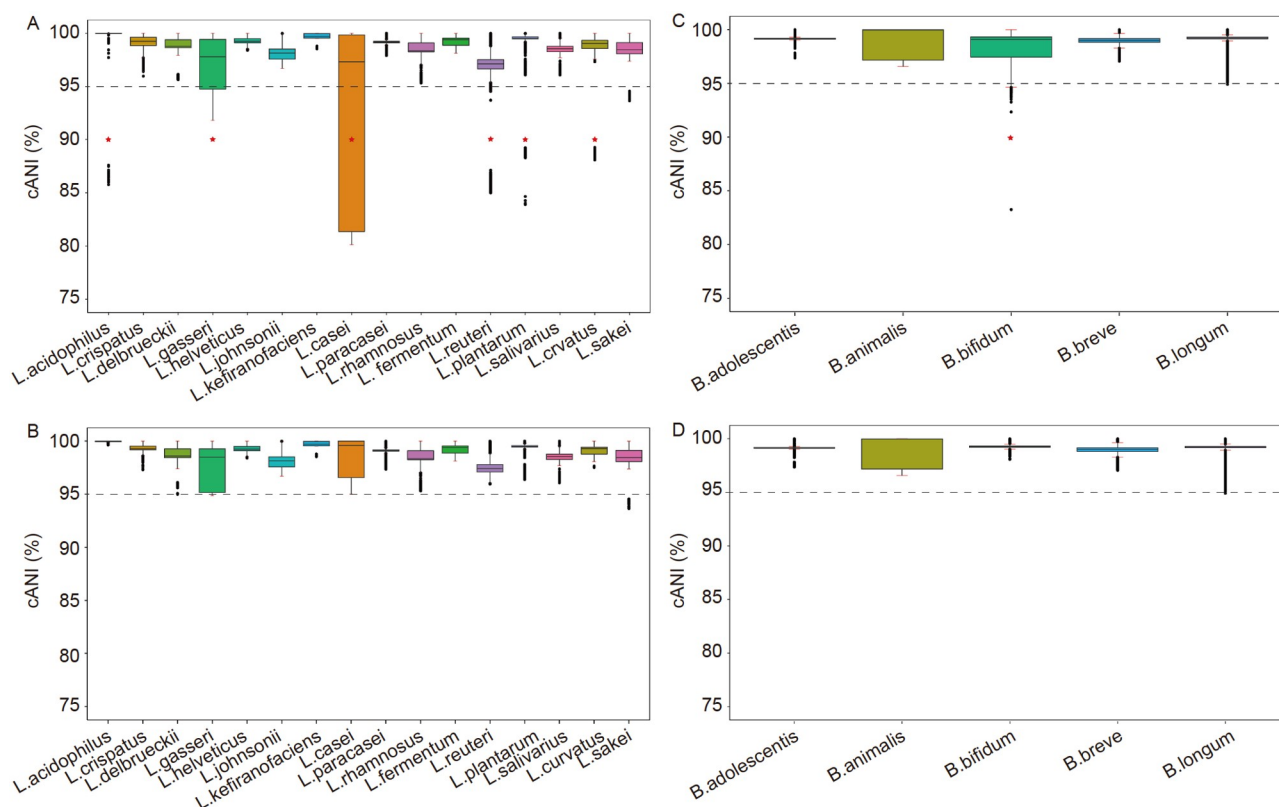


图1 《可用于食品的菌种名单》中的16种乳酸杆菌和5种双歧杆菌的种内cANI。A: 乳酸杆菌cANI的箱线图; B: 经过物种校正之后, 乳酸杆菌cANI的箱线图; C: 双歧杆菌cANI的箱线图; D: 经过物种校正之后, 双歧杆菌cANI的箱线图

Figure 1 The intra-species cANI values of 16 lactobacilli species and 5 *Bifidobacterium* species in the List of Microorganisms Used in Food. A: Box plot of the intra-species cANI values in the lactobacilli species (based on NCBI biosample); B: box plot of the intra-species cANI values in the lactobacilli species (after NCBI Taxonomy Check); C: box plot of the intra-species cANI values in the *Bifidobacterium* species (based on NCBI biosample); D: box plot of the intra-species cANI values in the *Bifidobacterium* species (after NCBI Taxonomy Check)

菌最相似, 1个基因组存在问题未找到最相似物种; 罗伊氏粘液乳杆菌的4株基因组分别与 *L. caviae*, *L. pontis* 和 *L. mucosae* 最相似; 干酪乳酪杆菌的12株基因组均与副干酪乳酪杆菌最相似。重新计算25株基因组的cANI值进一步验证了上述结果(图1B)。校正后的嗜酸乳杆菌、弯曲广布乳杆菌、植物乳植杆菌、罗伊氏粘液乳杆菌和干酪乳酪杆菌的种内最小cANI值分别为99.6%, 97.5%, 96.4%, 96.0%和95.0%, 均高于常规的95%阈值。上述结果表明, 在当前的分类学系统下, 已知的乳酸杆菌菌株信息中仍存在分类不准确的情况, 建议利用基因组并结合表型再次进行校正。此外, 16种乳酸杆菌的种内菌株间最小cANI各不相同, 在93.6%~99.6%之间, 分布范围较广, 建议在进行分类鉴定的时候, 不同乳酸杆菌采用不同的cANI阈值, 以提高种水平分类鉴定的准确性。还需注意的是, 清酒广布乳杆

菌的种内最小cANI为93.6%, 低于研究人员大多采用的95%阈值, 该菌种还存在进一步划分的可能性。

2.3 双歧杆菌中不同物种的核心基因组比对分析

一般认为, 双歧杆菌是一类革兰氏阳性、无芽孢的严格厌氧细菌, 现有100余种双歧杆菌被分类鉴定^[33]。目前, 我国《可用于食品的菌种名单》中涉及5种双歧杆菌, 分别为青春双歧杆菌、动物双歧杆菌(动物亚种、乳亚种)、两歧双歧杆菌、短双歧杆菌和长双歧杆菌(长亚种、婴儿亚种)。本研究计算了5种双歧杆菌的种内cANI值(根据NCBI biosample信息确定基因组物种分类), 箱线图如图1C所示。双歧杆菌属的5个种中, 种内最小cANI值高于95%的种有青春双歧杆菌、动物双歧杆菌、短双歧杆菌和长双歧杆菌, 其中最高的为青春双歧杆菌, 种内最小cANI值为98.8%; 与

部分乳酸杆菌菌种的情况相似, 两歧双歧杆菌的种内 cANI 最小值仅为 83.3%, 存在菌株分类不准确的情况. 同样进行“分类学检查”验证, 发现两歧双歧杆菌种内的 1 个基因组与短双歧杆菌最为相似(网络版附表 2). 随后将这 1 个基因组的分类信息更改为短双歧杆菌, 重新计算 cANI, 结果显示, 两歧双歧杆菌的菌株最小 cANI 为 98.1%, 符合同一菌种的鉴定阈值. 经校正后的 5 种双歧杆菌的 cANI 结果箱线图如图 1D 所示, 5 种双歧杆菌的种内 cANI 值分布范围小于 16 种乳酸杆菌, 分布在 94.9%~98.1% 之间, 但相差仍然达到 3.2%, 同样建议不同双歧杆菌菌种在进行鉴定时, 采用不同的 cANI 阈值.

2.4 乳酸杆菌和双歧杆菌菌株间差异功能基因分析

不同物种益生菌具有不同的功能, 而同一物种益生菌中不同菌株的功能是有差异的, 导致这种差异的原因主要是不同菌株基因组中功能基因种类和数目的差异. 本研究选择基因组完成图多于 10 个的物种, 即 4 种乳酸杆菌(鼠李糖乳酪杆菌 32 株、副干酪乳酪杆菌 47 株、罗伊氏粘液乳杆菌 20 株、植物乳植杆菌 148 株)和 3 种双歧杆菌(短双歧杆菌 45 株、两歧双歧杆菌 12 株、长双歧杆菌 56 株)进行了泛基因组比对分析, 以解析不同物种中不同菌株的差异基因和功能的差异. 针对上述 7 种乳酸杆菌和双歧杆菌, 同样采用 Roary 软件进行同一物种不同菌株之间的比较分析, 分别获得每一株益生菌的特有基因、非必需基因和核心基因组, 结果见表 2 和网络版附表 3~9. 鼠李糖乳酪杆菌中 32 株菌的 cANI 平均值为 98.8%, 核心基因组为 1294 个基因, 每个菌株特有基因在 1~311 个之间; 副干酪乳酪杆菌中 47 株菌的 cANI 平均值为 99.3%, 核心基因组为 1070 个基因, 每个菌株特有基因在 1~221 个之间; 罗伊氏粘液乳杆菌中 20 株菌的 cANI 平均值为 97.8%, 核心基因组为 887 个基因, 每个菌株特有基因在 2~436 个之间; 植物乳植杆菌中 148 株菌的 cANI 平均值为 99.5%, 核心基因组为 706 个基因, 每个菌株特有基因在 1~289 个之间. 在双歧杆菌属中, 短双歧杆菌 45 株菌的 cANI 平均值为 99.1%, 核心基因组的基因数目为 1021 个, 每个菌株特有基因在 1~205 个之间; 两歧双歧杆菌中 12 株菌的 cANI 平均值为 99.3%, 核心基因组为 1181 个基因, 每个菌株特有基因在 1~242 个之间; 长双歧杆菌中 56 株菌的 cANI 平均值为 98.5%, 核心基因组为 687 个基因,

每个菌株特有基因在 1~213 个之间. 总体上看, 4 种乳酸杆菌的 cANI 平均值在 97.8%~99.5% 之间, 不同种的乳酸杆菌中含有菌株特有基因数目最多的是罗伊氏粘液乳杆菌的菌株 LL7, 含有菌株特有基因数目最少的有 2 株副干酪乳酪杆菌, 3 株鼠李糖乳酪杆菌和 4 株植物乳植杆菌; 3 种双歧杆菌的 cANI 平均值在 98.5%~99.3% 之间, 不同种的双歧杆菌中含有菌株特有基因数目最多的是两歧双歧杆菌的菌株 JCM7004, 含有菌株特有基因数目最少的有 2 株两歧双歧杆菌、5 株长双歧杆菌和 2 株短双歧杆菌.

为了深入解析同一物种不同菌株之间的进化关系, 本研究构建了以核心基因组为基础的不同物种菌株的进化树, 以便更加深入地解析益生菌的功能. 进化分析发现, 鼠李糖乳酪杆菌在菌株水平可分为 3 个亚群(亚群 A、亚群 B 和亚群 C), 亚群 A 中有 7 株菌的 cANI 值在 99.99% 以上, 其中菌株 SCT-10-10-60 和菌株 JCM1553, 菌株 BIO5326 和菌株 X253 之间的 cANI 值的差异分别为 1 个单核苷酸多态性(single nucleotide polymorphism, SNP)(百万分之三), 基因数目的差异均为 3 个; 亚群 C 中有 7 株菌的 cANI 值在 99.95% 以上, 其中菌株 ATCC53103、菌株 GG 和菌株 GG[ATCC53103]实际上是来源于同一株菌, 即 ATCC53103^[34,35], 但是可能因为保存地点和传代次数的原因, 三个菌株之间的 cANI 差异分别为 1 个 SNP(百万分之三)和 2 个 SNP(百万分之七), 基因数目的差异在 2~6 个之间, 这 3 株菌可能已是不同的菌株; 进一步分析发现, 菌株 ATCC53103 和菌株 4B15, 菌株 GG 和菌株 4B15, 菌株 GG[ATCC53103]和菌株 LR-B1, 菌株 GG[ATCC53103]和菌株 BIO6870, 菌株 LR-B1 和菌株 BIO6870 之间的 cANI 值差异为 1 个 SNP(百万分之三), 基因数目的差异分别为 48 个、42 个、5 个、3 个和 2 个(图 2A). 副干酪乳酪杆菌、罗伊氏粘液乳杆菌和植物乳植杆菌的泛基因组分析结果相似, 副干酪乳酪杆菌有 3 个亚群, 其中亚群 A 中的菌株 A-NA5 和菌株 S-NB 之间的 cANI 值差异为 1 个 SNP(百万分之二), 基因数目的差异为 14 个(网络版附图 2A); 罗伊氏粘液乳杆菌有 2 个亚群, 其中亚群 A 中的菌株 TK-F8A 和菌株 IRT, 菌株 TK-F8A 和菌株 ATG-F4 之间的 cANI 值差异为 10 个 SNP(百万分之十九), 基因数目的差异分别为 58 个和 8 个(网络版附图 2B); 植物乳植杆菌有 2 个亚群, 其中亚群 A 中的菌株 ST-III 和菌株 Y44, 菌株 SRCM100434 和菌株 SRCM100440 之间的 cANI 值

表 2 乳酸杆菌和双歧杆菌不同物种的核心基因组、非必需基因和菌株特有基因

Table 2 Core genes, dispensable genes and strain-specific genes distributed across lactobacilli and *Bifidobacterium* species

物种	核心基因	非必需基因	菌株特有基因
乳酸杆菌			
鼠李糖乳酪杆菌	1294	1272~1676	1~311
罗伊氏粘液乳杆菌	887	932~1334	2~436
副干酪乳酪杆菌	1070	1519~1975	1~221
植物乳植菌	706	1935~2772	1~289
双歧杆菌			
短双歧杆菌	1021	806~1218	1~205
两歧双歧杆菌	1181	557~765	1~242
长双歧杆菌	687	1050~1826	1~213

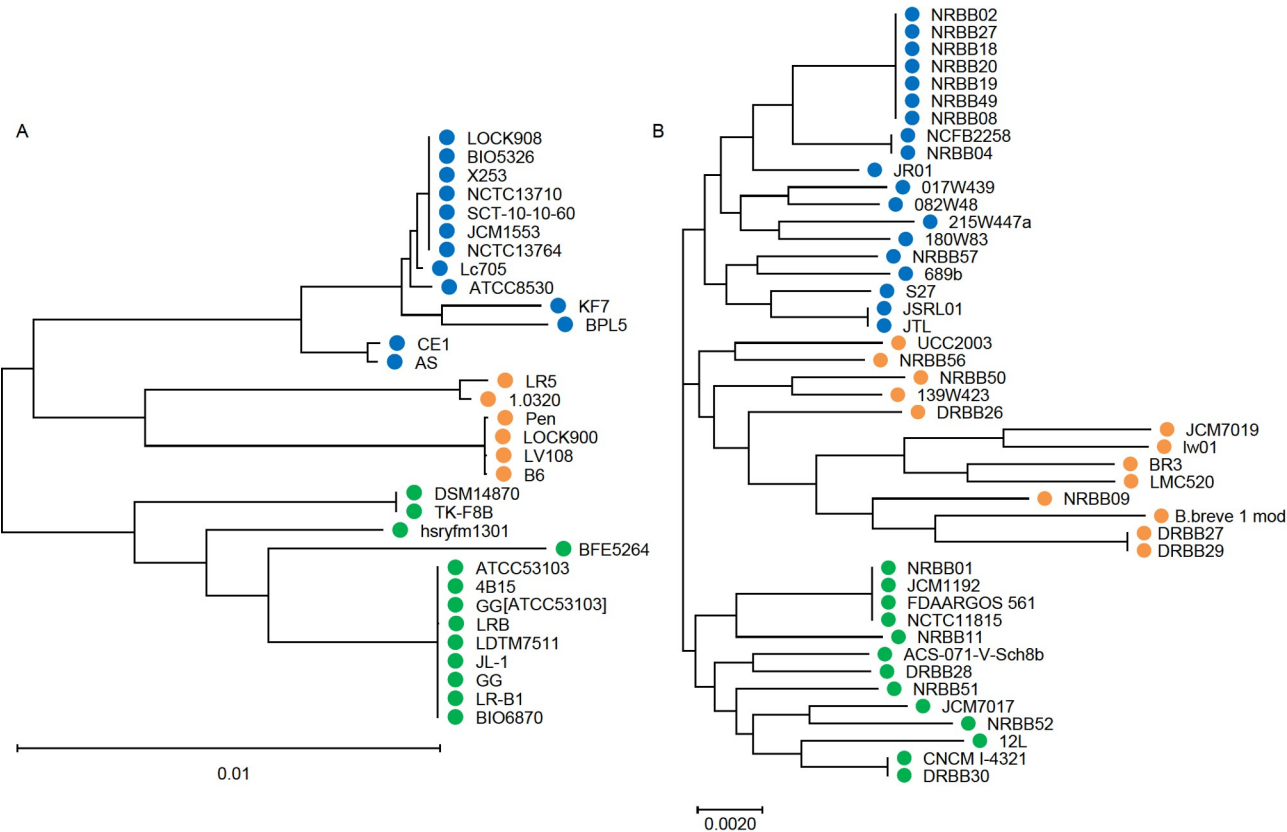


图 2 基于基因组完成图单拷贝核心基因的鼠李糖乳酪杆菌(A)和短双歧杆菌(B)的系统发育树。亚群用不同颜色的圆圈表示, 亚群A为蓝色, 亚群B为橙色, 亚群C为绿色

Figure 2 Phylogenetic tree based on single copy core genes of *L. rhamnosus* (A) and *B. breve* (B). Clade A is blue circle, clade B is orange circle, clade C is green circle

差异为1个SNP(百万分之二), 基因数目的差异均为4个(网络版附图2C)。双歧杆菌与乳酸杆菌进化分析结果相似, 短双歧杆菌在株水平可分为3个亚群(亚群A、

亚群B和亚群C), 亚群A中来源于同一个研究机构的7株菌株之间的cANI值均在99.99%以上, 其中菌株NRBB20和菌株NRBB27之间的cANI值差异只有1个

SNP(百万分之一), 基因数目的差异为1个, 而菌株NRBB27和菌株NRBB49之间的cANI值差异为7个SNP(百万分之八), 基因数目的差异也为1个(图2B). 两歧双歧杆菌有3个亚群, 其中亚群C中的菌株NCTC13001和菌株JCM1255之间的cANI值差异为20个SNP(百万分之二十五), 基因数目的差异为1个(网络版附图3A). 长双歧杆菌也有3个亚群, 其中亚群C中的菌株JCM11340和菌株JCM11342之间的cANI值差异为3个SNP(百万分之四), 基因数目的差异为9个(网络版附图3B). 总体上看, 乳酸杆菌最近的亚群差异基因数目为2~58个, 双歧杆菌最近的亚群差异基因数目1~9个, 相同种内菌株的cANI值越高, 差异基因数目越少, 但当核心基因组cANI>99.99%时, 差异基因数目与cANI无明显相关性. 此外, 本研究还对菌株特有基因和差异基因功能进行了分析, 发现不同菌株之间差异基因的功能各不相同, 再次表明乳酸杆菌和双歧杆菌相同种内不同菌株的功能差异大.

2.5 基于基因组的乳酸杆菌和双歧杆菌安全性评估

益生菌的安全性主要体现在潜在的致病性和耐药性. 本研究使用DIAMOND和RGI v5.2.0软件, 针对国家卫生健康委发布的《可用于食品的菌种名单》中的16种乳酸杆菌和5种双歧杆菌的3780个基因组(包含基因组草图和完成图), 进行了毒力相关基因和耐药基因分析. 将乳酸杆菌和双歧杆菌的菌株与VFDB数据库进行比对, 结果显示有5株菌含有毒力相关基因(表3), 其中乳酸杆菌中只有嗜酸乳杆菌中的菌株P2含有二型分泌系统毒力相关基因*orgC*; 双歧杆菌属中有4株菌含有毒力相关基因, 其中长双歧杆菌BIOML-A353含有血红素转运系统相关毒力相关基因*chu*和*shu*, 菌株

BIOML-A76含有侵入相关毒力相关基因*neuE*; 青春双歧杆菌中的菌株BIOML-A63含有*chu*和*shu*, 三型分泌系统毒力相关基因*gsp*, 侵入相关毒力相关基因*neu*和*kps*, 毒素相关基因*sen*, 菌株BIOML-A79含有毒素相关基因*ast*和*east1*. 通过查阅美国食品药品监督管理局一般公认安全(Generally Recognized as Safe, GARS)清单、欧洲食品安全局安全资格认证(Qualified Presumption of Safety, QPS)清单、产品成分表等信息, 上述含有毒力相关基因的5株菌并未用于食品生产, 且全部分离自健康人体肠道, 这些毒力相关基因可能是从其他菌株获得的. 此外, 菌株BIOML-A79含有耐热性肠毒素基因, 提示即使是在可用于食品名录上的益生菌种类, 也可能具有毒力相关基因, 如果发现携带有毒力相关基因的菌株, 需要用表型或动物实验进行进一步验证.

针对益生菌的耐药性, 本研究采用CARD和ResFinder数据库对16种乳酸杆菌及5种双歧杆菌的3780株菌的基因组进行了潜在的耐药基因分析. CARD数据库预测出50种耐药基因, 涉及13类抗菌药物. 在2247株乳酸杆菌中, 有360株菌含有耐药基因, 分布在副干酪乳酪杆菌、鼠李糖乳酪杆菌、植物乳植杆菌等10个种中, 最主要的为林可酰胺类耐药基因(99株)、四环素耐药基因(85株)和大环内酯类耐药基因(61株); 在1533株双歧杆菌中, 长双歧杆菌、青春双歧杆菌、短双歧杆菌和两歧双歧杆菌4个种的589株基因组含有耐药基因, 其中青春双歧杆菌中有324株菌(99.69%)含有利福霉素耐药基因, 两歧双歧杆菌中有120株菌(97.56%)含有莫匹罗星耐药基因(表4). ResFinder数据库预测出38种耐药基因, 涉及10类抗菌药物. 乳酸杆菌中, 有994株菌含有耐药基因, 分布在干酪乳酪杆菌、鼠李糖乳酪杆菌、植物乳植杆菌等11个种中, 最

表 3 乳酸杆菌和双歧杆菌的毒力基因分布
Table 3 Distribution of virulence genes of lactobacilli and Bifidobacterium strains

菌种	菌株	毒力基因
嗜酸乳杆菌	P2	<i>orgC</i>
长双歧杆菌	BIOML-A353	<i>chuA</i> , <i>chuS</i> , <i>shuA</i> , <i>shuS</i>
	BIOML-A76	<i>neuE</i>
青春双歧杆菌	BIOML-A63	<i>chu(A/S/T/U/V/W/X/Y)</i> , <i>gsp(D/K/L)</i> , <i>kps(C/D/E/F/M/S/T/U)</i> , <i>neu(A/B/C/D/E)</i> , <i>senB</i> , <i>shu(A/S/T/U/V/X)</i>
	BIOML-A79	<i>astA</i> , <i>east1</i>

表 4 乳酸杆菌和双歧杆菌的耐药基因分布(CARD数据库)

Table 4 Distribution of antibiotic resistance genes of lactobacilli and *Bifidobacterium* strains (CARD database)

菌种	抗菌药物
副干酪乳酪杆菌(<i>n</i> =242)	氟喹诺酮类(0.83%)、氨基糖苷类(0.41%)、四环素(0.41%)
鼠李糖乳酪杆菌(<i>n</i> =226)	四环素(1.33%)、单壳聚糖(0.44%)
植物乳植杆菌(<i>n</i> =599)	氨基糖苷类(1.01%)、二氨基嘧啶(0.17%)、糖肽(0.17%)、四环素(0.17%)
卷曲乳杆菌(<i>n</i> =156)	林可酰胺类(14.10%)、链霉素(10.89%)、大环内酯类(10.26%)、氨基糖苷类(3.85%)、氯霉素(0.64%)
德氏乳杆菌(<i>n</i> =103)	单壳聚糖(1.94%)、四环素(0.97%)
约氏乳杆菌(<i>n</i> =60)	大环内酯类(25.00%)、林可酰胺类(13.33%)、链球菌素(13.33%)、四环素(11.67%)、氨基糖苷类(3.33%)、核苷(1.67%)、氯霉素(1.67%)
唾液联合乳杆菌(<i>n</i> =193)	四环素(33.16%)、林可酰胺类(19.69%)、氨基糖苷类(13.99%)、大环内酯类(8.81%)、氯霉素(5.18%)、单壳聚糖(0.52%)、链霉素(0.52%)
弯曲广布乳杆菌(<i>n</i> =24)	林可酰胺类(4.17%)、四环素(4.17%)
清酒广布乳杆菌(<i>n</i> =56)	氨基糖苷类(3.57%)、四环素(1.79%)
罗伊氏粘液乳杆菌(<i>n</i> =208)	林可酰胺类(14.42%)、大环内酯类(6.25%)、氨基糖苷类(6.25%)、四环素(2.88%)、氯霉素(2.40%)、链霉素(2.40%)、二氨基嘧啶(0.48%)
青春双歧杆菌(<i>n</i> =325)	利福霉素(99.69%)、四环素(4.31%)、二氨基嘧啶(0.92%)、大环内酯类(0.92%)、氨基糖苷类(0.31%)、氯霉素(0.31%)
两歧双歧杆菌(<i>n</i> =123)	莫匹罗星(97.56%)、四环素(4.88%)、氟喹诺酮类(1.63%)
短双歧杆菌(<i>n</i> =123)	四环素(20.33%)、氨基糖苷类(7.32%)、大环内酯类(4.07%)、核苷(3.25%)、二氨基嘧啶(2.44%)
长双歧杆菌(<i>n</i> =871)	四环素(3.56%)、二氨基嘧啶(1.84%)、氨基糖苷类(0.80%)、氟喹诺酮类(0.80%)、大环内酯类(0.57%)、单壳聚糖(0.11%)、核苷(0.11%)、利福霉素(0.11%)

主要的为消毒剂耐药基因(575株)、四环素耐药基因(129株)和林可酰胺类耐药基因(100株); 双歧杆菌属的5个种的365株基因组含有耐药基因, 最主要的耐药基因表型为大环内酯类耐药基因(173株)和四环素抗性基因(161株)(表5)。上述菌株中, 包含长双歧杆菌BORI、两歧双歧杆菌BGN4、植物乳植杆菌299v等已经批准用于食品生产的菌株, 虽然检测到了耐药基因, 只是表明这些菌株具有潜在的耐药性, 需进一步实验验证是否有耐药表型, 构成食品安全问题。

3 讨论

乳酸杆菌和双歧杆菌使用历史悠久, 是当今应用最为广泛的益生菌菌种。综合基因组系统发育分析、氨基酸序列一致性、簇特异性基因、表型特征等信息, 国际上对乳酸杆菌和双歧杆菌的分类进行了修正, 我国也于2022年8月发布了《可用于食品的菌种名单》更新版。本研究分析了更新版菌种名单中涉及的16种乳酸杆菌和5种双歧杆菌不同种中菌株的cANI值, 对嗜酸乳杆菌、弯曲广布乳杆菌、植物乳植杆菌、罗

伊氏粘液乳杆菌、干酪乳酪杆菌和两歧双歧杆菌等6个菌种中的26个菌株分类进行了校正, 校正后不同菌种的最小cANI值各不相同, 分布在93.6%~99.6%之间。如果以95%作为所有益生菌种水平鉴定的阈值, 对于嗜酸乳杆菌(种内最小cANI值为99.6%)、卷曲乳杆菌(种内最小cANI值为97.3%)、发酵粘液乳杆菌(种内最小cANI值为98.1%)等核心基因平均核苷酸相似性较高或清酒广布乳杆菌(种内最小cANI值为93.6%)核心基因平均核苷酸相似性较低的菌种来说, 均可能造成错误鉴定。因此, 建议对不同益生菌菌种设定不同的种水平鉴定cANI阈值, 即种特异性核心基因组平均核苷酸相似性, 以提高益生菌种水平分类鉴定的准确性。

益生菌的有益功能具有菌株特异性, 不同菌株含有的基因各不相同, 例如, 鼠李糖乳酪杆菌的功能基因簇*spaCBA-srtC1*, 其编码的菌毛与肠道定植、调节人体免疫水平、抵抗致病菌等多种功能密切相关, 仅有27%的鼠李糖乳酪杆菌菌株含有该功能基因簇^[19]; 植物乳植杆菌的*ltaS*, *tag*, *dltX*, *MPE*等12个磷壁酸相关功能基因, 与宿主的免疫调节、抗炎等功能密切相关^[20], 在599个基因组中仅有两株含有上述12个基因

表 5 乳酸杆菌和双歧杆菌的耐药基因分布(ResFinder数据库)
Table 5 Distribution of antibiotic resistance genes of lactobacilli and Bifidobacterium strains (ResFinder database)

菌种	抗菌药物
干酪乳酪杆菌(n=18)	消毒剂(11.11%)
副干酪乳酪杆菌(n=242)	四环素(8.41%)、消毒剂(1.65%)、喹诺酮类(0.83%)
鼠李糖乳酪杆菌(n=226)	消毒剂(4.42%)、四环素(1.33%)、β-内酰胺类(0.44%)
植物乳植杆菌(n=599)	消毒剂(91.82%)、氨基糖苷类(0.33%)、四环素(0.17%)、甲氧苄啶(0.17%)
卷曲乳杆菌(n=156)	链霉素(21.15%)、林可酰胺类(14.10%)、大环内酯类(10.26%)、四环素(5.77%)、氯霉素(0.64%)
德氏乳杆菌(n=103)	β-内酰胺类(1.94%)、四环素(0.97%)
约氏乳杆菌(n=60)	链霉素(26.67%)、大环内酯类(21.67%)、四环素(21.67%)、林可酰胺类(13.33%)、氨基糖苷类(3.33%)、氯霉素(1.67%)
唾液联合乳杆菌(n=193)	四环素(38.34%)、林可酰胺类(20.21%)、大环内酯类(11.92%)、氨基糖苷类(8.81%)、氯霉素(5.18%)、β-内酰胺类(0.52%)、链霉素(0.52%)
弯曲广布乳杆菌(n=24)	消毒剂(4.17%)、林可酰胺类(4.17%)、四环素(4.17%)
清酒广布乳杆菌(n=56)	消毒剂(14.29%)、四环素(1.79%)
罗伊氏粘液乳杆菌(n=208)	林可酰胺类(14.42%)、四环素(12.02%)、氨基糖苷类(9.62%)、大环内酯类(7.21%)、氯霉素(2.88%)、链霉素(2.88%)、甲氧苄啶(0.48%)
青春双歧杆菌(n=325)	四环素(4.92%)、大环内酯类(4.00%)、氯霉素(0.31%)
动物双歧杆菌(n=91)	林可酰胺类(3.30%)
两歧双歧杆菌(n=123)	大环内酯类(26.02%)、四环素(20.33%)、喹诺酮类(1.63%)
短双歧杆菌(n=123)	大环内酯类(34.15%)、四环素(34.15%)、氨基糖苷类(7.32%)
长双歧杆菌(n=871)	大环内酯类(9.87%)、四环素(8.96%)、氨基糖苷类(0.92%)、喹诺酮类(0.69%)、β-内酰胺类(0.11%)、林可酰胺类(0.11%)

完整的功能基因通路。本研究针对4种乳酸杆菌和3种双歧杆菌的菌株完成图级别基因组进行的比较分析,结果显示,乳酸杆菌中不同物种的菌株特有基因在1~436个之间,双歧杆菌属中不同物种的菌株特有基因数量在1~242个之间,且功能各不相同,菌株间功能基因差异明显。

益生菌的菌株功能特异性需要在菌株水平对益生菌菌株进行鉴定,基于菌株的差异基因数量或基于核心基因的突变(SNP)数量来鉴定菌株可能是一种菌株鉴定的有效手段^[36]。有研究已在尝试使用SNP进行大肠杆菌菌株的鉴定,但仍未形成统一的标准^[37,38]。本研究针对4种乳酸杆菌和3种双歧杆菌属的菌株进行了核心基因SNP数量和基因组差异基因分析,在菌株核心基因组cANI值大于99.99%时,乳酸杆菌中不同物种的菌株差异基因数目在2~58个,双歧杆菌属中不同物种的菌株差异基因数目在1~9个之间。当cANI值大于99.99%时,乳酸杆菌和双歧

杆菌菌株核心基因SNP数量平均小于100个,且不同基因突变速率不同,使得以核心基因SNP数量作为菌株鉴定的方法难度较大。而差异基因数目与益生菌的核心特性-益生功能直接相关,且相对易于分析,因此,建议通过差异功能基因来进行益生菌菌株水平的鉴定。

此外,本研究的毒力相关基因和耐药分析结果显示,乳酸杆菌和双歧杆菌中的食品生产用菌株均未发现毒力相关基因,但这些常用益生菌来源的菌种中部分菌株(分离自健康人肠道)是存在毒力相关基因的;耐药基因分析结果显示,乳酸杆菌和双歧杆菌部分菌株含有多种耐药基因,其中超过90%的青春双歧杆菌、两歧双歧杆菌和植物乳植杆菌菌株分别携带有利福霉素、莫匹罗星和消毒剂类耐药基因。上述结果表明,即使是乳酸杆菌和双歧杆菌这种长期用于食品生产的菌种,部分菌株具有潜在的毒性和耐药性,应对每株菌进行严谨的验证。

致谢 感谢山东凤凰生物科技股份有限公司单宝龙、任宝涛和王静, 中科宜康(北京)生物科技股份有限公司, 中科微智(北京)生物科技股份有限公司对本文研究工作的大力支持。

参考文献

- 1 Food and Agricultural Organization of the United Nations and World Health Organization. Probiotics in Food: Health and Nutritional Properties and Guidelines for Evaluation. Rome: World Health Organization: Food and Agriculture Organization of the United Nations, 2006
- 2 Hill C, Guarner F, Reid G, et al. The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic. *Nat Rev Gastroenterol Hepatol*, 2014, 11: 506–514
- 3 Ren C, Faas M M, de Vos P. Disease managing capacities and mechanisms of host effects of lactic acid bacteria. *Crit Rev Food Sci Nutr*, 2021, 61: 1365–1393
- 4 Koutsoumanis K, Allende A, Alvarez-Ordóñez A, et al. Update of the list of QPS-recommended microbiological agents intentionally added to food or feed as notified to EFSA 16: suitability of taxonomic units notified to EFSA until March 2022. *EFSA J*, 2022, 20: e07408
- 5 Membership of the Pharmacopoeia Commission of the People's Republic of China. Pharmacopoeia of the People's Republic of China (2020) Volume III (in Chinese). Beijing: China Medical Science and Technology Press, 2020 [国家药典委员会. 中华人民共和国药典3部. 北京: 中国医药科技出版社, 2020]
- 6 Cleenwerck I, De Vos P. Polyphasic taxonomy of acetic acid bacteria: an overview of the currently applied methodology. *Int J Food Microbiol*, 2008, 125: 2–14
- 7 Kämpfer P, Glaeser S P. Prokaryotic taxonomy in the sequencing era - the polyphasic approach revisited. *Environ Microbiol*, 2012, 14: 291–317
- 8 Deng M K, Sun Y, Han W Q. Methods of identification of bacteria (in Chinese). *Prog Biomed Eng*, 2014, 35: 5 [邓梅葵, 孙迎, 韩雯晴. 细菌鉴定方法. 生物医学工程学进展, 2014, 35: 5]
- 9 Schleifer K H. Classification of Bacteria and Archaea: past, present and future. *Syst Appl Microbiol*, 2009, 32: 533–542
- 10 Mahato N K, Gupta V, Singh P, et al. Microbial taxonomy in the era of OMICS: application of DNA sequences, computational tools and techniques. *Antonie van Leeuwenhoek*, 2017, 110: 1357–1371
- 11 Zhou J L. Research on the identification, classification and evolutionary variation of bacteria based on whole genome (in Chinese). Dissertation for Master's Degree. Beijing: Academy of Military Sciences, 2020 [周江林. 基于全基因组的细菌鉴定分类与进化变异研究. 硕士学位论文. 北京: 军事科学院, 2020]
- 12 Kämpfer P. Systematics of prokaryotes: the state of the art. *Antonie van Leeuwenhoek*, 2012, 101: 3–11
- 13 Chun J, Rainey F A. Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. *Int J Syst Evol Microbiol*, 2014, 64: 316–324
- 14 Kim M, Oh H S, Park S C, et al. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol*, 2014, 64: 346–351
- 15 Wittouck S, Wuyts S, Meehan C J, et al. A genome-based species taxonomy of the *Lactobacillus* genus complex. *mSystems*, 2019, 4: e00264-19
- 16 Zheng J, Wittouck S, Salvetti E, et al. A taxonomic note on the genus *Lactobacillus*: description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. *Int J Syst Evol Microbiol*, 2020, 70: 2782–2858
- 17 Gupta R S, Patel S, Saini N, et al. Robust demarcation of 17 distinct *Bacillus* species clades, proposed as novel *Bacillaceae* genera, by phylogenomics and comparative genomic analyses: description of *Robertmurraya kyonggiensis* sp. nov. and proposal for an emended genus *Bacillus* limiting it only to the members of the Subtilis and Cereus clades of species. *Int J Syst Evol Microbiol*, 2020, 70: 5753–5798
- 18 Tettelin H, Massignani V, Cieslewicz M J, et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proc Natl Acad Sci USA*, 2005, 102: 13950–13955
- 19 Rossi F, Amadoro C, Pallotta M L, et al. Variability of genetic characters associated with probiotic functions in *Lactocaseibacillus* species. *Microorganisms*, 2022, 10: 1023
- 20 Liu Y W, Liu W H, Wu C C, et al. Psychotropic effects of *Lactobacillus plantarum* PS128 in early life-stressed and naïve adult mice. *Brain Res*, 2016, 1631: 1–12
- 21 Roe A L, Boyte M E, Elkins C A, et al. Considerations for determining safety of probiotics: A USP perspective. *Regul Toxicol Pharmacol*, 2022,

136: 105266

- 22 Probiotics Society of the Chinese Institute of Food Science and Technology. Scientific Consensus on Probiotics (2020) (in Chinese). J Chin Inst Food Sci Technol, 2020, 20: 5 [中国食品科学技术学会益生菌分会. 益生菌的科学共识(2020年版). 中国食品学报, 2020, 20: 5]
- 23 European Food Safety Authority (EFSA). EFSA statement on the requirements for whole genome sequence analysis of microorganisms intentionally used in the food chain. EFSA J, 2021, 19: e06506
- 24 Parks D H, Chuvochina M, Chaumeil P A, et al. A complete domain-to-species taxonomy for Bacteria and Archaea. Nat Biotechnol, 2020, 38: 1079–1086
- 25 Sanford R A, Lloyd K G, Konstantinidis K T, et al. Microbial taxonomy run amok. Trends Microbiol, 2021, 29: 394–404
- 26 Albert K, Rani A, Sela D A. Comparative Pangenomics of the mammalian gut commensal *Bifidobacterium longum*. Microorganisms, 2019, 8: 7
- 27 Carpi F M, Coman M M, Silvi S, et al. Comprehensive pan-genome analysis of *Lactiplantibacillus plantarum* complete genomes. J Appl Microbiol, 2022, 132: 592–604
- 28 Harris H M B, Bourin M J B, Claesson M J, et al. Phylogenomics and comparative genomics of *Lactobacillus salivarius*, a mammalian gut commensal. Microb Genomics, 2017, 3: e000115
- 29 Wang J, Zhang W Y. Phylogenomics analysis of *Lactobacillus paracasei* (in Chinese). Genom Appl Biol, 2022, 41: 1461–1474 [王静, 张文羿. 副干酪乳杆菌系统发育组学分析. 基因组学与应用生物学, 2022, 41: 1461–1474]
- 30 Zhou X, Yang B, Stanton C, et al. Comparative analysis of *Lactobacillus gasseri* from Chinese subjects reveals a new species-level taxa. BMC Genomics, 2020, 21: 119
- 31 Konstantinidis K T, Tiedje J M. Towards a genome-based taxonomy for prokaryotes. J Bacteriol, 2005, 187: 6258–6264
- 32 Chun J, Oren A, Ventosa A, et al. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. Int J Syst Evol Microbiol, 2018, 68: 461–466
- 33 Parte A C, Sardà Carbasse J, Meier-Kolthoff J P, et al. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. Int J Syst Evol Microbiol, 2020, 70: 5607–5612
- 34 Morita H, Toh H, Oshima K, et al. Complete Genome Sequence of the Probiotic *Lactobacillus rhamnosus* ATCC 53103. J Bacteriol, 2009, 191: 7630–7631
- 35 Kankainen M, Paulin L, Tynkkynen S, et al. Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pili containing a human-mucus binding protein. Proc Natl Acad Sci USA, 2009, 106: 17193–17198
- 36 Van Rossum T, Ferretti P, Maistrenko O M, et al. Diversity within species: interpreting strains in microbiomes. Nat Rev Microbiol, 2020, 18: 491–506
- 37 Spira B, de Almeida Toledo R, Maharjan R P, et al. The uncertain consequences of transferring bacterial strains between laboratories-RpoS instability as an example. BMC Microbiol, 2011, 11: 248
- 38 Kong L Y, Eyre D W, Corbeil J, et al. *Clostridium difficile*: investigating transmission patterns between infected and colonized patients using whole genome sequencing. Clin Infect Dis, 2019, 68: 204–209

Identification, safety and functional prediction of food-grade lactobacilli and *Bifidobacterium* based on whole genome sequences

CHEN Nan^{1,2,3}, GUO YaQiong^{1,4}, Lyu Na^{1,2}, Li FengQin⁵, CHEN LiJun^{6,7}, Xiang XueSong^{1,8},
WANG Xin^{1,9}, LIU LieGang^{1,10}, WANG HongWei¹, Zhi FaChao^{1,11}, HE Fang^{1,12}, CHEN Wei¹³,
YANG RuiFu^{1,14} & ZHU BaoLi^{1,2,15,16,17}

1 Committee of Probiotics-Prebiotics and Health, Chinese Nutrition Society, Beijing 100052, China;

2 Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China;

3 MicroBio (Beijing) Medical Technology Co., Ltd., Beijing 100068, China;

4 University of Chinese Academy of Sciences, Beijing 100049, China;

5 National Health Commission Key Laboratory of Food Safety Risk Assessment, China National Center for Food Safety Risk Assessment, Beijing 100021, China

6 National Engineering Research Center of Dairy Health for Maternal and Child, Beijing Sanyuan Foods Co. Ltd., Beijing 100163, China;

7 Beijing Engineering Research Center of Dairy, Beijing Technical Innovation Center of Human Milk Research, Beijing 100163, China;

8 National Institute for Nutrition and Health, Chinese Center for Disease Control and Prevention, Beijing 100050, China;

9 Institute of Food Research, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China;

10 School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China;

11 Guangdong Provincial Key Laboratory of Gastroenterology, Department of Gastroenterology, Institute of Gastroenterology of Guangdong Province, Nanfang Hospital, Southern Medical University, Guangzhou 510515, China;

12 West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu 610041, China;

13 School of Food Science and Technology, Jiangnan University, Wuxi 214122, China;

14 Beijing Institute of Microbiology and Epidemiology, Academy of Military Medicine, Beijing 100071, China;

15 Jinan Microecological Biomedicine Shandong Laboratory, Jinan 250117, China;

16 School of Basic Medicine, Southwest Medical University, Luzhou 646000, China;

17 Chinese National Human Genome Center/SinoGenoMax Co., Ltd., Beijing 100176, China

Probiotics are beneficial microorganisms for human health, including bacteria and fungi. The identification of probiotics has always been a pressing technical issue in the academic community. Traditional taxonomic methods can only identify at the species level, while genomic analysis can identify at the strain or species level. In this study, a whole-genome comparison analysis was performed on 16 strains of lactobacilli and 5 strains of *Bifidobacterium* listed in China's "List of Microorganisms Used in Food". The core genes average nucleotide identity (cANI) was applied to explore the method of reclassification and identification of species. The results showed that the cANI values of lactobacilli and *Bifidobacterium* had species-specificity, with cANI values of lactobacilli ranging from 93.6% to 99.6% and cANI values of *Bifidobacterium* ranging from 94.9% to 98.1%. At the same time, 25 strains of lactobacilli and one strain of *Bifidobacterium* were found to have taxonomic errors. The results of comparative genome analysis of 247 strains of lactobacilli and 113 strains of *Bifidobacterium* showed that the number of different genes between strains of lactobacilli and *Bifidobacterium* could reach up to 436, and the minimum cANI similarity value between strains was 1 single nucleotide polymorphisms (SNPs). The analysis of virulence genes and resistance genes in lactobacilli and *Bifidobacterium* showed that no virulence-related genes were found in all commercial strains, and more than 20% of strains carried different types of resistance genes. In conclusion, the cANI values of different species of lactobacilli and *Bifidobacterium* have significant differences and species specificity, which can be used for the identification of probiotic species, and the differences in gene and SNP numbers between different strains of the same species can be used for strain identification.

probiotic, pangenome, cANI, strain specific genes, SNP, virulence genes, resistance genes

doi: [10.1360/SSV-2023-0043](https://doi.org/10.1360/SSV-2023-0043)