



比较基因组学分析不同来源罗伊氏乳杆菌基因多样性及生境适应性

安晓娜, 李伟程, 于洁, 潘琳, 莫蓝馨, 姚彩青, 张和平*

内蒙古农业大学乳品生物技术与工程教育部重点实验室, 农业部奶制品加工重点实验室, 内蒙古呼和浩特 010018

摘要:【目的】研究 34 株不同来源罗伊氏乳杆菌(*Lactobacillus reuteri*)基因多样性及生境适应性机制, 了解该菌株在肠外生境及肠内生境中适应性的异同, 为 *L. reuteri* 优良菌株的开发提供理论基础。【方法】基于二代测序平台对 11 株源于发酵食品(酸马奶、酸粥)的 *L. reuteri* 进行测序, 并应用比较基因组学将其与发酵食品、酸面团、食草动物 *L. reuteri* 分离株基因组进行比较分析。【结果】分离自酸马奶、酸粥 *L. reuteri* 基因组大小平均为 2.14 Mb, GC 含量平均为 38.77%, 且同种来源分离株系统发育关系距离较近。泛-核心基因集分别包含 7242 个、969 个基因家族, 其中酸马奶分离株特异性基因最多为 459 个。功能注释结果显示不同来源菌株碳水化合物、氨基酸相关基因数量及种类差异较大, 仅在发酵食品和食草动物株中发现抗生素耐药性基因。注释到的碳水化合物活性酶中仅出现在发酵食品与食草动物分离株的为 GH3 (β -葡萄糖苷酶等)和 GH43 (β -木糖苷酶等), 特有的分别为 AA3 (纤维二糖脱氢酶等)和 GH66 (葡萄糖转移酶等)。【结论】不同来源 *L. reuteri* 具有广泛的基因多样性, 与生存环境密切相关。发酵食品分离株具有部分食草动物肠道菌株特性且具有其独特的环境适应性特征, 体现了与宿主有关的环境适应能力, 可加深对食品发酵和肠内环境中细菌生境适应性的理解。

关键词: 罗伊氏乳杆菌, 比较基因组学, 发酵食品, 生境适应性

罗伊氏乳杆菌(*Lactobacillus reuteri*)属专性异型发酵菌种, 能发酵糖产生 CO₂、乳酸、乙酸和乙醇。在人体和动物肠道、发酵食品中广泛存在。

大量研究表明 *L. reuteri* 在人体及动物中具有多种益生特性, 如可长期定殖于肠道中; 调节宿主免疫系统; 分泌广谱的抗菌化合物^[1]; 预防腹泻和

基金项目: 国家自然科学基金(31430066); 内蒙古自然科学基金(2018MS03041); 校优秀青年科学基金(2017XYQ-4)

*通信作者。Tel/Fax: +86-471-4319940; E-mail: hepingdd@vip.sina.com

收稿日期: 2019-07-31; 修回日期: 2019-09-20; 网络出版日期: 2020-03-09

结肠炎^[2]；降低婴儿急性腹痛患病率等^[3]。因此，*L. reuteri* 在功能性食品和医疗行业具有非常高的潜在应用价值。

L. reuteri 已作为标准模型来描述脊椎动物肠道共生体生态及进化机制^[4]。因此脊椎动物肠道内 *L. reuteri* 的生长代谢及进化机制已被广泛研究。Oh 等^[5]应用扩增片段长度多态性和多位点序列分析(MLSA)技术对分离自杂食动物(人,大鼠,小鼠,猪,鸡和火鸡)肠道或体内的 *L. reuteri* 进行了系统发育和群体遗传结构分析。结果显示不同宿主来源的 *L. reuteri* 基因特征不同,并且形成多个系统发育分支。Wegmann 等^[6]应用比较基因组学对来源于猪体内的 *L. reuteri* 进行了系统发育及进化历程分析。结果表明其基因组具有宿主专一性,其中编码细胞表面蛋白的基因具有宿主特异性,这可能是导致形成两类系统发育分支的原因。Zheng 等^[7]应用比较基因组学方法对 4 株面团分离株与 12 株分离自人类、猪、啮齿动物、家禽的 *L. reuteri* 探究其生境适应性机制。实验表明基于核心基因组和基因含量分别构建的系统发育树拓扑结构不同,此外约有 10% 的核心基因处于正向选择压力下,主要与能量转换与碳水化合物代谢有关。

同种菌的比较基因组学研究可以更深入地了解基因的改变、缺失或获得,这些变化会促使菌株在不同生态位中进化及增强环境适应性^[8]。目前为止,可在数据库中获得 120 多株 *L. reuteri* (人类,猪,鸡,大鼠,小鼠,面团和食草动物)的全基因组信息,但分离自发酵食品 *L. reuteri* 全基因组信息甚少。故本文通过比较基因组学和高通量测序结合,将分离自内蒙古地区发酵食品中 11 株 *L. reuteri* 与 NCBI 上 23 株食草动物来源 *L. reuteri*

进行比较,研究其基因多样性及生境适应性,从基因层面揭示 *L. reuteri* 在肠内生境及肠外生境适应能力机制,为 *L. reuteri* 优良菌株的开发提供理论基础。

1 材料和方法

1.1 菌株选取

本研究依托内蒙古农业大学“乳品生物技术与工程”教育部重点实验室乳酸菌菌种资源库(LABCC)中分离自内蒙古自治区传统酸马奶及酸粥的 11 株 *L. reuteri*, 并从 NCBI 网站 GenBank 数据库中获取 16 株源于食草动物^[9] (5 株山羊、4 株绵羊、4 株牛、3 株马肠道分离株)、7 株酸面团的 *L. reuteri* 全基因组信息。

1.2 菌株活化及基因组 DNA 提取

将真空冷冻干燥保藏的 11 株 *L. reuteri* 在脱脂乳中 37 °C 培养 24 h 活化后,在 MRS 液体培养基(CM0359B, OXOID)中继代培养。离心收集菌体后采用 TIANGEN 细菌基因组 DNA 试剂盒提取基因组 DNA,操作步骤参考试剂盒说明书。

1.3 DNA 质控及基因组重测序

完成基因组 DNA 抽提后,采用微量分光光度计测定 DNA 液的浓度,质量合格送交上海美吉生物医药科技有限公司进行 16S rRNA 基因测序。确定为目标菌株后进行全基因组重测序。测序平台为 Illumina HiSeq 2000 平台,选取 150 bp 的 DNA 片段构建 Pair-end 测序文库,高质量数据平均覆盖度在 500X 左右。

1.4 基因组拼接组装

获取的原始数据首先进行质量评估,过滤后运用软件包 SOAPdenovo v1.06 对高质量的 reads 进行

拼接, 并选取合适的 kmer 值对过滤后的数据进行拼接组装和单碱基校正^[10]。以基因组大小、Scaffold 数量、N50 长度、N90 长度、GC 含量值为指标对组装结果进行评估。

1.5 基因注释

将 34 株菌株蛋白序列与 COG^[11] (Clusters of orthologous groups, 直系同源基因簇数据库) 进行 BLASTp (E -value $<1e-10$, identity 值 $\geq 40\%$) 比对, 完成蛋白序列功能注释。

1.6 泛-核心基因集及系统发育分析

首先将 34 株 *L. reuteri* 的基因序列用 Prokka v1.12 软件^[12]进行基因注释, 获得 GFF 格式文件 (用于计算核心基因组和泛基因组)。使用 Roary 软件^[13]计算核心基因组和泛基因组数量。除此之外, 基于核心基因利用 MEGA7 软件构建邻接树 (Neighbor-joining tree, NJ)。参数 Bootstrap 值设为 1000。使用 Venny 2.1 软件绘制了韦恩图, 展现宿主特异性基因。使用 KEGG 数据库进行特殊基因所参与的代谢通路注释^[14]。

1.7 碳水化合物活性酶分析

下载并安装 hmmer 软件包^[15]以及下载 CAZyme

(Carbohydrate-Active Enzymes) 数据库; 对基因组序列进行碳水化合物活性酶的注释; 最后, 根据结果在网页上 (<http://www.cazy.org/>) 查询详细的碳水化合物活性酶家族信息。

1.8 数据统计分析

用 t 检验比较两两菌株之间的基因组大小和编码基因数量差异。采用 Mann whitney 检验不同来源菌株功能基因差异。

2 结果和分析

2.1 11 株发酵食品 *L. reuteri* 基因组特征

通过 Illumina HiSeq 高通量测序平台获得了 11 株来源于酸马奶和酸粥的 *L. reuteri* 全基因组信息, 其基因组大小为 1.90–2.95 Mb, 平均为 2.14 Mb; GC 含量大小为 38.69%–38.82%, 平均含量为 38.77% (表 1)。

2.2 泛基因组、核心基因组及宿主特异性基因分析

应用比较基因组学对 34 株不同分离源的 *L. reuteri* 进行了泛基因组和核心基因组注释。泛基因组共包含 7242 个基因家族, 核心基因组包含

表 1. 11 株 *L. reuteri* 基因组拼接质量与基本特征

Table 1. Quality and basic characteristics of genome splicing of 11 *L. reuteri* strains

Isolates	Origin	Scaffolds	N50/bp	N90/bp	Size/Mb	GC mol/%	GenBank accession No.
N1	Koumiss	241	18222	5082	2.92	38.82	VRYI00000000
N2	Koumiss	250	18298	5323	2.95	38.82	VTTF00000000
N3	Koumiss	233	18942	5262	1.91	38.81	VRYJ00000000
N4	Koumiss	237	18192	5410	1.95	38.80	VTTE00000000
N5	Koumiss	241	18978	5278	1.93	38.80	VTTD00000000
N6	Koumiss	190	20619	5674	1.96	38.71	V TTC00000000
N7	Koumiss	206	19569	5629	1.90	38.77	VTTB00000000
Z2	Acid gruel	250	24946	5743	2.04	38.76	VRYK00000000
Z3	Acid gruel	219	26794	5857	2.02	38.73	VRYL00000000
Z5	Acid gruel	262	16008	4348	2.02	38.69	VTTA00000000
Z6	Acid gruel	259	14960	4262	1.95	38.76	VRYM00000000

969 个基因家族。除此之外, 34 株 *L. reuteri* 共包含 4430 个附属基因(图 1-B), 此类基因为泛基因集中去掉核心及稀有基因(存在频率小于 0.05 的基因)后的基因, 可显示菌株基因组之间的特异性。可从图 1-A 看出 34 株 *L. reuteri* 附属基因的存在与缺失。为了解 969 个核心基因所涉及的功能, 进行了 COG 功能基因注释。结果显示约 37.04% 的基因编码与代谢有关的蛋白质, 除了未知功能的基因, 34 株 *L. reuteri* 核心基因组主要与翻译、核糖体结构和生物发生(J)(11.07%)、氨基酸转运和代谢(E)(10.36%)、复制、重组和修复(L)(7.10%)、碳水化合物转运和代谢(G)(6.42%)等相关功能有关(图 2)。

依据来源将 34 株 *L. reuteri* 分为食草动物、酸马奶、酸粥和酸面团 4 组, 并基于 4 类菌株

的核心基因绘制了 Venn 图(图 3)。结果显示酸马奶、酸粥、酸面团和食草动物分离株分别有 459、450、253、404 个特异性基因。经 KEGG 数据库注释后发现酸马奶分离株 459 个特异基因主要参与丙酮酸代谢、氨基糖和核苷酸糖代谢、脂肪酸降解、精氨酸生物合成等代谢通路。酸粥分离株 450 个特异基因主要参与群体感应、泛酸盐和辅酶 A 生物合成、双组分系统等代谢通路有关。酸面团分离株 253 个特异基因主要参与精氨酸生物合成、蛋白运输等代谢通路有关。食草动物分离株 404 个特异性基因主要参与卟啉和叶绿素代谢、叶酸生物合成、阿特拉津降解、氮代谢等代谢通路有关。这些结果表明不同来源的 *L. reuteri* 基因功能较为广泛且具有宿主特异性。

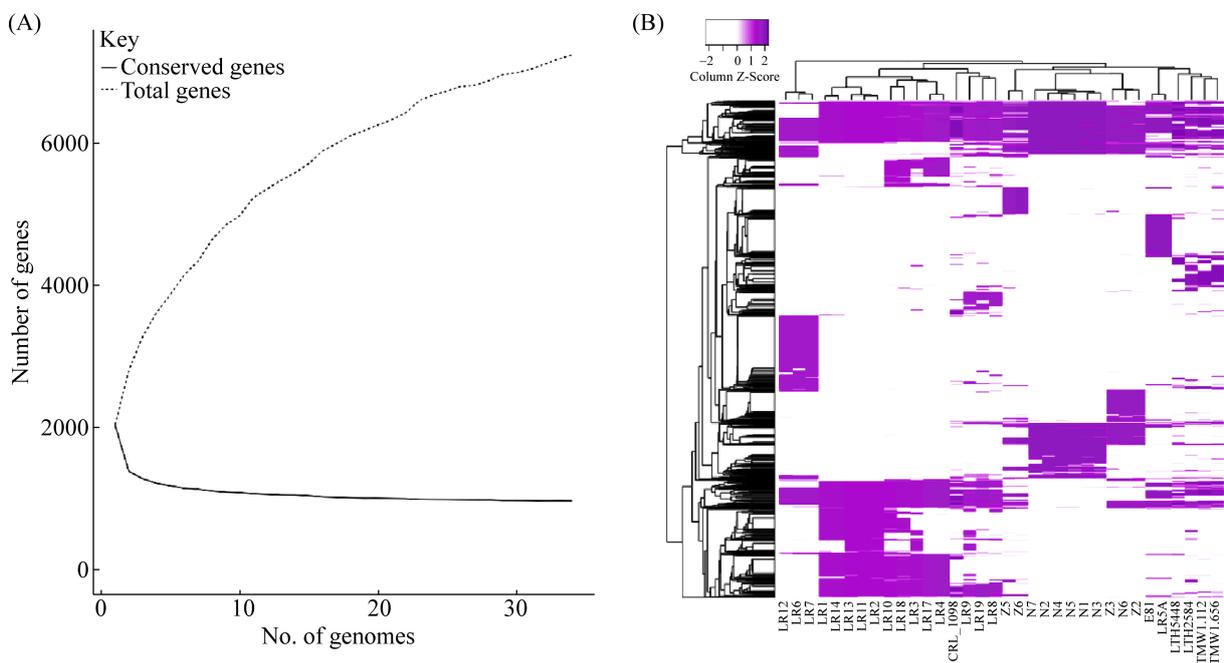


图 1. 泛-核心基因集变化趋势图(A)和附属基因集在 34 株 *L. reuteri* 中的存在与缺失(B)

Figure 1. Trend map of pan-core gene sets (A) and existence and deletion of accessory gene set in 34 *L. reuteri* strains (B).

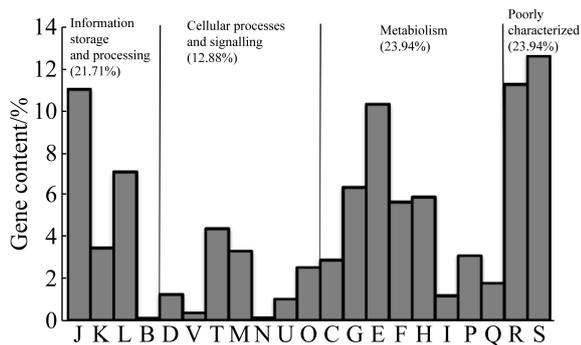


图 2. 核心基因 COG 注释结果

Figure 2. COG annotation results of core gene set. J: translation, ribosomal structure and biogenesis; B: chromatin structure and dynamics; K: transcription; L: replication, recombination and repair; D: cell cycle control, cell division, chromosome partitioning; V: defense mechanisms; T: signal transduction mechanisms; M: cell wall/membrane/envelope biogenesis; Z: cytoskeleton; U: intracellular trafficking, secretion, and vesicular transport; O: posttranslational modification, protein turnover, chaperones; C: energy production and conversion; G: carbohydrate transport and metabolism; E: amino acid transport and metabolism; F: nucleotide transport and metabolism; H: coenzyme transport and metabolism; I: lipid transport and metabolism; P: inorganic ion transport and metabolism; Q: secondary metabolites biosynthesis, transport and catabolism; R: general function prediction only; S: function unknown.

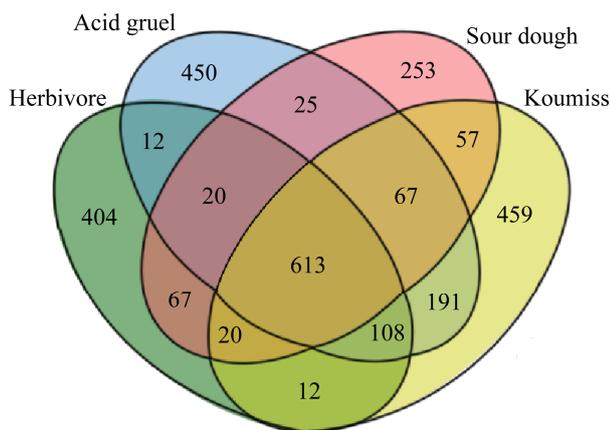


图 3. 基于四类不同来源菌株核心基因构建的 Venn 图
Figure 3. Venn maps based on core genes of four different strains.

2.3 系统发育分析及不同分支特异基因比较

为了确定 34 株 *L. reuteri* 的系统发育关系, 根据核心基因组构建了系统发育树(图 4-A)。结果表明食草动物分离株更靠近根部, 可能是其他分离株的祖先, 且相同来源的菌株距离较近。这些结果表明自然环境不同导致形成不同的系统发育分支。

为了找出各分支的特异基因, 按系统发育关系分为 a、b、c、d 四组, 其中, a 分支包含 7 株酸马奶分离株和 4 株酸粥分离株; b、c 分支分别包含 3 株面团分离株; d 分支包含 13 株食草动物分离株。对其进行两两之间宿主特异基因的比较。基于核心基因组构建了 Venn 图(图 4-B), 从图中可知 a、b、c、d 分别有 288、265、350、531 个特异基因, 共有基因为 653 个, 表明各分支间基因种类差异较大。其中 a 分支中主要与半乳糖代谢、氨基酸生物合成、丙酮酸代谢等代谢途径有关。b 分支主要与丙酮酸代谢、abc 转运蛋白等代谢途径有关。c 分支主要与酵母自噬、群体感应、糖酵解/糖异生、 α -亚麻酸代谢等途径有关。d 分支主要参与卟啉和叶绿素代谢、阿特拉津降解、新生霉素代谢、色氨酸代谢等途径。各来源菌株的代谢途径均能体现各宿主所处环境特征。

2.4 不同分支编码营养代谢物质基因的比较

2.4.1 编码碳水化合物相关基因: 通过对 a、b、c、d 分支中碳水化合物比较发现, a 分支(酸马奶、酸粥)中糖酵解/糖异生途径[15(参与酶的数量)]、磷酸戊糖途径(14)、半乳糖代谢途径(10)、淀粉和蔗糖代谢途径(7)、TCA 循环(6)中基因所编码的酶数量最多(图 5)。

L. reuteri 属于专性异型发酵菌, 能发酵糖产生

CO₂、乳酸、乙酸和乙醇。在糖酵解/糖异生途径中，a 分支编码相关酶的基因数量最多为 15 个，b、c、d 分支分别有 3 个、5 个、6 个，但 a 分支菌株糖酵解/糖异生途径并不完整，缺少磷酸果糖激酶 (EC 2.7.1.11)、己糖激酶 (EC 2.7.1.1)。糖酵解的产物丙酮酸，可被乳酸脱氢酶还原成乳酸。a 分支有 2 个编码 D-乳酸脱氢酶 (1dhD-1 和 1dhD-7)，因此能产生 D 型乳酸。除了乳酸脱氢酶外，还具有 2 个编码丙酮酸脱氢酶 (EC 1.2.4.1) (pdhA_1 和 pdhB_1) 基因，可催化丙酮酸转化生成乙酰 CoA，用于之后的三羧酸循环及磷酸戊糖代谢途径。b 分支中有 2 个编码 D-乳酸脱氢酶基因 (group_338 和 ldhA)；c、d 分支中各有一个分别为 group_337 和 group_3714。以上结果表明不同生存环境导致菌株

利用糖类物质的不同。

编码蛋白质和氨基酸相关基因：a 分支共有 12 类与氨基酸合成代谢相关途径，参与酶类的数量达 66 个。主要有丙氨酸、天冬氨酸和谷氨酸代谢，赖氨酸生物合成，谷胱甘肽代谢等。b 分支共有 4 类 7 个与氨基酸合成代谢途径相关基因：丙氨酸、甘氨酸、丝氨酸和苏氨酸代谢等。c 分支共有 6 类 11 个与氨基酸合成代谢途径相关基因，主要为赖氨酸生物合成、酪氨酸代谢、缬氨酸等；d 分支共有 7 类 20 个与氨基酸合成代谢途径相关基因：组氨酸代谢，赖氨酸生物合成，酪氨酸代谢等。结果显示酸马奶、酸粥分离株氨基酸代谢通路数量与种类最多，表明该分离源菌株具有较强利用氨基酸的能力，可能与其处于丰富营养物质的环境有关。

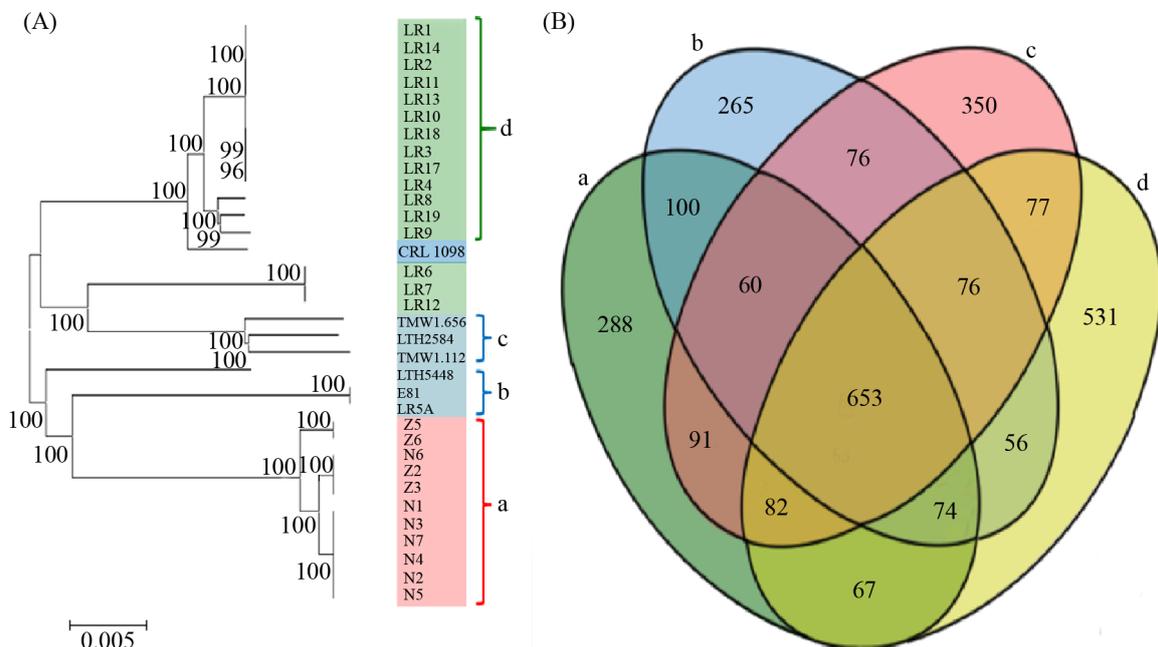


图 4. 34 株 *L. reuteri* 系统发育树(A)及不同系统发育分支共有-特异基因 Venn 图(B)

Figure 4. Phylogenetic relationship of 34 *L. reuteri* strains (A) and Venn map of common-specific genes in different phylogenetic branches (B). a: 7 koumiss isolates, 4 acid gruel isolates; b: 3 dough isolates; c: 3 dough isolates; d: 13 herbivore isolates.

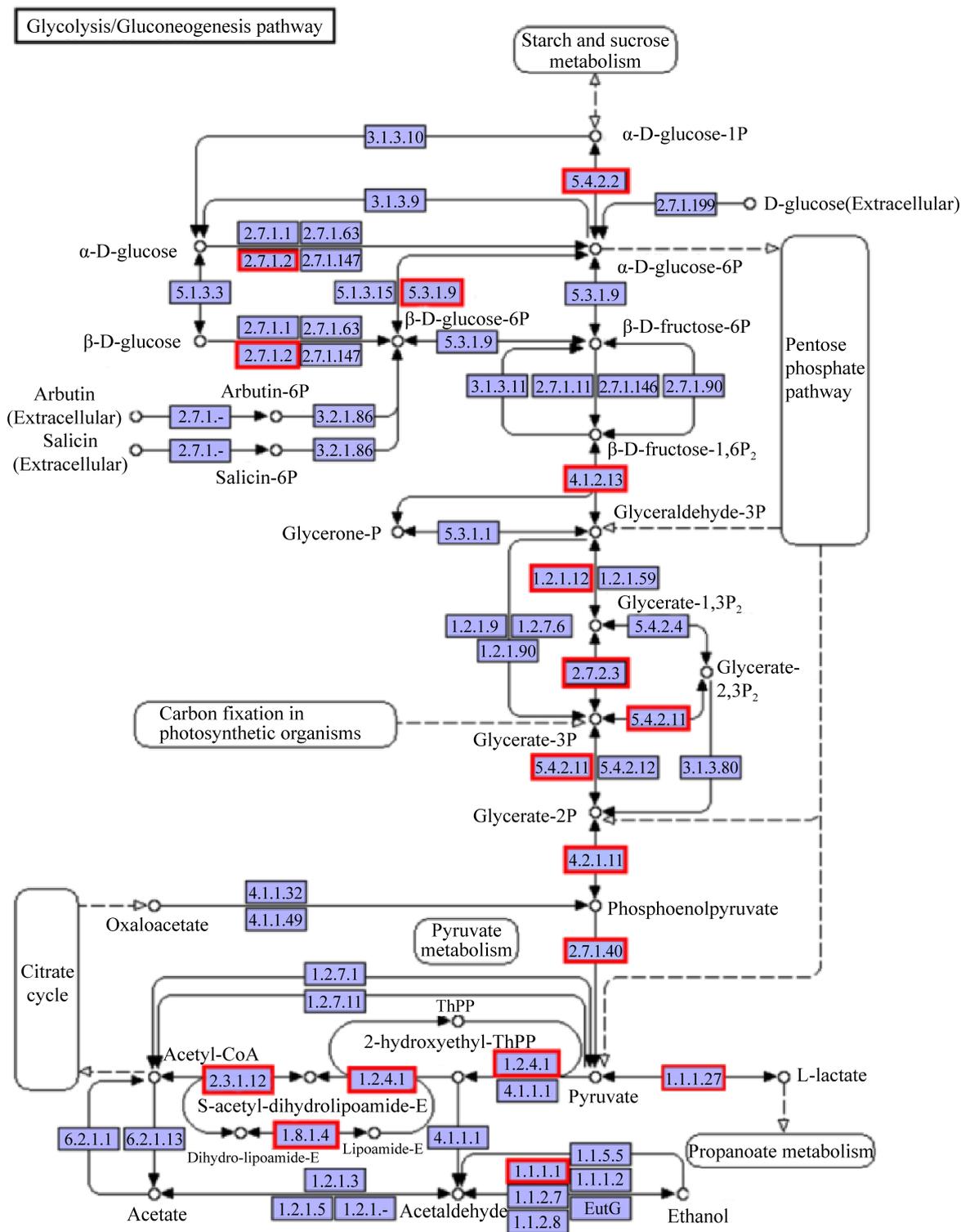


图 5. 糖酵解/糖异生代谢通路图

Figure 5. Glycolysis/gluconeogenesis pathway diagram.

2.4.2 其他代谢途径: 除上述代谢途径外, 34 株 *L. reuteri* 基因组中还具有编码丁酸代谢途径相关的酶, 其中 a 分支相关酶数量最多, 这可能表明酸马奶及酸粥中 *L. reuteri* 具有潜在代谢丁酸的能力。另外, 仅在 a 分支中发现了二甲苯降解、二恶英降解、萘降解、氯代烷烃和氯烯烃降解相关途径的酶。在抗生素耐药性方面, 仅在 a、d 分支中出现了 β -内酰胺抗性基因和万古霉素耐药性基因。

2.4.3 碳水化合物活性酶分析: 为在基因组水平上了解不同来源 *L. reuteri* 对碳水化合物的利用能力, 本研究通过碳水化合物活性酶数据库(CAZy)分析了 34 株 *L. reuteri* 的碳水化合物活性酶功能基因。34 株 *L. reuteri* 共注释到五大类碳水化合物活性酶的 51 个小类基因家族, 主要以 GHs 和 GTs

酶类的含量最为丰富(图 6)。其中有 20 个 GH 家族(20/152), GH3 和 GH43 只出现在酸马奶、酸粥和食草动物分离株中, GH42、GH65、GH68 在酸马奶、酸粥分离株中较多, GH23、GH25、GH66、GH73、GH109 在食草动物分离株中较多。仅在酸马奶、酸粥和食草动物分离株中出现的为 GH3、GH43。酸面团分离株中 GH 家族数量及种类都不及另两类分离株多。

GT 家族中共有 9 个家族(9/105), GT8、GT28 在酸马奶、酸粥分离株中数量较多; GT8 在酸面团分离株中较多; GT2、GT4、GT19、GT51 在食草动物组中较多。GT2 和 GT4 酶类(与蔗糖等二糖以及脂多糖、纤维素、壳聚糖的合成相关)在 34 株 *L. reuteri* 中含量较丰富, 可能与菌株所处环境有关, 并且表明具有较强的糖合成能力。

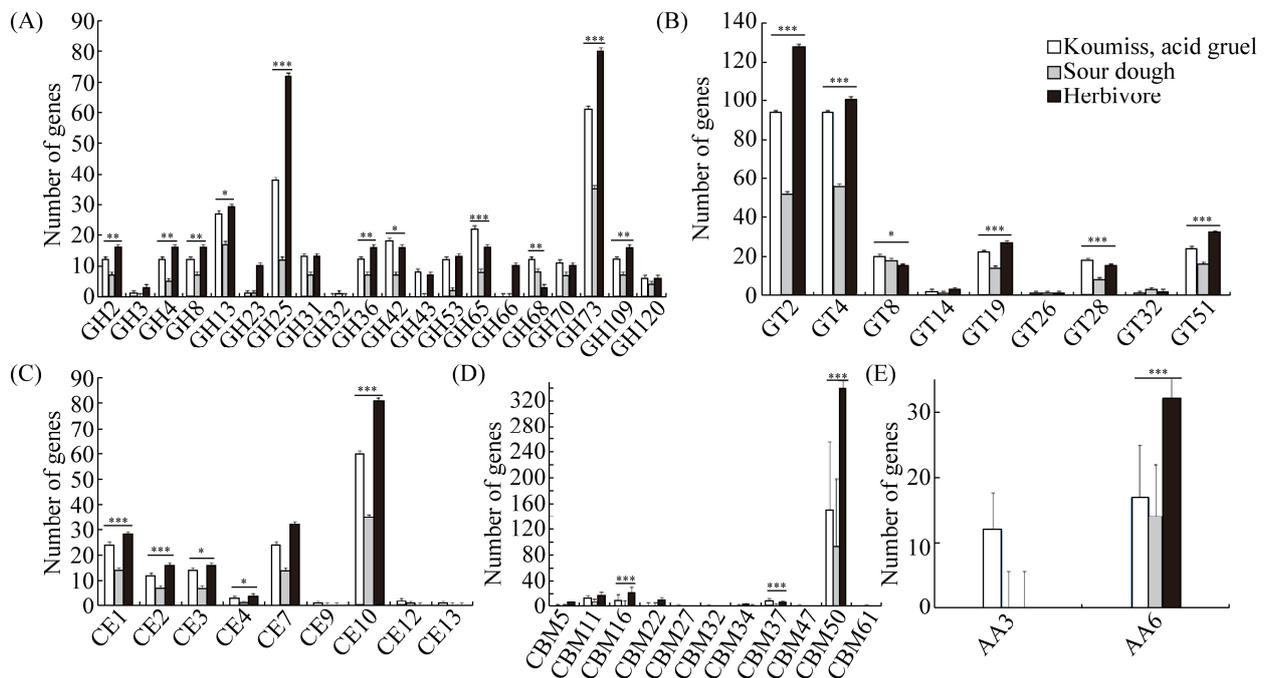


图 6. 碳水化合物活性酶注释结果

Figure 6. Annotation results of carbohydrate-active enzymes. A: GH family; B: GT family; C: CE family; D: CBM family; E: AA family. *: $P < 0.05$; **: $P < 0.01$; ***: $P < 0.001$.

CE 家族中共有 9 个家族(9/16), 比较后发现 CE9 和 CE13 是酸马奶、酸粥分离株中特有的; 食草动物分离株中 CE10 数量较多。CBM 家族中注释到 11 个家族(11/83), 其中 CBM27、CBM32、CBM47 是酸马奶、酸粥分离株所独有的, 而 CBM50 在食草动物分离株中最多, 平均数量为 21 个, 且 CBM5、CBM22 只出现在食草动物组中; CBM16 和 CBM37 只在酸马奶、酸粥和食草动物分离株中存在。AA 家族共有 2 个家族(2/14), AA3 是酸马奶、酸粥分离株所特有的。以上结果显示了不同来源的 *L. reuteri* 碳水化合物活性酶种类多样且具有宿主特异性。

3 讨论

本文研究了 4 种来源共 34 株 *L. reuteri* 的基因特征及生境适应性。比较了来源于酸马奶、酸粥、酸面团和食草动物 *L. reuteri* 的基因组特征, 泛基因集和核心基因集分别包含 7242、969 个基因家族, 且各类宿主特异性基因种类及数量差异较大, 主要编码与蛋白质代谢(37.04%)相关蛋白, 表明这些基因在基因组中处于不可或缺的位置, 在菌株生长和繁殖过程中发挥重要作用。基因组大小、G+C 含量表明不同来源的 *L. reuteri* 具有广泛的基因多样性。11 株酸马奶、酸粥分离株的 G+C 含量比酸面团和食草动物分离株略有偏高, 这可能与自然选择、变异、生活方式及所处环境等众多因素有关^[16]。34 株 *L. reuteri* 系统发育关系表明相同分离源具有较近的亲缘关系, 表明不同生存环境对 *L. reuteri* 的进化发挥重要的作用。故本文初步推断基因多样性可能与生存环境有关。

功能注释方面发现在酸马奶、酸粥分离株中具有较多与半乳糖代谢相关基因, 这可能因为马

乳中的乳糖含量高于牛乳、牦牛乳、羊乳, 且具有较多种类的氨基酸合成代谢途径, 说明这些菌株处于营养丰富环境并没有使其丧失一些物质的合成能力。此外, b、c 分支均为面团分离株, 面团中蛋白质主要为麦醇溶蛋白和麦谷蛋白, 其中麦醇溶蛋白主要由非极性氨基酸组成, 在 b、c 分支中具有相关基因。d 分支为食草动物分离株, 其特异性基因为与卟啉、叶绿素代谢及氨基酸生物合成相关基因。叶绿素是绿色植物中的主要物质, 研究发现光合细菌有叶绿素合成代谢基因^[17], 但在乳酸菌中还未曾报道。因为食草动物主要摄入树叶、草和其他植物, 只能利用非蛋白氮合成氨基酸、肽和蛋白质^[18], 因此为了满足食草动物的营养需求, *L. reuteri* 可能参与食草动物肠道内氨基酸的生物合成。在抗生素耐药性方面, a、d 分支中出现了 β -内酰胺抗性基因和万古霉素耐药性基因。这些特殊基因可能是由于宿主肠道抗生素耐药性菌群基因的水平转移导致的。抗生素滥用导致了不同环境的严重污染, 对微生物产生了越来越大的选择性压力, 从而导致抗生素耐药性增强^[19]。这些结果也反映出不同宿主 *L. reuteri* 的功能基因在各自宿主中的生存方式。

碳水化合物是生命细胞结构的主要成分及主要供能物质, 并且有调节细胞活动的重要功能。本文经比较了酸马奶、酸粥、酸面团和食草动物分离株碳水化合物活性酶相关基因, 结果表明 34 株 *L. reuteri* 共注释到 51 个碳水化合物活性酶家族, 其中共有 20 个 GH 家族, 9 个 GT 家族, 9 个 CE 家族, 9 个 CBM 家族, 2 个 AA 家族, 具有较为广泛的多样性。Sun 等^[20]基于 213 株乳杆菌模式菌株的比较基因组学研究表明目前已发现的所有乳杆菌基因组中, 共编码 48 个糖苷水解酶, 本研究中仅 *L. reuteri* 一个物种就具有其中的 20 种, 足以说

明 *L. reuteri* 的糖苷水解酶含量之丰富。GH3 和 GH43 只出现在酸马奶、酸粥和食草动物分离株中。GH3 编码 β -葡萄糖苷酶；木聚糖酶 1,4- β -木糖糖苷酶等，其广泛分布于细菌、真菌和植物中，具有纤维素生物量降解、能量代谢和病原体防御等多种功能^[21]。GH43 主要编码 β -木糖酶、 α -1-阿拉伯呋喃基葡萄糖苷酶、木聚糖酶等，木聚糖是半纤维素的组成成分，广泛分布于高等植物细胞壁中；GH25、GH73 (主要编码溶菌酶类，有抗菌、消炎、抗病毒等作用)在食草动物分离株中数量较多的原因可能为食草动物长期与外界环境接触导致致病菌增多，此类菌株为抵抗病原菌入侵而作出的进化选择；GH42、GH65、GH68 在酸马奶、酸粥分离株中较多，其中 GH42 主要编码低聚半乳糖和乳糖酶类，作用底物主要为乳糖和乳果糖，可用于降低 UHT 奶中乳糖含量^[22]；GH68 编码果聚糖蔗糖酶、 β -呋喃果糖苷酶、菊粉蔗糖酶，以蔗糖或果糖为底物产生菊糖或低聚果糖^[23]，具有多种有益功能如甜味取代剂、低卡路里、增加对致病菌的抵抗力等。GH66 只存在于食草动物组中，主要编码右旋糖酶，是细菌和真菌中主要酶类，是从植物中分离出来的^[24]，食草动物的长期植物饮食可能导致了肠道中 *L. reuteri* 菌株的特异性。AA3 仅在发酵食品分离株中存在，其主要为氧化还原酶类，细菌在氧气环境中生长，细胞内会产生部分还原的活性氧包括过氧化氢、超氧自由基和羟基^[25]。这些活性氧会引起 DNA、脂类和蛋白质的氧化，对细胞造成致命的损伤，表明发酵食品分离株抗氧化能力较强。Su 等^[26] 研究表明发酵食品中乳酸菌最初来源于肠道，而 GH3、GH43、CBM16、CBM37 是发酵食品和食草动物肠道分离株所共有的，说明这类来源的 *L. reuteri* 在适应肠外环境时仍然保留了部分肠道环境的特征；AA3 和 GH66 分别是发酵食品和食草

动物所特有的，这也说明菌株在适应过程中可能水平获得了某些适应性基因以此来更好地适应环境。

综上，这些结果表明不同来源 *L. reuteri* 具有广泛的基因多样性，且与其生存环境密切相关，可利用发酵食品作为模型来研究更为复杂的肠道环境。未来，随着在其他环境中 *L. reuteri* 菌株的分离，还可以比较更多不同环境中 *L. reuteri* 与其他宿主菌株的基因组差异，这将有力地说明影响其基因多样性及生境适应性的因素，为 *L. reuteri* 优良菌株的开发提供理论基础。

参考文献

- [1] Talarico TL, Dobrogosz WJ. Chemical characterization of an antimicrobial substance produced by *Lactobacillus reuteri*. *Antimicrobial Agents and Chemotherapy*, 1989, 33(5): 674–679.
- [2] Weizman Z, Asli G, Alsheikh A. Effect of a probiotic infant formula on infections in child care centers: comparison of two probiotic agents. *Pediatrics*, 2005, 115(1): 5–9.
- [3] Indrio F, Riezzo G, Raimondi F, Bisceglia M, Cavallo L, Francavilla R. The effects of probiotics on feeding tolerance, bowel habits, and gastrointestinal motility in preterm newborns. *The Journal of Pediatrics*, 2008, 152(6): 801–806.
- [4] Duar RM, Frese SA, Lin XB, Fernando SC, Burkey TE, Tasseva G, Peterson DA, Blom J, Wenzel CQ, Szymanski CM, Walter J. Experimental evaluation of host adaptation of *Lactobacillus reuteri* to different vertebrate species. *Applied and Environmental Microbiology*, 2017, 83(12): e00132.
- [5] Oh PL, Benson AK, Peterson DA, Patil PB, Moriyama EN, Roos S, Walter J. Diversification of the gut symbiont *Lactobacillus reuteri* as a result of host-driven evolution. *The ISME Journal*, 2010, 4(3): 377–387.
- [6] Wegmann U, MacKenzie DA, Zheng JS, Goesmann A, Roos S, Swarbreck D, Walter J, Crossman LC, Juge N. The pan-genome of *Lactobacillus reuteri* strains originating from the pig gastrointestinal tract. *BMC Genomics*, 2015, 16: 1023.
- [7] Zheng JS, Zhao X, Lin XB, Gänzle M. Comparative genomics *Lactobacillus reuteri* from sourdough reveals adaptation of an intestinal symbiont to food fermentations.

- Scientific Reports*, 2015, 5(1): 18234.
- [8] Francis F, Kim J, Ramaraj T, Farmer A, Rush MC, Ham JH. Comparative genomic analysis of two *Burkholderia glumae* strains from different geographic origins reveals a high degree of plasticity in genome structure associated with genomic islands. *Molecular Genetics and Genomics*, 2013, 288(3/4): 195–203.
- [9] Yu J, Zhao J, Song YQ, Zhang JC, Yu ZJ, Zhang HP, Sun ZH. Comparative genomics of the herbivore gut symbiont *Lactobacillus reuteri* reveals genetic diversity and lifestyle adaptation. *Frontiers in Microbiology*, 2018, 9: 1151.
- [10] Luo RB, Liu BH, Xie YL, Li ZY, Huang WH, Yuan JY, He GZ, Chen YX, Pan Q, Liu YJ, Tang JB, Wu GX, Zhang H, Shi YJ, Liu Y, Yu C, Wang B, Lu Y, Han CL, Cheung DW, Yiu SM, Peng SL, Xiaoqian Z, Liu GM, Liao XK, Li YR, Yang HM, Wang J, Lam TW, Wang J. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *Gigascience*, 2012, 1: 18.
- [11] Tatusov RL, Fedorova ND, Jackson JD, Jacobs RA, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*, 2003, 4(1): 41.
- [12] Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 2014, 30(14): 2068–2069.
- [13] 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. *Bioinformatics*, 2015, 31(22): 3691–3693.
- [14] Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 2016, 44(D1): D457–D462.
- [15] Marchin M, Kelly PT, Fang JW. Tracker: continuous HMMER and BLAST searching. *Bioinformatics*, 2005, 21(3): 388–389.
- [16] Wu H, Zhang Z, Hu SN, Yu J. On the molecular mechanism of GC content variation among eubacterial genomes. *Biology Direct*, 2012, 7: 2.
- [17] Yamagata H, Matoba R, Fujii T, Yukawa H. Application of photosynthetic bacteria for porphyrin production. *Studies in Surface Science and Catalysis*, 1998, 114: 475–478.
- [18] Allison MJ. Biosynthesis of amino acids by ruminal microorganisms. *Journal of Animal Science*, 1969, 29(5): 797–807.
- [19] Qian Y, Yue DM, Peng YK, Liu Y, Xiao L. Occurrence and distribution of antibiotic-resistant bacteria and transfer of resistance genes in Lake Taihu. *Microbes and Environments*, 2013, 28(4): 479–486.
- [20] Sun ZH, Harris HMB, McCann A, Guo CY, Argimon S, Zhang WY, Yang XW, Jeffery IB, Cooney JC, Kagawa TF, Liu WJ, Song YQ, Salvetti E, Wrobel A, Rasinkangas P, Parkhill J, Rea MC, O'Sullivan O, Ritari J, Douillard FP, Paul Ross R, Yang RF, Briner AE, Felis GE, de Vos WM, Barrangou R, Klaenhammer TR, Caufield PW, Cui YJ, Zhang HP, O'Toole PW. Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. *Nature Communications*, 2015, 6(1): 8322.
- [21] Macdonald SS, Blaukopf M, Withers SG. *N*-acetylglucosaminidases from CAZy family GH3 are really glycoside phosphorylases, thereby explaining their use of histidine as an acid/base catalyst in place of glutamic acid. *Journal of Biological Chemistry*, 2015, 290(8): 4887–4895.
- [22] Di Lauro B, Strazzulli A, Perugini G, La Cara F, Bedini E, Corsaro MM, Rossi M, Moracci M. Isolation and characterization of a new family 42 β -galactosidase from the thermoacidophilic bacterium *Alicyclobacillus acidocaldarius*: identification of the active site residues. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 2008, 1784(2): 292–301.
- [23] Ozimek LK, van Hijum SAFT, van Koningsveld GA, van der Maarel MJEC, van Geel-Schutten GH, Dijkhuizen L. Site-directed mutagenesis study of the three catalytic residues of the fructosyltransferases of *Lactobacillus reuteri* 121. *FEBS Letters*, 2004, 560(1/3): 131–133.
- [24] Kim YM, Kiso Y, Muraki T, Kang MS, Nakai H, Saburi W, Lang W, Kang HK, Okuyama M, Mori H, Suzuki R, Funane K, Suzuki N, Momma M, Fujimoto Z, Oguma T, Kobayashi M, Kim D, Kimura A. Novel dextranase catalyzing yeloisomaltooligosaccharide formation and identification of catalytic amino acids and their functions using chemical rescue approach. *Journal of Biological Chemistry*, 2012, 287(24): 19927–19935.
- [25] Sützl L, Laurent CVFP, Abrera AT, Schütz G, Ludwig R, Haltrich D. Multiplicity of enzymatic functions in the CAZy AA3 family. *Applied Microbiology and Biotechnology*, 2018, 102(6): 2477–2492.
- [26] Su MSW, Oh PL, Walter J, Gänzle MG. Intestinal origin of sourdough *Lactobacillus reuteri* isolates as revealed by phylogenetic, genetic, and physiological analysis. *Applied and Environmental Microbiology*, 2012, 78(18): 6777–6780.

Comparative genomics analysis of genetic diversity and habitat adaptability of *Lactobacillus reuteri* from different sources

Xiaona An, Weicheng Li, Jie Yu, Lin Pan, Lanxin Mo, Caiqing Yao, Heping Zhang*

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

Abstract: [Objective] We studied the genetic diversity and habitat adaptability mechanism of 34 strains of *Lactobacillus reuteri* from different sources. We aimed to compare the differences of their adaptability in extraintestinal and intrainestinal habitats to provide theoretical basis for the development of *Lactobacillus reuteri* strains. [Methods] Based on the second generation sequencing platform, We sequenced 11 *Lactobacillus reuteri* strains from fermented food (yogurt and sour porridge) and compared with the genomes of fermented food, sour dough and herbivore sources of *Lactobacillus reuteri* strains by comparative genomics. [Results] The average genome size and GC content of *L. reuteri* isolated from yoghurt and sour porridge were 2.14 Mb and 38.77% respectively, and the phylogenetic relationship between the isolates from the same source was close. Pan-core gene sets consisted of 7242 and 969 gene families, yoghurt isolates had the most specific genes (459). Functional annotation showed that the number and types of carbohydrate and amino acid related genes varied greatly among strains from different sources, and the antibiotic resistance genes were found only in fermented food and herbivores strains. The carbohydrate-active enzymes of GH3 (β -glucosidase etc.) and GH43 (β -xylosidase etc.) only appeared in fermented food herbivore isolates, and the specific ones were AA3 (cellobiose dehydrogenase etc.) and GH66 (dextranase etc.), respectively. [Conclusion] *Lactobacillus reuteri* from different sources has a wide range of genetic diversity and is closely related to the living environment. Fermented food isolates had the characteristics of some herbivorous intestinal strains and their unique environmental adaptability, and reflected the host environmental adaptability. It can deepen the understanding of food fermentation and bacterial habitat adaptability in intestinal environment.

Keywords: *Lactobacillus reuteri*, comparative genomics, fermented food, habitat adaptability

(本文责编: 李磊)

Supported by the National Natural Science Foundation of China (31430066), by the General Project of Inner Mongolia Natural Foundation (2018MS03041) and by the Training Program of Science Foundation for Outstanding Youth (2017XYQ-4)

*Corresponding author. Tel/Fax: +86-471-4319940; E-mail: hepingdd@vip.sina.com

Received: 31 July 2019; Revised: 20 September 2019; Published online: 9 March 2020