



基于比较基因组学揭示不同植物乳杆菌的遗传特征及菌株差异——以 *Lactobacillus plantarum* P9 和 Lp-6 研究为例

刘文君, 吕瑞瑞, 李伟程, 孙志宏, 陈永福, 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, 内蒙古乳品生物技术与工程重点实验室, 内蒙古 呼和浩特 010018

摘要: 【目的】植物乳杆菌(*Lactobacillus plantarum*, *L. plantarum*) 在食品、医药和动物养殖等多个领域均有应用。本文以 *L. plantarum* P9 和 Lp-6 为例, 解析 *L. plantarum* 遗传背景和基因组特征, 为其鉴定和开发奠定基础。【方法】本研究采用 PacBio SMRT 测序技术完成了 *L. plantarum* P9 和 Lp-6 全基因组测序, 结合已公开的 110 株 *L. plantarum* 全基因组数据和 1 株模式菌株 ATCC 14917^T 数据, 通过比较基因组学方法探究 *L. plantarum* 基因组的差异。【结果】*L. plantarum* P9 和 Lp-6 基因组大小分别为 3314.1 和 3482.5 kb, GC 含量(%) 分别为 44.38% 和 44.32%, 二者分别含有 8 个和 9 个质粒。113 株 *L. plantarum* 系统发育树结果显示, *L. plantarum* P9 与 *L. plantarum* ATCC 14917^T 遗传距离更近, *L. plantarum* Lp-6 更接近祖先群体分支。与 *L. plantarum* WCSF1 相比, 含有 *xerS* 等基因的 22.0 kb 基因组片段在 *L. plantarum* Lp-6 上发生了倒位, 在 *L. plantarum* P9 基因组中缺失; *L. plantarum* Lp-6 染色体插入含 *tagF* 等基因的 13.0 kb 片段; 包含 *gpmA* 等基因的 14.4 kb 基因片段插入到 *L. plantarum* P9 染色体中。【结论】通过比较基因组学方法解析 *L. plantarum* P9 和 Lp-6 遗传信息, 发现不同 *L. plantarum* 菌株的遗传特征存在差异。

关键词: 植物乳杆菌, 全基因组测序, 基因组图谱, 比较基因组学

植物乳杆菌 (*Lactobacillus plantarum*, *L. plantarum*) 是一种兼性异型的乳酸菌 (lactic acid bacterial, LAB), 其广泛存在于碳水化合物

高的环境中, 如乳制品、发酵肉制品和植物基质等多种环境^[1]。*L. plantarum* 作为人体肠道土著菌群具有很高的安全性, 是益生菌开发的潜力贮

基金项目: 内蒙古自治区科技重大专项(zd zx 2018018); 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。Tel: +86-471-4300593; E-mail: hepingdd@vip.sina.com

收稿日期: 2020-08-28; 修回日期: 2020-10-03; 网络出版日期: 2020-10-22

备^[2]。Nguyen 等^[3]研究发现部分 *L. plantarum* 有平衡肠道菌群、缓解炎症和降低血胆固醇等作用。Kim Eiseul 等^[4]发现 *L. plantarum* 可产生植物乳杆菌素, 对革兰氏阴性菌和阳性菌均有抗菌活性。由于 *L. plantarum* 特性优良使得它无论在发酵食品、医药领域和工业乳酸发酵都有广泛应用。

随着高通量技术不断发展, 测序成本和时间很大程度降低, 越来越多的菌株完成了全基因组测序^[5]。比较基因组学(comparative genomics)是以相关物种基因组的相似性为基础, 根据测序和基因组图谱, 通过基因组信息推测基因数目、功能及表达机制的研究^[6]。利用比较基因组学来构建物种核心基因组(core genome)和泛基因组(pan genome), 通过讨论物种的直系同源序列从而分析其群体遗传和表型演化^[5]。*L. plantarum* 作为乳酸菌组成的重要菌株, 通过比较基因组学方法揭示其丰富的基因组信息, 对食品工业发酵、医疗保健和农业生产至关重要。

本团队前期通过对 121 株食物来源的 *L. plantarum* 筛选到 1 株 *L. plantarum* P9, 其可在大鼠肠道内定殖、排泄有机磷农药和缓解肠道炎症^[7]。本团队于 2004 年从新疆酸马奶中分离得到 *L. plantarum* Lp-6, 后续实验发现该菌可对有害菌、致病菌具有一定抑菌特性。本研究对 *L. plantarum* Lp-6 和 *L. plantarum* P9 进行 PacBio SMRT 全基因组测序, 并从 NCBI Refseq 数据库下载了 110 株植物乳杆菌全基因组序列和 1 株模式菌株 *L. plantarum* ATCC 14917^T 基因组序列。以 113 株 *L. plantarum* 为研究对象, 采用比较基因组的方法从全基因组水平探究不同菌株之间基因组的差异, 解析不同 *L. plantarum* 菌株遗传异质性, 为 *L. plantarum* 应用奠定遗传学基础。

1 材料和方法

1.1 实验菌株

本研究测定的 2 株 *L. plantarum* 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), *L. plantarum* P9 分离自内蒙古自治区自然发酵酸粥, *L. plantarum* Lp-6 分离自新疆维吾尔自治区酸马奶。

截止 7 月 10 日为止, 将 NCBI (national coalition building institute, <https://www.ncbi.nlm.nih.gov/>) Refseq 数据库 112 株 *L. plantarum* 基因组完成图全部下载, 同时下载模式菌株 *L. plantarum* ATCC 14917^T。计算平均核苷酸一致性(average nucleotide identity, ANI)后, 去掉 2 株与模式菌株 ANI 值小于 95% 菌株。

1.2 主要试剂和仪器

Wizard[®] Genomic DNA Purification Kit (美国普洛麦格生物技术有限公司); MRS 培养基(Oxoid 公司); SMRT 建库试剂盒(DNA Template Prep Kit 1.0)和测序试剂(DNA/Polymerase Binding Kit P6 v2 和 DNA Sequencing Bundle 4.0) (美国太平洋生物科学公司); 电热恒温培养箱(北京一恒科技有限公司); 超微量紫外分光光度计(NanoDrop 公司); Qubit 2.0 荧光计(美国 Life Technologies); PicBio SMRT RS II 测序平台(美国 Pacific Biosciences 公司)。

1.3 菌株培养与基因组 DNA 的提取、测序

本实验所用菌株 *L. plantarum* P9 和 *L. plantarum* LP-6, 具体步骤如下: (1) 将其分别接种至 5 mL 的 MRS 液体培养基中, 37 °C, 无氧条件培养 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

到第三代的菌液以 3000 g/min 离心 10 min 收集菌泥；(4) 用 PBS 缓冲液将菌泥清洗 2 遍后，准备进行全基因组 DNA 的提取。全基因组 DNA 提取流程参考本实验室钟智等^[8]的方法。

对于提取后的 DNA 具体处理如下：(1) 使用 NanoDrop 微量紫外分光光度计及 0.6% 的琼脂糖凝胶电泳对所提取的 DNA 进行浓度、纯度和片段完整度进行检测；(2) 将符合要求的 DNA 合并到 2 mL 离心管中，加入 0.46×磁珠进行吸附；(3) 将废液吸出，并且用 75%乙醇对吸附 DNA 的磁珠快速洗涤两遍；(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 的质量浓度。通过回溶后 DNA 体积与浓度相乘获得 DNA 的最终质量，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 *L. plantarum* DNA 按照 PacBio SMRT 建库流程对提取的全基因组 DNA 建立 10 kb 文库，建库后根据上机流程采用 PacBio SMRT RS II 测序平台进行全基因组测序。每株菌使用一个 PacBio SMRT cell，平均可测得高于 150 倍的原始数据。

1.4 基因组测序和组装

将获取的原始数据进行质量评估，经过低质量 reads 的过滤、去除接头和引物后，获得高质量的 reads 序列，通过 SMRT[®] Portal (V2.7) 软件的方案 RS_HGAP_Assembly.3 进行质控和菌株基因组组装。使用 circlator (V1.5.5)^[9] 对下机数据进行环化得到基因组完成图。

1.5 比较基因组学分析

1.5.1 平均核算一致性(average nucleotide identity, ANI)和总核酸一致性(total nucleotide identity, TNI)值的计算：本研究中 *L. plantarum* 的 ANI 和

TNI 值的计算参考 Goris 等^[10]和 Chen 等^[11]方法，用自制 Perl 脚本进行计算。

1.5.2 全基因组圈图和功能基因组分析：将上述组装好的 *L. plantarum* P9 和 *L. plantarum* Lp-6 的核酸序列文件分别上传至 RAST (repaid annotation using subsystem technology; <http://rast.nmpdr.org/rast.cgi>) 进行注释，并下载对应的基因功能注释文件。将 *L. plantarum* P9 和 *L. plantarum* Lp-6 染色体与质粒拆分，与功能基因注释文件分别对应上传至 CGView Server (http://stothard.afns.ualberta.ca/cgview_server/)。统计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^[12] 软件对菌株基因组进行基因预测后，采用 Roary^[13] 软件识别核心基因集和附属基因集(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，核心基因是存在于所有菌株基因组中共有的基因；其中仅在一个菌株中出现的基因称为特有基因，所有菌株特有基因相加为特有基因之和。

1.5.4 系统发育树和共线性分析：利用 Roary 得到的核心基因序列使用 treebest 软件利用邻接法(neighbor-joining, NJ) 构建邻接树，探究 *L. plantarum* 的系统发育关系。使用 Mauve 软件^[14] 对 *L. plantarum* P9、*L. plantarum* Lp-6 和 *L. plantarum* WCSF1 的基因组序列进行共线性分析。

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。

1.6 菌株核酸登录号：*L. plantarum* P9 和 *L. plantarum* Lp-6 基因组序列已提交至 GeneBank 数据库，序列号为 PRJNA655563。

2 结果和分析

2.1 植物乳杆菌基因组信息

对于测序样品 *L. plantarum* P9 和 *L. plantarum* Lp-6 进行数据评估和序列组装, 得到基因组组装结果(表 1)和基因组基本特征(表 2), 将组装完成的序列拼接成基因组圈图(图 1)。结果显示, *L. plantarum* P9 基因组序列在组装后有 1 条环状染色体和 9 个质粒; 全基因组长度为 3314.2 kb, 其中环状染色体大小为 3067 kb, GC 含量(%)为 44.11%, 其中包含了 3147 个蛋白质编码序列 (protein-coding sequence, CDS)。*L. plantarum* Lp-6 基因组序列组装后由 1 条环状染色体和 8 个质粒

组成; 全基因组长度为 3482.5 kb, 其中环状染色体长度为 3101.3 kb, GC 含量(%)为 44.82%, 其中包含了 3436 个蛋白质编码序列 CDS。

L. plantarum 是乳酸菌中基因组最大的菌种之一, 全基因组长度一般在 3.2 Mb 至 3.4 Mb 之间^[15]。由基因组特征结果可知, *L. plantarum* P9 和 Lp-6 两株菌基因组特征具有一定差异, 全基因组长度分别为 3314.2 kb 和 3482.5 kb。*L. plantarum* P9 有 9 个质粒, 平均 GC 含量(%)为 38.19%, 其中 GC 含量(%)最高为 44.62%, 最低为 33.77%; *L. plantarum* Lp-6 含有 8 个质粒, 其平均 GC 含量(%)为 40.41%, GC 含量(%)最低为 36.86%, 最高为 41.22%。

表 1. 植物乳杆菌 P9 和 LP-6 的基因组组装结果统计

Table 1. Statistics of assembly results of *Lactobacillus plantarum* P9, LP-6 genome

Strains	Features	Number of scaffolds	Full length/bp	GC content/%	Strains	Features	Number of scaffolds	Full length/bp	GC content/%
<i>Lactobacillus plantarum</i> P9	Chromosome	1	3067028	44.11	<i>Lactobacillus plantarum</i> Lp-6	Chromosome	1	3101304	44.82
	Plasmid1	1	13709	44.62		Plasmid1	1	22072	40.45
	Plasmid2	1	91545	39.38		Plasmid2	1	49865	40.15
	Plasmid3	1	71514	38.64		Plasmid3	1	40043	41.22
	Plasmid4	1	52855	38.38		Plasmid4	1	104731	41.20
	Plasmid5	1	8884	35.21		Plasmid5	1	62199	41.01
	Plasmid6	1	39853	36.04		Plasmid6	1	62394	41.68
	Plasmid7	1	13703	37.36		Plasmid7	1	24406	36.86
	Plasmid8	1	6603	33.77	Plasmid8	1	15509	40.69	
Plasmid9	1	2206	40.34						

表 2. 植物乳杆菌 P9、LP-6 基因组基本特征

Table 2. General features of *Lactobacillus plantarum* P9, LP-6 genomes

Features	<i>Lactobacillus plantarum</i> P9	<i>Lactobacillus plantarum</i> Lp-6
Number of contigs	10	9
Bases/kb	3314.2	3482.5
Number of CDS	3147	3436
Number of tRNA	94	71
Number of tmRNA	1	1

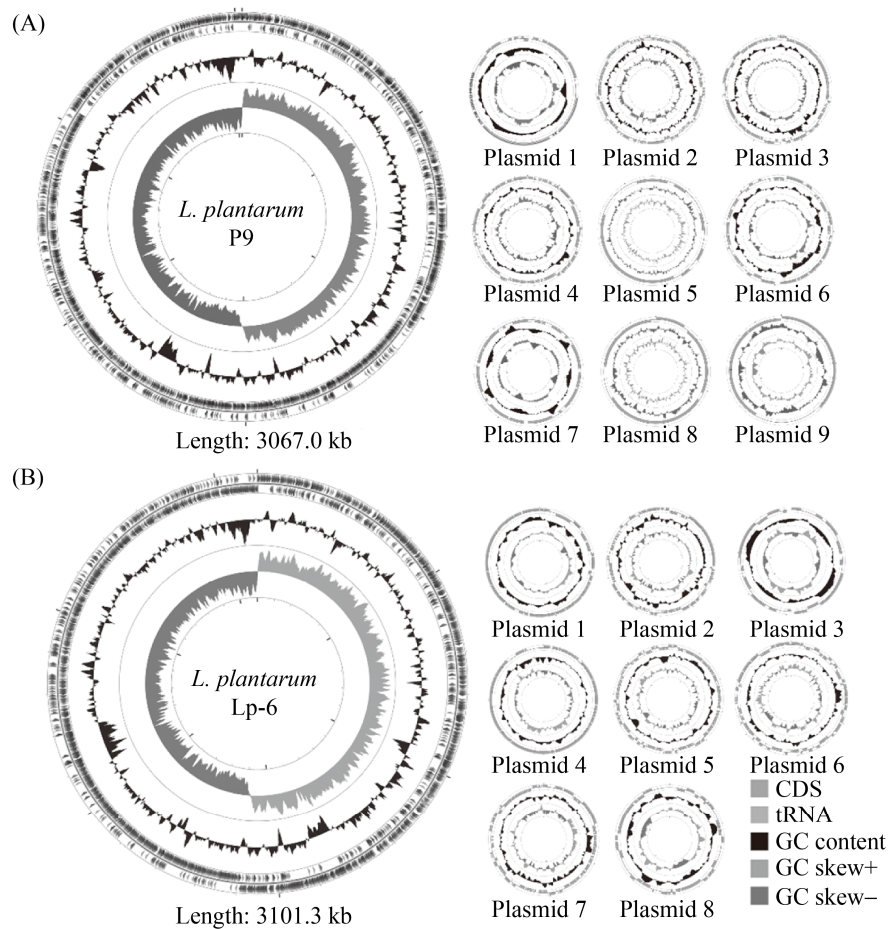


图 1. *Lactobacillus plantarum* P9 (A)、LP-6 (B) 的基因组圆形图谱
Figure 1. Circular map of *Lactobacillus plantarum* P9 (A), LP-6 (B) genome.

2.2 ANI 和 TNI 分析

ANI 是基因组的同源序列进行比对来判断菌株是否为同一种或亚种, TNI 是计算基因组之间核酸匹配比例, 为基因组数据集提供更高的区分能力^[16]。在比较基因组学分析过程中, ANI 可评估基因组间多态性的程度, 也可判断基因组之间的相似性, 一般认为 ANI 大于等于 95% 为同一种^[17]。

Zhong 等^[18]通过计算 161 株粪肠球菌 ANI 值发现不同菌株之间存在遗传差异。本研究对 113 株 *L. plantarum* 全基因组序列进行 ANI 和 TNI 分析, 并绘制了热图(图 2)。结果显示全部菌株之间 ANI

均大于 97.0%, TNI 大于 78.0%。*L. plantarum* P9 和 *L. plantarum* Lp-6 之间 ANI 为 98.5%, TNI 为 85.4%; *L. plantarum* P9 和模式菌株 *L. plantarum* ATCC 14917^T 间 ANI 为 99.0%, *L. plantarum* Lp-6 和 *L. plantarum* ATCC 14917^T 的 ANI 为 98.7%。*L. plantarum* P9 和 *L. plantarum* ATCC 14917^T 碱基和核酸匹配度更高, 不同 *L. plantarum* 基因组存在差异。

2.3 共线性分析

共线性分析是研究物种基因组间相关性, 基因组相关性越高其亲缘关系越近, 会出现共线性,

代表基因序列部分或全部保守^[19]。*L. plantarum* WCSF1 是从人类唾液中分离且于 2001 年完成首次测序, 在 2012 年完成了重新测序和重新注释^[20]。本实验中将该 *L. plantarum* WCSF1 菌株全基因组序列作为参考, 通过 Mauve 软件分析 *L. plantarum* Lp-6 和 *L. plantarum* P9 (图 3)。结果表明, *L. plantarum* Lp-6 和 *L. plantarum* P9 突变重组位置较少, 与 *L. plantarum* WCSF1 共线性及相似度

较好, *L. plantarum* Lp-6 和 *L. plantarum* P9 局部的插入、缺失和倒位发生在局部共线块上。

以 *L. plantarum* WCSF1 为参考, 观察到 22.0 kb 含有 *xerS*、*acm* 和 *xre* 等基因的片段在 *L. plantarum* Lp-6 发生了倒位现象, 而 *L. plantarum* P9 染色体上缺失该基因组片段, 这些基因主要涉及到转录调控、酪氨酸重组和溶菌酶合成等功能; *L. plantarum* Lp-6 染色体上发现 13.0 kb 特有基因组片段, 包含

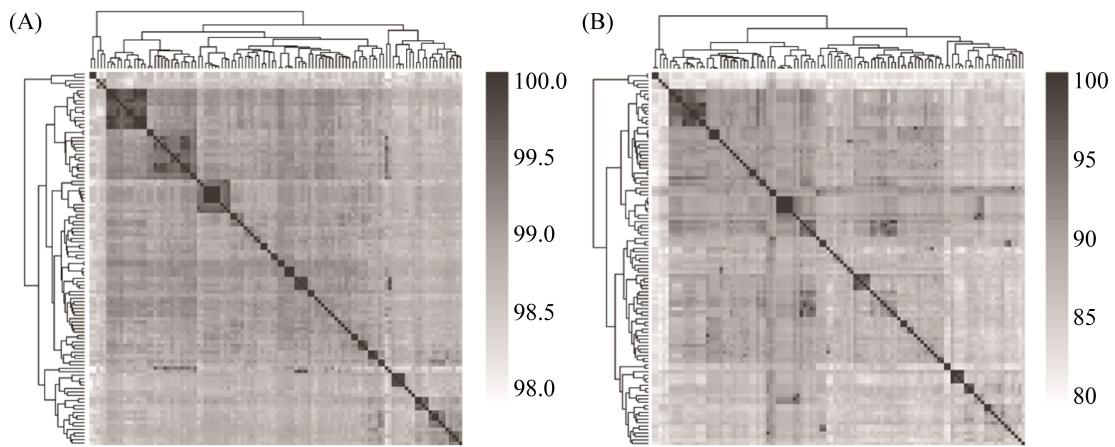


图 2. 111 株植物乳杆菌的 ANI (A)及 TNI (B)结果

Figure 2. Heatmap of ANI (A) and TNI (B) based on the sequences of 111 *L. plantarum*.

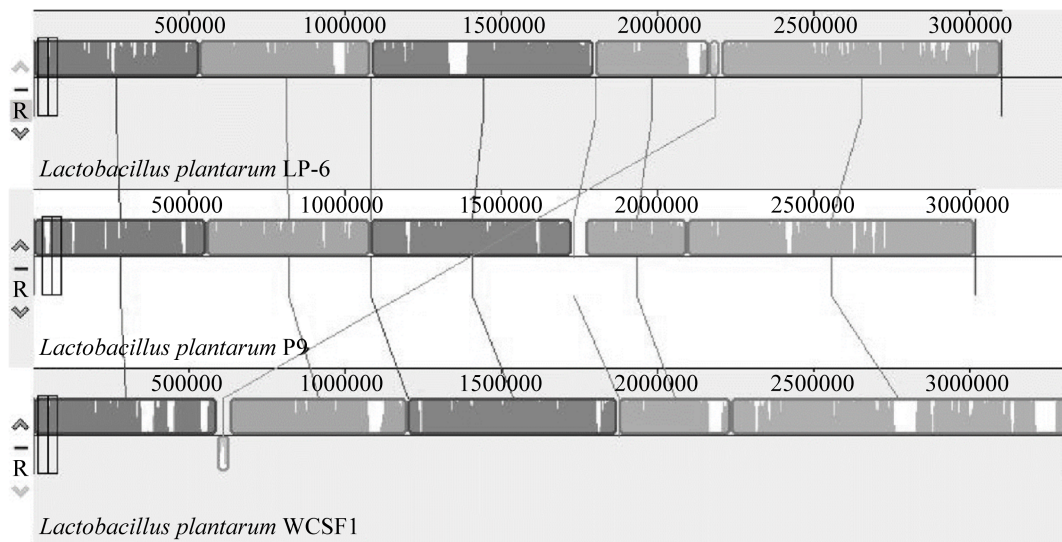


图 3. 植物乳杆菌共线性分析

Figure 3. Synteny analysis of *L. plantarum* using the genome of strain WCSF1 as the reference.

tagF、*ssbB* 和 *dnaC* 等基因，可能与磷酸转移、DNA 合成和重组有关；*L. plantarum* P9 染色体上 14.4 kb 基因片段观察到有 *gpmA*、*pstA* 和 *pstC* 等特有基因，这些基因均与磷酸盐代谢转运功能有关。

两株 *L. plantarum* 发生一定的基因组插入或缺失等现象。短基因片段的插入和倒位提示菌株在进化过程中可能发生了基因重组和转移现象^[21]，基因组的插入或缺失可能与开放性基因组的形成有关^[16]。*L. plantarum* Lp-6、*L. plantarum* P9 与参考菌株基因组差异表明了基因组在进化过程中发生了重组和转移，但该现象是否会导致 *L. plantarum* 形成开放性基因组仍有待研究。

2.4 泛-核心(pan-core)基因集的构建

为比较不同菌株间的基因差异，通过 Prokka 和 Roary 软件进行分析，以大于 90% 的氨基酸一致性为标准，统计核心基因和泛基因个数。113 株 *L. plantarum* 的泛基因集包括了 15473 个基因，其中核心基因 1482 个，特有基因 5773 个。由图 4 可知，随着基因组数量的增加，核心基因的个数逐渐趋于稳定，而泛基因个数仍然呈现增长的趋势。

研究表明，开放基因组特点为随着基因组数量的增加，核心基因趋于稳定，泛基因呈上升状态的曲线，则该基因组具有较高的遗传多样性^[22]。开放式基因组特点与本文 *L. plantarum* 的相同，表明 *L. plantarum* 为开放式基因组，故推测 *L. plantarum* 具有较高遗传异质性。

2.5 核心基因(core genome)系统发育分析

系统发育树可以直观地表现出同一个群体内不同个体间的遗传距离与进化关系，构建系统发育树是研究种群结构和物种进化的必要前提^[17]。

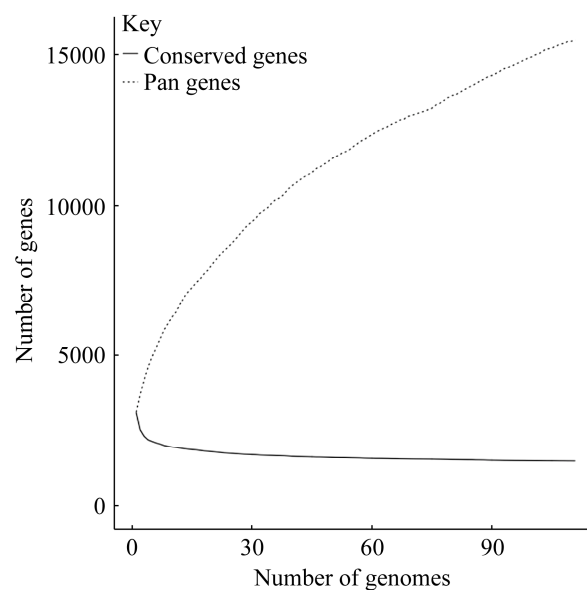


图 4. 泛基因集和核心基因集曲线

Figure 4. Pan-core gene family curve. The abscissa represents the number of strains, and the ordinate represents the number of genes.

为了进一步了解 *L. plantarum* P9、*L. plantarum* Lp-6 和其余 111 株植物乳杆菌的群体结构，本研究基于 1482 个核心基因的核酸序列，将近缘物种的 *L. plantarum* subsp. *argentoratensis* DSM16365^T 作为外群，通过邻接法(neighbor-joining)构建了系统发育树，bootstrap 值为 1000 (图 5)。113 株 *L. plantarum* 其中包括 75 株食品分离源、21 株动物肠道分离源、10 株其他分离源、4 株果蝇分离源及 3 株植物分离源。113 株 *L. plantarum* 在进化过程中由于其遗传多样性造成菌株之间的差异，由图 5 可知，其中分为 3 个主要分支，集中在一个分支表明亲缘关系更近。*L. plantarum* Lp-6 所在分支更靠近外群根部；*L. plantarum* ATCC14917^T 在距离根部最远的 H3 分支，该分支包括 63 株 *L. plantarum* 菌株，主要由 4 株果蝇分离株(占果蝇分离株 100%)、16 株动物肠道分离株(占动物

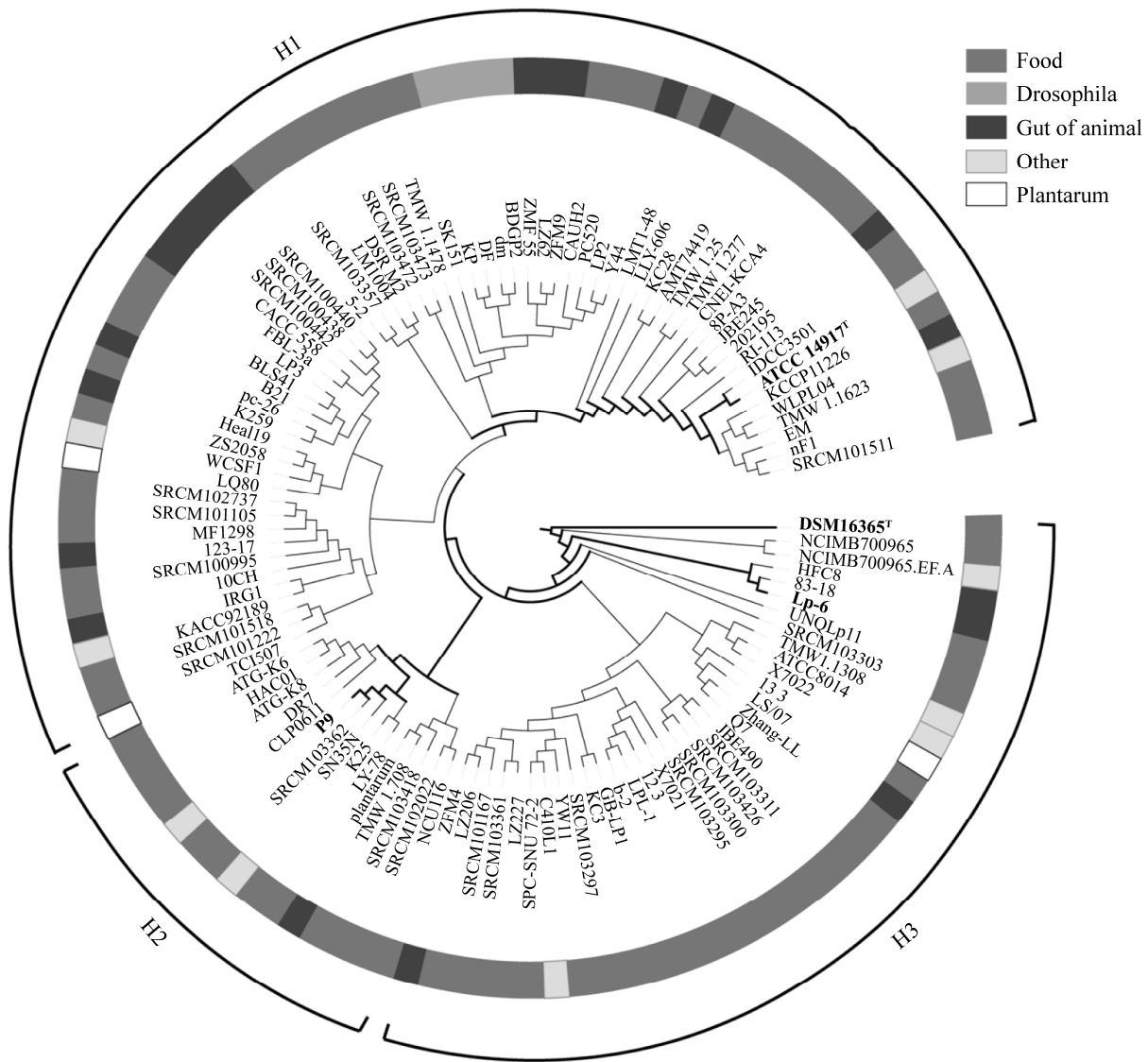


图 5. 基于核心基因序列的系统发育树

Figure 5. Phylogenetic tree based on core gene sequences.

肠道分离株 76%)和 37 株食品分离株(占食品分离株 49%); *L. plantarum* P9 在 H2 分支, 有 15 株 *L. plantarum*, 其中 12 株分离自食品源。*L. plantarum* Lp-6 所在 H3 群体, 可能是 *L. plantarum* P9 和 *L. plantarum* ATCC14917^T 的祖先群体。果蝇分离株和动物肠道分离株呈现明显的聚集性。本实验室前期刘亚华^[23]研究发现果蝇分离株由于分离环境

特殊, 细菌基因组会对环境适应或进化累积该环境相关的遗传变异信息, 从而导致菌株间遗传信息具有差异。宋宇琴^[17]通过对 200 株不同地区的德式乳杆菌保加利亚亚种研究发现, 不同地理环境生存的菌株为应对环境压力, 发生了适应性进化。但本文中分离自果蝇的植物乳杆菌较少, 是否由于适应性进化产生环境特异性基因仍有待研究。

2.6 基于全基因组测序对植物乳杆菌菌株水平的鉴定

2.6.1 113 株植物乳杆菌菌株水平鉴定：在 *L. plantarum* 菌株基因组水平上，对 113 株 *L. plantarum* 菌株通过 Excel 随机抽样函数随机挑选 6 株 *L. plantarum*。通过比较基因组学分析发现不同菌株基因组大小、质粒数和 GC 含量(%) 均有不同(表 3)；计算核酸一致性(ANI 和 TNI)显示出 6 株 *L. plantarum* 与模式菌株 *L. plantarum* ATCC 14917^T 核酸匹配度各有差异；以核心基因构建系统发育树结果显示 6 株 *L. plantarum* 亲缘关系远近不同而分布在不同分支；对 113 株 *L. plantarum* 菌株核心基因泛基因组进行分析，其中 6 株菌为例，不同菌株特有基因个数和基因注释功能各有差异，其中表 3 中特有基因数量仅为 Prokka 注释到功能的基因。综上所述，本研究可通过比较基因组学的方法，发现不同的菌株基因组上存在的差异，最终从基因组特征、核酸一致性、系统发育树和特有基因等角度实现 *L. plantarum* 菌株水平的鉴定，为工业和生产上 *L. plantarum* 的应用奠定遗传学基础。

2.6.2 植物乳杆菌 P9 和 Lp-6 差异分析：*L. plantarum* P9 和 *L. plantarum* Lp-6 为本实验室

前期研究筛选出的菌株，前期研究发现可能具有益生特性。本文以 *L. plantarum* P9 和 *L. plantarum* Lp-6 为例，通过比较基因组学发现，在基因组水平上 *L. plantarum* P9 比 *L. plantarum* Lp-6 基因组 GC 含量(%)小，但质粒数更多；从核酸一致性角度，*L. plantarum* P9 和 *L. plantarum* Lp-6 之间 ANI 为 98.5%，TNI 为 85.4%，表明二者均为 *L. plantarum* 菌株，但其核酸一致性存在一定差异，可能是在进化过程中不同分离源所导致；系统发育树结果显示，113 株 *L. plantarum* 可分为 3 个分支，*L. plantarum* P9 和 *L. plantarum* Lp-6 在不同分支，显示其亲缘关系较远，但与作为外群的菌株 *L. plantarum* subsp. *argentoratensis* DSM16365^T 距离较远，表明 *L. plantarum* P9 和 *L. plantarum* Lp-6 均为 *L. plantarum* subsp. *plantarum*。113 株 *L. plantarum* 菌株核心基因泛基因组分析结果显示，其中通过 Prokka 注释到功能的基因，*L. plantarum* P9 有 *lysS* 和 *hslO* 为代表的 14 个特有基因，*L. plantarum* Lp-6 有 *iap* 和 *nadC* 等 13 个特有基因(表 4)。在染色体结构上，通过共线性分析发现 *L. plantarum* P9 和 *L. plantarum* Lp-6 染色体中各自有部分基因片段发生基因重组和转移。

表 3. 部分植物乳杆菌菌株基因组信息

Table 3. Genomic information of some *Lactobacillus plantarum* strains

Strain names	Plasmids	Genome size	GC Content/%	Branches	The reference is <i>L. plantarum</i> ATCC 14917 ^T		Number of specific genes
					ANI/%	TNI/%	
dm	1	3325676	44.52	H1	99.26	92.53	1
SK151	1	3231249	44.56	H1	99.25	91.30	0
NCIMB700965.EF.A	6	3217131	44.53	H3	98.48	79.21	1
TMW 1.277	11	3400131	44.22	H1	99.38	87.92	3
SN35N	5	3415192	44.27	H2	98.98	86.40	3
SRCM103300	5	3388332	44.35	H3	98.82	85.86	8

表 4. Prokka 注释植物乳杆菌 P9 和 Lp-6 特有基因
Table 4. The unique genes of *L. plantarum* P9 and *L. plantarum* Lp-6 annotated by Prokka

<i>L. plantarum</i> P9	Number	<i>L. plantarum</i> Lp-6	Number
<i>rep</i>	1	<i>ydjI</i>	1
<i>nfrA</i>	2	<i>wbbI</i>	1
<i>murG</i>	1	<i>puuP</i>	1
<i>melR</i>	1	<i>pucI</i>	1
<i>lysS</i>	3	<i>plaP</i>	1
<i>ltaS</i>	2	<i>nadC</i>	2
<i>hslO</i>	3	<i>menE</i>	1
<i>hmoB</i>	1	<i>menB</i>	1
<i>hcaR</i>	1	<i>iap</i>	4
<i>gmuC</i>	1	<i>hepT</i>	1
<i>dusC</i>	1	<i>gadB</i>	2
<i>drpA</i>	1	<i>emrB</i>	1
<i>codB</i>	1	<i>alx</i>	1
<i>clpC</i>	1		

为进一步了解 *L. plantarum* P9 和 *L. plantarum* Lp-6 基因组的差异, 通过 Prokka 注释结果发现, *L. plantarum* P9 注释到 3093 个基因, 特有基因为 112 个, 其中注释到功能相关的基因有 14 个, *L. plantarum* Lp-6 注释到 3355 个基因, 特有基因 259 个, 有 13 个功能相关的基因, 其中二者与功能有关的特有基因在表中(表 4)列出, *L. plantarum* P9 特有基因 14 个, 主要参与 DNA 复制重组、NADPH 氧化还原酶、赖氨酸合成及其他生物合成及代谢途径; *L. plantarum* Lp-6 包括 13 个特有基因, 涉及膜转运功能、胺类、肽酶和酯酶的合成以及多种物质代谢功能调控, 其中有 4 个拷贝的 *iap* 基因参与肽链内切酶合成。蛋白酶和肽酶可降解肽类和氨基酸, 构成蛋白质水解系统^[24]。不同分离源的菌株在适应环境的过程中, 可能会获得有利于其在该环境下生存的环境特异性基因^[23]。两株菌特有基因功能具有较大差异, 但 *L. plantarum* Lp-6 注释到较多与蛋白水解相关的特有功能基因, 故推测 *L. plantarum* Lp-6 有更强的蛋白水解能力, 但还需进一步研究验证。

3 结论

随着测序技术不断发展以及生物信息学的不断进步, 越来越多 *L. plantarum* 基因组组装得到基因组完成图。本研究通过 PacBio SMRT 得到 *L. plantarum* Lp-6 和 *L. plantarum* P9 基因组完成图, 结合比较基因组学方法发现 *L. plantarum* Lp-6 和 *L. plantarum* P9 基因组大小、质粒数量和染色体结构均存在差异。*L. plantarum* 基因组特征除 GC 含量(%)外, 基因组大小、质粒数均有差异。系统发育树结果显示 *L. plantarum* P9 与 *L. plantarum* ATCC 14917^T 遗传距离更近, *L. plantarum* Lp-6 更接近祖先群体分支。与参考菌株 *L. plantarum* WCFS1 相比, *L. plantarum* Lp-6 约 22.0 kb 的基因片段发生倒位, 该片段在 *L. plantarum* P9 发生缺失, 主要包含有 *xerS*、*acm* 和 *xre* 等基因; *L. plantarum* Lp-6 缺失了大小约 13.0 kb 基因组片段, *L. plantarum* P9 基因组缺失约 14.4 kb, 包含 *ssbB*、*dnaC* 和 *gpmA* 等基因。*L. plantarum* 基因组特征除 GC 含量(%)外, 其他基因组特征均有差异。本文通过比较基因组学分析发现不同 *L. plantarum* 菌株存在差异, 以期为之后 *L. plantarum* 比较分析及其生产应用奠定一定遗传学基础。

参考文献

- [1] Siezen RJ, van Hylckama Vlieg JE. Genomic diversity and versatility of *Lactobacillus plantarum*, a natural metabolic engineer. *Microbial Cell Factories*, 2011, 10(S1): S3.
- [2] De Vries MC, Vaughan EE, Kleerebezem M, De Vos VW. *Lactobacillus plantarum*—survival, functional and potential probiotic properties in the human intestinal tract. *International Dairy Journal*, 2006, 16(9): 1018–1028.
- [3] Nguyen TDT, Kang JH, Lee MS. Characterization of *Lactobacillus plantarum* PH04, a potential probiotic

- bacterium with cholesterol-lowering effects. *International Journal of Food Microbiology*, 2007, 113(3): 358–361.
- [4] Kim E, Chang HC, Kim HY. Complete genome sequence of *Lactobacillus plantarum* EM, a putative probiotic strain with the cholesterol-lowering effect and antimicrobial activity. *Current Microbiology*, 2020, 77(8): 1871–1882.
- [5] Sun ZH. Application of population genomics in the research of lactic acid bacteria. *Journal of Chinese Institute of Food Science and Technology*, 2017, 17(8): 12–18. (in Chinese)
孙志宏, 群体基因组学在乳酸菌研究中的应用. *中国食品学报*, 2017, 17(8): 12–18.
- [6] Kalinowski J, Bathe B, Bartels D, Bischoff N, Bott M, Burkovski A, Dusch N, Eggeling L, Eikmanns BJ, Gaigalat L, Goesmann A, Hartmann M, Huthmacher K, Krämer R, Linke B, McHardy AC, Meyer F, Möckel B, Pfeifferle W, Pühler A, Rey DA, Rückert C, Rupp O, Sahn H, Wendisch VF, Wiegräbe I, Tauch A. The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. *Journal of Biotechnology*, 2003, 104(1/3): 5–25.
- [7] Li CK, Ma YZ, Mi ZH, Huo R, Zhou TT, Hai HC, Kwok LY, Sun ZH, Chen YF, Zhang HP. Screening for *Lactobacillus plantarum* strains that possess organophosphorus pesticide-degrading activity and metabolomic analysis of phorate degradation. *Frontiers in Microbiology*, 2018, 9: 2048.
- [8] Zhong Z, Sun TS, Chen YF. Genomic insights into the high exopolysaccharides-producing bacterium *Streptococcus thermophilus* ND-07. *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. *中国乳品工业*, 2018, 46(4): 9–11,21.
- [9] Hunt M, De Silva N, Otto TD, Parkhill J, Keane JA, Harris SR. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biology*, 2015, 16: 294.
- [10] Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology*, 2007, 57(1): 81–91.
- [11] Chen JP, Yang XW, Chen JW, Cen Z, Guo CY, Jin T, Cui YJ. SISP: a fast species identification system for prokaryotes based on total nucleotide identity of whole genome sequences. *Infectious Diseases and Translational Medicine*, 2015, 1(1): 30–55.
- [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] Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, 2004, 14(7): 1394–1403.
- [15] Chen C, Ren J, Zhou FF, Liu ZM, Guo BH. Comparative genomic analysis of *Lactobacillus plantarum*. *China Biotechnology*, 2013, 33(12): 35–44. (in Chinese)
陈臣, 任婧, 周方方, 刘振民, 郭本恒. 植物乳杆菌的比较基因组学研究. *中国生物工程杂志*, 2013, 33(12): 35–44.
- [16] Sun ZH, Harris HMB, McCann A, Guo CY, Argimón 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, Ross RP, 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: 8322.
- [17] 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. 内蒙古农业大学博士学位论文, 2018.
- [18] Zhong Z, Kwok LY, Hou QC, Sun Y, Li WC, Zhang HP, Sun ZH. Comparative genomic analysis revealed great plasticity and environmental adaptation of the genomes of *Enterococcus faecium*. *BMC Genomics*, 2019, 20(1): 602.
- [19] Song XM, Li HB, Du LX. Comparative genomics: research fields and applications. *Chemistry of Life*, 2006, 26(5): 425–427. (in Chinese)
宋雪梅, 李宏滨, 杜立新. 比较基因组学及其应用. *生命的化学*, 2006, 26(5): 425–427.
- [20] Liu CJ, Wang R, Gong FM, Liu XF, Zheng HJ, Luo YY, Li XR. Complete genome sequences and comparative genome analysis of *Lactobacillus plantarum* strain 5-2 isolated from fermented soybean. *Genomics*, 2015, 106(6): 404–411.
- [21] Li WX, Yang L, Nan WL, Lu J, Zhang SW, Ujiroghene OJ, Pang XY, Lv JP. Whole-genome sequencing and genomic-based acid tolerance mechanisms of *Lactobacillus delbrueckii* subsp. *bulgaricus* LJJ. *Applied Microbiology and Biotechnology*, 2019, 104(6): 1–12.
- [22] Wang Y, Zhao J, Sun ZH, Sun TS, Zhang HP. Comparative genomics of genetic diversity and defense system in

- Streptococcus thermophilus*. *Acta Microbiologica Sinica*, 2020, 60(5): 924–938. (in Chinese)
- 王宇, 赵洁, 孙志宏, 孙天松, 张和平. 基于比较基因组学分析嗜热链球菌的遗传多样性和防御系统. *微生物学报*, 2020, 60(5): 924–938.
- [23] Liu YH, Li WC, Yu ZJ, Zhao FY, Zhao J, Sun ZH, Zhang HP. Population genetics of *Lactobacillus plantarum* isolated from different environments. *Microbiology China*, 2019, 46(12): 3388–3401. (in Chinese)
- 刘亚华, 李伟程, 余中节, 赵飞燕, 赵洁, 孙志宏, 张和平. 不同分离源植物乳杆菌的群体基因组分析. *微生物学通报*, 2019, 46(12): 3388–3401.
- [24] Hao P, Zheng HJ, Yu Y, Ding GH, Gu WY, Chen ST, Yu ZH, Ren SX, Oda M, Konno T, Wang SY, Li X, Ji ZS, Zhao ZP. Complete sequencing and pan-genomic analysis of *Lactobacillus delbrueckii* subsp. *bulgaricus* reveal its genetic basis for industrial yogurt production. *PLoS One*, 2011, 6(1): e15964.

Comparative genomics revealed genetic characteristics of different *Lactobacillus plantarum* strains: using P9 and Lp-6 strains as examples

Wenjun Liu, Ruirui Lv, Weicheng Li, Zhihong Sun, Yongfu Chen, Heping Zhang*

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 010018, Inner Mongolia Autonomous Region, China

Abstract: [Objective] This study aimed to analyze the genetic characteristics of *Lactobacillus* (*L.*) *plantarum* P9 and Lp-6, to provide information for future identification of *L. plantarum*. [Methods] Whole genome sequencing of *L. plantarum* strains P9 and Lp-6 was performed with PacBio SMRT sequencing and comparatively analyzed together with 110 genome sequences to identify characteristics usable in differentiating between *L. plantarum* strains. [Results] The genome length of P9 and Lp-6 strains was 3314.2 kb and 3482.5 kb, respectively. The GC content (%) of Lp-6 and P9 strains was 44.38% and 44.32%, and they contained 8 and 9 plasmids, respectively. Phylogenetic analysis showed a close genetic distance between *L. plantarum* P9 and ATCC 14917, and Lp-6 was closer to the ancestral branch. Comparing with *L. plantarum* WCSF1, the 22 kb-genome fragment of *xerS* was inverted in Lp-6 and was missing in P9 strain. A 13 kb-genomic fragment with the *tagF* gene was inserted in the Lp-6 genome; a 14.4 kb-gene fragment (containing the *gpmA* gene) was inserted in the P9 genome. [Conclusion] Through comparative genomic analysis of P9 and Lp-6 together with other *L. plantarum*, some specific genomic features were identified, which could be used for differentiation between strains.

Keywords: *Lactobacillus plantarum*, whole genome sequencing, genome atlas, comparative genomics

(本文责编: 张晓丽)

Supported by the Science and Technology Major Projects of Inner Mongolia Region (zdzx2018018), by the Inner Mongolia Science and Technology Project (201802063) and by the China Agriculture Research System of MOF and MARA

*Corresponding author. Tel: +86-471-4300593; E-mail: hepingdd@vip.sina.com

Received: 28 August 2020; Revised: 3 October 2020; Published online: 22 October 2020