



## 不同来源植物乳杆菌基因组结构和功能差异的比较研究

刘超楠<sup>1</sup>, 王金铭<sup>1</sup>, 南韦肖<sup>2</sup>, 司华哲<sup>2</sup>, 张新宇<sup>1</sup>, 蔡熙烜<sup>1</sup>, 刘晗璐<sup>1\*</sup>

<sup>1</sup> 中国农业科学院特产研究所, 吉林 长春 130112

<sup>2</sup> 吉林农业大学动物科学技术学院, 吉林 长春 130118

**摘要:**【目的】本试验研究不同来源植物乳杆菌(*Lactobacillus plantarum*)基因特点以及在不同环境下其基因多样性, 探究 2 株 *L. plantarum* A8 和 P9 在肠道生境及植物表面适应性的异同, 为优良菌株的开发提供理论基础。【方法】本研究对从动物肠道和植物表面分离获得的 *L. plantarum* A8 和 *L. plantarum* P9 的基因组进行分析, 利用第二代测序技术(NextGeneration Sequencing, NGS), 基于 Illumina NovaSeq 测序平台, 同时利用第三代单分子测序技术, 基于 PacBio Sequel 测序平台, 对 *L. plantarum* A8 和 *L. plantarum* P9 进行测序。采用 Carbohydrate-active enzymes (CAZy)、Koyto encyclopedia of genes and genomes (KEGG)和 Clusters of orthologous genes (COG)数据库对基因组进行功能注释; 采用 CGView 软件绘制菌株的基因组环形图谱。应用比较基因组学与已经公开发表的其他 *L. plantarum* 基因组进行比较分析。【结果】由研究可知 *L. plantarum* A8 和 *L. plantarum* P9 基因组大小存在差异, 通过构建系统发育树发现 2 株菌与其他来源的 *L. plantarum* 分在同一分支, 并且 *L. plantarum* P9 与母乳来源的 *L. plantarum* WLPL04 菌株距离最近, 而 *L. plantarum* A8 与 *L. paraplantarum* DSM10667 距离最近。通过基因家族分析可知, 2 株菌共有基因为 2643 个, 其中包括一些抗应激蛋白如热休克蛋白、冷休克蛋白。*L. plantarum* A8 和 P9 独特基因分别为 321 和 336 个, *L. plantarum* A8 中独特基因主要参与 DNA 复制、ABC 转运系统(ABC transfer system)、PTS 系统(phosphotransferase system)、磷酸盐转运系统、氨基酸生物合成等代谢通路; *L. plantarum* P9 的独特基因以参与碳水化合物的运输和代谢基因居多, 例如 *rpiA* 基因、*lacZ* 基因、*FruA* 基因等。【结论】通过比较基因组学方法解析 *L. plantarum* 的基因组信息, 发现动物肠道来源的 *L. plantarum* 具有较好的氨基酸转运能力, 植物表面附着的 *L. plantarum* 菌株具有较好碳水化合物利用能力, 从而为益生菌的开发与利用提供理论依据。

**关键词:** 植物乳杆菌, 比较基因组学, 基因组多样性

基金项目: 国家重点研发计划(2018YFC1706602); 吉林省科技发展计划(20200301019RQ)

\*通信作者。Tel: +86-431-81919856; E-mail: liuhanlu@caas.cn

收稿日期: 2020-12-28; 修回日期: 2021-02-08; 网络出版日期: 2021-08-04

*L. plantarum* 是一种兼性厌氧乳杆菌, 形态呈圆端短杆状, 无芽孢, 革兰氏染色呈阳性, 触酶阴性, 能够在 10–45 °C、pH 4.5–9.5 范围内生长, 最适生长条件为 30–35 °C、pH 6.5 左右, 属兼性异型乳酸发酵<sup>[1]</sup>。*L. plantarum* 来源安全, 拥有广泛的代谢多样性, 被用作各种发酵食品中的功能性发酵剂, 如奶酪、植物性食品、肉制品、葡萄酒和啤酒等<sup>[2]</sup>。大量研究表明, 健康成人摄取 *L. plantarum* 后可增强个体的获得性免疫<sup>[3]</sup>, *L. plantarum* 可通过 NF- $\kappa$ B 和 P38 MAPK 代谢信号通路调控促炎性细胞因子, 进而发挥正向的免疫调节作用<sup>[4]</sup>。*L. plantarum* 还能够使过敏原原肌球蛋白免疫活性减弱, 从而为研究食物过敏原的活性控制及低致敏性水产加工制品的开发提供理论参考<sup>[5]</sup>。另外 *L. plantarum* EPS-YW11 在加工乳制品中具有用作功能性试剂的潜力, 这种乳制品能够增强结构稳定性和生物活性, 例如降低胆固醇、抗氧化剂和抗生物膜<sup>[6]</sup>。因此, *L. plantarum* 对食品工业和人体健康有着多重有益的功效, 具有很高的应用价值。

比较基因组分析是生物信息学中的一种常见工具, 能够识别菌株及其来源之间的联系, 并评估特定物种的基因分布<sup>[7]</sup>, 对不同生境 *L. plantarum* 进行基因组学分析, 不仅可揭示菌株本身的遗传特性, 对深入了解并挖掘其生物学功能也具有重要意义。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。本研究通过进行全基因组测序获得 *L. plantarum* A8 和 P9 详细和准确的基因组遗传信息, 对 2 株细菌差异性进行比较, 并与其他不同分离生境的同种菌株进行比较分析, 一方面可以

丰富 *L. plantarum* 基因组信息, 另一方面为判断 *L. plantarum* A8 和 P9 是否具有工业应用益生菌的潜力提供参考。

## 1 材料和方法

### 1.1 实验用菌株及序列信息

本实验室自主分离保藏的 *L. plantarum* A8、P9 分别分离自中国农业科学院特产研究所毛皮动物实验基地水貂肠道和内蒙古正蓝旗育草站披碱草植株。参与比较基因组的信息均从 GenBank 数据库中获取, 包括其他不同来源的 8 株 *L. plantarum* 序列以及 12 株乳杆菌的基因组信息, 具体信息见表 1。

### 1.2 基因组测序和草图组装

利用第二代测序技术, 基于 Illumina NovaSeq 测序平台进行测序<sup>[8]</sup>。再进行第三代单分子测序数据拼装: 将 Pacbio 获得的下机数据使用 HGAP v4、CANU v1.7.1 软件进行拼装, 得到 contig 序列<sup>[9]</sup>。将二代的高质量数据(high quality reads, HQ reads)使用 pilon v1.18 软件对三代 contig 结果进行校正, 最终拼接得到完整序列<sup>[10]</sup>。

### 1.3 基因组特征预测

采用 GeneMarkS v4.32 软件对全基因序列进行基因预测<sup>[11]</sup>, 采用 tRNAscan-SE v1.3.1、Barrnap v0.9 预测全基因组中的 tRNA 和 rRNA 基因<sup>[12]</sup>。再利用 CRISPR finder v20170509<sup>[13]</sup>和 IslandViewer 4<sup>[14]</sup>预测全基因组中的移动原件 DRs (正向重复序列)、Spacers(间隔区)和基因岛。使用 COG<sup>[15]</sup>、KEGG<sup>[16]</sup>、GO<sup>[17]</sup>对细菌基因组进行快速注释, 并得到相应的注释信息。最后采用 CGView v1.1.0.50 绘制该基因组的圈图, 用 Photoshop CS v8.0 对该图进行编辑<sup>[18]</sup>。

表 1. 参与比较基因组学分析的所有菌株信息  
Table 1. Information on all strains involved in comparative genomics analysis

Strain	Size/Mb	Origin	Proteins	Level	BioSample
<i>Lactobacillus salsicarnum</i> TMW 1.2098	2.42824	Sausage	5194	Scaffold	SAMN11864207
<i>Lactobacillus halodurans</i> TMW 1.2172	2.84476	Sausage	3253	Contig	SAMN11864209
<i>Lactobacillus plantarum</i> WLPL04	3.141 59	Breast milk	2881	Complete genome	SAMN04092618
<i>Lactobacillus plantarum</i> -plantarum	3.24663	Intestine	2865	Complete genome	SAMN02603936
<i>Lactobacillus murinus</i> CR147	2.29045	Mouse gut	38225	Complete genome	SAMN07690924
<i>Lactobacillus quenuiae</i> HV_6	1.57521	Bee	1509	Contig	SAMN08331817
<i>Lactobacillus plantarum</i> LZ95	3.32246	Human feces	3026	Complete genome	SAMN03854682
<i>Lactobacillus plantarum</i> A8	3.03756	Mink Gut	2864	Complete genome	CP058733–CP058739
<i>Lactobacillus plantarum</i> WCFS1	3.34862	Human saliva	3 063	Complete genome	SAMEA3138345
<i>Lactobacillus plantarum</i> DF	3.69731	Drosophila	3430	Complete genome	SAMN04346870
<i>Lactobacillus plantarum</i> ST-III	3.30794	Pickle	3020	Complete genome	SAMN02603896
<i>Lactobacillus plantarum</i> ATCC 8014	3.30947	Corn silage	3030	Complete genome	SAMN07837328
<i>Lactobacillus paraplantarum</i> DSM10667	3.39575	Beer	6217	Scaffold	SAMN02369508
<i>Lactobacillus musae</i> NBRC 112868	2.35779	Banana	2291	Contig	SAMD00154527
<i>Lactobacillus plantarum</i> 5-2	3.23765	Fermented soybeans	2991	Complete genome	SAMN02953961
<i>Lactobacillus pontis</i> LP475	1.71477	Rye sourdough	8112	Complete genome	SAMN12784819
<i>Lactobacillus pentosiphilus</i> IWT25	2.72391	Silage	9162	Scaffold	SAMD00042398
<i>Lactobacillus curieae</i> CCTCC M2011381	2.09586	Brine of stinky tofu	8029	Complete genome	SAMN03175170
<i>Lactobacillus kefir</i> DH5	2.48903	Yogurt	17317	Complete genome	SAMN09381536
<i>Lactobacillus mali</i> LM596	2.61195	Environment	2341	Complete genome	SAMN12784790
<i>Lactobacillus harbinensis</i> M1	3.59219	Naturally fermented tofu whey	3082	Complete genome	SAMN09915110
<i>Lactobacillus plantarum</i> P9	3.18547	Grass	2950	Complete genome	CP058675–CP058681

## 1.4 比较基因组学分析

**1.4.1 系统发育比较:** 基于 16S rRNA 利用 MEGA 7.0 软件构建邻接树(Neighbor-joining tree, NJ), 参数 Bootstrap 值设为 1000。

**1.4.2 泛基因和核心基因组分析:** 首先, 下载参考基因组的蛋白质序列, 根据蛋白质序列长度进行筛选, 去除序列长度低于 50 个氨基酸的序列。将所有待分析的蛋白质序列合并成一个文件, 以该数据集为基础构建数据库, 同时以该数据集作为查询, 进行 all-VS-all blastp 分析, 序列比对的阈值设置为  $1e-10$ 。将序列比对的结果采用 orthomcl (version 2.0.8) 软件进行处理, 其中序列比

对的长度设置为 70%, 采用 MCL 对基因家族进行聚类, 聚类所采用的 I (Inflation) 设置为 1.5<sup>[19]</sup>。最后, 采用自制的 Perl 脚本对聚类得到的结果进行整理和统计。

**1.4.3 全基因组序列比对:** 下载参考基因组序列, 为了使所有基因组的起始位点能对齐, 首先以其中一个基因组为参考, 对其他所有基因组的起点序列进行调整。然后采用 Mauve (version 2.3.1) 构建获得该基因组和参考基因组的序列比对结果<sup>[20]</sup>。

下载参考基因组序列, 采用 MUMmer (version 3.22) 软件构建两株菌的全基因组序列进行比对<sup>[21]</sup>。

## 2 结果和分析

### 2.1 基因组测序和草图组装

*L. plantarum* A8 和 *L. plantarum* P9 经二代测序获得高质量 reads 数目分别为 7894222 个和 7412316 个。三代测序获得总序列数量分别为 173702 个和 197466 个, 基因组长度分别为 1428045177 bp 和 1617664031 bp, 具体信息如表 2。

### 2.2 基因组特征预测

**2.2.1 基因组基本信息:** 根据 *L. plantarum* 研究进展可知全基因组大小为 2.91–3.70 Mb, GC 百分含量大约为 44.2%–45.1%, 基因数目从 2860 到 3599 不等, 通过测序的组装信息可以看出 A8 和 P9 符合 *L. plantarum* 基因组特征。同时本试验发现两者之间存在差异, 即 *L. plantarum* P9 的基因组较 *L. plantarum* A8 大并且其编码的基因数量较 A8 多

86 个, 而 *L. plantarum* A8 的 GC 含量和 tRNA 数量均比 *L. plantarum* P9 多, 统计详情见表 3。

**2.2.2 短回文重复序列分析:** 短回文重复序列 (clustered regularly interspaced short palindromic repeats, CRISPRs) 重复片段的间隔区可以整合到宿主基因组中, 同时也可作为一种记忆元件<sup>[22]</sup>。本次在样品 A8 中共预测到 2 条存疑的 CRISPRs 结构, 总长度为 215 bp, 而在样品 P9 中共预测到 1 条置信的 CRISPRs 结构, 长度为 1517 bp, 详细信息见表 4。

表 2. 二、三代测序数据序列统计

Table 2. Two and three generation sequencing data sequence statistics

Sample	HQ reads	Total sequence number	Total sequence length/bp
A8	7894222	173702	1428045177
P9	7412316	197466	1617664031

表 3. A8、P9 基因组基本信息

Table 3. Basic information about the genome of A8 and P9

Sort	Origin	Gene length	Sequence type	G+C content	Open reading frame	rRNA	tRNA	ncRNA	Plasmid
A8	Animal	3037561 bp	circular	44.81%	2864	16	71	0	6
P9	Plant	3185467 bp	circular	44.65%	2950	16	67	155	6

表 4. 短回文重复序列基本信息比对

Table 4. Comparison of basic information of clustered regularly interspaced short palindromic repeats

CRISPRs	A8	P9
Number of CRISPRs	2	1
Seq ID	chr	chr
CRISPRs type	Questionable	Confirmed
Length/bp	215	1517
Genome/%	0.0071	0.0476

**2.2.3 基因岛分析:** 利用 IslandViewer 4 软件对 2 株 *L. plantarum* 基因岛进行分析, 基因岛通常与基因的获得和丢失有关, 一些获得的基因聚集在一起形成一个原本不属于自己的短区域, 在这些区域中通常发现适应性功能<sup>[23]</sup>。由图 1 可知, A8 中含有 15 个基因岛, P9 中含有 18 个基因岛。并且 2 株试验菌基因岛所在位置稍有不同, 但主要参与复制和修复、氨基酸代谢、翻译等基本功能。如图 1 所示 A 为 A8 基因岛圈图, B 为 P9 基因岛圈图。

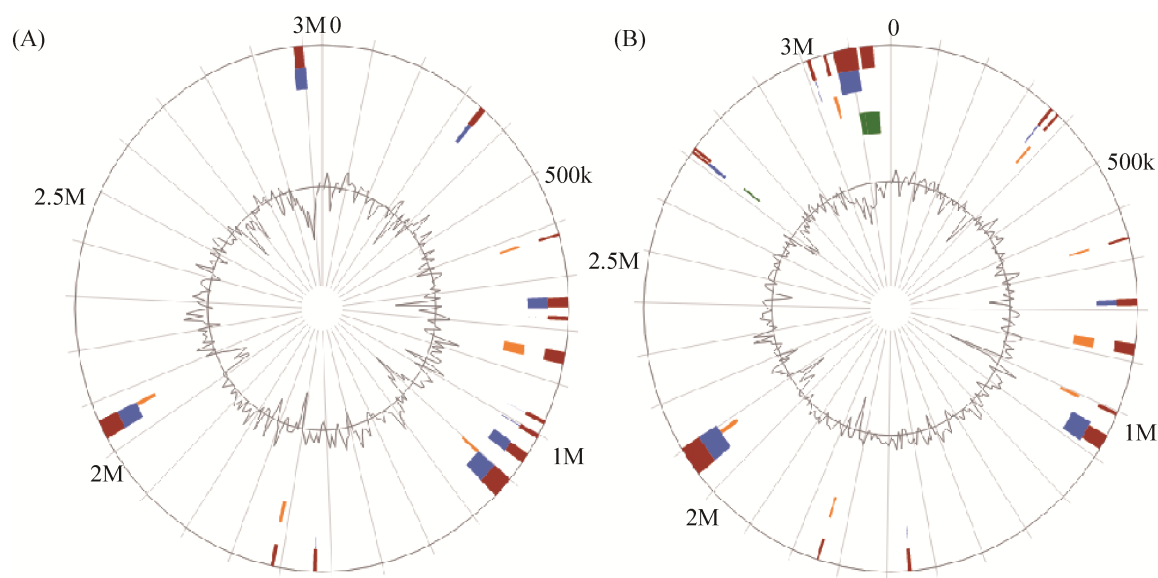


图 1. 基因岛圈图比较

Figure 1. Comparison of gene Island circles. A: the A8 gene island circle map; B: the P9 gene island circle map. Red represents the predicted by at least one method, blue represents that the results of IslandPath-DIMOB predicted, yellow represents the predicted results of SIGI-HMM, green represents the predicted results of IslandPick, light blue represents Islander's forecast results.

## 2.3 比较基因组学分析

**2.3.1 进化树分析:** 为了探究 2 株菌与其他 12 株乳杆菌之间的进化关系, 将本实验室所得的 2 株菌与其他不同来源的 *L. plantarum*, 以及其他乳杆菌的 16S rRNA 构建系统发育树(图 2)。发现 2 株菌与其他 *L. plantarum* 在同一分支水平上, 其中菌株 A8、P9 与菌株 WLPL04 和菌株 DSM10667 有最接近的亲缘关系。

**2.3.2 不同来源 *L. plantarum* 韦恩图分析:** 泛基因组主要由核心基因、辅助基因和特异基因组成<sup>[24]</sup>。将本实验室分离所得的 2 株 *L. plantarum* 与 GenBank 中所选择的 8 株 *L. plantarum* 进行分析, 10 株菌平均基因数为 3178 个, GC 百分含量为 44.1%–44.8%。10 株菌共有的基因为 2250 个, A8 独特的基因有 167 个, P9 独特的基因 190 个

(图 3)。十菌株各自的独特基因以参与复制、重组和修复功能的基因为主, 但 A8 还含有参与转录过程的抗阻蛋白, 以及 ABC 超家族 ATP 结合盒转运蛋白膜蛋白。P9 则含有自己独特的 PTS 系统组成成分, 用于碳水化合物的运输。

**2.3.3 不同来源 *L. plantarum* COG 注释分析:** 十株 *L. plantarum* 基因组大部分基因参与转录(10%)、氨基酸转运和代谢(8%)、碳水化合物运输和代谢(8%)途径, 其余基因多参与复制、重组与修复(5%)、细胞壁/膜/包膜生物发生(6%)、核苷酸转运和代谢(5%)等过程。结果表明核心基因所参与的功能大多与细胞生长增殖密切相关, 推测这些基因可能是在细菌长期进化过程中, 在环境的作用下获得了稳定遗传(图 4)。

**2.3.4 菌株 A8、P9 共线性分析:** Mauve 分析区

块的上、下代表正义链和反义链, 相近颜色区块代表推测的同源区块。分析发现内部无基因组重排, 仅有少量的易位, 表明菌株在进化过程中发生过小范围的基因重组和转移。共线性主要反应

了基因组间的结构性变异, 反映出基因组之间的编码顺序和结构的同源性<sup>[25]</sup>。通过上述共线性分析, 可以发现这两株菌基因的序列和顺序非常相似, 具有很好的同源性和共线性(图 5)。

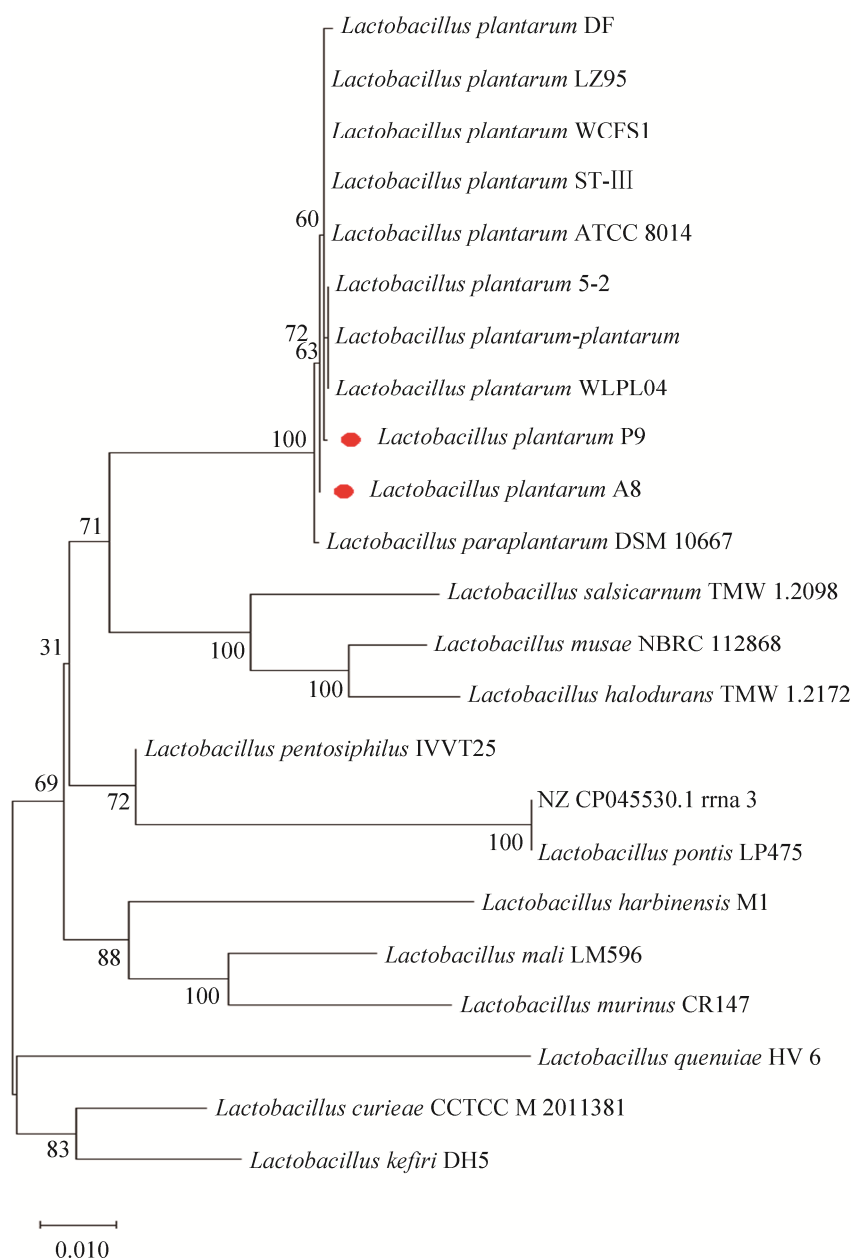


图 2. 16S rRNA 进化树分析

Figure 2. The construction of a phylogenetic tree based on the 16S rRNA sequence can be roughly divided into 5 main branches. The red markers are the strains isolated in our laboratory, and the rest of the reference strain information comes from the GenBank database.

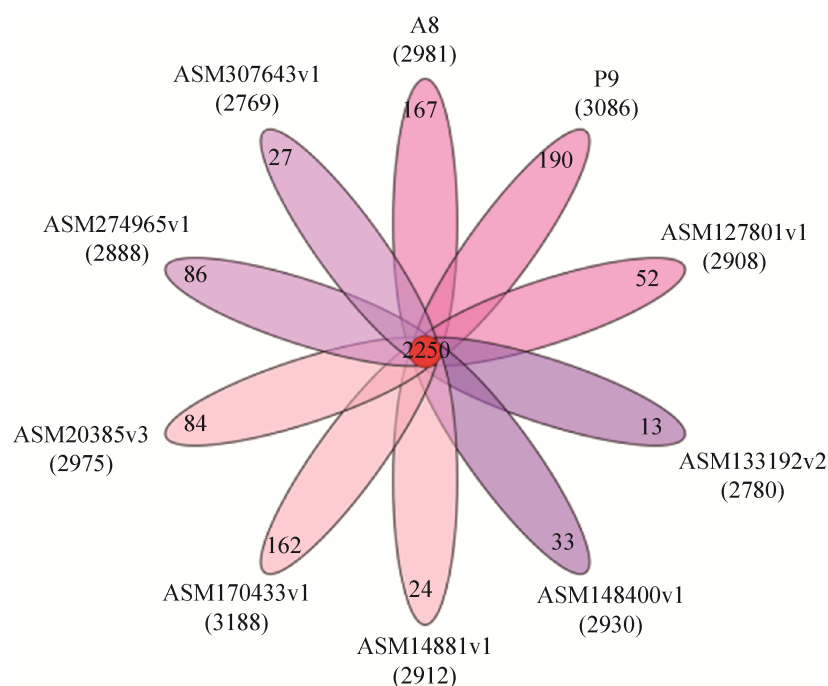


图 3. 十株菌基因家族分析

Figure 3. Gene family analysis of ten strains. The red part of the figure represents the core genes of all strains, and the remaining colors represent the number of unique genes of each strain.

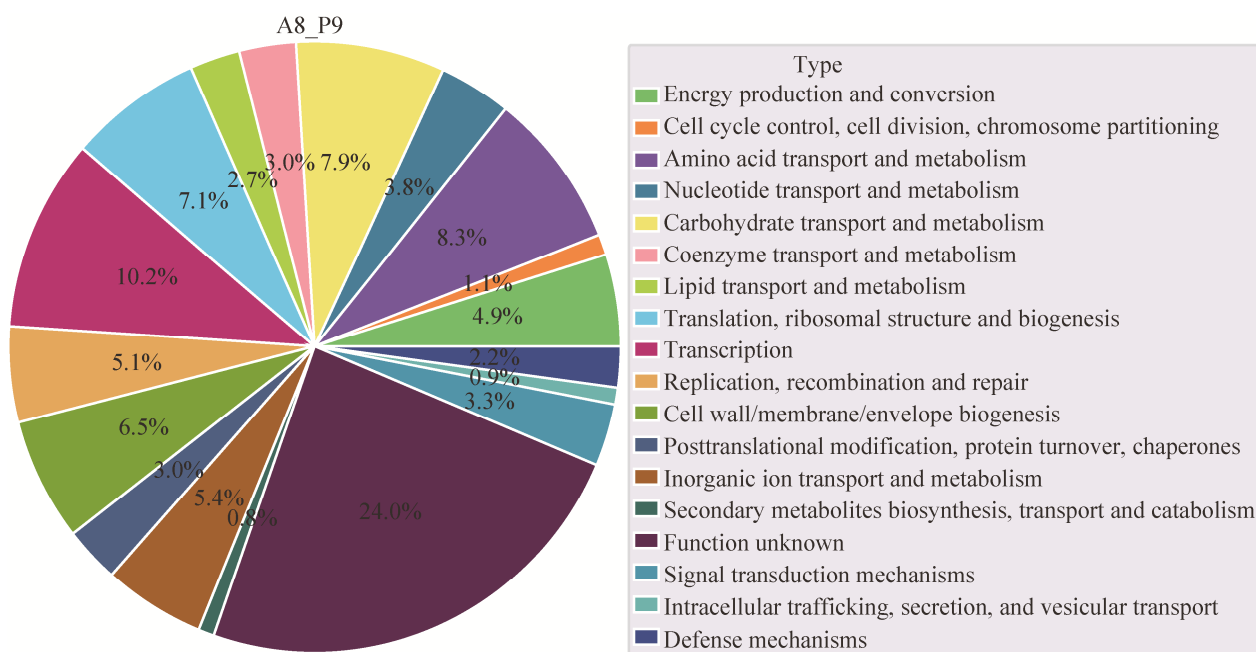


图 4. 十株菌 COG 注释饼图

Figure 4. As the picture shows that the COG annotation pie chart of ten strains, each color in the figure represents the proportion of each function in the genome.



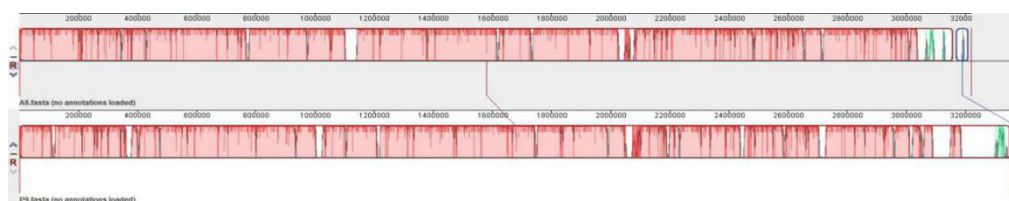


图 5. Mauve 分析

Figure 5. Mauve analysis. The upper block represents strain A8, the lower block represents strain P9.

**2.3.5 菌株 A8、P9 基因家族分析：**两株菌总共预测到 6020 个基因，其中 A8 有 2964 个基因，P9 菌株有 3056 个基因。它们共有的基因有 2643 个，菌株 A8、P9 各自独特基因分别为 321 和 413 个(图 6)。

**2.3.6 菌株 A8、P9 基因组 COG 注释：**如图 7-A 所示，菌株 A8、P9 的核心基因除占比最大的功能未知基因外，其中 9.8%基因参与转录(K)、8.9%参与碳水化合物的运输和代谢(G)、氨基酸转运和代谢(E)占比为 7.8%。由此可知已知功能的核心基因主要参与物种的基本生物学功能和主要表型特

征。核心基因还包括一些抗应激蛋白，比如热休克蛋白(heat shock proteins, HSPs)和冷休克蛋白(cold shock protein, Csp)，它是连续合成并表达以克服冷休克期间冷休克的有害作用的蛋白质<sup>[26]</sup>(表 5)。并且在两株菌的质粒中发现具有相同功能的核心基因，比如参与 DNA 修复和重组的基因 *umuC*，以及锰离子转运蛋白 *MntH*。

A8 基因组的 321 个独特基因注释到 COG 的 16 个功能类别中，除未知功能基因外，参与复制、重组和修复的基因占 22.7%，另参与转录和防御机制的分别占 8.8%和 6.4%，如图 7-B 所示。再通过 KEGG 数据库注释后发现其独特基因主要参与 DNA 复制、ABC 转运系统、PTS 系统、磺酸盐转运系统、氨基酸生物合成等代谢通路，其中包括博来霉素水解酶(表 6)。并且通过 KEGG 注释到功能的基因包括以 *antB*、*ssbL* 为代表的 27 个特有基因(表 7)。

P9 独特基因注释到 COG 数据库中的基因有 336 个，除未知功能基因外，其中以参与碳水化合物的运输和代谢的基因居多，有 62 个，占基因总数的 18.4%(表 6)。而参与复制、重组和修复的基因有 56 个，占 16.7%，参与转录的有 37 个基因，占比 11%，如图 7-C 所示。将其在 KEGG 数据库注释后发现这些基因主要参与磷酸肌醇代谢、磷酸戊糖途径、PTS 系统、蛋白运输等代谢通路。通过 KEGG 注释到功能的基因包括以 *iolG*、*cbi* 为代表的 32 个特有基因(表 7)。

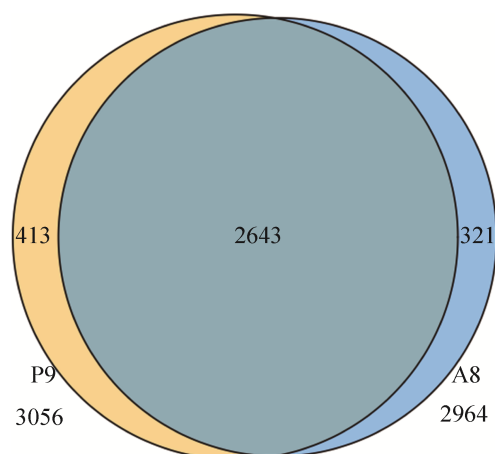


图 6. 基因家族分析

Figure 6. Analysis of gene family. The overlapping part of the figure represents the shared genes of the two strains. The yellow part on the left is the unique gene of strain P9, and the blue part on the right is the unique gene of strain A8.



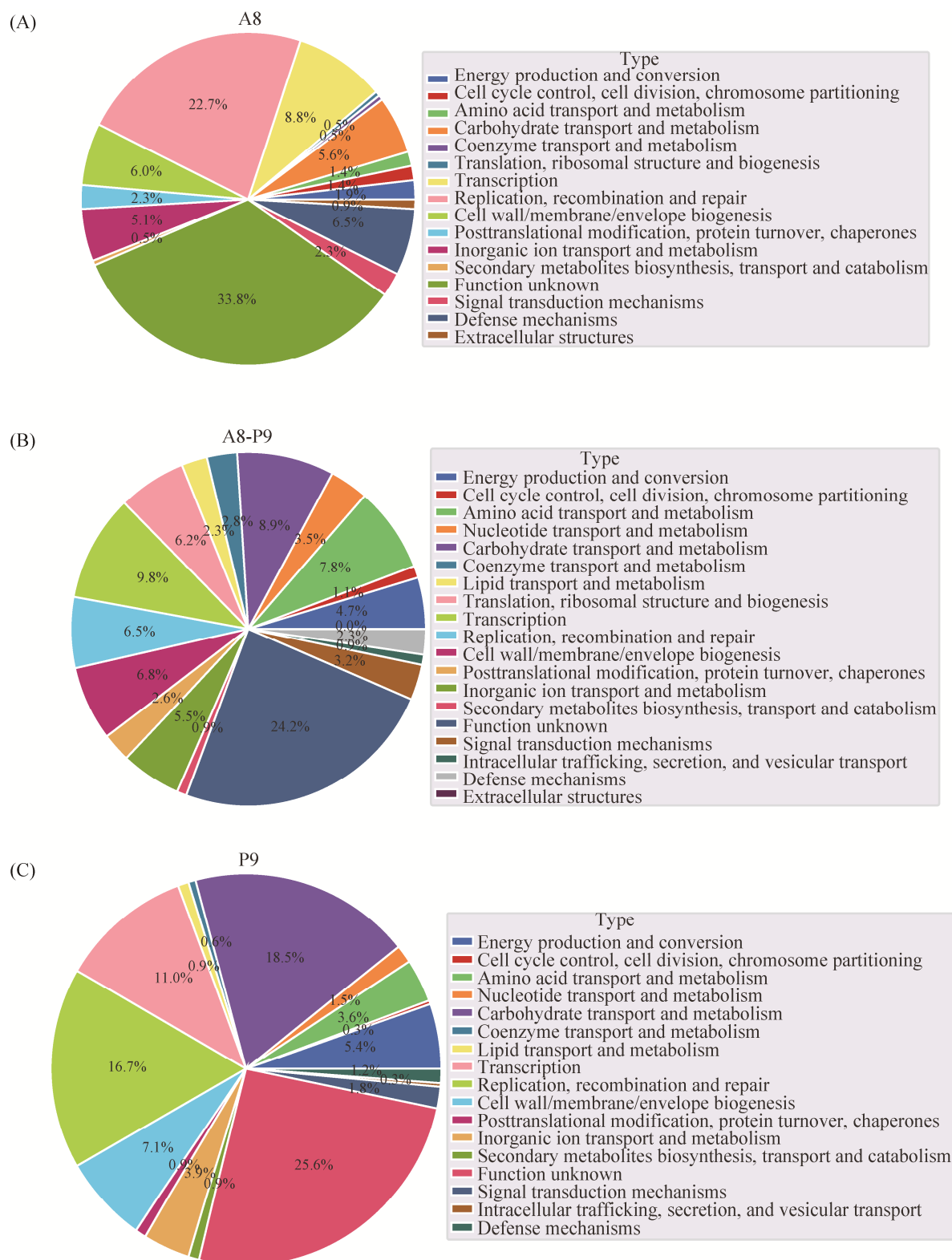


图 7. 基因 COG 注释图

Figure 7. COG annotation map of gene. A: the COG annotation of the core genes of the two strains; B: the unique gene annotation of the A8 genome; C: the unique gene annotation of the P9 genome.

表 5. 菌株 A8、P9 核心基因

Table 5. Core genes of strain A8 and P9

Gene name	Type of gene	Gene copy number
<i>HSPs</i>	Heat shock proteins	1
<i>cspA</i>	Cold shock protein	3
<i>umuC</i>	DNA polymerase V	1
<i>MntH</i>	Manganese transport protein	1

表 6. 菌株 A8、P9 独特基因 KEGG 注释

Table 6. Unique gene KEGG annotation of strain A8 and P9

KEGG	Number of genes	
	A8	P9
Transporter	5	4
Sulfur metabolism	10	6
DNA replication	3	4
Replication and repair	2	5
Carbohydrate metabolism	6	21

表 7. 菌株 A8、P9 独特基因

Table 7. Unique genes of strain A8 and P9

A8	Number of genes	P9	Number of genes
<i>bdh</i>	1	<i>cbi</i>	2
<i>ssuA</i>	1	<i>cob</i>	1
<i>NitT</i>	1	<i>hem</i>	2
<i>antB</i>	2	<i>cys</i>	1
<i>ssb</i>	2	<i>rpiA</i>	1
<i>yxeP</i>	1	<i>lacZ</i>	1
<i>coiA</i>	1	<i>sgaA</i>	1
<i>xtmB</i>	1	<i>FruA</i>	1
<i>recT</i>	1	<i>iolG</i>	3
<i>mscL</i>	1	<i>iolT</i>	1

### 3 讨论

*L. plantarum* 是一种在自然界广泛存在的典型的革兰氏阳性细菌, 已被广泛应用于生产和科学研究。*L. plantarum* A8 和 P9 全基因组序列的获得不仅有助于丰富基因组数据库, 还将为探索不同生境的同种细菌的应用提供理论依据。CRISPRs 是一类独特的 DNA 正向重复序列家族,

广泛存在于原核生物基因组中, 目前的研究认为它能够给宿主提供某种获得性免疫<sup>[27]</sup>。CRISPRs 序列及其相关基因构成抵御外来入侵者重要的防御系统, 在抵抗外源 DNA 感染中发挥作用<sup>[28]</sup>。本试验通过全基因组测序发现 *L. plantarum* A8 和 P9 均含有 CRISPR 序列, 这可能说明以上 2 株菌在应对噬菌体感染及质粒接合转移时可以发挥积极作用<sup>[29]</sup>。基因水平转移是细菌进化的重要表现, 是细菌获得外源性功能基因、应对选择压力的另一种方式<sup>[30]</sup>。菌株 A8 和 P9 分别存在 15 和 18 个基因岛, 并且在两者基因岛区域中均存在寡肽 OPP 转运蛋白(oligopeptide ABC transport)基因簇, 而这种基因簇广泛分布于细菌、真菌和真核生物体内, 可以穿过细胞膜转运物质, 从而帮助机体更好地适应外界环境生活<sup>[31]</sup>。*L. plantarum* 基因组移动原件不同, 表明其基因组信息存在差别, 推断其差别的原因可能与生存环境有关。

本试验基于 16S rRNA 进行系统进化分析发现, 同种/不同来源的 *L. plantarum* 被聚类在一个分支, 而其他不同物种乳杆菌则聚类在其他分支。与 P9 距离最近是来源于母乳中的菌株 WLPL04, 该菌株也具有肠道生存能力(酸和胆汁盐耐受性以及模拟胃肠道中的生存能力)和益生菌特性(病原体的抑制, 病原体的抗粘附性, 以及对 Caco-2 细胞的 SDS 有害作用和炎性应激的防护)<sup>[32]</sup>。将与 2 株菌距离较近的同种不同来源的其他 10 株 *L. plantarum* 进行比较发现, 其基因组大小、G+C 百分含量表明不同来源的 *L. plantarum* 具有广泛的基因多样性。基因家族分析中核心基因主要负责物种的基本生物学功能和主要表型特征, 非必需基因和特异基因主要参与次级代谢途径<sup>[33]</sup>。本试验中菌株 A8 和 P9 的独特基因有超过

一半存在于它们各自的质粒中, 考虑到生存环境和营养物质的特殊性等情况, 推测这些特有基因非常有可能来自基因的水平转移, 或者被噬菌体侵染而获得一些特殊的基因。而 COG 注释核心基因表明, *L. plantarum* 可以利用或产生更多种类的碳水化合物, 在恶劣环境中有更好的适应性、细胞定殖和生存能力<sup>[34]</sup>。

比较 2 株 *L. plantarum* 的基因组特征, 泛基因组和核心基因集分别包含 6020、2643 个基因, 其中以 9.8% 参与转录的基因占比最大, 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。另外, 生物调控、信号传导、生物膜形成的基因也被注释到, 该类基因有利于触发菌体的群体感应, 在培养基中菌体形成了很明显的生物膜, 膜的形成加速了生物体间的物质和信息交流, 有利于细菌适应复杂的环境<sup>[35]</sup>。

两个基因组序列都包含编码热休克蛋白和冷休克蛋白的基因, 这些冷休克蛋白可能是 *L. plantarum* 菌株对胁迫条件适应的一种体现, 并且在罗伊氏乳杆菌、大肠杆菌等许多细菌基因组中均有发现<sup>[36]</sup>。此外, 2 个基因组中还涉及辅助因子和维生素代谢的基因。这些注释基因主要参与钴胺素(维生素 B12)的生物合成, 还有叶酸生物合成的代谢以及硫胺素生物合成。在这些基因中, 有一个由 2 个 *cbi* 基因(*cbiM* 和 *cbiO*)、1 个 *cob* 基因(*cobC*)、2 个 *hem* 基因(*hemH*、*hemK*)和 1 个 *cys* 基因(*cysE*、*cysK*、*cysS*、*cysC*、*cys*)组成的与钴胺素生物合成相关的基因群, 这表明 *L. plantarum* A8、P9 能够产生钴胺素, 为 2 株菌作为潜在的益生菌提供了理论依据<sup>[37]</sup>。另外研究表明质粒可以携带多种功能基因, 比如赋予抗生素抗性和有毒

重金属抗性的基因、毒力决定因子以及增强修复 DNA 损伤功能的基因<sup>[38]</sup>。而我们在两株菌的质粒中发现了参与 DNA 修复和重组的基因 *umuC*, 它具有 DNA 聚合酶功能, 从而可以修复细胞 DNA 损伤。还发现了锰离子转运蛋白(MntH), 它可以利用细胞内外的质子梯度作为驱动力, 摄取环境中的锰离子进入细胞, 并且  $\text{Fe}^{2+}$ 、 $\text{Mn}^{2+}$  等金属离子可作为辅酶参与各种氧化还原过程, 影响细菌的毒性和致病性<sup>[39]</sup>。这些结果表明抗应激蛋白和质粒中存在的损伤修复以及离子转运蛋白作为核心基因, 在细菌应对外界不良环境时发挥重要作用。

氨基酸是乳酸菌生长发育过程中必需的营养因素之一, 乳酸菌既具有从头合成氨基酸的能力, 也会从环境中摄取蛋白质后分解用以自身蛋白质的合成、代谢能的储藏和产生以及辅助因子的循环使用。而相比于其他乳杆菌菌株, *L. plantarum* 似乎不包含编码胞外蛋白酶 *Prt* 的基因<sup>[40]</sup>, 而我们的 2 株 *L. plantarum* 的基因组也不含有编码胞外蛋白酶基因 *Prt*。但 A8 基因组中的独特基因通过 KEGG 数据库注释后发现, 其独特基因含有多条参与氨基酸生物合成代谢的通路, 其中含有一种博来霉素水解酶。博来霉素水解酶是一种进化高度保守的胞质半胱氨酸肽酶, 根据作用特点, 半胱氨酸水解酶分为 3 大类: 半胱天冬酶、钙蛋白酶和木瓜蛋白酶, 而博来霉素水解酶为木瓜蛋白酶家族的重要成员<sup>[41]</sup>, 并且该酶已知的活性是使糖肽博来霉素代谢失活。除此之外, 作为细胞内的水解酶类, 还参与免疫系统的抗原呈递和水解老年痴呆 A $\beta$  致病蛋白<sup>[42]</sup>。因为食草性动物主要采食树叶、牧草或者其他植物, 只能利用非蛋白氮合成氨基酸、肽和蛋白质, 因此为了满足食

草动物的营养需求, 推测 *L. plantarum* A8 可能参与食草性动物肠道内氨基酸的生物合成。

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。

*L. plantarum* P9 基因组中超过 18.4% 的注释基因 (62 个基因) 涉及碳水化合物代谢。注释的基因涉及利用各种碳源, 例如, 编码核糖 5-磷酸异构酶 A 的 *rpiA* 基因涉及戊糖磷酸途径, 编码  $\beta$ -半乳糖苷酶的 *lacZ* 基因负责将乳糖分解为单糖涉及半乳糖代谢, 编码磷酸转移酶系统的果糖特异性通透性的 *FruA* 基因涉及果糖转运、果糖和甘露糖代谢, 从而可以使用不同的碳源如戊糖、半乳糖、果糖和蔗糖。*L. plantarum* P9 含有丰富的碳水化合物利用基因可能与菌株所处环境有关, 表明具有较强的糖合成能力。

采用高通量测序技术对 *L. plantarum* A8 和 P9 进行全基因组测序, 菌株 A8、P9 进行比较基因分析, 这两株菌基因的序列和顺序非常相似, 具有很好的同源性和共线性。独特基因功能注释发现 *L. plantarum* A8 在氨基酸代谢方面具有明显优势, 而 *L. plantarum* P9 则在碳水化合物利用方面表现突出。本试验结果提供了关于在该物种中观察到的遗传多样性信息, 两株菌独特基因的不同显示了菌株基因组对于外界环境的适应性变化, 并且两株菌具有蛋白水解和氨基酸代谢能力, 为筛选具有优质应用型益生菌添加剂菌株提供理论基础。

## 参 考 文 献

- [1] Widodo W, Ariani AL, Widiyanto D, Haltrich D. Genomic comparison of *Lactobacillus casei* AP and *Lactobacillus plantarum* DR131 with emphasis on the butyric acid biosynthetic pathways. *Microorganisms*, 2021, 9(2): 425.
- [2] Jeon S, Jung J, Kim K, Yoo D, Lee C, Kang J, Cho K, Kang DK, Kwak W, Yoon SH, Kim H, Cho S. Comparative genome analysis of *Lactobacillus plantarum* GB-LP3 provides candidates of survival-related genetic factors. *Infection, Genetics and Evolution*, 2017, 53: 218–226.
- [3] Levit R, Savoy de Giori G, de Moreno de LeBlanc A, LeBlanc JG. Protective effect of the riboflavin-overproducing strain *Lactobacillus plantarum* CRL2130 on intestinal mucositis in mice. *Nutrition*, 2018, 54: 165–172.
- [4] Xiao Y, Ye JQ, Li XC, Li XH, Shi WZ, Lu Y. Study on the reduction effect of *Lactobacillus plantarum* on the immunological activity of tropomyosin. *Food Science*, 2021, 3(16): 1–15. (in Chinese)  
肖叶, 叶精勤, 李晓晨, 李晓晖, 施文正, 卢瑛. 植物乳杆菌(*Lactobacillus plantarum*)对过敏原原肌球蛋白免疫活性的消减作用. *食品科学*, 2021, 3(16): 1–15.
- [5] Nambiar RB, Sellamuthu PS, Perumal AB, Sadiku ER, Phiri G, Jayaramudu J. Characterization of an exopolysaccharide produced by *Lactobacillus plantarum* HM47 isolated from human breast milk. *Process Biochemistry*, 2018, 73: 15–22.
- [6] Chin CS, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, Cramer GR, Delledonne M, Luo CY, Ecker JR, Cantu D, Rank DR, Schatz MC. Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods*, 2016, 13(12): 1050–1054.
- [7] Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng QD, Wortman J, Young SK, Earl AM. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*, 2014, 9(11): e112963.
- [8] Besemer J, Lomsadze A, Borodovsky M. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Research*, 2001, 29(12): 2607–2618.
- [9] Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, 1997, 25(5): 955–964.
- [10] Grissa I, Vergnaud G, Pourcel C. CRISPRfinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Research*, 2007, 35(suppl\_2):

- W52–W57.
- [11] Bertelli C, Laird MR, Williams KP, Simon Fraser University Research Computing Group, Lau BY, Hoad G, Winsor GL, Brinkman FSL. IslandViewer 4: expanded prediction of genomic Islands for larger-scale datasets. *Nucleic Acids Research*, 2017, 45(W1): W30–W35.
  - [12] Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 2015, 12(1): 59–60.
  - [13] Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang YJ, Wishart DS. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research*, 2016, 44(W1): W16–W21.
  - [14] Stothard P, Wishart DS. Circular genome visualization and exploration using CGView. *Bioinformatics*, 2005, 21(4): 537–539.
  - [15] Fischer S, Brunk BP, Chen F, Gao X, Harb OS, Iodice JB, Shanmugam D, Roos DS, Stoeckert CJ Jr. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. *Current Protocols in Bioinformatics*, 2011, 35(1): 6.12.1–6.12.19.
  - [16] Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, 2004, 14(7): 1394–1403.
  - [17] Delcher AL, Salzberg SL, Phillippy AM. Using MUMmer to identify similar regions in large sequence sets. *Current Protocols in Bioinformatics*, 2003, 00(1): 10.3.1–10.3.18.
  - [18] Fei YT, Li L, Zheng Y, Liu DM, Zhou QX, Fu L. Characterization of *Lactobacillus amylolyticus* L6 as potential probiotics based on genome sequence and corresponding phenotypes. *LWT*, 2018, 90: 460–468.
  - [19] Yang Y, An HR, Zhai ZY, Wang GH, Li JX, Hao YL. Complete genome sequence of *Lactobacillus helveticus* CAUH18, a potential probiotic strain originated from koumiss. *Journal of Biotechnology*, 2016, 224: 18–19.
  - [20] Gao Y, Liu YJ, Sun MY, Zhang HP, Mu GQ, Tuo YF. Physiological function analysis of *Lactobacillus plantarum* Y44 based on genotypic and phenotypic characteristics. *Journal of Dairy Science*, 2020, 103(7): 5916–5930.
  - [21] Huang D, Yang B, Chen Y, Stanton C, Ross RP, Zhao JX, Zhang H, Chen W. Comparative genomic analyses of *Lactobacillus rhamnosus* isolated from Chinese subjects. *Food Bioscience*, 2020, 36: 100659.
  - [22] Zhang DC, Yan YR, Que HY, Yang TT, Cheng XX, Ding SJ, Zhang XM, Cheng W. CRISPR/Cas12a-mediated interfacial cleaving of hairpin DNA reporter for electrochemical nucleic acid sensing. *ACS Sensors*, 2020, 5(2): 557–562.
  - [23] Jatuponwiphat T, Namrak T, Supataragul A, Nitisinprasert S, Nakphaichit M, Vongsangnak W. Comparative genome analysis reveals metabolic traits associated with probiotics properties in *Lactobacillus reuteri* KUB-AC5. *Gene Reports*, 2019, 17: 100536.
  - [24] Feng CJ, Zhang FX, Wang BN, Zhang LY, Dong YS, Shao YY. Genome-wide analysis of fermentation and probiotic trait stability in *Lactobacillus plantarum* during continuous culture. *Journal of Dairy Science*, 2020, 103(1): 117–127.
  - [25] Wang S, Bai YP, Li PH, Yang L, Wang XL. Genome-wide identification and expression analysis of the dof (DNA binding with one finger) protein family in monocot and dicot species. *Physiological and Molecular Plant Pathology*, 2019, 108: 101431.
  - [26] Shi YY, Yan LW, Zhao JX, Chen W, Zhang H, Hang F. The Co-culture technology applied in milk fermentation by *Lactobacillus plantarum*. *Science and Technology of Food Industry*, 2020, 41(6): 351–359. (in Chinese)  
石阳阳, 严利文, 赵建新, 陈卫, 张灏, 杭锋. 协同发酵在植物乳杆菌发酵乳中的应用研究进展. *食品工业科技*, 2020, 41(6): 351–359.
  - [27] Zhang HC, Duan YF, Xu LM, Cai F, Lv N, Zhu BL, Chang XD. Whole genome sequencing of two *Lactobacillus plantarum* strains isolated from human milk. *Microbiology China*, 2020, 47(6): 1876–1887. (in Chinese)  
张欢畅, 段云峰, 徐黎明, 蔡峰, 律娜, 朱宝利, 常学东. 两株母乳源植物乳杆菌的全基因组测序分析. *微生物学通报*, 2020, 47(6): 1876–1887.
  - [28] Jung J, Kim K, Yoo D, Lee C, Kang J, Cho K, Kang DK, Kwak W, Yoon SH, Sohn H, Kim H, Cho S. Comparative genomic analysis of *Lactobacillus plantarum* GB-LP4 and identification of evolutionarily divergent genes in high-osmolarity environment. *Genes & Genomics*, 2018, 40(2): 217–223.
  - [29] Wang YY, Chen C, Ai LZ, Zhou FF, Zhou ZM, Wang L, Zhang H, Chen W, Guo BH. Complete genome sequence of the probiotic *Lactobacillus plantarum* ST-III. *Journal of Bacteriology*, 2011, 193(1): 313–314.
  - [30] Sun MH, Deng M, Bai Y. Research progress on the probiotic

- properties of *Lactobacillus reuteri* and the treatment of ulcerative colitis. *Modern Digestion & Intervention*, 2019, 24(12): 1493–1496. (in Chinese)
- 孙梦晗, 邓敏, 白杨. 罗伊氏乳杆菌益生特性及治疗溃疡性结肠炎研究进展. *现代消化及介入诊疗*, 2019, 24(12): 1493–1496.
- [31] Fan H, Huo R, Zhao JN, Zhou TT, Zha MS, Kwok LY, Zhang HP, Chen YF. Microbial diversity analysis of jiaohe from Xilingol, Inner Mongolia. *Journal of Dairy Science*, 2020, 103(7): 5893–5905.
- [32] Bernardeau M, Vernoux JP, Henri-Dubernet S, Guéguen M. Safety assessment of dairy microorganisms: The *Lactobacillus* genus. *International Journal of Food Microbiology*, 2008, 126(3): 278–285.
- [33] Shi QH, Qin B, You S. Advances in the application of enzymes as therapeutic drugs. *Microbiology China*, 2020, 47(7): 2193–2206. (in Chinese)
- 时清华, 秦斌, 游松. 酶作为治疗性药物的研究进展. *微生物学通报*, 2020, 47(7): 2193–2206.
- [34] Butorac K, Banić M, Novak J, Leboš Pavunc A, Uroić K, Durgo K, Oršolić N, Kukolj M, Radović S, Scalabrin S, Žučko J, Starčević A, Šušćević J, Kos B. The functional capacity of plantaricin-producing *Lactobacillus plantarum* SF9C and S-layer-carrying *Lactobacillus brevis* SF9B to withstand gastrointestinal transit. *Microbial Cell Factories*, 2020, 19(1): 1–16.
- [35] Yusof HM, Rahman NA, Mohamad R, Zaidan UH, Samsudin AA. Biosynthesis of zinc oxide nanoparticles by cell-biomass and supernatant of *Lactobacillus plantarum* TA4 and its antibacterial and biocompatibility properties. *Scientific Reports*, 2020, 10: 19996.
- [36] Wang JZ, Jiang HY, Yang R, Zhang SB, Zhao WJ, Hu JT, Jiang YL, Yang WT, Huang HB, Shi CW, Yang GL, Wang HL, Wang CF, Feng N. Construction and evaluation of recombinant *Lactobacillus plantarum* NC<sub>8</sub> delivering one single or two copies of G protein fused with a DC-targeting peptide (DCpep) as novel oral rabies vaccine. *Veterinary Microbiology*, 2020, 251: 108906.
- [37] Shi X, Hu C, Cai S, Tao X, Zhou Y, Smidt H, Ye B. Protective effects of *Lactobacillus plantarum* strain P1 against toxicity of the environmental oestrogen di-n-butyl phthalate in rats. *Beneficial Microbes*, 2020, 11(8): 803–813.
- [38] Gerlach VL, Aravind L, Gotway G, Schultz RA, Koonin EV, Friedberg EC. Human and mouse homologs of *Escherichia coli* DinB (DNA polymerase IV), members of the UmuC/DinB superfamily. *Proceedings of the National Academy of Sciences of the United States of America*, 1999, 96(21): 11922–11927.
- [39] Makui H, Roig E, Cole ST, Helmann JD, Gros P, Cellier MFM. Identification of the *Escherichia coli* K-12 Nramp orthologue (MntH) as a selective divalent metal ion transporter. *Molecular Microbiology*, 2000, 35(5): 1065–1078.
- [40] Huang E, Kim S, Park H, Park S, Ji Y, Todorov SD, Lim SD, Holzapfel WH. Modulation of the gut microbiome and obesity biomarkers by *Lactobacillus plantarum* KC28 in a diet-induced obesity murine model. *Probiotics and Antimicrobial Proteins*, 2021, 13(3): 677–697.
- [41] Liu Q, Ni XQ, Wang Q, Peng ZR, Niu LL, Wang HS, Zhou Y, Sun H, Pan KC, Jing B, Zeng D. *Lactobacillus plantarum* BSGP201683 isolated from giant Panda feces attenuated inflammation and improved gut microflora in mice challenged with enterotoxigenic *Escherichia coli*. *Frontiers in Microbiology*, 2017, 8: 1885.
- [42] Ye ZW, Guo TF, Tang C, Yuan Y, Zhao Y, Lu J, Lin JF, Guo LQ. Transcriptional analysis for cholesterol-lowering effects of marine *Lactobacillus plantarum* Lp10 isolated from kelp. *LWT*, 2021, 139: 110563.

# Comparative study on the structure and function of *Lactobacillus plantarum* from different ecology

Chaonan Liu<sup>1</sup>, Jinming Wang<sup>1</sup>, Weixiao Nan<sup>2</sup>, Huazhe Si<sup>2</sup>, Xinyu Zhang<sup>1</sup>, Xiheng Cai<sup>1</sup>, Hanlu Liu<sup>1\*</sup>

<sup>1</sup> Institute of Special Animal and Plant Sciences, Chinese Academy of Agricultural Sciences, Changchun 130112, Jilin Province, China

<sup>2</sup> College of Animal Science and Technology, Jilin Agricultural University, Changchun 130118, Jilin Province, China

**Abstract: [Objective]** This experiment studies the genetic characteristics of *Lactobacillus plantarum* from different sources and its genetic diversity in different environments, and explores the similarities and differences in the adaptability of two strains of *L. plantarum* A8 and P9 in the animal intestine and plant surfaces, so as to provide a theoretical basis for the development of excellent strains. **[Methods]** This study analyzed the genomes of *L. plantarum* A8 and *L. plantarum* P9 isolated from the Animal intestines and plant surface. The research uses Next Generation Sequencing (NGS), based on the Illumina NovaSeq sequencing platform, and using the third-generation single-molecule sequencing technology, based on the PacBio Sequel sequencing platform to sequence *L. plantarum* A8 and *L. plantarum* P9. Carbohydrate-active enzymes (CAZy), Koyto encyclopedia of genes and genomes (KEGG), and Clusters of orthologous genes (COG) databases are used to annotate the genome function; CGView software is used to draw the genome circle map of the strain. At the same time, comparative genomics was used to compare with other published *L. plantarum* genomes. **[Results]** The research shows that the *L. plantarum* A8 and *L. plantarum* P9 genomes were different in genome size, the phylogenetic tree showed that the two strains were in the same branch with *L. plantarum* from other sources, and *L. plantarum* P9 was the closest to *L. plantarum* WLPL04 from breast milk, while *L. plantarum* A8 was the closest to *L. paraplantarum* DSM 10667. According to gene family analysis, there are 2643 common genes in the two strains, including some anti-stress proteins, such as heat shock proteins and cold shock proteins. There are 321 and 336 unique genes in plantarum A8 and P9, respectively. The unique genes in *L. plantarum* A8 are mainly involved in DNA replication, ABC transport system, PTS system, sulfonate transport system, amino acid biosynthesis and other metabolic pathways; The unique genes of *L. plantarum* P9 are mainly involved in carbohydrate transport and metabolism, such as *rpiA* gene, *lacZ* gene, *fruA* gene and so on. **[Conclusion]** By analyzing the genomic information of *L. plantarum* with comparative genomics method, it was found that *L. plantarum* from animal intestine has better amino acid transport ability, and the *L. plantarum* strain attached to the surface of the plant has good carbohydrate utilization ability, which provides a theoretical basis for the development and utilization of probiotics.

**Keywords:** *Lactobacillus plantarum*, comparative genomics, genome diversity

(本文责编: 李磊)

Supported by the National Key R&D Project (2018YFC1706602) and by the Jilin Province Science and Technology Development Plan Project (20200301019RQ)

\*Corresponding author. Tel: +86-431-81919856; E-mail: liuhanlu@caas.cn

Received: 28 December 2020; Revised: 8 February 2021; Published online: 4 August 2021