



研究报告

双歧杆菌属天然质粒系统分类

李洪飞¹ 孙大庆^{*1} 张桂芳¹ 庄柯瑾² 王坤² 郭庆丰³

1 黑龙江八一农垦大学国家杂粮工程技术研究中心 黑龙江 大庆 163319

2 黑龙江八一农垦大学食品学院 黑龙江 大庆 163319

3 黑龙江八一农垦大学理学院 黑龙江 大庆 163319

摘要:【背景】以往双歧杆菌质粒分类学研究较少,双歧杆菌属质粒系统分类方法缺失。【目的】建立双歧杆菌属天然质粒系统分类和鉴定方法,促进质粒在双歧杆菌生物学研究中的理解和应用。【方法】利用质粒复制起始蛋白进化树和基因组共线性分析方法,对目前所有已测序的双歧杆菌属天然质粒进行系统分类研究。【结果】双歧杆菌所有已知天然质粒可以划分为6个不同类型的质粒家族和3个独特的复合质粒,家族III和家族VI质粒可进一步划分为不同的亚型类群。家族III质粒是双歧杆菌属天然质粒的主要类型和优势家族。家族VI质粒成员亚型分类最丰富。在质粒家族水平,2种方法的分类结果完全一致。【结论】本文揭示了所有分析质粒之间的系统分类关系,建立了双歧杆菌属天然质粒系统的分类标准、方法和体系,可为今后双歧杆菌天然质粒分类和鉴定提供重要的理论参考和分类依据。

关键词: 双歧杆菌属,质粒分类,复制起始蛋白,基因组共线性,复制子

Phylogenetic classification of natural plasmids in *Bifidobacterium*LI Hongfei¹ SUN Daqing^{*1} ZHANG Guifang¹ ZHUANG Kejin² WANG Kun²
GUO Qingfeng³

1 National Coarse Cereals Engineering Research Center, Heilongjiang Bayi Agricultural University, Daqing, Heilongjiang 163319, China

2 College of Food Science, Heilongjiang Bayi Agricultural University, Daqing, Heilongjiang 163319, China

3 College of Science, Heilongjiang Bayi Agricultural University, Daqing, Heilongjiang 163319, China

Abstract: [Background] In the past, there were few studies on the taxonomy of bifidobacteria plasmids, and the phylogenetic classification method of *Bifidobacterium* plasmid was missing. [Objective] In order

Foundation items: Youth Science Foundation of Heilongjiang Province (QC2014C020); Training Plan of Young Innovative Talents in Colleges and Universities of Heilongjiang Province (UNPYSC2017112); Science Research Foundation of Heilongjiang Bayi Agricultural University (XDB202009); Heilongjiang Bayi Agricultural University Support Program for San Heng San Zong (ZRCPY202005); Advantage and Characteristic of Discipline of Coarse Production and Processing in Heilongjiang Province ([2018]4)

***Corresponding author:** E-mail: sundaqing1979@163.com

Received: 18-02-2020; **Accepted:** 20-04-2020; **Published online:** 09-06-2020

基金项目: 黑龙江省青年科学基金(QC2014C020); 黑龙江省普通本科高等学校青年创新人才培养计划(UNPYSC2017112); 黑龙江八一农垦大学科研启动计划(XDB202009); 黑龙江八一农垦大学三横三纵支持计划(ZRCPY202005); 黑龙江省“杂粮生产与加工”优势特色学科资助项目(黑教联[2018]4号)

***通信作者:** E-mail: sundaqing1979@163.com

收稿日期: 2020-02-18; **接受日期:** 2020-04-20; **网络首发日期:** 2020-06-09

to establish a method for phylogenetic classification and identification of natural plasmids in *Bifidobacterium*, and increase understanding and application of plasmids in the research of bifidobacteria biology. **[Methods]** We conducted to phylogenetic classification of all sequenced plasmids in *Bifidobacterium* by the analysis of phylogenetic relationship of replication initiation protein and genome colinearity. **[Results]** All known natural plasmids of *Bifidobacterium* were divided into six different types of plasmid families and three unique complex plasmids. Plasmids of type III and VI family were further divided into different subtype groups. Plasmids of the type III and VI family can be further divided into different subtype groups. Plasmids of the type III family were the main type and dominant family of natural plasmids in *Bifidobacterium*. Plasmid subtypes of type VI family were most abundantly classified. At the plasmid family level, the classification results of the two methods were completely identical. **[Conclusion]** These results comprehensively and accurately revealed the phylogenetic classification relationship between all the plasmids that were analyzed, and established the phylogenetic classification standard and method of natural plasmids in *Bifidobacterium*, which provided important theoretical references and classification basis for the future classification and identification of natural plasmids in *Bifidobacterium*.

Keywords: *Bifidobacterium*, plasmid classification, replication initiation protein, genome colinearity, replicon

双歧杆菌常发现于人和动物的消化道、生殖道和口腔等环境中, 是人和动物体内重要的共生菌。双歧杆菌可以通过抑制致病菌生长、增强免疫应答、抑制肿瘤血管生成和改善营养等方式, 在维持和调节体内微生态平衡、促进宿主健康方面发挥着广泛且积极的生理作用^[1-3]。双歧杆菌被证实可在几种实体瘤中选择性定位和增殖, 可用于癌症基因治疗中高度特异的基因递送载体^[4-5]。研究表明, 双歧杆菌与人和动物饮食、健康和疾病紧密相关^[6-7]。目前, 双歧杆菌已在食品和医药行业广泛应用, 产品种类众多, 具有重要的经济价值和社会意义。

目前, 双歧杆菌属已有 70 个菌(物)种和 10 个亚种被明确分离和鉴定 (<http://www.bacterio.net/bifidobacterium.html>)。基于美国国家生物技术信息中心(National Center for Biotechnology Information, NCBI)的 RefSeq 数据库(截止 2019-10-10), 双歧杆菌属共有 38 个质粒基因组序列被收录, 其中 23 个质粒来自长双歧杆菌(*Bifidobacterium longum*), 15 个质粒来自双歧杆菌属其他物种。这些数据表明, 双歧杆菌属已知天然质粒数量稀少, 资源比较匮乏, 多数物种可能不携带天然质粒, 质粒分布存在明显的物种依赖性。此外, 以往双歧杆菌质粒研究主要涉及质粒分离、鉴定和载体构建^[8], 而关于双歧杆菌质粒分类的研究很少, 至今在双歧杆菌“属”水平上开展的质粒系统分

类学研究未见报道。

质粒分类一直是细菌质粒研究的基础与核心问题。由于质粒遗传多样性丰富、变异率高, 至今没有一种通用且准确的方法可以将所有天然质粒进行分类^[9]。目前, 使用最广、最常见的分类方法仍是早期研究建立的质粒不相容性分类法^[10]。该方法分类和鉴定质粒时, 需要进行烦琐的不相容性功能验证实验, 费时费力、效率低下。同时, 质粒不相容性与复制、分配、维持多个质粒功能相关, 分子机制复杂多样, 至今没有完全阐明, 这进一步降低了该方法的有效性和准确性。因此, 该方法已不能满足后基因组时代大量新测序质粒的分类和鉴定需求。近年来, 以质粒复制子^[11-12]或松弛酶^[13]作为分子标记的质粒系统分类方法已有报道。前者理论上适用于所有天然质粒, 但实际中该方法只在有害菌(肠杆菌科^[12]、鲍曼不动杆菌^[14]、肠球菌^[15]、葡萄球菌^[16])中得到了发展和应用, 而在有益菌质粒中未见报道; 后者由于靶基因限制, 只适用于可转移质粒的分类, 因此应用范围受限^[13]。

双歧杆菌和乳杆菌是益生菌研究和应用的典型代表, 但它们的亲缘关系(双歧杆菌属于放线菌门, 乳杆菌属于厚壁菌门)和染色体 GC 含量(双歧杆菌约 60%, 乳杆菌约 40%)相差甚远, 因此天然质粒的多样性和分类很可能也显著不同。之

前, 基于质粒复制起始蛋白(Replication Initiation Protein, Rep)和/或质粒全基因组序列, 我们建立了2种不同的天然质粒系统分类方法, 并成功应用于乳杆菌属天然质粒分类^[17-19]。但该方法能否适用于亲缘关系较远和高GC含量的双歧杆菌或其他细菌质粒分类还不得而知。本文以所有测序的双歧杆菌天然质粒作为研究对象, 通过质粒Rep蛋白进化树分析和质粒基因组共线性分析, 探索建立双歧杆菌天然质粒系统分类方法、分类标准和分类体系, 为今后双歧杆菌天然质粒分类、鉴定和应用提供重要的理论参考和分类依据。同时, 本研究可进一步评价质粒Rep蛋白和质粒基因组共线性分类法对高GC含量细菌天然质粒的分类效果, 以期为今后建立一种通用、高效的细菌天然质粒分类方法提供更加充实的理论基础和实验依据。

1 材料与方法

1.1 质粒基因组数据

所有质粒基因组序列数据来自NCBI的Nucleotide和RefSeq数据库, 序列详细信息见表1。

1.2 方法

1.2.1 质粒基因组序列的收集、整理和初步分析

检索Nucleotide和RefSeq数据库中所有收录的双歧杆菌质粒基因组序列, 筛查获得完整基因组质粒及其Rep蛋白编码情况。对全部Rep蛋白序列进行BLAST序列比对分析, 鉴定其保守结构域, 初步确定蛋白家族。利用DNASTar 7.1.0 EditSeq软件对核苷酸或氨基酸序列进行存储、编辑等处理。

1.2.2 核苷酸和氨基酸多序列比对和重复序列分析

利用MEGA 7.0^[20]软件中Muscle法进行核苷酸或氨基酸多序列比对, 比对参数为默认设置。DNA重复序列分析利用DNASTar 7.1.0 GeneQuest软件进行, 最小序列长度设置为8 bp, 间距设置为小于200 bp。

1.2.3 质粒Rep蛋白系统进化树分析

利用MEGA 7.0软件中Neighbor-Joining (N-J)法进行质粒Rep蛋白系统进化树构建。系统进化树检验采用Bootstrap法, 自展值设置为1 000, 其他参数采用默认设置。

表1 Nucleotide和RefSeq数据库收录的双歧杆菌属质粒基因组序列

Table 1 Genome sequences of *Bifidobacterium* plasmids in the Nucleotide and RefSeq databases

菌种 Species	质粒 Plasmid	登录号 Accession No.	菌种 Species	质粒 Plasmid	登录号 Accession No.
<i>B. asteroides</i>	pCIBAO89	NC_010908.1	<i>B. longum</i>	pMG1	NC_006997.1
<i>B. bifidum</i>	pB80	NC_011332.1	<i>B. longum</i>	p6043A	NC_010857.1
<i>B. breve</i>	pBR3	NZ_CP010414.1	<i>B. longum</i>	p6043B	NC_010861.1
<i>B. breve</i>	pCIBb1	NC_002133.1	<i>B. longum</i>	pFI2576	NC_011139.1
<i>B. breve</i>	pB21a	NC_010930.1	<i>B. longum</i>	pSP02	NC_019200.1
<i>B. breve</i>	pMP7017	KM406416.1	<i>B. longum</i>	pNAL8M	NC_025161.1
<i>B. catenulatum</i>	pBC1	NC_007068.1	<i>B. longum</i>	pNAL8L	NC_025162.1
<i>B. choerinum</i>	pBC	NZ_CP018045.1	<i>B. longum</i>	pBL505	NZ_CM011649.1
<i>B. gallinarum</i>	p1CACC514	NZ_CP035465.1	<i>B. longum</i>	pDOJH10L	NC_004252.1
<i>B. kashiwanohense</i>	pBBKW-1	NC_021875.1	<i>B. longum</i>	pDOJH10S	NC_004253.1
<i>B. kashiwanohense</i>	pBBKW-2	NC_021876.1	<i>B. longum</i>	pBLO1	NC_004943.1
<i>B. longum</i>	pRY68	NZ_CP010454.1	<i>B. longum</i>	p1-5B1	KP691640.1
<i>B. longum</i>	pBIF10	DQ093580.1	<i>B. longum</i>	p1-6B2	KP691635.1
<i>B. longum</i>	pEK13	KP691637.1	<i>B. longum</i>	p157F-NC1	NC_015053.1
<i>B. longum</i>	pKJ36	NC_002635.1	<i>B. longum</i>	p157F-NC2	NC_015066.1
<i>B. longum</i>	pB44	NC_004443.1	<i>B. longum</i>	BLNIAS_P1	NC_017220.1
<i>B. longum</i>	pNAC3	NC_004768.1	<i>B. longum</i>	BLNIAS_P2	NC_017222.1
<i>B. longum</i>	PNAC2	NC_004769.1	<i>B. pseudocatenulatum</i>	p4M	NC_003527.1
<i>B. longum</i>	pNAC1	NC_004770.1	<i>B. pseudolongum</i>	pASV479	NC_010877.1
<i>B. longum</i>	pKJ50	NC_004978.1	<i>B. sp.</i>	pBIFA24	NC_010164.1
<i>B. longum</i>	pTB6	NC_006843.1			

1.2.4 质粒基因组共线性分析

利用 Mauve 2.4.0^[21]软件进行质粒基因组共线性分析, Match Seed Weight 为 13, 其他参数均为默认设置。基因组共线性分析中, 质粒基因组默认由小到大排列。

2 结果与分析

2.1 双歧杆菌天然质粒基本特征

经检索和筛查分析, 目前 Nucleotide 数据库收录双歧杆菌质粒 DNA 序列 103 个(包括 RefSeq 数据库收录的 38 个双歧杆菌质粒 DNA 序列), 其中

41 个质粒为完整基因组序列。统计分析发现, 双歧杆菌 41 个完整质粒基因组总长 387.77 kb, 携带 399 个蛋白编码基因, 大小范围为 2.11–190.18 kb, 平均 GC 含量为 61.39%, 大多数(27/41)质粒来自 *B. longum*, 大多数(60/70)物种尚未发现天然质粒(表 2)。这些结果表明, 双歧杆菌质粒普遍较小, 大型质粒匮乏, 平均 GC 含量接近染色体, 质粒分布具有明显的物种依赖性, *B. longum* 是双歧杆菌属中天然质粒多样性最丰富的物种。检索和 BLAST 分析发现, 38 个双歧杆菌质粒共编码 42 个

表 2 双歧杆菌属质粒基因组和 Rep 蛋白基本特征

Table 2 General characteristics of plasmid genome and Rep in *Bifidobacterium*

质粒 Plasmid	大小 Size (kb)	GC 含量 GC content (mol%)	编码蛋白 Coding protein	Rep 家族/大小 Rep family/Size (aa)	质粒 Plasmid	大小 Size (kb)	GC 含量 GC content (mol%)	编码蛋白 Coding protein	Rep 家族/大小 Rep family/Size (aa)
pCIBAO89	2.111	52.34	1	RepA_C/346	pBIFA24	4.892	61.77	3	Rep_3/244
pFI2576	2.197	61.9	4	Rep*/306	p157F-NC1	4.895	61.9	5	Rep_3/303
pBC1	2.54	63.66	3	Rep*/315	p6043A	4.896	61.85	3	Rep_3/303
pRY68	2.638	66.3	2	Rep_3/299	pSP02	4.896	61.93	4	Rep_3/303
pBBKW-2	2.92	62.6	6	Rep*/307	pB80	4.898	61.88	3	Rep_3/303
pNAL8L	3.489	58.96	5	Rep_3/303	pNAL8M	4.91	61.87	7	Rep_3/303
pNAC1	3.538	58.79	4	Rep_3/319	pKJ50	4.96	61.83	3	Rep_3/275
pB44	3.624	65.12	2	Rep_3/299	pB21a	5.206	56.26	5	Rep_1/565
pTB6	3.624	65.09	4	Rep_3/230	pCIBb1	5.75	56.9	3	#Rep_1/380
p157F-NC2	3.624	64.96	3	Rep_3/299	BLNIAS_P2	6.23	61.06	3	N
p1-6B2	3.624	64.96	2	Rep_3/299	pEK13	7.05	60.68	7	Rep_3/303
pKJ36	3.625	65.13	3	Rep_3/230	pBBKW-1	7.716	64.85	6	Rep_3/297
pBLO1	3.626	64.81	2	Rep_3/297					RepA_C/366
pDOJH10S	3.661	66.18	3	N	p1CACCC51	48.72	64.23	6	RepA_C/370
p6043B	3.68	65.11	2	Rep_3/297	pBIF10	9.275	41.7	9	Rep_3/289
pMG1	3.682	65.07	2	Rep_3/255	pDOJH10L	10.073	62.18	9	Rep_3/297; Rep_3/303; RepA_C/326
pNAC2	3.684	64.69	3	Rep_3/297					
pBL505	3.751	65.42	3	Rep_3/299	pNAC3	10.224	62.03	8	RepA_C/370
p1-5B1	3.919	61.06	7	Rep_3/319	pBC	11.012	55.68	11	#Rep*/305; #Rep_trans/383
BLNIAS_P1	4.233	64.38	2	N					
p4M	4.488	53.1	2	Viral_Rep/341	pMP7017	190.178	57.43	228	RepA_C/405
pASV479	4.815	59.27	6	#Rep_1/378					
pBR3	4.891	61.91	5	Rep_3/303					

注: Rep_1: pfam01446; Rep_3: pfam01051; RepA_C: pfam04796; Rep*: pfam03090 和 pfam08708; Rep_trans: pfam02486; Viral_Rep: pfam02407; N: 未编码已知 Rep 蛋白; #: 本研究根据序列同源性推定的 Rep 蛋白

Note: Rep_1: pfam01446; Rep_3: pfam01051; RepA_C: pfam04796; Rep*: pfam03090 and pfam08708; Rep_trans: pfam02486; Viral_Rep: pfam02407; N: No known Rep has been encoded; #: Putative Rep by sequence similarity in this study

Rep 蛋白, 其中 35 个质粒只含有 1 个 Rep 蛋白, 3 个复合质粒含有多个 Rep 蛋白; 其余 3 个质粒没有编码已知的 Rep 蛋白; 所有 Rep 蛋白根据氨基酸序列中的保守结构域可以初步划分为 6 个不同的蛋白家族, 分别称为 Rep_1、Rep_3、RepA_C、Rep*、Rep_trans 和 Viral_Rep (表 2), 这表明双歧杆菌大多数(38/41)质粒含有 Rep 蛋白, 这些 Rep 蛋白序列具有一定的同源性和差异性, 因此 Rep 蛋白可以为双歧杆菌质粒系统分类提供良好的分子标记和依据。

2.2 双歧杆菌质粒 Rep 蛋白进化树分类

基于 Rep 蛋白氨基酸序列的同源性, 构建了双歧杆菌质粒 Rep 蛋白进化树, 结果见图 1。根据进化树拓扑结构和主要分支点置信度, 双歧杆菌质粒所有 Rep 蛋白明显聚类为 7 个不同的蛋白家族, 除了质粒 pBIF10 编码的 Rep 蛋白(AAZ79481.1), Rep 蛋白进化树分类结果与前文 Rep 蛋白保守结构域分类结果一致。根据 Rep 蛋白与质粒的对应关系, 35 个编码单一 Rep 蛋白质粒根据 Rep 蛋白进化树可以明确划分为 6 个不同类型的质粒家族, 其中质粒类型家族 III 是双歧杆菌天然质粒的主要类型和优势家族, 根据进化树家族 III 分支的拓扑结构和主要分支点置信度, 家族 III 质粒可以进一步明确划分为 2 个亚型类群(图 1)。此外, 尽管 3 个复合质粒不能划分到单一类型的质粒家族, 但其可以明确划分到非单一类型的质粒家族。例如, 复合质粒 pDOJH10L 含有 Rep_3 (WP 011067901.1; WP 008783729.1)和 RepA_C (WP 080504066.1)两类 Rep 蛋白, 因此其可以同时归属于质粒家族 III 和家族 VI (图 1)。实际上, 复合质粒很可能是由 2 个甚至更多不同质粒发生重组融合形成的, 属于质粒进化的特殊阶段, 因此, 该方法对复合质粒的分类结果恰好体现了复合质粒基因组中不同复制子序列的进化起源, 这进一步证明该方法对双歧杆菌质粒分类的有效性和准确性。总之, 上述分析结果表明, 双歧杆菌质粒 Rep 蛋白进化树分析比 Rep 蛋白保守结构域分析具有更高的分辨率和更好的分类效果, 可以对双歧杆菌所有编码 Rep 蛋白天然质粒进行简单、有效和准确的系统分类, 因此 Rep 蛋

白序列可以作为双歧杆菌所有编码 Rep 蛋白天然质粒系统分类理想的分类标准和分子标记。

2.3 双歧杆菌质粒基因组共线性分类

由于复制子是天然质粒基因组唯一必需的遗传元件, 因此利用基因组共线性分析结果进行质粒系统分类时, 同时将具有同源性复制子作为质粒家族分类的重要原则和标准。基于质粒基因组共线性和复制子序列分析, 双歧杆菌属全部 41 个完整质粒系统分类结果如图 2 所示。

2.3.1 双歧杆菌属质粒家族 I

经基因组共线性和复制子序列分析, 发现质粒 pASV479、pCIBb1、pB21a3 含有高度同源的复制子和比较保守的基因组共线性, 与其他双歧杆菌质粒没有显示同源性区域(图 2)。它们的复制子均由同源的单链起点、双链起点和 *rep* 基因组成, Rep 蛋白均含有保守结构域 Rep_1 (pfam01446)。这些特征与已知滚环型复制 pC194 质粒家族^[22]复制子高度相似。同时, O'Riordan 等研究已证明, 质粒 pCIBb1 采用滚环机制进行复制^[23]。因此, 基于基因组共线性和复制子序列特征, 这 3 个质粒可以明确划分为同一个类型质粒家族, 采用滚环机制复制, 命名为双歧杆菌属质粒家族 I。

2.3.2 双歧杆菌属质粒家族 II

基因组共线性分析显示, 质粒 p4M 只与质粒 pBIF10 在复制子以外区域显示了一定的同源性, 与其他双歧杆菌质粒没有显示任何同源性(图 2)。复制子序列分析显示, 质粒 p4M 的 Rep 蛋白含有保守结构域 Viral_Rep (pfam02407), 与 GenBank 数据库中众多来自动、植物单链 DNA 环状病毒的 Rep 蛋白显示高度一致性。目前, 关于质粒 p4M 只有一篇文献报道^[24], 该研究表明, 质粒 p4M 编码的 Rep 蛋白与多种动物环形病毒编码的 Rep 蛋白同属于一个滚环型复制的类质粒 Rep 蛋白家族, 该家族蛋白很可能由跨物种的重组事件产生。基于基因组共线性和复制子序列特征, 质粒 p4M 可以划分为一个独立类型的质粒家族, 可能采用环形病毒相似的滚环机制复制, 命名为双歧杆菌属质粒家族 II。

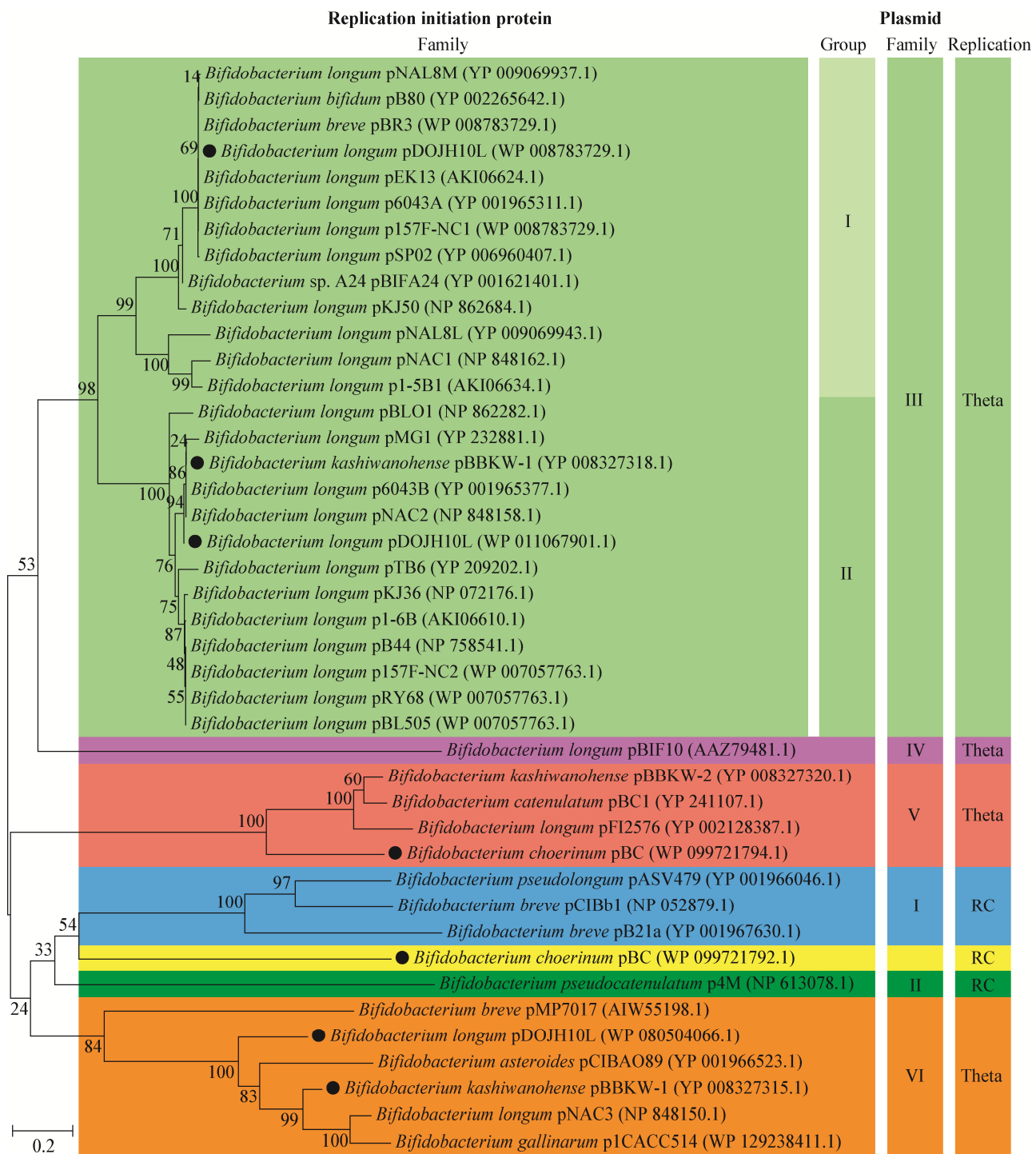


图 1 双歧杆菌属质粒 Rep 蛋白进化树

Figure 1 Phylogenetic tree of the plasmid Reps in *Bifidobacterium*

注: ●: 复合质粒 Rep 蛋白; RC: Rolling-Circle; 括号中字符为质粒编码 Rep 蛋白氨基酸序列登录号; 分支点数字表示置信度; 标尺表示遗传距离

Note: Black dot (●) indicates complex plasmid; RC indicates rolling-circle; The characters in parentheses are accession numbers of the plasmid-encoded Rep amino acid sequence; Branch point number indicates confidence; Scale indicates evolution distance

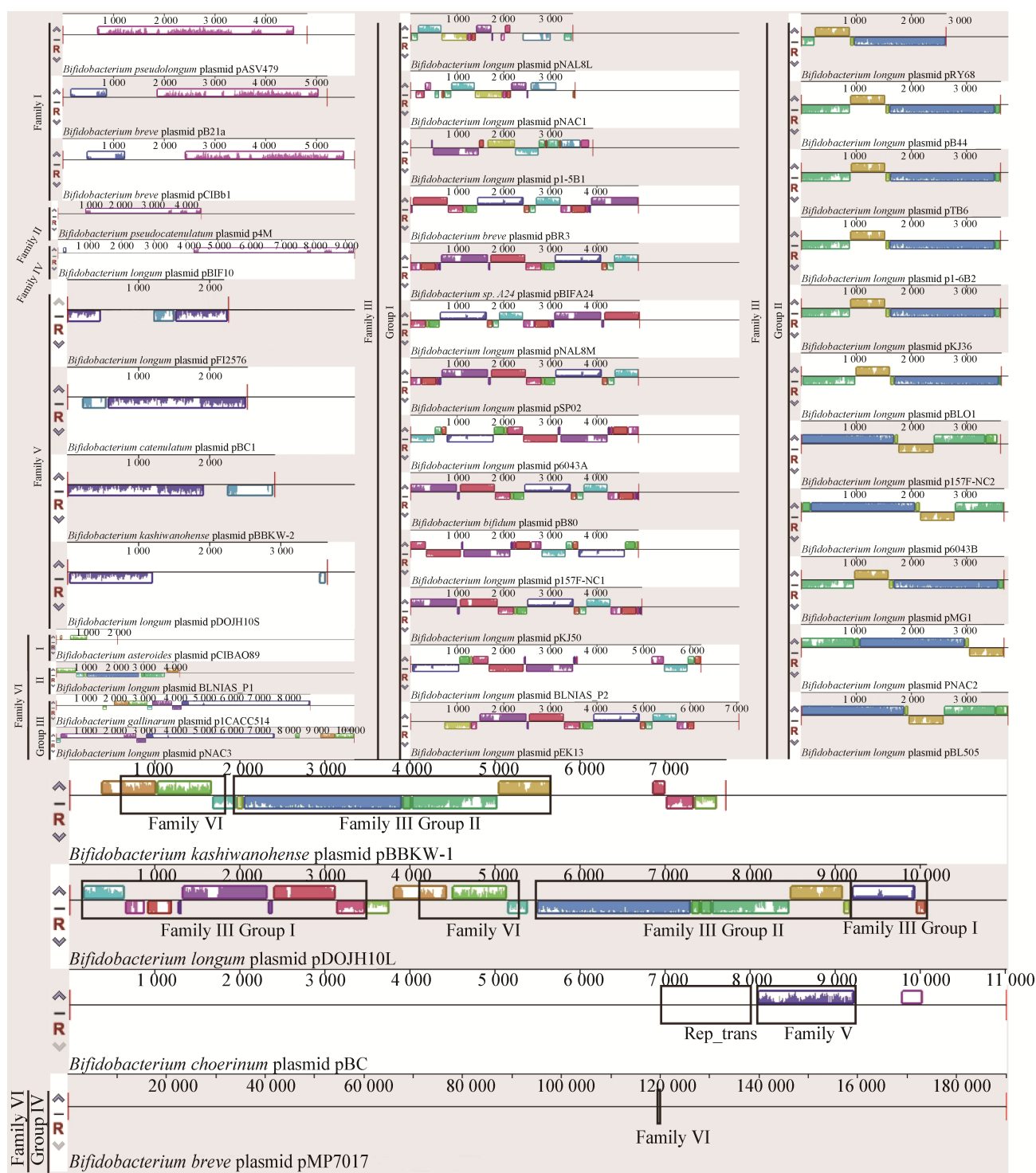


图2 双歧杆菌属质粒基因组共线性

Figure 2 Genome colinearity of the *Bifidobacterium* plasmids

注: *: 复合质粒; 黑框: 同源性区域

Note: Asterisk (*) indicates complex plasmid; Black rimmed indicates homology region

2.3.3 双歧杆菌属质粒家族 III

经基因组共线性和复制子序列分析, 发现 25 个质粒含有高度相似的复制子, 除复制子区域外, 这些质粒基因组共线性明显分为 2 个高度保守的亚型类群(图 2)。复制子序列分析显示, 这些质粒复制子均由 *rep* 基因及其上游的重复子序列组成, Rep 蛋白均含有保守结构域 Rep_3 (pfam01051), 重复子均由 3×10 bp 或者 4×10 bp 和 (3.0–5.5)×22 bp 两组串联的正向重复序列组成。这些特征与已知 Theta 型复制 pUCL287 质粒家族^[25]高度相似。因此, 基于基因组共线性和复制子特征, 这些质粒可以明确划分为同一个类型的质粒家族, 可能采用 Theta 机制复制, 命名为双歧杆菌属质粒家族 III。根据基因组共线性的明显差异, 可以进一步划分为 2 个亚型类群。另外, 基因组共线性分析发现, 尽管数据库中质粒 BLNIAS_P2 没有注释 *rep* 基因, 但实际上该质粒编码了家族 III 质粒高度同源的重复子和 *rep* 基因, 只是其 *rep* 基因编码区都发生了移码突变, 不能产生完整的 Rep 蛋白。此外, 尽管家族 III 所有质粒都含有一个典型的 Theta 型复制子序列, 但质粒 pNAL8L^[26]、pNAC1^[27]、pKJ36^[28]、pMG1^[29]、pSP02^[30]、pKJ50^[28,31]的 DNA 杂交实验结果显示, 这些质粒复制过程中都存在单链中间体, 而单链中间体是滚环复制质粒的典型特征^[32]。因此, 这些质粒真实的复制类型和机制仍然存疑, 这些质粒的相关报道也都提及和论述了这一现象。这提示, 家族 III 质粒可能存在一种未知的质粒复制机制。

2.3.4 双歧杆菌属质粒家族 IV

基因组共线性分析显示, 质粒 pBIF10 只与家族 II 质粒 p4M 在复制子以外区域显示了一定的同源性, 而与其他双歧杆菌质粒几乎没有显示同源性(图 2)。复制子序列分析显示, 质粒 pBIF10 复制子由 *rep* 基因及其上游的重复子序列组成, Rep 蛋白含有保守结构域 Rep_3 (pfam01051), 重复子由 3.5×20 bp 一组串联的正向重复序列组成。这些结果表明, 尽管质粒 pBIF10 的复制子结构和 Rep 蛋白结构域与家族 III 质粒相似, 但该质粒 Rep 蛋白序

列(图 1)、重复子结构和基因组共线性(图 2)均与家族 III 质粒明显不同, 因此质粒 pBIF10 可以划分为一个独立类型的质粒家族, 推测采用 Theta 机制复制^[33], 命名为双歧杆菌属质粒家族 IV。

2.3.5 双歧杆菌属质粒家族 V

基因组共线性和复制子序列分析显示, pFI2576 等 4 个质粒含有高度同源的复制子和高度保守的基因组区域, 与其他双歧杆菌质粒没有显示同源性区域(图 2)。它们的复制子由 *rep* 基因及其下游的复制起点序列组成, Rep 蛋白均含有保守结构域 pfam03090 和 pfam08708。复制子特征与已知的 Theta 型复制 ColE3 质粒家族^[34]高度相似。同时, 质粒 pFI2576^[35]、pBBKW-2^[8]和 pDOJH10S^[36]已经被实验证实采用 Theta 机制进行复制。此外, 基因组共线性分析发现, 质粒 pDOJH10S 实际上编码了 *rep* 基因, 只是该基因发生了突变, 形成了假基因。基于上述基因组共线性和复制子特征, 这 4 个质粒可以明确划分为新的质粒家族, 采用 Theta 机制复制, 命名为双歧杆菌属质粒家族 V。

2.3.6 双歧杆菌属质粒家族 VI

基因组共线性和复制子序列分析显示, pCIBAO89 等 5 个质粒虽然基因组大小和共线性差异显著, 但它们具有一个同源的复制子区域(图 2)。该复制子由 *rep* 基因及其上游的重复子序列组成, Rep 蛋白均含有保守结构域 RepA_C (pfam04796), 重复子序列、结构和间隔表现出较低的保守性。目前这 5 个质粒的复制机制没有得到实验验证, 但根据质粒 pCIBAO89 和 pNAC3 的报道推测, 它们可能采用 Theta 机制进行复制^[27,37]。尽管这 5 个质粒的基因组大小和共线性差异显著, 但基因组中同源性复制子的存在表明, 它们很可能来自同一个祖先质粒, 之后发生了分歧进化。因此, 基于同源性复制子特征, 这 5 个质粒可以划分为同一个类型的质粒家族, 可能采用 Theta 机制复制, 命名为双歧杆菌属质粒家族 VI; 同时, 基于基因组共线性的明显差异, 可以进一步划分为 4 个不同的亚型类群(图 2)。另外, 该家族不但包括双歧杆菌中最小的质粒

pCIBAO89 和唯一的巨型质粒 pMP7017, 并且该家族质粒亚型数量也是双歧杆菌中最多的, 因此质粒家族 VI 很可能是双歧杆菌天然质粒中基因水平转移等重组进化事件发生最频繁的质粒家族。此外, 与质粒 BLNIAS_P2 和 pDOJH10S 情况一样, 基因组共线性分析发现, 质粒 BLNIAS_P1 同样含有一个 *rep* 假基因。

2.3.7 双歧杆菌属复合质粒

基因组共线性和复制子序列分析显示, 复合质粒 pBBKW-1 同时含有质粒家族 III 和质粒家族 VI 类型的复制子, 复合质粒 pBC 含有质粒家族 V 同源性复制子之外, 还编码了一个独特的 Rep_{trans} 家族 Rep 蛋白, 复合质粒 pDOJH10L 甚至同时含有 3 个完整的复制子(图 2)。这些结果表明, 这些复合质粒很可能在近期由 2 个不同类型质粒发生重组进化事件融合形成, 这与质粒 pBBKW-1 和 pDOJH10L 的报道结果^[8,36]一致。实际上, 复合质粒只是天然质粒进化过程的一个特殊阶段, 并非生存常态。鉴于复合质粒含有 2 个以上不同家族质粒的复制子区域, 不能划分到单一类型的质粒家族, 因此它们可以作为独特的质粒类型进行单独分类。

2.4 双歧杆菌属天然质粒系统分类结果和特征

综合质粒 Rep 蛋白系统进化树和基因组共线性分析结果, 双歧杆菌属所有 41 个天然质粒可以划分为 6 个不同类型和起源的质粒家族和 3 个独特的复合质粒, 家族 III 和家族 VI 质粒可进一步划分为不同的亚型类群(图 1 和图 2)。家族 I、II 质粒可能采用滚环机制复制, 家族 III–VI 质粒可能采用 Theta 机制复制。家族 III 质粒成员数量最多, 是双歧杆菌属天然质粒的主要类型和优势家族。家族 VI 质粒成员大小和亚型多样性最丰富, 是双歧杆菌属天然质粒系统进化关系最复杂的质粒家族。在质粒家族水平 2 种方法分类结果完全一致, 在亚家族水平基因组共线性分类法可以进行更加细致的亚型分类(例如家族 VI 质粒的亚型分类, 图 2)。2 种独立方法的分类结果互为证明和验证, 因此这些结果准确、有效地揭示了所有分析质粒之间的系统进化关

系和对应的分类学地位, 成功地建立了双歧杆菌属天然质粒系统分类标准、方法和体系, 对今后双歧杆菌天然质粒分类和鉴定具有重要的参考价值和指导意义。

3 讨论与结论

本文通过 Rep 蛋白进化树和基因组共线性 2 种分类方法, 对目前已测序的双歧杆菌属天然质粒进行了系统分类研究。研究结果显示, 质粒 Rep 蛋白进化树分类法可将双歧杆菌所有 38 个编码 Rep 蛋白天然质粒划分为 6 个不同类型的质粒家族和 3 个独特的复合质粒, 家族 III 质粒可进一步划分为 2 个亚型类群(图 1); 质粒基因组共线性分类法可将双歧杆菌所有 41 个天然质粒划分为 6 个不同类型的质粒家族和 3 个独特的复合质粒, 家族 III 和家族 VI 质粒可进一步划分为 2 个和 4 个亚型类群(图 2)。除了 3 个 *rep* 基因突变质粒, 2 种方法分类结果在质粒家族水平完全一致, 这有力地证明了分类结果和分类方法的有效性、准确性和可靠性。比较 2 种方法的分类结果可知, 基因组共线性分类法的通用性和分辨率(亚型分类更细致)明显优于 Rep 蛋白进化树分类法。这些结果证明, 基因组共线性分类法可以对双歧杆菌所有天然质粒进行全面、细致和可靠的系统分类, 因此是双歧杆菌天然质粒一种比较理想的系统分类方法。同时, 质粒基因组共线性分类法使用时需要同时进行质粒复制子序列特征分析和比较, 这一定程度上增加了分析时间和操作难度, 尤其对于初学者和复制机制不够清晰的质粒类型。比较而言, 质粒 Rep 蛋白进化树分类法更加简单、快速和实用, 更适合用于质粒类型的初步鉴定。总之, 质粒 Rep 蛋白进化树和基因组共线性分类法各有优点和不足, 应根据具体的研究目的、对象和范围, 合理地选择 1 种或 2 种分析策略和分类方法, 扬长避短, 才可能获得最佳的质粒分类或鉴定结果。

随着 DNA 测序技术的快速发展, 测序的质粒数量迅速增加, 这对质粒分类和鉴定提出了更高的

要求,同时也促使基于基因(组)系统进化关系的质粒系统分类成为可能。前文已经提到,早期建立的质粒不相容性表型分类法已经无法满足后基因组时代大量质粒的分类和鉴定需求。近年来,基于质粒保守序列的系统分类方法得到了较快的发展,其中复制子和松弛酶分类法最具有代表性^[11-13]。遗传学研究显示,复制子是质粒中唯一普遍存在的遗传单元,但质粒复制子中并不存在普遍保守的基因,复制子遗传元件的组成和结构复杂且多样^[13]。因此,尽管复制子是质粒系统分类中唯一具有普遍性和通用性的遗传标记,但至今质粒复制子分类法只是一些有害菌中得到了发展和应用,而在益生菌质粒分类研究中未见报道。本研究使用的 Rep 蛋白进化树分类法实际上是质粒复制子分类法的简化,其省去了复制子中其他序列的分析和鉴定,只保留了复制子中最常见的 Rep 蛋白,因此提高了分类方法的简便性和实用性。松弛酶是质粒水平转移必需的关键酶。一些研究基于松弛酶的系统发育关系对可移动质粒实现了简单、高效的系统分类,但该方法对于不可转移质粒无效,因而影响了该方法的广泛应用^[13]。

另一方面,与单个保守基因或蛋白相比,理论上基于全基因组序列同源性的系统分类方法具有更好的准确性和可靠性,并且具有优良的全面性和通用性。以往研究显示,由于质粒可以跨物种频繁发生水平转移,突变和重组率高,不存在共有基因和性状,因此质粒分类一直是质粒研究领域的难点和挑战。目前质粒分类还没有一种通用、准确的方法,仍无法采用双名制命名和阶元分类系统。这直接造成了质粒分类、命名和鉴定的混乱,严重阻碍了质粒基础和应用研究的发展。利用质粒 Rep 蛋白进化树和基因组共线性分类法,我们前期对厚壁菌门低 GC 含量的乳杆菌属天然质粒进行了成功的系统分类^[17-19]。本研究结果显示,这 2 种分类法同样适用于放线菌门高 GC 含量双歧杆菌属天然质粒的系统分类,进一步验证和拓展了这些分类方法的有效性和通用性。本研究与之前乳杆菌属质粒研究结

果一致表明,质粒基因组共线性分类法比 Rep 蛋白进化树分类法具有更高的准确性和可靠性,并且表现出良好的全面性和通用性。因此,该方法克服了以往复制子分类法(包括 Rep 蛋白进化树分类法)和松弛酶分类法的不足和瓶颈,理论上适用于所有细菌天然质粒的系统分类。从这个角度而言,质粒基因组共线性分类法的建立和发展,不仅为今后乳杆菌和双歧杆菌天然质粒的系统分类鉴定提供了新的分类方法,而且为建立一种通用、准确的细菌天然质粒分类方法奠定了重要的研究基础和依据。因此,该方法可能对今后所有细菌天然质粒的系统分类、分类阶元界定、系统进化、双名制命名和分类鉴定研究具有普遍的指导意义和潜在的应用价值。同时,质粒基因组共线性分类法主要利用生物信息学软件 Mauve 在电脑中进行操作,实现了较高的通量(一次性分析,双歧杆菌 41 个质粒,乳杆菌家族 5 质粒 116 个^[19])和自动化分析,基本解决了高通量测序中大量质粒分类和鉴定的难题,满足了后基因组时代天然质粒分类的高通量和自动化的需求。然而,质粒基因组共线性分类法仍存在一些不足,在使用过程中,其分析时间和分类效率仍有待于进一步优化、改进和提高。

REFERENCES

- [1] Russell DA, Ross RP, Fitzgerald GF, Stanton C. Metabolic activities and probiotic potential of bifidobacteria[J]. International Journal of Food Microbiology, 2011, 149(1): 88-105
- [2] Arbolea S, Watkins C, Stanton C, Ross RP. Gut bifidobacteria populations in human health and aging[J]. Frontiers in Microbiology, 2016, 7: 1204
- [3] Ku S, Park MS, Ji GE, You HJ. Review on *Bifidobacterium bifidum* BGN4: functionality and nutraceutical applications as a probiotic microorganism[J]. International Journal of Molecular Sciences, 2016, 17(9): 1544
- [4] Fu GF, Li X, Hou YY, Fan YR, Liu WH, Xu GX. *Bifidobacterium longum* as an oral delivery system of endostatin for gene therapy on solid liver cancer[J]. Cancer Gene Therapy, 2005, 12(2): 133-140
- [5] Tang W, He YF, Zhou SC, Ma YP, Liu GL. A novel *Bifidobacterium infantis*-mediated TK/GCV suicide gene therapy system exhibits antitumor activity in a rat model of bladder cancer[J]. Journal of Experimental & Clinical

- Cancer Research, 2009, 28(1): 155
- [6] Ruiz L, Delgado S, Ruas-Madiedo P, Sánchez B, Margolles A. Bifidobacteria and their molecular communication with the immune system[J]. *Frontiers in Microbiology*, 2017, 8: 2345
- [7] Mahony J, Lugli GA, Van sinderen D, Ventura M. Impact of gut-associated bifidobacteria and their phages on health: two sides of the same coin?[J]. *Applied Microbiology and Biotechnology*, 2018, 102(5): 2091-2099
- [8] Lee JH, O'Sullivan DJ. Sequence analysis of two cryptic plasmids from *Bifidobacterium longum* DJO10A and construction of a shuttle cloning vector[J]. *Applied and Environmental Microbiology*, 2006, 72(1): 527-535
- [9] Orlek A, Stoesser N, Anjum MF, Doumith M, Ellington MJ, Peto T, Crook D, Woodford N, Sarah Walker A, Phan H, et al. Plasmid classification in an era of whole-genome sequencing: application in studies of antibiotic resistance epidemiology[J]. *Frontiers in Microbiology*, 2017, 8: 182
- [10] Hedges RW, Datta N. Plasmids determining I pili constitute a compatibility complex[J]. *Journal of General Microbiology*, 1973, 77(1): 19-25
- [11] Carattoli A, Bertini A, Villa L, Falbo V, Hopkins KL, Threlfall EJ. Identification of plasmids by PCR-based replicon typing[J]. *Journal of Microbiological Methods*, 2005, 63(3): 219-228
- [12] Carloni E, Andreoni F, Omiccioli E, Villa L, Magnani M, Carattoli A. Comparative analysis of the standard PCR-based replicon typing (PBRT) with the commercial PBRT-KIT[J]. *Plasmid*, 2017, 90: 10-14
- [13] Fernandez-Lopez R, Redondo S, Garcillan-Barcia MP, De La Cruz F. Towards a taxonomy of conjugative plasmids[J]. *Current Opinion in Microbiology*, 2017, 38: 106-113
- [14] Bertini A, Poirel L, Mugnier PD, Villa L, Nordmann P, Carattoli A. Characterization and PCR-based replicon typing of resistance plasmids in *Acinetobacter baumannii*[J]. *Antimicrobial Agents and Chemotherapy*, 2010, 54(10): 4168-4177
- [15] Jensen LB, Garcia-Migura L, Valenzuela AJS, Løhr M, Hasman H, Aarestrup FM. A classification system for plasmids from enterococci and other Gram-positive bacteria[J]. *Journal of Microbiological Methods*, 2010, 80(1): 25-43
- [16] Lozano C, García-Migura L, Aspiroz C, Zarazaga M, Torres C, Aarestrup FM. Expansion of a plasmid classification system for Gram-positive bacteria and determination of the diversity of plasmids in *Staphylococcus aureus* strains of human, animal, and food origins[J]. *Applied and Environmental Microbiology*, 2012, 78(16): 5948-5955
- [17] Sun DQ, Li HF, Song DW, Yang J. Classification of natural plasmids in *Lactobacillus plantarum*[J]. *Food Science*, 2017, 38(12): 69-74 (in Chinese)
孙大庆, 李洪飞, 宋大巍, 杨健. 植物乳杆菌天然质粒分
类[J]. *食品科学*, 2017, 38(12): 69-74
- [18] Sun DQ, Li HF, Yang J, Song DW. Classification of natural plasmids in *Lactobacillus brevis*[J]. *Food Science*, 2018, 39(10): 173-178 (in Chinese)
孙大庆, 李洪飞, 杨健, 宋大巍. 短乳杆菌天然质粒分类
[J]. *食品科学*, 2018, 39(10): 173-178
- [19] Sun DQ, Song DW, Li HF. Research on the Genomics of Natural Plasmids of *Lactobacillus*[M]. Beijing: China Light Industry Press, 2018: 187-189 (in Chinese)
孙大庆, 宋大巍, 李洪飞. 乳杆菌属天然质粒基因组学研
究[M]. 北京: 中国轻工业出版社, 2018: 187-189
- [20] Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets[J]. *Molecular Biology and Evolution*, 2016, 33(7): 1870-1874
- [21] Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements[J]. *Genome Research*, 2004, 14(7): 1394-1403
- [22] Seery LT, Nolan NC, Sharp PM, Devine KM. Comparative analysis of the pC194 group of rolling circle plasmids[J]. *Plasmid*, 1993, 30(3): 185-196
- [23] O'Riordan K, Fitzgerald GF. Molecular characterisation of a 5.75-kb cryptic plasmid from *Bifidobacterium breve* NCFB 2258 and determination of mode of replication[J]. *FEMS Microbiology Letters*, 1999, 174(2): 285-294
- [24] Gibbs MJ, Smeianov VV, Steele JL, Upcroft P, Efimov BA. Two families of Rep-like genes that probably originated by interspecies recombination are represented in viral, plasmid, bacterial, and parasitic protozoan genomes[J]. *Molecular Biology and Evolution*, 2006, 23(6): 1097-1100
- [25] Benachour A, Frère J, Novel G. pUCL287 plasmid from *Tetragenococcus halophilus* (*Pediococcus halophilus*) ATCC 33315 represents a new theta-type replicon family of lactic acid bacteria[J]. *FEMS Microbiology Letters*, 1995, 128(2): 167-175
- [26] Guglielmetti S, Karp M, Mora D, Tamagnini I, Parini C. Molecular characterization of *Bifidobacterium longum* biovar longum NAL8 plasmids and construction of a novel replicon screening system[J]. *Applied Microbiology and Biotechnology*, 2007, 74(5): 1053-1061
- [27] Corneau N, Émond É, LaPointe G. Molecular characterization of three plasmids from *Bifidobacterium longum*[J]. *Plasmid*, 2004, 51(2): 87-100
- [28] Park MS, Lee KH, Ji GE. Isolation and characterization of two plasmids from *Bifidobacterium longum*[J]. *Letters in Applied Microbiology*, 1997, 25(1): 5-7
- [29] Park MS, Moon HW, Ji GE. Molecular characterization of plasmid from *Bifidobacterium longum*[J]. *Journal of Microbiology and Biotechnology*, 2003, 13(3): 457-462
- [30] Álvarez-Martin P, Życka-Krzesińska J, Bardowski J, Mayo

- B. Sequence analysis of plasmid pSP02 from *Bifidobacterium longum* M62 and construction of pSP02-derived cloning vectors[J]. *Plasmid*, 2013, 69(2): 119-126
- [31] Park MS, Shin DW, Lee KH, Ji GE. Sequence analysis of plasmid pKJ50 from *Bifidobacterium longum*[J]. *Microbiology*, 1999, 145(3): 585-592
- [32] Ruiz-Masó JA, Machón C, Bordanaba-Ruisecco L, Espinosa M, Coll M, Del Solar G. Plasmid rolling-circle replication[J]. *Microbiology Spectrum*, 2015, 3(1): PLAS-0035-2014
- [33] Ma YP, Xie TT, Hu QW, Qiu ZY, Song FZ. Sequencing analysis and characterization of the plasmid pBIF10 isolated from *Bifidobacterium longum*[J]. *Canadian Journal of Microbiology*, 2015, 61(2): 124-130
- [34] Aoki K, Shinohara M, Itoh T. Distinct functions of the two specificity determinants in replication initiation of plasmids ColE2-P9 and ColE3-CA38[J]. *Journal of Bacteriology*, 2007, 189(6): 2392-2400
- [35] Moon GS, Wegmann U, Gunning AP, Gasson MJ, Narbad A. Isolation and characterization of a theta-type cryptic plasmid from *Bifidobacterium longum* FI10564[J]. *Journal of Microbiology and Biotechnology*, 2009, 19(4): 403-408
- [36] Takahata M, Toh H, Nakano A, Takagi M, Murakami M, Ishii Y, Takizawa T, Tanabe S, Morita H. Complete sequence analysis of two cryptic plasmids from *Bifidobacterium kashiwanohense* JCM 15439 (type strain) isolated from healthy infant feces[J]. *Animal Science Journal*, 2014, 85(2): 158-163
- [37] Cronin M, Knobel M, O'Connell-Motherway M, Fitzgerald GF, Van Sinderen D. Molecular dissection of a bifidobacterial replicon[J]. *Applied and Environmental Microbiology*, 2007, 73(24): 7858-7866

2021 年中国微生物学会及各专业委员会学术活动计划表(3-1)

序号	活动名称	主办/协办单位	时间	人数	地点	联系方式
1	第四届姑苏医学论坛——脓毒症精准诊断与治疗等学术会议	中国微生物学会微生物毒素专业委员会	2021 年 1 月 22-24 日	120	江苏 苏州	吴瑶 13466596992
2	微生物肥料产业创新与绿色优品认定培训班	中国微生物学会农业微生物学专业委员会	2021 年 4-5 月	60	北京	马鸣超 Mamingchao@caas.cn
3	第八届全国微生物基因组学学术研讨会	中国微生物学会农业微生物学专业委员会	2021 年 4-5 月	200	湖北 武汉	孙明 m98sun@mail.hzau.edu.cn
4	微生物基因组培训班	中国微生物学会农业微生物学专业委员会	2021 年 4-5 月	100	武汉	孙明 m98sun@mail.hzau.edu.cn
5	第 12 届传染病防控基础研究与应用技术学术交流会	中国微生物学会分析微生物学专业委员会	2021 年 6 月	1 000	待定	宋亚军 010-66948562
6	黄酒研讨会	中国微生物学会酿造分会	2021 年 5 月 10 日	30	北京	魏艳丽 15910408405
7	第九届地质微生物学学术研讨会	中国微生物学会地质微生物学专业委员会	2021 年 5 月 28-30 日	400	北京	李高远 15901174283
8	第一届世界病毒学会年会	中国微生物学会病毒学专业委员会	2021 年 6 月	2 000	线上	吴莹(中国地区) 15901455682
9	兽医微生物与生物制品学学术论坛	中国微生物学会	2021 年 6 月 18-20 日	300	河南 郑州	朱良全 丁家波 010-62103675
10	微生物与人类健康 2021	中国微生物学会医学微生物与免疫学专业委员会	2021 年 7 月 16-18 日	160	重庆	饶先才 15223315285
11	中国微生物学会第八届全国农业微生物研究及产业化研讨会暨第十七届全国杀虫微生物学术研讨会	中国微生物学会农业微生物学专业委员会	2021 年 7 月	200	广东 广州	杨凯 yangkai@mail.sysu.edu.cn