

根瘤菌基因组结构的多样性及其与系统发育的关系*

郑君芳^{1**} 陈文峰² 刘桂荣^{1,3} 隋新华²
朱万孚¹ 陈文新² Randal N. Johnston⁴ 刘树林^{1,3**}

(北京大学基础医学院微生物学系 北京 100083)¹ (中国农业大学生物学院微生物系 北京 100094)²
(加拿大卡尔加里大学微生物与感染病学系)³ (加拿大卡尔加里大学生化与分子生物学系)⁴

摘要: 研究探讨依据基因组结构对根瘤菌进行系统发育分类的可行性。采用 *I-CeuI* 酶切及脉冲场电泳分析确定 25 株根瘤菌模式菌株的基因组结构, 据此对根瘤菌进行聚群, 并与 16S rRNA 聚群结果相比较。依据基因组结构的近似或不同, 25 株根瘤菌共分为 11 个基因组型 (genome type, GT)。这些基因组型与依据 16S rRNA 序列的聚群结果大体一致, 其不同之处显示了两种方法各自的特点。因此, 基因组的物理结构, 即本研究中 *rrn* 操纵子的数量及其在基因组上的位置, 可作为系统发育关系的标志而用于根瘤菌的分类。

关键词: 根瘤菌, 基因组结构, 系统发育分析, 16S rRNA

中图分类号: Q93 文献标识码: A 文章编号: 0253-2654 (2003) 04-0072-05

DIVERSITY OF GENOME STRUCTURE AMONG RHIZOBIA AND ITS CORRELATION WITH PHYLOGENY EVALUATED BY *I-CeuI*

ZHENG Jun-Fang¹ CHEN Wen-Feng² LIU Gui-Rong^{1,3} SUI Xin-Hua²
ZHU Wan-Fu¹ CHEN Wen-Xin² Randal N. Johnston⁴ LIU Shu-Lin^{1,3}

(Department of Microbiology, Peking University Health Sciences Center, Beijing 100083)¹

(College of Biological Science, China Agricultural University, Beijing 100094)²

(Department of Microbiology and Infectious Diseases, University of Calgary, Calgary, AB, Canada)³

(Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, AB, Canada)⁴

Abstract: This study has attempted to categorize the rhizobia based on their genomic features, hoping to use them as phylogenetic markers for natural classification of rhizobia. Genome structures of 25 rhizobial type strains were resolved by *I-CeuI* cleavage and subsequent pulsed-field gel electrophoresis (PFGE). Then they were divided into groups according to the similarity and differences of their genome structures. The 25 rhizobial strains were grouped into 11 genome types (GTs). These rhizobial strains were also clustered by 16S rRNA sequence comparison. The two methods had a general agreement, with each also reflecting different aspect of the genomes. It can be concluded that physical structure of the bacterial genome, here specifically the copy number and genomic distribution of genes coding for ribosomal RNA, can unambiguously be revealed by the endonuclease *I-CeuI* and used as phylogenetic marker to categorize rhizobia according to evolutionary relationship.

Key words: Rhizobia, Genome structure, Phylogenetic analysis, 16S rRNA

由于根瘤菌具有共生结瘤固氮的重要生物学特性, 其资源调查和分类的研究一直是近百年来研究的热点。目前根瘤菌的分类体系是以 16S rRNA 序列比较分析技术为中

* 教育部教育振兴行动计划专项基金资助项目 (985 项目)

国家教育部留学回国人员科研启动基金资助项目 (2002)

加拿大科研基金 CIHR (MOP-47817), MRC (GOP-38106), NSERC (216912-00)

** 联系人

收稿日期: 2002-09-17, 修回日期: 2002-11-25

心的多相分类体系。虽然 16S rRNA 序列的使用使根瘤菌分类开始建立在系统发育基础上^[1], 但 16S rRNA 序列存在基因横向转移和重组现象^[2], 而且用它推论的系统发育关系因所采用的算法不同而不同^[3]。所以, 只有找出更为稳定的基因组特征进行比较, 才能更好地从系统发育的角度对根瘤菌进行自然分类。

最近研究结果表明, 细菌基因组物理结构在进化上高度保守, 可以用作系统发育标志对细菌进行聚群^[4]。基因组结构中能反映细菌系统发育的信息包括编码核糖体 RNA 基因的数量和其在基因组的分布, 而核酸内切酶 I-*CeuI* 可明确反映这一特征^[4,5]。I-*CeuI* 是由藻类 *Chlamydomonas eugametos* 核糖体大亚单位 *rrl* 基因的 I 型内含子 (intron) 编码的核酸内切酶^[6,7], 其酶切位点为 *rrl* 基因的一段 26 bp 的序列。本研究即用此技术来对根瘤菌进行聚群, 并与 16S rRNA 分类结果相比较, 进而评价基因组物理结构方法学在根瘤菌系统发育研究中的意义及其在根瘤菌自然分类上的应用前景。

1 材料与方法

1.1 菌株、酶和化学试剂

所用菌株为根瘤菌模式菌株 25 株及完成全序列分析的 4 株根瘤菌 (表 1), 20% 甘

表 1 根瘤菌株

菌株	宿主	来源	GenBank 登录号	对应图 2 的序号
<i>Agrobacterium tumefaciens</i> C58 [*]	<i>Prunus pseudocerasus</i>	美国	AF008688	19
<i>A. tumefaciens</i> LAM12048	未知	荷兰	AJ389904	20
<i>Allorhizobium undicola</i> LMG11875 ^T	<i>Neptunia natans</i>	比利时	Y17047	11
<i>Bradyrhizobium elkanii</i> USDA76 ^T	<i>Glycine max</i>	美国	U35000	6
<i>B. japonicum</i> USDA110 [*]	<i>Glycine max</i>	美国	BA000040	5
<i>Mesorhizobium loti</i> MAFF303099 [*]	<i>Lotus</i>	日本	BA000012	1
<i>M. huakuii</i> CCGAU2609 ^T	<i>Astragalus sinicus</i>	南京 [*]	D12797	3
<i>M. mediterraneum</i> USDA3392 ^T	<i>Cicer arietinum</i>	西班牙	L38825	2
<i>Rhizobium etli</i> CFN42 ^T	<i>Phaseolus vulgaris</i>	墨西哥	U28916	23
<i>R. gallicum</i> USDA2918 ^T	<i>Phaseolus vulgaris</i>	法国	U86343	29
<i>R. giardinii</i> H152 ^T	<i>Phaseolus vulgaris</i>	法国	U86344	12
<i>R. hainanense</i> CCGAU57015 ^T	<i>Desmodium sinuatum</i>	海南 [*]	U71078	7
<i>R. huautlense</i> SO2 ^T	<i>Sesbania herbacea</i>	墨西哥	AF025852	28
<i>R. indigoferae</i> CCGAU 71042 ^T	<i>Indigofera amblyantha</i>	陕西 [*]	AF364068	26
<i>R. leguminosarum</i> USDA2370 ^T	<i>Pisum sativum</i>	美国	U29386	25
<i>R. mongolense</i> USDA 1844 ^T	<i>Medicago ruthenica</i>	内蒙古 [*]	U89817	9
<i>R. tropici</i> IIType A CFN299 ^T	<i>Phaseolus vulgaris</i>	巴西	X67233	10
<i>R. tropici</i> IIType B CIAT899 ^T	<i>Phaseolus vulgaris</i>	哥伦比亚	X67234	8
<i>R. yanglingense</i> CCGAU71623 ^T	<i>Guzdenstaedia multiflora</i>	甘肃 [*]	AF003375	24
<i>Sinorhizobium arboris</i> HAMBI1552 ^T	<i>Prosopis chilensis</i>	苏丹	Z78204	22
<i>S. fredii</i> USDA205 ^T	<i>Glycine soja</i>	河南 [*]	X67231	16
<i>S. kostiensense</i> HAMBI 1489 ^T	<i>Acacia senegal</i>	苏丹	Z78203	18
<i>S. kummerowiae</i> CCGAU71714 ^T	<i>Kummerowia stipulacea</i>	陕西 [*]	AF364067	4
<i>S. medicae</i> USDA1037 ^T	<i>Medicago truncatula</i>	法国	L39882	15
<i>S. meliloti</i> 1021 [*]	<i>Alfalfa</i>		AL591688	27
<i>S. meliloti</i> USDA1002 ^T	<i>Medicago sativa</i>	美国	X67222	21
<i>S. saheli</i> LMG7837 ^T	<i>Sesbania pachycarpa</i>	塞内加尔	X68390	13
<i>S. terangaie</i> LMG7834 ^T	<i>Acacia laeta</i>	塞内加尔	X68387	14
<i>S. xinjiangense</i> CCGAU110 ^T	<i>Glycine max</i>	新疆 [*]	AF250354	17

注: T 模式菌株, # 全序列分析完成的菌株, * 中国省或城市, CCGAU 中国农业大学菌种库, HAMBI 芬兰赫尔辛基大学应用化学和微生物系菌保中心, IAM 日本东京大学应用微生物研究所, LMG Collection of the Laboratorium voor Microbiologie en Microbiele Genetics, Rijksuniversiteit B-9000 Gent Belgium, USDA 美国农业部

油-80℃保藏于中国农业大学根瘤菌保藏中心 (Culture Collection of Beijing Agricultural University, CCBAU)。YMA 培养基 28℃培养。主要的工具酶 *I-CeuI* 购自 New England Biolabs, 蛋白酶 K 购自 Boehringer Mannheim, 其它化学试剂购自 Sigma 试剂公司。

1.2 DNA 提取

为得到完整的基因组 DNA, 整个提取的操作过程都在琼脂糖凝胶中进行^[8,9]。经蛋白酶 K 除蛋白后, 基因组 DNA 置于 TE 缓冲液中待用。

1.3 脉冲场凝胶电泳

I-CeuI 酶切后的基因组 DNA 经 Bio-Rad CHEF DRII 脉冲场电泳分离。

1.4 系统发育学分析

研究中分析的 16S rRNA 序列来自 GenBank 数据库, 用 PHYLIP 软件包得出系统发育树状图。各菌株菌号及其 16S rRNA 全序列在 GenBank 中的登录号见表 1。

2 结果

2.1 根瘤菌具有不同的基因组结构

如图 1 所示, 根瘤菌有较多样性的 *I-CeuI* 酶切电泳图谱。25 株根瘤菌的 *I-CeuI* 酶切图谱的模式图 (图 2) 表现出 11 种基因组型 (genome types, GTs), 有些基因组型根据菌株间的细微差异被进一步分成亚型 (图 2 和表 1)。本研究中所用的根瘤菌表现出 1~

6 个 *I-CeuI* 酶切片段, 即 1~6 个 *rrn* 操纵子。图 2 和表 1 还包括 4 个完成全序列测定的菌株作为参照。

2.2 16S rRNA 序列比较及其与基因组型 (GT) 聚群的关系

2.2.1 相同分类群内及不同分类群内菌株间的相似 *I-CeuI* 酶切图谱: 根瘤菌虽然具有相当基因组多样性的结构, 但在相同分类学种或属内, 一些菌株的染色体物理结构高度保守。如 *Sinorhizobium* 属、*Bradyrhizobium* 属和 *Mesorhizobium* 属包含的种具有几乎相同的 *I-CeuI* 酶切图谱 (图 2)。*Rhizobium* 属包含的种除 *R. giardinii* 外, 其余种大致可分为两种基本相似的 *I-CeuI* 酶切图谱。虽然相同的 *I-CeuI* 酶切图谱通常揭示保守的基因组结构和这些细菌之间密切的系统发育关系^[4,5], 但近似而不完全相同的酶切图谱能在多大程度上反映细菌间的系统发育关系则仍需其它方法如 16S rRNA 序列比较来证明。本研究的重要发现之一是某些属于不同分类种或属的菌株有近似甚至相同的 *I-CeuI* 酶切图谱。其中最显著的例子是 *R. giardinii* 与 *Sinorhizobium* 的多个种的菌株具有共同的 *I-CeuI* 酶切图谱 (GT XIIa; 图 2)。另外, 还有一些不同分类群的菌株 *I-CeuI* 酶切图谱仅有细微差别。如 GT XII 有四个亚型, 分别为 XIIa, XIIId, XIIg 和 XIIh, 这些基因组亚型包括 *Agrobacterium*, *Rhizobium* 和 *Sinorhizobium* 3 个属的菌株。不同分类群间具有近似的 *I-CeuI* 酶切图谱很可能是由于这些细菌系统发育关系密切, 也有可能只是巧合, 这就需要其它方法例如 16S rRNA 序列比较

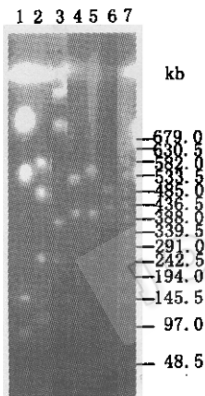


图 1 根瘤菌模式菌株 DNA 经 *I-CeuI* 酶切及脉冲场凝胶电泳 (PFGE)

后的胶图

电泳条件: 30-120s, 5.4V/cm, 16h; 50-60s, 5.4V/cm, 16h; 80-120s, 5.4V/cm, 16h。条带 1, *Salmonella typhimurium* LT2, 作为另一个分子量标记; 2, *Allorhizobium undicola* LMG11875^T; 3, *A. tumefaciens* IAM12048; 4, *R. yan-glingense* CCBAU71623^T; 5, *Rhizobium etli* CFN42^T; 6 *R. gallicum* USDA2918^T; 7 酵母染色体 PFG 分子量标记

分析对此做出澄清。

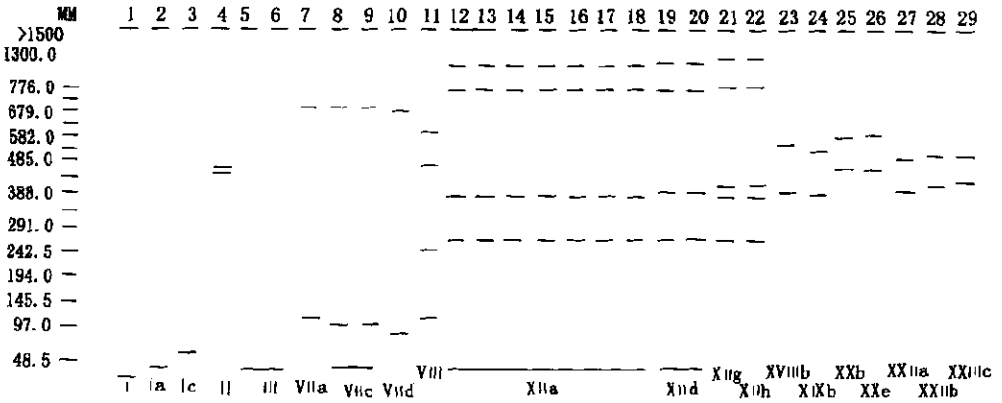


图 2 根瘤菌模式菌株 DNA 经 *I-CeuI* 酶切及脉冲场凝胶电泳 (PFGE) 后带型的模式图

其中包括 4 株全序分析完成的根瘤菌株作为参照, 序号与表 1 第 5 栏对应, 带型底部罗马数字为菌株所属的基因组型 (GTs)

2.2.2 16S rRNA 序列比较及其与基因组型 (GT) 聚群的关系 16S rRNA 序列分析将阐明上述具有相同基因组型的不同分类群是否是相同系统发育群 (“种 species” 或 “属 genus”) 的成员。上述 29 株菌的 16S rRNA 系统发育树 (图 3) 有 6 个主要分支。分支 1 仅包含 *Rhizobium* 属 9 个种的模式菌株。此分支包括五个 “分支 1 特有的” GTs: VII, XVIIIb, XIXb, XX 和 XXIIIc。分支 2 的菌株之间关系更密切, 成员间有极短的子分支和距离, 包括 *Sinorhizobium* 属的几乎所有种 (10 个种中的 8 个种) 的模式菌株。而这些种 GT 也几乎相同, 皆为 XII, 包括 XIIa, XIIg 和 XIIh。另外, 此分支还包含一个 II 和一个 XXIIa。XII 株还在紧密相邻的另外两个分支中出现 (分支 3 和 6)。分支 3 仅包含 1 株, 即 *R. giardinii*。虽然它在分类群上属于 *Rhizobium*, 但却有典型的 *Sinorhizobium* 属的基因组结构, 提示它和 *Sinorhizobium* 属有密切的系统发育关系, 而 16S rRNA 比较分析亦表明这些不同分类群细菌的密切系统发育关系。分支 4 仅包含 *Mesorhizobium* 属的模式菌株, 其 GT 为分支 4 特有的 I, 包括 3 个亚型, 即 I, Ia 和 Ic。分支 5 是 *Bradyrhizobium* 分支, 其包含的两株菌 16S rRNA 和 GT 都关系密切。分支 6 情况不同, 它不是仅包含一个分类学属的细菌, 而是包含来自 3 个属的细菌, 即 *Agrobacterium*, *Allorhizobium* 和 *Rhizobium*。

3 讨论

本研究所用的 *I-CeuI* 酶切技术通过反映核糖体 RNA 基因的数量和其在基因组的分布, 来反映全基因组的信息, 从而从系统发育角度反映全基因组特征^[8,9], 据此对根瘤菌进行系统发育分类。其结果不受基因横向转移影响, 而且不需要繁琐的分子克隆手段即可在普通实验室进行, 可针对每个菌株从系统发育角度高效和方便地对细菌进行分群。

4 株全基因组序列分析完成的根瘤菌证明了基因组结构分类的正确性。25 株根瘤

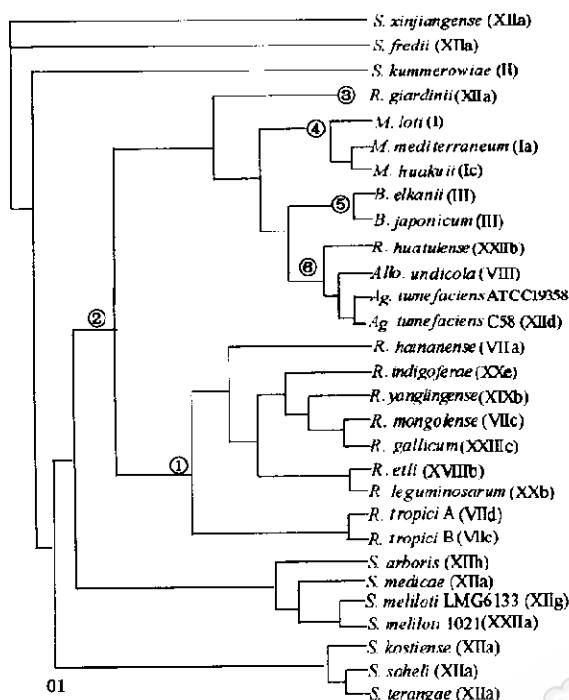


图3 根瘤菌模式菌株 16S rRNA 基因全序列构建的系统发育树。序列的 GenBank 登录号见表 1

其包含的部分基因组型 (XII、XXII) 与分支 2 在亚型水平上有重叠。同时, 基因组分型结果也有类似的分类情况。

所以, 染色体的物理结构, 即本研究中的 *rrn* 操纵子的数量和基因组位置, 可用作系统发育标志, 将根瘤菌依据进化关系分类成“系统发育种或属”, 借助 *rrn* 特征进行系统发育分类将是表明细菌基因组总体特征的强有力的系统学方法。

致谢 承蒙本实验室研究生龚俊和进修生钟有添老师在部分实验中的帮助, 特此致谢。

参考文献

- [1] Woese C R. Microbiol Rev, 1987, 51 (2): 221 ~ 271.
- [2] van Berkum P, Terzefewok Z, Paulin L, et al. J Bacteriol, 2003, 185 (10): 2988 ~ 2998.
- [3] Young J M. Int J Syst Evol Microbiol, 2001, 51: 945 ~ 953.
- [4] Liu S L, Schryvers A B, Sanderson K E, et al. J Bacteriol, 1999, 181 (21): 6747 ~ 6755.
- [5] Liu S L, Hessel A, Sanderson K E. Proc Natl Acad Sci U S A, 1993, 90 (14): 6874 ~ 6878.
- [6] Gauthier A, Turnel M, Lemieux C. Curr Genet, 1991, 19 (1): 43 ~ 47.
- [7] Marshall P, Lemieux C. Gene, 1991, 104 (2): 241 ~ 245.
- [8] Liu S L, Sanderson K E. J Bacteriol, 1992, 174 (5): 1662 ~ 1672.
- [9] Liu S L, Sanderson K E. Bacterial genome analysis. In: Methods in Microbiology: Bacterial Pathogenesis. N. Y.: Academic Press, 1998. 371 ~ 381.

菌模式菌株基因组结构分型与依据 16S rRNA 全序列的聚群基本一致, 但也各有特点。在依据 16S rRNA 数据建构的系统发育树的 6 个主要分支中 (图 3), 5 个分支分别由单一属的菌株组成, 其中 4 个分支的基因组型之间没有重叠, 充分表明基因组结构进化上高度保守, 可用作某种水平上细菌种系分离的系统发育标志。值得注意的是, *R. giardinii* 是分支 3 唯一的种, 在 16S rRNA 树中, 它不和其它 *Rhizobium* 属的种聚群, 而是和分支 2 密切相关, 并和大多数分支 2 的种具有相同的基因组亚型 XIIa。此结果强有力地表明基因组结构可作为系统发育标志, 并被 16S rRNA 数据所支持。分支 6 是一个例外, 由 3 个属的细菌组成,