

## 植物内生固氮菌系统发育进化新进展

黄淑芬<sup>1</sup> 郜晨<sup>1</sup> 刘丽辉<sup>1</sup> 谭志远<sup>1</sup> 彭桂香<sup>2\*</sup>

(1. 华南农业大学农学院 广东 广州 510642)

(2. 华南农业大学资源环境学院 广东 广州 510642)

**摘要:** 在植物内生固氮菌系统发育进化关系研究中,常用的方法有形态学与蛋白质水平法、数值分类和自动化鉴定法、化学分类法、分子遗传学方法等。本文简要介绍了常用方法的关键技术,并归纳了它们的优缺点。生物学的研究进入基因组时代后,随着高通量 DNA 测序技术在微生物学领域应用的迅速发展,全基因组测序被应用到微生物系统发育进化研究中,然而目前并未发现对已测全基因组序列的植物内生固氮菌进行系统总结。本文在对已测序植物内生固氮菌进行归纳的基础上,又详细研究了基于基因组数据的几种具有代表性的新方法(ANI 分析法、最大唯一匹配指数法、核心基因组分析、组分矢量法、基因流动性分析),并结合目前系统发育进化研究常用方法,对植物内生固氮菌系统发育进化研究趋势进行总结和展望,旨在使植物内生固氮菌的系统发育进化关系研究在精确度、可靠性等方面有所突破。

**关键词:** 内生固氮菌, 系统发育进化, 高通量测序, 基因组分析方法

## Research methods and trend of phylogenetic evolution of endophytic diazotrophs

HUANG Shu-Fen<sup>1</sup> GAO Chen<sup>1</sup> LIU Li-Hui<sup>1</sup> TAN Zhi-Yuan<sup>1</sup> PENG Gui-Xiang<sup>2\*</sup>

(1. College of Agriculture, South China Agricultural University, Guangzhou, Guangdong 510642, China)

(2. College of Resources and Environment, South China Agricultural University, Guangzhou, Guangdong 510642, China)

**Abstract:** To study phylogenetic relationships of endophytic nitrogen fixing bacteria in plants, common methods include morphology and protein level identification, numerical classification and automatic identification, chemical classification and identification, and molecular genetic identification. This paper introduces the key techniques of common methods, and discusses their advantages and disadvantages. With the development of high-throughput DNA sequencing technology in the field of microbiology, the whole genome sequencing has been applied to the study

**Foundation items:** National Natural Science Foundation of China (31370052); Natural Science Foundation of Guangdong Province (2014A030313459, 2014A050503058); National Natural Science Foundation of China-Guangdong Joint Fund (U1401234); Tobacco Monopoly Bureau Project of Guangdong Province (Guangdong Tobacco Department [2012] 26)

**\*Corresponding authors:** Tel: 86-20-85285852; E-mail: gxpeng@scau.edu.cn

**Received:** March 21, 2017; **Accepted:** May 18, 2017; **Published online** ([www.cnki.net](http://www.cnki.net)): June 27, 2017

**基金项目:** 国家自然科学基金项目(31370052); 广东省科技项目(2014A030313459, 2014A050503058); NSFC-广东联合基金项目(U1401234); 广东省烟草专卖局(公司)科技项目(粤烟科[2012]26)

**\*通信作者:** Tel: 86-20-85285852; E-mail: gxpeng@scau.edu.cn

**收稿日期:** 2017-03-21; **接受日期:** 2017-05-18; **网络首发日期**([www.cnki.net](http://www.cnki.net)): 2017-06-27

of microbial phylogenetic evolution. However, there is no systematic review of nitrogen-fixing endophytes whose genomes have been sequenced. A preliminary study on measuring whole genome sequences of endophytic diazotrophs. Several representative new methods based on genomic data (ANI analysis, the maximal unique match index, core genome analysis, component vector method, gene flow analysis) were summarized. Based on the current research methods of phylogenetic evolution, research trends of the development and evolution of Endophytic Nitrogen Fixing Bacteria were summarized and prospected. The purpose of this paper is to make the phylogenetic evolution of Endophytic Nitrogen Fixing Bacteria more accuracy and reliability.

**Keywords:** Endophytic diazotrophs, Phylogenetic evolution, High-throughput sequencing, Genome analysis methods

植物内生固氮菌是指那些定殖在植物内部与植物宿主联合固氮的固氮菌<sup>[1]</sup>。植物内生固氮菌占据着植物组织内有利于营养供应和微环境适宜的生态位,可以有效地拮抗病原微生物的生长,较根外环境更有利于形成高效固氮体系,进而促进作物的生长及产量的提高<sup>[2-4]</sup>。到目前为止,已有文献报道的内生固氮菌均为内生固氮细菌,尚未见到有关内生固氮真菌的报道<sup>[5]</sup>。植物内生固氮菌的种类不同,为寄主植物提供氮源的效率、产生植物激素类物质、促进宿主的生理变化也不同<sup>[6]</sup>。对从植物体内分离出来的固氮菌进行系统发育进化关系研究是生物固氮的基础之一,只有通过科学的方法,准确而有效地确定每一种内生固氮菌的分类学地位,才能更好地进行生物固氮机理的研究以及功能应用的开发。

## 1 常用的研究方法

常用的植物内生固氮菌系统发育进化研究方法主要是从4个水平进行:形态学与蛋白质水平法;数值分类法和自动化鉴定;化学分类鉴定法;分子遗传学鉴定法。

### 1.1 形态学与蛋白质水平法

传统的系统发育进化研究方法主要依赖于表型分析和个体形态学观察。形态学经常辅以生化试验来对植物内生固氮菌进行系统发育进化分析。生化试验是根据内生菌不同菌种在培养过程中所产生的新陈代谢产物各异而表现出不同的生长特性<sup>[7]</sup>。如糖(醇)类代谢试验、脂肪酸类和蛋白质代谢试验、矿物质元素利用试验、酶活性<sup>[8]</sup>、植物促

生物质试验等。对内生固氮菌蛋白质水平的鉴定技术,主要是蛋白质图谱分析<sup>[9]</sup>。蛋白质图谱分析主要采用聚丙烯酰胺凝胶电泳(Polyacrylamide gel electrophoresis, PAGE)和 SDS-PAGE。采用全细胞 SDS-PAGE 对高度标准条件下培养的细胞的可溶性蛋白质进行分析,可以得到反映生物基因组成的蛋白质图谱信息<sup>[10]</sup>,这项技术也是一种分群和比较大质量相近菌株的较好方法,该方法具有一定的可靠性,一般用于属内的分群<sup>[11]</sup>。形态学与蛋白质水平法存在一些缺点,例如重现率低、某些技术的不定性、低识别能力以及相似的表型特征并不等同于相似或者关系密切的基因型。所以对于许多表型特征难以区别的菌种,仅通过传统方法无法准确进行系统发育进化分析,该方法操作较繁琐,实验周期也较长<sup>[12]</sup>。

### 1.2 数值分类法和自动化鉴定

数值分类法是近 20 年来发展起来的理论,它应用大量已知菌对相关生化试验反应出现的频率得出数据进行分析,根据相似系数大小判断细菌种属间的亲缘性<sup>[13]</sup>,自动化鉴定系统便是采用数值分类原理建立的。随着微电子、计算机、分子生物学等先进技术向植物内生固氮菌生物学领域的渗透和多学科的交叉,对植物内生固氮菌进行快速系统发育进化分析有了突破性的进展<sup>[14]</sup>。BIOLOG 微生物自动分析系统是美国 BIOLOG 公司推出的一套运用数值分类原理的自动鉴定系统<sup>[15]</sup>,该系统主要根据细菌对糖、醇、酸、醋、胺和大分子聚合物等 95 种碳源的利用情况进行分析并鉴定。细菌利用碳

源进行呼吸时, 会将四哇类氧化还原染色剂(TV)从无色还原成紫色, 从而在鉴定微平板(96孔板)上形成该菌株特征性的反应模式或“指纹图谱”。鉴定板由读数仪自动读取吸光值, 软件自动判断结果为阴/阳性或边界值, 自动与数据库对比, 给出鉴定结果。预测微生物模型自动化鉴定是运用微生物学、工程数学以及统计学进行数学建模, 并与应用计算机学相结合组成的一种鉴定系统<sup>[16]</sup>, 它的目的在于用数学的语言描述微生物在不同环境条件下的生长、死亡情况。这些环境条件包括 pH、水分活度、温度、气体环境等。预测微生物自动鉴定的核心在于建立完善的数学模型, 然后利用美国 BIOLOG 公司开发的一种新的微生物鉴定方法——代谢指纹法, 运用特征选择和优化技术, 得到细菌鉴定板上每孔的特征曲线。对于得到的特征-时间曲线, 再运用预测微生物学, 拟合出预测微生物生长模型<sup>[17]</sup>, 结合统计学原理, 对数据进行过滤和匹配, 从而实现了对细菌的全自动鉴定。

### 1.3 化学分类鉴定法

植物内生固氮菌化学分类法通过化学测定, 获得各植物内生固氮菌组分的化学数据<sup>[18]</sup>, 依据得到的化学成分数据在种的水平上对植物内生固氮菌做出精确鉴定, 属的分类主要测定各种氨基酸组分。另外, 还有全细胞水解液糖型分析、脂肪酸分析、磷酸类脂成分分析、枝菌酸分析、醌类分析和光合色素成分分析等<sup>[19]</sup>, 常使用红外光谱、高效液相色谱等新技术。红外光谱法早在 20 世纪 50 年代起就开始运用于区分不同的植物内生固氮菌<sup>[20]</sup>。随着现代干涉型红外光谱仪、傅立叶变换技术以及计算机的发展, 这一研究领域又有了新的进展。傅里叶变换红外光谱法以未损伤细胞的傅里叶变换红外光谱的特殊指纹区为基础, 光谱反映的是整个细胞组成分子的振动特征, 也就是蛋白质、核酸等物质的特征, 因此可以区分生化信息上的差别<sup>[21]</sup>。高效液相色谱法(HPLC)是 20 世纪 60 年代末发展起来的一种分离分析新技术<sup>[22]</sup>, 可直接测定细菌 DNA

的碱基组成和化学组分, 可对细菌进行 DNA 的 G+C%含量测定<sup>[23]</sup>、分枝菌酸分析、醌类分析。化学分类法在植物内生固氮菌系统发育进化研究中具有分离效率高、选择性好、检测灵敏度高和分析速度快等优点。

### 1.4 分子遗传学分类鉴定法

自 20 世纪 60 年代开始, 分子遗传学和分子生物学技术的迅速发展使植物内生固氮菌系统发育进化研究进入了分子生物学时代, 目前主要的遗传学分析方法, 包括 DNA 的 G+C%、核酸杂交技术、16S rRNA 基因序列分析、多位点序列分型、全基因组测序以及核酸指纹图谱等<sup>[24]</sup>。染色体 DNA 的 G+C%含量即 DNA 的碱基组成<sup>[25]</sup>, 由于不受菌龄的影响, 已成为植物内生固氮菌系统发育进化的重要指标; 热变性温度测定法( $T_m$  法)操作简便、精确度高、重复性好, 被广为采用<sup>[23]</sup>。核酸杂交技术根据碱基互补配对原理, 将 2 条不同来源的单链核酸进行复性以鉴定菌株间的亲缘关系<sup>[26]</sup>, 用于 DNA-DNA (Southern 杂交)、DNA-RNA、RNA-RNA (Northern 杂交) 和 PNA-DNA 杂交(PNA 为肽核酸)。DNA-DNA 杂交适用于种水平的研究, 而 DNA-RNA 杂交<sup>[23]</sup>用于属和属以上水平的分类研究。16S rRNA 基因广泛存在于原核生物的细胞中, 保留了大量的信息量又便于扩增和测序, 其基因序列由恒定区和可变区组成, 可变区序列因不同植物内生固氮菌而异, 恒定区序列基本保守, 通常利用恒定区序列设计引物, 结合 PCR 技术, 将 16S rRNA 基因序列扩增出来, 再利用可变区序列的差异来对不同属种的植物内生固氮菌进行分类鉴定<sup>[27]</sup>, 现已成为大多数植物内生固氮菌系统发育进化研究的常用方法和描述新分类单元的必要指标<sup>[28]</sup>。

## 2 高通量测序技术在植物内生固氮菌中的应用

作为最重要的分子生物学分析方法之一, DNA 测序为遗传信息的揭示和基因表达调控等基础生

物学研究提供了重要的数据。自 2005 年以来,以 Roche 公司的 454 技术、Illumina 公司的 Solexa 技术和 ABI 公司的 SOLiD 技术为标志的高通量测序技术相继诞生。虽然高通量测序技术建立的时间不长,但发展非常快,已经应用于基因组,包括测序和表观基因组学以及功能基因组学研究的许多方面<sup>[29]</sup>。高通量测序技术堪称测序技术发展历程的一个里程碑,该技术可以对数百万个 DNA 分子进行同时测序。这使得对一个物种的转录组和基因组进行细致全貌的分析成为可能,将基因组学水平的研

究带入了一个新的时期,也使经典分子生物学家对基因组学的认识和思考上升到一个新的水平。高通量测序技术开启了植物内生固氮菌的基因组测序。2006 年 Krause 等<sup>[30]</sup>对第一株内生固氮菌 *Azoarcus* sp. strain BH72 完成了测序工作,但在植物内生固氮菌的系统发育进化研究中,可用的基因组数据并不多,这给研究工作带来了一定的困难,鉴于此,有必要对已测全基因组数据的植物内生固氮菌进行系统总结。目前已完成全基因组测序的植物内生固氮菌汇总见表 1。

表 1 已测全基因组序列的植物内生固氮菌

Table 1 List of nitrogen-fixing endophytes whose genomes have been sequenced

编号 Code	菌株 Strain	作者及年份 Author & Year	参考文献 References
1	<i>Azoarcus</i> sp. strain BH72	Krause et al., 2006	[30]
2	<i>Klebsiella pneumoniae</i> 342	Fouts et al., 2008	[31]
3	<i>Gluconacetobacter diazotrophicus</i> Pal5	Bertalan et al., 2009	[32]
4	<i>Azospirillum</i> sp. B510	Kaneko et al., 2010	[33]
5	<i>Herbaspirillum seropedicae</i> strain SmR1	Pedrosa et al., 2011	[34]
6	<i>Enterobacter</i> sp. strain SST3	Gan et al., 2012	[35]
7	<i>Enterobacter</i> sp. strain SP1	Zhu et al., 2012	[36]
8	<i>Burkholderia</i> sp. strain KJ006	Kwak et al., 2012	[37]
9	<i>Enterobacter cloacae</i> subsp. Cloacae strain ENHKU01	Liu et al., 2012	[38]
10	<i>Enterobacter radicincitans</i> DSM16656(T)	Witzel et al., 2012	[39]
11	<i>Enterobacter</i> sp. strain R4-368	Madhaiyan et al., 2013	[40]
12	<i>Rhizobium</i> sp. strain IRBG74	Crook et al., 2013	[41]
13	<i>Enterobacter cloacae</i> P101	Humann et al., 2013	[42]
14	<i>Herbaspirillum frisingense</i> GSF30	Straub et al., 2013	[43]
15	<i>Methylobacterium</i> sp. strain L2-4	Madhaiyan et al., 2014	[44]
16	<i>Bacillus pumilus</i>	Jeong et al., 2014	[45]
17	<i>Paenibacillus</i> sp. P22	Hanak et al., 2014	[46]
18	<i>Pantoea agglomerans</i>	Gan et al., 2014	[47]
19	<i>Klebsiella variicola</i> DX120E	Lin et al., 2015	[48]
20	<i>Kosakonia oryzae</i> K0348	Meng et al., 2015	[49]
21	<i>Raoultella terrigena</i> R1Gly	Schicklberger et al., 2015	[50]
22	<i>Kosakonia sacchari</i> strain BO-1	Shinjo et al., 2016	[51]
23	<i>Mangrovibacter</i> sp. strain MP23	Behera et al., 2016	[52]

### 3 基于基因组的几种新方法

高通量 DNA 测序技术的飞速发展使其在微生物学领域的应用也不断成熟, 比表型和基因序列分析的全基因组更准确的数据被应用到微生物系统发育研究, 全基因组测序带来了植物内生固氮菌的海量信息, 研究者们无法完全按照常规的方法对其进行系统发育进化研究<sup>[53]</sup>, 随之也涌现出一些基于基因组数据的分析方法, 使植物内生固氮菌系统发育进化研究在精确度、可靠性等方面取得了重要的突破。

#### 3.1 ANI 分析

ANI (Average nucleotide identity) 分析又名平均核苷酸一致性分析, 是指 2 个基因组之间同源基因的相似性<sup>[54-55]</sup>。平均核苷酸一致性是基于物种全基因组序列, 通过分析比较同源基因序列来判定物种间遗传关联性的重要参数。ANI 值可以通过 2 种运算方法得出, 一种是以 MUMmer 运算法则为基础 (ANIm), 另一种是以 BLASTn 方法为基础 (ANIb), 相比之下后者应用更为广泛<sup>[53]</sup>。普遍认为亲缘关系较近的种群 ANI 值至少为 70%–75%, 而定义一个种的 ANI 值需要达到 95%–96% 以上, 并且引起研究者们关注的是这些 ANI 的数值与“金标准”——DDH 有紧密的对应关系<sup>[56-57]</sup>。平均核苷酸一致性具有使用方便、耗时少、较低错误率、高分辨率的优势<sup>[58]</sup>。目前使用 JSpecies 软件计算 ANI 值是一个优选的方法<sup>[59]</sup>。如 Zhang 等采用多基因序列 (如 SMC00019-truA-thrA) 分析的 ANI 值比较, 以 94% 或 96% 作为区分不同种的阈值标准, 很好地将不同内生固氮菌区分开来<sup>[60]</sup>。平均核苷酸同源性分析逐渐被一些学者认为可以作为代替 DDH 的下一代金标准候选方法<sup>[61]</sup>。

#### 3.2 最大唯一匹配指数法

最大唯一匹配在基因序列比对中可以发挥重要的作用。最大唯一匹配可以从相互重叠的序列片段中重构 DNA 的完整序列, 也可在各种试验条件下从探测数据中决定物理和基因图存贮并查看和比

较数据库中的 DNA 序列来判断 2 个或多个序列的相似性。最大唯一匹配指数法 MUMi (Maximal unique matches index, MUMi) 是以 2 个基因组间的最大唯一配对数为基础并结合生物信息软件来计算基因组距离的方法, 可用于植物内生固氮菌的种内比较。MUMi 值的变动区间在 0–1 之间, MUMi 值越小则代表这 2 个基因组之间的亲缘关系越近。许多研究发现 MUMi 在衡量基因组亲缘关系时与 ANI 和 DDH 有比较好的关联性, 0.33±0.03 的 MUMi 值对应于 95%±0.5% 的 ANI 值和 70% 的 DDH 值<sup>[62]</sup>。Ormeño-Orrillo 等<sup>[63]</sup>在对根瘤菌进行研究中采用了最大唯一匹配指数的研究方法, 并确定了 *Rhizobium* sp. PRF 81、*Rhizobium tropici* CIAT899 与其他根瘤菌的亲缘关系, 其研究结果表明, *Rhizobium* sp. PRF 81 和 CIAT899 为亲缘关系很近的菌株, 而 *Rhizobium rhizogenes* K84 则应该属于其他的分类群。最大唯一匹配指数法在研究种内相近菌株的差异性方面具有很高的灵敏度<sup>[62]</sup>, 可在植物内生固氮菌种内相似菌株的鉴别中发挥作用; 此外这种方法还具有方便快捷的优点, 同时与 DDH 和 ANI 也存在紧密的对应关系, 因此在植物内生固氮菌系统发育进化研究中具有较大的优势<sup>[64]</sup>。

#### 3.3 核心基因组分析

核心基因组 (Core genome) 是指在菌株基因组中不易发生水平转移、较为稳定、包括看家基因在内的基因集, 这些基因大都具有种属特异性<sup>[65-66]</sup>。由于核心基因组的进化比较缓慢并且受种内重组的影响比较小, 因此可以将其用于鉴定植物内生固氮菌菌株之间的亲缘关系中<sup>[67]</sup>。核心基因组的构建方法主要是利用 BLAST 工具进行相似性比对, 在比对的同时设定研究所需要的阈值<sup>[68]</sup>。核心基因组法在植物内生固氮菌系统发育进化研究中的具体应用通常是先将核心基因组数据进行比对, 并据此绘制出内生固氮菌的系统发育树。谢剑波<sup>[69]</sup>通过将固氮菌株与非固氮菌株的核心基因组比较后发现, 固氮菌株具有 9 个特有的固氮基因。Kaas 等<sup>[70]</sup>在研

究了大肠杆菌的核心基因组数据后发现,由核心基因组建立的系统发育关系具有优异的分辨率,同时也指出使用核心基因组建树的方法应该普及并应用于大肠杆菌的系统发育分析中。基于核心基因组方法研究的进展表明核心基因组分析可广泛应用于属及属以上植物内生固氮菌的研究中,也可用于揭示各内生固氮菌群体内的进化关系。目前对于核心基因组分析方法的争议主要在于应该如何定义核心基因组,定义核心基因组的具体界限应该是什么,这些基因到底需要保守到什么程度才可以被列入到核心基因组的范畴中<sup>[71]</sup>。鉴于此,核心基因组方法在植物内生固氮菌系统发育进化中的应用还需进一步完善<sup>[72]</sup>。

### 3.4 K 串组分矢量法

K 串组分矢量法(K-string)是通过计算蛋白质序列或 DNA 序列中寡肽的出现频率来推断基因组相关性的一种方法<sup>[73]</sup>。在一个长度为 N 的 DNA 或者 RNA 序列中长度为 K 的连续结构即为 K 串,对于基因序列有  $4^K$  个可能的 K 串<sup>[74]</sup>。K 可以选择从 1-N 之间的任意数。短 K 值侧重于反映不同种之间的共同特征,而长的 K 值重点在于强调种的特异性<sup>[75]</sup>。在应用此方法构建的系统发育树中,当 K=5 或 6 时是最适合用于细菌系统发育构建的 K 值,并且也有学者指出没有必要选择大于 7 的 K 值<sup>[76]</sup>。2004 年,初步开始构建了这一方法应用的网络平台是 CVTree (<http://tlife.fudan.edu.cn/cvtree/>),这一平台在 2009 年得到了进一步的完善,从而为研究人员利用这一方法进行系统发育进化分析提供了技术支持。Chan 等将熵原理应用到组分矢量方程中,并对原有方法进行了合理的调整改善,从而使组分矢量法更加准确、应用范围也更加广泛<sup>[77]</sup>。Qi 等<sup>[73]</sup>利用组分矢量法分析了原核生物的系统发育关系,得到的结果与利用 16S rRNA 基因分析相似。K 串组分矢量法在系统发育进化分析中取得了重要的成就,不但适用范围比较广,而且还可用于细菌、古菌及真菌的系统发育进化分析<sup>[73]</sup>。然而 K 串组

分矢量法也存在一些问题,研究发现并不是所有的 K 串都可以有效地用于种系发生树的建立<sup>[78]</sup>,另外水平基因的转移也会影响到 K 串组分矢量法的准确性<sup>[75]</sup>。

### 3.5 基因流动性分析

基因流动性  $\phi$  (Genome fluidity)是指在一组个数为 N 的基因组中,独特基因数和总体基因的比值,这是属于一种基因容量的比对方法,可以简单地认为是基因组之间非重叠部分的度量,此方法最初主要用于衡量微生物种群内基因的多样性<sup>[79-80]</sup>。基因流动性  $\phi$  值为 0.1 的时候,代表一对基因组有 10% 的基因是不同的,有 90% 的基因是相同的,由此可见,基因流动性的数值越小代表对应基因组的特异性程度就越小<sup>[81]</sup>。Lan 等<sup>[82]</sup>介绍了一个基因组信息的简易比较平台 POGO-DB (<http://pogo.ece.drexel.edu/>),基因流动性是此平台所使用的比对方法之一。Zwick 等<sup>[81]</sup>运用此方法计算了蜡状芽孢杆菌(*Bacillus cereus*)的基因流动性,结果显示蜡状芽孢杆菌(*Bacillus cereus*)的基因流动性值  $\phi=0.22$ ,说明这个属的分类地位处于世界性物种如脑膜炎奈瑟氏菌(*Neisseria meningitidis*)或者大肠杆菌(*Escherichia coli*) ( $\phi=0.3$ ) 和金黄色葡萄球菌(*Staphylococcus aureus*) ( $\phi=0.15$ )之间。基因流动性分析这一方法的优点在于它不会因提供的基因序列过短或是鉴定基因同源性方法的不同而影响其结果的精确性<sup>[79]</sup>,但是基因流动性的准确性同样也会受水平基因转移的影响<sup>[77]</sup>,在一定程度上制约了这一方法在系统发育进化分析中的应用。

## 4 基于基因组分析方法的发展趋势

在系统发育进化分析研究中,以上论述的这几种基于基因组数据的方法,它们的应用领域各有侧重,但在微生物系统发育进化分析中均有大量成功的实例。DNA 杂交分析因与平均核苷酸一致性分析有良好的关联性而有望成为下一代的金标准<sup>[61]</sup>。最大唯一匹配指数法和平均核苷酸一致性分析同为计算基因组指数的方法,与 DNA 杂交同样也存在

一定的联系这一特性增加了其在系统发育进化应用中的优势<sup>[71]</sup>。核心基因组分析因为所选的基因都具有较强的保守性,因此在系统发育进化分析中会有比较高的分辨率和准确性。涉及到基因含量的2种方法——基因流动性分析和K串组分矢量法目前在植物内生固氮菌系统发育进化中的应用还比较少,但是由于其具有算法简便和免对比的优点,因而也具有一定的应用前景。在这5种方法中核心基因组分析和平均核苷酸一致性分析的应用相对比较成熟<sup>[71]</sup>。以上所涉及到的方法大部分是用于可培养植物内生固氮菌的,但是到目前为止,因为局限于技术手段,仍有大量不可培养的植物内生菌我们无法获得。对这些不可培养的内生菌,传统的系统发育进化研究方法是行不通的,需要运用宏基因组学的方法。宏基因组学<sup>[83]</sup>以环境中微生物基因组的总和为研究对象,从而规避了传统方法中绝大部分微生物不能培养的缺陷,因此近年来在环境微生物学研究中得到了广泛应用。宏基因组学不仅克服了微生物难以培养的困难,而且还可以结合生物信息学的方法,揭示微生物之间、微生物与寄主植物、微生物与环境之间相互作用的规律,大大拓展了微生物学的研究思路与方法,为从群落结构水平上全面认识微生物的生态特征和功能开辟了新的途径。

## 5 总结与展望

植物内生固氮菌系统发育进化研究的最终目标是建立一个可以反映“自然秩序”的系统。因此,植物内生固氮菌系统发育进化分析的技术在探索和争论中不断提高。目前已出现的基于基因组的新方法已经在很大程度上修正了之前存在的偏差,但是这些方法仍存在较大的改善空间。另一方面,由于目前对基因组序列信息所包含的生物学意义了解得不够透彻,所以在应用这些基因组数据时应当更加慎重。但由于以基因组数据为基础的这些方法具有快速、准确、节省人力等优点,相信会得到更广泛的应用。目前,对于不熟悉或从未发表的新菌种,很难采用现有的一种或少数几种方法进行系统

发育进化分析;对于不可培养的内生固氮菌,更不能用传统的方法进行系统发育进化分析。在今后的研究中,为了提高植物内生固氮菌系统发育进化分析的合理性和准确性,除了对已有方法改进完善并且探索新方法之外,也需要考虑将这些基因组数据的方法同形态学、生理生化特性等表型标记方法联合使用,同时还要探索和开发基于宏基因组学的新方法。因此,植物内生固氮菌系统发育进化分析必须在充分利用和完善现有的鉴定技术基础上,更进一步开发基于宏基因组大数据的新技术手段,使用传统与现代相结合以及不断发掘新方法以适应需要。随着植物内生固氮菌系统发育进化研究方法的不断建立和完善,将为科研工作者提供简便、快速、准确、微量、灵敏和成本低廉的检测方法及更先进、更全面的检测手段。

## REFERENCES

- [1] Prakamhang J, Minamisawa K, Teamtaisong K, et al. The communities of endophytic diazotrophic bacteria in cultivated rice (*Oryza sativa* L.)[J]. *Applied Soil Ecology*, 2009, 42(2): 141-149
- [2] Sturz AV, Christie BR, Nowak J. Bacterial endophytes: potential role in developing sustainable systems of crop production[J]. *Critical Reviews in Plant Sciences*, 2000, 19(1): 1-30
- [3] Yuan H, Hui Y, Fang Y, et al. Molecular characterization and phylogenetic analysis of endophytic nitrogen-fixing bacteria in *Oryza australiensis*[J]. *Chinese Journal of Applied and Environmental Biology*, 2014, 20(4): 571-577
- [4] Gupta G, Panwar J, Jha PN. Natural occurrence of *Pseudomonas aeruginosa*, a dominant cultivable diazotrophic endophytic bacterium colonizing *Pennisetum glaucum* (L.) R. Br.[J]. *Applied Soil Ecology*, 2013, 64: 252-261
- [5] Qin LP, Huang SL, Li YR. Research progress in endophytic diazotroph[J]. *Chinese Agricultural Science Bulletin*, 2005, 21(2): 150-152, 159 (in Chinese)  
覃丽萍, 黄思良, 李杨瑞. 植物内生固氮菌的研究进展[J]. *中国农学通报*, 2005, 21(2): 150-152, 159
- [6] Broek AV, Vanderleyden J. Genetics of the *Azospirillum*-plant association[J]. *Critical Reviews in Plant Sciences*, 1995, 14(5): 445-466
- [7] Chaudhary HJ, Peng GX, Hu M, et al. Genetic diversity of endophytic diazotrophs of the wild rice, *Oryza alta* and identification of the new diazotroph, *Acinetobacter oryzae* sp. nov.[J]. *Microbial Ecology*, 2012, 63(4): 813-821
- [8] Tan ZY, Fu QM, Peng GX, et al. Identification and characterization of endophytic diazotrophs isolated from *Cymbopogon caesius* and *Miscanthus floridulus*[J]. *Chinese Journal of Applied & Environmental Biology*, 2013, 19(4):

- 643-649 (in Chinese)  
谭志远, 傅琴梅, 彭桂香. 青香茅和五节芒内生固氮菌的分离与生理生化鉴定[J]. 应用与环境生物学报, 2013, 19(4): 643-649
- [9] Di Cagno R, De Angelis M, Calasso M, et al. Proteomics of the bacterial cross-talk by quorum sensing[J]. Journal of Proteomics, 2012, 74(1): 19-34
- [10] Wang LT, Lee FL, Tai CJ, et al. Comparison of *gyrB* gene sequences, 16S rRNA gene sequences and DNA-DNA hybridization in the *Bacillus subtilis* group[J]. International Journal of Systematic and Evolutionary Microbiology, 2007, 57(Pt 8): 1846-1850
- [11] Tan ZY, Xu XD, Wang ET, et al. Phylogenetic and genetic relationships of *Mesorhizobium tianshanense* and related rhizobia[J]. International Journal of Systematic Bacteriology, 1997, 47(3): 874-879
- [12] Deng MK, Sun Y, Han WQ. Methods of identification of bacteria[J]. Progress in Biomedical Engineering, 2014(2): 84-88 (in Chinese)  
邓梅葵, 孙迎, 韩雯晴. 细菌鉴定方法[J]. 生物医学工程学进展, 2014(2): 84-88
- [13] Li Y, Liu SL, Yi QQ, et al. Dynamic of soil microorganisms from root region of ginseng with different growing years[J]. Agricultural Sciences & Technology, 2009, 10(6): 141-143 (in Chinese)  
李勇, 刘时轮, 易茜茜. 不同栽培年限的人参根际区土壤微生物区系变化[J]. 农业科学与技术, 2009, 10(6): 141-143
- [14] Chen XQ, Chen Q, Zhang SR, et al. Taxonomy and BOX-PCR analysis of free-living diazotrophs isolated from soils in Liusha River Valley[J]. Journal of Agro-Environment Science, 2006, 25(S1): 528-532 (in Chinese)  
陈晓琴, 陈强, 张世熔, 等. 流沙河流域土壤内生固氮菌数值分类及 BOX-PCR 研究[J]. 农业环境科学学报, 2006, 25(S1): 528-532
- [15] Sellyei B, Wehmann E, Makrai L, et al. Evaluation of the Biolog system for the identification of certain closely related *Pasteurella* species[J]. Diagnostic Microbiology and Infectious Disease, 2011, 71(1): 6-11
- [16] Ballesté E, Bonjoch X, Belanche L, et al. Molecular indicators used in the development of predictive models for microbial source tracking[J]. Applied and Environmental Microbiology, 2010, 76(6): 1789-1795
- [17] Leordeanu M, Sukthankar R, Hebert M. Unsupervised learning for graph matching[J]. International Journal of Computer Vision, 2012, 96(1): 28-45
- [18] Walls I, Scott VN. Use of predictive microbiology in microbial food safety risk assessment[J]. International Journal of Food Microbiology, 1997, 36(2/3): 97-102
- [19] Tan ZY, Peng GX, Xu PZ, et al. Diversity and high nitrogenase activity of endophytic diazotrophs isolated from *Oryza rufipogon* griff[J]. Chinese Science Bulletin, 2009, 54(16): 2839-2848
- [20] Wu H. Experiment for microorganism research by fourier transform infrared spectrometry[J]. China Resources Comprehensive Utilization, 2009, 27(5): 23-25 (in Chinese)  
吴海. 傅立叶变换红外光谱技术对微生物研究的探索实验[J]. 中国资源综合利用, 2009, 27(5): 23-25
- [21] Chen SN. Effects *azotobacter* on the micro-ecosystem associated with the rhizosphere of maize and wheat[D]. Yangling: Doctoral Dissertation of Northwest A & F University, 2012 (in Chinese)  
陈胜男. 自生固氮菌对玉米、小麦根际微生态调控机制的研究[D]. 杨凌: 西北农林科技大学博士学位论文, 2012
- [22] Lu YM, Chen XH, Mei J, et al. Biogenic amines in Chinese soy sauce[J]. Food Control, 2009, 20(6): 593-597
- [23] Yan AM, Chen WX. DNA-DNA hybridization analysis of three new *Rhizobia* Groups[J]. Journal of China Agricultural University, 2000, 5(1): 14-20 (in Chinese)  
阎爱民, 陈文新. 三个根瘤菌新群的 DNA-DNA 杂交分析[J]. 中国农业大学学报, 2000, 5(1): 14-20
- [24] Álvarez-Sánchez B, Priego-Capote F, De Castro MDL. Metabolomics analysis I. Selection of biological samples and practical aspects preceding sample preparation[J]. TrAC Trends in Analytical Chemistry, 2010, 29(2): 111-119
- [25] Sofu A, Ekinci FY. Bacterial diversity dynamics of traditional Turkish Ezine Cheese as evaluated by PCR-DGGE and SSCP analysis[J]. International Journal of Dairy Technology, 2016, 69(4): 592-600
- [26] Moon SB, Lee SS. Erratum to: *Roseovarius algicolicus* sp. nov., isolated from culture fluid of *Cochlodinium polykrikoides*[J]. Antonie van Leeuwenhoek, 2016, 109(1): 169
- [27] Peng GX, Wu Z, Luo HF, et al. *Enterobacter oryzae* sp. nov., a nitrogen-fixing bacterium isolated from the wild rice species *Oryza latifolia*[J]. International Journal of Systematic and Evolutionary Microbiology, 2009, 59(7): 1650-1655
- [28] Kim O, Cho YJ, Lee K, et al. Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species[J]. International Journal of Systematic and Evolutionary Microbiology, 2012, 62(3): 716-721
- [29] Mardis ER. Next-generation DNA sequencing methods[J]. Annual Review of Genomics and Human Genetics, 2008, 9(1): 387-402
- [30] Krause A, Ramakumar A, Bartels D, et al. Complete genome of the mutualistic, N(2)-fixing grass endophyte *Azoarcus* sp. strain BH72[J]. Nature Biotechnology, 2006, 24(11): 1385-1391
- [31] Fouts DE, Tyler HL, DeBoy RT, et al. Complete genome sequence of the N<sub>2</sub>-fixing broad host range endophyte *Klebsiella pneumoniae* 342 and virulence predictions verified in mice[J]. PLoS Genetics, 2008, 4(7): e1000141
- [32] Bertalan M, Albano R, De Pádua V, et al. Complete genome sequence of the sugarcane nitrogen-fixing endophyte *Gluconacetobacter diazotrophicus* Pal5[J]. BMC Genomics, 2009, 10: 450
- [33] Kaneko T, Minamisawa K, Isawa T, et al. Complete genomic structure of the cultivated rice endophyte *Azospirillum* sp. B510[J]. DNA Research, 2010, 17(1): 37-50
- [34] Pedrosa FO, Monteiro RA, Wasseem R, et al. Genome of *Herbaspirillum seropedicae* strain SmR1, a specialized diazotrophic endophyte of tropical grasses[J]. PLoS Genetics, 2011, 7(5): e1002064
- [35] Gan HM, McGroty SE, Chew TH, et al. Whole-genome sequence of *Enterobacter* sp. strain SST3, an endophyte isolated from Jamaican sugarcane (*Saccharum* sp.) stalk tissue[J]. Journal of



- Bacteriology, 2012, 194(21): 5981-5982
- [36] Zhu B, Chen MY, Lin L, et al. Genome sequence of *Enterobacter* sp. strain SP1, an endophytic nitrogen-fixing bacterium isolated from sugarcane[J]. Journal of Bacteriology, 2012, 194(24): 6963-6964
- [37] Kwak MJ, Song JY, Kim SY, et al. Complete genome sequence of the endophytic bacterium *Burkholderia* sp. strain KJ006[J]. Journal of Bacteriology, 2012, 194(16): 4432-4433
- [38] Liu WY, Chung MK, Wong CF, et al. Complete genome sequence of the endophytic *Enterobacter cloacae* subsp. *cloacae* Strain ENHKU01[J]. Journal of Bacteriology, 2012, 194(21): 5965
- [39] Witzel K, Gwinn-Giglio M, Nadendla S, et al. Genome sequence of *Enterobacter radicincitans* DSM16656(T), a plant growth-promoting endophyte[J]. Journal of Bacteriology, 2012, 194(19): 5469
- [40] Madhaiyan M, Peng N, Ji LG. Complete genome sequence of *Enterobacter* sp. strain R4-368, an endophytic N-fixing gammaproteobacterium isolated from surface-sterilized roots of *Jatropha curcas* L.[J]. Genome Announcements, 2013, 1(4): e00544-13
- [41] Crook MB, Mitra S, Ané JM, et al. Complete genome sequence of the sesbania symbiont and rice growth-promoting endophyte *Rhizobium* sp. strain IRBG74[J]. Genome Announcements, 2012, 1(6): e00934-13
- [42] Humann JL, Wildung M, Pouchnik D, et al. Complete genome of the switchgrass endophyte *Enterobacter cloacae* P101[J]. Standards in Genomic Sciences, 2014, 9(3): 726-734
- [43] Straub D, Rothballer M, Hartmann A, et al. The genome of the endophytic bacterium *H. frisingense* GSF30<sup>T</sup> identifies diverse strategies in the *Herbaspirillum* genus to interact with plants[J]. Frontiers in Microbiology, 2013, 4: 168
- [44] Madhaiyan M, Chan KL, Ji LH. Draft genome sequence of *Methylobacterium* sp. strain L2-4, a leaf-associated endophytic N-fixing bacterium isolated from *Jatropha curcas* L.[J]. Genome Announcements, 2014, 2(6): e1306-e1314
- [45] Jeong H, Choi SK, Kloepper JW, et al. Genome sequence of the plant endophyte *Bacillus pumilus* INR7, triggering induced systemic resistance in field crops[J]. Genome Announcements, 2014, 2(5): e01093-14
- [46] Hanak AM, Nagler M, Weinmaier T, et al. Draft genome sequence of the growth-promoting endophyte *Paenibacillus* sp. P22, isolated from *Populus*[J]. Genome Announcements, 2014, 2(2): e00276-14
- [47] Gan HY, Gan HM, Savka MA, et al. Whole-genome sequences of 13 endophytic bacteria isolated from Shrub Willow (*Salix*) grown in Geneva, New York[J]. Genome Announcements, 2014, 2(3): e00288-14
- [48] Lin L, Wei CY, Chen MY, et al. Complete genome sequence of endophytic nitrogen-fixing *Klebsiella variicola* strain DX120E[J]. Standards in Genomic Sciences, 2015, 10: 22
- [49] Meng XF, Bertani I, Abbruscato P, et al. Draft genome sequence of rice endophyte-associated isolate *Kosakonia oryzae* KO348[J]. Genome Announcements, 2015, 3(3): e00594-15
- [50] Schicklberger M, Shapiro N, Loqué D, et al. Draft genome sequence of *Raoultella terrigena* R1Gly, a diazotrophic endophyte[J]. Genome Announcements, 2014, 3(3): e00607-e00615
- [51] Shinjo R, Uesaka K, Ihara K, et al. Complete genome sequence of *Kosakonia sacchari* strain BO-1, an endophytic diazotroph isolated from a sweet potato[J]. Genome Announcements, 2016, 4(5): e00868-16
- [52] Behera P, Vaishampayan P, Singh NK, et al. The draft genome sequence of *Mangrovibacter* sp. strain MP23, an endophyte isolated from the roots of *Phragmites karka*[J]. Genomics data, 2016, 9: 128-129
- [53] Chun J, Rainey FA. Integrating genomics into the taxonomy and systematics of the *Bacteria* and *Archaea*[J]. International Journal of Systematic and Evolutionary Microbiology, 2014, 64(Pt 2): 316-324
- [54] Konstantinidis KT, Tiedje JM. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead[J]. Current Opinion in Microbiology, 2007, 10(5): 504-509
- [55] Kim M, Oh HS, Park SC, et al. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes[J]. International Journal of Systematic and Evolutionary Microbiology, 2014, 64(Pt 2): 346-351
- [56] Goris J, Konstantinidis KT, Klappenbach JA, et al. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities[J]. International Journal of Systematic and Evolutionary Microbiology, 2007, 57(Pt 1): 81-91
- [57] Rodriguez-R LM, Konstantinidis KT. Bypassing cultivation to identify bacterial species[J]. Microbe, 2014, 9(3): 111-118
- [58] Liu B, Hu GP, Tang WQ. Characteristic of average nucleotide identity (ANI) based on the whole genomes from *Bacillus* species in *Bacillus*-like Genus[J]. Fujian Journal of Agricultural Sciences, 2013, 28(9): 833-843 (in Chinese)  
刘波, 胡桂萍, 唐唯其. 基于全基因组的芽孢杆菌平均核苷酸同源性(ANI)分析[J]. 福建农业学报, 2013, 28(9): 833-843
- [59] Richter M, Rosselló-móra R. Shifting the genomic gold standard for the prokaryotic species definition[J]. Proceedings of the National Academy of Sciences of the United States of America, 2009, 106(45): 19126-19131
- [60] Zhang YM, Tian CF, Sui XH, et al. Robust markers reflecting phylogeny and taxonomy of rhizobia[J]. PLoS One, 2012, 7(9): e44936
- [61] Chen WF. Progress and perspective of systematics of rhizobia[J]. Microbiology China, 2016, 43(5): 1095-1100 (in Chinese)  
陈文峰. 根瘤菌系统学研究进展与展望[J]. 微生物学通报, 2016, 43(5): 1095-1100
- [62] Deloger M, El Karoui M, Petit MA. A genomic distance based on MUM indicates discontinuity between most bacterial species and genera[J]. Journal of Bacteriology, 2009, 191(1): 91-99
- [63] Ormeño-Orrillo E, Menna P, Almeida LGP, et al. Genomic basis of broad host range and environmental adaptability of *Rhizobium tropici* CIAT 899 and *Rhizobium* sp. PRF81 which are used in inoculants for common bean (*Phaseolus vulgaris* L.)[J]. BMC Genomics, 2012, 13(1): 735
- [64] Tiller RV, Gee JE, Frace MA, et al. Characterization of novel *Brucella* strains originating from wild native rodent species in north Queensland, Australia[J]. Applied and Environmental Microbiology, 2010, 76(17): 5837-5845
- [65] Schleifer KH. Classification of *Bacteria* and *Archaea*: past,

- present and future[J]. *Systematic and Applied Microbiology*, 2009, 32(8): 533-542
- [66] Lefebvre T, Bitar PDP, Suzuki H, et al. Evolutionary dynamics of complete *Campylobacter* pan-genomes and the bacterial species concept[J]. *Genome Biology and Evolution*, 2010, 2(1): 646-655
- [67] Leekitcharoenphon P, Lukjancenko O, Friis C, et al. Genomic variation in *Salmonella enterica* core genes for epidemiological typing[J]. *BMC Genomics*, 2012, 13(1): 88
- [68] Sun ZH, Harris HMB, McCann A, et al. Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera[J]. *Nature Communications*, 2015, 6: 8322
- [69] Xie JB. Isolation, identification and comparative genomics of nitrogen-fixing *Paenibacillus*[D]. Beijing: Doctoral Dissertation of China Agricultural University, 2015 (in Chinese)  
谢剑波. 固氮类芽孢杆菌的分离鉴定及比较基因组学研究[D]. 北京: 中国农业大学博士学位论文, 2015
- [70] Kaas RS, Friis C, Ussery DW, et al. Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes[J]. *BMC Genomics*, 2012, 13: 577
- [71] Hui WY, Zhang HP. Application of genomic analysis in Microbial Taxonomy[J]. *Microbiology China*, 2016, 46(5): 1136-1142 (in Chinese)  
惠文彦, 张和平. 基因组分析方法在微生物分类学中的应用[J]. *微生物学通报*, 2016, 46(5): 1136-1142
- [72] Li YY. Systematic taxonomy of bacteria belonging to the genus *Kosakonia* based on the whole genome sequences[D]. Hangzhou: Master's Thesis of Zhejiang University, 2016 (in Chinese)  
李媛媛. 基于全基因组序列系统分类 *Kosakonia* 属的细菌[D]. 杭州: 浙江大学硕士学位论文, 2016
- [73] Qi J, Wang B, Hao BI. Whole proteome prokaryote phylogeny without sequence alignment: A *K*-string composition approach[J]. *Journal of Molecular Evolution*, 2004, 58(1): 1-11
- [74] Elloumi M, Zomaya AY. New Developments in Processing of Degenerate Sequences[A]//Antoniou P, Lliopoulos CS. Algorithms in Computational Molecular Biology: Techniques, Approaches and Applications. Hoboken: John Wiley & Sons, Inc., 2010: 73-90
- [75] Zuo GH, Zhao X, Hao BL. *Shigella* strains are not clones of *Escherichia coli* but sister species in the Genus *Escherichia*[J]. *Genomics, Proteomics & Bioinformatics*, 2013, 11(1): 61-65
- [76] Qi J, Luo H, Hao B, et al. CVTree: a phylogenetic tree reconstruction tool based on whole genomes[J]. *Nucleic Acids Research*, 2004, 32(W1): W45-W47
- [77] Chan JZM, Halachev MR, Loman NJ, et al. Defining bacterial species in the genomic era: insights from the genus *Acinetobacter*[J]. *BMC Microbiology*, 2012, 12: 302
- [78] Yuan JB, Zhu QM, Liu B. Phylogenetic and biological significance of evolutionary elements from metazoan mitochondrial genomes[J]. *PLoS One*, 2014, 9(1): e84330
- [79] Kislyuk AO, Haegeman B, Bergman NH, et al. Genomic fluidity: an integrative view of gene diversity within microbial populations[J]. *BMC Genomics*, 2011, 12(1): 32
- [80] Snipen LG, Ussery DW. A domain sequence approach to pangenomics: applications to *Escherichia coli*[J]. *F1000research*, 2012, 1: 19
- [81] Zwick ME, Joseph SJ, Didelot X, et al. Genomic characterization of the *Bacillus cereus* sensu lato species: backdrop to the evolution of *Bacillus anthracis*[J]. *Genome Research*, 2012, 22(8): 1512-1524
- [82] Lan YM, Morrison JC, Hershberg R, et al. POGO-DB-a database of pairwise-comparisons of genomes and conserved orthologous genes[J]. *Nucleic Acids Research*, 2014, 42(D1): D625-D632
- [83] Martínez-Porchas M, Vargas-Albores F. Microbial metagenomics in aquaculture: a potential tool for a deeper insight into the activity[J]. *Reviews in Aquaculture*, 2017, 9(1): 42-56