

Pseudomonas stutzeri A1501 基因组结构及功能注释

燕永亮¹, 杨剑², 窦岳坦¹, 平淑珍¹, 王忆平³, 金奇², 林敏¹

(1. 中国农业科学院生物技术研究所, 北京 100081; 2. 病毒基因工程国家重点实验室, 北京 100176;
3. 北京大学生命科学学院, 北京 100871)

摘要:采用全基因组“shotgun”方法完成了固氮斯氏假单胞菌 A1501 的全基因组序列测定,并进行了基因组结构与功能注释分析。A1501 基因组全长 4 567 418 bp,含有 4 146 个 ORFs。该基因组中已鉴定了 42 个编码转座酶的重复序列,这些序列的存在预示着转座现象在 A1501 菌中非常活跃,预示该菌与其他生物之间基因交流可能比较频繁。比较基因组表明,为了适应特定的生存环境,假单胞菌在基因组结构和遗传信息容量上产生了明显的分化。此外,基因组分析鉴定了 A1501 环境适应的遗传基础,包括物质转运、信号传导和趋化系统等,这些系统是细菌能够在根际土壤环境中保持竞争力以及能够与水稻形成高效联合固氮体系的关键。A1501 基因组的完成为进一步开展功能基因组学和蛋白质组学研究奠定了基础。

关键词:基因组测序;斯氏假单胞菌 A1501;联合固氮;比较基因组学

中图分类号:Q811.4 **文献标识码:**A **文章编号:**1008-0864(2008)05-0045-08

Genome Structure and Functional Annotation of Nitrogen-fixing *Pseudomonas stutzeri* A1501

YAN Yong-liang¹, YANG Jian², DOU Yue-tan¹, PING Shu-zhen¹, WANG Yi-ping³, JIN Qi², LIN Min¹

(1. Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Beijing 100081;

2. State Key Laboratory for Molecular Virology and Genetic Engineering, Beijing 100176;

3. College of Life Sciences, Peking University, Beijing 100871, China)

Abstract:The genome sequencing of *Pseudomonas stutzeri* A1501 was accomplished by using the "shotgun" strategy, then the genome structure and gene function annotation analysis were also carried out. The genome of *P. stutzeri* A1501 is composed of 4 567 418 bp, with 4 146 ORFs. In the genome, 42 copies of repeat sequences encoding multiple transposases were identified, implying transposition in A1501 was very active and the lateral gene transfer was frequently between A1510 and other organisms. Comparative genomics showed that in order to adapt to the niche lifestyle, *Pseudomonas* strains modulate their genome structure and genome contents. The genetic basis of A1501 including the transport system, signal transduction system and the chemotaxis system etc. are identified by genome analysis. These systems are key points for bacteria to remain competitive in the rhizosphere environment, and to construct an associative high-nitrogen-fixing system with rice. The accomplishment of A1501 genome is a foundation for further transcriptomic and proteomics study.

Key words:genome sequencing, *Pseudomonas stutzeri* A1501, associative nitrogen fixation, comparative genomics

氮素是影响农业生产的重要元素之一。自然界中可被作物直接利用的氮素主要来自生物固氮。固氮微生物种类繁多,根据固氮生物与植物之间的相互关系,通常将固氮体系分为自生固氮、

共生固氮和联合固氮三种。与以根瘤菌为代表的共生固氮菌不同,联合固氮菌不能形成特异位化结构,微生物通过定植于植物根表或侵入内根际进行固氮作用。联合固氮的研究开展的较晚,仅有

收稿日期:2008-09-02;修回日期:2008-09-18

基金项目:国家 973 计划项目(2007CB707805、2007CB109203 及 2001CB108904),国家 863 计划项目(2007AA021304 和 2006AA020101)及国家自然科学基金项目(30470047 和 30200007)资助。

作者简介:燕永亮,博士,主要从事固氮微生物分子遗传学及功能基因组学研究。E-mail: yongliangyan@yahoo.com.cn。

通讯作者:林敏,研究员,博士,博士生导师,主要从事固氮微生物及特殊环境微生物的分子生物学和基因工程研究。
Tel:010-82106145;E-mail: linmin57@vip.163.com

30 多年的历史,但是由于联合固氮菌在改善禾本科植物氮素营养方面的潜在价值,特别是具有植物内生作用的联合固氮菌的利用有可能为非豆科植物打开一条“体内固氮”的新途径,所以对于这类菌的分子遗传和基因工程研究已经成为该领域的研究热点之一^[1-3]。

斯氏假单胞菌 A1501 (*Pseudomonas stutzeri*) 于 20 世纪 80 年代分离自我国南方水稻根际土壤^[4]。该菌在无氮和微好氧条件下具有良好的固氮活性,而且固氮产物能够迅速被水稻直接吸收利用^[5,6]。30 多年来,科学家们对该菌的生理生化以及遗传特性进行了深入研究,先后克隆了斯氏假单胞菌 A1501 中的固氮酶相关基因以及固氮调节基因 *ntrC*^[7,8] 等。最近的研究表明,该菌的固氮遗传信息分布在一个长 49kb, 编码 59 个基因的 DNA 区域上,该区域基因的突变可引起 A1501 固氮能力的减弱甚至丧失^[9]。A1501 固氮基因的表达受到一般氮代谢 *ntr* 系统和固氮基因表达调控系统 *nifLA* 的严谨调控。通过酵母双杂交和体外免疫共沉淀研究发现,A1501 的 NifA 和 NifL 蛋白可通过分子间直接接触而发生互作,进而开启和关闭固氮基因的表达^[10],而 PII 蛋白 GlnK 可能通过与 NifA 竞争 NifL 的作用位点从而影响固氮基因的表达^[11]。除固氮作用外,在厌氧条件下,A1501 菌还可以乳酸为碳源, NH_4^+ 为唯一氮源进行反硝化作用,将 NO_3^- 还原为 N_2 ^[12]。

联合固氮菌与宿主水稻间建立的是一种“松散”的共生关系,其相互作用缺乏严格的专一性,同时,联合固氮效率还受到三个主要限制因素的制约,即铵抑制、能量限制和根表竞争,研究难度较大,需要借助新的研究平台寻求突破。此外,A1501 菌兼具了生物固氮、硝化、反硝化等多种氮代谢途径,这些代谢途径之间的网络调控也需要从系统学的角度进行诠释。在这种背景下,我们开展了斯氏假单胞菌 A1501 全基因组序列测定工作,该基因组测序的完成不仅可以加深我们对 A1501 菌遗传背景的了解与研究,同时也为进一步开展该菌的转录组学和功能基因组学研究奠定基础。

1 材料与方 法

1.1 菌株与载体

斯氏假单胞菌 A1501 株为中国农业科学院

生物技术研究所保存,该菌于 1980 年分离自中国南方水稻根际土壤,经纯化、复壮保存。

pUC19,购自 Promega 公司。经限制性内切酶 *Sma* I 酶切、去磷酸化处理得到线性化载体,用于构建基因组随机 DNA 文库。

pGEM-T Easy 质粒,购自 Promega 公司,用于 PCR 产物的克隆。

E. coli DH5 α ,本实验室保存用于质粒的扩增与转化。

1.2 工具酶及各种试剂盒

*Eco*R I、*Hind* III、*Bam*H I、*Kpn* I、*Pst* I、*Sma* I 等限制性内切酶、T4 DNA 连接酶、碱性磷酸化酶 CIP 均购于 New England Biolabs 公司;T4 DNA Polymerase 购于 TaKaRa 公司;RNaseA, Proteinase K 购于 Promega 公司;Taq DNA 聚合酶以及 LA Taq DNA 聚合酶均购于 TaKaRa 公司;QIAprep 8 Miniprep kit 为质粒提取试剂盒,QIAEX II kit 用于 DNA 凝胶回收,购自 QIAGEN 公司;Millipore kit 用于 DNA 测序模板提取,购自 Millipore 公司;BIG-DIY kit 为 DNA 测序试剂盒,购自 PE 公司。

1.3 各种化学试剂

X-gal、IPTG 等购于 Promega 公司;Tris、ED-TA、Agrose、SDS 购于 GIBCO BRL 公司;溴化乙锭、氨苄青霉素购于 Sigma 公司。胰蛋白胨 (Tryptone) 和酵母浸出物 (Yeast extract) 购自 OX-IOD 公司,其他常用化学试剂为国产试剂,均为分析纯。

1.4 实验方法

采用全基因组“shotgun”的方法,首先将 A1501 染色体 DNA 超声破碎,回收不同目的片段,构建插入片段分别为 2~4 kb(S 文库)和 4~6 kb(L 文库)的基因组随机文库。对 L 文库进行两端测序,打造基因组的框架,对 S 文库进行单向测序,填充框架内信息。采用 ABI 3700 测序仪进行大规模测序,测序总量达到基因组 6~7 倍的覆盖率。使用华盛顿大学基因组中心的 Phil Green 和 Brent Ewing 等开发的 Phred-Phrap^[13,14] 软件包进行序列的拼接工作。使用 glimmer 软件 2.01 版进行^[15] ORFs 的预测及与现有 COGs 和 NR 蛋白序列库作 BLASTP 比对。

1.5 核酸序列登记号

斯氏假单胞菌 A1501 全基因组序列已提交

GenBank, 登记号为 CP000304。

2 结果与分析

2.1 *P. stutzeri* A1501 基因组序列测定

2.1.1 染色体随机文库的构建 *P. stutzeri* A1501 全基因组序列测定采用“shotgun”法, 染色体 DNA 提取纯化后, 采用超声波破碎断裂, 本研究超声波处理的条件为: 输出功率 60W, 脉冲时间 2 s, 脉冲次数为 3 次, 此时染色体 DNA 片段完全被打断, 呈现一条弥散的带, 其中 80% 的片段长度范围集中在 1 ~ 10 kb 内, 符合文库构建的要求。

2.1.2 大规模基因组测序 染色体破碎后, 分别切胶回收 2 ~ 4 kb 和 4 ~ 6 kb 的 DNA 片段, 末端补平后与末端去磷酸化处理过的 pUC19 载体连接, 转化大肠杆菌 DH5 α , 构建基因组随机文库。使用 ABI 3700 测序仪对制备好的 37 152 个测序模板进行测序, 其中对 S 文库模板进行单向测序, 对 L 文库模板进行双向测序。本研究共进行了 53 568 个测序反应, 得到 50 408 个有效读长 (reads), 测序成功率为 94.1%, reads 的平均有效长度为 609 bp, 累加总长度为 30.7 Mbp, 约为基因组的 7 倍覆盖率。采用 Phred-Phrap 软件包对所获得的 50 408 个有效 reads 进行序列的拼接, 获得 126 个 contigs (多 reads 拼接成的序列群), 在形成的 126 个 Gap (缺口) 中, 114 个为序列缺口 (有质粒跨过该区域, 具有明确位置关系), 12 个为物理缺口 (位置关系不明确, 文库中不包括此区域)。通过 primer walking 和随机 PCR 的方法进行补 Gap 反应, 获得 A1501 全长基因组序列。随后, 为了保证序列的准确以及染色体框架的完整, 本研究通过模板克隆染色体定位以及低质量区域重复测序等方法对染色体框架以及测序低质量区域进行了验证。

2.2 *P. stutzeri* A1501 基因组功能注释分析

2.2.1 *P. stutzeri* A1501 染色体基本信息 斯氏假单胞菌 A1501 的全基因组是一条环形染色体, 没有质粒的存在, 这与该菌染色体的 PFGE 电泳结果相符 (数据未发表)。序列分析表明, A1501 菌基因组具有以下基本特征: 基因组全长 4 567 418 bp, 预测出 4 146 个 ORFs, G + C 含量为 63.8%, 有 4 套 16S/23S/5S rRNA 操纵子区域

和 59 个 tRNA 基因。此外, 基因组中共鉴定了 42 处重复序列 (图 1), 这些重复序列分别编码 1 ~ 2 个转座酶, 具有典型的 IS 序列特征, 但是具体功能需要进一步验证。G - C/G + C 倾斜曲线是反应基因组中碱基分布均一性的参数, 在 A1501 菌基因组中, 以复制原点为中心, 该曲线呈现明显的对称分布, 表明了 A1501 菌染色体中碱基分布均匀, 具有良好的对称性。

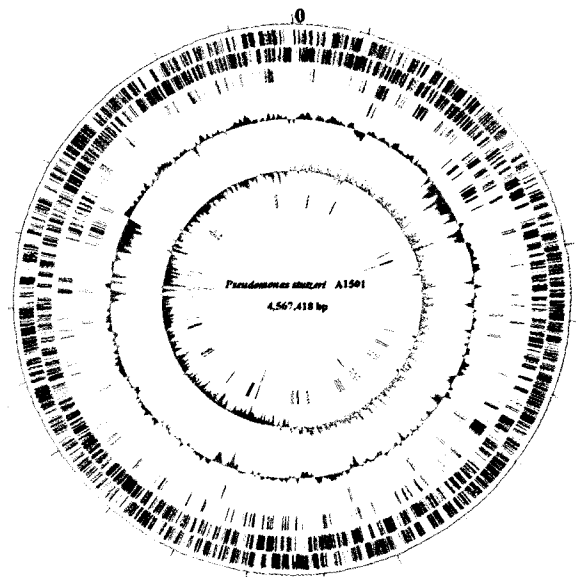


图 1 斯氏假单胞菌 A1501 的染色体基因组环形图谱

Fig. 1 Circular representation of the *P. stutzeri* A1501 genome.

注: 每个刻度表示基因组上 200 kb, 0 代表复制起始点。由外向内每环分别代表基因组上不同的相关信息, 第一和第二环分别代表正链及负链所编码的 ORFs, 不同颜色代表不同功能, 与 COGs 功能组统一; 第三环代表 A1501 菌株特异基因在染色体上的分布; 第四环表示重复序列 (RP) 在染色体上的分布; 第五环代表基因组的 G + C 含量, 绿色表示该区域含量 > 63.8% (平均值), 深绿色代表 < 63.8%; 第六环代表 G - C/G + C 倾斜曲线, 浅黄色表示大于零, 橙黄色表示小于零; 第七环代表 tRNA 在基因组上的分布; 第八环代表 rRNA 操纵子的分布。

Circular representation of the *P. stutzeri* A1501 genome. Outer scale is marked in 200 kb. (From the outer to the inner concentric circle) Circle 1 and 2, genes encoded by leading and lagging strands with color codes for functions by COG categories. Circle 3, distribution of no hits (specific to A1501) genes. Circle 4, distribution of the RP sequence. Circle 5 and 6, G + C content and GC skew (G - C/G + C) with a window size of 10 kb, respectively. Circle 7 and 8, distribution of tRNA genes and rRNA operons, respectively.

根据编码基因与已知功能基因的同源关系, 可以将 A1501 菌的 4 146 个基因划分为三类: 与其他生物中已知功能基因具有同源性、与已报道的未知功能的基因同源以及与所有已报道基因都

没有同源性。其中第一类基因有 3 366 个,占 ORFs 总数的 81.2%;第二类基因有 473 个,占 ORFs 总数的 11.4%;第三类基因,即 A1501 基因组独特的 ORFs 有 316 个,占 ORFs 总数的 7.6%。这些基因的编码蛋白质按照 COGs (Cluster of orthologous groups of proteins) 的标准可以划分为功能组(20 个)、未知功能组(1 个)和假设功能组(1 个)三类,共 22 个组(表 1)。由表 1 可以看出,在 A1501 中,能量代谢是其生命活动中最活

跃的部分,总共涉及全基因组中 1 409 个 ORFs,占整个基因组 ORFs 的 34%;有 192 个基因参与了外界信号经细胞膜向胞内的传导过程;597 个基因参与了遗传信息的复制、转录及翻译;由于基因的表达受到很多内、外界因素的影响,所以转录的调节是一个非常复杂的过程,有大量调节因子参与其中,因此,在遗传信息的遗传及传递过程中,参与转录及调节的基因是最多的,有 223 个。

表 1 斯氏假单胞菌 A1501 基因组编码蛋白按 COGs 分类表
Table 1 Classification of the A1501 proteins according to the COGs standard.

| 功能组分类 Functional groups | ORFs |
|--|------|
| 信息保存以及传递 Information storage and processing | |
| [J] 翻译,核糖体结构和生物合成 Translation, ribosomal structure and biogenesis | 181 |
| [A] RNA 修饰 RNA processing and modification | 1 |
| [K] 转录及调节 Transcription and regulation | 223 |
| [L] 染色体复制、重组以及修复 Replication, recombination and repair | 193 |
| [B] 染色体结构以及动力 Chromatin structure and dynamics | 1 |
| 细胞过程及信号传递 Cellular processes and signaling | |
| [D] 细胞周期控制、细胞分裂及染色体分裂 Cell cycle control, cell division, chromosome partitioning | 32 |
| [V] 防卫机制 Defense mechanisms | 46 |
| [T] 信号转导机制 Signal transduction mechanisms | 191 |
| [M] 细胞壁/膜/外鞘的生物合成 Cell wall/membrane/envelope biogenesis | 173 |
| [N] 细胞的运动 Cell motility | 112 |
| [U] 胞内运输以及分泌 Intracellular trafficking, secretion, and vesicular transport | 89 |
| [O] 翻译后修饰,蛋白折叠及分子伴侣 Posttranslational modification, protein turnover, chaperones | 162 |
| 物质代谢 Metabolism | |
| [C] 能量的产生与转化 Energy production and conversion | 252 |
| [G] 碳水化合物运输及代谢 Carbohydrate transport and metabolism | 162 |
| [E] 氨基酸转运及代谢 Amino acid transport and metabolism | 312 |
| [F] 核酸转运及代谢 Nucleotide transport and metabolism | 72 |
| [H] 辅酶转运及代谢 Coenzyme transport and metabolism | 132 |
| [I] 脂类转运及代谢 Lipid transport and metabolism | 142 |
| [P] 无机离子转运及代谢 Inorganic ion transport and metabolism | 230 |
| [Q] 次生物质的生物合成、转运及代谢 Secondary metabolites biosynthesis, transport and catabolism | 107 |
| 功能不明确 Poorly characterized | |
| [R] 仅有功能提示 General function prediction only | 429 |
| [S] 功能未知 Function unknown | 280 |

2.2.2 *P. stutzeri* A1501 基因组中的重复序列及转座酶 在 A1501 菌的基因组中鉴定了 42 个重复序列区域(不包括 16S/23S/5S 区域)。这 42 个重复序列分为 10 种类型,编号分别为 RP1 ~ RP10,长度从 1 110 bp 到 2 639 bp 不等(表 2),总长度为 61 728 bp,占整个基因组的 1.35%。它

们在基因组上的分布没有明显的规律,拷贝数也不尽相同,其中 RP6 的拷贝数最多,有 10 处,包括 4 处完整的以及 6 处不完整区域,而 RP2、RP5 和 RP7 在基因组中都只有 2 个重复。

从图 1 中可以看出,这 42 处 RP 序列存在区域的 G + C 含量明显要低于基因组其他区域

表 2 A1501 菌基因组中 RP 的类型及分布

Table 2 Number and the protein productions of the repeat sequences in *P. stutzeri* A1501.

| 名称 Name | 长度 (bp) Length (bp) | ORFs | 拷贝数(完整 + 部分) Copy number (intact + partial) |
|------------|---------------------------|-------------------|---|
| RP1 | 2 016 | 2 ^a | 6 |
| RP2 | 1 687 | TnpA ^b | 2 |
| RP3 | 1 110 | TnpA ^b | 5 |
| RP4 | 1 306 | TnpA ^b | 3 |
| RP5 | 1 198 | 2 ^a | 2 |
| RP6 | 2 639 | 2 ^a | 10(4 + 6) |
| RP7 | 2 605 | 2 ^a | 2 |
| RP8 | 2 372 | 2 ^a | 4(3 + 1) |
| RP9 | 1 231 | 1 ^a | 5(3 + 2) |
| RP10 | 1 113 | 1 ^a | 3 |

注: a 表示转后酶; b 表示 TnpA 家族的转座子,

Note: a indicates transposases; b indicates TnpA family transposon.

的 G + C 含量。对于基因组中 G + C 含量的不均一性, 有两种解释: 一是认为具有明显不同 G

+ C 含量的区段, 可能是在近代进化中有着不同的来源; 第二种解释就是这些区域往往是特殊功能区。对 RP 序列的分析表明, 这些序列都可以含有 1 ~ 2 个 ORFs, 这些 ORFs 分别与已知的转座酶同源, 其中 RP2、RP3 和 RP4 根据其转座酶特性可以归属于 TnpA 家族的转座子, 由此可见, 在 A1501 菌的基因组中, 转座现象是非常活跃的, 预示着该菌与其他生物之间基因交流可能比较频繁。

A1501 菌基因组的平均 G + C 含量为 63.8%, 但是在 A1501 菌中有几个比较大的低 G + C 含量区域(图 1), 涉及到多个 ORFs 及功能区(表 3)。特别是 PST0570-0710 和 PST3343-3476 区域, 长度都大于 100 kb, 且具有一个明显的共同点, 即该区域所涉及的 ORFs 多为功能未知, 而且非常频繁地出现编码转座酶的 RP 序列。同样的情况, 在编号为 1, 2, 7, 8 的区域也有 RP 序列的存在, 预示着这些基因可能通过基因的水平转移来源于其他微生物种群。

表 3 A1501 菌基因组中 G + C 含量明显偏低的区域及编码蛋白功能描述

Table 3 Functional description of the DNA regions with low G + C percent in *P. stutzeri* A1501.

| 编号 Number | ORFs | (G + C)% | 长度(kb) Total length(kb) | 功能描述 Functional description | 是否涉及转座酶 Related to transposases |
|--------------|----------------|----------|-------------------------------|--|------------------------------------|
| 1 | PST0224 - 0259 | 59.8 | 35 | 砷酸盐还原酶, I 型限制 - 修饰系统 Arsenate reductase, type I restriction-modification system | 是/Yes |
| 2 | PST0552 - 0559 | 56.6 | 7 | Type4 菌毛合成蛋白 Type4 fimbrial biogenesis protein | 是/Yes |
| 3 | PST0570 - 0710 | 57.4 | 137 | 聚羟基脂肪酸合成, 功能未知蛋白 Polyhydroxyalkanoate synthesis, hypothetical protein | 是/Yes |
| 4 | PST0763 - 0810 | 58.2 | 34 | 核糖体组成蛋白 Ribosomal protein | 否/No |
| 5 | PST0968 - 0975 | 55.2 | 7 | Type4 菌毛合成蛋白 Type4 fimbrial biogenesis protein | 否/No |
| 6 | PST1050 - 1070 | 59 | 19 | Type4 菌毛合成蛋白, 核糖体组成蛋白, 泛醌细胞色素 C 还原酶 Type4 fimbrial biogenesis protein, ribosomal protein and ubiquinol-cytochrome C reductase | 否/No |
| 7 | PST1464 - 1489 | 58.4 | 30 | I 型限制 - 修饰系统, 功能未知蛋白 Type I restriction-modification system, hypothetical protein | 是/Yes |
| 8 | PST1859 - 1879 | 58.4 | 20 | 琥珀酸盐脱氢酶 Succinate dehydrogenase | 是/Yes |
| 9 | PST3343 - 3476 | 56 | 141 | 转录调控因子、转座酶, 功能未知蛋白 Transcriptional regulator, transposase, hypothetical protein | 是/Yes |

2.2.3 *P. stutzeri* A1501 与其他假单胞菌的比较基因组学 *P. stutzeri* A1501 是斯氏假单胞种内第一个完成基因组测定的菌株。在此之前,假单胞属基因组测序已全面展开,共有 10 多个菌株已完成了基因组测定,这些菌株分别属于 5 个不同的种。选择其中具有代表性的假单胞菌的基因组序列与 A1501 基因组进行了比较,它们是人类机会致病菌铜绿假单胞菌(*P. aeruginosa*) PAO1^[16]、植物共生荧光假单胞(*P. fluorescens*) pf-5^[17]、土壤腐生菌恶臭假单胞(*P. putida*) KT2440^[18]、植物病原菌丁香假单胞(*P. syringae*) DC3000^[19] 和虫媒假单胞菌(*P. entomophila*) L48^[20]。

斯氏假单胞菌 A1501 的基因组比其他已完成基因组测序的假单胞菌的基因组小得多,只有其他五个代表菌株基因组大小的 64.6% ~ 77.6%。相比其他的假单胞菌,在基因组水平上 A1501 与铜绿假单胞菌的同源性最高。A1501 基因组中的 2 770 个基因(66.8%)都可以在铜绿假单胞菌 PAO1 基因组上找到同源基因。通过基因组的比较,A1501 基因组中 1 997 个基因是所有已完成基因组测定的假单胞菌所共有的,它们共同构成了假单胞菌的基因组骨架(backbone)。在铜绿假单胞菌 PAO1 的基因组中所有涉及毒力与致病性的基因在 A1501 中都没有发现,比如 III 型和 VI 型分泌系统、2 种群体感应信号分子的合成、藻酸盐聚合物的合成以及铁载体和抗生素的合成等。这可能是 A1501 能够侵入植物体内,但并不造成植物病害的一个原因。

有 487 个基因是 A1501 基因组所独缺,但是存在于其他假单胞菌中。这些基因中,有 163 个基因编码未知功能蛋白,另外 324 个基因的编码蛋白具有可预测的功能。这些基因主要包括:一个由 14 个基因组成的脓菌素(一种细菌分泌的抗菌物)合成基因簇、一系列编码细胞色素 O 泛醌氧化酶的基因、由 10 个基因组成的藻酸盐合成基因簇、由 12 个基因组成的 NADH 脱氢酶 I 合成的基因簇、参与维生素 B 和甜菜碱合成代谢的一系列基因,多个参与不同种类氨基酸转运及代谢的基因,包括精氨酸和组氨酸等,另外还有几个参与氢氰酸代谢的关键基因。

2.2.4 *P. stutzeri* A1501 环境适应的遗传基础

为了适应作物根际的生活环境,A1501 菌进化出了多种环境适应的机制,其中包括能够使细菌在

环境中获取足够营养物质的转运蛋白系统、使细菌向植物根际定向移动的趋化系统以及在水田环境中低氧分压下的固氮系统和厌氧条件下的无氧呼吸系统-反硝化系统等。固氮系统和反硝化系统的分子基础已经在相关的杂志上发表^[7,9,21],本文通过注解分析 A1501 的基因组,对 A1501 的物质转运、信号感应和趋化系统进行了研究。

P. stutzeri A1501 基因组中携带了 302 个编码转运蛋白的基因,主要负责向细胞内部输送氨基酸、糖类、脂肪酸等营养物质和无机物质,该系统保证了 A1501 菌与外界环境的营养物质和无机盐离子交换。

已在 A1501 基因组中鉴定了大量参与基因表达调控的基因,包括 137 个转录调节因子和 24 对二元调节系统(环境感受器)。大多数的转录调节蛋白根据其调控特性可以归类到不同调控蛋白家族,分别为 LysR (36 个)、AraC (21 个)、TetR (13 个)、MerR (11 个)、GntR (7 个)、Cro/C1 (5 个)、AsnC (3 个)、ArsR (3 个)、MarI (4 个)和 IclR (3 个)。二元调节系统主要包括感受器(sensor kinase)和应答调节因子(response regulator)两部分,通过对外界信息的感应与传递,这些系统可赋予斯氏假单胞菌 A1501 对环境变化产生相应应答的能力。大量的调节基因以及环境应答因子的存在是该菌能够具有很好的环境适应能力的基础。

趋化作用在细菌与植物根际联合体系的形成中起着重要作用。A1501 的趋化系统非常复杂,包括运动器官-鞭毛的合成及运动、外界信号的感应、信号的传递和信号的去除等多个过程。细菌对外界环境的感应受细胞膜受体的控制,其机制涉及多种特异蛋白质的甲基化、去甲基化作用。在 A1501 基因组中,与细菌运动及趋化性相关的基因有 112 个,在这些基因中,鞭毛结构蛋白是主体部分,有 37 个基因参与了鞭毛的生物合成。此外,在 A1501 基因组中还鉴定了 10 个趋化转导器(transducer)、3 个趋氧受体蛋白(*aer*, *aer-2*)、13 个趋化蛋白、6 个甲基接受趋化蛋白以及 2 个趋化蛋白甲基转移酶基因。

在 A1501 菌的基因组中还发现了 22 个与 IV 型菌毛的生物合成以及运动相关的基因(鉴定为 *pil* 和 *fim* 基因)。细菌的菌毛具有粘附于细胞表面的能力。菌毛根据其形态、数量、分布、粘附性

和抗原性可分为 I~VI 6 型,其中 I 型和 IV 型可粘附于人或动物血红细胞而引起血凝^[22]。血凝根据能否被 D-甘露糖抑制而分为甘露糖敏感(MS)和甘露糖耐受(MR)两类。其中 IV 型菌毛多为 MR 型菌毛。MR 型菌毛具有较强的表面粘附能力,又称为定居因子抗原。由此,我们可以推断,A1501 菌可能是通过菌毛的粘附作用而定居于植物根部,从而建立起联合固氮体系。

3 讨论

迄今为止,已有多达 10 余种假单胞菌的基因组完成了序列测定。对这些基因组的分析表明,为了适应特定的生存环境,不同的物种在基因组结构和遗传信息容量上产生了明显的分化。在假单胞菌中,A1501 基因组是最小的,只有其他假单胞菌基因组大小的 64.6%~77.6%。比较基因组数据表明 A1501 菌相对于病原菌铜铝假单胞 PAO1 和丁香假单胞菌 DC3000 缺少了 III 型和 VI 型分泌系统,这些系统是细菌侵入寄主细胞所必需的^[23]。该系统的缺失导致 A1501 菌不能侵入到水稻根部组织,只能粘附和定植在水稻根表。除此之外,A1501 基因组还缺少群体感应信号分子的合成、藻酸盐聚合物以及铁载体和抗生素的合成等基因,这些基因的缺失可能是 A1501 菌在进化过程中由于环境选择压的改变而逐渐丢失的。

A1501 菌基因组中 42 个编码重组酶的重复序列说明 A1501 菌与外界环境的基因交换非常频繁。试验已证明多个斯氏假单胞菌具有接纳和整合外源 DNA 的能力^[24,25]。这也就不难解释为什么 A1501 菌中含有一个长达 49 kb 的固氮基因岛的存在^[9],而其他假单胞菌中并没有固氮基因的发现。基因组分析揭示了 A1501 菌环境适应的遗传基础,这些系统是细菌能够在根际土壤环境中保持竞争力以及能够与水稻形成高效联合固氮体系的关键。

A1501 菌基因组的完成为在新的高度全面认识和分析该菌的固氮体系、各种代谢网络调节和寻找碳代谢及氮代谢之间的偶联作用及其分子机理提供了最原始的数据,为进一步开展功能基因组学和蛋白质组学研究奠定了基础,此外,A1501 基因全序列的完成为克隆获取新型的高效抗逆基

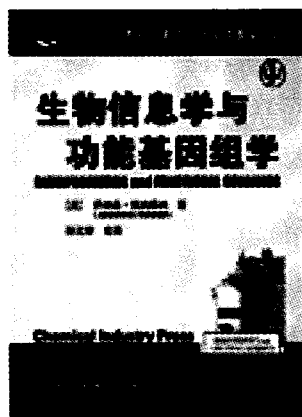
因提供了资源^[26],具有重要的理论价值和应用前景。

参 考 文 献

- [1] 尤崇杓. 固氮酶的结构和功能研究进展[J]. 农业生物技术学报, 1995, 3: 1-13.
- [2] 尤崇杓. 水稻根际联合固氮[M]. 北京: 农业出版社, 1990, 179-189.
- [3] 尤崇杓, 姜涌明, 宋鸿遇. 生物固氮[M]. 北京: 科学出版社, 1987.
- [4] 丘元盛, 周淑萍, 莫小真, 等. 水稻根联合固氮细菌的研究 I. 菌种的分离和鉴定[J]. 微生物学报, 1981, 21: 468-472.
- [5] 丘元盛, 周淑萍, 莫小真, 等. 稻根联合固氮细菌的研究 II. 粪产碱菌 A-15 和阴沟肠杆菌 E-26 的固氮特性[J]. 微生物学报, 1981, 21: 468-472.
- [6] Vermeiren H, Willems A, Schoofs G, et al. The rice inoculant strain A15 is a nitrogen-fixing *Pseudomonas stutzeri* strain [J]. Syst. Appl. Microbiol., 1999, 22: 215-224.
- [7] Desnoues N, Lin M, Guo X, et al. Nitrogen fixation genetics and regulation in a *Pseudomonas stutzeri* strain associated with rice [J]. Microbiology, 2003, 149: 2251-2262.
- [8] 林敏, 尤崇杓, Elmerich C. 固氮粪产碱菌 *nif*、*ntr* 和 *gln* 基因鉴定及 *nifA* 克隆[J]. 农业生物技术学报, 1995, 3: 47-51.
- [9] Yan Y, Yang J, Dou Y, et al. Nitrogen fixation island and rhizosphere competence traits in the genome of the root-associated *Pseudomonas stutzeri* A1501 [J]. Proc. Natl. Acad. Sci. USA. 2008, 105 (21): 7564-7569.
- [10] Xie Z, Dou Y, Ping S, et al. Interaction between NifL and NifA in the nitrogen fixing *Pseudomonas stutzeri* A1501 [J]. Microbiology, 2006, 152: 3535-3542.
- [11] He S, Chen M, Xie Z, et al. Involvement of GlnK, a P(II) protein, in control of nitrogen fixation and ammonia assimilation in *Pseudomonas stutzeri* A1501 [J]. Arch Microbiol., 2008, DOI 10.1007/s00203-008-0354-x.
- [12] Lin M, You C B. Denitrification and nitrogen fixation by *Alcaligenes faecalis* [J]. Acta Agriculturae Nucleatae Sinica, 1987, 1: 1-7.
- [13] Ewing B, Hillier L, Wendl M C, et al. Base-calling of automated sequencer traces using phred I. accuracy assessment [J]. Genome Res., 1998, 8: 175-185.
- [14] Ewing B, Green P. Base-calling of automated sequencer traces using phred II. error probabilities [J]. 1998, Genome Res., 8: 186-194.
- [15] Delcher A L, Bratke K A, Powers E C, et al. Identifying bacterial genes and endosymbiont DNA with Glimmer [J]. Bioinformatics, 2007, 23: 673-679.
- [16] Stover C K, Pham X Q, Erwin A L, et al. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen [J]. Nature, 2000, 406: 959-964.
- [17] Paulsen T, Press M, Ravel J, et al. Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5 [J]. Nat. Biotechnol., 2005, 23: 873-878.

- [18] Nelson K E, Weinel C, Paulsen T, *et al.*. Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440 [J]. *Environ. Microbiol.*, 2002, 4: 799 - 808.
- [19] Buell R C, Joardar V, Lindeberg M, *et al.*. The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000 [J]. *Proc. Natl. Acad. Sci. USA*, 2003, 100: 10181 - 10186.
- [20] Vodovar N, Vallenet D, Cruveiller S, *et al.*. Complete genome sequence of the entomopathogenic and metabolically versatile soil bacterium *Pseudomonas entomophila* [J]. *Nat. Biotechnol.*, 2005, 23: 873 - 878.
- [21] Yan Y, Yang J, Chen L, *et al.*. Structural and functional analysis of denitrification genes in *Pseudomonas stutzeri* A1501 [J]. *Science in China Ser. C Life Sciences*, 2005, 48 (6): 585 - 592.
- [22] 闻玉梅. 现代医学微生物学[M]. 上海:上海医科大学出版社, 1999, 383 - 393.
- [23] Pallen J M, Chaudhuri R, Henderson R I. Genomic analysis of secretion systems [J]. *Current Opinion in Microbiology*, 2003, 6: 519 - 527.
- [24] Sikorski J, Teschner N, Wackernagel W. Highly different levels of natural transformation are associated with genomic subgroups within a local population of *Pseudomonas stutzeri* from soil [J]. *Appl. Environ. Microbiol.*, 2002, 68: 865 - 873.
- [25] Lorenz M G, Sikorski J. The potential for intraspecific horizontal gene exchange by natural genetic transformation: sexual isolation among genomovars of *Pseudomonas stutzeri* [J]. *Microbiology*, 2000, 146: 3081 - 3090.
- [26] Liang A, Sha J, Lu W, *et al.*. A single residue mutation of 5-enoylpyruvylshikimate-3-phosphate synthase in *Pseudomonas stutzeri* enhances resistance to the herbicide lyphosate [J]. *Biotechnol. Lett.*, DOI 10.1007/s10529-008-9703-8, 2008.

【新书推介】



《生物信息学与功能基因组学》

(美) 乔纳森·佩夫斯纳 著 孙之荣 译 化学工业出版社

出版日期: 2006.7

I S B N: 7502583831

定 价: 95.00 元

开 本: 16 开

页 数: 706 页

本书内容全面,很好地将生物信息学与功能基因组学的理论和实践应用结合起来,非常重视生物信息学的具体应用技巧。

本书还具有如下特色:①作者权威。本书作者是霍普金斯医学院教授,在国际上有很高的声誉和权威性。②实践性强。每一章都有问题集、与生物信息学有关的 web 操作训练以及相应的

web 链接,还列出了可以免费获取的生物信息学软件和作者推荐的读物。③范例典型。本书从头到尾都用一个基因和它的蛋白产物作为例子进行讲解。这个蛋白是视黄醇结合蛋白(retinol-binding protein, RBP)。④图文并茂。本书在论述的同时配以大量的图片,直观、形象、通俗易懂。