



研究报告

苏云金芽胞杆菌 LM1212 菌株和突变株的全基因组序列比较分析

佟蕾¹ 张睿彬^{1,2} 彭琦¹ 张杰¹ Lereclus Didier² 宋福平^{*1}

1 中国农业科学院植物保护研究所 植物病虫害生物学国家重点实验室 北京 100193

2 INRA, UMR1319 Micalis, Jouy-en-Josas 78350, France

摘要:【背景】LM1212 菌株是昆虫病原菌苏云金芽胞杆菌(*Bacillus thuringiensis*, Bt)中的一员,其芽胞和晶体分别产生于芽胞形成细胞和晶体产生细胞中,具有独特的细胞分化表型。与野生株 LM1212 相比,突变株 LM1212-DB 芽胞细胞比例明显降低并产生更高比例的晶体产生细胞,这使得 LM1212-DB 菌株成为研究晶体产生细胞形成机制和提高菌株杀虫活性的绝佳实验材料。【目的】比较 LM1212 菌株和 LM1212-DB 菌株的基因组差异,以便于揭示导致这两个菌株表型差异的原因。【方法】利用单分子测序技术(single molecular real-time, SMRT)和 Pacbio RS II 测序平台对两个菌株进行全基因组测序,对染色体和质粒、双组分信号系统和插入序列等进行差异分析,并构建表型特性相关基因的系统发育树。【结果】基因组分析发现,LM1212 和 LM1212-DB 菌株均含有丰富的插入序列和双组分信号系统,暗示两个菌株极易发生基因重排且具有较强的环境适应性。与 LM1212 菌株相比,突变株 LM1212-DB 中发生了染色体和质粒片段缺失、质粒重排、质粒拷贝数变异。进一步分析缺失基因的功能发现,一些环境胁迫响应基因(如 *sigB*)和芽胞形成相关基因(如 *abrB*)等缺失;通过分析质粒拷贝数变异发现,具有增加晶体细胞比例功能的转录因子 CpcR 所在质粒的拷贝数增加 1 个,同时对 CpcR 的进化分析发现,与其亲缘关系最近的基因的从属菌株也产生与 LM1212 菌株相似的细胞分化表型。这些重要功能基因的缺失和拷贝数变异可能是导致两个菌株表型差异的原因。此外,突变株 LM1212-DB 缺失 I 型限制-修饰系统,这使得突变株 LM1212-DB 与野生菌株 LM1212 相比具有更好的外源 DNA 兼容性。【结论】突变株 LM1212-DB 染色体和质粒的结构变异可能是导致与野生株 LM1212 表型差异的潜在原因,这将为研究 LM1212 菌株的晶体细胞分化机制提供指导方向。

关键词: 苏云金芽胞杆菌, 基因组测序, 比较基因组, 表型差异

Foundation item: National Natural Science Foundation of China (31530095)

***Corresponding author:** Tel: 86-10-62815923; E-mail: fpsong@ippcaas.cn

Received: 31-03-2020; **Accepted:** 06-05-2020; **Published online:** 19-05-2020

基金项目: 国家自然科学基金(31530095)

***通信作者:** Tel: 010-62815923; E-mail: fpsong@ippcaas.cn

收稿日期: 2020-03-31; **接受日期:** 2020-05-06; **网络首发日期:** 2020-05-19

Comparative analysis of two complete genome sequences from *Bacillus thuringiensis* strain LM1212 and its mutant

TONG Lei¹ ZHANG Rui-Bin^{1,2} PENG Qi¹ ZHANG Jie¹ Lereclus Didier²
SONG Fu-Ping^{*1}

1 State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing 100193, China

2 INRA, UMR1319 Micalis, Jouy-en-Josas 78350, France

Abstract: **[Background]** The LM1212 strain, a member of the insect pathogen bacterium *Bacillus thuringiensis* (Bt), has a unique cell differentiation phenotype that spores and crystals are produced in spore-forming cells and crystal-producing cells, respectively. Compared to the wild-type LM1212, the mutant strain LM1212-DB formed a reduced proportion of spore-forming cells and a higher proportion of crystal-producing cells, which provided an excellent experimental material for studying the mechanism of crystal-producing cell formation. **[Objective]** In this study, we tried to reveal the potential reason for the phenotypic difference between these two strains by investigating their genomic differences. **[Methods]** The whole genomes of the two strains were sequenced by using the single molecular real-time (SMRT) technology on the Pacific Bioscience (Pacbio) RS II sequencing platform. The differences in chromosome, plasmid, two-component system and insertion sequence between two strains were analyzed, and a phylogenetic tree based on a gene related to phenotypic characteristics was constructed. **[Results]** Genomic analysis showed that both strains contained abundant insertion sequences and two-component systems, suggesting that two strains were prone to rearrange genes for environmental adaptation. Fragment deletion within chromosome and plasmid, rearrangement and copy number variation for plasmid occurred in the mutant LM1212-DB. Some environmental stress response genes such as *sigB* and sporulation-related genes such as *abrB* were found to be absent and one copy number of plasmid carrying the transcription factor CpcR increased in the genome of LM1212-DB. The evolutionary analysis of CpcR indicated that the strain carrying the *cpcR* homologous gene also had a cell differentiation phenotype similar to that of LM1212. The deletion of these important functional genes and the decreasing of the copy number may be responsible for the phenotypic differences between the two strains. In addition, LM1212-DB lacked type I restriction-modification system, which might endow it a better exogenous DNA compatibility, compared to the wild-type strain LM1212. **[Conclusion]** Structural variation of chromosomes and plasmids lead to phenotypic differences between LM1212 and LM1212-DB, which would provide a clue for the study on differentiation mechanism of LM1212 cells.

Keywords: *Bacillus thuringiensis*, Genome sequencing, Comparative genome, Phenotypic differences

苏云金芽胞杆菌(*Bacillus thuringiensis*, Bt)是一种具有多种生态位的革兰氏阳性细菌,其最显著的特征是在产芽胞的过程中会产生由杀虫晶体蛋白(Cry蛋白或Cyt蛋白)组成的伴胞晶体^[1]。由于杀虫晶体蛋白具有特异性杀虫活性,目前Bt已成为应用最成功的微生物杀虫剂^[2]。苏云金芽胞杆菌染色体为环状DNA,基因组大小为5.0 Mb左右,(G+C)mol%含量一般为32%–35%。Bt菌株包含至少一个质粒DNA,大多数的杀虫晶体蛋白基因定

位于质粒DNA上。迄今为止,在NCBI网站上已公布了594株Bt菌株的全基因组序列,这使得高通量测序结合生物信息学分析的方法在Bt基因组数据中深度挖掘各类型基因变得切实可行,如杀虫晶体蛋白基因、营养期杀虫蛋白基因^[3]、分泌性杀虫蛋白基因^[4]、对癌细胞有特异性活性的基因、苏云金素合成基因簇、双效菌素合成基因簇等。因此,Bt菌株的应用范围也被不断拓宽^[2]。

苏云金芽胞杆菌的普遍存在性和高度适应能

力可能是由于其具有响应复杂环境变化的生物功能元件。例如,双组分信号系统可通过磷酸基团的级联反应对环境的波动,比如营养缺乏、氧限制及冷/热刺激等,做出快速而有力的反应^[5-6],并调控细菌的多种生理行为,包括运动性、趋化性、芽胞形成过程、生物膜形成以及毒素合成等^[7-9]。再如,插入序列单元具有高移动性和插入能力,可通过剪切和插入导致基因的突变或移动,从而影响基因的表达。插入序列元件多与信号转导基因或转运蛋白基因相邻,这表明插入序列的转座倾向于影响与环境适应性相关的功能基因的表达^[10]。

典型的 Bt 菌株的芽胞和晶体产生于同一细胞内,其杀虫晶体基因的表达由芽胞依赖型启动子介导^[11]。与此形成鲜明对比的是,LM1212 菌株的杀虫晶体基因仅在非芽胞细胞中表达且不依赖于芽胞的形成;此外,LM1212 菌株具有独特的细胞分化表型,其芽胞和晶体分别产生于芽胞形成细胞和晶体产生细胞中,其芽胞形成细胞和晶体产生细胞的比例约为 6:4^[12]。前期研究鉴定到 LM1212 菌株中存在一种新型的转录因子 CpcR,目前发现其具有增加晶体细胞比例的功能^[13],这表明 CpcR 是 LM1212 菌株细胞分化机制的影响因素之一。为了进一步阐明 LM1212 菌株细胞分化的机制,马丽霞等^[14]通过高温处理的方法进行质粒消除实验,得到了一株表型稳定的突变菌株 LM1212-DB,该菌株具有更少的芽胞形成细胞和更多的晶体产生细胞。这种能使晶体产生细胞比例显著增加的突变株既可作为研究细胞分化机制的实验材料,又可作为提高 Bt 菌株杀虫晶体蛋白产量的基础菌株。实验室前期研究表明,与 LM1212 相比,LM1212-DB 突变株的质粒大小差异显著且鉴定到位于质粒上的几个毒素基因已缺失,这意味着二者的基因组结构出现变异,为进一步解析二者之间的差异,对 LM1212 与 LM1212-DB 菌株进行了全基因组测序^[14]。本研究比较分析了芽胞形成细胞和晶体产生细胞分化比

例具有明显差异的 LM1212 和 LM1212-DB 菌株的全基因组序列,重点分析了两个菌株的全基因组、双组分信号系统、插入序列等方面的差异,探求导致两个菌株表型产生差异的原因。本研究结果以期与研究 LM1212 菌株细胞分化机制提供理论依据,进而为 Bt 制剂的定向遗传改良奠定基础。

1 材料与方法

1.1 供试菌株

LM1212 菌株和突变株 LM1212-DB 均为实验室保存。

1.2 主要试剂和仪器

蛋白胨、氯化钠、酵母提取物, Sigma 公司; RNase A, 北京全式金生物技术有限公司。低温离心机, Eppendorf 公司; pH 计, Thermo Scientific 公司。

1.3 培养基

LB 液体培养基(g/L): NaCl 10.0, Tryptone 10.0, Yeast extract 5.0, pH 7.0。

1.4 基因组制备与测序

将 LM1212 菌株和突变株 LM1212-DB 以 30 °C、200 r/min 的培养条件培养过夜后,按照 Stobdan 等^[15]描述的方法提取并检测基因组 DNA,将其送至北京赛墨百合生物科技有限公司进行测序。利用 SMRT 技术和 Pacbio RS II 测序平台进行测序^[16-17]。

1.5 测序数据的组装及注释

由于单分子测序技术具有高读长、数据信号捕捉困难且随机性强等特征,因此,在得到原始数据后需要进行严格的数据质量过滤和前处理。利用层次基因组装软件包 HGAP 对过滤后的高质量 Clean reads 进行组装。利用 Prodigal^[18]、Infernal^[19]和 RNAmmer^[20]等软件包对编码序列(coding sequence, CDS)、tRNA 和 rRNA 进行预测。使用 Rfam^[21]、Swissprot^[22]、COG^[23]和 KEGG^[24]数据库对全基因组序列进行基因的功能预测与注释。

1.6 插入序列鉴定及比较基因组分析

使用在线网站 ISFINDER^[25] (<https://www-is.biotoul.fr/>)对 LM1212 和 LM1212-DB 全基因组序列进行插入序列的鉴定与分析。利用 MUMmer 3.0^[26]软件进行比较基因组分析。

1.7 CpcR 同源序列系统发育树构建

使用在线网站 NCBI-BLASTp (<https://blast.ncbi.nlm.nih.gov/Blast>)进行 CpcR 同源序列比对。利用 MEGA 7.0 软件构建系统发育树。

2 结果与分析

2.1 LM1212 菌株和突变株 LM1212-DB 的全基因组概况

LM1212 菌株和突变株 LM1212-DB 的全基因组序列已上传至 GenBank, 登录号分别为 CP024771-CP024779^[13]和 CP051013-CP051018。根据基因组注释结果显示, LM1212 菌株和 LM1212-DB 菌株的质粒数量、质粒拷贝数以及包含的杀虫基因种类均存在差异。LM1212 菌株包含

1 个环状的染色体、8 个环状的质粒和 14 种杀虫基因, LM1212-DB 菌株包含 1 个环状的染色体、5 个环状的质粒和 10 种杀虫基因(表 1)。

2.2 LM1212 菌株和突变株 LM1212-DB 的染色体结构变异及功能分析

对 LM1212 菌株和突变株 LM1212-DB 的染色体基因组进行比较分析发现, 位于 LM1212 菌株染色体基因组的 2 607 455 bp 与 2 682 965 bp 之间的片段在 LM1212-DB 菌株中缺失。染色体基因组的缺失片段共含有 69 个开放阅读框(open reading frame, ORF): 4 个 ORF 被注释为转座酶, 值得注意的是, 位于缺失片段的首末均为同一个转座酶, 这可能是导致该片段缺失的原因; 38 个 ORFs 被注释为未知功能的假想蛋白基因; 31 个 ORFs 被注释为各类别的功能基因, 可将这些缺失的功能基因分为 2 类: 一类基因是在基因组未缺失区域仍存在与此基因功能相似或相同的基因; 另一类基因是在 LM1212-DB 菌株中仅存在唯一的拷贝且丢失(表 2)。

表 1 LM1212 和 LM1212-DB 菌株基因组概况

Table 1 Genome landscapes of LM1212 and LM1212-DB strain

菌株	染色体/质粒名称	碱基数	拷贝数	杀虫基因类别
Strains	Chromosome/Plasmid name	Sequence size (bp)	Copy number	Insecticidal genes types
LM1212	Chromosome	5 705 934	1	—
	pLM248	248 002	3	<i>cry32Wa1</i> , <i>cry32Va1</i> , <i>cry35-like</i> , <i>cry35-like2</i> , <i>cry41Ca1</i> , <i>cry69Aa-like</i> , <i>cryBP1</i> , <i>cytAa-like</i> , <i>cytCa-like</i>
	pLM158	157 839	3	<i>cry45ba1</i> , <i>cry74Aa1</i> , <i>nt32-like</i>
	pLM113	112 709	2	<i>cry32-like</i> , <i>mtx2-like</i>
	pLM35	35 121	3	—
	pLM28	28 658	2	—
	pLM14	14 193	10	—
	pLM13	13 118	6	—
	pLM5	4 975	8	—
LM1212-DB	Chromosome	5 629 931	1	—
	pLM241	241 137	4	<i>cry41Ca1</i> , <i>cry45ba1</i> , <i>cry74Aa1</i> , <i>cry69Aa-like</i> , <i>cryBP1</i> , <i>cytAa-like</i> , <i>cytCa-like</i> , <i>nt32-like</i>
	pLM113	112 715	2	<i>cry32-like</i> , <i>mtx2-like</i>
	pLM35	35 175	4	—
	pLM14	14 375	3	—
	pLM14-1	14 195	2	—

注: —: 无杀虫基因

Note: —: No insecticidal gene

表 2 LM1212-DB 菌株染色体基因组中缺失基因列表

Table 2 List of the absent genes in the chromosome of LM1212-DB strain

名称 Name	蛋白长度 Protein length (aa)	功能 Function	COG 分类 COG category
LM_02887	325	鸟苷酸环脱氢酶 Ornithine cyclodeaminase	E
LM_02876	471	藻酸盐乙酰转移酶 Alginate acetyltransferase	M
LM_02926	506	I 型限制-修饰系统酶 M Type I restriction-modification system M subunit	V
LM_02927	424	I 型限制-修饰系统酶 S Type I restriction-modification system S subunit	V
LM_02928	1 031	I 型限制-修饰系统酶 R Type I restriction-modification system R subunit	V
LM_02909	117	XRE 家族转录因子 XRE family transcription factors	K
LM_02916	225	双组分信号系统转录因子 Two-component signaling system transcription factors	T
LM_02906	197	Sigma B	K

针对第二类在 LM1212-DB 菌株中完全缺失的基因进行功能分析,未发现与细胞分化过程相关的基因,但发现缺失基因的功能多注释为环境适应性和防御机制相关的基因。按照基因组注释的功能,可将完全缺失的基因分为 4 类:(1) 参与响应外界环境胁迫的转录因子,如 SigmaB (LM_02906)、XRE 家族转录因子(LM_02909)和未知功能的双组分信号系统的转录因子(LM_02916);(2) 与脯氨酸合成途径密切相关的关键酶鸟苷酸环脱氢酶(LM_02887)^[27];(3) 与细胞壁合成相关的藻酸盐乙酰转移酶(LM_02876)^[28];(4) 参与细菌防御机制的 I 型限制-修饰系统酶 M、S 和 R (LM_02926、LM_02927 和 LM_02928)^[29]。

综上所述,突变株 LM1212-DB 的染色体基因组上共缺失了 38 个未知功能的假想蛋白,其中可能存在着与细胞分化过程相关的基因,但具体的生物学功能有待进一步验证。此外,突变株缺失了一些与响应环境胁迫相关的基因,如可使芽胞杆菌更好地应对环境压力、能量损耗或低温的转录因子 Sigma B^[30],以及可以通过参与合成脯氨酸来影响细菌渗透调节的鸟苷酸环脱氢酶,由此可以看出,突变株对极端环境的适应能力可能变

弱。突变株还缺失了 1 个可以对藻酸盐进行修饰的藻酸盐乙酰转移酶基因,但在突变株中并未发现藻酸盐合成的基因簇。I 型限制-修饰系统的缺失则会导致突变株对外源 DNA 具有更好的兼容性,因此该突变株更适用于进行分子遗传操作。

2.3 LM1212 菌株和突变株 LM1212-DB 的质粒差异分析

利用 MUMmer 3.0 软件对 LM1212-DB 菌株的 5 个质粒与 LM1212 菌株的 8 个质粒进行序列比对发现,LM1212-DB 菌株的质粒发生重排以及部分质粒或片段发生缺失:(1) 质粒重排及片段缺失:位于 LM1212 菌株 pLM248 质粒上的 108 222 bp 片段与位于 pLM158 质粒上的两段序列长度分别为 40 681 bp 和 95 786 bp 的片段发生重排,在 LM1212-DB 菌株中形成了一个新的质粒 pLM241 (图 1,标注绿色的片段),同时 LM1212 菌株 pLM248 和 pLM158 质粒上未参与重排的片段在 LM1212-DB 菌株中缺失(图 1,未标注颜色的片段);(2) 质粒缺失:LM1212 菌株中的 pLM13 和 pLM5 在 LM1212-DB 菌株中缺失,LM1212 菌株中的 pLM113、pLM35 和 pLM14 在 LM1212-DB 菌株中未发生重排和缺失,LM1212 菌株中的 pLM14 质粒在 LM1212-DB 菌株中命名为 pLM14-1。

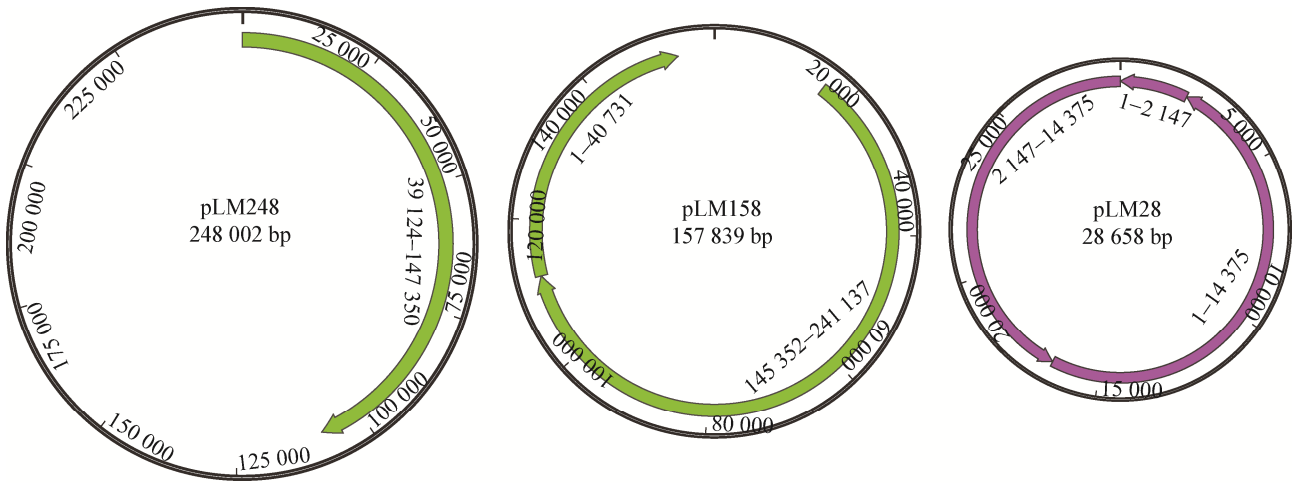


图 1 质粒重排示意图

Figure 1 Diagram of plasmid rearrangement

注: 最内侧的数字表示在 LM1212-DB 菌株中基因的位置。

Note: The innermost number indicates the gene location in LM1212-DB strain.

除了发生质粒重排、质粒片段缺失和质粒缺失的情况, 还发生了质粒拷贝数的改变: LM1212 菌株中的 pLM248 质粒和 pLM158 质粒均为 3 个拷贝, 而在 LM1212-DB 菌株中的 pLM241 质粒为 4 个拷贝(表 1), 这相当于位于 pLM248 质粒和 pLM158 质粒上的未缺失基因在 LM1212-DB 菌株中均增加了 1 个拷贝; LM1212 菌株中的 pLM35 质粒为 3 个拷贝, 而在 LM1212-DB 菌株中 pLM35 质粒增加了 1 个拷贝, 为 4 个拷贝(表 1); LM1212-DB 菌株中的 pLM14 质粒源于 LM1212 菌株中 pLM28 质粒的一半(图 1, 标注粉色的片段), 在 LM1212-DB 菌株中的 pLM14 质粒为 3 个拷贝(表 1), 比 LM1212 菌株中 pLM28 质粒增加了 1 个拷贝; LM1212 菌株 pLM14 质粒为 10 个拷贝, 而在 LM1212-DB 菌株中的 pLM14-1 质粒为 2 个拷贝, 相当于减少了 8 个拷贝。

2.4 LM1212-DB 菌株质粒缺失片段的生物学功能分析

在 LM1212-DB 菌株中缺失的质粒片段共包含 184 个 ORF, 包括位于 pLM248 质粒上的 *orf118-orf269*、位于 pLM158 质粒上的 *orf1-orf15* 和 *orf172-orf177*、位于 pLM13 质粒上的 *orf1-9*、位于 pLM5 质粒上的 *orf1-5*。通过 COG 功能富集分析发现, 78 个基因编码的蛋白与复制、重组和修复的功能有关, 其中大多数基因被注释为各类

型的转座酶或质粒重组酶; 8 个基因编码的蛋白与细胞壁或细胞膜的形成有关, 其中包括 3 个细胞壁相关蛋白、2 个 Rhs 家族蛋白、1 个可切割肽聚糖侧链的 N-乙酰胞壁酰-L-丙氨酸酰胺酶、1 个糖基转移酶和 1 个与离子转运系统相关的蛋白; 6 个基因编码的蛋白与信号转导机制有关; 4 个基因编码的蛋白与防御机制有关; 3 个基因编码了与芽胞相关的蛋白, 其中包括 1 个芽胞衣相关蛋白和 2 个芽胞萌发蛋白(图 2); 在 LM1212-DB 菌株的缺失片段中包含了 5 个杀虫晶体蛋白, 分别为 Cry35-like (pLM248_00251)、Cry35-like2 (pLM248_00249)、Cry69Aa-like (pLM248_00196)、Cry32Wal (pLM248_00197)和Cry32Val (pLM248_00227)。

综上所述, 与基因组的缺失片段相比, 质粒的缺失片段含有更多的未知功能基因, 对于这一部分基因功能的研究势必推动 LM1212 菌株细胞分化机制的研究。LM1212 菌株的质粒上存在着丰富的转座酶和质粒重组酶, 这表明 LM1212 菌株的质粒极易发生重排。这是 LM1212 菌株的一个重要特性, 因此在缺失片段中也包含着大量的转座酶和质粒重组酶。此外, 缺失片段中包含 2 个芽胞衣相关蛋白和 1 个芽胞萌发蛋白, 芽胞衣相关蛋白在芽胞形成过程的晚期发挥作用, 芽胞萌发蛋白在芽胞萌发初期发挥作用, 因此二者并不会影响芽胞

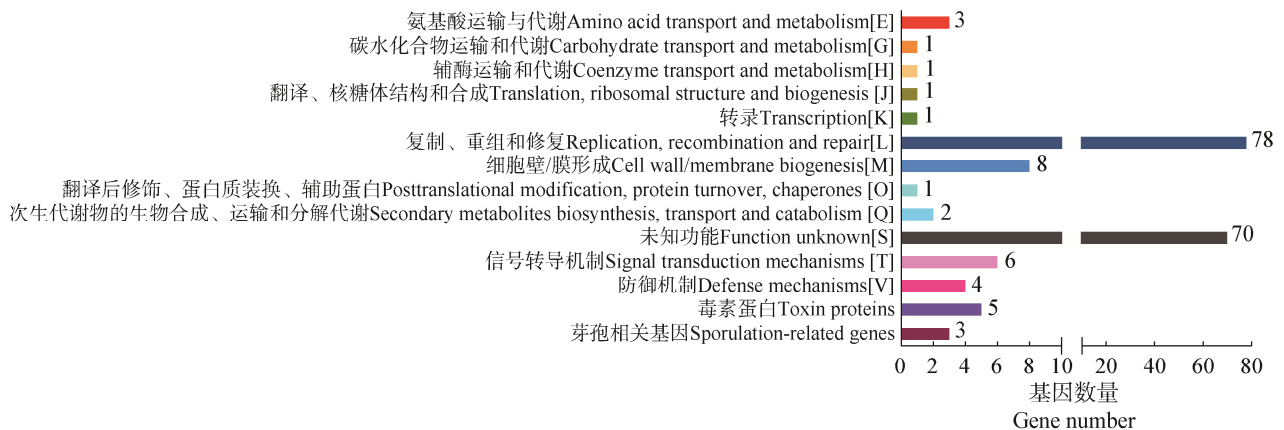


图 2 缺失基因的 COG 功能注释

Figure 2 COG functional annotation of the absent genes

形成过程。值得注意的是，缺失片段中还包含了一个与芽胞形成相关的转录因子 AbrB (pLM248_00154)。在枯草芽胞杆菌中 AbrB 可通过抑制 *sigH* 表达间接调控 Spo0A 的磷酸化水平，从而影响芽胞形成过程^[31]。因此，转录因子 AbrB 的缺失可能是导致两个菌株产生表型差异的重要原因。

2.5 LM1212-DB 菌株中拷贝数增加的转录因子 CpcR 的分布情况及系统发育树

在 LM1212 菌株中，*cpcR* 在 pLM248 质粒上为 3 个拷贝。在 LM1212-DB 菌株中，*cpcR* 在 pLM241 质粒上为 4 个拷贝。拷贝数增加意味着 CpcR 的表达量增加，从而导致 LM1212-DB 菌株增加晶体细胞比例的效果更为明显，这暗示 CpcR 很可能是导致 LM1212-DB 菌株和 LM1212 菌株表型差异的重要影响因素之一。为了进一步探究 *cpcR* 是否是导致这种表型差异的特异性基因，以 CpcR 的氨基酸序列为参考序列，在 NCBI 数据库中进行序列比对，以氨基酸序列相似性大于 50% 为标准，共检索到 88 个同源基因，这些同源基因分属于 10 个物种，且均被注释为细菌双组分信号系统中 OmpR 家族转录因子，均含有信号接收结构域和 DNA 结合结构域。进一步分析发现，这 10 个物种大多属于厚壁菌门，其中多为蜡样芽胞杆菌族的细菌，如苏云金芽胞杆菌、蜡样芽胞杆菌 (*Bacillus cereus*)、细胞毒素芽胞杆菌 (*Bacillus cytotoxicus*) 等；此外，类芽胞杆菌属的 *Paenibacillus contaminans*、葡萄

球菌属的 *Thermoflavimicrobium dichotomicum* 以及梭菌属的 *Clostridium cylindrosporum* 等细菌中也鉴定到 *cpcR* 的同源基因。但是，虽然 *cpcR* 的同源基因在苏云金芽胞杆菌的近缘物种(一些产胞细菌)中均有分布，但这并不意味着 *cpcR* 在产胞细菌中保守，相反，*cpcR* 基因的分布与物种的相关性很差。这一结论可通过包含 *cpcR* 同源基因的菌株占该物种中已完成基因组测序的比例即可看出：10 个物种中包含 *cpcR* 同源基因的菌株仅占 88 株，而这些菌株所在物种中完成基因组测序的菌株高达 2 662 株 (表 3)，这说明 *cpcR* 不是芽胞杆菌属菌株的核心基因，其分布很可能与环境或生态位特异性有关。

CpcR 的保守性与物种相关性差也在其系统发育树中体现出来。88 个菌株中 CpcR 的同源序列存在物种内多个菌株序列 100% 一致的情况，因此在保证序列的物种信息不损失的前提下，选取可代表全部同源序列特征的 21 个序列进行系统发育树构建，建树结果显示，根据序列的亲缘关系，CpcR 的同源序列被聚类成 6 个分支，每个分支中的序列不具备物种特异性，例如苏云金芽胞杆菌中的一些亚种在红色、黑色、蓝色分支中均有出现。此外，在苏云金芽胞杆菌 *yunnanensis* 亚种 *Bacillus thuringiensis* BGSC4AM1 中鉴定到与 CpcR 氨基酸序列相似性高达 85% 的序列 WP_086410509.1，且有文献记载该菌株具有与 LM1212 菌株相似的表型^[32-33]，这也侧面印证了 *cpcR* 可能是引起两个菌株表型差异的特异性基因 (图 3)。

表 3 CpcR 同源序列分布情况

Table 3 The distribution of CpcR homologous sequence

菌株名称 Strain name	NCBI 已测序菌株数目 Number of sequencing strains in NCBI	包含 CpcR 同源序列菌株数目 Number of strains containing CpcR homologous genes
<i>Bacillus thuringiensis</i>	594	11
<i>Bacillus cereus</i>	1 113	62
<i>Bacillus toyonensis</i>	216	1
<i>Bacillus cytotoxicus</i>	14	1
<i>Bacillus wiedmannii</i>	158	3
<i>Bacillus manliponensis</i>	1	1
<i>Bacillus sp.</i>	563	6
<i>Thermoﬂavimicrobium dichotomicum</i>	1	1
<i>Paenibacillus contaminans</i>	1	1
<i>Clostridium cylindrosporium</i>	1	1
总计 Total	2 662	88

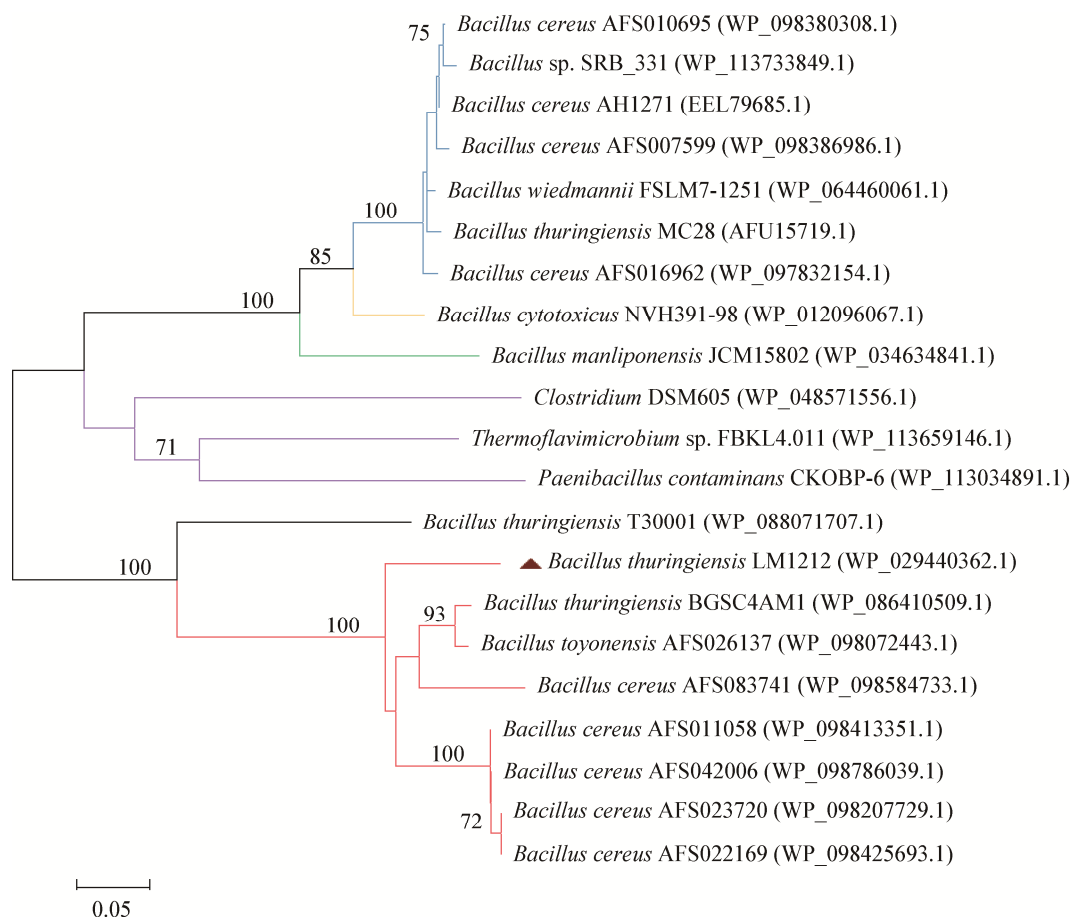


图 3 基于 *cpcR* 同源序列的系统发育树

Figure 3 Phylogenetic tree of *cpcR* homologous sequences

注: 括号里的序号为菌株 GenBank 登录号; 分支上的数字为 Bootstrap 值; 比例尺长度表示序列的差异度。

Note: The numbers in parentheses is the strain accession number in GenBank; The robustness of branching is indicated by bootstrap percentages calculated for 1 000 subsets; Bar indicates 5% estimated sequence divergence.

2.6 LM1212 菌株和突变株 LM1212-DB 双组分信号系统分析

典型的双组分信号系统包含一个具有保守激酶核心区域的组氨酸激酶(histidine kinase, HK)和一个具有保守调控区域的应答调控蛋白(response regulator, RR)。以 HK-RR 基因对为筛选标准对全基因组数据进行挖掘和筛选, 结果发现: LM1212 菌株共编码了 36 个 HK-RR 基因对, 其中包含 19 个可确定功能的 HK-RR 基因对, 其具体的功能详见表 4; 其余的 HK-RR 基因对仍需要进一步的实验确定其生物学功能。与 LM1212 菌株不同的是, 突变株 LM1212-DB 中缺失了 1 个 RR 基因(表 2), 且该 RR 基因为未知功能的孤儿基因, 并未找到与其配对的 HK 基因。

2.7 LM1212 菌株和突变株 LM1212-DB 的插入序列分析

插入序列单元是普遍存在于细菌染色体和质

粒中并且具有插入能力的小分子 DNA 片段。在确保完整性的前提下, LM1212 菌株染色体和质粒上分布着属于 9 个家族的 56 个完整的插入序列单元, 其中 24 个位于染色体上, 17 个位于 pLM248 质粒上, 7 个位于 pLM158 质粒上, 7 个位于 pLM113 质粒上(表 5)。YBT-1520 菌株在染色体上仅含有 4 个完整的 *ISBth13* 拷贝, 而 LM1212 菌株的染色体上含有 19 个完整的 *ISBth13* 拷贝。*IS240A* 发现于苏云金芽胞杆菌以色列亚种^[34], LM1212 菌株 pLM248 质粒上含有 13 个完整的 *IS240A* 拷贝。LM1212 菌株蕴含丰富的插入序列单元可能是导致原核微生物基因组重排的重要因素。

与 LM1212 菌株相比, 在 LM1212-DB 菌株中共缺失了 6 个插入序列单元, 其中包括 2 个 *ISBth13*、3 个 *IS240A* 和 1 个 *IS231N*。

表 4 LM1212 菌株基因组中的双组分信号系统

Table 4 Two-component system in the genome of LM1212 strain

基因位置 Gene locus	蛋白名称 Protein name	功能 Function
LM_00819/LM_00820	PhoR/PhoP	磷酸盐代谢 Phosphate metabolism
LM_03600/LM_03599	ResE/ResD	呼吸作用 Respiration
LM_01788/LM_01789	VicK/VicR	维持细胞壁稳态 Maintain cell wall homeostasis
LM_00673/LM_00674	GraS/GraR	响应外界压力 Response to external stress
LM_01114/LM_01113		细胞壁代谢 Cell wall metabolism
LM_04865/LM_04866		
LM_06077/LM_06078		
LM_01380/LM_01379	VanS/VanR	万古霉素抗性 Vancomycin resistance
LM_02513/LM_02514	CitS/CitT	柠檬酸转运 Citric acid transport
LM_02534/LM_02535		
LM_01764/LM_01763	LytS/LytT	丙酮酸代谢 Pyruvate metabolism
LM_04434/LM_04433	AgrC/AgrA	毒力基因表达 Virulence gene expression
LM_01458/LM_01459	ComP/ComA	细胞感受态 Cellular competence
LM_01675/LM_01674	DesK/DesR	细胞膜流动性 Cell membrane fluidization
LM_03561/LM_03562	LiaS/LiaR	感应抗生素压力 Response to antibiotic stress
LM_04277/LM_04276	SalK/SalR	分泌系统 Secretion system
LM_03762/LM_03761	CheA/ CheY	细菌趋化性系统 Bacterial chemotactic system
LM_03883/LM_03884	YcbA/YcbB	谷氨酰胺的利用 Utilization of glutamine
LM_04326/LM_04327		

表 5 LM1212 菌株基因组中的插入序列基因信息

Table 5 Insertion sequences in the genome of LM1212 strain

插入序列 Insertion sequence	家族 Family	插入序列长度 Insertion sequence length (bp)	染色体 Chromosome	Plasmid pLM248	Plasmid pLM158	Plasmid pLM113
IS <i>Bth166</i>	IS110	1 729	3	4	0	1
IS <i>Bth13</i>	IS110	1 588	19	0	1	1
IS <i>Bth15</i>	IS200/IS605	1 490	1	0	0	0
IS <i>Bth16</i>	IS200/IS605	1 821	1	0	1	0
IS231 <i>N</i>	IS4	1 654	0	0	1	1
IS231 <i>Y</i>	IS4	1 645	0	0	0	1
IS <i>Bce9</i>	IS4	1 597	0	0	0	1
IS240 <i>A</i>	IS6	861	0	13	0	0
IS240 <i>B</i>	IS6	861	0	0	4	2
总计 Total			24	17	7	7

3 讨论与结论

LM1212 菌株是一株具有独特表型的苏云金芽胞杆菌, 可分化成芽胞形成细胞和晶体产生细胞, 比例约为 6:4^[12]。这样的特性在 *Bt* 菌株 *yunnanensis* 亚种中也曾被发现, 但产生这种现象的分子机制尚未阐明^[32,35]。与野生株 LM1212 相比, 突变株 LM1212-DB 晶体产生细胞的比例更高, 芽胞形成细胞比例更低, 为 LM1212 菌株表型的分子机制解析奠定了重要的基础。此外, 与 LM1212 菌株相比, LM1212-DB 菌株中晶体产生细胞比例的增加可大大提高 *Bt* 杀虫剂的产量和纯度, 从而提高 *Bt* 制剂的防治效果。本研究从 LM1212 和 LM1212-DB 菌株基因组入手, 结合比较基因组学和生物信息学分析的手段, 探究二者细胞表型变化的原因。从染色体和质粒变异、双组分信号系统、插入序列等方面对其表型差异原因进行解析。

本研究发现, 与野生株 LM1212 相比, LM1212-DB 菌株具有染色体片段缺失、质粒缺失、质粒上部分片段缺失、质粒间重组及质粒拷贝数变化等生物学现象。基于全基因组数据的分析发现, 这两个菌株具有丰富的插入序列和重组酶。研究发现, 插入序列具有多种转座机制可插

入到基因编码区导致基因缺失或倒置, 也可插入到基因的启动子区抑制下游基因的表达^[3]。重组酶则可以把重组分子重新拼接, 导致基因重排及质粒重组等现象。因此, 两个菌株中丰富的插入序列和重组酶很可能是导致两个菌株基因组结构变异的原因。伴随着基因组结构的变异, 很多影响菌株分化表型的重要基因也会随之缺失或拷贝数增加, 从而导致其功能受到影响, 进而对菌株的表型产生影响。

基于基因组比较分析的结果, 我们推测有 5 个方面的因素导致了 LM1212-DB 和 LM1212 菌株表型的差异: (1) LM1212-DB 菌株中缺少了一些参与环境胁迫响应相关的基因, 如响应环境胁迫的转录因子 Sigma B^[30,36]以及 XRE 家族转录因子等。这些环境胁迫响应基因的缺失, 可能导致启动芽胞形成的信号转导途径被部分阻断, 从而大大降低了芽胞形成细胞的比例。事实上, 芽胞形成过程是 *Bt* 菌应对恶劣环境的一种机制, 而环境胁迫响应基因的缺失很可能导致芽胞形成的决策途径被阻断, 从而影响芽胞形成起始。例如在枯草芽胞杆菌中已经证实, 芽胞形成起始的开关是 Spo0A 的磷酸化, 而 Spo0A 的磷酸化需要来自 Spo0E 传递的磷酸信号^[37]。Spo0E 的启动子中有一个是

Sigma B 依赖型的, 这种由 Sigma B 激活 Spo0E 启动子活性的方式是启动 Spo0A 磷酸化的重要途径^[38-39]。这意味着, LM1212-DB 菌株中 Sigma B 的缺失也可能导致芽胞形成分子开关无法正常打开, 从而抑制了 LM1212-DB 菌株的芽胞形成。(2) 与芽胞形成相关的重要转录因子 AbrB 等基因的缺失, 可能与表型差异密切相关, 但仍需进一步的实验进行验证。(3) LM1212-DB 菌株质粒拷贝数的变化, 导致某些关键基因的拷贝数也发生了变化, 如在 LM1212-DB 菌株中晶体形成细胞转录因子 CpcR 增加了 1 个拷贝, 从而导致 CpcR 的表达量升高, 增加晶体细胞比例的作用更为明显, 导致菌株表型发生变化。(4) 缺失片段含有大量未知功能的基因, 其中某些基因可能参与或调控了芽胞形成过程, 从而导致表型差异。(5) 由于 LM1212-DB 菌株中缺失了许多重要的调控因子, 使 LM1212-DB 菌株中与表型相关的调控模块网络瘫痪, 进而使受其抑制的基因在突变体中表达量上调, 从而间接导致表型差异。

本研究还对 LM1212 和 LM1212-DB 菌株的特异性基因进行挖掘, 也让我们深入了解了这两个菌株的特性, 对其应用具有指导意义。基于全基因组数据的分析发现, 这两个菌株具有丰富的双组分信号系统以及插入序列, 暗示该菌株可响应复杂的环境变化。本研究筛选得到的 36 对双组分信号系统, 如在枯草芽胞杆菌中已经解析的调控细胞分化的转录因子 DegU、ComA 和 Spo0A 等^[38-39], 这表明在 LM1212 菌株中也存在着复杂的细胞分化调控网络。同时, 发现基因组中也存在 HK 和 RR 的孤儿基因, 因此这两个菌株可能存在一个 HK 可磷酸化多个 RR 或一个 RR 可被多个 HK 磷酸化的复杂磷酸传递网络, 各双组分信号系统间可能也存在着信号交谈(cross talk)的现象。此外, 对转录因子 CpcR 进行进化分析时发现, 该基因属于双组分信号系统 OmpR 家族转录因子, 这意味着在 LM1212 和 LM1212-DB 菌株中还存在着一个

与 CpcR 互作的 HK, 而且该基因不是产胞细菌的核心基因, 应该是伴随环境进化过程中的水平转移基因; 与此同时, 与 LM1212 表型相近的 Bt 菌株 *yunnanensis* 亚种中存在着与 CpcR 氨基酸序列相似性高达 85% 的同源基因, 这意味着 CpcR 和其互作的激酶可能是导致 LM1212 菌株细胞分化表型的重要原因之一, 而对 CpcR 互作激酶的寻找将是有利于 LM1212 细胞分化机制解析的重要方向。此外, 对 LM1212 菌株的插入序列单元的分布进行分析发现, 在 LM1212 菌株中, 与插入序列相邻的基因多为杀虫晶体蛋白基因, 这表明插入序列和杀虫晶体蛋白基因可能存在着共同进化的关系。除了杀虫晶体蛋白基因外, 部分相邻的基因被注释为假想蛋白, 进一步解析这些假想蛋白的功能将有助于阐明插入序列与进化之间的关系。LM1212-DB 菌株的染色体上缺失了 I 型限制-修饰系统的操纵子, 这使得 LM1212-DB 菌株与野生菌株相比具有更好的外源 DNA 兼容性。

通过对苏云金芽胞杆菌 LM1212 和 LM1212-DB 菌株进行基因组分析发现, 两个菌株表型差异的主要原因可能是由于染色体和质粒变异引发了菌株核心基因及特异性基因之间的系统平衡被打破, 进而导致菌株细胞分化的表型趋向于一个极端。上述结果仍需开展进一步实验研究, 有利于明确 Bt 细胞分化的分子机制以及 Bt 菌株资源和基因挖掘的研究方向。本研究通过对两个菌株的基因组进行比较分析, 从基因组层面解析了两个菌株表型差异的原因, 为细菌细胞分化的机制研究奠定了基础, 也对 Bt 制剂的遗传改良具有重要的指导意义。

REFERENCES

- [1] Bravo A, Likitvivanavong S, Gill SS, et al. *Bacillus thuringiensis*: a story of a successful bioinsecticide[J]. *Insect Biochemistry and Molecular Biology*, 2011, 41(7): 423-431
- [2] Jouzani GS, Valijanian E, Sharafi R. *Bacillus thuringiensis*: a successful insecticide with new environmental features and tidings[J]. *Applied Microbiology and Biotechnology*,

- 2017, 101(7): 2691-2711
- [3] Estruch JJ, Warren GW, Mullins MA, et al. Vip3A, a novel *Bacillus thuringiensis* vegetative insecticidal protein with a wide spectrum of activities against lepidopteran insects[J]. Proceedings of the National Academy of Sciences of the United States of America, 1996, 93(11): 5389-5394
- [4] Donovan WP, Engleman JT, Donovan JC, et al. Discovery and characterization of Sip1A: a novel secreted protein from *Bacillus thuringiensis* with activity against coleopteran larvae[J]. Applied Microbiology and Biotechnology, 2006, 72(4): 713-719
- [5] Aguilar PS, Hernandez-Arriaga AM, Cybulski LE, et al. Molecular basis of thermosensing: a two-component signal transduction thermometer in *Bacillus subtilis*[J]. The EMBO Journal, 2001, 20(7): 1681-1691
- [6] Jordan S, Hutchings MI, Mascher T. Cell envelope stress response in gram-positive bacteria[J]. FEMS Microbiology Reviews, 2008, 32(1): 107-146
- [7] Jiang M, Shao WL, Perego M, et al. Multiple histidine kinases regulate entry into stationary phase and sporulation in *Bacillus subtilis*[J]. Molecular Microbiology, 2000, 38(3): 535-542
- [8] Lyon GJ, Novick RP. Peptide signaling in *Staphylococcus aureus* and other gram-positive bacteria[J]. Peptides, 2004, 25(9): 1389-1403
- [9] Szurmant H, Ordal GW. Diversity in chemotaxis mechanisms among the bacteria and archaea[J]. Microbiology and Molecular Biology Reviews, 2004, 68(2): 301-319
- [10] Zaghoul L, Tang C, Chin HY, et al. The distribution of insertion sequences in the genome of *Shigella flexneri* strain 2457T[J]. FEMS Microbiology Letters, 2007, 277(2): 197-204
- [11] Deng C, Peng Q, Song FP, et al. Regulation of *cry* gene expression in *Bacillus thuringiensis*[J]. Toxins, 2014, 6(7): 2194-2209
- [12] Deng C, Slamti L, Raymond B, et al. Division of labour and terminal differentiation in a novel *Bacillus thuringiensis* strain[J]. The ISME Journal, 2015, 9(2): 286-296
- [13] Zhang RB, Slamti L, Tong L, et al. The stationary phase regulator CpcR activates *cry* gene expression in non-sporulating cells of *Bacillus thuringiensis*[J]. Molecular Microbiology, 2020, 113(4): 740-754
- [14] Ma LX, Peng Q, Didier L, et al. Effect of plasmid curing on cell differentiation of *Bacillus thuringiensis* strain LM1212[J]. Microbiology China, 2017, 44(3): 574-582 (in Chinese)
- 马丽霞, 彭琦, Didier L, 等. 苏云金芽胞杆菌 LM1212 质粒缺失对细胞分化的影响[J]. 微生物学通报, 2017, 44(3): 574-582
- [15] Stobdan T, Kaur S, Singh A. Cloning and nucleotide sequence of a novel *cry* gene from *Bacillus thuringiensis*[J]. Biotechnology Letters, 2004, 26(14): 1153-1156
- [16] Eid J, Fehr A, Gray J, et al. Real-time DNA sequencing from single polymerase molecules[J]. Science, 2009, 323(5910): 133-138
- [17] Faino L, Seidl MF, Datema E, et al. Single-molecule real-time sequencing combined with optical mapping yields completely finished fungal genome[J]. Microbiology, 2015, 6(4): e00936-15
- [18] Hyatt D, Chen GL, LoCascio PF, et al. Prodigal: prokaryotic gene recognition and translation initiation site identification[J]. BMC Bioinformatics, 2010, 11(1): 119
- [19] Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches[J]. Bioinformatics, 2013, 29(22): 2933-2935
- [20] Lagesen K, Hallin P, Rødland EA, et al. RNAmmer: consistent and rapid annotation of ribosomal RNA genes[J]. Nucleic Acids Research, 2007, 35(9): 3100-3108
- [21] Nawrocki EP, Burge SW, Bateman A, et al. Rfam 12.0: updates to the RNA families database[J]. Nucleic Acids Research, 2015, 43: D130-D137
- [22] Magrane M, Consortium U. UniProt knowledgebase: a hub of integrated protein data[J]. Database, 2011, 2011: bar009
- [23] Tatusova T, Ciufo S, Fedorov B, et al. RefSeq microbial genomes database: new representation and annotation strategy[J]. Nucleic Acids Research, 2014, 42(D1): D553-D559
- [24] Kanehisa M, Araki M, Goto S, et al. KEGG for linking genomes to life and the environment[J]. Nucleic Acids Research, 2008, 36(S1): D480-D484
- [25] Siguier P, Perochon J, Lestrade L, et al. ISfinder: the reference centre for bacterial insertion sequences[J]. Nucleic Acids Research, 2006, 34(S1): D32-D36
- [26] Kurtz S, Phillippy A, Delcher AL, et al. Versatile and open software for comparing large genomes[J]. Genome Biology, 2004, 5(2): R12
- [27] Schindler U, Sans N, Schröder J. Ornithine cyclodeaminase from octopine Ti plasmid Ach5: identification, DNA sequence, enzyme properties, and comparison with gene and enzyme from nopaline Ti plasmid C58[J]. Journal of Bacteriology, 1989, 171(2): 847-854
- [28] Franklin MJ, Ohman DE. Identification of *algI* and *algJ* in the *Pseudomonas aeruginosa* alginate biosynthetic gene cluster which are required for alginate O acetylation[J]. Journal of Bacteriology, 1996, 178(8): 2186-2195
- [29] Raleigh EA, Brooks JE. Restriction modification systems: where they are and what they do[A]//de Bruijn FJ, Lupski

- JR, Weinstock GM. Bacterial Genomes[M]. Boston: Springer, 1998: 78-92
- [30] Hecker M, Pané-Farré J, Uwe V. SigB-dependent general stress response in *Bacillus subtilis* and related gram-positive bacteria[J]. Annual Review of Microbiology, 2007, 61(1): 215-236
- [31] Fujita M, Sadaie Y. Feedback loops involving Spo0A and AbrB in *In vitro* transcription of the genes involved in the initiation of sporulation in *Bacillus subtilis*[J]. The Journal of Biochemistry, 1998, 124(1): 98-104
- [32] Srinivas G, Vennison SJ, Sudha SN, et al. Unique regulation of crystal protein production in *Bacillus thuringiensis* subsp. *yunnanensis* is mediated by the cry protein-encoding 103-megadalton plasmid[J]. Applied and Environmental Microbiology, 1997, 63(7): 2792-2797
- [33] Zhang MS, Zhu DK, Wang MS. Transposases in bacterial insertion sequences and their transposition mechanisms[J]. Chinese Journal of Biochemistry and Molecular Biology, 2018, 34(10): 1057-1064 (in Chinese)
张孟思, 朱德康, 汪铭书. 细菌插入序列的转座酶和转座机制[J]. 中国生物化学与分子生物学报, 2018, 34(10): 1057-1064
- [34] Rosso ML, Delécluse A. Distribution of the insertion element IS240 among *Bacillus thuringiensis* strains[J]. Current Microbiology, 1997, 34(6): 348-353
- [35] Ohba M, Aizawa K. Crystals of *Bacillus thuringiensis* subsp. *yunnanensis* are produced only in asporogenous cells[J]. Journal of Invertebrate Pathology, 1986, 48(2): 254-256
- [36] Kint N, Janoir C, Monot M, et al. The alternative sigma factor σ^B , plays a crucial role in adaptive strategies of *Clostridium difficile* during gut infection[J]. Environmental Microbiology, 2017, 19(5): 1933-1958
- [37] Burbulys D, Trach KA, Hoch JA. Initiation of sporulation in *B. subtilis* is controlled by a multicomponent phosphorelay[J]. Cell, 1991, 64(3): 545-552
- [38] López D, Kolter R. Extracellular signals that define distinct and coexisting cell fates in *Bacillus subtilis*[J]. FEMS Microbiology Reviews, 2010, 34(2): 134-149
- [39] Reder A, Gerth U, Hecker M. Integration of σ^B activity into the decision-making process of sporulation initiation in *Bacillus subtilis*[J]. Journal of Bacteriology, 2012, 194(5): 1065-1074