



## 奥利万星中间体 A82846B 产生菌拟孢囊菌 CCTCC M2020063 全基因组序列分析与 *mbtH* 类基因的验证

程雯<sup>1</sup>, 蒲桂洪<sup>1</sup>, 牛国清<sup>2</sup>, 邹祥<sup>1\*</sup>

<sup>1</sup>西南大学药学院, 重庆 400715

<sup>2</sup>西南大学生物技术中心, 重庆 400715

**摘要:** 【目的】分析荒漠拟孢囊菌 CCTCC M2020063 中 A82846B 合成的代谢途径和关键基因。

【方法】使用 Illumina 二代测序和 PacBio 三代测序技术对荒漠拟孢囊菌 CCTCC M2020063 进行全基因组测序, 利用 Glimmer 预测编码序列, 使用 HPLC 和 LCMS 鉴定次级代谢产物, 使用 antiSMASH 5.0 软件预测次级代谢产物合成基因簇。利用 Geneious 软件对 A82846B 合成相关基因进行分析, 对其中的 *mbtH* 类基因着重分析。【结果】本实验菌株鉴定为荒漠拟孢囊菌 (*Kibdelosporangium aridum*), 基因组中有一条线性染色体, 无质粒, 序列全长 12475688 bp, GC 含量为 66.27%, 有 11900 个开放阅读框, 共有 47 个基因簇。该菌株具有合成 A82846B 的能力, 且生物合成相关基因位于 Cluster32, 包含 33 个基因, *mbtH* 类基因 *gene07864* 的过表达促进 A82846 的合成, 提升了 26.42%, 卤化酶基因为 *gene07859*, 与万古霉素、巴利霉素的卤化酶相似度较高。【结论】本研究对荒漠拟孢囊菌 CCTCC M2020063 进行了基因组序列分析, 获得了 A82846B 生物合成相关的功能基因信息, 为 A82846B 的代谢途径和工程改造提供了强有力的基础。

**关键词:** 荒漠拟孢囊菌, 奥利万星, A82846B, 全基因组组装, *mbtH* 类基因, 卤化酶基因

荒漠拟孢囊菌是一种具有工业价值的抗生素产生菌<sup>[1]</sup>, 拟孢囊菌属的基因组中含有大量次级代谢产物生物合成基因簇, 具有合成糖肽类抗生素等多种活性代谢产物的潜力<sup>[2-3]</sup>。目前报道有荒漠拟孢囊菌<sup>[4]</sup>和东方拟无枝酸菌

(*Amycolatopsis orientalis*)<sup>[5]</sup>两者可产奥利万星 (oritavancin) 的中间体 A82846B。与东方拟无枝酸菌相比, 荒漠拟孢囊菌的相关研究甚少, 特别是缺乏菌株的相关遗传信息分析。A82846 (chloroeremomycin) 的合成需要 5 种氨基酸, 其中

\*通信作者。Tel: +86-23-68251225; Fax: +86-23-68251048; E-mail: zhx1030@swu.edu.cn

收稿日期: 2021-02-19; 修回日期: 2021-06-17; 网络出版日期: 2021-07-07

$\beta$ -羟酪氨酸、对羟基苯甘氨酸和间二羟基苯甘氨酸为胞内合成的非蛋白质氨基酸，在非核糖体肽合成酶 CepA、CepB、CepC 的作用下按 D-亮氨酸、 $\beta$ -羟酪氨酸、L-天门冬酰胺、对羟基苯甘氨酸、对羟基苯甘氨酸、 $\beta$ -羟酪氨酸和间二羟基苯甘氨酸的先后顺序组装成线性的七肽骨架，随后在氧化交联酶 CepE、CepF、CepG 的作用下发生交联反应，成环后形成更稳定的骨架结构，经 3 个糖基转移酶 GtfA、GtfB、GtfC 作用下在第 4 位氨基上连接一个葡萄糖基和一个万古胺基，在第 6 位氨基上连接一个万古胺基，合成过程中还有其他的修饰酶参与作用，最终合成具

有抗菌活性的 A82846<sup>[6]</sup>。A82846B 与万古霉素结构相似，但糖基侧链不同，并且万古霉素的卤化酶能催化两个位点均发生卤化反应，合成途径中不会出现单氯化的中间体<sup>[7]</sup>，而 A82846B 的合成会产生含一个氯原子取代的 A82846A 和无氯原子取代的 A82846C (见图 1-A 中 R<sub>1</sub>、R<sub>2</sub> 位点)，给 A82846B 的分离纯化造成了的困难。

随着万古霉素耐药菌的出现，以及最低抑菌浓度的不断升高，万古霉素正在失去其对耐甲氧西林金黄色葡萄球菌(methicillin-resistant *Staphylococcus aureus*, MRSA)感染的临床疗效，开发新型的糖肽类抗生素十分紧要<sup>[9]</sup>。2014 年 FDA

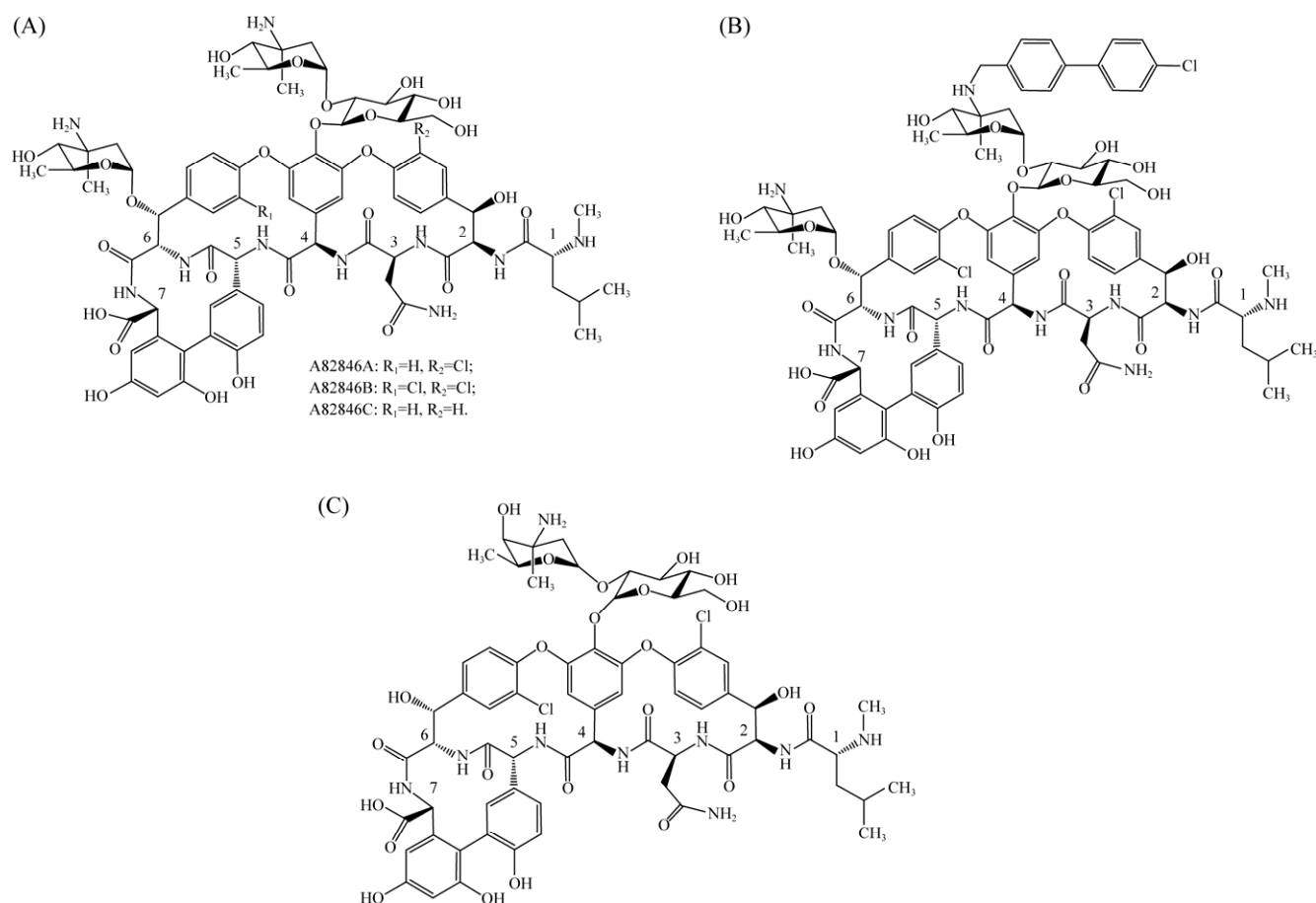


图 1. A82846 (A)、奥利万星(B)和万古霉素(C)的结构示意图<sup>[8]</sup>

Figure 1. The structure of A82846 (A), oritavancin (B) and vancomycin (C)<sup>[8]</sup>.

批准了第二代糖肽类抗生素奥利万星上市<sup>[10]</sup>, 作为一种万古霉素类似物, 奥利万星是 FDA 批准用于由敏感革兰阳性菌引起的急性细菌性皮肤和皮肤结构感染 (acute bacterial skin and skin-structure infections, ABSSSIs) 治疗的首个和唯一的单剂治疗方案的抗菌药物<sup>[11-13]</sup>, 也被 FDA 快速批准为急性细菌性皮肤软组织感染 (skin and soft tissue infections, SSTI) 适应症的合格感染性疾病产品 (qualified infectious disease product, QIDP)<sup>[14]</sup>。目前主要通过微生物来源的 A82846B 进行化学修饰合成奥利万星<sup>[15]</sup>, 因此, A82846B 的微生物合成研究对奥利万星的产业化十分重要。

本课题组对荒漠拟孢囊菌 CCTCC M2020063

的全基因组进行序列分析, 对该菌株的进化关系及 A82846B 合成相关的 *mbtH* 基因和卤化酶有初步的了解。此次基因组研究获得了有关荒漠拟孢囊菌 CCTCC M2020063 菌株的遗传信息, 对后续 A82846B 的代谢调控具有重要的指导意义, 对 A82846B 的工业化生产有着极大的价值。

## 1 材料和方法

### 1.1 菌株、质粒和引物

奥利万星中间体 A82846B 产生菌荒漠拟孢囊菌 CCTCC M2020063 保藏于西南大学药学院发酵工程研究室, 株号 ZH04。其余菌株、质粒和引物见表 1。

表 1. 实验所用菌株、质粒和引物  
Table 1. Strains, plasmids and primers used in the study

Strains/plasmids/ primers	Description	Notes
<b>Strains</b>		
<i>K. aridum</i> ZH04	The original strain capable of producing A82846B	Our laboratory
ZH04-E	<i>K. aridum</i> ZH04 containing pSET-E, Apr <sup>r</sup>	This study
ZH04-M	<i>K. aridum</i> ZH04 containing pSET-M, Apr <sup>r</sup>	This study
<b><i>E. coli</i></b>		
ET12567 (pUZ8002)	Donor strain for intergenetic conjugation between <i>E. coli</i> and <i>Streptomyces</i> , Kan <sup>r</sup> , Chl <sup>r</sup>	From Professor Niu Guoqing of Biotechnology Center of Southwest University
ETZ-E	ET12567 (pUZ8002) containing pSET-E, Kan <sup>r</sup> , Chl <sup>r</sup> , Apr <sup>r</sup>	This study
ETZ-M	ET12567 (pUZ8002) containing pSET-M, Kan <sup>r</sup> , Chl <sup>r</sup> , Apr <sup>r</sup>	This study
<b>Plasmids</b>		
pSET152	<i>E. coli-Streptomyces</i> shuttle vector, Apr <sup>r</sup>	From Professor Niu Guoqing of Biotechnology Center of Southwest University
pSET-E	pSET152 derivative vector containing ermE* promoter and fd terminator	This study
pSET-M	<i>gene07864</i> inserted into <i>Nde</i> I / <i>Eco</i> R I sites of pSET-E	This study
<b>Primers</b>		
M-pSET152-F	TCCACAGGAGGACCCATATGTCGGTCTGAAGATTTCGATGTTG	Amplification of <i>mbtH</i> gene for construction of pSET-M
M-pSET152-R	CGGCCGCGGTACCGGAATTCTCAGGCCGGGTGGTGC	
PSET-YZ-F	GGATCTGTGCACGCGGTC	Verification of recombinant
PSET-YZ-R	GGGTCCGAACAGGAGAGCG	

## 1.2 培养基及基因组 DNA 的提取

ZH04 斜面和平板培养基(g/L): 大豆蛋白胨 10、可溶性淀粉 20、酵母浸粉 10、pH 6.80; TSA 培养基(g/L): 胰蛋白胨 15、大豆蛋白胨 5、NaCl 5、pH 7.0±0.2; 种子培养基(g/L): 葡萄糖 10、麦芽糊精 15、黄豆粉 10、碳酸钙 2、pH 7.2±0.2; 发酵培养基(g/L): 葡萄糖 20、麦芽糊精 25、黄豆粉 10、碳酸钙 4、糖蜜 1、pH 7.2±0.2, 培养温度 28 °C; 大肠杆菌使用 LB 培养基培养(g/L): 蛋白胨 10、酵母浸粉 5、NaCl 5。

提取 ZH04 基因组 DNA: 在平板上刮取 ZH04 菌体接种至 TSA 液体培养基中, 220 r/min 下培养 48 h, 用于基因组 DNA 的制备。使用细菌基因组 DNA 提取试剂盒(天根生化科技(北京)有限公司), 使用 TBS-380 荧光计(Turner BioSystems Inc., Sunnyvale, CA)测定 DNA 浓度, 采用高质量的 DNA 样品( $OD_{260/280}=1.8-2.0$ , DNA 总量  $\geq 15 \mu\text{g}$ , 浓度  $\geq 50 \text{ ng}/\mu\text{L}$ )构建片段文库。

## 1.3 基因组测序、组装及基因预测

利用 PacBio RS II 和 Illumina 平台生成的数据进行生物信息学分析。所有分析均在上海美吉生物的 I-Sanger 云平台(www.i-sanger.com)上进行。

**1.3.1 基因组测序:** 基因组测序使用 PacBio RS II 单分子实时测序(SMRT)和 Illumina 测序平台组合, Illumina 的数据被用来评估基因组重复度, 评估基因组大小、是否存在质粒及污染, 同时保证更完整更精确的组装。

**1.3.2 基因组组装:** 最初的 raw data 是以 fastq 格式储存起来的, 为了使后续的组装更加准确, 对原始数据进行质量剪切, 得到高质量的 clean data。利用 canu 及 HGAP 软件进行 PacBio 数据组装, 将 reads 组装成 contigs, 得到完整的染色

体和质粒基因组。最后利用 illumina 测序数据对组装结果进行校正。

**1.3.3 基因预测:** 利用 Glimmer 对基因组中的编码序列进行预测, 质粒基因采用 GeneMarkS 软件预测, tRNAscan-SE 进行 tRNA 预测, Barrnap 进行 rRNA 预测。

## 1.4 次级代谢合成分析及 A82846B 代谢途径分析

通过 antiSMASH 5.0 软件对样本的次级代谢产物合成基因簇进行预测, 使用 Geneious 软件对合成 A82846B 的相关基因进行比对, 定位 A82846B 合成基因的位置, 使用软件 DNAMAN 8 和 NCBI 在线蛋白比对功能, 对基因组中的卤化酶进行同源性分析。

## 1.5 *mbtH* 类基因的过表达

**1.5.1 质粒 pSET-M 的构建:** 使用 Phanta Max Super-Fidelity DNA Polymerase 在 ZH04 的基因组 DNA 上扩增 *gene07864* 基因片段, 将质粒 pSET-E 从 *Nde* I 和 *Eco*R I 两个限制性酶切位点处切开得到线性载体, 使用 FastPure Gel DNA Extraction Mini Kit 试剂盒将 *gene07864* 基因片段和 pSET-E 线性载体回收, 将扩增的目的片段和双酶切后的线性化载体进行重组得到含有 *gene07864* 基因的质粒 pSET-M。

**1.5.2 改造菌株的构建及发酵:** 将质粒 pSET-E、pSET-M 分别转入 ET12567 (pUZ8002)感受态中, 得到 ETZ-E、ETZ-M, 按链霉菌-大肠杆菌结合转移的操作步骤, 与 ZH04 进行结合转移实验, 筛选出正确的结合子, 分别命名为 ZH04-E、ZH04-M。将 ZH04、ZH04-E、ZH04-M 三株菌在平板培养基上培养 6 d, 接种至种子培养基, 220 r/min 下振荡培养 48 h。以 10%的接种量转接

种子液至发酵培养基中, 振荡培养 6 d 结束发酵, 检测发酵液。

## 1.6 A82846 的测定

**1.6.1 对照品溶液的配制:** 精密称取 A82846B 对照品适量, 加入 ddH<sub>2</sub>O 溶解, 配置成 2 g/L 的溶液, 再逐步稀释得到 1500 mg/L、1000 mg/L、800 mg/L、600 mg/L、200 mg/L、20 mg/L 的溶液, 使用 0.22 μm 的滤膜过滤, 方可进行高效液相色谱分析。将 A82846B 对照品溶液浓度( $x$ )与出峰面积( $f(x)$ )作图, 得到标准曲线  $f(x)=13068.4122x+25656.9474$ ,  $R^2=0.9999$ 。

**1.6.2 供试品溶液的制备:** 用氢氧化钠溶液将发酵液 pH 调至 10.5–10.7, 放置 1 h, 12000 r/min 离心 10 min 去除菌体, 再加入稀盐酸调 pH 至 7.2–7.5, 经多次离心及 0.22 μm 滤膜过滤即可进行高效液相色谱分析。

**1.6.3 高效液相色谱条件:** 色谱柱: Agilent SB-C8 4.6 mm×150 mm, 3.5 μm; 柱温: 40 °C; 流动相: 以 0.1% 三氟乙酸溶液为流动相 A, 乙腈为流动相 B; 流速: 1.0 mL/min; 检测波长: 240 nm; 进样量: 20 μL; 洗脱程序: 0→22 min, A:B=95:5→85:15; 22→24 min, A:B=85:15→95:5; 24→30 min, A:B=95:5→95:5。

**1.6.4 液相色谱质谱联用条件:** 色谱柱: Agilent SB-C8 4.6 mm×150 mm, 3.5 μm; 柱温: 40 °C; 流动相: 以 0.3% 甲酸溶液为流动相 A, 乙腈为流动相 B; 流速: 1.0 mL/min; 检测波长: 240 nm; 进样量: 30 μL; 洗脱程序: 0→22 min, A:B=98:2→90:10; 22→25 min, A:B=90:10→98:2; 25→40 min, A:B=98:2→98:2。离子源: ESI; DL 温度: 250 °C; 雾化气体流量: 2.50 L/min。

## 2 结果和分析

### 2.1 菌株的基因组组装和进化分析

**2.1.1 菌株的基因组组成特征:** Illumina Hiseq 二代测序质控数据显示, 插入序列长 399 bp, 原始 reads 读长 150 bp, 经过一系列的质量剪切之后得到的高质量 reads, 质控后的 reads 用来进行基因组组装。PacBio 三代测序结果显示, 总的 Reads 数量为 85036, 平均 reads 读长 6212.23 bp, 总碱基长 528262860 bp。

完整基因组信息显示, 该菌株为荒漠拟孢囊菌, 其基因组中不存在质粒, 有一条线性染色体, 全长 12475688 bp, 所有碱基的平均 GC 含量 66.27%。预测有 11900 个编码基因, 基因总长度 11324571 bp, 编码基因占基因组百分比为 90.77%, 平均长度 951.64 bp, 其中有 35.29% 的基因长度在 1000 bp 以上。基因组中有 68 个 tRNA, 共 23 种类型。rRNA 共有 9 个, 其中 16S rRNA、23S rRNA 和 5S rRNA 各 3 个。使用 Tandem Repeats Finder 软件进行串联重复序列预测, 有 355 个重复序列, 总长 58370 bp, 占基因组的 0.47%。Circos 基因组圈图见图 2。

**2.1.2 菌株的进化分析:** 通过自动化检索, 在数据库中搜索与样本亲缘关系较近的基因组, 利用基因组单拷贝同源基因构建系统进化树, 基于 16S rRNA 基因和 31 个看家基因(*dnaG*、*frr*、*infC*、*nusA*、*pgk*、*pyrG*、*rplA*、*rplB*、*rplC*、*rplD*、*rplE*、*rplF*、*rplK*、*rplL*、*rplM*、*rplN*、*rplP*、*rplS*、*rplT*、*rpmA*、*rpoB*、*rpsB*、*rpsC*、*rpsE*、*rpsI*、*rpsJ*、*rpsK*、*rpsM*、*rpsS*、*smpB*、*tsf*)的样本进化分析, 显示与本菌株的近源物种为 *Kibdelosporangium phytohabitans* KLBMP1111 (图 3, 4)。

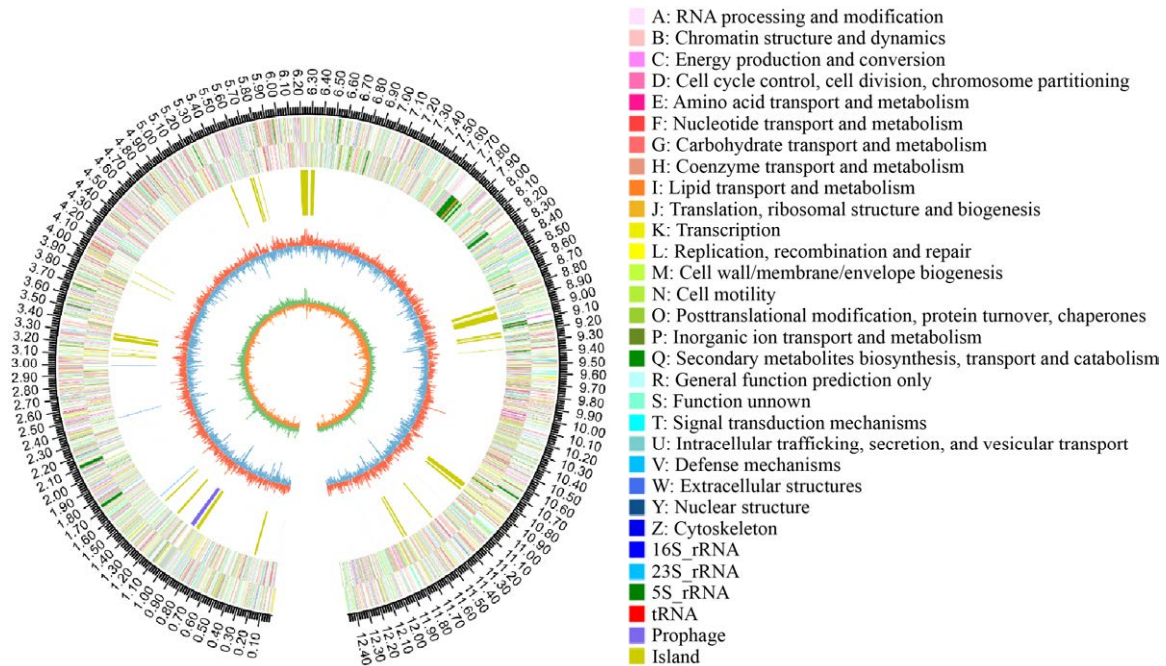


图 2. 荒漠拟孢囊菌 CCTCC M2020063 基因组 Circos 图

Figure 2. Circos map of *Kibdelosporangium aridum* CCTCC M2020063 genome. The outermost circle of the graph is the identification of the genome size; the second circle and the third circle are the CDS on the positive and negative strands; the fourth circle is rRNA and tRNA; the fifth circle is the GC content; the innermost circle is the GC-Skew value.

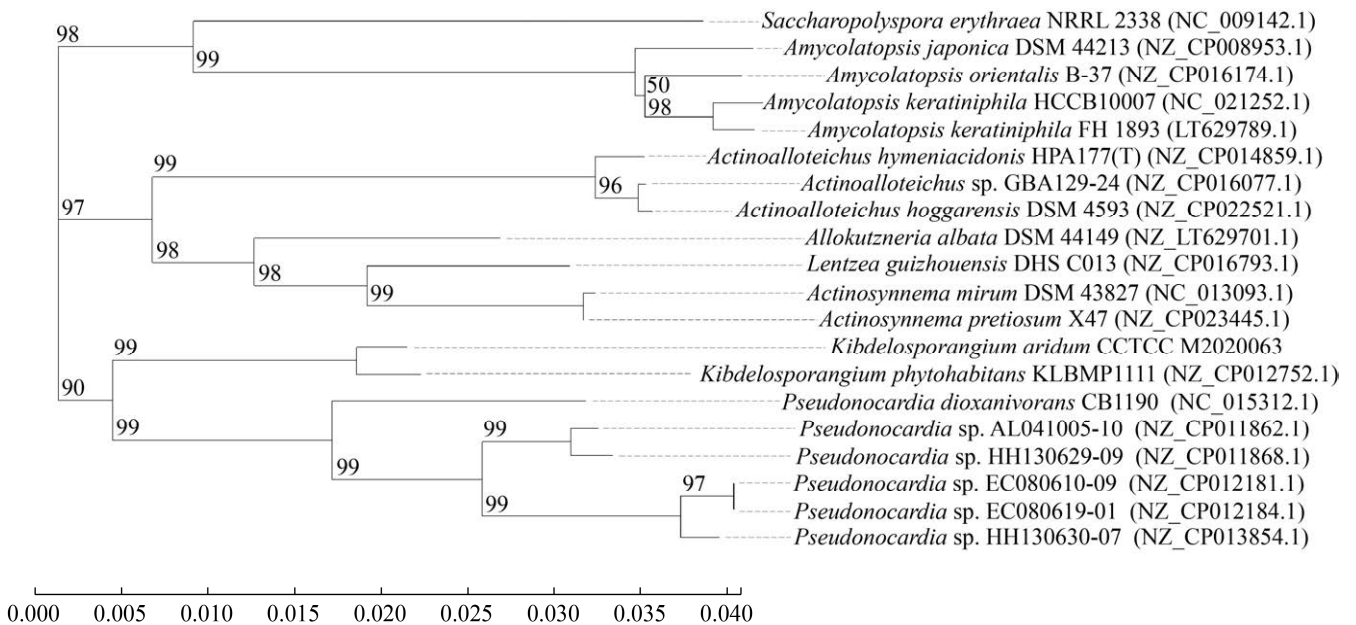


图 3. 基于 16S rRNA 基因序列的系统进化树

Figure 3. Phylogenetic tree of 16S rRNA genes. The sequence number in brackets is the NCBI reference sequence number. The number on the branch point represents the reliability of the data, the larger the value, the more reliable it is. Branch length indicates evolutionary distance. The shorter the horizontal line, the higher the homology.

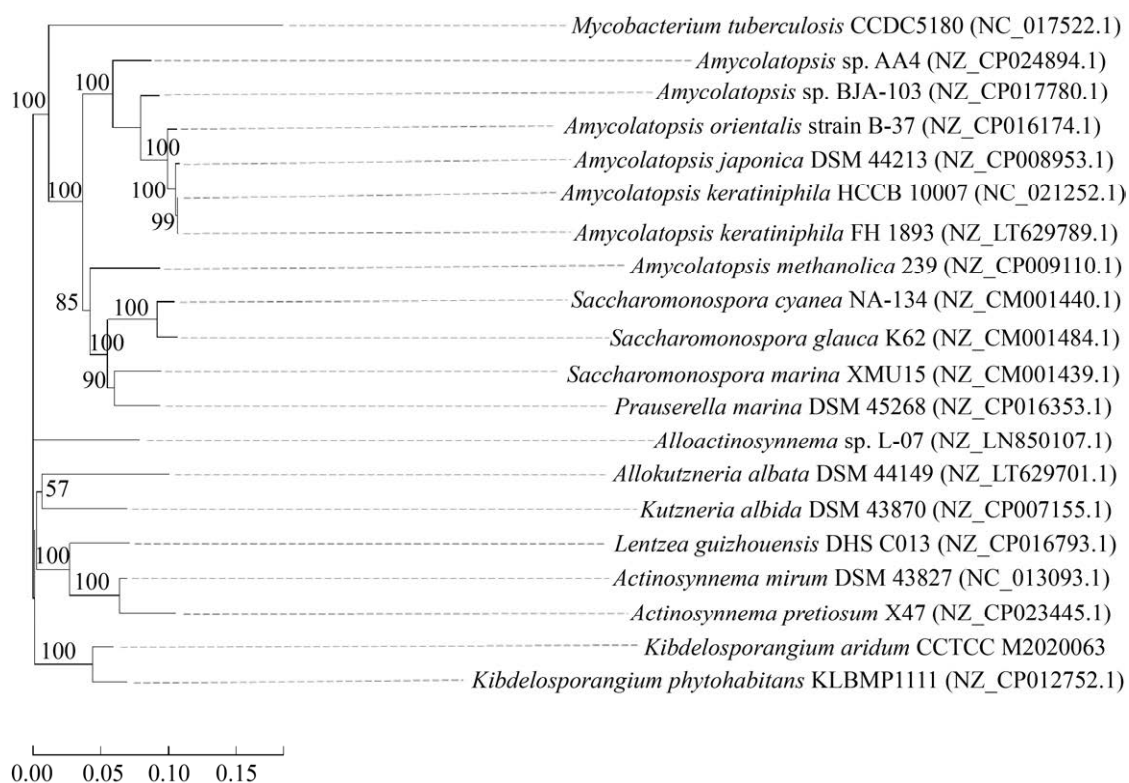


图 4. 基于看家基因的样本进化分析图

Figure 4. Phylogenetic tree of house-keeping genes. The sequence number in brackets is the NCBI reference sequence number. The number on the branch point represents the reliability of the data, the larger the value, the more reliable it is. Branch length indicates evolutionary distance. The shorter the horizontal line, the higher the homology.

## 2.2 A82846B 的合成情况分析

与 KEGG BRITE 数据库(<https://www.kegg.jp/kegg/brite.html>)比对发现, 万古霉素类抗生素(ko01055)合成通路在荒漠拟孢囊菌 CCTCC M2020063 中较为完整(图 5), 从图中可以看出合成 A82846B 的通路完整。

如图 6 所示, A82846B 标准品和荒漠拟孢囊菌 CCTCC M2020063 发酵液样品中的某一物质在高效液相色谱图中同一时间出峰。ESI-MS 给出 A82846B 标准品和发酵液样品的  $m/z$  均为 1593.35  $[M+H]^+$ , 表明二者分子量一致, 为同一种物质, 从 A82846B 的分子式( $C_{73}H_{88}N_{10}O_{26}Cl_2$ )计算其分子量为 1592.43, 因此确定荒漠拟孢囊菌 CCTCC M2020063 的发酵液样品中有 A82846B

存在, 荒漠拟孢囊菌 CCTCC M2020063 确实具备合成 A82846B 的能力。

## 2.3 A82846B 合成基因分析

利用 antiSMASH 5.0 软件(<https://antismash.secondarymetabolites.org/#!/start>)预测出全基因组中共有 47 个次级代谢产物合成基因簇, 根据 A82846B 合成途径中的关键基因, 如非核糖体肽合成酶基因 *cepA*、*cepB*、*cepC*<sup>[16]</sup>, 细胞色素 P450 单加氧酶基因 *cepE*、*cepF*、*cepG*, 以及 3 种糖基转移酶基因 *gtfB*、*gtfA* 和 *gtfC*<sup>[17-18]</sup>等, 定位到 A82846B 生物合成基因位于 Cluster32 (129715 bp), 因此 Cluster32 为 A82846B 合成基因簇, 其中还存在一些全局调控蛋白和未知蛋白的基因。

## Biosynthesis of vancomycin group antibiotics

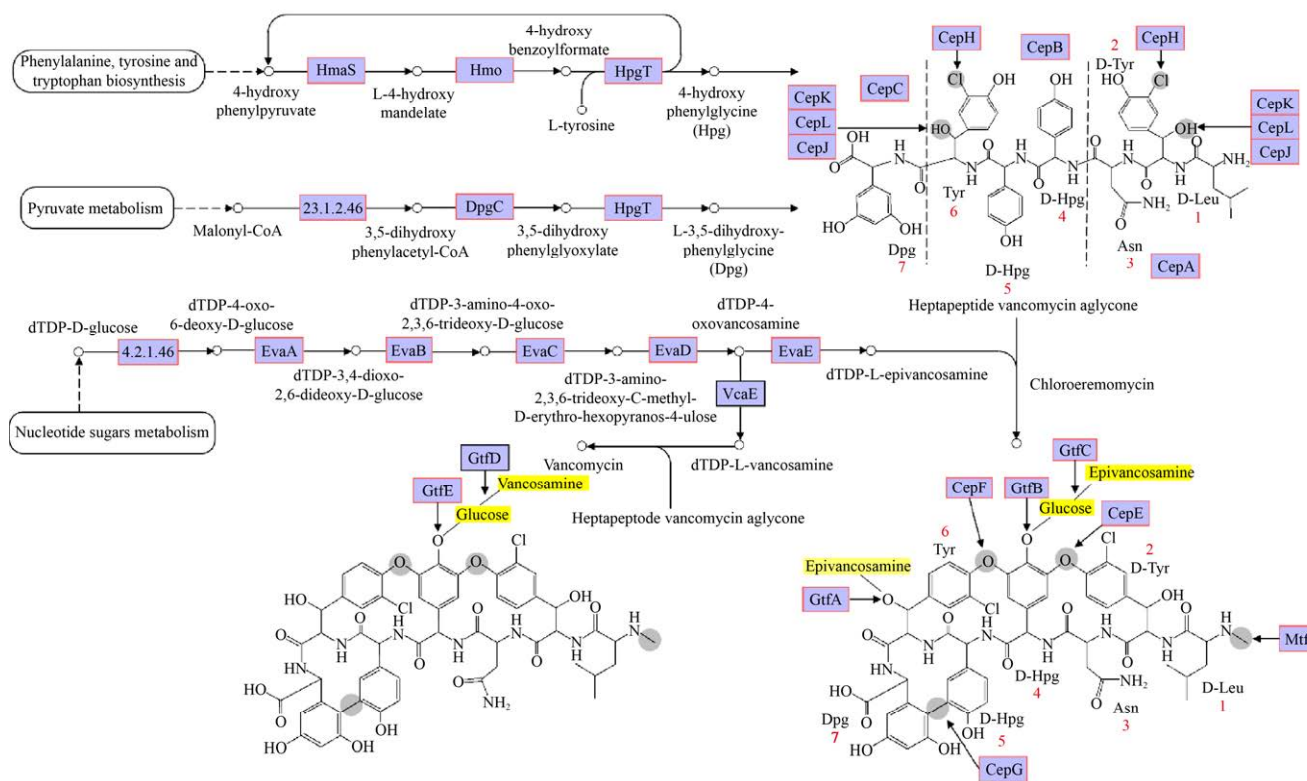


图 5. 万古霉素类抗生素生物合成通路

Figure 5. Biosynthetic pathway of vancomycin group antibiotics. The red border indicates the presence of the gene in the strain.

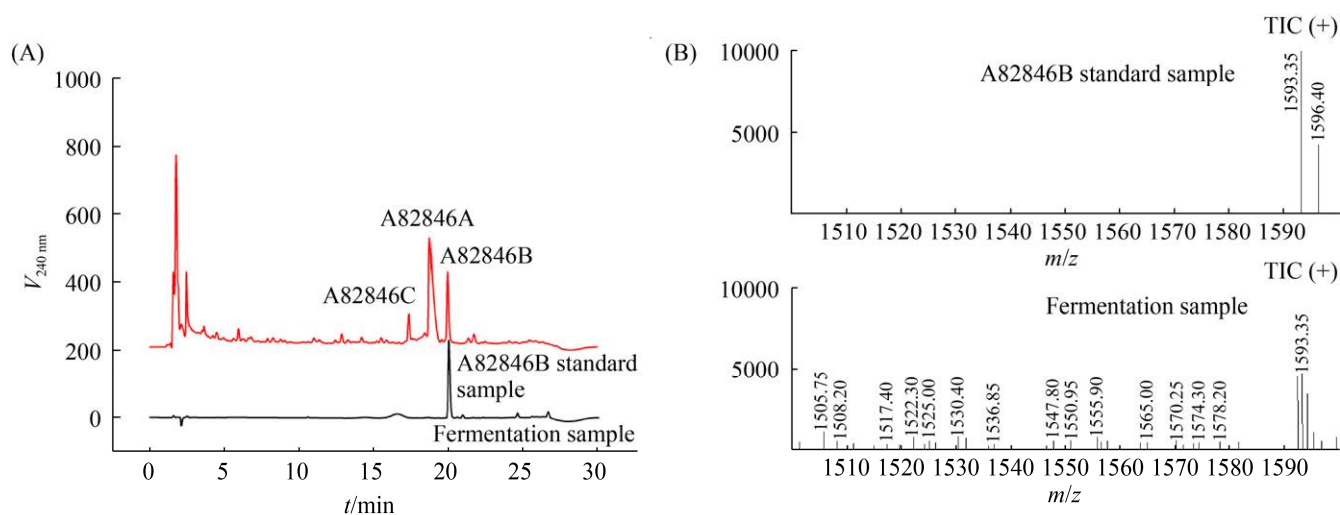


图 6. A82846B 标准品和荒漠拟孢囊菌 CCTCC M202063 发酵液样品的 HPLC (A) 和 LCMS (B) 检测

Figure 6. The HPLC (A) and LCMS (B) detection of A82846B standard sample and the fermentation sample of *Kibdelosporangium aridum* CCTCC M2020063.



为了了解荒漠拟孢囊菌 CCTCC M2020063 基因组中 A82846B 合成基因的详细信息, 对相关基因进行一一比对。东方拟无枝酸菌中有 30 个基因明确属于 Chloroeremomycin 合成通路<sup>[6]</sup>, 其中有 1 个基因其蛋白功能未知(CepD)。根据报道中连续的 30 个基因, 确定本菌株中

有连续的 33 个基因(8163604–8164407)属于 ZH04-chloroeremomycin 的合成通路, 其中有 6 个基因无注释信息。通过基因的顺序以及 Geneious 软件比对蛋白序列, 确定 *gene07869*、*gene07854*、*gene07849* 分别对应蛋白 CepM、CepI、CepL (表 2)。此外, PD 蛋白和 TyrA2

表 2. ZH04-chloroeremomycin 合成基因簇中基因的蛋白序列比对

Table 2. Protein sequence alignment of ZH04-chloroeremomycin biosynthetic cluster

Chloroeremomycin biosynthetic cluster	Accession number	Metabolic pathway	ZH04-chloroeremomycin biosynthetic cluster	Identity/%	Note
PD	CAA11792.1	HPG biosynthesis	TyrA2 (7871)	100.0	
/		Unknown function	/ (7870)	/	
CepM	CAA11793.1	ABC transporter	CepM (7869)	100.0	
CepA	CAA11794.1	NRPS	CepA (7868)	100.0	
/		Unknown function	/ (7867)	/	
CepB	CAA11795.1	NRPS	CepB (7866)	100.0	
CepC	CAA11796.1	NRPS	CepC (7865)	100.0	
CepD	CAA11799.1	NRPS	MbtH (7864)	98.2	MbtH-like protein
CepE	CAA11797.1	Oxidative cross-linking	CepE (7863)	100.0	
CepF	CAA11798.1	Oxidative cross-linking	CepF (7862)	100.0	
/		Unknown function	/ (7861)	/	
CepG	CAA11791.1	Oxidative cross-linking	CepG (7860)	100.0	
CepH	CAA11780.1	Halogenation	CepH (7859)	100.0	
GtfA	CAA11774.1	Glucosyl transferase	GtfA (7858)	99.7	
GtfB	CAA11775.1	Glucosyl transferase	GtfB (7857)	100.0	
GtfC	CAA11776.1	Glucosyl transferase	GtfC (7856)	100.0	
EvaC	CAA11777.1	Sugar biosynthesis	EvaC (7855)	100.0	
CepI	CAA11778.1	Epimerization	CepI (7854)	100.0	
MtfA	CAA11779.1	N-methylation	MtfA (7853)	100.0	
HpgT	CAA11790.1	HPG biosynthesis	HpgT (7852)	100.0	
CepJ	CAA11784.1	$\beta$ -OH-tyrosine biosynthesis	CepJ (7851)	100.0	
CepK	CAA11773.1	$\beta$ -OH-tyrosine biosynthesis	CepK (7850)	100.0	
CepL	CAA11772.1	$\beta$ -OH-tyrosine biosynthesis	CepL (7849)	100.0	
Hmas	CAA11761.1	HPG biosynthesis	Hmas (7848)	100.0	
Hmo	CAA11762.1	HPG biosynthesis	Hmo (7847)	100.0	
EvaA	CAA11763.1	Sugar biosynthesis	EvaA (7846)	100.0	
EvaE	CAA11764.1	Sugar biosynthesis	EvaE (7845)	100.0	
EvaB	CAA11782.1	Sugar biosynthesis	EvaB (7844)	100.0	
EvaD	CAA11781.1	Sugar biosynthesis	EvaD (7843)	100.0	
DpgA	CAA11765.1	DHPG biosynthesis	DpgA (7842)	100.0	
DpgB	CAA11766.1	DHPG biosynthesis	DpgB (7841)	100.0	
DpgC	CAA11787.1	DHPG biosynthesis	DpgC (7840)	100.0	
DpgD	CAA11767.1	DHPG biosynthesis	DpgD (7839)	100.0	

蛋白相同, CepD 蛋白和 MbtH 蛋白同源性达到 98.2%, 说明 CepD 是一种 MbtH 类蛋白, 参与非核糖体肽合成酶的反应。结果表明已报道的 Chloroeremomycin 合成通路中的 30 个基因均在 Cluster32 中找到了相对应的基因, 该菌株具有完整的 A82846B 合成基因, 但其中仍有 *gene07867*、*gene07870*、*gene07861* 的功能未知, 这 3 个基因的序列较短, 猜测是菌株自然进化造成的基因片段插入。

下载已知的卤化酶蛋白序列<sup>[19]</sup>, 与 Cluster32 中卤化酶基因 *gene07859* 的蛋白序列比对, 使用 MEGA 6 进行绘图分析(图 7)。比对结果表明荒漠

拟孢囊菌 CCTCC M2020063 的卤化酶与卤化酶 CepH<sup>[5]</sup> (CAA11780) 完全一致, 同源性为 100%, 与巴利霉素卤化酶 BhaA<sup>[20]</sup> (CAA76550) 和万古霉素卤化酶 Vhal<sup>[21]</sup> (AEI58871) 高度相似, 同源性分别为 95.5% 和 93.7%。

## 2.4 *mbtH* 类基因分析

MbtH 类蛋白是一种小蛋白(约 10 kDa), 主要与一些细菌非核糖体肽合成酶(NRPS)的腺苷化结构域非共价结合, 促进 NRPS 酶的折叠、稳定性和活性<sup>[22-23]</sup>。MbtH 类蛋白对于腺苷化步骤是非常必要的, 在东方拟无枝酸菌中, *mbtH* 类基因(*vcm11*)的缺失突变株完全丧失了产生万古

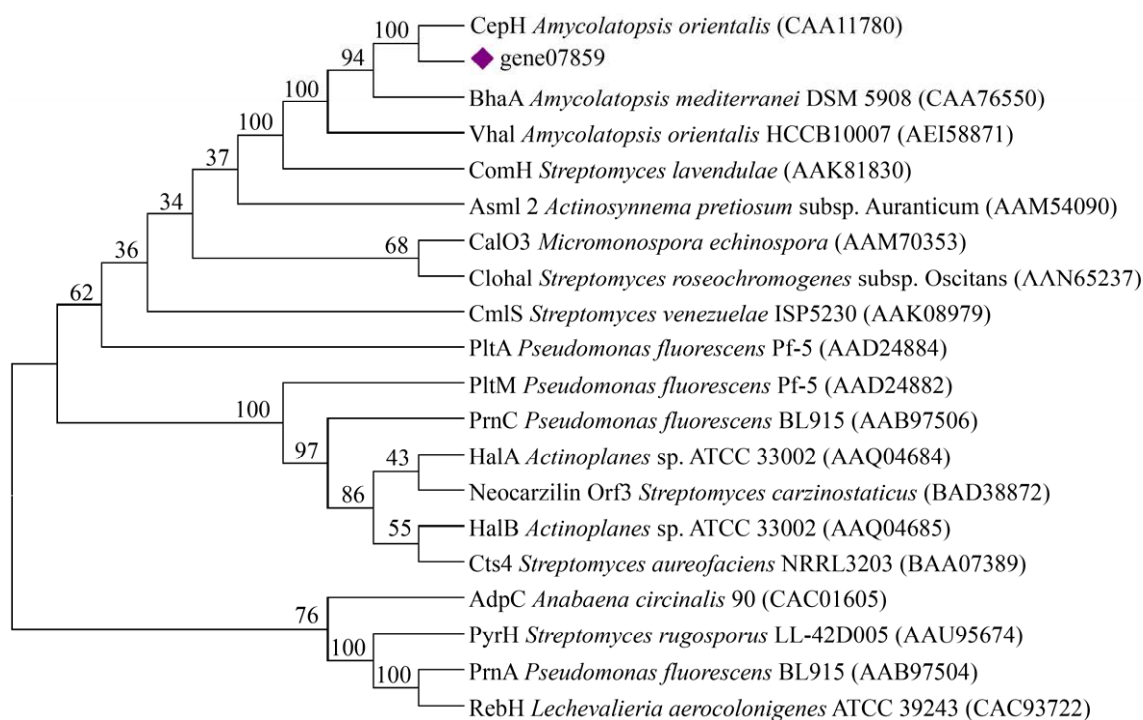


图 7. 黄素依赖型卤化酶的系统进化树分析

Figure 7. Phylogenetic analysis of flavin-dependent halogenase. The sequence number in brackets is the NCBI reference sequence number. The number on the branch point represents the reliability of the data, the larger the value, the more reliable it is. Branch length indicates evolutionary distance. The shorter the horizontal line, the higher the homology.

霉素的能力, 回补后产量几乎完全恢复到野生型菌株的水平, 过表达 *mbtH* 类基因(同源基因 *vcm11* 或非同源基因 *cloY*)均能提高万古霉素的产量<sup>[24]</sup>, 因此 *mbtH* 类基因是一个提高非核糖体肽产量的基因工程靶点。

ZH04-chloroeremomycin 合成通路中 *gene07864* 为 *mbtH* 类基因。为了验证 *mbtH* 类基因对 A82846B 的生物合成的相关性, 对 *gene07864* 进行过表达, 发酵结果见图 8。过表达 *mbtH* 类基因(*gene07864*)对 A82846A 和 A82846C 具有明显的增产效果, 分别提升了 21.66% ( $1020.55 \pm 67.28$  mg/L) 和 94.70% ( $446.95 \pm 30.09$  mg/L), 而 A82846B 的产量降低了 20.37% ( $198.92 \pm 28.02$  mg/L), 整体而言促进了 A82846 的生产, 总产量提升了 26.42%。由于 *mbtH* 类基因是在非核糖体肽生物合成步骤中发挥作用的, 影响七肽骨架的合成, 过表达 *mbtH* 类基因(*gene07864*)能够促进 A82846 总生产的提高,

但其中 A82846B 产量降低的原因尚不能明确, 从该现象可以推测 A82846 生物合成过程中前期的通量发生变化, 会对卤化酶的催化反应有所影响, 造成 A、B、C 三个组分的比例发生波动。

### 3 讨论

A82846B 作为奥利万星合成的关键中间体, 目前微生物发酵是唯一合成途径。若想要借助基因工程的方法来提升产量, 改善杂质的产生, 深入了解 A82846B 的产生菌及其生物合成信息是十分必要的。经全基因组测序, 鉴定本实验室所藏菌种为荒漠拟孢囊菌, 具有丰富的次级代谢产物合成潜力, 可以合成天然糖肽类抗生素 A82846B, 并且该生物合成基因簇位于基因组的 Cluster32 中。经过比对, 了解到在荒漠拟孢囊菌 CCTCC M2020063 中, 参与合成 A82846B 的基因与东方拟无枝酸菌中公布的基因有少许差异,

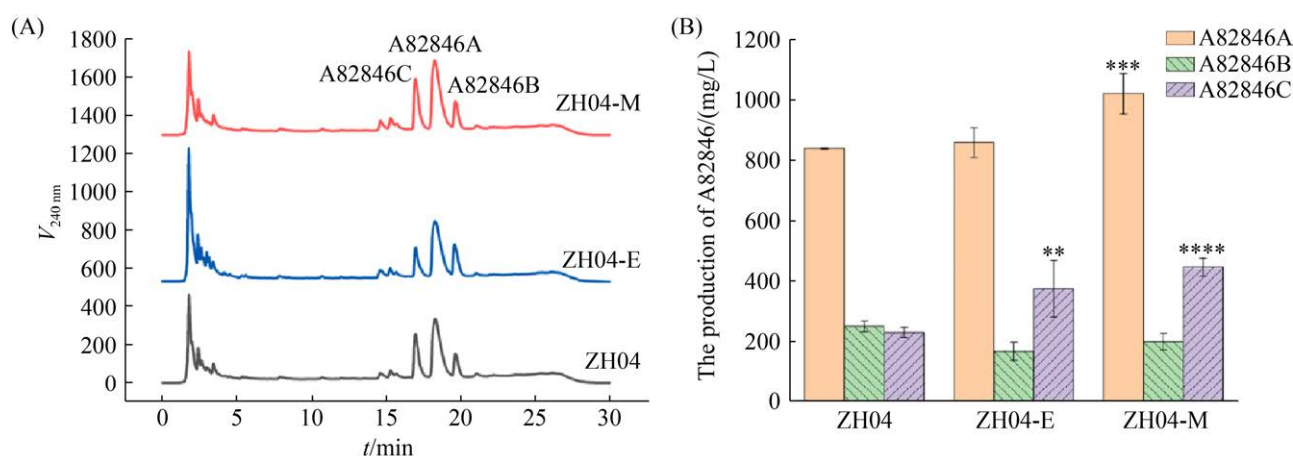


图 8. 在 ZH04 菌株中过表达 *mbtH* 类基因

Figure 8. Overexpression of *mbtH*-like gene in ZH04. A, B: HPLC analysis of A82846 components production (Error bars represent means $\pm$ standard deviation. Significant differences analysis was carried out, \*\*\*\*:  $P < 0.0001$ , \*\*\*:  $P < 0.001$ , \*\*:  $P < 0.01$ ); C: comparison of A82846 yield.

但不影响产物的合成,可能是菌属的差异或适应环境带来的改变。根据全基因组分析的结果,对 A82846B 生物合成基因簇中的 *mbtH* 类基因 *gene07864* 进行过表达,结果表明其对 A82846 总产量有明显的提升作用。由于 MbtH 类蛋白参与七肽骨架的组装,然而 3 个组分在过表达该基因后产量并非一同提升,该结果可能是因为 3 个组分的合成能力与卤化反应的酶息息相关,并且该反应存在一定瓶颈,即便前期合成步骤的通量增加,也依旧受卤化反应限制。

已知过表达卤化酶基因能够提高 A82846B 的产量,并且杂质组分的比例大幅度降低<sup>[5]</sup>,说明现有的菌株中卤化反应的催化能力较低,是造成未氯化或单氯化的杂质(A82846C、A82846A)含量较高的主要原因,A82846B 和 A82846A、A82846C 之间的比例与卤化酶的表达量密切相关,但仅是增加卤化酶基因的拷贝数并不能彻底消除杂质。目前卤化酶的反应机制尚不明确,造成组分结构差异的决定因素是什么,这些需要继续深入研究。要想根除杂质的生成,我们可以从参与卤化反应的多个酶或者从 A82846B 生物合成的整个级联反应过程去思考问题。目前看来单独对卤化酶基因进行研究是不够的,还需充分了解 A82846 的合成机制,这对 A82846B 的工业化生产有重要作用。

全文主要对荒漠拟孢囊菌 CCTCC M2020063 基因组进行了组装和分析,重点关注了 A82846B 生物合成相关基因,初步分析了 *gene07859* 为 A82846B 生物合成基因簇中的卤化酶基因。在原始菌株中过表达了可以促进七肽骨架合成的 *mbtH* 类基因,A82846 的产量得到了提

升,表明 *mbtH* 类基因可以作为一个较好的基因工程靶点。由于 A82846B 产生菌的相关研究较少,荒漠拟孢囊菌是一种稀有放线菌,其基因组信息极为缺乏。通过本文的研究,有助于理解荒漠拟孢囊菌相关的生物学信息,为进一步促进 A82846B 的次级代谢合成奠定基础。

## 参 考 文 献

- [1] Xiong YW, Jiang JH, Qin S. Recent advances on the genus *Kibdelosporangium*. *Acta Microbiologica Sinica*, 2019, 59(5): 799–808. (in Chinese)  
熊有威, 蒋继宏, 秦盛. 拟孢囊菌属放线菌的研究进展. *微生物学报*, 2019, 59(5): 799–808.
- [2] Ratnayake R, Lacey E, Tennant S, Gill JH, Capon RJ. Isokibdelones: novel heterocyclic polyketides from a *Kibdelosporangium* sp. *Organic Letters*, 2006, 8(23): 5267–5270.
- [3] Tsunakawa M, Komiyama N, Tenmyo O, Tomita K, Kawano K, Kotake C, Konishi M, Oki T. New antiviral antibiotics, cycloviracins B1 and B2. I. production, isolation, physico-chemical properties and biological activity. *The Journal of Antibiotics*, 1992, 45(9): 1467–1471.
- [4] Tian X, Huang H, Hu HF. Enhancing A82846B production by artificial *attB*-assisted overexpression of *orf10-orf11* genes in *Kibdelosporangium aridum* SIPI-3927. *AMB Express*, 2020, 10(1): 1–10.
- [5] Wang WY, Yang SB, Wu YJ, Shen XF, Chen SX. Enhancement of A82846B yield and proportion by overexpressing the halogenase gene in *Amycolatopsis orientalis* SIPI18099. *Applied Microbiology and Biotechnology*, 2018, 102(13): 5635–5643.
- [6] Hubbard BK, Walsh CT. Vancomycin assembly: nature's way. *Angewandte Chemie International Edition*, 2003, 42(7): 730–765.
- [7] Schmartz PC, Zerbe K, Abou-Hadeed K, Robinson JA. Bis-chlorination of a hexapeptide-PCP conjugate by the halogenase involved in vancomycin biosynthesis. *Organic & Biomolecular Chemistry*, 2014, 12(30): 5574–5577.

- [8] Marcone GL, Binda E, Berini F, Marinelli F. Old and new glycopeptide antibiotics: from product to gene and back in the post-genomic era. *Biotechnology Advances*, 2018, 36(2): 534–554.
- [9] Blaskovich MAT, Hansford KA, Butler MS, Jia ZG, Mark AE, Cooper MA. Developments in glycopeptide antibiotics. *ACS Infectious Diseases*, 2018, 4(5): 715–735.
- [10] Mattox J, Belliveau P, Durand C. Oritavancin: a novel lipoglycopeptide. *The Consultant Pharmacist*, 2016, 31(2): 86–95.
- [11] Flüh G, Seifert H, Kaasch AJ. Oritavancin: an update. *Future Microbiology*, 2018, 13(7): 727–729.
- [12] Brade KD, Rybak JM, Rybak MJ. Oritavancin: a new lipoglycopeptide antibiotic in the treatment of gram-positive infections. *Infectious Diseases and Therapy*, 2016, 5(1): 1–15.
- [13] Rosenthal S, Decano AG, Bandali A, Lai D, Malat GE, Bias TE. Oritavancin (Orbactiv): a new-generation lipoglycopeptide for the treatment of acute bacterial skin and skin structure infections. *Pharmacy and Therapeutics*, 2018, 43(3): 143–179.
- [14] Tirupathi R, Areti S, Salim SA, Palabindala V, Jonnalagadda N. Acute bacterial skin and soft tissue infections: new drugs in ID armamentarium. *Journal of Community Hospital Internal Medicine Perspectives*, 2019, 9(4): 310–313.
- [15] Chen SX, Yang SB, Wang WY. Fermentation medium and production process of A82846B, the precursor of oritavancin. China: 107557415A. Jan. 09, 2018.
- [16] Van Wageningen AM, Kirkpatrick PN, Williams DH, Harris BR, Kershaw JK, Lennard NJ, Jones M, Jones SJ, Solenberg PJ. Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic. *Chemistry & Biology*, 1998, 5(3): 155–162.
- [17] Mulichak AM, Losey HC, Lu W, Wawrzak Z, Walsh CT, Garavito RM. Structure of the TDP-*epi*-vancosaminyltransferase GtfA from the chloroeremomycin biosynthetic pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 2003, 100(16): 9238–9243.
- [18] Mulichak AM, Losey HC, Walsh CT, Garavito RM. Structure of the UDP-glucosyltransferase GtfB that modifies the heptapeptide aglycone in the biosynthesis of vancomycin group antibiotics. *Structure*, 2001, 9(7): 547–557.
- [19] Dorrestein PC, Yeh E, Garneau-Tsodikova S, Kelleher NL, Walsh CT. Dichlorination of a pyrrolyl-S-carrier protein by FADH<sub>2</sub>-dependent halogenase PltA during pyoluteorin biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102(39): 13843–13848.
- [20] Pelzer S, Süßmuth R, Heckmann D, Recktenwald J, Huber P, Jung G, Wohlleben W. Identification and analysis of the balhimycin biosynthetic gene cluster and its use for manipulating glycopeptide biosynthesis in *Amycolatopsis mediterranei* DSM5908. *Antimicrobial Agents and Chemotherapy*, 1999, 43(7): 1565–1573.
- [21] Xu L, Huang H, Wei W, Zhong Y, Tang B, Yuan H, Zhu L, Huang WY, Ge M, Yang S, Zheng HJ, Jiang WH, Chen DJ, Zhao GP, Zhao W. Complete genome sequence and comparative genomic analyses of the vancomycin-producing *Amycolatopsis orientalis*. *BMC Genomics*, 2014, 15(1): 1–18.
- [22] Schomer RA, Park H, Barkei JJ, Thomas MG. Alanine scanning of YbdZ, an MbtH-like protein, reveals essential residues for functional interactions with its nonribosomal peptide synthetase partner EntF. *Biochemistry*, 2018, 57(28): 4125–4134.
- [23] Zwahlen RD, Pohl C, Bovenberg RAL, Driessen AJM. Bacterial MbtH-like proteins stimulate nonribosomal peptide synthetase-derived secondary metabolism in *Filamentous Fungi*. *ACS Synthetic Biology*, 2019, 8(8): 1776–1787.
- [24] Lee KS, Lee BM, Ryu JH, Kim DH, Kim YH, Lim SK. Increased vancomycin production by overexpression of MbtH-like protein in *Amycolatopsis orientalis* KFCC10990P. *Letters in Applied Microbiology*, 2016, 63(3): 222–228.

# Oritavancin intermediate A82846B-producing strain *Kibdelosporangium aridum* CCTCC M2020063 whole genome sequence analysis and *mbtH*-like gene verification

Wen Cheng<sup>1</sup>, Guihong Pu<sup>1</sup>, Guoqing Niu<sup>2</sup>, Xiang Zou<sup>1\*</sup>

<sup>1</sup> College of Pharmaceutical Sciences, Southwest University, Chongqing 400715, China

<sup>2</sup> Biotechnology Research Center, Southwest University, Chongqing 400715, China

**Abstract:** [Objective] We analyzed the metabolic pathways and key genes of A82846B synthesis of *Kibdelosporangium aridum* CCTCC M2020063. [Methods] We used the second-generation sequencing technology Illumina and the third-generation sequencing technology PacBio to perform whole-genome sequencing of *Kibdelosporangium aridum* CCTCC M2020063. We also used Glimmer to predict the coding sequence. To identify the secondary metabolite, high performance liquid chromatography and liquid chromatograph mass spectrometer was applied. To predict the secondary metabolite synthesis gene cluster, AntiSMASH 5.0 software was applied as well. Besides, focusing on the analysis of its *mbtH*-like gene, geneious software was adopted in this paper to analyze A82846B synthesis-related genes. [Results] Sequence analysis showed that the strain belonged to *Kibdelosporangium aridum*, containing a plasmid-free linear genome chromosome with the full-length gene sequence of 12475688 bp and the GC content of 66.27%. This linear genome chromosome had 11900 open reading frames and a total of 47 gene clusters. This strain was capable to synthesize A82846B, and the biosynthesis-related genes were located in Cluster32, which contained 33 genes. Overexpression of the *mbtH*-like gene *gene07864* promoted the synthesis of A82846 by 26.42%. The halogenase gene was *gene07859*, being highly similar to the halogenase of vancomycin and balimycin. [Conclusion] In this study, we carried out the genome sequence analysis of *Kibdelosporangium aridum* CCTCC M2020063, and we also obtained the information of A82846B biosynthesis-related functional genes, which provided a strong basis for the metabolic pathway and engineering modification of A82846B.

**Keywords:** *Kibdelosporangium aridum*, oritavancin, A82846B, whole genome assembly, *mbtH*-like gene, halogenase gene

(本文责编: 张晓丽)

\*Corresponding author. Tel: +86-23-68251225; Fax: +86-23-68251048; E-mail: zhx1030@swu.edu.cn

Received: 19 February 2021; Revised: 17 June 2021; Published online: 7 July 2021