

• 工业细胞科学 •

**陈实** 楚天学者特聘教授、科技部中青年科技创新领军人才、青年千人计划专家、湖北省医学领军人才。先后硕博毕业于华中农业大学和上海交通大学；在麻省理工学院及哈佛大学进行博士后研究，2011年回到武汉大学任教。任 *Microbiol Res* 和 *BMC Systems Biology* 副主编。以通讯或共同通讯作者在 *Proc Natl Acad Sci USA*、*Nat Commun*、*Angew Chem Int Ed*、*Nucleic Acids Res*、*FEMS Microbiol Rev* 等发表论文。主持基金委重点国际合作项目等，担任 973 课题组长。致力于表观遗传、基因组修饰与编辑、合成生物学与药物发现等研究。



## 基因组学技术解码天然产物合成

池湜甜, 陈实

武汉大学 药学院 组合生物合成与新药发现教育部重点实验室, 湖北 武汉 430071

池湜甜, 陈实. 基因组学技术解码天然产物合成. 生物工程学报, 2019, 35(10): 1889–1900.

Chi HT, Chen C. Genomics approaches decode natural products synthesis. Chin J Biotech, 2019, 35(10): 1889–1900.

**摘要:** 天然产物一直以来都是新药发现的重要来源。自 20 世纪末以来, 随着组学技术的不断发展, 许多生物的基因组被破译并解析, 发现基因组中潜藏着众多未知的天然产物生物合成基因簇, 而这些基因簇在实验室生长条件下无法表达或低表达。因此, 需要综合运用多种学科深入挖掘生物中潜藏的具有新型结构和生物活性的天然产物, 使其广泛应用于人类的生产生活。文中将从天然产物合成基因簇的挖掘、“沉默”天然产物合成途径的激活和生物底盘构建 3 个方面简述基因组学技术在天然产物挖掘中的研究进展。

**关键词:** 基因组学, 天然产物, 基因挖掘, 生物底盘

## Genomics approaches decode natural products synthesis

Haotian Chi, and Shi Chen

Key Laboratory of Combinatorial Biosynthesis and Drug Discovery, Ministry of Education, School of Pharmaceutical Sciences, Wuhan University, Wuhan 430071, Hubei, China

**Abstract:** Novel natural products have always been the most important sources for discovery of new drugs. Since the end of the 20th century, advances in genomics technology have contributed to decode and analyze numerous genomes, revealing

**Received:** March 18, 2019; **Accepted:** July 29, 2019

**Supported by:** National Natural Science Foundation of China (Nos. 31720103906, 31520103902).

**Corresponding author:** Shi Chen. Tel/Fax: +86-27-68756643; E-mail: shichen@whu.edu.cn

国家自然科学基金 (Nos. 31720103906, 31520103902) 资助。

remarkable potential for production of new natural products in organisms. However, this potential is hampered by laboratory culture conditions. Therefore, the integration of all these new advances is necessary to unveil these treasures, addressing the rise in resistance to antibiotics. In this review, we discuss the strategies of genome mining, inducing the expression of silent biosynthetic gene clusters and construction of biological chassis.

**Keywords:** genomics, natural products, genome mining, biological chassis

20 世纪 40 年代, 第一个抗生素青霉素被发现并成功运用于临床以后, 已有数千种结构多样的天然产物被发现并广泛应用于人类的生产生活中<sup>[1]</sup>。天然产物已成为抗生素、免疫抑制剂、抗恶性细胞增生剂、抗高血压和抗病毒等药物的重要来源<sup>[2-3]</sup>。在过去的 35 年中, 美国 FDA 批准的药品超过半数都是源于天然产物<sup>[4]</sup>。因此, 挖掘生物中的天然产物是制药产业的重要支柱之一。

传统观点认为生物只能产生数量有限的次级代谢产物。但随着 DNA 测序技术和生物信息学的快速发展, 链霉菌和曲霉属真菌的第一个基因组序列被公布<sup>[5-6]</sup>, 生物潜藏的天然产物合成能力被发现。生物信息学分析预测链霉菌基因组, 发现其可以产生约 150 000 多种潜在的抗菌药物<sup>[7-8]</sup>, 然而到 2005 年只有约 7 600 个天然产物被确认<sup>[4]</sup>。在标准的实验室生长条件下, 多数次级代谢产物都未能表达或是低表达。这些“沉默”的生物合成基因簇 (Biosynthetic gene clusters, BGCs) 可能编码众多结构新颖的生物活性分子, 而这些天然产物可以广泛用于医药或是畜牧业<sup>[9]</sup>。因此, 挖掘并唤醒这些“沉默”的次级代谢途径可以极大地丰富现有的天然产物库, 为研发更具临床应用价值的药物, 解决愈发严重的抗生素耐药性问题提供来源<sup>[10]</sup>。

组学时代的来临, 基因组学、转录组学、蛋白质组学和代谢组学等学科研究的不断深入, 有利于挖掘并激活生物中“沉默”的天然产物合成途径。本文将从天然产物合成基因簇的挖掘、“沉默”天然产物合成途径激活和生物底盘构建 3 个方面介绍组学技术在天然产物挖掘中的研究进展。

## 1 天然产物合成基因簇的挖掘

16S 核糖体 RNA (Ribosomal RNA, rRNA) 的分析数据显示, 只有 1% 的细菌可以在实验室条件下进行培养<sup>[11]</sup>。因此为了获得活性丰富的天然产物, 就需要突破传统基于生物活性寻找天然产物的方法, 寻找新的方法。宏基因组学则直接从环境 DNA (Environment DNA, eDNA) 中获取隐藏的编码天然产物的基因序列<sup>[12]</sup>。基于快速便捷的测序技术和累积的丰富数据信息, 这种新方法可以打破传统筛选模式, 挖掘更多结构多样、生物活性优良天然产物。

### 1.1 宏基因组学

宏基因组学的挖掘工作流程主要分为两种。传统方法需要构建巨大的 cosmid 文库, 并对单个 eDNA 克隆进行功能筛选。筛选克隆的方法较为多样, 如可视的检测信号、生物活性、报告/生物传感器或特定的催化反应。最新发展的方法是基于生物合成中保守 DNA 序列的相似性, 选择性地获取环境样本中潜在的生物合成基因簇。随后将筛选得到的基因簇进行异源表达, 最终获得新的化合物<sup>[13]</sup>。

传统方法中, 宏基因组文库是通过提取 eDNA、克隆并连接到穿梭载体上, 然后转化到合适的异源宿主中而构建得到的。筛选阳性克隆的主要方法是观察表型或色谱分离鉴定, 也可以用报告/生物传感器筛选系统, 如代谢调节表达系统 (Metabolite-regulated expression, METREX)<sup>[14]</sup>和基质诱导基因表达筛选系统 (Substrate-induced gene

expression screening, SIGEX)<sup>[15]</sup>。然而,这种方法发现的天然产物较少,结构相对比较简单,如抗菌色素 (Violacein)<sup>[16]</sup>、靛玉红 (Indirubin) 及其同分异构体 (Indigo)<sup>[17]</sup>、N-酰基酪氨酸<sup>[18]</sup>等。

最新发展的方法是从环境样品中提取粗 eDNA, 通过聚合酶链反应 (Polymerase chain reaction, PCR) 扩增子特异地针对天然产物生物合成基因簇内的序列进行筛选。其中 PCR 扩增子的混合物称为天然产物序列标签 (Natural product sequence tags, NPSTs)。基于 NPSTs 的系统发育关系预测工具, 如天然产物多样性预测工具 (Environmental surveyor of natural product, eSNaPD)<sup>[19]</sup>和天然产物结构域检测工具 (Natural product domain search, NaPDos)<sup>[20]</sup>, 可以对 DNA 序列标签进行系统分类和生物合成来源评价, 并与数据库进行比较, 重组潜在的目标生物合成基因簇。新方法构建的宏基因组文库是由重组的生物合成基因簇所组成, 可以排除大多数与生物合成途径无关的 DNA 序列。蛋白酶体抑制剂 landepoxin A 是通过该方法筛选得到, 它是通过比较酮合成酶 (Ketosynthase, KS) 结构域序列标签和已知的环氧酮生物合成基因簇的 KS 结构域, 从宏基因组样本中识别环氧酮蛋白酶抑制剂的衍生物而发现<sup>[21]</sup>。Brady 等通过该方法筛选到色氨酸二聚体 hydroxysporine 和 reductasporine<sup>[22]</sup>、细胞毒性蕈环霉素 arimetamycin A<sup>[23]</sup>、抗生素 tetarimycin A<sup>[24]</sup>、malacidins A 和 B<sup>[25]</sup>、蛋白酶体抑制剂 clarepoxcins A-E、landepoxcins A 和 B<sup>[21]</sup>。

## 1.2 生物信息学

自 20 世纪末对生物体进行基因组分析以来, 已有超过 7 000 个微生物的基因组信息被报道。在这之前, 人们认为产生次级代谢物的微生物只含有少数几种 BGCs。然而, 阿维链霉菌和天蓝链霉菌的基因组解码后发现 20–37 个未知的生物合成

基因簇<sup>[6,26–27]</sup>。生物信息学分析显示天然产物生物合成基因簇超过总基因组的 5%–7%<sup>[28]</sup>。丝状真菌构巢曲霉 *Aspergillus nidulans* 的基因组具有更大的潜力, 推测有 56 条次级代谢途径<sup>[29]</sup>。此外, 分析不同生态位的生物群落, 如人类微生物群<sup>[30]</sup>和环境样本<sup>[31]</sup>, 进一步揭示了天然产物的多样性。

随着新一代测序技术的应用, DNA 序列分析的处理速度和准确性显著提高, 成本大幅度降低。获得的大量基因组分析数据让公共数据库很好地建立起来, 利用 BLAST 和 FASTA 可以在一定程度上推断出未知基因的功能<sup>[32–33]</sup>。目前, 除了利用同源性预测功能外, 隐马尔可夫模型 (Hidden Markov model, HMM) 分析蛋白质家族 (Protein family, Pfam) 被广泛应用<sup>[34]</sup>。专门分析预测 BGCs 产物的统计模型 antiSMASH 也被公开<sup>[35]</sup>。由于计算机处理能力的显著提高, 蛋白质功能预测的精确度也有所提升, BGCs 的评估更加精准<sup>[36]</sup>。因此, 从基因组序列数据中挖掘次级代谢生物合成基因簇的方法被广泛应用。

分析基因组中可能存在的次级代谢生物合成酶的编码基因是挖掘新型天然产物的经典方法。虽然次级代谢物结构丰富多样, 但其生物合成机理是相对保守的, 尤其是核心酶的氨基酸序列相似度非常高<sup>[37]</sup>。经典挖掘方法运用了反向遗传学原理, 即以一个或多个“参考”酶的基因序列来确定目标生物基因组序列中的同源基因。常用软件有 BLAST<sup>[38]</sup>、DIAMOND<sup>[39]</sup>、HMMer<sup>[40]</sup> 和 antiSMASH, 其中 antiSMASH 的 3.04 版本可以识别 44 种不同的基因簇类型<sup>[35,41–42]</sup>。经典挖掘方法找到的生物合成基因簇通常都是较容易识别的, 如泰斯巴汀 (Teixobactin)<sup>[43]</sup>、卷曲霉素 (Cypemycin)<sup>[44]</sup>、瑞斯托霉素 (Ristomycin)<sup>[45]</sup>等。

最新发展的方法不再专注于参与合成的单一基因, 而是通过分析部分或全部基因簇、抗性或调

控基因来深入挖掘新型天然产物。Cimermancic 等运用双态 HMM 建立了一个寻找 BGCs 的算法<sup>[46]</sup>。首先用 677 个已验证的次级代谢生物合成基因簇和随机选取的非次级代谢生物合成基因序列中的 Pfam 域字符串验证该模型。然后用该概率模型对 1 154 个原核生物基因组进行筛选, 得到 33 000 个可能的 BGCs, 其中有 10 700 个 BGCs 的评分很高<sup>[46]</sup>。这些分析数据表明挖掘微生物中的新型天然产物潜力巨大。

随着人们对天然产物生物合成途径的认识不断深入, 发现 BGCs 不仅含有编码天然产物合成的生化酶, 还有调控元件、转运蛋白和抗性基因等。因此, 逐渐发展出基于抗性或是调控因子的天然产物挖掘方法。Wright 等验证了对糖肽和安莎霉素类抗生素耐受的微生物更有可能产生类似的化合物, 并开发了基于抗生素耐药性的挖掘平台用于分离特定的抗生素产生菌<sup>[47-48]</sup>。Moore 等则将这种方法进一步改良优化, 开发了一种靶向基因组挖掘方法<sup>[49]</sup>; 他们筛选了 86 株海洋放线菌属 *Salinispora* 的菌株, 确定了一个在细菌脂肪酸合酶附近的非常规 PKS-NRPS 基因簇。通过克隆、异源表达和突变分析, 发现该基因簇与已知的脂肪酸合成酶抑制剂硫拉霉素的生物合成有关。因此, 将假定的耐药基因与“沉默”的次级代谢基因簇相关联, 是一种挖掘天然产物的有效途径。

## 2 “沉默”天然产物合成途径的激活

宏基因组学、生物信息学解析基因组序列发现的 BGCs 在特定生长条件下没有表达或是表达量低检测不到, 这些 BGCs 被称为“沉默”天然产物生物合成基因簇。目前主要有两种策略可以激活这些“沉默”BGCs。一是随机激活, 主要有对生长条件的经验优化<sup>[50]</sup>、添加化学诱导剂<sup>[51-52]</sup>、微量金属离子<sup>[53]</sup>、提供外源性小分子<sup>[54]</sup>、核糖体工程<sup>[55-57]</sup>、与其他生物共培养等方法。二是基于基因组序列的

靶向激活<sup>[58]</sup>, 主要针对调节基因<sup>[59]</sup>。

### 2.1 随机激活

随机激活通过高通量筛选获得目的菌株。改变菌株的生长环境, 如温度、pH、共培养或添加化学诱导剂等, 均可诱导 BGCs 表达<sup>[50-51,53-54]</sup>。在 *Streptomyces* spp. 培养基中加入低浓度稀土元素如铈和镧, 不仅提高了已知抗生素的产量同时诱导了新型天然产物的产生<sup>[60-61]</sup>。2016 年, 筛选含有 30 569 个小分子的化合物库, 发现了 19 个新的化学诱导剂<sup>[62]</sup>。这些诱导剂能够使天蓝色链霉菌菌落的色素沉着, 而色素的产生又通常与天然产物的生成有关。其中 ARC2 被认为是一种极具潜力的化学诱导剂, 其衍生物 CI-ARC 用于筛选 50 株不同的放线菌, 有至少 23% 的 BGCs 被激活, 还有 3 种罕见的对细菌和真核生物具有活性的天然产物<sup>[62-63]</sup>。烟曲霉菌 *Aspergillus fumigatus* 与链霉菌 *Streptomyces rapamycinicus* 进行共培养激活了烟曲霉菌中一个聚酮类 BGC 的表达, 发现了一种新型多酚类天然产物并命名为烟环烷(Fumicyclines)<sup>[64]</sup>。这类化合物对 *S. rapamycinicus* 的生长有一定抑制作用, 并可以在防御真菌时发挥作用。

变铅青链霉菌 *Streptomyces lividans* 中 S12 的突变激活抗生素 actinorhodin 的生物合成途径, 而该 BGC 在变铅青链霉菌中通常是沉默的<sup>[65]</sup>。进一步研究发现编码 RNA 聚合酶 (RNA polymerase, RNAP)、30S 核糖体蛋白 S12 和其他核糖体蛋白的基因突变可以上调特定 BGCs 的转录和翻译。鸟苷四磷酸 (Guanosine tetraphosphate, ppGpp) 是细菌适应环境压力非常重要的信号素, 也可以诱导抗生素的生物合成<sup>[66]</sup>。S12 是细菌核糖体小亚单位的组成部分, 参与翻译起始并决定翻译的准确性。ppGpp 可以与 RNAP 结合, 参与控制 rRNA 的生物合成同时调节 RNAP 在不同启动子上启动转录的能力<sup>[67]</sup>。因此怀疑 RNAP 的突变可以使其模仿与 ppGpp 结合的构象, 由此激活沉默 BGCs 的转

录<sup>[68-69]</sup>。有研究使用作用于核糖体的抗生素链霉素和庆大霉素以及作用于 RNAP 的抗生素利福平分别对不同的放线菌进行筛选,使编码 S12 的基因 *rpsL* 和编码 RNAP 的  $\beta$ -亚基的基因 *rpoB* 自发突变。最后发现中杀菌素链霉菌 *Streptomyces mauvecolor* 的突变株产生了 8 个抗菌天然产物,随后的结构研究显示这是一种新型抗生素 piperidamycins<sup>[55]</sup>。假单胞菌、芽孢杆菌、分支杆菌等微生物的核糖体突变都可以激活生物合成基因簇<sup>[61,65]</sup>。此外,全局调控因子的改变也会诱导某些“沉默”BGCs 的表达。如将天蓝色链霉菌中多效调控基因 *absA1* 的等位基因导入到异源链霉菌 *Streptomyces flavopersicus* 中,诱导 *S. flavopersicus* 产生了具有广谱抗菌活性的 pulvomycin<sup>[70]</sup>。

## 2.2 靶向激活

靶向激活相较于随机激活具有更好的可控性和可预测性。该方法需要设计和实施特定的策略来激活目标 BGCs, 通量低于随机激活。已知天然产物生物合成基因簇内或附近编码转录调控因子的基因会影响天然产物的表达<sup>[58-59,71]</sup>。因此,诱导激活因子的表达或敲除编码抑制因子的基因都是非常有效的激活“沉默”BGCs 的方法。

在产二素链霉菌 *Streptomyces ambofaciens* 基因组中发现一个编码 PKS 的基因簇<sup>[72]</sup>。生物信息学分析预测该 BGC 的代谢产物是一种含有新型碳骨架的聚酮化合物,但 qRT-PCR 分析显示该基因簇在实验室生长条件下低表达。分析发现基因簇内的 *samR0484* 基因可能编码特异性转录激活因子,该因子属于 LAL 家族调控因子。过表达 *samR0484* 可诱导该 PKS 和其他生物合成基因的表达,发现了一类新型的大环内酯类抗生素 stambomycins<sup>[72]</sup>。TetR 家族抑制蛋白含有一个螺旋的 DNA 结合结构域和一个响应小分子配体的受体结构域。在特定的响应配体缺失时, TetR 二聚体会与 DNA 结合阻止下游基因的转录<sup>[73]</sup>。失活 TetR 家族抑制蛋白可以

激活多杰霉素 (Jadomycin)<sup>[74-75]</sup>、卡那霉素 (Kinamycin)<sup>[76]</sup>、aurici<sup>[77]</sup>、ceolimycin<sup>[78]</sup> 和 gaburedins<sup>[79]</sup>等多种天然产物的表达。

目前使用天然启动子、修饰启动子和合成启动子来挖掘天然产物是非常有效的靶向激活方法<sup>[58]</sup>。启动子可能是激活“沉默”BGCs 和提高天然产物产量所必需的。因此有研究小组不断丰富启动子库<sup>[80]</sup>。他们已经成功合成了 56 个启动子,将其分为强、中、弱 3 种不同强度的启动子,并验证在多个放线菌菌株中均有活性<sup>[81]</sup>。这种合成启动子库为未来更好地靶向激活生物合成途径提供了更多选择。

## 3 生物底盘构建

多数生物在实验室生长条件下较难培养,遗传背景复杂且不易进行基因操作,其他代谢途径也会影响目标 BGC 的表达。因此,构建具有精简化基因组的生物底盘,对于挖掘天然产物至关重要<sup>[82]</sup>。目前构建生物底盘主要有“自上而下”和“自下而上”两种策略。

### 3.1 “自上而下”构建策略

基于对基因组认识的不断深入,“自上而下”主要是去除赘余基因,降低非必要代谢途径和复杂调控网络的干扰。优化细胞代谢途径提高对底物和能量的利用率,增加底盘的可预测性和可控性<sup>[83]</sup>。

“自上而下”的构建思路主要是通过比较分析来自不同生物的基因组,并结合已有的实验数据,分析出非必需基因<sup>[84]</sup>。然后,使用质粒和线性 DNA 介导、位点特异性重组酶、转座子和 CRISPR/Cas 系统等方法进行缺失,构建基因组精简的底盘。目前,在大肠杆菌<sup>[85-89]</sup>、链霉菌<sup>[82,90-92]</sup>、枯草芽孢杆菌<sup>[93-95]</sup>和假单胞菌<sup>[96-98]</sup>等生物中开展了底盘构建工作。以阿维链霉菌 SUKA17 为例,研究者使用 Cre-loxP 重组系统对阿维链霉菌基因组进行分步缺失,去除了约 20%的基因组,消除了 67%

的 IS 序列和 78% 的转座基因。此外, SUKA17 也是首个用于异源表达次级代谢生物合成途径的放线菌底盘<sup>[92]</sup>。SUKA17 基因组的稳定性增加且可以在基础培养基中正常生长, 不产阿维链霉菌内源次级代谢物如阿维霉素、寡霉素、菲律宾菌素和萜类化合物等。目前, 已有超过 30 多种不同的生物合成基因簇在 SUKA17 中进行表达, 如核苷类、多肽类、氨基糖苷类等<sup>[91,99]</sup>。多数 BGCs 异源导入到底盘后可直接检测到目标代谢物, 少数合成基因簇需要导入调控基因。如将 75 kb 含有 *pladienolide* 整个生物合成基因簇的 BAC 克隆插入到 SUKA17 中, 并没有产生 *pladienolide*。转录分析表明, *pladienolide* 生物合成基因簇和 SUKA17 基因组中并没有相关的调控基因可以激活该代谢途径。在导入调控基因 *pldR* 后, 突变株可以合成 *pladienolide*。此外, SUKA17 可以使经密码子优化后的合成基因有效表达, 产生植物萜类合成中间前体物质。阿维链霉菌基因敲除后的底盘更适用于表达多种不同的次级代谢生物合成基因簇, 可能因为精简化的底盘能为新引入的异源代谢途径提供充足的生物合成前体物<sup>[100]</sup>。这些充分说明生物底盘的基因组稳定性好且系统兼容性强, 为天然产物的挖掘提供良好的基础。

### 3.2 “自下而上”构建策略

“自下而上”构建策略则是尝试从头合成生物底盘。随着 DNA 合成、测序和移植技术的不断发展, 使从头合成复杂且长的 DNA 序列成为可能<sup>[101]</sup>。该方法主要使用 PCR 组装具有同源序列的短片段, 最终构建得到全合成生物底盘<sup>[102-103]</sup>。

2004 年, 使用 PCR 技术成功合成了一个长 32 kb 的聚酮合酶基因簇<sup>[103]</sup>。虽然合成价格昂贵且错误率高, 但这一成果使制备人工细胞成为可能。2008 年, Venter 等合成了第一个完整的生物体基因组尿道支原体 *M. genitalium* JCVI-1.0<sup>[101]</sup>。该全合成基因组长 582 970 bp, 所含基因与野生型

尿道支原体 *M. genitalium* G37 几乎完全相同<sup>[104]</sup>。构建流程为化学合成 101 个 5–7 kb 具有重叠序列的片段, 并通过测序进行了验证。然后, 通过体外重组得到约 24 kb、72 kb (1/8 基因组) 和 144 kb (1/4 基因组) 的中间片段。最后在酿酒酵母中组装得到完整的基因组。

计算机设计好基因组序列, 通过合成、组装得到 1.08 Mb 的丝状支原体基因组 JCVI-syn1.0<sup>[104-105]</sup>。将其移植到山羊支原体受体细胞中, 产生仅受该合成染色体控制的丝状支原体细胞。JCVI-syn1.0 的组装是通过体外酶策略和酵母体内重组相结合完成的。JCVI-syn1.0 具有预期的表型特征, 能够持续自我复制。随后基于 syn1.0 基因组, 设计并构建了 531 kb 的基因组 JCVIsyn3.0, 该基因组含有 473 个基因, 编码 438 个蛋白和 35 个带注释的 RNA<sup>[106]</sup>。合成基因组 JCVIsyn3.0 比任何已知的可以在无菌培养中生长的生物体的基因组都要小。首先, 利用 Tn5 诱变技术来识别必要基因、准必要基因和非必要基因。随后, 验证准必要基因。最后, 通过设计、合成和测试, 构建得到最小基因组 JCVIsyn3.0。JCVIsyn3.0 是一个多功能底盘, 可以用于研究生命的核心功能和探索全基因组。这样将为未来天然产物的挖掘提供更多可能。

## 4 展望

20 世纪末, 生物基因组测序的实现让我们了解到基因组中天然产物生物合成基因簇比预想的更多。实质上只有很小部分的天然产物被开发, 多数在实验室生长条件下是保持“沉默”的。随着生物信息学、基因测序、分子遗传学等学科快速发展, 天然产物合成基因簇的挖掘、代谢途径的激活和生物底盘的构建在天然产物挖掘中得到了广泛应用。我们可以综合使用各种策略高效地挖掘并唤醒生物中“沉默”的生物合成基因簇 (图 1)。

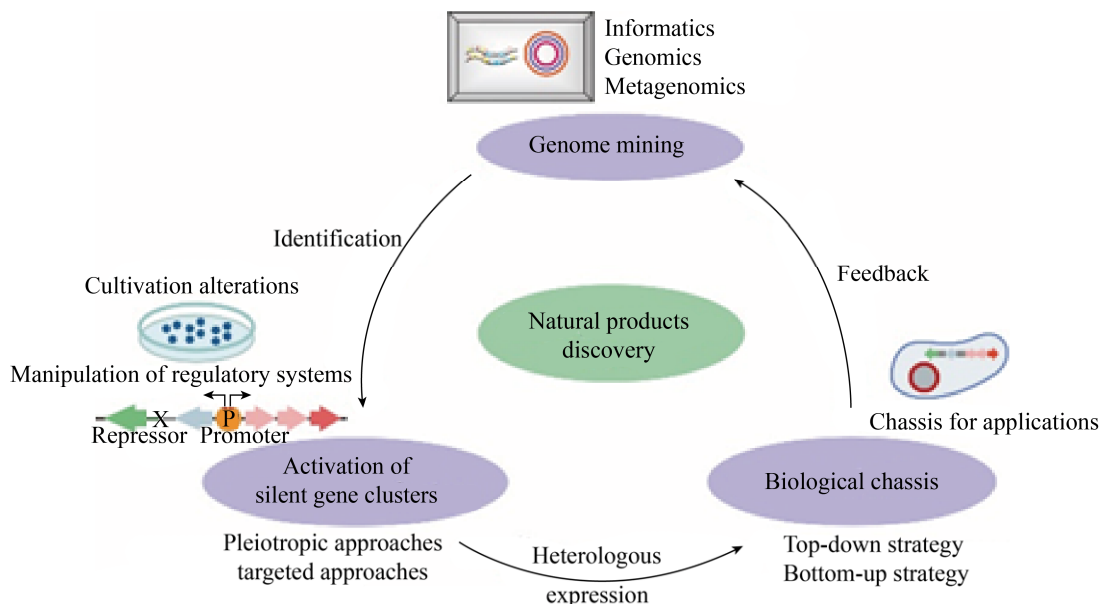


图 1 基因组学技术解析天然产物合成的流程

Fig. 1 Schematic illustration of natural products synthesis using genomics approaches.

天然产物合成基因簇的挖掘中，宏基因组学揭示了未经培养的生物可以产生结构复杂并具有生物活性的天然产物，证明 eDNA 是一个丰富资源。然而宏基因组学仍需要突破两个局限性问题，一是如何从土壤样本中分离得到宏基因组学的 DNA 样本库。因为收集含有新型生物合成途径的 eDNA 需要花费大量时间，并且需要懂得基础的生态学知识。二是如何将宏基因组文库中的 BGCs 进行异源表达。多数 BGCs 在异源宿主中是沉默的，可能是受密码子、稀有 tRNAs、毒性或稳定性等因素的影响。在挖掘到可能的天然产物生物合成途径后，需要随机激活或靶向激活“沉默”基因簇的表达。通常随机激活不需要预先了解 BGC 的机理，且可以进行高通量的筛选，但对激活结果却无法预估。靶向激活则需要预先对通路有所了解，其通量较低但是具有可控性和可预测性。此外，生物底盘的构建无论对于基因组挖掘还是激活“沉默”基因簇的表达都非常重要。目前有“自上而下”和“自下而上”两种策略。其中“自上而下”是通过去除冗余基因得到精简化的生物底盘。“自

下而上”则是从头合成基因组。天然产物挖掘工作充满挑战，但随着组学技术的不断发展，将会有更多具有新型结构的天然产物被发现并广泛用于人们的生产生活中。

## REFERENCES

- [1] Jones D, Metzger HJ, Schatz A, et al. Control of gram-negative bacteria in experimental animals by Streptomycin. *Science*, 1944, 100(2588): 103–105.
- [2] Staunton J, Weissman KJ. Polyketide biosynthesis: a millennium review. *Nat Prod Rep*, 2001, 18(4): 380–416.
- [3] Meier JL, Burkart MD. Proteomic analysis of polyketide and nonribosomal peptide biosynthesis. *Curr Opin Chem Biol*, 2011, 15(1): 48–56.
- [4] Bérdy J. Bioactive microbial metabolites. *J Antibiot (Tokyo)*, 2005, 58(1): 1–26.
- [5] Machida M, Asai K, Sano M, et al. Genome sequencing and analysis of *Aspergillus oryzae*. *Nature*, 2005, 438(7071): 1157–1161.
- [6] Ōmura S, Ikeda H, Ishikawa J, et al. Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. *Proc Natl Acad*



- Sci USA, 2001, 98(21): 12215–12220.
- [7] Girot E. Nurse education in the new millennium. A review of current reports. *Br J Perioper Nurs*, 2001, 11(8): 352–361.
- [8] Challinor VL, Bode HB. Bioactive natural products from novel microbial sources. *Ann N Y Acad Sci*, 2015, 1354(1): 82–97.
- [9] Challis GL, Hopwood DA. Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species. *Proc Natl Acad Sci USA*, 2003, 100(S2): 14555–14561.
- [10] Baltz RH. Gifted microbes for genome mining and natural product discovery. *J Ind Microbiol Biotechnol*, 2017, 44(4/5): 573–588.
- [11] Vartoukian SR, Palmer RM, Wade WG. Strategies for culture of ‘unculturable’ bacteria. *FEMS Microbiol Lett*, 2010, 309(1): 1–7.
- [12] Brady SF, Simmons L, Kim JH, et al. Metagenomic approaches to natural products from free-living and symbiotic organisms. *Nat Prod Rep*, 2009, 26(11): 1488–1503.
- [13] Lok C. Mining the microbial dark matter. *Nature*, 2015, 522(7556): 270–273.
- [14] Williamson LL, Borlee BR, Schloss PD, et al. Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. *Appl Environ Microbiol*, 2005, 71(10): 6335–6344.
- [15] Uchiyama T, Abe T, Ikemura T, et al. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nat Biotechnol*, 2005, 23(1): 88–93.
- [16] Brady SF, Chao CJ, Handelsman J, et al. Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. *Org Lett*, 2001, 3(13): 1981–1984.
- [17] Lim HK, Chung EJ, Kim JC, et al. Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli*. *Appl Environ Microbiol*, 2005, 71(12): 7768–7777.
- [18] Brady SF, Chao CJ, Clardy J. New natural product families from an environmental DNA (eDNA) gene cluster. *J Am Chem Soc*, 2002, 124(34): 9968–9969.
- [19] Reddy BVB, Milshteyn A, Charlop-Powers Z, et al. eSNaPD: a versatile, web-based bioinformatics platform for surveying and mining natural product biosynthetic diversity from metagenomes. *Chem Biol*, 2014, 21(8): 1023–1033.
- [20] Ziemert N, Podell S, Penn K, et al. The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS ONE*, 2012, 7(3): e34064.
- [21] Owen JG, Charlop-Powers Z, Smith AG, et al. Multiplexed metagenome mining using short DNA sequence tags facilitates targeted discovery of epoxyketone proteasome inhibitors. *Proc Natl Acad Sci USA*, 2015, 112(14): 4221–4226.
- [22] Chang FY, Ternei MA, Calle PY, et al. Targeted metagenomics: finding rare tryptophan dimer natural products in the environment. *J Am Chem Soc*, 2015, 137(18): 6044–6052.
- [23] Kang HS, Brady SF. Arimetamycin a: improving clinically relevant families of natural products through sequence-guided screening of soil metagenomes. *Angew Chem Int Ed Engl*, 2013, 52(42): 11063–11067.
- [24] Kallifidas D, Kang HS, Brady SF. Tetarimycin A, an MRSA-active antibiotic identified through induced expression of environmental DNA gene clusters. *J Am Chem Soc*, 2012, 134(48): 19552–19555.
- [25] Hover BM, Kim SH, Katz M, et al. Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. *Nat Microbiol*, 2018, 3(4): 415–422.
- [26] Bentley SD, Chater KF, Cerdeño-Tárraga AM, et al. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, 2002, 417(6885): 141–147.
- [27] Ikeda H, Ishikawa J, Hanamoto A, et al. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol*, 2003, 21(5): 526–531.
- [28] Ikeda H. Natural products discovery from micro-organisms in the post-genome era. *Biosci, Biotechnol, Biochem*, 2017, 81(1): 13–22.
- [29] Yaegashi J, Oakley BR, Wang CCC. Recent advances in genome mining of secondary metabolite biosynthetic gene clusters and the development of



- heterologous expression systems in *Aspergillus nidulans*. *J Ind Microbiol Biotechnol*, 2014, 41(2): 433–442.
- [30] Mousa WK, Athar B, Merwin NJ, et al. Antibiotics and specialized metabolites from the human microbiota. *Nat Prod Rep*, 2017, 34(11): 1302–1331.
- [31] Milshcheyn A, Schneider JS, Brady SF. Mining the metabiome: identifying novel natural products from microbial communities. *Chem Biol*, 2014, 21(9): 1211–1223.
- [32] Ziemert N, Alanjary M, Weber T. The evolution of genome mining in microbes - a review. *Nat Prod Rep*, 2016, 33(8): 988–1005.
- [33] Medema MH, Fischbach MA. Computational approaches to natural product discovery. *Nat Chem Biol*, 2015, 11(9): 639–648.
- [34] Finn RD, Coghill P, Eberhardt RY, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*, 2016, 44(D1): D279–D285.
- [35] Weber T, Blin K, Duddela S, et al. antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res*, 2015, 43(W1): W237–W243.
- [36] Machado H, Tuttle RN, Jensen PR. Omics-based natural product discovery and the lexicon of genome mining. *Curr Opin Microbiol*, 2017, 39: 136–142.
- [37] Weber T, Welzel K, Pelzer S, et al. Exploiting the genetic potential of polyketide producing streptomycetes. *J Biotechnol*, 2003, 106(2/3): 221–232.
- [38] Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. *J Mol Biol*, 1990, 215(3): 403–410.
- [39] Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*, 2015, 12(1): 59–60.
- [40] Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*, 2011, 39(S2): W29–W37.
- [41] Medema MH, Blin K, Cimermancic P, et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res*, 2011, 39(S2): W339–W346.
- [42] Blin K, Medema MH, Kazempour D, et al. antiSMASH 2.0-a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res*, 2013, 41(S1): W204–W212.
- [43] Ling LL, Schneider T, Peoples AJ, et al. A new antibiotic kills pathogens without detectable resistance. *Nature*, 2015, 517(7535): 455–459.
- [44] Claesen J, Bibb M. Genome mining and genetic analysis of cypemycin biosynthesis reveal an unusual class of posttranslationally modified peptides. *Proc Natl Acad Sci USA*, 2010, 107(37): 16297–16302.
- [45] Spohn M, Kirchner N, Kulik A, et al. Overproduction of Ristomycin A by activation of a silent gene cluster in *Amycolatopsis japonicum* MG417-CF17. *Antimicrob Agents Chemother*, 2014, 58(10): 6185–6196.
- [46] Cimermancic P, Medema MH, Claesen J, et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell*, 2014, 158(2): 412–421.
- [47] Thaker MN, Wang WL, Spanogiannopoulos P, et al. Identifying producers of antibacterial compounds by screening for antibiotic resistance. *Nat Biotechnol*, 2013, 31(10): 922–927.
- [48] Thaker MN, Waglechner N, Wright GD. Antibiotic resistance-mediated isolation of scaffold-specific natural product producers. *Nat Protoc*, 2014, 9(6): 1469–1479.
- [49] Tang XY, Li J, Millán-Aguñaga N, et al. Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. *ACS Chem Biol*, 2015, 10(12): 2841–2849.
- [50] Rutledge PJ, Challis GL. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat Rev Microbiol*, 2015, 13(8): 509–523.
- [51] Seyedsayamdost MR. High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. *Proc Natl Acad Sci USA*, 2014, 111(20): 7266–7271.
- [52] Moore JM, Bradshaw E, Seipke RF, et al. Use and discovery of chemical elicitors that stimulate biosynthetic gene clusters in *Streptomyces* bacteria.

- Methods Enzymol, 2012, 517: 367–385.
- [53] Locatelli FM, Goo KS, Ulanova D. Effects of trace metal ions on secondary metabolism and the morphological development of streptomycetes. *Metallomics*, 2016, 8(5): 469–480.
- [54] Okada BK, Seyedsayamdost MR. Antibiotic dialogues: induction of silent biosynthetic gene clusters by exogenous small molecules. *FEMS Microbiol Rev*, 2017, 41(1): 19–33.
- [55] Hosaka T, Ohnishi-Kameyama M, Muramatsu H, et al. Antibacterial discovery in actinomycetes strains with mutations in RNA polymerase or ribosomal protein S12. *Nat Biotechnol*, 2009, 27(5): 462–464.
- [56] Ochi K, Hosaka T. New strategies for drug discovery: activation of silent or weakly expressed microbial gene clusters. *Appl Microbiol Biotechnol*, 2013, 97(1): 87–98.
- [57] Tanaka Y, Kasahara K, Hirose Y, et al. Activation and products of the cryptic secondary metabolite biosynthetic gene clusters by rifampin resistance (*rpoB*) mutations in actinomycetes. *J Bacteriol*, 2013, 195(13): 2959–2970.
- [58] Rebets Y, Brötz E, Tokovenko B, et al. Actinomycetes biosynthetic potential: how to bridge *in silico* and *in vivo*? *J Ind Microbiol Biotechnol*, 2014, 41(2): 387–402.
- [59] Rigali S, Anderssen S, Naômé A, et al. Cracking the regulatory code of biosynthetic gene clusters as a strategy for natural product discovery. *Biochem Pharmacol*, 2018, 153: 24–34.
- [60] Kawai K, Wang GJ, Okamoto S, et al. The rare earth, scandium, causes antibiotic overproduction in *Streptomyces* spp. *FEMS Microbiol Lett*, 2007, 274(2): 311–315.
- [61] Tanaka Y, Hosaka T, Ochi K. Rare earth elements activate the secondary metabolite-biosynthetic gene clusters in *Streptomyces coelicolor* A3(2). *J Antibiot (Tokyo)*, 2010, 63(8): 477–481.
- [62] Craney A, Ozimok C, Pimentel-Elardo SM, et al. Chemical perturbation of secondary metabolism demonstrates important links to primary metabolism. *Chem Biol*, 2012, 19(8): 1020–1027.
- [63] Pimentel-Elardo SM, Sørensen D, Ho L, et al. Activity-independent discovery of secondary metabolites using chemical elicitation and cheminformatic inference. *ACS Chem Biol*, 2015, 10(11): 2616–2623.
- [64] König CC, Scherlach K, Schroeckh V, et al. Bacterium induces cryptic meroterpenoid pathway in the pathogenic fungus *Aspergillus fumigatus*. *Chembiochem*, 2013, 14(8): 938–942.
- [65] Shima J, Hesketh A, Okamoto S, et al. Induction of actinorhodin production by *rpsL* (encoding ribosomal protein S12) mutations that confer streptomycin resistance in *Streptomyces lividans* and *Streptomyces coelicolor* A3(2). *J Bacteriol*, 1996, 178(24): 7276–7284.
- [66] Bibb MJ. Regulation of secondary metabolism in streptomycetes. *Curr Opin Microbiol*, 2005, 8(2): 208–215.
- [67] Srivatsan A, Wang JD. Control of bacterial transcription, translation and replication by (p)ppGpp. *Curr Opin Microbiol*, 2008, 11(2): 100–105.
- [68] Xu J, Tozawa Y, Lai C, et al. A rifampicin resistance mutation in the *rpoB* gene confers ppGpp-independent antibiotic production in *Streptomyces coelicolor* A3(2). *Mol Genet Genomics*, 2002, 268(2): 179–189.
- [69] Artsimovitch I, Patlan V, Sekine SI, et al. Structural basis for transcription regulation by alarmone ppGpp. *Cell*, 2004, 117(3): 299–310.
- [70] McKenzie NL, Thaker M, Koteva K, et al. Induction of antimicrobial activities in heterologous streptomycetes using alleles of the *Streptomyces coelicolor* gene *absA1*. *J Antibiot (Tokyo)*, 2010, 63(4): 177–182.
- [71] Zerikly M, Challis GL. Strategies for the discovery of new natural products by genome mining. *Chembiochem*, 2009, 10(4): 625–633.
- [72] Laureti L, Song LJ, Huang S, et al. Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*. *Proc Natl Acad Sci USA*, 2011, 108(15): 6258–6263.
- [73] Orth P, Schnappinger D, Hillen W, et al. Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nat Struct Biol*, 2000, 7(3): 215–219.
- [74] Xu GM, Wang J, Wang LQ, et al. “Pseudo”

- $\gamma$ -butyrolactone receptors respond to antibiotic signals to coordinate antibiotic biosynthesis. *J Biol Chem*, 2010, 285(35): 27440–27448.
- [75] Zhang YY, Zou ZZ, Niu GQ, et al. *JadR* and *jadR2* act synergistically to repress jadomycin biosynthesis. *Sci China Life Sci*, 2013, 56(7): 584–590.
- [76] Bunet R, Song LJ, van Mendes M, et al. Characterization and manipulation of the pathway-specific late regulator AlpW reveals *Streptomyces ambofaciens* as a new producer of Kinamycins. *J Bacteriol*, 2011, 193(5): 1142–1153.
- [77] Lee HN, Huang JQ, Im JH, et al. Putative TetR family transcriptional regulator SCO1712 encodes an antibiotic downregulator in *Streptomyces coelicolor*. *Appl Environ Microbiol*, 2010, 76(9): 3039–3043.
- [78] Takano E, Kinoshita H, Mersinias V, et al. A bacterial hormone (the SCB1) directly controls the expression of a pathway-specific regulatory gene in the cryptic type I polyketide biosynthetic gene cluster of *Streptomyces coelicolor*. *Mol Microbiol*, 2005, 56(2): 465–479.
- [79] Sidda JD, Song LJ, Poon V, et al. Discovery of a family of  $\gamma$ -aminobutyrate ureas via rational derepression of a silent bacterial gene cluster. *Chem Sci*, 2014, 5(1): 86–89.
- [80] Siegl T, Tokovenko B, Myronovskiy M, et al. Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes. *Metab Eng*, 2013, 19: 98–106.
- [81] Seghezzi N, Amar P, Koebmann B, et al. The construction of a library of synthetic promoters revealed some specific features of strong *Streptomyces* promoters. *Appl Microbiol Biotechnol*, 2011, 90(2): 615–623.
- [82] Liu R, Deng ZX, Liu TG. *Streptomyces* species: ideal chassis for natural product discovery and overproduction. *Metab Eng*, 2018, 50: 74–84.
- [83] Nepal KK, Wang GJ. Streptomycetes: surrogate hosts for the genetic manipulation of biosynthetic gene clusters and production of natural products. *Biotechnol Adv*, 2019, 37(1): 1–20.
- [84] Foley PL, Shuler ML. Considerations for the design and construction of a synthetic platform cell for biotechnological applications. *Biotechnol Bioeng*, 2010, 105(1): 26–36.
- [85] Umenhoffer K, Fehér T, Balikó G, et al. Reduced evolvability of *Escherichia coli* MDS42, an IS-less cellular chassis for molecular and synthetic biology applications. *Microb Cell Fact*, 2010, 9: 38.
- [86] Iwadate Y, Honda H, Sato H, et al. Oxidative stress sensitivity of engineered *Escherichia coli* cells with a reduced genome. *FEMS Microbiol Lett*, 2011, 322(1): 25–33.
- [87] Yu BJ, Sung BH, Koob MD, et al. Minimization of the *Escherichia coli* genome using a Tn5-targeted Cre/loxP excision system. *Nat Biotechnol*, 2002, 20(10): 1018–1023.
- [88] Csörgő B, Fehér T, Tímár E, et al. Low-mutation-rate, reduced-genome *Escherichia coli*: an improved host for faithful maintenance of engineered genetic constructs. *Microb Cell Fact*, 2012, 11: 11.
- [89] Couto JM, McGarrity A, Russell J, et al. The effect of metabolic stress on genome stability of a synthetic biology chassis *Escherichia coli* K12 strain. *Microb Cell Fact*, 2018, 17: 8.
- [90] Huang H, Zheng GS, Jiang WH, et al. One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*. *Acta Biochim Biophys Sin (Shanghai)*, 2015, 47(4): 231–243.
- [91] Ikeda H, Shin-Ya K, Omura S. Genome mining of the *Streptomyces avermitilis* genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters. *J Ind Microbiol Biotechnol*, 2014, 41(2): 233–250.
- [92] Komatsu M, Uchiyama T, Ōmura S, et al. Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proc Natl Acad Sci USA*, 2010, 107(6): 2646–2651.
- [93] Zweers JC, Barák I, Becher D, et al. Towards the development of *Bacillus subtilis* as a cell factory for membrane proteins and protein complexes. *Microb Cell Fact*, 2008, 7: 10.
- [94] Reuß DR, Altenbuchner J, Mäder U, et al. Large-scale reduction of the *Bacillus subtilis* genome: consequences for the transcriptional network, resource allocation, and metabolism.

- Genome Res, 2017, 27(2): 289–299.
- [95] Koo BM, Kritikos G, Farelli JD, et al. Construction and analysis of two genome-scale deletion libraries for *Bacillus subtilis*. Cell Syst, 2017, 4(3): 291–305.E7.
- [96] Belda E, van Heck RGA, José Lopez-Sanchez M, et al. The revisited genome of *Pseudomonas putida* KT2440 enlightens its value as a robust metabolic chassis. Environ Microbiol, 2016, 18(10): 3403–3424.
- [97] Nickel PI, de Lorenzo V. *Pseudomonas putida* as a functional chassis for industrial biocatalysis: from native biochemistry to trans-metabolism. Metab Eng, 2018, 50: 142–155.
- [98] Nickel PI, Chavarría M, Danchin A, et al. From dirt to industrial applications: *Pseudomonas putida* as a Synthetic Biology chassis for hosting harsh biochemical reactions. Curr Opin Chem Biol, 2016, 34: 20–29.
- [99] Komatsu M, Komatsu K, Koiwai H, et al. Engineered *Streptomyces avermitilis* host for heterologous expression of biosynthetic gene cluster for secondary metabolites. ACS Synth Biol, 2013, 2(7): 384–396.
- [100] Yamada Y, Arima S, Nagamitsu T, et al. Novel terpenes generated by heterologous expression of bacterial terpene synthase genes in an engineered *Streptomyces* host. J Antibiot (Tokyo), 2015, 68(6): 385–394.
- [101] Moya A, Gil R, Latorre A, et al. Toward minimal bacterial cells: evolution vs. design. FEMS Microbiol Rev, 2009, 33(1): 225–235.
- [102] Cello J, Paul AV, Wimmer E. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. Science, 2002, 297(5583): 1016–1018.
- [103] Kodumal SJ, Patel KG, Reid R, et al. Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. Proc Natl Acad Sci USA, 2004, 101(44): 15573–15578.
- [104] Gibson DG, Benders GA, Andrews-Pfannkoch C, et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. Science, 2008, 319(5867): 1215–1220.
- [105] Gibson DG, Benders GA, Axelrod KC, et al. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. Proc Natl Acad Sci USA, 2008, 105(51): 20404–20409.
- [106] Hutchison III CA, Chuang RY, Noskov VN, et al. Design and synthesis of a minimal bacterial genome. Science, 2016, 351(6280): aad6253.

(本文责编 陈宏宇)