

特约评述

DOI: 10.12211/2096-8280.2022-073

进化与大数据导向生物信息学在天然产物研究中的发展及应用

张凡忠^{1,2}, 相长君^{1,2,3}, 张骊驊^{1,2}

(¹ 西湖大学理学院化学系, 浙江省功能分子精准合成重点实验室, 浙江 杭州 310030; ² 浙江西湖高等研究院, 理学研究所, 浙江 杭州 310024; ³ 复旦大学化学系, 上海 200243)

摘要: 自然界亿万年的进化孕育出了丰富的天然产物资源, 进而为药物研发提供了巨大的分子宝库。进化导向生物信息学方法在微生物天然产物研究中发挥着越来越重要的作用。微生物基因组数据的快速增长为生物合成基因簇的大数据分析以及进化分析提供了新机遇, 不仅让我们对天然产物全景图有了更清晰的认识, 还能够揭示天然产物的进化规律, 利用进化分析方法和大数据资源挖掘新型的药物先导天然产物, 理解生物合成酶, 甚至设计改造生物合成体系创造非天然分子。本文综述了近年来进化和大数据导向生物信息学应用于天然产物研究中的相关进展, 强调了进化与大数据在生物合成酶的功能预测、进化机理、基因挖掘以及生物合成改造方面的应用, 最后分析了目前面临的问题并对未来发展趋势进行了展望。

关键词: 天然产物; 进化; 基因挖掘; 生物合成改造; 生物信息学

中图分类号: Q31; Q819 **文献标志码:** A

Advances and applications of evolutionary analysis and big-data guided bioinformatics in natural product research

ZHANG Fanzhong^{1,2}, XIANG Changjun^{1,2,3}, ZHANG Lihan^{1,2}

(¹Key Laboratory of Precise Synthesis of Functional Molecules of Zhejiang Province, Department of Chemistry, School of Science, Westlake University, Hangzhou 310030, Zhejiang, China; ²Institute of Natural Sciences, Westlake Institute for Advanced Study, Hangzhou 310024, Zhejiang, China; ³Department of Chemistry, Fudan University, Shanghai 200243, China)

Abstract: Nature has invented a myriad of natural products through billions of years of evolution. Natural products own unique structural features selected by evolutionary pressure and serve as a treasure trove for drug discovery. The rapid growth of microbial genomic data now provides new opportunities for evolutionary and big data analysis of biosynthetic gene clusters, which not only gives us a clearer picture about the global landscape of natural products, but also enables us to reveal the evolutionary trajectory of natural products. Such holistic understanding of natural products

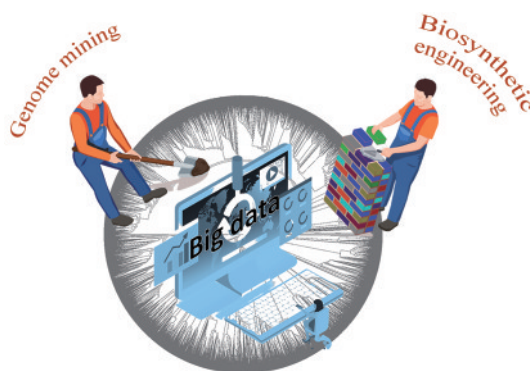
收稿日期: 2022-12-08 修回日期: 2023-02-21

基金项目: 国家自然科学基金 (22177092, 82003631, 82104026); 浙江省领军型创新创业团队 (2020R01004); 杭州市科技发展计划 (20201203B122)

引用本文: 张凡忠, 相长君, 张骊驊. 进化与大数据导向生物信息学在天然产物研究中的发展及应用[J]. 合成生物学, 2023, 4(4): 629-650

Citation: ZHANG Fanzhong, XIANG Changjun, ZHANG Lihan. Advances and applications of evolutionary analysis and big-data guided bioinformatics in natural product research[J]. Synthetic Biology Journal, 2023, 4(4): 629-650

can facilitate the phylogeny-guided genome mining, allow better functional prediction of biosynthetic enzymes, and even open the door to biosynthetic redesign to create non-natural molecules by evolution-guided engineering. The core essence of evolutionary and large-scale bioinformatics lies in that it visualizes the entire sequence space and their distribution of a particular analyte family. Therefore, big data-driven bioinformatics has the potential to answer some challenging questions such as “How many natural products remain to be discovered?”, and “How long can natural products discovery be sustainable?” This review summarizes recent advances in the application of evolution and big data-guided bioinformatics for natural products research from several perspectives including: ① natural product discovery; ② functional and structural prediction of biosynthetic enzymes and their products; ③ bioengineering, with the emphasis on the assembly line enzymes such as polyketide synthases and non-ribosomal peptide synthetases. Due to the modular domain architecture they have, the assembly line enzymes have been the main targets for genome mining. The phylogenetic analysis of their domains has shown to be a powerful and effective way to predict their enzymatic function and substrate specificity. Recently, the evolutionary mechanism of the assembly line enzymes has been investigated, and several evolution-guided engineering strategies were shown to have much higher efficiency for the assembly line reprogramming, providing a potential breakthrough for the bioproduction of complex polyketides and peptides. Non-modular enzymes are also discussed with selected representative examples. Finally, we present current challenges and future prospects of big data-driven natural products research.



Keywords: natural products; evolution; genome mining; biosynthetic engineering; bioinformatics

进化是促进生命出现、发展和多样化的过程^[1]。生命的进化本质上是基因信息的进化，而天然产物生物合成相关的酶由基因信息编码，因此也离不开进化的力量^[1-3]。植物和微生物为适应自然环境创造出许多天然产物。过去一个多世纪里，天然产物作为先导分子在医药健康和农业生产中发挥着重要作用，包括已被用作药物造福人类的青霉素、红霉素以及万古霉素等。生物信息学的预测分析表明，迄今为止，只有3%的细菌来源天然产物被发现，即使是链霉菌这样被高度研究的类群仍然含有许多未知的天然产物^[4]。

基因测序技术的发展带来基因组数据的快速

增长(图1)，大规模基因组学、代谢组学和功能研究的系统数据即“大数据”与生物信息学的结合为天然产物研究带来了“技术革命”。传统的天然产物研究高度依赖于化合物分离纯化，过去我们只能通过积累分离得到的单体化合物来了解天然产物。如今，我们已处在向天然产物全景图可视化过渡的阶段。基于大数据和生物信息学的现代天然产物研究使我们对天然产物的分子多样性、丰度以及分布有了宏观认识，不仅让我们了解到微生物天然产物库的巨大挖掘潜力，也可以指导具有临床或商业价值的新分子的发现，提高天然产物发现的效率^[1-2, 5-8]。由于已测序的真菌基因组

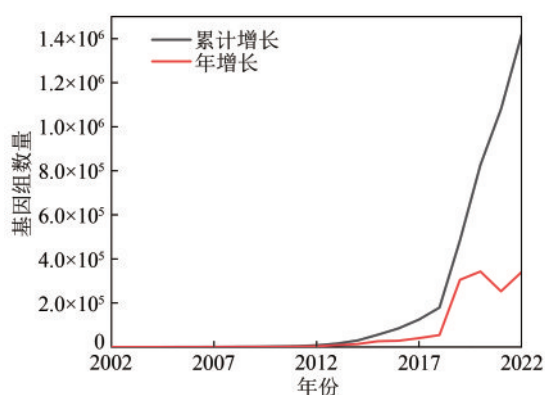


图1 2002—2022年（近20年）NCBI数据库中细菌基因组数量的增长趋势

Fig. 1 Growing trends for the number of bacterial genomes in the NCBI database from 2002 to 2022 (in the last two decades)

数量相比细菌要少很多（截至2023年1月，NCBI数据库中细菌基因组数为1 420 776，而真菌为28 183，只占细菌基因组数量的2%），进行大数据分析主要以细菌为主，因此本文主要阐述细菌天然产物研究，但也包括了部分真菌研究。

对天然产物生物合成机理及相关酶的生化特性的认识促进了进化分析在酶功能预测方面的应用，进而指导酶的改造以及生物合成途径的改造^[6, 9]。目前基于进化进行的天然产物研究主要集中于以下几个方面：①利用进化导向方法发现新的天然产物（化合物结构预测）；②通过进化分析预测酶的功能；③通过改造生物合成体系创造出我们需要的产物。因此，本文将聚焦于基于大数

据的进化导向生物信息学方法在天然产物发现和酶工程研究中的运用进展，并对这些工具、方法在天然产物研究领域的发展进行展望。

1 基于进化和大数据的天然产物研究策略

基因挖掘是在没有化学结构的前提下，基于遗传信息来预测和分离活性天然产物^[10-13]。微生物基因组挖掘方法振兴了抗生素的研究，但这些方法依赖于先前识别的生物合成酶的序列相似性搜索，这种经验性质限制了所探索的化学空间。近年来，天然产物研究人员通过将进化原理融入基因组分析来寻找新的路径^[14-15]。利用进化和大数据进行的天然产物研究一方面是基于系统发育距离预测功能相似度：当靶标蛋白序列与已知化合物的编码序列距离较远且形成不同进化分支时，倾向于产生具有新核心结构的新产物；当靶标序列与已知化合物编码序列相邻时，可能会产生与已知化合物变化不大的新产物 [图2(a)]^[10, 15]。另一方面可以依据目标序列是否与已知序列聚集在一起，或属于一个新的分支，或是一个罕见的异常值，实现分布模式和多样性的可视化，进而展示一个全景图 [图2(b)]。

已有不少基于进化的生物信息学工具可用于天然产物的挖掘，例如ARTS (Antibiotic Resistant

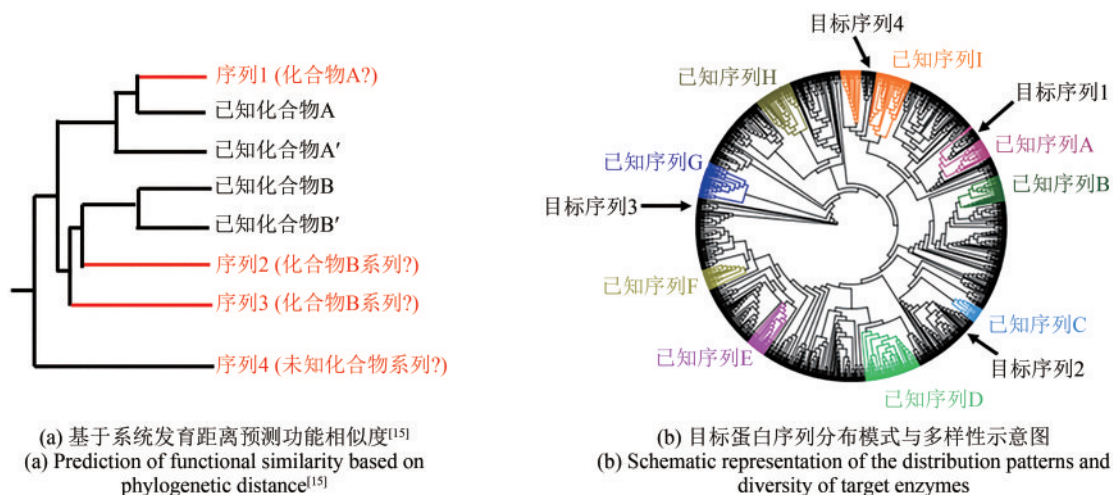


图2 进化分析的核心观念

Fig. 2 A core concept of the phylogenetic analysis

Target Seeker)^[16-17]、NaPDoS (Natural Product Domain Seeker)/NaPDoS2^[18-19]、EvoMining^[20]、Big-SCAPE 和 CORASON^[21-22]等。ARTS 和 EvoMining 面向进化上相关的基因组, 针对保守的生物合成基因簇 (biosynthetic gene cluster, BGC) 进行预测和聚类分析。其中 ARTS 以抗性基因为靶标, 通过将管家基因、已知抗性基因与邻近 BGC 联系起来, 自动筛选序列数据, 挖掘作用方式新颖的抗生素, 还可以用于比较相似的 BGC 及其假定的抗性基因; 而 Evomining 基于酶的基因复制和底物特异性扩展的进化方式开发了一种基因挖掘方法, 可以检测某些类型的管家基因的同源基因, 并比较每个酶家族的平均数量和系统发育距离, 因此可以用于直观地展示天然产物生物合成酶的起源与进化方向。NaPDoS 是将 PCR 产物、基因组或宏基因组数据快速提取和分组, 分析目标 KS (ketosynthase) 或者 C (condensation) 结构域在进化树上的位置, 从细菌遗传数据中推断次生代谢物的新颖性和潜力。Big-SCAPE 是面向进化信息未知的多个基因组, 以 MIBiG 数据库^[23]的基因簇作为参考分析 antiSMASH^[24] 预测的基因簇, 构建序列相似性网络, 并把这些基因簇分为不同的基因簇家族, 然后用 CORASON 解释每个基因簇家族中不同基因簇的进化关系。这些生物信息学的工具同是基于进化原理, 但是可以针对不同类型的基因以及用于实现不同的目的。

目前利用进化方法研究得最好的一类天然产物合成酶是模块型酶, 如聚酮合酶 (polyketide synthase, PKS) 和非核糖体肽合成酶 (non-ribosomal peptide synthetase, NRPS)。根据目标基因簇类型, 我们对进化导向研究的进展进行了分类讨论, 主要包括 PKS、NRPS 以及其他非模块型酶。

2 进化和大数据导向的 PKS 研究

聚酮化合物是一大类具有广泛结构和功能的生物活性天然产物, 临床上使用的多种药物都属于聚酮类, 例如红霉素、阿维菌素、四环素等。聚酮化合物由聚酮合酶 (PKS) 负责生物合成。目前已知的细菌的 PKS 有 3 类 (I 型、II 型和 III 型)。

其中, I 型 PKS 由多个模块形成组装线, 每个模块都包含核心结构域 KS、AT (acyltransferase)、ACP (acyl carrier protein) 协同催化聚酮链延伸的一个循环, 部分模块含有 KR (ketoreductase)、DH (dehydratase)、ER (enoylreductase) 等结构域对聚酮进行不同程度的修饰^[25] [图 3(a)]。II 型 PKS 通过聚酮合酶 (KS/KS_α) 和链长因子 (CLF/KS_β) 催化乙酸单元的迭代缩合反应, 随后通过还原、环化和脱水反应形成多环芳香族骨架 [图 3(b)]^[26]。III 型 PKS, 也称查尔酮合成酶样 PKS, 属于同型二聚体酶, 本质上是迭代作用的缩合酶^[27-28]。许多研究者尝试从进化的角度去理解自然界中 PKS 基因与聚酮化合物结构之间的关系, 主要是由于: 首先, 对天然 PKS 多样性的探索有可能发现新的生物活性聚酮; 其次, PKS 的多模块结构也为研究多个同源但功能不同的蛋白进化提供了一个独特的例子; 最后, 更好地理解自然界聚酮多样化形成机制可以为 PKS 改造开辟新的途径^[29]。

2.1 PKS 的基因进化机制

通过对 PKS 不同结构域 (KS、AT 和 KR 等) 的进化分析, 目前认为导致 PKS 装配线多样化的进化过程主要包括基因复制、水平基因转移、基因转换以及重组 [图 4(a)]^[29]。I 型 PKS 的模块化几乎都来源于单个祖先模块的多个副本^[30]。重复模块为基因重组提供了理想的平台, 可以导致产物化学结构的相应变化^[30]。除了基因重组, 基因进化过程中 DNA 序列可能从一个同源区域非交互地转移到另一个同源区域, 从而使这些同源序列同质化, 即发生基因转换。基因转换广泛存在于 I 型 PKS 中^[31]。对 PKS 的这些“自然重编程”事件的分析可能有助于开发生物活性化合物的生物组合设计^[30]。

cis-AT PKS 的进化被认为主要通过模块复制以及整个组装线的水平或垂直获取^[32]。相反, *trans*-AT 系统具有明显的重组和嵌合形成新的基因簇的趋势^[33]。进化树上相近的 *trans*-AT PKS 的 KS 往往催化相似结构的底物^[33], 而对于 *cis*-AT PKS, 同一个基因簇的 KS 结构域往往具有高序列相似性^[34]。Ikuro Abe 等通过对 4 个氨基多醇 (neomediomycin

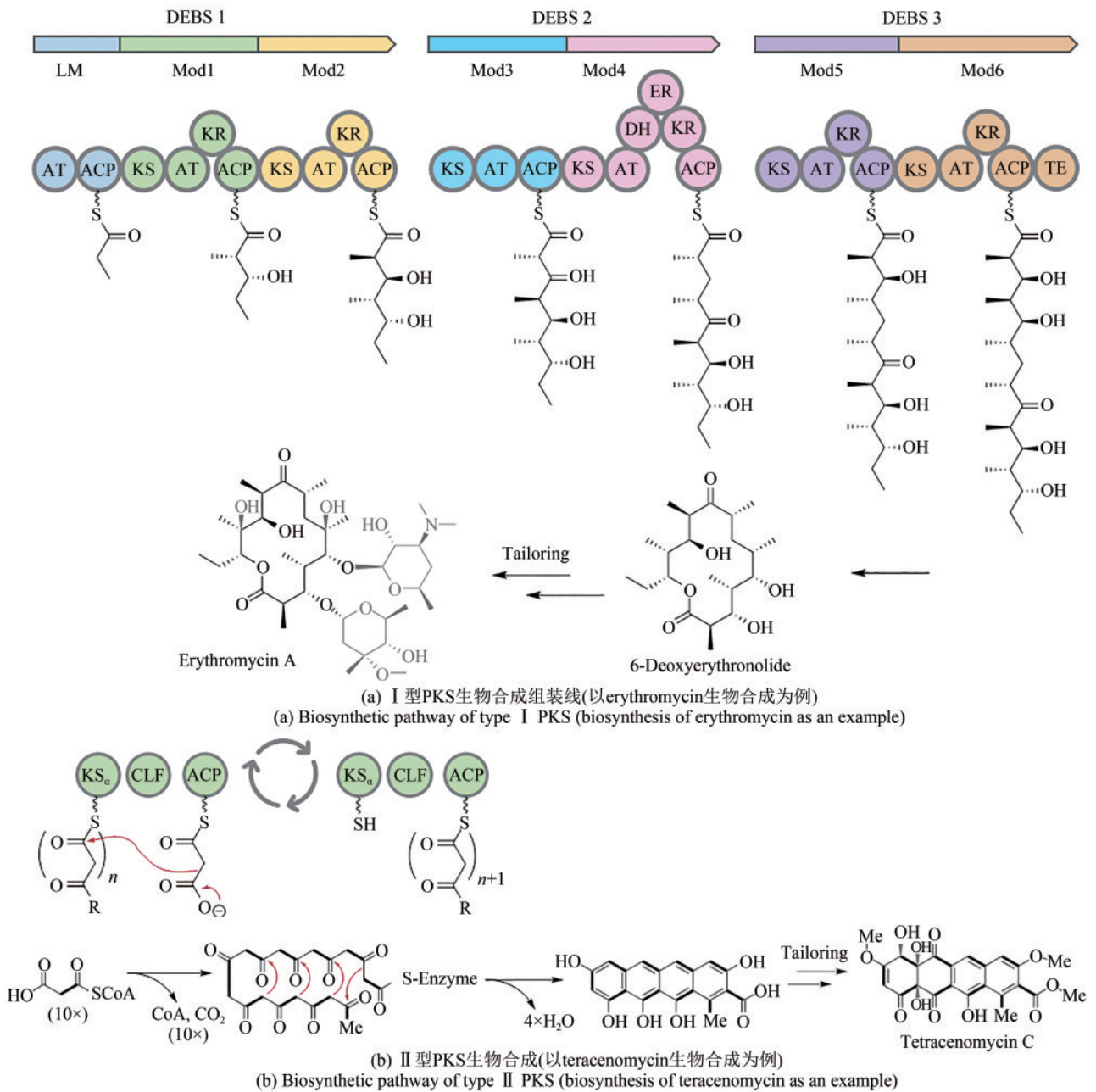


图3 PKS生物合成

Fig. 3 Biosynthetic pathway of PKS

B、mediomycin、ECO-02301、tetrafibricin) 的基因簇分析发现KS结构域表现出与上游模块的ACP更近的进化关系, 这表明 *cis*-AT PKS 中KS与上游模块的ACP及修饰结构域作为整个单元进行重组, 自然的重组发生在KS与AT或AT与修饰结构域之间^[35-36]。Adrian Keatinge-Clay 随后的分析表明 *trans*-AT PKS 的KS结构域也表现出与上游模块的ACP更近的进化关系^[37]。这与KS门控角色相符,

因为KS底物的结构由上游模块的组成所决定。这也形成了装配线PKS模块的重新定义AT-(DH-KR-ER)-ACP-KS, 而非传统的KS-AT-(DH-KR-ER)-ACP [图4(b)]^[36]。

不同于KS, KR结构域在系统发育树上基于产物羟基的立体构型分类^[38], 在 *trans*-AT PKS 中还依据其他修饰结构域的存在进行分类^[39]。一种观点认为这是由于基因转换不需要影响PKS模块的

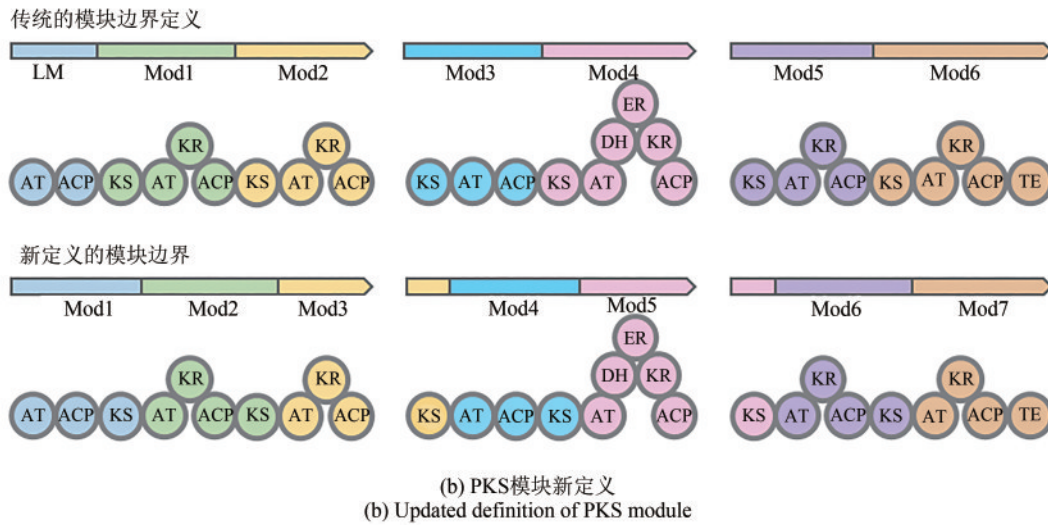
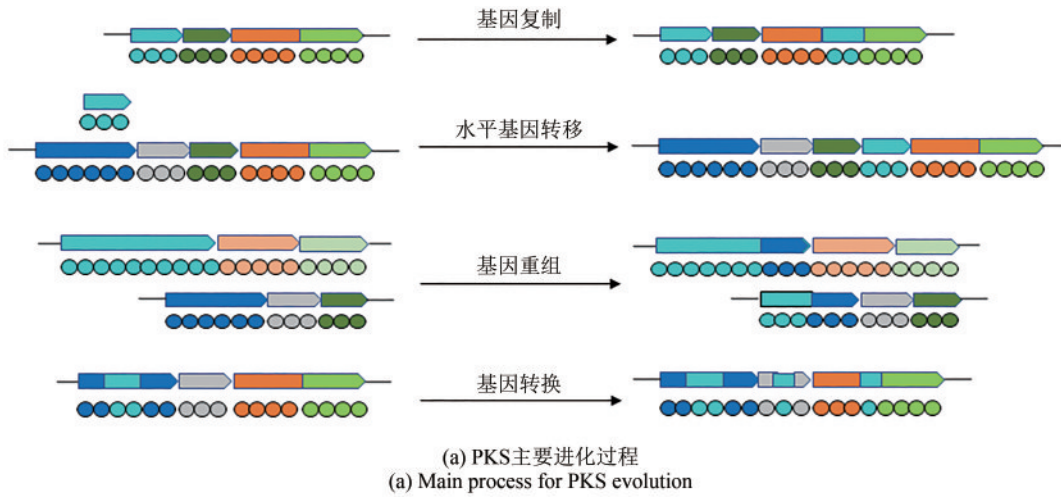


图4 PKS进化机理

Fig. 4 Evolutionary mechanism of PKS

所有结构域，所以修饰结构域并不总是与KS共进化^[29]。这也就意味着，在模块内部同样存在着结构域自然重组的位点。我们注意到，导致遗传序列同质化的协同进化在PKS这样的重复遗传区域经常发生^[40]，而这会掩盖进化轨迹。因此，要阐明PKS的确切进化过程还需要进行仔细分析。

II型PKS基因簇的核心KS和CLF基因被认为来源于一个古老的KS复制^[41]，它们可能在放线菌诞生之前就从FabF（脂肪酸途径中的KS同源蛋白）的共同祖先分化而来，然后在很少甚至没有基因交换的情况下共同进化^[42]。与KS树相比，CLF有更清晰的分支结构，且CLF的进化分支与聚酮结构单元数目（而非总碳数）的关系更紧密^[5]。除了KS和CLF基因，II型PKS基因簇中的

聚酮还原酶与环化酶被认为是从其他系统交换到基因簇中，随后进化出PKS特异性功能^[42]。

2.2 PKS的功能预测

进化分析最重要的应用之一是区分旁系同源（由重复产生）和直系同源（由物种形成产生）的基因或蛋白。一般情况下，同源与相同的功能相关^[43]。因此能够利用系统基因组方法从进化亲缘来预测序列功能，即对功能未知的基因根据它们相对于已知基因的系统发育位置进行功能预测。

对PKS的功能预测目前主要针对KS结构域，I型*trans*-AT PKS的KS结构域^[30, 32]、II型PKS的CLF结构域分支^[5, 42]等均能形成与其底物的化学结构紧密相关的进化分支，因此可用于化合物结

构预测以及同工酶的预测。KS结构域序列不仅可以用来区分聚酮与脂肪酸、烯二炔和多不饱和脂肪酸，而且还可以用来区分不同类型的聚酮，如 *cis*-AT PKS 和 *trans*-AT PKS、PKS/NRPS 杂合型，NaPDoS 即是利用这一原理进行酶功能和化合物结构预测^[18-19]。此外，II型PKS中修饰酶的功能也可以通过进化分析预测，如KR的区域选择性和环化酶的环化方式^[42, 44]。

除了根据已知功能的酶来推测同家族相似酶的底物和功能外，还可以依据基因的共同进化来研究未知酶。BGC的发展经历了不同的进化过程，如基因组内的复制、重排、域/模块/基因交换和水平基因转移^[30, 40]。同一个簇中那些相互作用的酶需要共同进化的过程，以保持适当的相互作用^[6]。因此还可以根据与之有相互作用的酶来预测未知酶的功能。Jörn Piel 课题组^[45]通过对 *trans*-AT PKS 中KS的进化分析，发现了一个包含TE_B类（负责O-乙酰化）结构域的KS分支。根据产物结构及缺失的HGTGT活性位点，推测这些KS均为非延伸的KS₀，尽管催化的聚酮结构不同，但TE_B模块具有生物化学上的一致性。

细菌I型PKS中除了KS结构域可以形成与底物化学结构密切相关的进化分支，AT结构域在进化树上也形成两个主要分支，分别具有接收丙二酰-CoA和甲基丙二酰-CoA的特异性^[46]。AT结构域对于丙二酰-CoA和甲基丙二酰-CoA的特异性识别可以通过序列中两个特征区域预测，HAFH和GHS(I/V)G序列表明其接收丙二酰-CoA，而YASH和GHSQG表明其接收甲基丙二酰-CoA^[47-49]，这一发现已长期被用于区分这些AT的底物选择性。

2.3 PKS的基因挖掘

KS结构域的进化分析用于化合物结构预测以及同工酶的预测是PKS基因挖掘的理论基础。对芳香聚酮BGC中的KS_α和CLF（即KS_β）进行的进化分析表明，KS_α和CLF的进化树结构和分支模式都很相似，而且以对应聚酮化合物的链长聚类，因此KS_α和CLF可作为理想的进化标签代表整个基因簇^[42]。Sean Brady 课题组^[50-51]以CLF序列作为

进化标签，从土壤微生物组中扩增相关基因，与已知CLF基因做序列比对并进行进化分析，发现了许多与已知序列在同一分支的不同亚支的序列，通过将对应的基因簇在菌株 *Streptomyces albus* J1074中异源表达发现了结构新颖且活性显著的多酚和蒽环霉素类化合物（1~3，结构见图5）。最近，本课题组^[5]在II型PKS中应用全局基因组挖掘，从一个与已知基因簇距离较远的进化分支上发现了oryzanaphthopyrans（4，结构见图5），此外，也基于大数据的进化分析描绘了II型PKS的分布、丰度以及多样性的全景图。戈惠明组^[52]利用CLF的进化模式结合抗性基因靶向挖掘四环素类化合物，发现了高度糖基化的四环素海南霉素（5，结构见图5）。这些研究或不需要培养微生物或在微生物培养前已能够预测化合物的结构新颖性和生物活性水平，体现了进化导向基因挖掘的优势。

在I型PKS中，*trans*-AT PKS的KS也能形成与其底物的化学结构紧密相关的进化分支^[53]。Jörn Piel 课题组^[54]利用KS数据库检索发现了与mis PKS（misakinolides PKS）进化上非常接近的序列，最终鉴定其产物为二聚的大环内酯luminaolide B（6，结构见图5）。通过研究misakinolides、scytophyacin以及luminaolides的生物合成和进化关系，发现它们的基因簇来源于共同祖先，通过缺失或获得PKS上游或末端序列实现结构多样化。为了实现*trans*-AT PKS产物的结构预测，理解*trans*-AT PKS的生物合成基础和进化模式，Jörn Piel 和 Marnix Medema 课题组^[55]开发了在线工具transATor和transPACT。TransATor输入PKS序列，预测KS底物特异性及对应的聚酮核心结构，利用这一工具他们发现了tartrolon类化合物和leptolyngbyalide。TransPACT是*trans*-AT PKS注释和比较工具，可以自动形成KS的功能分支，识别不同PKS组装链共有的连续模块。他们利用transPACT从GenBank中得到1782个*trans*-AT PKS的基因簇并用antiSMASH进行分析，随后提取KS序列进行进化分析，根据计算生成的模块共享网络和进化树进行基因挖掘，最终发现了新的*trans*-AT PKS产物secimide（7）、gynuellalide（8）、spliceostatin L（9）等（结构见图5），并探究了相似化学结构在序列上的相关

性^[56]。这些研究表明了可以通过对 *trans*-AT PKS 中 KS 结构域的进化分析来指导结构新颖的聚酮化合物的挖掘，同时也为 PKS 改造以生产非天然的 *trans*-AT PKS 聚酮产物提供了基础。辛志宏课题组^[57]利用 KS 序列的进化分析挖掘植物内生真菌中 I 型 PKS 产物，发现了具有抗菌活性的天然色素 talafun (10, 结构见图5)。该研究表明使用高度保守的 KS 结构域作为进化标记，可以快速连接真菌基因信息和化学结构，并作为高通量测序技术的常规方法应用于实践。

除了 PKS 本身，同一个簇中的共同进化基因同样可以用于基因挖掘。烯二炔是一类由 I 型 PKS 产生的线性多烯，具有极高活性，常作为临床试验中的抗体药物配合物^[58]。Shen Ben 课题组^[58-59]以两组不同的烯二炔生物合成基因 E5/E 和 E/E10 为靶标通过实时定量 PCR 从 3400 株菌株中挖掘烯二炔化合物。通过 PCR 他们发现了 81 株具有烯二炔聚酮合酶基因的菌，同时对基因 E 的进化分析表明许多簇与已知的是不同的。为确认这一结果，

他们对 31 株代表性菌株进行了基因组测序，对其中的相关基因簇进行 GNN (genome neighborhood network) 分析，发现了与已知基因簇明显不同的基因簇，最终通过分离鉴定发现了活性化合物 tiancimycin A (11, 结构见图5)^[58-59]。以上研究为挖掘更多烯二炔类化合物或利用 PKS 合成烯二炔同系物奠定了一定基础。

2.4 PKS 生物合成改造

事实上，自从三十多年前 I 型 PKS 的多模块化特征被确定，研究人员就尝试利用模块和结构域的重新组合来产生新的非天然聚酮化合物^[60]。由于其多模块的特性，PKS 提供了一个通用的合成平台，例如可以作为合成特定有机酸的有效方法^[61]。然而，在对 PKS 装配线进行改造的一些早期尝试中，交换或删减一些结构域和模块经常导致酶活性显著降低甚至是无活性^[52]，推测这与蛋白相互作用^[53]以及底物选择

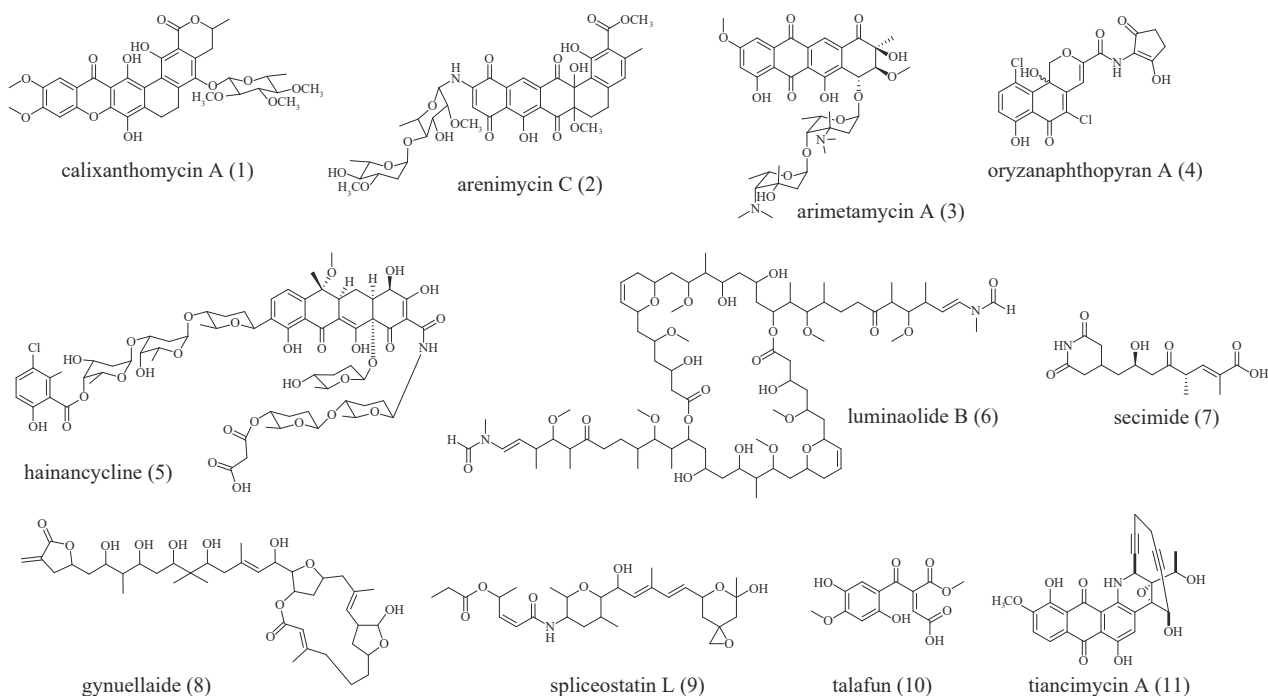


图5 进化导向基因挖掘获得的聚酮化合物1~11分子结构

化合物1~5为II型PKS基因挖掘的芳香聚酮；化合物6~9为*trans*-AT PKS基因挖掘产物；

化合物10为真菌I型PKS基因挖掘产物；化合物11为烯二炔类聚酮化合物

Fig. 5 The structure of polyketides molecules 1 to 10 obtained by phylogeny-guided genome mining

Compounds 1 to 5 were aromatic polyketides discovered by genome mining of type II PKS. Compounds 6 to 9 were discovered by genome mining of *trans*-AT PKS. Compound 10 was discovered by genome mining of fungal type I PKS. Compounds 11 were enediynes

性^[54]有关。

同时,越来越多的证据表明,对组装线系统的进化有更好的理解可以进一步提高改造这些系统的能力[图6(a)]^[29, 35, 62]。Adrian Keatinge-Clay课题组^[63-65]根据新定义的模块边界,即在KS和AT之间选择切点构建了多个杂合PKS,目标产物的产量得到明显的增高(相比于根据传统定义构建的杂合PKS有10~48倍的提高)[图6(b)]。Christian Hertweck课题组^[66]分析几条聚酮合成基因簇中各个模块KS的系统发育树,通过几种不同方式的剪切融合,证明在自然进化过程中模块的增加或删减可能发生在KS-AT连接处,并通过分析基因簇中P450修饰酶的底物特异性,进一步推测出PKS的进化顺序。同样地,除了上述KS-AT连接处,post-AT连接处同样被证明为有效的模块融合和结构域交换切点^[67-69]。类似的利用自然偏好的位点进行“剪切-粘贴”的策略可能也适用于*trans*-AT PKS的改造^[54]。

Jay Keasling课题组^[69-72]通过序列比对,把KR或KR-DH-ER整体在模块间进行交换,体外和体

内实验证明了这种策略的可行性[图6(c)],这也进一步表明还原结构域可能为进化过程中潜在的重组单元。另一方面,一些研究通过尝试对AT^[73]和KR^[74]活性口袋进行多点突变以逆转其选择性,然而当这些突变被用于整个模块,仍无法专一地得到预期产物,可以推测PKS并不是单纯依赖点突变的方式,而是通过基因重组导致的结构域交换去改变结构域的选择性^[29]。

虽然PKS单个或多个结构域甚至整个模块的结构已经通过X射线单晶衍射和低温冷冻电镜技术解析^[75-81],揭示了蛋白质相互作用在聚酮链延伸各个阶段的重要性以及一些重要的蛋白相互作用位点,但PKS催化过程中各个结构域潜在的协同作用未被完全揭示,基于三维结构的理性设计改造仍困难重重。因此,从自然进化的角度分析整个PKS或各个结构域的进化关系,推测自然重组发生的位点,以此作为人工改造的切点,并根据进化关系选择合适的候选PKS进行拼接构建杂合PKS,这为研究者提供了进化导向的PKS改造新思路。

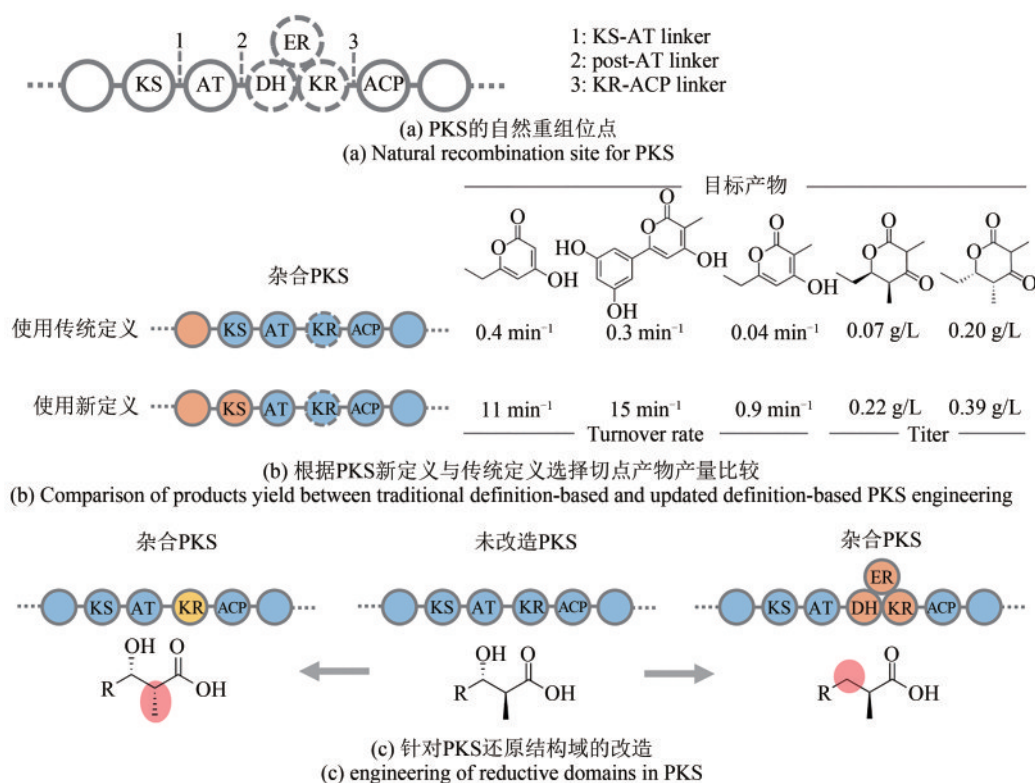


图6 进化导向的PKS改造

Fig. 6 Evolution-guided engineering for PKS

3 进化和大数据导向的NRPS研究

非核糖体肽合成酶 (NRPS) 是来自细菌和真菌的多模块酶或酶复合体, 其催化生成的肽类化合物很多具有重要的生物活性, 其中一些被临床使用如环孢素、万古霉素、达托霉素等^[82-84]。根据合成酶整体结构的不同, NRPS通常被分为 I 型和 II 型^[85]。I 型 NRPS 是大型模块化复合体, 以类似于 I 型 PKS 的流水线方式生成肽类化合物, 每个模块主要包含 C (condensation)、A (adenylation)、T (thiolation, 也称载体蛋白) 3 个结构域或其他修饰型结构域如 E (epimerization)。II 型 NRPS 蛋白通常是独立的酶或两个结构域协同形成独特的氨基酸衍生物^[85]。NRPS 合成肽类化合物过程中, A 结构域选择特定的氨基酸单体, 由 ATP 激活形成氨酰基-AMP, 然后转移到载体蛋白 T 上。C 结构域缩合被激活的氨酰基 (肽基) 硫酯, 通过形成酰胺键进行链延伸。同 PKS 一样, NRPS 对挖掘活性分子、研究酶催化和蛋白相互作用等具有重要意义。

3.1 NRPS 进化机理

与 PKS 类似, NRPS 的进化过程中自然发生的基因重组发挥了重要作用。非核糖体肽的分化主要由 A 结构域或子结构域的重组驱动^[86]。A 结构域内的重组发生在 A_{core} 的可变部分从而调节底物, 但结构域之间的相互作用以及 A_{sub} 基本不受影响^[62, 87]。

NRPS 的另一核心结构域是 C 结构域, 根据立体选择性的不同可以分为¹C_L、^DC_L和起始 C 结构域 (C_S或 starterC) 3 种类型。尽管¹C_L和起始 C 结构域由于某些序列的显著差异导致了底物的差异 (氨基和 β-羟基羧酸), 它们在进化树上的关系似乎比其他亚型更密切^[88]。已有研究表明 C 结构域对立体化学的选择与 E 结构域功能相关^[89]。在细菌 NRPS 中 PCPE-E-^DC_L几乎是普遍保守的, 表明尽管在进化史上发生了无数的基因组复制、插入、删除和重组事件, E-^DC_L连接域仍保持着强大的选择压力^[89]。

3.2 NRPS 功能预测

根据 NRPS 的进化机理, NRPS 组装线上的结构域序列与产物的化学结构之间有直接的关系, 这种关系使从 DNA 序列预测肽类化合物的化学结构成为可能。

1991 年, Stachelhaus 和 Marahiel 报道了 A 结构域底物特异性预测的开创性工作。他们关注到 A 结构域的系统进化树与底物类别有较强的关联性 (图 7), 其次发现在 A_{core} 区域形成底物结合口袋的 10 个关键氨基酸序列 (即 Stachelhaus 密码), 与它们接收的底物具有高度相关性^[90]。随后, 一些 NRPS A 结构域的底物预测工具被开发利用, 如 NRPSpredictor2^[91] 和 SANDPUMA^[92] 等。这些预测工具的发展一方面为基因挖掘发现新型天然产物提供巨大帮助, 另一方面也可为 NRPS 装配线改造寻找合适的候选基因。

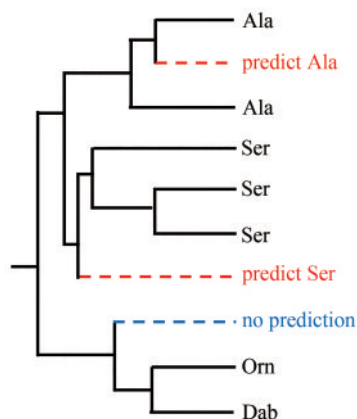


图7 进化导向方法预测 A 结构域底物示意图
Fig. 7 Prediction for substrates of A domain using phylogeny-guided method

NRPS 的 A 结构域相当于“底物水平”的进化信号, 可以用来预测底物特异性, 而 NRPS 的 C 结构域相当于“途径水平”的进化信号, 可以用来预测类似分子的 BGC 模式^[93]。C 结构域超家族除了原始 C 结构域, 还包括其他几个同样属于 NRPS 结构域的成员如 C_S、^DC_L、¹C_L、E (差向异构化)、Cyc (杂环化)、^{Dual}C (差向异构化/缩合) 和^{modAA}C (脱氢氨基酸相关), 它们由于功能不同在进化树上形成了明显分支 (图 8)^[89, 94], 因此可以利用 C 结构域的进化分析进行 NRPS 中相关结构域的功能预测。

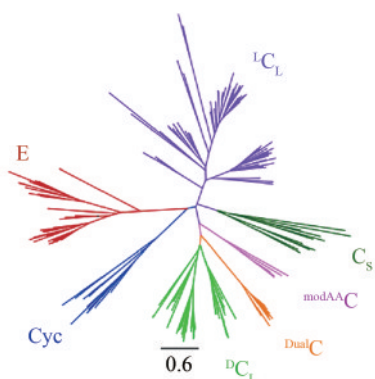


图8 C结构域超家族的无根进化树^[89]

Fig. 8 Unrooted phylogenetic tree of the C-domain superfamily^[89]

3.3 NRPS基因挖掘

NRPS中A结构域和C结构域的底物选择性以及不同模块的催化顺序决定了氨基酸的连接顺序,意味着非核糖体肽类化合物的结构可以与NRPS序列直接关联起来,因此可以利用A结构域或C结构域的进化分析进行NRPS的基因挖掘。

钙离子依赖型抗生素是一类需要钙离子才能发挥活性的环肽。已知的这类化合物都有保守的

Asp⁴-X-Asp⁶-Gly⁷片段促进其与钙离子的结合。基于此, Sean Brady课题组^[95]通过PCR从土壤eDNA中扩增NRPS的A结构域,并对扩增得到的序列用eSNaPD进行分析。标签序列的进化树中Asp⁴结构域有许多进化分支与已知BGC距离较远,这暗示土壤微生物组中有未知的钙离子依赖抗生素存在。通过异源表达,他们鉴定出一类新的钙离子依赖抗生素malacidins(12,结构见图9)。虽然基于宏基因组的抗生素发现方法仍处于起步阶段,但是上述研究中规模化 and 自动化方法为高效挖掘隐藏在宏基因组中的抗生素,以及为对抗抗生素耐药性提供了一种潜在的强大方法。

2020年, Gerard Wright课题组^[96]收集了71个糖肽类抗生素的基因簇,利用基因簇中的C结构域构建进化树预测出可能有新的生物活性的糖肽类化合物,最终发现两个具有新的功能的糖肽类抗生素corbomycin(13)和complestatin(14)(结构见图9),并阐明它们是通过与肽聚糖结合,阻断了自溶素(自溶素是生长过程中细胞壁重构所必需的肽聚糖水解酶)作用从而抑制细菌生长。近期, Gerard Wright课题组^[97]利用糖肽抗生素指纹

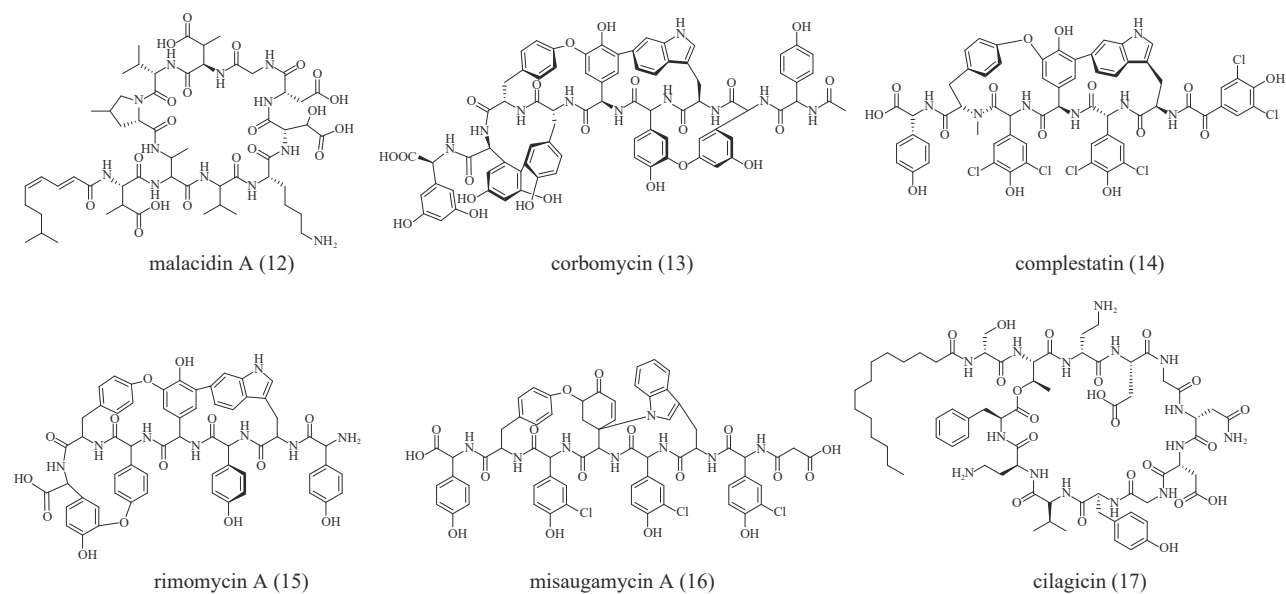


图9 进化导向基因挖掘获得的肽类化合物12~17分子结构

(化合物12为基于NRPS的A结构域基因挖掘发现的钙离子依赖抗生素;化合物13~16为基于NRPS的C结构域基因挖掘发现的糖肽类抗生素;化合物17为基于NRPS预测发现的脂肽类化合物)

Fig. 9 The structure of peptide molecules 12~17 obtained by phylogeny-guided genome mining

(Compounds 12 was calcium-dependent antibiotic discovered by genome mining of A domain; compounds 13~16 belonged to glycopeptide family of antibiotics discovered by genome mining of C domain; compound 17 was lipopeptide antibiotic discovered by NRPS prediction)

序列扩充候选 BGC 后, 通过进化导向基因挖掘及异源表达鉴定了 5 个新的 type V 糖肽类抗生素 (type V GPA) rimomycins (15) 和 misaugamycins (16) (结构见图 9), 并证明它们的作用机理也是通过阻止自溶素活性而抑制细胞分裂。这些发现拓展了 type V GPA 的化学多样性, 为药物开发提供了新的化学骨架, 展示了基于进化的生物信息学平台在挖掘糖肽类抗生素化学“暗物质”中的巨大应用潜力。

NRPS 的准确预测也促进了不依赖于传统分离技术的生物活性肽的开发。Sean Brady 课题组^[98]利用对脂肽的生物信息学预测, 通过化学合成获得了具有较强抗菌活性的脂肽 cilagicin (17, 结构见图 9), 并阐明它是通过阻断细胞壁生物合成中两种必需的十一烯基磷酸发挥抗菌作用。该研究基于 C₅ 结构域的进化树分析, 发现了孤儿 BGC, 然后将化合物结构预测与化学合成结合得到对应产物, 避免了目标基因簇不表达或产物产量低等问题。

3.4 NRPS 及 NRPS-PKS 杂合的生物合成改造

3.4.1 NRPS

目前针对 NRPS 的生物合成改造已有不少尝试, 主要可以分为以下几类: ①取代 A 结构域或 A-T 结构域, 从而改变延伸单元^[99]; ②改变 A 结构域的底物结合口袋^[100-101]; ③C-A 或者 C-A-T 结构域交换^[102]。

通过 A 结构域取代, David Ackerley 课题组^[86]高效地获得了高产量的 pyoverdine 修饰肽, 确定了允许的 A 结构域重组边界 [图 10(a)]。Jörn Piel 课题组^[103]对 hormaomycin 合成基因簇中 7 个 A 结构域核苷酸序列分析发现, 除了与底物识别口袋有关的约 400 个碱基对, 其余序列展现出了超过 90% 的相似度, 这暗示着潜在的自然重组位点。基于这种猜测, 利用从自然重组推断的序列边界, 以 HrmO 第 3 个 A 结构域作为模板构建了 3 个嵌合体, 体外试验表明 A 结构域的底物特异性被成功转移, 并且仍保持较高的转化率。基于同样的策略,

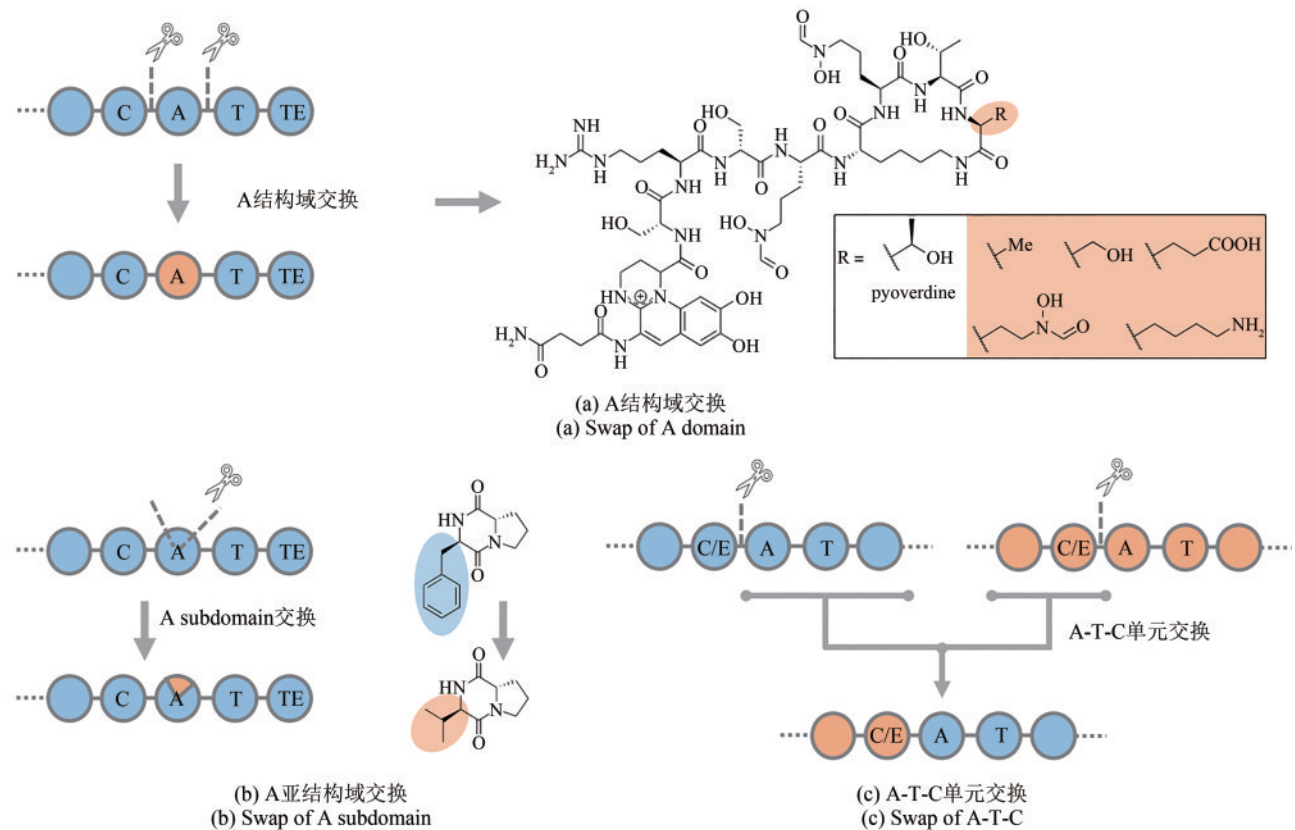


图 10 进化导向的 NRPS 改造

Fig. 10 Evolution-guided engineering for NRPS

Donald Hilvert 课题组^[104]把9个不同的底物特异性移植到 gramicidin S 合成基因簇的 GrsA 模块中 [图 10(b)]。

C 结构域和 C-A 连接域影响 A 结构域催化活性和底物选择性。与此类只改变 A 结构域核心序列的策略不同, Helge Bode 课题组^[105-108]将 A-T-C 或 A-T-C/E 定义为交换单元, 提出了 XU (exchange unit) 以及 XUC (exchange unit condensation domain) 的概念。他们认为 C-A 连接域是理想的重组位点, 因为序列比对表明 A-T 和 T-C 连接域保守性低, 并且在催化循环中可能参与重要的蛋白相互作用, 而 C 与 A 结构域之间主要依赖疏水作用。依据这种交换单元构建的杂合 NRPS 虽然相比于野生型显示出降低的产量, 但仍可以得到足够用于活性分析的目标产物 [图 10(c)], 且在少数例子中实现了与野生型同等的产量。有趣的是, 这种交换单元类似于 PKS 模块边界的新定义, 考虑到 C 结构域同样具有门控的功能, 可以推测 C 结构域与上游 A-T 结构域进化关系上可能具有一定的关联性。

3.4.2 NRPS/PKS 杂合

聚酮和多肽类化合物具有截然不同的骨架, 而 NRPS 和 PKS 杂合大大促进了天然产物类型的多样化。对 NRPS/PKS 组装线进行改造是产生新型生物活性分子的一种有效方法。在真菌中, 高度还原型聚酮合酶 (HR-PKS) 可以与 NRPS 形成杂合, 合成以吡啶酮骨架为代表的一系列真菌聚酮化合物。Hideaki Oikawa 课题组^[109]分析了 NCBI 真菌基因组中 884 条 PKS-NRPS 杂合酶, 发现了酶系统进化树的分支与产物分子骨架之间有明显的对应关系, 为真菌 PKS-NRPS 基因簇产物的分布与结构多样性提供了宏观见解。

细菌 PKS-NRPS 杂合酶是合成博来霉素等的模块型酶。尽管 PKS 和 NRPS 的进化机制仍存在许多疑点^[29], 但通过进化分析发现一些自然重组位点, 可以作为模块改造和结构域融合的切点。双内酯缩肽类化合物 antimycins 由 NRPS-PKS 杂合酶合成, 序列分析表明其可能与三内酯 JBIR-06、四内酯 neoantimycin A 由同一个祖先进化而来。Ikuro Abe 课题组^[110]受此启发, 进一步推测出自然重组发生的位点, 对 JBIR-06 和 neoantimycin A 合成酶模块进行增减, 通过异源表达实现了对内酯缩肽

类化合物环尺寸大小的控制。这项研究通过分析 NRPS-PKS 杂合酶自然重组进化的过程, 证实上述 PKS 和 NRPS 改造策略 (2.4 和 3.4.1 小节) 同样适用于 NRPS-PKS 杂合酶体系^[110]。

4 进化和大数据导向的非模块型酶研究

4.1 非模块型生物合成酶基因挖掘

4.1.1 RiPP (ribosomally-synthesized and post-translationally modified peptide) 合成酶

不同于聚酮、非核糖体肽类化合物等, RiPP 的生物合成途径缺乏共同的生物合成特征, 其基因簇难以进行可靠的生物信息学预测^[111]。针对 RiPP 的预测工具可以依赖于前体肽特征或者修饰酶。丁伟课题组^[112]通过深度学习来探究 RiPP 生物合成的底层逻辑, 提出了基于 BERT 预训练模型的组合模型 BERiPP (bidirectional language model for enhancing the performance of identification of RiPP precursor peptide)。BERiPP 能够在不考虑基因组背景的情况下无差别地识别 RiPP 前体肽, 并对前导肽裂解位点进行预测, 为高通量挖掘新的 RiPP 提供了思路。

YcaO 是 RiPP 中已知的一个修饰酶, 催化噁唑林、噁唑、脒和硫代酰胺形成, 与 TfuA 蛋白共同作用可形成硫代酰胺。Andrew Truman 课题组在放线菌中筛选了 229 个 TfuA 同系蛋白, 并使用一个新的基因挖掘工具 RiPPER 检索了这些蛋白的潜在 BGC 以及 743 条多肽, 得到了 74 个不同的多肽网络, 然后使用 MultiGeneBlast^[113]对每个网络对应的基因簇进行比较, 最终发现一类新的硫代酰胺类化合物 thiovarsolins, 并且证明 TfuA 的进化与前体肽相似度有很强的相关性^[111]。套索肽 (lassopeptide) 的生物合成通常需要两种酶: 套索环化酶和前体肽酶。Douglas Mitchell 课题组^[114]开发了一种用于 BGC 识别的算法 RODEO, 输出 BGC 和多肽的图形、表格, 然后对其进行分析, 针对套索肽进行了优化的 RODEO 归纳了 GenBank 中所有潜在的套索肽基因簇, 根据长度和基本序列特征对得到的套索肽进行评估, 并将其分为预测的前导区和核心区, 构建了 1315 个套索肽前体

的序列相似性网络,进而鉴定出6种新的套索肽。这一发现扩展了套索肽的多样性,也为未来套索肽的基因组挖掘工作提供了一个框架。

此外,基于机器学习技术的进步,Nathan Magarvey课题组^[115]开发了DeepRiPP,集成了基因组和代谢组学数据,使用机器学习来自动发现和分离新的RiPP。DeepRiPP通过3个模块实现:识别独立于基因组结构和邻近生物合成基因的RiPP,优先选择编码新化合物的基因座,从复杂的细菌提取物中自动分离出相应的产物。他们利用DeepRiPP对来自463株菌株的10 498个提取物的数据库进行大规模比较代谢组学分析,最终发现了3种新型RiPP,结构与平台预测的完全一致。DeepRiPP提高了RiPP基因挖掘效率,展示了机器学习技术在微生物基因大数据挖掘中的应用前景。

4.1.2 萜类合酶

萜类化合物是真菌及植物中常见的重要天然产物类型,由IPP(isopentenyl diphosphate)和DMAPP(dimethylallyl diphosphate)为底物形成单萜、倍半萜和二萜等生物合成的线性前体,再由萜类合酶催化多样的环化反应形成复杂的碳骨架结构。与植物和真菌萜类合成酶相比,细菌萜类合酶总体上序列相似度很低。Dickschat和Garbeva课题组^[116]对链霉菌进行了全基因组系统发育分析,比较了萜类合酶基因在链霉菌中的分布,并对这些萜类合酶进行进化分析,研究发现这些酶的进化与链霉菌的进化并不一致,这表明基因水平转移可能是链霉菌萜类合酶基因分布的重要机制。同时,他们发现链霉菌的萜类合酶在进化树上可分为10类,其中土臭素(geosmin)合酶最为丰富。为探究细菌萜类合酶和真菌萜类合酶的进化关系,Chen Feng课题组^[117]对908个真菌萜类合酶和1535个细菌萜类合酶进行进化分析,研究表明真菌同样通过基因水平转移从细菌中获得萜类合酶。此外,近年来越来越多的证据表明基因水平转移在萜类合成前体生物合成途径的进化中也起着重要作用^[118]。

二萜是由4个异戊二烯单位构成的萜类化合物,广泛分布于植物界,其含氧衍生物很多具有较强的生物活性,如紫杉醇、雷公藤内酯等。微生物代谢产物中也发现有二萜类化合物,但与植

物来源二萜合酶相比,来自菌类的二萜合酶的研究较少。为了从公共数据库中挖掘潜在的二萜合酶编码序列,刘宏伟课题组^[119]利用EriG蛋白(猴头菇中形成cyathane骨架的环化酶,属于UbiA超家族)序列作为探针进行基因组挖掘,通过序列聚类分析和进化树分析,发现了细菌和真菌中与UbiA相关的新家族二萜环化酶(cluster 11)。通过在大肠杆菌中表达鉴定了7个新的二萜环化酶,并确定了其对应产物的结构,其中包括一个新的具有不同寻常骨架的二萜lydicene。这项研究丰富了细菌和真菌中二萜环化酶的多样性,更新了UbiA超家族成员,为微生物二萜合酶在生物催化和代谢工程中的应用提供了新的机遇。

近期,Dickschat和Abe课题组^[120]开展了真菌二倍半萜生物合成酶的系统进化研究。二倍半萜是由C端异戊烯转移酶(prenyltransferase, PT)和N端I型萜类合酶两个结构域组成(terpene synthase, TS)的嵌合萜类合酶(PTTS)催化合成,该课题组利用18个PTTS的TS功能区域构建系统进化树,发现PTTS形成6类主要分支,对应不同的环化产物^[120]。刘天罡课题组^[120]进一步拓展了PTTS进化树分析,结合基因挖掘揭示了6类主要分支大致对应异戊二烯线性前体的两大环化模式,即第四-第五双键成环的Type A反应以及第三-第四双键成环的Type B反应。利用PTTS进化树的基因挖掘,刘天罡课题组^[121]首次发现了三种真菌来源三萜合酶及其三萜产物。其中三萜合酶MpMS和CgCS催化的环化模式是不符合上述两类的新机制,表明基于序列分析的萜类合酶功能预测的局限以及其丰富的催化可塑性。

4.1.3 其他基因

在PKS中,催化Claisen缩合反应形成聚酮主链的KS属于硫解酶超家族^[122]。从进化角度来说,这个超家族的成员具有与分支相同的功能簇,都从类似于古菌硫解酶类硫解酶祖先分化而来。考虑到它们进化的多样性和结构的相似性,Ramon Gonzalez课题组^[123]假设硫解酶超家族中除了PKS以外还有酶能够催化迭代Claisen缩合反应合成聚酮骨架。他们通过合成具有代表性的聚酮化合物,如内酯(三乙酸内酯)、烷基间苯二酚酸(alkylresorcinolic acids)、烷基间苯二酚、羟基苯

甲酸和烷基酚等证明了这一途径的可行性。这一发现可以扩展到其他硫解酶，以进一步阐明它们的结构和功能关系，并将它们的生物合成潜力用于PKS研究^[123]。

末端炔是一种广泛应用于有机合成、医药科学、材料科学和生物化学的功能性物质，在微生物体内可以由特殊的去饱和酶——乙炔酶催化形成^[124]。Zhang Wenjun 课题组^[125-126]阐明了JamA、JamB、JamC在末端炔jamaicamide生物合成中的功能，以炔基形成关键酶*jamB*基因作为探针对其序列相似基因进行进化分析，筛选新的炔烃基因簇，发现了由TtuA、TtuB和TtuC组成的新的末端炔生物合成机制，从而扩大了末端炔生物合成研究模型，在合成和化学生物学中有广泛的应用前景。

吡啉吡啉是一种被用作抗癌药物先导的天然产物，其核心结构是由嗜铬吡咯酸(chromopyrrolic acid, CPA)合酶催化的两分子氧化的色氨酸二聚形成。对土壤宏基因组中的CPA合酶同源基因进行系统发育分析，发现了新的吡啉色氨酸boregomyocins A~D、erdasporine A~B以及reductasporine等^[15, 127-128]。

安莎霉素是临床上重要的天然产物家族，这类化合物的显著结构特征是存在一个芳香核心，来自共同的前体3-氨基-5-羟基苯甲酸(AHBA)。对链霉菌AHBA合酶同源基因进行进化分析，最终从6株菌中发现了25个安莎霉素，包括8个新的juanlimycin和neoansamycin等化合物^[15]。

进化导向基因挖掘方法对化合物生物合成特征基因如催化末端炔形成的乙炔酶、催化吡啉吡啉形成的嗜铬吡咯酸合酶以及催化安莎霉素形成的AHBA合酶进行进化分析，从中挖掘出具有相同官能团的活性化合物，该方法使得利用antiSMASH或ClusterFinder^[129]无法直接检测到的特殊化合物BGC的检索成为可能，为发展基于进化的生物信息学途径提供了可行性验证。

4.2 靶向抗性基因的天然产物挖掘

天然产物挖掘的主要目标之一就是发现新的作用模式的抗生素，以抵抗病原菌的多重耐药性。

为避免产生的抗生素对自身的伤害，微生物进化出了几种避免自身毒性的耐药策略，包括产物修饰、底物运输和结合、靶标复制或修饰，这些耐药修饰由位于抗生素BGC附近的抗性基因编码^[130]。基因簇中抗性基因的存在可以作为预测该途径合成的天然产物生物活性的窗口。基于自我抵抗基因识别的天然产物发现有助于弥补活性导向和基因组导向方法在天然产物发现和功能分配中的缺口^[10, 131]。近年来，也有一些利用进化思想靶向抗性基因进行的天然产物挖掘，例如前文中提到的ARTS即是靶向抗性基因的进化导向基因挖掘工具^[16-17]。利用ARTS探测模式，Nadine Ziemert课题组分析了所有已知的BGC和可用的细菌基因组（其中包含已知的耐药靶基因），除了MIBiG数据库中26个已知的基因簇，还检测到22个具有抗性靶标的基因簇，体现了进化导向基因挖掘的应用潜力。

转录调控因子*tetR/marR*和抗性转运因子如*tetA*是四环素生物合成中常见的一对抗性基因，其中抗性转运蛋白是位于细胞膜上的四环素/金属离子逆向转运蛋白，而调控蛋白TetR是四环素诱导阻遏蛋白。基于四环素BGC的抗性机制，戈惠明课题组^[52]以TetR/MarR-转运蛋白作为挖掘四环素类天然产物的指标，结合链长因子CLF的系统发育分析进一步细化，发现了25个不同的四环素基因簇，最终分离出一种新的四环素海南环素。这种同时靶向抗性基因和II型PKS基因的基因挖掘方法为特异性、高效地挖掘新颖的抗生素提供了可能。

4.3 进化导向的非模块型酶改造

基于进化分析所启发的生物合成酶改造并不局限于PKS和NRPS这类多模块化酶。Tobias Erb课题组^[132]通过将烯酰辅酶A羧化/还原酶(ECR)与其他中链脱氢/还原酶(MDR)进行聚类分析(图11)，发现丙酰辅酶A合酶(PCS)以及古菌烯酰还原酶(AER)可能与ECR由共同的祖先进化而来，并具有潜在的CO₂结合口袋和羧化功能。在一定CO₂浓度下，PCS和AER均展现出较弱的羧化功能，但主要是还原产物(大于95%)。结合三

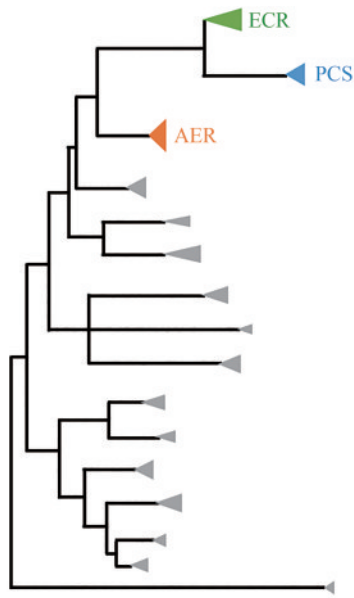


图 11 中链脱氢/还原酶 (MDR) 进化分析
Fig. 11 Phylogenetic analysis of MDR

维结构模型, 作者进一步分析了 CO_2 结合口袋的序列, 并对关键氨基酸进行突变。通过增强 CO_2 的结合和阻止水进入口袋, 成功唤醒了 PCS 和 AER 的羧化功能, 羧化产物占比均提高了约 20 倍, 成为主要产物 (图 12)。

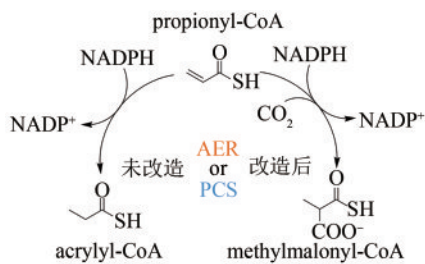


图 12 古菌烯酰还原酶 (AER) 和丙酰辅酶 A 合酶 (PCS) 的改造

Fig. 12 Engineering for AER and PCS

上述的研究是通过对同源蛋白的比较来阐明特定蛋白功能的进化, 这种方法通常称为“水平的”。因为它们是基于对某一进化阶段的分析, 用于分析的蛋白都是在现存物种中发现的蛋白质。而系统发育算法为序列分析增加了垂直维度, 可以根据现存序列的进化追溯到共同的祖先^[133]。祖先序列重建 (ancestral sequence reconstruction, ASR) 即是实现从现代序列 (即现存序列) 推断原始序列的强大工具^[134]。ASR 的一个基本元素是系统发育树的计算, 其叶子是所选的现存序列, 与

系统发育树的根相关的重构序列代表了所研究序列的共同祖先。如果这个序列编码了一种蛋白质, 就可以通过基因合成技术“复活”这个祖先蛋白, 并借助生物化学实验来研究其生化特性。ASR 还可以推导出树中所有内部节点的序列, 进一步阐明进化过程^[133]。

5 总结与展望

BGC 可以通过广泛的水平基因转移有效地分散, 甚至跨越门的边界, 因此天然产物研究中基于进化的发现策略在增加新颖性方面有力地补充了传统方法。基于进化原理的生物信息学分析工具的开发和应用产生了日益增长的遗传 (基因)、催化 (蛋白) 和化学 (化合物结构) 数据库, 并推动天然产物研究进入现代大数据时代, 使得天然产物全景图可视化成为可能^[1-2, 6]。通过这些策略获得的天然产物不仅加深了对天然生物活性分子合成方式的认识, 而且丰富了生物活性化合物库。

随着天然产物相关研究的数量和质量的生长, 人工智能分析方法 (如机器学习) 的应用潜力也在增加, 将进化导向的方法与人工智能结合将是该领域的发展方向之一。而成功的机器学习方法需要高质量的训练数据, 未来可能需要跨实验室甚至国际的协调努力, 以标准化的方式生成数据集并进行管理^[135], 而这依赖于生物信息学的发展^[136]。

基于进化及大数据进行的天然产物挖掘不仅本质上受到已测序基因组数据量与范围的限制, 目前还仍然面临一些挑战: ①许多含有目标分子 BGC 的微生物在实验室条件下不可培养, 或者目标基因簇不表达。目前对于这一问题的解决, 主要是依赖于异源表达等分子生物学技术的发展, 因此, 提高异源表达效率等技术方面的改进将促进进化导向的天然产物挖掘。②预测基因簇产物的生物活性仍然困难。目前只能通过靶向生物活性分子类似物或者抗性基因的天然产物挖掘提高发现活性分子的概率。探寻化合物结构与生物活性在进化中的关系或许能够为进化导向天然产物挖掘发现新的活性分子提供更多机会。③萜类、

生物碱等的BGC很难展现化合物的结构特征，基因簇产物的分子结构预测仍有较大挑战。对于这些非模块型BGC，还需要增加基因簇与其产物的表征数量，并仔细分析每一步生物合成酶的进化特征和规律。

酶改造方面，对于模块型酶的合理改造是其发现以来的重要目标。在PKS及NRPS改造的早期尝试中，交换或删减一些结构域和模块经常导致酶活性显著降低甚至是无活性，而对组装线系统的进化分析推断出自然重组发生的位点，可以指导人工酶改造的设计。目前，基于进化分析得出的PKS新模块定义以及NRPS的XUC概念为模块型酶的改造提供了理论依据，这些概念的应用以及进一步优化将推动合成生物学的发展。此外，这种思路不局限于PKS和NRPS这类多结构域、多模块的酶，将进化与大数据分析相结合也将为其他酶的改造提供新的思路。

自然界千万年来基于各种进化机制创造了丰富多样的生物合成途径和天然产物，人类认识自然界的脚步也从未停歇。通过生物信息学研究生物合成酶进化的机制，挖掘其活性产物用于医药健康或者农业生产领域，结合大数据分析描绘天然产物全景图，或者利用自然界的规则和元件，从模仿进化的角度去设计改造生物合成酶以满足人类的需求，这正是发现自然并改造自然的过程。

参 考 文 献

- [1] CHEVRETTE M G, GAVRILIDOU A, MANTRI S, et al. The confluence of big data and evolutionary genome mining for the discovery of natural products[J]. *Natural Product Reports*, 2021, 38(11): 2024-2040.
- [2] CHEVRETTE M G, GUTIÉRREZ-GARCÍA K, SELEM-MOJICA N, et al. Evolutionary dynamics of natural product biosynthesis in bacteria[J]. *Natural Product Reports*, 2020, 37(4): 566-599.
- [3] JENSEN P R. Natural products and the gene cluster revolution [J]. *Trends in Microbiology*, 2016, 24(12): 968-977.
- [4] GAVRILIDOU A, KAUTSAR S A, ZABURANNYI N, et al. Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes[J]. *Nature Microbiology*, 2022, 7(5): 726-735.
- [5] CHEN S C, ZHANG C, ZHANG L H. Investigation of the molecular landscape of bacterial aromatic polyketides by global analysis of type II polyketide synthases[J]. *Angewandte Chemie International Edition*, 2022, 61(24): e202202286.
- [6] ADAMEK M, ALANJARY M, ZIEMERT N. Applied evolution: phylogeny-based approaches in natural products research [J]. *Natural Product Reports*, 2019, 36(9): 1295-1312.
- [7] PANDE S, KOST C. Bacterial unculturability and the formation of intercellular metabolic networks[J]. *Trends in Microbiology*, 2017, 25(5): 349-361.
- [8] ZIEMERT N, ALANJARY M, WEBER T. The evolution of genome mining in microbes - a review[J]. *Natural Product Reports*, 2016, 33(8): 988-1005.
- [9] WALKER J M. *Engineering natural products biosynthesis*[M]. New York: Humana Imprint, 2022.
- [10] 杨谦, 程伯涛, 汤志军, 等. 基因组挖掘在天然产物发现中的应用和前景[J]. *合成生物学*, 2021, 2(5): 697-715.
YANG Q, CHENG B T, TANG Z J, et al. Applications and prospects of genome mining in the discovery of natural products[J]. *Synthetic Biology Journal*, 2021, 2(5): 697-715.
- [11] ZERIKLY M, CHALLIS G L. Strategies for the discovery of new natural products by genome mining[J]. *ChemBioChem*, 2009, 10(4): 625-633.
- [12] SCHERLACH K, HERTWECK C. Mining and unearthing hidden biosynthetic potential[J]. *Nature Communications*, 2021, 12(1): 3864.
- [13] BAUMAN K D, BUTLER K S, MOORE B S, et al. Genome mining methods to discover bioactive natural products[J]. *Natural Product Reports*, 2021, 38(11): 2100-2129.
- [14] CRUZ-MORALES P, KOPP J F, MARTÍNEZ-GUERRERO C, et al. Phylogenomic analysis of natural products biosynthetic gene clusters allows discovery of arseno-organic metabolites in model *Streptomyces*[J]. *Genome Biology and Evolution*, 2016, 8(6): 1906-1916.
- [15] KANG H S. Phylogeny-guided (meta)genome mining approach for the targeted discovery of new microbial natural products[J]. *Journal of Industrial Microbiology & Biotechnology*, 2017, 44(2): 285-293.
- [16] ALANJARY M, KRONMILLER B, ADAMEK M, et al. The antibiotic resistant target seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery[J]. *Nucleic Acids Research*, 2017, 45(W1): W42-W48.
- [17] DIRENÇ MUNGAN M, ALANJARY M, BLIN K, et al. ARTS 2.0: feature updates and expansion of the antibiotic resistant target seeker for comparative genome mining[J]. *Nucleic Acids Research*, 2020, 48(W1): W546-W552.
- [18] 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[J]. *PLoS One*, 2012, 7(3): e34064.
- [19] KLAU L J, PODELL S, CREAMER K E, et al. The Natural

- Product Domain Seeker version 2 (NaPDoS2) webtool relates ketosynthase phylogeny to biosynthetic function[J]. *Journal of Biological Chemistry*, 2022, 298(10): 102480.
- [20] SÉLEM-MOJICA N, AGUILAR C, GUTIÉRREZ-GARCÍA K, et al. EvoMining reveals the origin and fate of natural product biosynthetic enzymes[J]. *Microbial Genomics*, 2019, 5(12): e000260.
- [21] NAVARRO-MUÑOZ J C, SELEM-MOJICA N, MULLOWNEY M W, et al. A computational framework to explore large-scale biosynthetic diversity[J]. *Nature Chemical Biology*, 2020, 16(1): 60-68.
- [22] MEDEMA M H, KOTTMANN R, YILMAZ P, et al. Minimum information about a biosynthetic gene cluster[J]. *Nature Chemical Biology*, 2015, 11(9): 625-631.
- [23] KAUTSAR S A, BLIN K, SHAW S, et al. MIBiG 2.0: a repository for biosynthetic gene clusters of known function[J]. *Nucleic Acids Research*, 2020, 48(D1): D454-D458.
- [24] BLIN K, SHAW S, KAUTSAR S A, et al. The antiSMASH database version 3: increased taxonomic coverage and new query features for modular enzymes[J]. *Nucleic Acids Research*, 2021, 49(D1): D639-D643.
- [25] RAWLINGS B J. Type I polyketide biosynthesis in bacteria (part A—erythromycin biosynthesis) [J]. *Natural Product Reports*, 2001, 18(2): 190-227.
- [26] HERTWECK C, LUZHETSKYY A, REBETS Y, et al. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork[J]. *Natural Product Reports*, 2007, 24(1): 162-190.
- [27] ABE I, MORITA H. Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases[J]. *Natural Product Reports*, 2010, 27(6): 809-838.
- [28] YU D Y, XU F C, ZENG J, et al. Type III polyketide synthases in natural product biosynthesis[J]. *IUBMB Life*, 2012, 64(4): 285-295.
- [29] NIVINA A, YUET K P, HSU J, et al. Evolution and diversity of assembly-line polyketide synthases[J]. *Chemical Reviews*, 2019, 119(24): 12524-12547.
- [30] JENKE-KODAMA H, DITTMANN E. Evolution of metabolic diversity: insights from microbial polyketide synthases[J]. *Phytochemistry*, 2009, 70(15-16): 1858-1866.
- [31] NIVINA A, HERRERA PAREDES S, FRASER H B, et al. GRINS: genetic elements that recode assembly-line polyketide synthases and accelerate their diversification[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2021, 118(26): e2100751118.
- [32] JENKE-KODAMA H, SANDMANN A, MÜLLER R, et al. Evolutionary implications of bacterial polyketide synthases[J]. *Molecular Biology and Evolution*, 2005, 22(10): 2027-2039.
- [33] NGUYEN T, ISHIDA K, JENKE-KODAMA H, et al. Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection [J]. *Nature Biotechnology*, 2008, 26(2): 225-233.
- [34] LOPEZ J V. Naturally mosaic operons for secondary metabolite biosynthesis: variability and putative horizontal transfer of discrete catalytic domains of the epothilone polyketide synthase locus[J]. *Molecular Genetics and Genomics*, 2004, 270(5): 420-431.
- [35] ZHANG L H, HASHIMOTO T, QIN B, et al. Characterization of giant modular PKSs provides insight into genetic mechanism for structural diversification of aminopolyl polyketides [J]. *Angewandte Chemie International Edition*, 2017, 56(7): 1740-1745.
- [36] KEATINGE-CLAY A T. Polyketide synthase modules redefined[J]. *Angewandte Chemie International Edition*, 2017, 56(17): 4658-4660.
- [37] VANDER WOOD D A, KEATINGE-CLAY A T. The modules of *trans*-acyltransferase assembly lines redefined with a central acyl carrier protein[J]. *Proteins*, 2018, 86(6): 664-675.
- [38] CAFFREY P. Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases [J]. *Chembiochem*, 2003, 4(7): 654-657.
- [39] VANDER WOOD D A, KEATINGE-CLAY A T. The modules of *trans*-acyltransferase assembly lines redefined with a central acyl carrier protein[J]. *Proteins*, 2018, 86(6): 664-675.
- [40] MEDEMA M H, CIMERMANCIC P, SALI A, et al. A systematic computational analysis of biosynthetic gene cluster evolution: lessons for engineering biosynthesis[J]. *PLoS Computational Biology*, 2014, 10(12): e1004016.
- [41] RIDLEY C P, LEE H Y, KHOSLA C. Evolution of polyketide synthases in bacteria[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105(12): 4595-4600.
- [42] HILLENMEYER M E, VANDOVA G A, BERLEW E E, et al. Evolution of chemical diversity by coordinated gene swaps in type II polyketide gene clusters[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2015, 112(45): 13952-13957.
- [43] GABALDÓN T, KOONIN E V. Functional and evolutionary implications of gene orthology[J]. *Nature Reviews Genetics*, 2013, 14(5): 360-366.
- [44] FRITZSCHE K, ISHIDA K, HERTWECK C. Orchestration of discoid polyketide cyclization in the resistomycin pathway[J]. *Journal of the American Chemical Society*, 2008, 130(26): 8307-8316.
- [45] FRALEY A E, DIETERICH C L, MABESOONE M F J, et al. Structure of a promiscuous thioesterase domain responsible for branching acylation in polyketide biosynthesis[J]. *Angewandte Chemie International Edition*, 2022, 61(39): e202206385.
- [46] SCHWECKE T, APARICIO J F, MOLNÁR I, et al. The biosyn-

- thetic gene cluster for the polyketide immunosuppressant rapamycin[J]. Proceedings of the National Academy of Sciences of the United States of America, 1995, 92(17): 7839-7843.
- [47] HAYDOCK S F, APARICIO J F, MOLNÁR I, et al. Divergent sequence motifs correlated with the substrate specificity of (methyl) malonyl-CoA: acyl carrier protein transacylase domains in modular polyketide synthases[J]. FEBS Letters, 1995, 374(2): 246-248.
- [48] APARICIO J F, MOLNÁR I, SCHWECKE T., et al. Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase[J]. Gene, 1996, 169(1): 9-16.
- [49] KAKAVAS S J, KATZ L, STASSI D. Identification and characterization of the niddamycin polyketide synthase genes from *Streptomyces caelestis*[J]. Journal of Bacteriology, 1997, 179(23): 7515-7522.
- [50] KANG H S, BRADY S F. Mining soil metagenomes to better understand the evolution of natural product structural diversity: pentangular polyphenols as a case study[J]. Journal of the American Chemical Society, 2014, 136(52): 18111-18119.
- [51] KANG H S, BRADY S F. Arimetamycin A: improving clinically relevant families of natural products through sequence-guided screening of soil metagenomes[J]. Angewandte Chemie International Edition, 2013, 52(42): 11063-11067.
- [52] LI L Y, HU Y L, SUN J L, et al. Resistance and phylogeny guided discovery reveals structural novelty of tetracycline antibiotics[J]. Chemical Science, 2022, 13(43): 12892-12898.
- [53] ZIEMERT N, JENSEN P R. Phylogenetic approaches to natural product structure prediction[J]. Methods in Enzymology, 2012, 517: 161-182.
- [54] UEOKA R, URIA A R, REITER S, et al. Metabolic and evolutionary origin of actin-binding polyketides from diverse organisms[J]. Nature Chemical Biology, 2015, 11(9): 705-712.
- [55] HELFRICH E J N, UEOKA R, DOLEV A, et al. Automated structure prediction of *trans*-acyltransferase polyketide synthase products[J]. Nature Chemical Biology, 2019, 15(8): 813-821.
- [56] HELFRICH E J N, UEOKA R, CHEVRETTE M G, et al. Evolution of combinatorial diversity in *trans*-acyltransferase polyketide synthase assembly lines across bacteria[J]. Nature Communications, 2021, 12(1): 1422.
- [57] GUO J, RAN H M, ZENG J, et al. Tafuketide, a phylogeny-guided discovery of a new polyketide from *Talaromyces funiculosus* Salicorn 58[J]. Applied Microbiology and Biotechnology, 2016, 100(12): 5323-5338.
- [58] SHEN B, HINDRA, YAN X H, et al. Enediynes: exploration of microbial genomics to discover new anticancer drug leads[J]. Bioorganic and Medicinal Chemistry Letters, 2015, 25(1): 9-15
- [59] YAN X H, GE H M, HUANG T T, et al. Strain prioritization and genome mining for enediyne natural products[J]. mBio, 2016, 7(6): e02104-e02116.
- [60] WEISSMAN K J. Genetic engineering of modular PKSs: from combinatorial biosynthesis to synthetic biology[J]. Natural Product Reports, 2016, 33(2): 203-230.
- [61] 曹晨凯, 李佳隆, 张科春. 人工代谢途径合成有机醇有机酸的研究进展[J]. 合成生物学, 2021, 2(6): 902-919.
- CAO C K, LI J L, ZHANG K C. Progress in artificial metabolic pathways for biosynthesis of organic alcohols & acids[J]. Synthetic Biology Journal, 2021, 2(6): 902-919.
- [62] BOOTH T J, BOZHÜYÜK K A J, LISTON J D, et al. Bifurcation drives the evolution of assembly-line biosynthesis[J]. Nature Communications, 2022, 13(1): 3498.
- [63] HIRSCH M, FITZGERALD B J, KEATINGE-CLAY A T. How *cis*-acyltransferase assembly-line ketosynthases gatekeep for processed polyketide intermediates[J]. ACS Chemical Biology, 2021, 16(11): 2515-2526.
- [64] MIYAZAWA T, HIRSCH M, ZHANG Z C, et al. An *in vitro* platform for engineering and harnessing modular polyketide synthases[J]. Nature Communications, 2020, 11(1): 80.
- [65] MIYAZAWA T, FITZGERALD B J, KEATINGE-CLAY A T. Preparative production of an enantiomeric pair by engineered polyketide synthases[J]. Chemical Communications, 2021, 57(70): 8762-8765.
- [66] PENG H Y, ISHIDA K, SUGIMOTO Y, et al. Emulating evolutionary processes to morph aureothin-type modular polyketide synthases and associated oxygenases[J]. Nature Communications, 2019, 10(1): 3918.
- [67] YUZAWA S, DENG K, WANG G, et al. Comprehensive *in vitro* analysis of acyltransferase domain exchanges in modular polyketide synthases and its application for short-chain ketone production[J]. ACS Synthetic Biology, 2017, 6(1): 139-147.
- [68] YUZAWA S, MIRSIAGHI M, JOCIC R, et al. Short-chain ketone production by engineered polyketide synthases in *Streptomyces albus*[J]. Nature Communications, 2018, 9: 4569.
- [69] ZARGAR A, VALENCIA L, WANG J, et al. A bimodular PKS platform that expands the biological design space[J]. Metabolic Engineering, 2020, 61: 389-396.
- [70] ENG C H, YUZAWA S, WANG G, et al. Alteration of polyketide stereochemistry from anti to syn by a ketoreductase domain exchange in a type I modular polyketide synthase subunit[J]. Biochemistry, 2016, 55(12): 1677-1680.
- [71] HAGEN A, POUST S, DE ROND T, et al. Engineering a polyketide synthase for *in vitro* production of adipic acid[J]. ACS Synthetic Biology, 2016, 5(1): 21-27.
- [72] ZARGAR A, LAL R, VALENCIA L, et al. Chemoinformatic-guided engineering of polyketide synthases[J]. Journal of the American Chemical Society, 2020, 142(22): 9896-9901.
- [73] MUSIOL-KROLL E M, WOHLLEBEN W. Acyltransferases

- as tools for polyketide synthase engineering[J]. *Antibiotics*, 2018, 7(3): 62.
- [74] KWAN D H, TOSIN M, SCHLÄGER N, et al. Insights into the stereospecificity of ketoreduction in a modular polyketide synthase[J]. *Organic & Biomolecular Chemistry*, 2011, 9(7): 2053-2056.
- [75] BAGDE S R, MATHEWS I I, FROMME J C, et al. Modular polyketide synthase contains two reaction chambers that operate asynchronously[J]. *Science*, 2021, 374(6568): 723-729.
- [76] COGAN D P, ZHANG K M, LI X Y, et al. Mapping the catalytic conformations of an assembly-line polyketide synthase module[J]. *Science*, 2021, 374(6568): 729-734.
- [77] HERBST D A, JAKOB R P, ZÄHRINGER F, et al. Mycoferosic acid synthase exemplifies the architecture of reducing polyketide synthases[J]. *Nature*, 2016, 531(7595): 533-537
- [78] ZHENG J T, GAY D C, DEMELER B, et al. Divergence of multimodular polyketide synthases revealed by a didomain structure[J]. *Nature Chemical Biology*, 2012, 8(7): 615-621.
- [79] GAY D, YOU Y O, KEATINGE-CLAY A, et al. Structure and stereospecificity of the dehydratase domain from the terminal module of the rifamycin polyketide synthase[J]. *Biochemistry*, 2013, 52(49): 8916-8928.
- [80] DUTTA S, WHICHER J R, HANSEN D A, et al. Structure of a modular polyketide synthase[J]. *Nature*, 2014, 510(7506): 512-517.
- [81] WHICHER J R, DUTTA S, HANSEN D A, et al. Structural rearrangements of a polyketide synthase module during its catalytic cycle[J]. *Nature*, 2014, 510(7506): 560-564.
- [82] FELNAGLE E A, JACKSON E E, CHAN Y A, et al. Nonribosomal peptide synthetases involved in the production of medically relevant natural products[J]. *Molecular Pharmaceutics*, 2008, 5(2): 191-211.
- [83] SIEBER S A, MARAHIEL M A. Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics[J]. *Chemical Reviews*, 2005, 105(2): 715-738.
- [84] SÜSSMUTH R D, MAINZ A. Nonribosomal peptide synthesis—principles and prospects[J]. *Angewandte Chemie International Edition*, 2017, 56(14): 3770-3821.
- [85] JAREMKO M J, DAVIS T D, CORPUZ J C, et al. Type II non-ribosomal peptide synthetase proteins: structure, mechanism, and protein-protein interactions[J]. *Natural Product Reports*, 2020, 37(3): 355-379.
- [86] CALCOTT M J, OWEN J G, ACKERLEY D F. Efficient rational modification of non-ribosomal peptides by adenylation domain substitution[J]. *Nature Communications*, 2020, 11(1): 4554.
- [87] BAUNACH M, CHOWDHURY S, STALLFORTH P, et al. The landscape of recombination events that create nonribosomal peptide diversity[J]. *Molecular Biology and Evolution*, 2021, 38(5): 2116-2130.
- [88] RAUSCH C, HOOF I, WEBER T, et al. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution[J]. *BMC Ecology and Evolution*, 2007, 7: 78.
- [89] WHEADON M J, TOWNSEND C A. Evolutionary and functional analysis of an NRPS condensation domain integrates β -lactam, D-amino acid, and dehydroamino acid synthesis[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2021, 118(17): e2026017118.
- [90] STACHELHAUS T, MOOTZ H D, MARAHIEL M A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases[J]. *Chemistry & Biology*, 1999, 6(8): 493-505.
- [91] RÖTTIG M, MEDEMA M H, BLIN K, et al. NRSPredictor2—a web server for predicting NRPS adenylation domain specificity[J]. *Nucleic Acids Research*, 2011, 39(suppl_2): W362-W367.
- [92] CHEVRETTE M G, AICHELER F, KOHLBACHER O, et al. SANDPUMA: ensemble predictions of nonribosomal peptide chemistry reveal biosynthetic diversity across Actinobacteria [J]. *Bioinformatics*, 2017, 33(20): 3202-3210.
- [93] ZIEMERT N, LECHNER A, WIETZ M, et al. Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2014, 111(12): E1130-E1139.
- [94] PATTESON J B, FORTINEZ C M, PUTZ A T, et al. Structure and function of a dehydrating condensation domain in nonribosomal peptide biosynthesis[J]. *Journal of the American Chemical Society*, 2022, 144(31): 14057-14070.
- [95] HOVER B M, KIM S H, KATZ M, et al. Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens[J]. *Nature Microbiology*, 2018, 3(4): 415-422.
- [96] CULP E J, WAGLECHNER N, WANG W L, et al. Evolution-guided discovery of antibiotics that inhibit peptidoglycan remodelling[J]. *Nature*, 2020, 578(7796): 582-587.
- [97] XU M, WANG W L, WAGLECHNER N, et al. Phylogeny-informed synthetic biology reveals unprecedented structural novelty in type V glycopeptide antibiotics[J]. *ACS Central Science*, 2022, 8(5): 615-626.
- [98] WANG Z Q, KOIRALA B, HERNANDEZ Y, et al. Bioinformatic prospecting and synthesis of a bifunctional lipopeptide antibiotic that evades resistance[J]. *Science*, 2022, 376(6596): 991-996.
- [99] CALCOTT M J, OWEN J G, LAMONT I L, et al. Biosynthesis of novel Pyoverdines by domain substitution in a nonribosomal peptide synthetase of *Pseudomonas aeruginosa*[J]. *Applied and Environmental Microbiology*, 2014, 80(18): 5723-

- 5731.
- [100] THIRLWAY J, LEWIS R, NUNNS L, et al. Introduction of a non-natural amino acid into a nonribosomal peptide antibiotic by modification of adenylation domain specificity[J]. *Angewandte Chemie International Edition*, 2012, 51(29): 7181-7184.
- [101] KRIES H, WACHTEL R, PABST A, et al. Reprogramming nonribosomal peptide synthetases for clickable amino acids[J]. *Angewandte Chemie International Edition*, 2014, 53(38): 10105-10108.
- [102] NGUYEN K T, RITZ D, GU J Q, et al. Combinatorial biosynthesis of novel antibiotics related to daptomycin[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2006, 103(46): 17462-17467.
- [103] CRÜSEMANN M, KOHLHAAS C, PIEL J. Evolution-guided engineering of nonribosomal peptide synthetase adenylation domains[J]. *Chemical Science*, 2013, 4(3): 1041-1045.
- [104] KRIES H, NIQUILLE D L, HILVERT D. A subdomain swap strategy for reengineering nonribosomal peptides[J]. *Chemistry & Biology*, 2015, 22(5): 640-648.
- [105] BOZHÜYÜK K A J, LINCK A, TIETZE A, et al. Modification and *de novo* design of non-ribosomal peptide synthetases using specific assembly points within condensation domains[J]. *Nature Chemistry*, 2019, 11(7): 653-661.
- [106] BOZHÜYÜK K A J, WATZEL J, ABBOOD N, et al. Synthetic zippers as an enabling tool for engineering of non-ribosomal peptide synthetases[J]. *Angewandte Chemie International Edition*, 2021, 60(32): 17531-17538.
- [107] KRANZ J, WENSKI S L, DICHTER A A, et al. Influence of condensation domains on activity and specificity of adenylation domains[EB/OL]. *bioRxiv*, 2021, DOI: 10.1101/2021.08.23.457306[2023-02-01]. <https://www.biorxiv.org/content/10.1101/2021.08.23.457306v1>.
- [108] BOZHÜYÜK K A J, FLEISCHHACKER F, LINCK A, et al. *De novo* design and engineering of non-ribosomal peptide synthetases[J]. *Nature Chemistry*, 2018, 10(3): 275-281.
- [109] MINAMI A, UGAI T, OZAKI T, et al. Predicting the chemical space of fungal polyketides by phylogeny-based bioinformatics analysis of polyketide synthase-nonribosomal peptide synthetase and its modification enzymes[J]. *Scientific Reports*, 2020, 10: 13556.
- [110] AWAKAWA T, FUJIOKA T, ZHANG L H, et al. Reprogramming of the antimycin NRPS-PKS assembly lines inspired by gene evolution[J]. *Nature Communications*, 2018, 9: 3534.
- [111] SANTOS-ABERTURAS J, CHANDRA G, FRATTARUOLO L, et al. Uncovering the unexplored diversity of thioamidated ribosomal peptides in Actinobacteria using the RiPPER genome mining tool[J]. *Nucleic Acids Research*, 2019, 47(9): 4624-4637.
- [112] 吕靖伟, 邓子新, 张琪, 等. 基于深度学习识别 RiPPs 前体肽及裂解位点[J]. *合成生物学*, 2022, 3(6), 1262-1276.
- LYU J W, DENG Z X, ZHANG Q, et al. Identification of RiPPs precursor peptides and cleavage sites based on deep learning[J]. *Synthetic Biology Journal*, 2022, 3(6), 1262-1276.
- [113] MEDEMA M H, TAKANO E, BREITLING R. Detecting sequence homology at the gene cluster level with MultiGeneBlast [J]. *Molecular Biology and Evolution*, 2013, 30(5): 1218-1223.
- [114] TIETZ J I, SCHWALEN C J, PATEL P S, et al. A new genome-mining tool redefines the lasso peptide biosynthetic landscape [J]. *Nature Chemical Biology*, 2017, 13(5): 470-478.
- [115] MERWIN N J, MOUSA W K, DEJONG C A, et al. DeepRiPP integrates multiomics data to automate discovery of novel ribosomally synthesized natural products[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2020, 117(1): 371-380.
- [116] MARTÍN-SÁNCHEZ L, SINGH K S, AVALOS M, et al. Phylogenomic analyses and distribution of terpene synthases among *Streptomyces*[J]. *Beilstein Journal of Organic Chemistry*, 2019, 15: 1181-1193.
- [117] JIA Q D, CHEN X L, KÖLLNER T G, et al. Terpene synthase genes originated from bacteria through horizontal gene transfer contribute to terpenoid diversity in fungi[J]. *Scientific Reports*, 2019, 9: 9223.
- [118] AVALOS M, GARBEVA P, VADER L, et al. Biosynthesis, evolution and ecology of microbial terpenoids[J]. *Natural Product Reports*, 2022, 39(2): 249-272.
- [119] YANG Y L, ZHANG S S, MA K, et al. Discovery and characterization of a new family of diterpene cyclases in bacteria and fungi[J]. *Angewandte Chemie International Edition*, 2017, 56(17): 4749-4752.
- [120] CHEN R, JIA Q D, MU X, et al. Systematic mining of fungal chimeric terpene synthases using an efficient precursor-providing yeast chassis[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2021, 118(29): e2023247118.
- [121] TAO H, LAUTERBACH L, BIAN G K, et al. Discovery of non-squalene triterpenes[J]. *Nature*, 2022, 606(7913): 414-419.
- [122] JIANG C G, KIM S Y, SUH D Y. Divergent evolution of the thiolase superfamily and chalcone synthase family[J]. *Molecular Phylogenetics and Evolution*, 2008, 49(3): 691-701.
- [123] TAN Z G, CLOMBURG J M, CHEONG S, et al. A polyketoacyl-CoA thiolase-dependent pathway for the synthesis of polyketide backbones[J]. *Nature Catalysis*, 2020, 3(7): 593-603.
- [124] SHANKLIN J, GUY J E, MISHRA G, et al. Desaturases: emerging models for understanding functional diversification of diiron-containing enzymes[J]. *Journal of Biological Chemistry*, 2009, 284(28): 18559-18563.
- [125] ZHU X J, LIU J, ZHANG W J. *De novo* biosynthesis of termi-

- nal alkyne-labeled natural products[J]. *Nature Chemical Biology*, 2015, 11(2): 115-120.
- [126] ZHU X J, SU M, MANICKAM K, et al. Bacterial genome mining of enzymatic tools for alkyne biosynthesis[J]. *ACS Chemical Biology*, 2015, 10(12): 2785-2793.
- [127] CHANG F Y, BRADY S F. Discovery of indolotryptoline anti-proliferative agents by homology-guided metagenomic screening[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2013, 110(7): 2478-2483.
- [128] CHANG F Y, TERNEI M A, CALLE P Y, et al. Targeted metagenomics: finding rare tryptophan dimer natural products in the environment[J]. *Journal of the American Chemical Society*, 2015, 137(18): 6044-6052.
- [129] CIMERMANCIC P, MEDEMA M H, CLAESEN J., et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters[J]. *Cell*, 2014, 158(2): 412-421.
- [130] O'NEILL E C, SCHORN M, LARSON C B, et al. Targeted antibiotic discovery through biosynthesis-associated resistance determinants: target directed genome mining[J]. *Critical Reviews in Microbiology*, 2019, 45(3): 255-277.
- [131] YAN Y, LIU N, TANG Y. Recent developments in self-resistance gene directed natural product discovery[J]. *Natural Product Reports*, 2020, 37(7): 879-892.
- [132] BERNHARDSGRÜTTER I, SCHELL K, PETER D M, et al. Awakening the sleeping carboxylase function of enzymes: engineering the natural CO₂-binding potential of reductases[J]. *Journal of the American Chemical Society*, 2019, 141(25): 9778-9782.
- [133] SIKOSEK T. *Computational methods in protein evolution*[M]// *Methods in molecular biology*. New York: Humana Press, 2019: 1064-3745.
- [134] HARMS M J, THORNTON J W. Analyzing protein structure and function using ancestral gene reconstruction[J]. *Current Opinion in Structural Biology*, 2010, 20(3): 360-366.
- [135] CECH N B, MEDEMA M H, CLARDY J. Benefiting from big data in natural products: importance of preserving foundational skills and prioritizing data quality[J]. *Natural Product Reports*, 2021, 38(11): 1947-1953.
- [136] JEON J, KANG S, KIM H U. Predicting biochemical and physiological effects of natural products from molecular structures using machine learning[J]. *Natural Product Reports*, 2021, 38(11): 1954-1966.



通讯作者: 张骊驎(1989—),男,特聘研究员,博士生导师。研究方向为天然产物分子多样性与生物合成进化,以及进化导向生物合成改造等。

E-mail: zhanglihan@westlake.edu.cn



第一作者: 张凡忠(1991—),女,博士后。研究方向为微生物天然产物分离鉴定,细菌中PKS和NRPS的进化导向基因挖掘。

E-mail: zhangfanzhong@westlake.edu.cn