

特约评述

DOI: 10.12211/2096-8280.2023-098

基因组深度挖掘驱动微生物萜类化合物高效发现

雷茹, 陶慧, 刘天罡

(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)

摘要: 萜类天然产物广泛分布于动物 (包括海洋无脊椎动物)、植物、微生物中, 具有复杂的化学结构和丰富的生物活性。人们通过从植物和微生物中直接分离提取的方式获得了大量萜类天然产物, 然而随着越来越多化合物被发现, 使用基于自然筛选的传统挖掘方式很难获得新的萜类天然产物。随着基因组测序技术和合成生物学使能技术的不断发展, 我们进入了基因组挖掘驱动天然产物发现的新时代, 萜类天然产物的挖掘也进入了“井喷式”发现新阶段。针对基因组挖掘在微生物萜类天然产物发现方面的应用, 本文综述了近年来使用的主要研究策略和最新研究进展, 介绍了多种高效微生物底盘、基因组深度挖掘策略、人工智能与自动化平台等驱动的萜类化合物挖掘的最新研究进展, 讨论了基因组挖掘萜类天然产物面临的挑战, 展望了未来萜类化合物创新发现的发展趋势。通过在多种微生物中强化前体供应途径, 人们打造了多个萜类化合物合成底盘, 突破了异源合成萜类天然产物时“产量低”和“产物难获取”的瓶颈; 针对萜类天然产物生物合成基因簇或萜类合酶进行深度挖掘, 可以有效地解决“重复发现”和“集中度低”的难题; 随着人工智能和自动化技术在合成生物学领域的发展和应用, 萜类化合物的发现也进入了高通量智能发现时期, 显著地改善了“研究通量低”的现状, 高效获得了大量新结构萜类天然产物。在未来, 更多萜类化合物将开发成药物、进入工业化生产应用, 更多萜类“暗物质”会走进我们视野。

关键词: 萜类天然产物; 萜类合酶; 微生物底盘; 基因组挖掘; 人工智能; 自动化高通量平台

中图分类号: Q819 文献标志码: A

Deep genome mining boosts the discovery of microbial terpenoids

LEI Ru, TAO Hui, LIU Tiangang

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

Abstract: The natural products terpenoids are widely distributed in animals (marine invertebrates), plants, microorganisms, with diverse molecular structures for bioactivities. A large number of terpenoids have been extracted directly from plants and microorganisms. However, traditional methods based on natural screening face challenges in

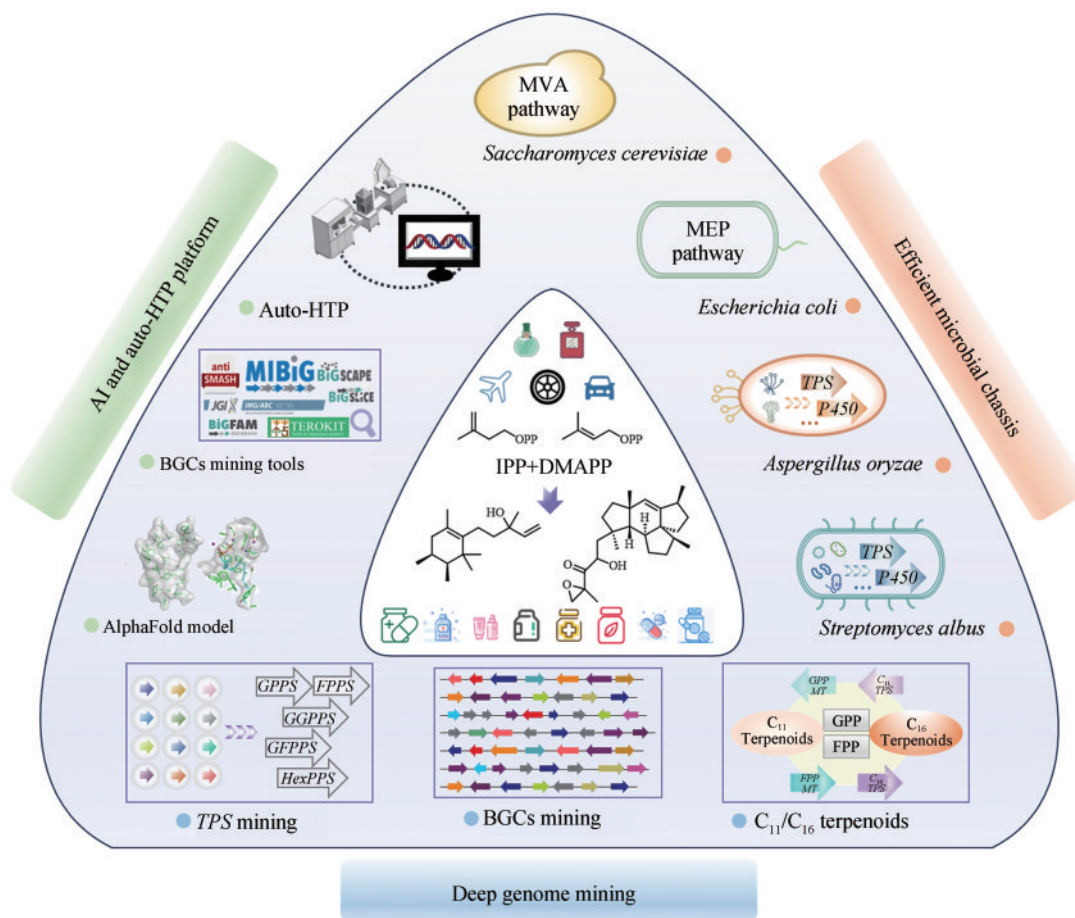
收稿日期: 2023-12-01 修回日期: 2024-02-22

基金项目: 国家重点研发计划 (2023YFA0916200)

引用本文: 雷茹, 陶慧, 刘天罡. 基因组深度挖掘驱动微生物萜类化合物高效发现[J]. 合成生物学, 2024, 5(3): 507-526

Citation: LEI Ru, TAO Hui, LIU Tiangang. Deep genome mining boosts the discovery of microbial terpenoids [J]. Synthetic Biology Journal, 2024, 5(3): 507-526

discovering new terpenes due to the increasing number of known compounds at large quantities. The advent of next-generation sequencing and synthetic biology technologies marks the onset of the era of genome mining-driven natural product discovery, particularly in the exploration of new terpenoids. However, challenges persist in this regard, such as low efficiencies, interference of known compounds, and limited data throughput. In this review, we focus on recent advances in terpenoid discovery *via* microbial genome mining strategies, including the use of the precursor supplying microbial chassis (*Escherichia coli*, *Saccharomyces cerevisiae*, *Aspergillus oryzae*, *Streptomyces albus*, etc.), the microbial resources from extreme geographical environments, deep genome mining, and terpene mining platforms driven by artificial intelligence and automation techniques. To produce more terpenoids using heterologous hosts, multiple microbial chassis with enhanced precursor supply have been developed to improve their production yields and thus facilitate the discovery of structurally unique terpenoids. With the growing understanding of terpene biosynthesis machinery, the deep mining of terpenoid biosynthetic gene clusters and terpene synthases can effectively address issues related to repeated and irrelevant discoveries. Furthermore, the integration of artificial intelligence and automation platform with synthetic biology has ushered in the high-throughput intelligent discovery of terpenoids, which significantly improves the research and enables the discovery of numerous terpenoids with new structures. Finally, we address challenges and future directions for genome mining based terpenoid discovery. Driven by synthetic biology and artificial intelligence, a new chapter for the discovery of terpenoids and other natural products will open. We are looking forward to seeing more terpenoids to be developed as drugs and valuable chemicals in the future.



Keywords: terpenoid natural products; terpene synthases; microbial chassis; genome mining; artificial intelligence; automated high-throughput platform

萜类化合物是一类化学结构上最多样化的天然产物家族，有超过18万个成员^[1]，它们广泛存在于自然界中，具有多种生理和生态功能，在食品、药品、化妆品、保健品、农业、生物能源与生物材料等领域有着广泛应用^[2-3]。

萜类化合物以异戊烯基焦磷酸（isopentenyl diphosphate, IPP）和二甲基烯丙基焦磷酸（dimethylallyl diphosphate, DMAPP）作为通用前体，在细菌、原生动植物和植物细胞的质体中，主要通过2-C-甲基-D-赤藓糖醇-4-磷酸（2-C-methyl-D-erythritol-4-phosphate, MEP）途径生成；而在真核生物和古菌中，主要通过甲羟戊酸（mevalonate, MVA）途径生成。异戊烯基转移酶（prenyltransferase, PT）将IPP和DMAPP组装成不同链长的异戊二烯单元——香叶基焦磷酸（geranyl pyrophosphate, GPP）、法尼基焦磷酸（farnesyl pyrophosphate, FPP）、香叶基香叶基焦磷酸（geranylgeranyl diphosphate, GGPP）、香叶基法尼基焦磷酸（geranylarnesyl pyrophosphate, GFPP）、六聚异戊烯基焦磷酸（hexaprenyl diphosphate, HexPP），并经萜类合酶（terpene synthase, TPS）作用生成具有多个手性中心的碳氢化合物核心骨架——单萜（monoterpenoid, C₁₀）、倍半萜（sesquiterpenoid, C₁₅）、二萜（diterpenoid, C₂₀）、二倍半萜（sesterterpenoid, C₂₅）、三萜（triterpenoid, C₃₀）、四萜（tetraterpenoid, C₄₀）以及多萜（polyterpenoid, >C₄₀），这些萜类化合物骨架在细胞色素P450氧化酶（cytochromes P450, CYP）、甲基转移酶、酰基转移酶、糖基转移酶等一系列后修饰酶的作用下进一步形成结构和功能上丰富多样的萜类化合物，如青蒿素（artemisinin）、紫杉醇（paclitaxel; taxol）、脱落酸（abscisic acid）、截短侧耳素（pleuromutilin）等^[4-7]。此外，人们还发现了一类非典型萜类化合物，它们通过非常规方式进行生物合成^[8-9]，主要包括C₁₁萜类化合物（如2-methylisborneol和2-methylenebornane）以及C₁₆萜类化合物（如sodorifen）^[10]。其中，C₁₁萜类前体为2-甲基-GPP，C₁₆萜类以索多芬基焦磷酸（presodorifenyl diphosphate, PSPP）为前体，两者由相应甲基转移酶对GPP或FPP进行甲基化修饰产生^[11-13]。常规萜类化合物和非常规萜类化合物共同组成海量萜类化合物库（图1）。

迄今为止，研究人员已开发了天然产物直接分离提取^[14-16]、沉默基因（簇）同源激活^[17-19]以及异源表达^[20-22]三大策略，将其用于微生物萜类化合物的挖掘。三大策略共同推进了萜类天然产物的大量发现，然而已表征的萜烯合酶和萜类化合物骨架仍然只是大自然巨大萜类宝库中的冰山一角。如何破解萜类化合物挖掘的“三低”瓶颈（产量低、结构新颖性低和研究通量低），并进一步开发出高效的萜类天然产物挖掘策略，成为目前研究的热点。

随着基因组学的发展，人们将挖掘对象从植物样本提取和微生物大规模发酵，一定程度上转移至萜类天然产物生物合成基因（簇），通过构建高效的前体供应底盘，可以促进萜类天然产物的异源合成；为了拓宽萜类天然产物的挖掘范畴，人们还将目光投向了深海、极地等极端地理环境中的微生物资源，为寻找天然产物提供了更多可能；同时，合成生物学与组合生物合成的迭代更新，也为天然产物的生物合成创建了新的路径，扩展了萜类的合成空间；高通量自动化平台与底盘的结合，显著提升了异源表达的研究通量，加速了萜类化合物的发现；人工智能的快速发展及其在合成生物学领域的应用，进一步提升了基因挖掘的准确度和效率，开创了天然产物的智能化挖掘模式，引领着更多萜类“暗物质”的创新发现。针对基因组挖掘驱动的微生物萜类天然产物发现研究，本文就如何实现新型萜类化合物的高效挖掘与创新发现展开探究，主要介绍高效前体供应微生物底盘、生物合成基因（簇）深度挖掘、人工智能与自动化平台在驱动萜类化合物高效发现方面取得的最新研究进展。

1 用于萜类化合物挖掘的高效微生物底盘

次级代谢产物生物合成基因簇（biosynthetic gene cluster, BGC）的异源表达驱动了系列天然产物的发现，有效地规避了同源激活策略存在的缺陷及本底菌株难获得的问题，被广泛地应用于天然产物挖掘及生物合成研究^[20]。异源表达通常选择具有完善遗传操作系统以及明确遗传背景的微

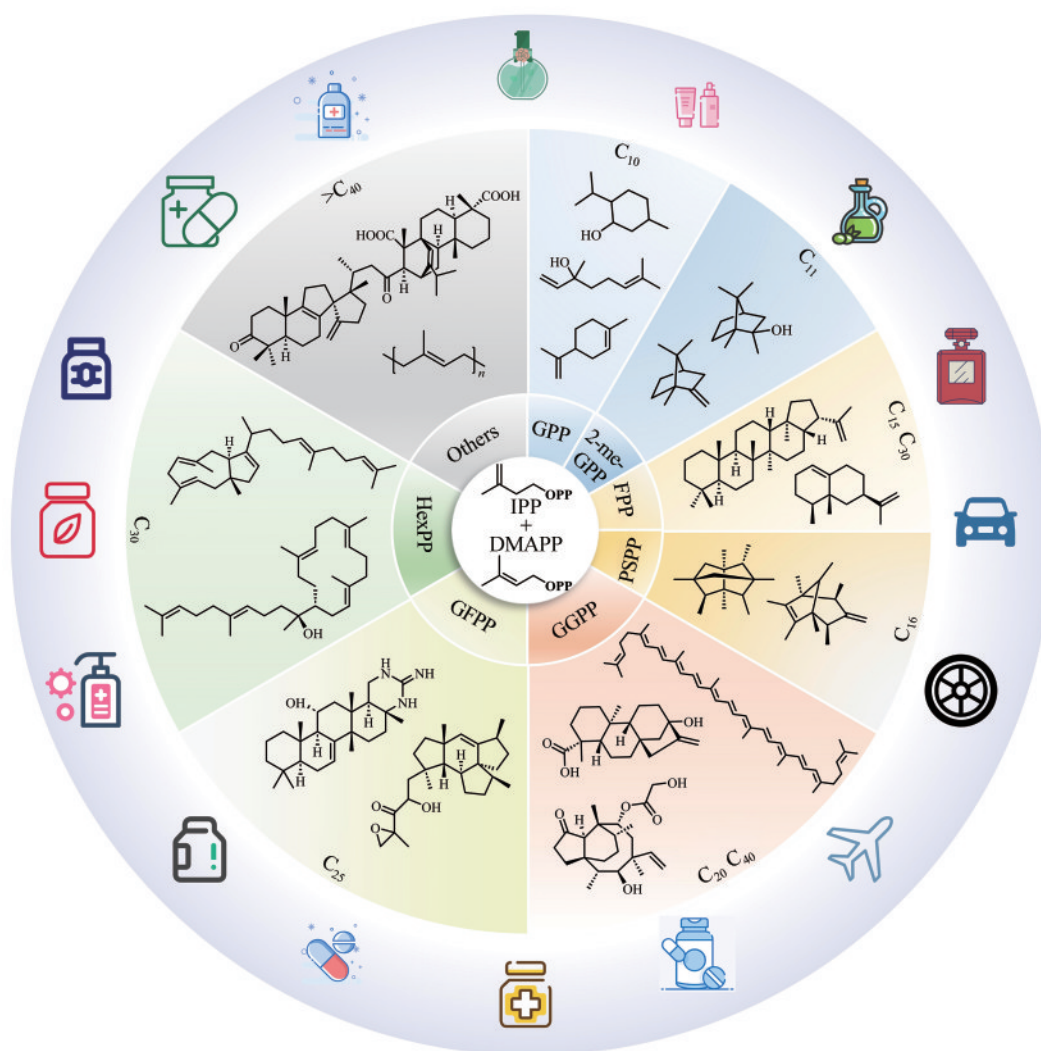


图1 萜类化合物的种类及代表性萜类化合物

Fig. 1 Types and structures of representative terpenoids

生物作为底盘细胞，通过异源表达各生物合成元件或基因（簇）来阐明其功能。受限于异源表达宿主的合成天然产物的能力，仅能对其中高丰度的产物进行获取和结构鉴定，而低丰度的化合物在很大程度上难以获取或被忽略^[23]。因此，构建高效的微生物底盘细胞，可以在微生物体内进行代谢流的重新分配，进而提高目标产物的滴度，实现萜类化合物的绿色制造。

国内外多个课题组在微生物萜类化合物底盘开发方面取得了突破性进展。如刘天罡教授团队创新性地提出“定向合成代谢”理念^[24]，通过合成途径的体外优化结果指导体内MVA代谢途径的改造，搭建了高效供应IPP和DMAPP前体的大肠杆菌（*Escherichia coli*）底盘，实现了法尼烯^[24]、

番茄红素^[25]、虾青素^[26]、紫杉二烯^[27]等萜类化合物的高效合成，充分证明了前体高效供给底盘的通用性和高效性。在此基础上，团队进一步提出“基于微生物高产底盘的天然产物创新挖掘”策略^[28]，阐述了“一类钥匙开一把锁”的研究理念和“近源性、同类性和完备性”的原则，通过使用近源宿主解决基因（簇）异源表达适配性差和产量低的问题^[29]，并相继构建了大肠杆菌^[30]、酿酒酵母（*Saccharomyces cerevisiae*）^[31-33]、丝状真菌米曲霉（*Aspergillus oryzae*）^[29]等萜类前体高效供应底盘，加速了单萜、倍半萜、二萜、二倍半萜、三萜等的发掘和生物合成机理解析。其他课题组也相应开发了毕赤酵母（*Komagataella phaffii*，原名 *Pichia pastoris*）^[34-35]、解脂耶氏酵母（*Yarrowia*

lipolytica)^[4, 36-39]、构巢曲霉 (*Aspergillus nidulans*)^[40]、白色链霉菌 (*Streptomyces albus*)^[41-43] 等微生物底盘, 实现了一系列高价值萜类化合物的发现和高产, 下文将针对其中主要的萜类合成微生物底盘及典型应用进行介绍, 并将主要微生物底盘优缺点汇总于表1。

1.1 大肠杆菌底盘

大肠杆菌因其遗传操作简单、培养周期短等优势, 在萜类化合物挖掘中备受青睐, 采用过表达大肠杆菌内源MEP途径中的关键酶^[44]或引入异源MVA途径的策略, 大幅加快了新型萜类化合物生物合成元件及产物的挖掘进度。如Bian等^[30]通过对不同异戊二烯前体合成模块和多功能萜类合酶模块进行组合, 在大肠杆菌中搭建了一个基于MVA途径的高效萜类合成平台, 充分释放了丝状真菌来源二萜合酶FgGS和二倍半萜合酶FgMS的合成潜能, 检测到包括多个新骨架在内的单萜、倍半萜、二萜以及二倍半萜化合物50种, 鉴定了8个新的萜类化合物。Reddy等^[45]在引入MVA途

径的大肠杆菌底盘中, 对不同细菌来源的22个潜在的I型TPS进行筛选, 鉴定了15个细菌来源TPS的功能, 具有单萜合酶、倍半萜合酶和二萜合酶活性和广泛的底物宽泛性。

在大肠杆菌中异源表达萜类生物合成基因, 特别是CYP时, 往往面临着很大的挑战, 我们应该如何改造大肠杆菌底盘, 使其可以用于挖掘包含CYP的萜类BGC呢? 最近Lin等^[46]给出了一个解决方案: 通过在大肠杆菌中引入与前体供应、电子传递和调控相关的31个重组基因, 使其拥有可以适配不同CYP的氧化还原电子传递系统, 用于挖掘萜类CYP及其氧化产物。使用该平台表征位于萜类环化酶基因附近的64种细菌CYP, 发现了40个具有功能的CYP, 将它们与17个萜类环化酶进行组合, 构建了1088条合成途径, 产生了大量氧化的萜类化合物。值得一提的是, 对比植物中可以修饰*ent-kaurene*骨架的44种酶, 其中16种可以被新发现的细菌CYP替代。这些成果彰显了该平台在挖掘细菌萜类CYP与合成萜类衍生物方面的强大应用潜力, 可以用于高通量探索萜类化合物的化学结构多样性。

表1 用于萜类化合物挖掘的主要微生物底盘

Table 1 Major microbial chassis for terpenoid mining

主要微生物底盘	优点	缺点
<i>Escherichia coli</i>	遗传操作简单 培养周期短 具有内源MEP途径	不具有跨膜结构域, 可能导致功能蛋白的错误折叠、失去活性甚至降解 不适合表达真菌和植物来源的萜类合成基因(簇)
<i>Saccharomyces cerevisiae</i>	遗传操作相对简单 重组效率高 具有完整的细胞器与内膜系统 具有内源MVA途径	存在密码子偏好性 无法正确识别和剪切内含子 不适合表达真菌来源的萜类合成基因(簇)
<i>Aspergillus oryzae</i>	具有强大的蛋白分泌特性 功能酶后修饰能力强大 可以正确地识别和剪切内含子 具有内源MVA途径 适合表达真菌或植物来源的萜类合成基因(簇)	遗传操作复杂 培养周期长
<i>Streptomyces albus</i>	功能酶后修饰能力强大 具有内源MEP途径 适合表达细菌来源的萜类合成基因(簇)	遗传操作复杂 培养周期长

注: 大肠杆菌和白色链霉菌具有内源MEP途径, 前者多用来表达细菌来源的萜类合酶, 后者适合表达细菌来源的萜类合成基因(簇); 酿酒酵母和米曲霉具有内源MVA途径, 前者多用来表达真菌或植物来源的萜类合酶, 后者适合表达真菌或植物来源的萜类合成基因簇。

Note: *Escherichia coli* and *Streptomyces albus* possess an endogenous MEP pathway. The former is mainly used to express terpene synthases from bacteria, and the latter is suitable for expressing terpene biosynthetic gene clusters from bacteria. Both *Saccharomyces cerevisiae* and *Aspergillus oryzae* possess an endogenous MVA pathway. The former is mainly used to express terpene synthases from fungi or plants, and the latter is suitable for expressing terpene biosynthetic gene clusters from fungi or plants.

1.2 酿酒酵母底盘

酿酒酵母具有重组效率高、食用安全等特点，是萜类合成的强大平台^[47]。酿酒酵母具有内源的MVA途径，因此仅调节关键酶的表达强度，如增加主要限速酶HMGR拷贝数，即能显著提高前体供应能力。Bian等^[31]在酿酒酵母中过表达*tHMG1*，构建了倍半萜底盘YZL141，随后过表达*ERG20*（FPP合成酶基因）用于探究禾谷镰刀菌来源倍半萜合成酶FgJ03939（FgFS）的合成潜力，最终分离得到8个倍半萜化合物，其中包括3个5/7双元环和5/6/3三元环稀有倍半萜新骨架化合物，进一步丰富了倍半萜化合物的骨架组成；该底盘还被用于研究嵌合二萜合酶CgDS的合成潜能^[32]，在此基础上首次表征了丝状真菌来源Clade II-D进化分支萜类合酶的C1-III-IV产物环化模式。Chen等^[48]进一步开发YZL141底盘应用潜能，通过引入GFPP合酶（GFPPS）搭建了一个二倍半萜酵母底盘，使用连接肽GSTGS将嵌合型萜类合酶PTTS与GFPPS融合，增强二倍半萜合酶代谢通量，成功实现了34个活性PTTS的鉴定，挖掘到两种新型环状二倍半萜化合物。

基于酿酒酵母内源MVA途径全面强化的高产底盘，可以快速筛选萜类合酶，实现低滴度化合物的异源高效合成。如李晓东等^[49]通过在酵母基因组串联整合MVA途径，构建了倍半萜高效底盘，随后异源表达真菌来源倍半萜合酶ec38-cs，实现了 α -新丁香三环烯在酵母中的首次从头合成。Siemon等^[33]在酿酒酵母中系统表达整个MVA途径，构建了高产前体IPP和DMAPP的酵母底盘JCR27，通过异源表达禾谷镰刀菌来源倍半萜合酶FgJ02895并增加拷贝数，高效合成了骨架化合物guaia-6,10(14)-diene，与有机合成方法相结合，驱动了具有显著抗肾癌活性的倍半萜明星分子englerin A的高效合成，实现当时报道的最高产率。Dong等^[50]首先调整关键基因表达，将*tHMG1*整合至酵母染色体上，敲除旁路角鲨烯合酶基因*ERG9*得到酵母基底DH02，并进一步将*ERG20-mERG20*（*ERG20* F96C突变，具有GGPPS功能）融合基因整合到DH02染色体中，构建了一个高效供给GGPP的二萜底盘。Mukherjee等^[51]基于过氧化物酶体富含乙酰辅酶A的特性，在酿酒酵母中

通过靶向定位MVA途径到细胞质和过氧化物酶体，建立了双定位协同的酵母底盘，可应用于单萜、倍半萜、三萜等萜类合成。Ma等^[52]通过前体加强、诱导调控等策略构建了GPP、FPP和GGPP的多前体萜类通用酵母合成平台。这些强化前体供应底盘的构建与优化，为萜类化合物的批量化挖掘奠定了良好基础。

1.3 丝状真菌底盘

丝状真菌具有可以正确识别和剪切真菌来源内含子的能力，蛋白分泌能力强大，是一个适合真菌来源萜类生物合成基因（簇）异源表达的真核宿主，在工业生产方面具有广泛的应用，如黑曲霉（*Aspergillus niger*）被用于有机酸和酶制剂的工业发酵^[53-54]，土曲霉（*Aspergillus terreus*）是降血脂药物洛伐他汀的工业生产菌^[55-56]，米曲霉常在酿造行业被用来制作酱油、清酒等食品^[57]。因此，打造一个强化萜类前体合成的丝状真菌底盘，用于高效挖掘真菌来源的萜类化合物，是一个具有较好应用前景的基因挖掘策略。Yuan等^[29]以具有合成萜类及其衍生物天然优势的*A. oryzae* NSAR1为宿主，将MVA途径和限速步骤*tHMG1*基因在基因高表达位点进行加强表达，搭建了两个通用且高效的*A. oryzae*萜类高产底盘菌株*A. oryzae* AO-S85（基因多拷贝随机插入型）和*A. oryzae* AO-S95（基因高表达位点同源重组型）。该策略可以突破底盘适配性问题困扰，便于其他真菌来源萜类化合物的快速挖掘和高效合成。Lin等^[40]在构巢曲霉*A. nidulans*中建立了一个天然产物异源表达平台，首先通过敲除8个高表达BGC以降低复杂次级代谢物背景，再加入7个可选择的标记以促进基因筛选，开发了遗传去重复菌株*A. nidulans* LO11098，该菌株具备异源表达烟曲霉来源聚酮二萜BGC和其他物种中沉默BGC的潜力。

1.4 链霉菌底盘

链霉菌是最具代表性的放线菌之一，具有产生丰富的天然产物的能力^[43, 58-59]，被广泛应用于以氨基酸为基础的天然产物生产^[60]。随着复杂的基因编辑系统在模式链霉菌中的建立，链霉菌也被

开发为萜类化合物合成底盘，用于挖掘新的细菌萜类化合物^[43]。其中，白色链霉菌、变铅青链霉菌 (*Streptomyces lividans*) 等被广泛用作异源表达宿主^[61]。Myronovsky 等^[62]以通用菌株 *S. albus* J1074 为出发菌株，通过 14 个步骤删除 15 个 BGC，得到了一个具有干净次级代谢背景的底盘 *S. albus* Del14，并通过增加 *phiC31* 整合位点 (*attB* 位点) 的数量来提高基因整合效率，是用于异源表达生产天然产物的新一代优良底盘。类似地，Ahmed 等^[60]敲除了菌株 *S. lividans* TK24 的 9 个 BGC 得到 *S. lividans* ΔYA9，是链霉菌底盘的又一有益补充。以上具有简化代谢背景的改良底盘，有助于产物的检测和纯化，为天然产物的高效发现奠定了基础。而 Hu 等^[42]则探索将 *S. albus* J1074 打造为一个更适用于萜类化合物生产的通用底盘。首先利用番茄红素的产量水平直观评价四种链霉菌底盘对萜类前体的供应能力，并探究了发酵条件和启动子对萜类产量的影响，结果表明链霉菌底盘能够提供足够的萜类合成前体。研究团队最终基于白色链霉菌 *S. albus* J1074，通过启动子强度调控建立了一套简便高效的细菌萜类 BGC 表达平台 *S. albus* J1074M，不仅可以用于细菌来源萜类 BGC 高效表达，还可以用于细胞色素 P450 等萜类后修饰基因的表达，是未来探索细菌萜类化合物的重要底盘。

2 基因组深度挖掘驱动萜类化合物系统发现

随着基因组学和生物信息学的兴起和蓬勃发展，微生物天然产物挖掘从依赖生物样本的自然筛选发现阶段，进阶到组学大数据驱动的智能筛选新征程。根据在线基因数据库 JGI GOLD (<https://gold.jgi.doe.gov/>)，截至 2023 年 11 月 30 日，已有 418 036 种细菌、47 625 种真核生物、5099 种古细菌等共计 488 845 种生物体的基因组测序工作完成或正在进行^[63]。随着“后基因组时代”的到来，基因组挖掘助力了大量萜类化合物及相关生物合成 BGC 的发现与表征，取得了一系列突破性进展，不仅加速了天然产物挖掘进程，还为萜类乃至其他天然产物的研究提供了系统高效的新方案。萜

类化合物的结构多样性由种类和功能多样的酶决定，首先在核心萜类合酶的作用下形成萜类化合物核心骨架，随后经系列后修饰酶的作用后，形成结构复杂多样的萜类化合物。基于萜类生物合成的特点，人们不仅针对核心酶和其所在 BGC 开展了基因组深度挖掘，还对非常规萜类和特殊生境萜类化合物开展了高效挖掘，发现了大量新颖的萜类化合物。

2.1 萜类合酶的精准挖掘

萜类化合物的化学结构新颖性在很大程度上取决于核心萜类合酶，因此基于核心酶序列的微生物基因组挖掘，可以快速实现具有新颖结构的萜类化合物的发现。如 Yang 等^[64]以鸟巢烷二萜环化酶 EriG 蛋白序列为探针进行基因组挖掘，发现了在细菌和真菌中广泛存在的 UbiA 类二萜环化酶新家族，并且成功鉴定了该家族中七个新的二萜环化酶，还确定了相应的产物结构，其中包含新骨架二萜化合物 lydicene。Sun 等^[65]从丝状真菌绿色木霉 (*Trichoderma viride*) 中发现了一种新颖的 I 型倍半萜合成酶 Tvi09626，其产物是一种新颖 Brasilane 型 5/6 双环倍半萜化合物。Chen 等^[66]完成担子菌白囊耙齿菌 (*Irpex lacteus*) 的全基因组测序，通过生物信息学分析从中锁定了一个潜在的 tremulane 倍半萜合成酶 *III*S，其在米曲霉中异源表达可以产生四种新的倍半萜 iltremulanols A~D，该工作填补了 tremulane 骨架生物合成研究的空白。Li 等^[67]通过基因组挖掘，基于序列相似性网络分析 (sequence similarity network, SSN) 在白网链霉菌 (*Streptomyces albireticuli*) 中发现了自然界中首个反式 eunicellane 萜类合酶 AlbS，进一步利用 AlbS 和顺式 eunicellane 萜类合酶 Bnd4^[68]进行标记实验，揭示了细菌中顺式和反式 eunicellane 形成机理的区别。

随着海量基因组数据的产生和数字化实验数据的实时共享，研究人员不再局限于挖掘单个物种的基因组数据，而是从数以万计的基因组等数据库中大规模批量化分析目标基因，极大程度提高了研究通量。Chen 等^[48]筛选了来自 NCBI、Uniprot 和 JGI 数据库中已经完成基因组测序的 477 株真菌，从中筛选出了 227 个嵌合型萜类合酶

PTTS; 借助二倍半萜前体高效供给的酿酒酵母底盘, 并结合自动化高通量 (automated high-throughput, Auto-HTP) 平台, 批量挖掘了34个PTTS, 成功鉴定了24个二萜和二倍半萜产物, 其中包括两个结构新颖的二倍半萜化合物 *sesterevisene* 和 *sesterorbiculene*, 如图2所示。值得关注的是, 在此项工作之前, 仅有大约20个PTTS进行了功能表征, 因此该研究挖掘的PTTS数量超过了此前已知的PTTS数量。这

些基于萜类合酶的基因组挖掘成果, 不仅凸显了微生物基因组是新型萜类化合物的宝贵来源, 丰富了萜类化合物的资源库, 还进一步拓展了我们对于萜类合酶的认识。

2.2 萜类BGC的深度挖掘

萜类化合物的BGC不仅包含萜类合酶, 还拥有多种萜类骨架后修饰酶, 其中萜类合酶决定萜

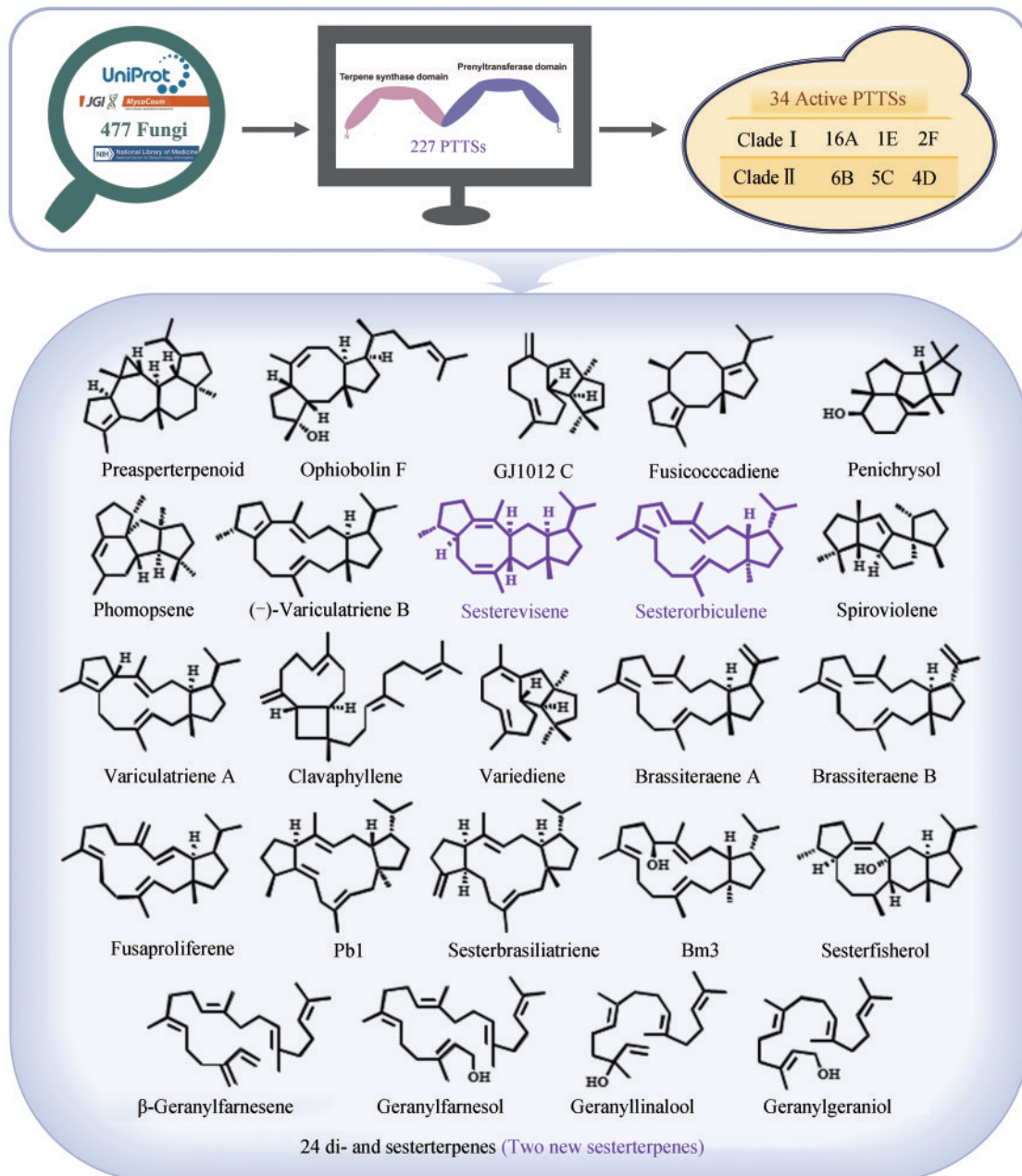
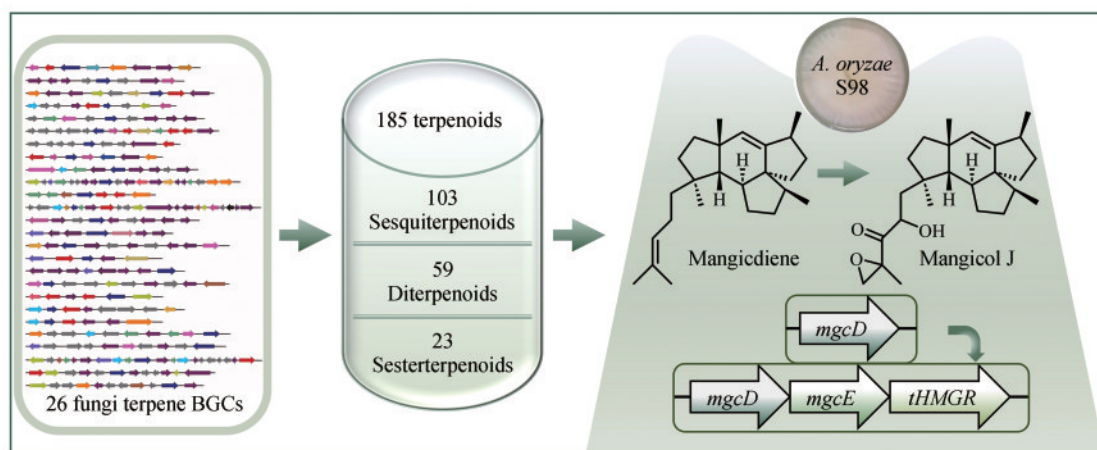


图2 基于批量挖掘真菌嵌合型萜类合酶PTTS发现24种萜类化合物^[48]

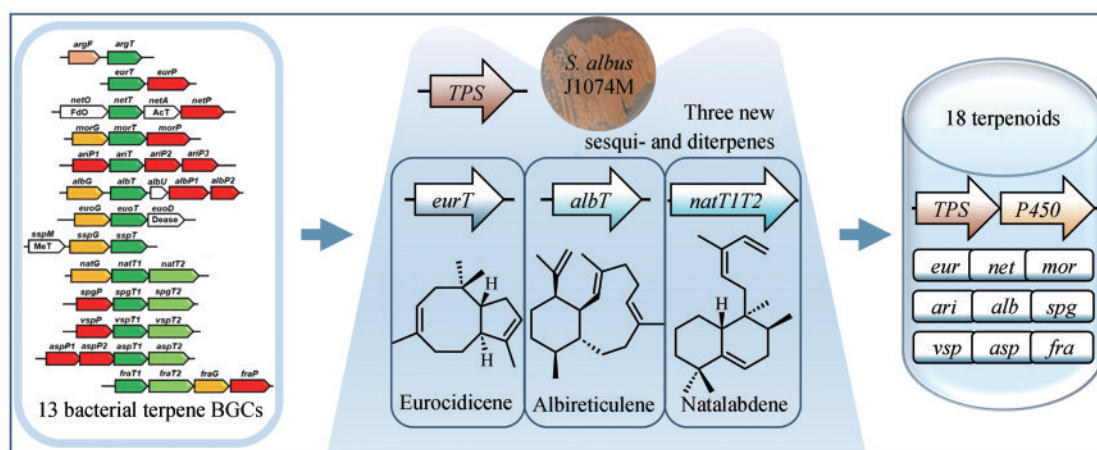
Fig. 2 Discovery of 24 terpenoids through the systematic mining of fungal chimeric terpene synthases PTTS^[48]

类核心骨架的结构，后修饰酶则进一步增加其结构的多样性，因此基于萜类合酶对其所在BGC开展深度挖掘可以显著提升萜类化合物的多样性。微生物基因组富含大量萜类BGC，人们不仅可以挖掘单个萜类BGC发现新骨架萜类化合物，还可以针对多个基因组数据开展BGC的规模化挖掘，批量发现萜类化合物。如Wang等^[69]在嗜热放线菌(*Amycolatopsis arida*)中发现了一个BGC-*ari*，在模式链霉菌中异源表达，得到了三种具有6/7/5三环结构的新颖eunicellane型二萜衍生物aridacins A~C；Li等^[70]在委内瑞拉链霉菌(*Streptomyces venezuela*)中鉴定了一个萜类BGC-*ven*，经异源表达后发现了两种二萜产物Venezuelaenes A和B，均

具有结构新颖的5/5/6/7四环骨架，揭示了单个萜类BGC的合成潜力。Yuan等^[29]对五株基因组测序的丝状真菌进行生物信息学分析，通过I型TPS保守结构域DDXXD/E和NSE/DTE，对含有不同生物合成基因的萜类BGC进行筛选。通过在米曲霉底盘中进行模块化组合重构了5种真菌来源的39个I型萜类BGC(26个BGC有产物)，获得了185个萜类化合物，其中包含103个倍半萜、59个二萜和23个二倍半萜。该工作有效地解决了产物集中度低的问题，不仅构建了一个具有丰富产物结构的萜类化合物库，还分离得到了一系列mangicol类产物(mangicols H~L)，其中二倍半萜mangicol J具有高效抗炎活性，开发的潜力巨大[图3(a)]。



(a) 26个丝状真菌萜类BGC的批量挖掘与其中二倍半萜mangicol J的发现^[29]
(a) Large-scale mining of 26 terpene BGCs from filamentous fungi and the discovery of mangicol J^[29]



(b) 13个细菌萜类BGC的批量挖掘与3个萜类新骨架的发现^[42]
(b) Large-scale mining of 13 terpene BGCs from bacteria and the discovery of three new terpene skeletons^[42]

图3 基于基因组数据的萜类BGC规模化挖掘

Fig. 3 Large-scale mining of terpene BGCs based on the genomic data

相比于真菌，人们较少发现来源于细菌的萜类天然产物，但随着细菌基因组信息爆发式的增长，大量功能未知的萜类BGC浮出水面，因此人们将目光投向了挖掘细菌来源的萜类化合物，如Hu等^[42]使用已知萜类合酶的隐马尔可夫模型(hidden Markov model, HMM)为探针，从282株细菌的基因组中确定了756个潜在的萜类合酶，进一步基于序列相似性网络SSN分析，锁定了16个萜类BGC(13个BGC有产物)，随后在高效萜类前体供应底盘*S. albus*中对其进行表征，通过产物的分离和结构鉴定，得到共计31个萜类化合物，包括3个新颖的萜类骨架分子和18个新颖的萜类衍生物，进一步丰富了细菌萜类化合物库[图3(b)]。

2.3 非常规萜类化合物的创新挖掘

应用合成生物学原理，将不同生物合成模块进行组合生物合成，还可以构建非天然生物合成途径，实现萜类化合物的人工智造。如Tang等^[71]将来自不同真菌的 $\alpha\beta$ -佛手柑素合酶与多功能P450单加氧酶进行组合生物合成，得到了一系列

具有新骨架的倍半萜类化合物。Tsukada等^[72]通过基因组挖掘在5个真菌基因组中发现了5个潜在合成含萜烷的二萜吡咯酮化合物(DDP)的BGC，随后通过在米曲霉中逐步异源表达生物合成基因，重建了5条天然途径和1条分流途径，并且通过组合生物合成构建了4条非天然拓展途径，成功得到了22种DDP，其中15个为新化合物，为扩大萜类化合物的化学多样性提供了新路径。

Duan等^[73]针对C₁₆萜类化合物sodorifen展开研究，发现其BGC至少包含一个萜类合酶TPS和一个甲基转移酶MT，在C-甲基转移酶*SpSodMT*作用下，FPP进行甲基化修饰合成C₁₆前体PSPP，并进一步在*SpSodTPS*作用下生成sodorife。因此作者将*SpSodMT*和*SpSodTPS*的蛋白序列作为探针，在NCBI RefSeq基因组数据库中进行搜索，发现739种已测序细菌基因组中的BGC都编码*SpSodMT*与*SpSodTPS*同源蛋白；为了研究这些BGC的功能，从上述基因簇中随机选取了40个BGC在强化FPP前体供应的酿酒酵母底盘进行异源表达，发现所有可溶的MT都会产生相同的C₁₆前体PSPP，其在18种TPS的进一步催化下，可以形成47种C₁₆萜类化合物，如图4所示。这类不同

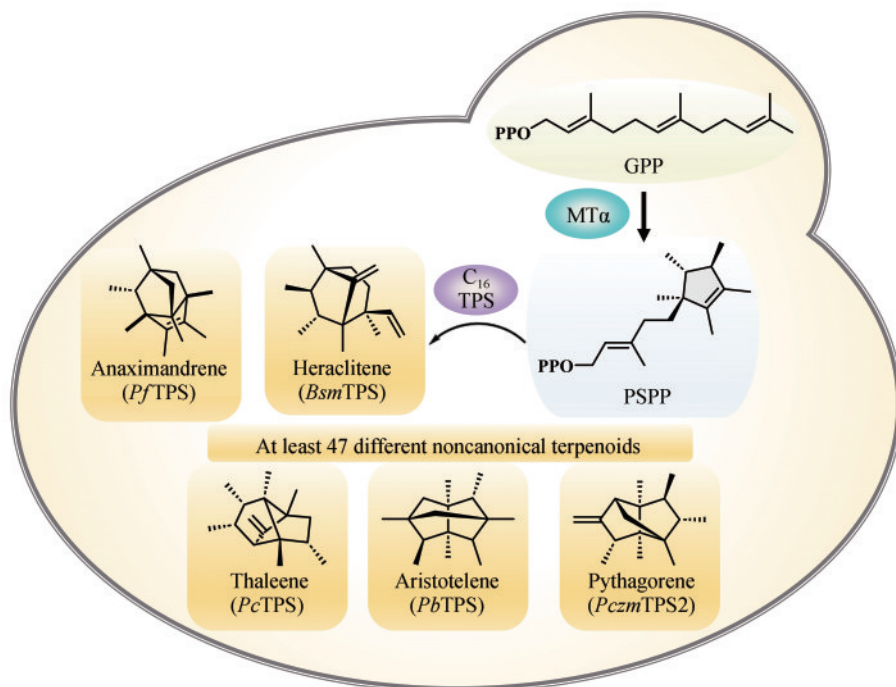


图4 基于酵母底盘挖掘细菌来源非常规C₁₆萜类化合物^[73]

Fig. 4 Yeast chassis for discovering noncanonical C₁₆ terpenes from bacteria^[73]

寻常的非经典 C_{16} 萜类化合物大多数具有复杂且新颖的双环和三环骨架，充分展示了细菌中非常规碳前体模块在扩展萜烯生物合成方面的巨大潜力。

2.4 特殊生境萜类化合物的挖掘

相比于陆地等普通生境，深海、沙漠、极地、盐沼、热液喷口、冷泉、冰川和永久冻土等极端环境^[74]为微生物提供了特殊的生态环境，微生物已经进化出区别于普通生境微生物的独特代谢机制来应对极端的环境压力^[75]，这些代谢途径可产生多种结构特殊、生物活性显著、药理性质优良天然产物^[76-79]。其中，海洋萜类化合物具有优良的药用特性，特别是抗菌^[15]、抗炎^[80]和抗癌^[81]活性。

海洋是天然产物的巨大宝库，Carroll等连续多年对每年发表的有关海洋天然产物的文献进行综述，如2021年报告的416篇文献中描述了1425个新化合物^[82]，2020年报告的420篇文献中描述了1407个新化合物^[83]，2019年报告的440篇论文中有1490个新化合物^[84]，可见海洋微生物是结构和活性新颖萜类等天然产物的丰富来源，然而目前关于海洋天然产物的研究表明海洋环境中微生物的多样性及天然产物的丰富性还有待进一步探究。基因组挖掘使这些“微生物暗物质”^[85]重见光明，如Guo等^[86]在研究中获得了一株海洋真菌 *Aspergillus ustus* 094102，该菌株代谢产生大量的萜类化合物，包含具有良好抗肿瘤耐药活性的二倍半萜蛇孢菌素。进一步分析发现该菌株基因组序列中还存在包含 BGC-*Au11189* 在内的多个双功能萜类合成基因（簇），随后在米曲霉中对双功能萜类合酶 *Au11189* 及其上游细胞色素 P450 酶 *Au11188* 进行异源表达，共分离鉴定了9个化合物，全部为新颖的二倍半萜结构，其中2个具有抗肿瘤活性，为新药研究提供了新的先导物。此外，Zhang等^[87]从海洋来源的真菌黄曲霉 (*Aspergillus flavipes*) 中鉴定出一个 I 型双功能二萜合成酶 *tndC*，并在酿酒酵母中表征了该酶，发现了具有 5-8-6 稠环结构的 *talarodiene* 型萜类新骨架。这些研究成果不仅展示了海洋微生物是新颖萜类化合物的重要来源，还暗示着深海等极端环境仍有待进一步挖掘和探索。

3 萜类化合物的智能发现与自动化批量发现

3.1 工程改造萜类合酶拓展萜类化合物的多样性

通过多轮突变和筛选进行酶的定向进化 (directed evolution)，以及依据序列和结构信息对酶的关键位点进行理性设计 (rational design)，可以实现对萜类合酶的工程改造，进而产生新的萜类化合物^[88-89]。如Li等^[90]基于蛋白结构对二萜合酶 *VenA* 进行了理性改造，使其拥有可以容纳更长底物的活性口袋，将 *VenA* 的底物谱从 GPP、FPP 和 GGPP 进一步拓展到 GFPP，获得了包含单萜、倍半萜、二萜以及二倍半萜在内的多个萜类化合物，对萜烯合酶的催化能力进行了丰富和扩展。

哥本哈根大学的 Kampranis 教授团队在非常规萜类化合物合成方面进行了探究，他们使用非常规模块扩展了萜类的生物合成。2018年，Ignea等^[91]通过融合表达 ERG20p 和 PIGPPMT (假单胞菌 GPP 甲基转移酶) 合成了 C_{11} 前体 2-甲基-GPP，进一步通过定点突变将单萜合酶转换为特异性合成 C_{11} 非常规萜类的专用酶，在酵母中成功地合成了40个 C_{11} 萜类化合物。2022年，Ignea等^[92]借助相似的研究思路，在酵母中成功地合成了一系列 C_{16} 非天然萜类分子。他们首先将甲基转移酶 *SpSodMT* 引入在工程酵母中，建立 FPP 的甲基化途径；随后通过点突变方法进行蛋白质理性设计，构建了一个包含55种不同 *SpSodMT* 单残基突变体文库；这些模块进一步被萜烯合酶识别用于构建非常规 C_{16} 萜烯骨架，再经由细胞色素 P450 酶修饰，最终合成了28种具有不同结构和不同氧化程度的非常规 C_{16} 萜类化合物，显著地增加了萜类化合物的结构多样性。

3.2 自动化高通量平台助力萜类化合物的批量挖掘

在 BGC 批量化挖掘和酶工程改造规模化筛选的背景下，海量的工程化试错实验需求和传统的劳动密集型研究弊端凸显，自动化高通量平台在合成生物学的推广应用也应运而生。2014年，赵惠民教授团队参与设计、搭建、运行了全自动合

成生物设施 iBioFAB; 2019年, 该团队将 iBioFAB 自动化系统与机器学习算法相结合形成 BioAutomata 系统, 显著提升了番茄红素产生菌株的筛选效率^[93]; 2022年, 该团队进一步搭建了多功能自动化 PlatmidMake 平台, 以高通量的方式完成质粒无痕构建^[94]。

自动化高通量平台打破萜类化合物研究的传统模式, 显著提升了挖掘通量和效率, 如 Leferink 等^[95]搭建了一种适用于筛选不同的单萜合酶文库的 Auto-HTP 工作站, 通过使用与 GC-MS 耦合的机器人液体处理平台, 可以实现数据的自动提取分析, 是对现有萜类合酶定向进化高通量筛选的有益补充。依托高效的 Auto-HTP 平台, Yuan 等^[29]进行了质粒和菌株的批量化构建, 大大加快了实验进程, 经过两轮自动化构建, 实现了批量扩增 580 个片段、构建 173 个质粒和 208 个米曲霉突变株。Yuan 等^[96]在对羊毛甾烷型四环三萜类化合物灵芝酸 (GA) 的研究中, 建立了一个高通量自动化迭代筛选平台, 通过 CYP 表达载体构建、功能性 CYP 高通量体内筛选、酵母宿主代谢工程这三个模块对 158 个灵芝来源的 CYP 进行功能鉴定, 发现了多个可以形成 II 型灵芝酸的关键 CYP, 破解了 GA 生物合成的关键节点, 展示了高通量筛选平台的巨大应用潜力。目前全球范围内已建成或在建多个大型工程化自动化设施平台, 用于合成生物学研究支持和应用变革, 如国外的 Amyris 公司、Ginkgo Bioworks 公司、Zymergen 公司、Transcriptic 公司等^[97]; 国内有浙江大学杭州国际科创中心生物与分子智造研究院, 已经建设完成并上线合成生物学自动化装置 iBioFoundry^[98]; 中国科学院天津工业生物技术研究所牵头建设的国家合成生物技术创新中心, 以及中国科学院深圳先进技术研究院正在牵头建设的“深圳合成生物研究重大科技基础设施”等^[99], 这些高通量自动化铸造平台未来在萜类等天然产物挖掘方面必将大放异彩。

3.3 萜类 BGC 新型挖掘工具和方法的更新换代

微生物拥有丰富的次级代谢途径和大量相关的 BGC, 而如何从大量候选基因中定位到感兴趣的基因则具有挑战性。生物信息学研究结果显示,

自然界中尚有海量“沉默的”BGC 有待挖掘。伴随着高通量测序方法的发展和海量基因组数据的产生, 各种基因组挖掘方法和工具也被开发, 以快速发现和表征负责合成新型萜类产物的隐藏 BGC, 打开萜类化合物发现之旅的新大门^[100-101]。

作为最庞大的天然产物类群, 萜类化合物的化学文库、生物合成途径及其应用方面的知识仍然是有限并且彼此缺乏关联的, 专门用于研究萜类化合物的工具较少, 这严重阻碍了萜类化合物的发现和开发, 全面的萜类化合物数据库亟待开发。巫瑞波教授团队在 2020 年建立了萜类化合物综合数据库 TeroKit^[1], 为萜类化合物的化学和生物学研究提供了广阔的平台。2022 年, 该团队基于 TeroKit 还开发了萜类化合物数据库 TeroMOL^[102], 用于收集萜类及其衍生物, 以可视化方式分析萜类化合物的骨架和生物来源。2023 年, 此团队进一步开发了 TeroENZ 和 TeroMAP 模块^[103], 前者收集了全面的萜类生物合成相关酶的综合数据, 后者可用于萜类化合物生源代谢网络的可视化分析, 两者均已整合至 TeroKit 平台。截至 2023 年 11 月, TeroKit 已经收录了 18 万多个萜类化合物分子、1.3 万多个萜类相关酶、4200 多个反应、1.4 万多个生物来源、3900 多个靶点信息, 全面反映萜类化合物的化学空间、生物活性和生物合成途径。

现将微生物萜类 BGC 挖掘和核心酶预测常用工具和数据库汇总于表 2。

3.4 智能学习驱动更多萜类“暗物质”的解锁

近年来, 机器学习、深度学习等人工智能概念正在为合成生物学领域注入新活力, 也为萜类等天然产物带来底层理解、途径设计和创新挖掘的新变革^[125-126]。其中 AlphaFold 是一种基于机器学习的新颖方法, 将蛋白质结构的物理和生物知识结合起来, 利用多序列比对设计深度学习算法^[127]。从 2018 诞生到 2020 年 AlphaFold 2 版本迭代, 再到 2023 年 AlphaFold 3 新版本发布, AlphaFold 驱动蛋白质折叠预测工具, 为大分子结构设计提供了强大的驱动力, 在萜类等天然产物的挖掘方面也具有巨大应用潜力。

刘天罡教授团队在利用高效酿酒酵母底盘挖

表2 萜类BGC挖掘的网络工具和数据库^[104-106]
 Table 2 Tools and databases for mining terpene BGCs^[104-106]

数据库或网络工具	简介	网址	参考文献
BGC 数据库			
antiSMASH Database	细菌、古细菌和真菌等基因组中可检测的高质量 BGC 集合	https://antismash-db.secondarymetabolites.org/	[107]
BiG-FAM	微生物生物合成基因簇家族(GCF)模型集合	https://bigfam.bioinformatics.nl/home	[108]
IMG-ABC	微生物 BGC 的集成基因组图谱	https://img.jgi.doe.gov/cgi-bin/abc-public/main.cgi	[109]
MIBiG	已知生物合成基因簇及其分子产物的信息集合	https://mibig.secondarymetabolites.org/	[110]
BGC 网络识别工具			
antiSMASH	用于自动识别和分析微生物次级代谢产物基因簇	https://antismash.secondarymetabolites.org/	[111]
ARTS	细菌基因组靶向挖掘,筛选潜在新型抗生素靶点	https://arts.ziemertlab.com/	[112]
BiG-SCAPE	用于构建 BGC 的序列相似性网络并将其分组到 GCF 中	https://git.wageningenur.nl/medema-group/BiG-SCAPE/wikis/home	[113]
BiG-SLiCE	聚集大量 BGC,通过序列相似性网络构建 GCF 模型	https://github.com/medema-group/bigslice	[114]
ClusterFinder	预测基因组中的 BGC	https://github.com/petercim/ClusterFinder	[115]
DeepBGC	使用深度学习的方法预测细菌和真菌基因组中的 BGC	https://github.com/Merck/deepbgc	[116]
e-DeepBGC	基于深度学习的方法,引入了 Pfam 信息,预测细菌基因组中 BGC	—	[117]
RL-BGC	基于 Pfam 蛋白质家族结构域和功能注释的强化学习方法,精准获得真菌候选 BGC	https://github.com/bioinfoUQAM/RL-bgc-components	[118]
PRISM4	基于微生物基因组序列,识别 BGC 并生成结构预测的组合文库	https://prism.adapsyn.com/	[119]
SMURF	用于真菌基因组 BGC 和途径预测分析	http://smurf.jevi.org/index.php	[120]
TeroKit	萜类化合物的化学空间、生物活性和生物合成途径的在线检索和分析	http://terokit.qmclab.com/	[1]
酶功能预测工具			
CLEAN	启用对比学习的酶注释,能够对未表征的酶实现功能预测	https://clean.platform.moleculemaker.org/configuration	[121]
ECRECEr	基于深度学习实现酶催化功能预测	https://ecrecer.biodesign.ac.cn/	[122]
PfamScan	根据 Pfam HMM 数据库进行酶功能预测分析	https://www.ebi.ac.uk/Tools/pfa/pfamscan/	[123-124]

掘萜类天然产物过程中,发现两种丝状真菌来源 I 型嵌合萜类合酶 TvTS 和 MpMS,能够直接以 IPP 和 DMAPP 或者 HexPP^[128] 为底物合成三萜核心骨架,这一发现颠覆了所有三萜化合物都经由角鲨烯合成的固有认知。在充分解析 I 型嵌合三萜合酶的结构-功能关系后,建立了基于 AlphaFold 和底物分子对接的精准挖掘策略:利用 AlphaFold2 批量预测潜在的萜类合酶的三维结构,结合底物分子对接,成功预测并筛选获得另外两个三萜合酶(CgCS 和 PTTC074),它们都能够合成新三个非角鲨烯三萜化合物 colleterpenol,夯实了 I 型嵌合萜类合酶催化 HexPP 环化生成三萜骨架的普遍性机理^[129]。这项三萜化合物生物合成路径的颠覆

性发现,不仅拓展了萜类天然产物生物合成机制的多样性,也凸显了人工智能用于拓宽新型天然产物深度挖掘和精准发现的潜力。

4 展望

微生物高产底盘的构建助力了大量新颖萜类化合物的高效发现,充分释放了萜类合酶的合成潜能,提升了低丰度新骨架萜类天然产物的检出率,显著地加快了萜类化合物的基因组挖掘进程。随着合成生物学底盘构建技术和策略的不断精进,人们已经在多种模式微生物和非模式微生物中搭

建了相应的前体高效供应平台用于特定类型萜类的高效合成,如单萜底盘、倍半萜底盘、二萜底盘、二倍半萜底盘等。值得注意的是,微生物的物质代谢和能量平衡受到严格而又复杂的代谢网络调控,底盘中前体供应并不是越多越好,还需要考虑中间产物积累产生细胞毒性的问题。如Ma等^[52]在构建萜类酵母底盘时实现了“生长”与“生产”的调控:首先利用葡萄糖诱导调控系统调控酵母生长和产物合成的代谢流;随后通过引入外源性两步磷酸化利用途径(IUP)实现IPP/DMAPP高效合成,减少与宿主自身生长资源的竞争,构建了GPP、FPP和GGPP的多种前体高效供给的萜类通用性合成平台。萜类的生物合成途径较长,涉及的酶可能定位于不同的物理空间,在底盘构建时如何避免中间体的损失和副反应的生成呢?多酶组装策略的使用有助于解决上述问题,目前人们已经陆续开发了RIAD-RIDD^[130]、mPKSeal^[131]等多酶组装策略,将功能各异的酶级联组装成一个整体可以形成“底物通道”,使萜类合成前体更多地走向下游目标途径,进一步强化底盘中萜类前体的供应。

基因组测序技术的飞速发展,驱动了海量的基因序列数字资源库的创建;基因组挖掘工具的迭代更新,使人们可以从数据库中快速精准锁定到目标BGC;高效精准的基因编辑技术的开发和不断精进,为探索萜类天然产物提供了先进的使能技术,如CRISPR/Cas9^[132]、CRISPR/Cas12a(Cpf1)^[133-134]、CRISPR/Cas13^[135]等。中国科学院天津工业生物技术研究所还开发了首个面向微生物基因编辑序列设计的云平台——AutoESD(<https://autoesd.biodesign.ac.cn/>),云平台拥有强大的高通量自动化性能,能够在几分钟内并行处理数百个设计任务^[136]。基于多学科交融的基因组挖掘策略,促进了萜类BGC和新型萜类产物的快速、高效、精准挖掘。然而,萜类化合物的基因组挖掘还存在一些挑战:

(1)虽然已经开发了多种生物信息学工具,用来在众多的基因序列中识别具有潜在价值的BGC,但相比于聚酮类(PKS)和非核糖体肽类(NRPS)天然产物,萜类BGC挖掘可用的甚至专用的数据库及挖掘工具还较少,BGC预测的精准

性还有待加强。

(2)萜类BGC很难展现化合物的结构特征,BGC产物的分子结构预测仍有较大挑战。目前已经有研究通过创建代谢基因组数据集,将天然产物与对应的BGC相匹配^[137],但该方法主要应用于细菌代谢产物及其生物合成研究^[138-140],在具体应用中也主要围绕NRPS、PKS等化合物^[141],萜类化合物很少有应用案例;2023年,Caesar等^[142]将代谢基因组学首次应用于真菌天然产物预测,但是主要针对的是已知BGC的检测。对于沉默BGC以及新颖BGC,未来仍需要很多探索才能使用BGC实现萜类化合物的精准预测。

(3)迄今为止,绝大多数萜类化合物是从植物和真菌中分离出来的,只有不到1.5%来自细菌^[143],而且相比于海量的微生物资源,大自然中可培养微生物只占很小一部分^[144],未培养微生物基因组中还蕴含着大量未知的BGC。研究人员也在不断扩宽研究范畴,如从深海等极端环境收集微生物资源开展BGC深度挖掘,然而目前缺乏这类特殊微生物的专用工具,包括遗传操作系统、合成与调控元件、底盘细胞等^[145],亟需解决上述问题以便最大程度开发特殊生境微生物的天然产物生物合成潜能。

(4)人工智能为合成生物学带来新的变革,处于学科交融的浪潮,机器学习、深度学习等人工智能也对我们提出了更高的要求,需要更加与时俱进的思想,以及更快的适应与融入,但不可否认,是挑战也是机遇,将生物底层设计、生物大数据与人工智能等有机结合,将给萜类天然产物的挖掘开启全新篇章。

参 考 文 献

- [1] ZENG T, LIU Z H, ZHUANG J Y, et al. TeroKit: a database-driven web server for terpenome research[J]. *Journal of Chemical Information and Modeling*, 2020, 60(4): 2082-2090.
- [2] FORDJOUR E, MENSAH E O, HAO Y P, et al. Toward improved terpenoids biosynthesis: strategies to enhance the capabilities of cell factories[J]. *Bioresources and Bioprocessing*, 2022, 9(1): 6.
- [3] AVALOS M, GARBEVA P, VADER L, et al. Biosynthesis, evolution and ecology of microbial terpenoids[J]. *Natural Product Reports*, 2022, 39(2): 249-272.

- [4] ARNESEN J A, JACOBSEN I H, DYKJÆR J D, et al. Production of abscisic acid in the oleaginous yeast *Yarrowia lipolytica*[J]. FEMS Yeast Research, 2022, 22(1): foac015.
- [5] PADDON C J, WESTFALL P J, PITERA D J, et al. High-level semi-synthetic production of the potent antimalarial artemisinin [J]. Nature, 2013, 496(7446): 528-532.
- [6] WU Q Y, HUANG Z Y, WANG J Y, et al. Construction of an *Escherichia coli* cell factory to synthesize taxadien-5 α -ol, the key precursor of anti-cancer drug paclitaxel[J]. Bioresources and Bioprocessing, 2022, 9(1): 82.
- [7] ALBERTI F, KHAIRUDIN K, DAVIES J A, et al. Biosynthesis of pleuromutilin congeners using an *Aspergillus oryzae* expression platform[J]. Chemical Science, 2023, 14(14): 3826-3833.
- [8] HOU A W, LAUTERBACH L, DICKSCHAT J S. Enzymatic synthesis of methylated terpene analogues using the plasticity of bacterial terpene synthases[J]. Chemistry, 2020, 26(10): 2178-2182.
- [9] HOU A W, DICKSCHAT J S. Using terpene synthase plasticity in catalysis: on the enzymatic conversion of synthetic farnesyl diphosphate analogues[J]. Chemistry, 2021, 27(63): 15644-15649.
- [10] VON REUSS S H, KAI M, PIECHULLA B, et al. Octamethylbicyclo[3.2.1]octadienes from the rhizobacterium *Serratia odorifera*[J]. Angewandte Chemie International Edition, 2010, 49(11): 2009-2010.
- [11] KOMATSU M, TSUDA M, OMURA S, et al. Identification and functional analysis of genes controlling biosynthesis of 2-methylisoborneol[J]. Proceedings of the National Academy of Sciences of the United States of America, 2008, 105(21): 7422-7427.
- [12] WANG C M, CANE D E. Biochemistry and molecular genetics of the biosynthesis of the earthy odorant methylisoborneol in *Streptomyces coelicolor*[J]. Journal of the American Chemical Society, 2008, 130(28): 8908-8909.
- [13] VON REUSS S, DOMIK D, LEMFACK M C, et al. Sodorifen biosynthesis in the rhizobacterium *Serratia plymuthica* involves methylation and cyclization of MEP-derived farnesyl pyrophosphate by a SAM-dependent C-methyltransferase[J]. Journal of the American Chemical Society, 2018, 140(37): 11855-11862.
- [14] HU Z X, YE Y, ZHANG Y H. Large-scale culture as a complementary and practical method for discovering natural products with novel skeletons[J]. Natural Product Reports, 2021, 38(10): 1775-1793.
- [15] RYU M J, HWANG S, KIM S, et al. Meroindenon and merochlorins E and F, antibacterial meroterpenoids from a marine-derived sediment bacterium of the genus *Streptomyces* [J]. Organic Letters, 2019, 21(15): 5779-5783.
- [16] ZHAI G F, CHEN S H, SHEN H J, et al. Bioactive monoterpenes and polyketides from the ascidian-derived fungus *Diaporthe* sp. SYSU-MS4722[J]. Marine Drugs, 2022, 20(9): 553.
- [17] LIU H, PU Y H, REN J W, et al. Genetic dereplication driven discovery of a tricinolonol acid biosynthetic pathway in *Trichoderma hypoxylon*[J]. Organic & Biomolecular Chemistry, 2020, 18(28): 5344-5348.
- [18] KIM S H, LU W L, AHMADI M K, et al. Atolypenes, tricyclic bacterial sesterterpenes discovered using a multiplexed *in vitro* Cas9-TAR gene cluster refactoring approach[J]. ACS Synthetic Biology, 2019, 8(1): 109-118.
- [19] SHEN X T, MO X H, ZHU L P, et al. Unusual and highly bioactive sesterterpenes synthesized by *Pleurotus ostreatus* during coculture with *Trametes robiniophila* Murr[J]. Applied and Environmental Microbiology, 2019, 85(14): e00293-19.
- [20] CHIANG C Y, OHASHI M, TANG Y. Deciphering chemical logic of fungal natural product biosynthesis through heterologous expression and genome mining[J]. Natural Product Reports, 2023, 40(1): 89-127.
- [21] MENG X F, FANG Y, DING M Y, et al. Developing fungal heterologous expression platforms to explore and improve the production of natural products from fungal biodiversity[J]. Biotechnology Advances, 2022, 54: 107866.
- [22] MATSUDA Y, MITSUHASHI T, LEE S K, et al. Astellifadiene: structure determination by NMR spectroscopy and crystalline sponge method, and elucidation of its biosynthesis[J]. Angewandte Chemie International Edition, 2016, 55(19): 5785-5788.
- [23] 程术, 邓子新, 卞光凯, 等. 萜类高效合成平台的搭建与萜类产物批量挖掘[J]. 生命科学, 2019, 31(5): 449-457.
- CHENG S, DENG Z X, BIAN G K, et al. Construction of high-efficient terpenoid platform and the application in terpenoid discovery[J]. Chinese Bulletin of Life Sciences, 2019, 31(5): 449-457.
- [24] ZHU F Y, ZHONG X F, HU M Z, et al. *In vitro* reconstitution of mevalonate pathway and targeted engineering of farnesene overproduction in *Escherichia coli*[J]. Biotechnology and Bioengineering, 2014, 111(7): 1396-1405.
- [25] ZHU F Y, LU L, FU S, et al. Targeted engineering and scale up of lycopene overproduction in *Escherichia coli*[J]. Process Biochemistry, 2015, 50(3): 341-346.
- [26] MA T, ZHOU Y J, LI X W, et al. Genome mining of astaxanthin biosynthetic genes from *Sphingomonas* sp. ATCC 55669 for heterologous overproduction in *Escherichia coli*[J]. Biotechnology Journal, 2016, 11(2): 228-237.
- [27] BIAN G K, YUAN Y J, TAO H, et al. Production of taxadiene by engineering of mevalonate pathway in *Escherichia coli* and endophytic fungus *Alternaria alternata* TPF6[J].

- Biotechnology Journal, 2017, 12(4): 1600697.
- [28] BIAN G K, DENG Z X, LIU T G. Strategies for terpenoid overproduction and new terpenoid discovery[J]. Current Opinion in Biotechnology, 2017, 48: 234-241.
- [29] YUAN Y J, CHENG S, BIAN G K, et al. Efficient exploration of terpenoid biosynthetic gene clusters in filamentous fungi[J]. Nature Catalysis, 2022, 5: 277-287.
- [30] BIAN G K, HAN Y C, HOU A W, et al. Releasing the potential power of terpene synthases by a robust precursor supply platform[J]. Metabolic Engineering, 2017, 42: 1-8.
- [31] BIAN G K, HOU A W, YUAN Y J, et al. Metabolic engineering-based rapid characterization of a sesquiterpene cyclase and the skeletons of fusariumdiene and fusagramineol from *Fusarium graminearum*[J]. Organic Letters, 2018, 20(6): 1626-1629.
- [32] BIAN G K, RINKEL J, WANG Z Q, et al. A clade II -D fungal chimeric diterpene synthase from *Colletotrichum gloeosporioides* produces dolasta-1(15),8-diene[J]. Angewandte Chemie International Edition, 2018, 57(48): 15887-15890.
- [33] SIEMON T, WANG Z Q, BIAN G K, et al. Semisynthesis of plant-derived englerin A enabled by microbe engineering of guaia-6,10(14)-diene as building block[J]. Journal of the American Chemical Society, 2020, 142(6): 2760-2765.
- [34] ZUO Y M, XIAO F, GAO J C, et al. Establishing *Komagataella phaffii* as a cell factory for efficient production of sesquiterpenoid α -santalene[J]. Journal of Agricultural and Food Chemistry, 2022, 70(26): 8024-8031.
- [35] LUO G J, LIN Y, CHEN S T, et al. Overproduction of patchoulol in metabolically engineered *Komagataella phaffii* [J]. Journal of Agricultural and Food Chemistry, 2023, 71(4): 2049-2058.
- [36] LI Z J, WANG Y Z, WANG L R, et al. Advanced strategies for the synthesis of terpenoids in *Yarrowia lipolytica*[J]. Journal of Agricultural and Food Chemistry, 2021, 69(8): 2367-2381.
- [37] MA Y R, WANG K F, WANG W J, et al. Advances in the metabolic engineering of *Yarrowia lipolytica* for the production of terpenoids[J]. Bioresource Technology, 2019, 281: 449-456.
- [38] ZHANG T L, YU H W, YE L D. Metabolic engineering of *Yarrowia lipolytica* for terpenoid production: tools and strategies[J]. ACS Synthetic Biology, 2023, 12(3): 639-656.
- [39] ZHANG G, WANG H, ZHANG Z, et al. Metabolic engineering of *Yarrowia lipolytica* for terpenoids production: advances and perspectives[J]. Critical Reviews in Biotechnology, 2022, 42(4): 618-633.
- [40] LIN S Y, OAKLEY C E, JENKINSON C B, et al. A heterologous expression platform in *Aspergillus nidulans* for the elucidation of cryptic secondary metabolism biosynthetic gene clusters: discovery of the *Aspergillus fumigatus* sartorypyrone biosynthetic pathway[J]. Chemical Science, 2023, 14(40): 11022-11032.
- [41] TARASOVA E V, LUCHNIKOVA N A, GRISHKO V V, et al. Actinomycetes as producers of biologically active terpenoids: current trends and patents[J]. Pharmaceuticals, 2023, 16(6): 872.
- [42] HU Y L, ZHANG Q, LIU S H, et al. Building *Streptomyces albus* as a chassis for synthesis of bacterial terpenoids[J]. Chemical Science, 2023, 14(13): 3661-3667.
- [43] MYRONOVSKIY M, LUZHETSKYY A. Heterologous production of small molecules in the optimized *Streptomyces* hosts[J]. Natural Product Reports, 2019, 36(9): 1281-1294.
- [44] AJIKUMAR P K, XIAO W H, TYO K E, et al. Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*[J]. Science, 2010, 330(6000): 70-74.
- [45] REDDY G K, LEFERINK N G H, UMEMURA M, et al. Exploring novel bacterial terpene synthases[J]. PLoS One, 2020, 15(4): e0232220.
- [46] LIN G M, VOIGT C A. Design of a redox-proficient *Escherichia coli* for screening terpenoids and modifying cytochrome P450s[J]. Nature Catalysis, 2023, 6: 1016-1029.
- [47] NAVALE G R, DHARNE M S, SHINDE S S. Metabolic engineering and synthetic biology for isoprenoid production in *Escherichia coli* and *Saccharomyces cerevisiae*[J]. Applied Microbiology and Biotechnology, 2021, 105(2): 457-475.
- [48] 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.
- [49] 李晓东, 杨成帅, 王平平, 等. 构建酿酒酵母细胞工厂从头合成倍半萜类化合物 α -新丁香三环烯和 β -石竹烯[J]. 合成生物学, 2021, 2(5): 792-803.
- LI X D, YANG C S, WANG P P, et al. Production of sesquiterpenoids α -neoclovene and β -caryophyllene by engineered *Saccharomyces cerevisiae*[J]. Synthetic Biology Journal, 2021, 2(5): 792-803.
- [50] DONG H, CHEN S, ZHU J X, et al. Enhance production of diterpenoids in yeast by overexpression of the fused enzyme of ERG20 and its mutant mERG20[J]. Journal of Biotechnology, 2020, 307: 29-34.
- [51] MUKHERJEE M, BLAIR R H, WANG Z Q. Machine-learning guided elucidation of contribution of individual steps in the mevalonate pathway and construction of a yeast platform strain for terpenoid production[J]. Metabolic Engineering, 2022, 74: 139-149.
- [52] MA Y S, ZU Y X, HUANG S W, et al. Engineering a universal and efficient platform for terpenoid synthesis in yeast[J]. Proceedings of the National Academy of Sciences of the

- United States of America, 2023, 120(1): e2207680120.
- [53] MEYER V, FIEDLER M, NITSCHKE B, et al. The cell factory *Aspergillus* enters the big data era: opportunities and challenges for optimising product formation[J]. *Advances in Biochemical Engineering/Biotechnology*, 2015, 149: 91-132.
- [54] TONG Z Y, ZHENG X M, TONG Y, et al. Systems metabolic engineering for citric acid production by *Aspergillus niger* in the post-genomic era[J]. *Microbial Cell Factories*, 2019, 18(1): 28.
- [55] HUANG X N, CHEN M, LU X F, et al. Direct production of itaconic acid from liquefied corn starch by genetically engineered *Aspergillus terreus*[J]. *Microbial Cell Factories*, 2014, 13: 108.
- [56] HUANG X N, TANG S, ZHENG L H, et al. Construction of an efficient and robust *Aspergillus terreus* cell factory for monacolin J production[J]. *ACS Synthetic Biology*, 2019, 8(4): 818-825.
- [57] ICHISHIMA E. Development of enzyme technology for *Aspergillus oryzae*, *A. sojae*, and *A. luchuensis*, the national microorganisms of Japan[J]. *Bioscience, Biotechnology, and Biochemistry*, 2016, 80(9): 1681-1692.
- [58] VAN BERGEIJK D A, TERLOUW B R, MEDEMA M H, et al. Ecology and genomics of Actinobacteria: new concepts for natural product discovery[J]. *Nature Reviews Microbiology*, 2020, 18(10): 546-558.
- [59] 董雷, 韩嘉瑞, 李帅, 等. 链霉菌最新研究进展[J]. *微生物学报*, 2023, 63(5): 1815-1832.
- DONG L, HAN J R, LI S, et al. The latest research progress of streptomycetes[J]. *Acta Microbiologica Sinica*, 2023, 63(5): 1815-1832.
- [60] AHMED Y, REBETS Y, ESTÉVEZ M R, et al. Engineering of *Streptomyces lividans* for heterologous expression of secondary metabolite gene clusters[J]. *Microbial Cell Factories*, 2020, 19(1): 5.
- [61] BALTZ R H. Genetic manipulation of secondary metabolite biosynthesis for improved production in *Streptomyces* and other actinomycetes[J]. *Journal of Industrial Microbiology & Biotechnology*, 2016, 43(2-3): 343-370.
- [62] MYRONOVSKIY M, ROSENKRÄNZER B, NADMID S, et al. Generation of a cluster-free *Streptomyces albus* chassis strains for improved heterologous expression of secondary metabolite clusters[J]. *Metabolic Engineering*, 2018, 49: 316-324.
- [63] MUKHERJEE S, STAMATIS D, LI C T, et al. Twenty-five years of Genomes OnLine Database (GOLD): data updates and new features in v.9[J]. *Nucleic Acids Research*, 2023, 51(D1): D957-D963.
- [64] 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.
- [65] SUN X, CAI Y S, YUAN Y J, et al. Genome mining in *Trichoderma viride* J1-030: discovery and identification of novel sesquiterpene synthase and its products[J]. *Beilstein Journal of Organic Chemistry*, 2019, 15: 2052-2058.
- [66] CHEN R, FENG T, LI M, et al. Characterization of tremulane sesquiterpene synthase from the basidiomycete *Irpex lacteus* [J]. *Organic Letters*, 2022, 24(31): 5669-5673.
- [67] LI Z N, XU B F, KOJASOY V, et al. First *trans*-eunicellane terpene synthase in bacteria[J]. *Chem*, 2023, 9(3): 698-708.
- [68] ZHU C X, XU B F, ADRESSA D A, et al. Discovery and biosynthesis of a structurally dynamic antibacterial diterpenoid [J]. *Angewandte Chemie International Edition*, 2021, 60(25): 14163-14170.
- [69] WANG Z Y, YANG Q, HE J Y, et al. Cytochrome P450 mediated cyclization in eunicellane derived diterpenoid biosynthesis[J]. *Angewandte Chemie International Edition*, 2023, 62(45): e202312490.
- [70] LI Z, JIANG Y Y, ZHANG X W, et al. Fragrant venezuelaenes A and B with A 5-5-6-7 tetracyclic skeleton: discovery, biosynthesis, and mechanisms of central catalysts[J]. *ACS Catalysis*, 2020, 10(10): 5846-5851.
- [71] TANG M C, SHEN C, DENG Z X, et al. Combinatorial biosynthesis of terpenoids through mixing-and-matching sesquiterpene cyclase and cytochrome P450 pairs[J]. *Organic Letters*, 2022, 24(26): 4783-4787.
- [72] TSUKADA K, SHINKI S, KANEKO A, et al. Synthetic biology based construction of biological activity-related library of fungal decalin-containing diterpenoid pyrones[J]. *Nature Communications*, 2020, 11(1): 1830.
- [73] DUAN Y T, KOUTSAVITI A, HARIZANI M, et al. Widespread biosynthesis of 16-carbon terpenoids in bacteria [J]. *Nature Chemical Biology*, 2023, 19(12): 1532-1539.
- [74] ROTHSCHILD L J, MANCINELLI R L. Life in extreme environments[J]. *Nature*, 2001, 409: 1092-1101.
- [75] SHU W S, HUANG L N. Microbial diversity in extreme environments[J]. *Nature Reviews Microbiology*, 2022, 20(4): 219-235.
- [76] GARETH JONES E B, PANG K L, ABDEL-WAHAB M A, et al. An online resource for marine fungi[J]. *Fungal Diversity*, 2019, 96(1): 347-433.
- [77] GONÇALVES M F M, ESTEVES A C, ALVES A. Marine fungi: opportunities and challenges[J]. *Encyclopedia*, 2022, 2(1): 559-577.
- [78] SHABANA S, LAKSHMI K R, SATYA A K. An updated review of secondary metabolites from marine fungi[J]. *Mini Reviews in Medicinal Chemistry*, 2021, 21(5): 602-642.
- [79] JIANG M H, WU Z E, GUO H, et al. A review of terpenes from marine-derived fungi: 2015-2019[J]. *Marine Drugs*, 2020,

- 18(6): 321.
- [80] YANG X W, PENG K, LIU Z, et al. Strepsesquitriol, a rearranged zizaane-type sesquiterpenoid from the deep-sea-derived actinomycete *Streptomyces* sp. SCSIO 10355[J]. *Journal of Natural Products*, 2013, 76(12): 2360-2363.
- [81] KIM D, LEE E J, LEE J, et al. Antartin, a cytotoxic zizaane-type sesquiterpenoid from a *Streptomyces* sp. isolated from an Antarctic marine sediment[J]. *Marine Drugs*, 2018, 16(4): 130.
- [82] CARROLL A R, COPP B R, DAVIS R A, et al. Marine natural products[J]. *Natural Product Reports*, 2023,40(2):275-325.
- [83] CARROLL A R, COPP B R, DAVIS R A, et al. Marine natural products[J]. *Natural Product Reports*, 2022,39(6):1122-1171.
- [84] CARROLL A R, COPP B R, DAVIS R A, et al. Marine natural products[J]. *Natural Product Reports*, 2021,38(2):362-413.
- [85] SCHULTZ J, MODOLON F, PEIXOTO R S, et al. Shedding light on the composition of extreme microbial dark matter: alternative approaches for culturing extremophiles[J]. *Frontiers in Microbiology*, 2023, 14: 1167718.
- [86] GUO J J, CAI Y S, CHENG F C, et al. Genome mining reveals a multiproduct sesterterpenoid biosynthetic gene cluster in *Aspergillus ustus*[J]. *Organic Letters*, 2021, 23(5): 1525-1529.
- [87] ZHANG P, WU G W, HEARD S C, et al. Identification and characterization of a cryptic bifunctional type I diterpene synthase involved in talaronoid biosynthesis from a marine-derived fungus[J]. *Organic Letters*, 2022, 24(38): 7037-7041.
- [88] 曲戈, 朱彤, 蒋迎迎, 等. 蛋白质工程:从定向进化到计算设计[J]. *生物工程学报*, 2019, 35(10): 1843-1856.
- QU G, ZHU T, JIANG Y Y, et al. Protein engineering: from directed evolution to computational design[J]. *Chinese Journal of Biotechnology*, 2019, 35(10): 1843-1856.
- [89] 卞佳豪, 杨广宇. 人工智能辅助的蛋白质工程[J]. *合成生物学*, 2022, 3(3): 429-444.
- BIAN J H, YANG G Y. Artificial intelligence-assisted protein engineering[J]. *Synthetic Biology Journal*, 2022, 3(3): 429-444.
- [90] LI Z, ZHANG L L, XU K W, et al. Molecular insights into the catalytic promiscuity of a bacterial diterpene synthase[J]. *Nature Communications*, 2023, 14(1): 4001.
- [91] IGNEA C, PONTINI M, MOTAWIA M S, et al. Synthesis of 11-carbon terpenoids in yeast using protein and metabolic engineering[J]. *Nature Chemical Biology*, 2018, 14(12): 1090-1098.
- [92] IGNEA C, RAADAM M H, KOUTSAVITI A, et al. Expanding the terpene biosynthetic code with non-canonical 16 carbon atom building blocks[J]. *Nature Communications*, 2022, 13(1): 5188.
- [93] HAMEDIRAD M, CHAO R, WEISBERG S, et al. Towards a fully automated algorithm driven platform for biosystems design[J]. *Nature Communications*, 2019, 10(1): 5150.
- [94] ENGHAD B, XUE P, SINGH N, et al. PlasmidMaker is a versatile, automated, and high throughput end-to-end platform for plasmid construction[J]. *Nature Communications*, 2022, 13(1): 2697.
- [95] LEFERINK N G H, DUNSTAN M S, HOLLYWOOD K A, et al. An automated pipeline for the screening of diverse monoterpene synthase libraries[J]. *Scientific Reports*, 2019, 9(1): 11936.
- [96] YUAN W, JIANG C J, WANG Q, et al. Biosynthesis of mushroom-derived type II ganoderic acids by engineered yeast[J]. *Nature Communications*, 2022, 13(1): 7740.
- [97] 崔金明, 张炳照, 马迎飞, 等. 合成生物学研究的工程化平台[J]. *中国科学院院刊*, 2018, 33(11): 1249-1257.
- CUI J M, ZHANG B Z, MA Y F, et al. Engineering platforms for synthetic biology research[J]. *Bulletin of Chinese Academy of Sciences*, 2018, 33(11): 1249-1257.
- [98] 卢挥, 张芳丽, 黄磊. 合成生物学自动化装置 iBioFoundry 的构建与应用[J]. *合成生物学*, 2023, 4(5): 877-891.
- LU H, ZHANG F L, HUANG L. Establishment of iBioFoundry for synthetic biology applications[J]. *Synthetic Biology Journal*, 2023, 4(5): 877-891.
- [99] 张亨, 冷梦甜, 金帆, 等. 合成生物研究重大科技基础设施概述[J]. *合成生物学*, 2022, 3(1): 184-194.
- ZHANG T, LENG M T, JIN F, et al. Overview on platform for synthetic biology research at Shenzhen[J]. *Synthetic Biology Journal*, 2022, 3(1): 184-194.
- [100] SCHERLACH K, HERTWECK C. Mining and unearthing hidden biosynthetic potential[J]. *Nature Communications*, 2021, 12(1): 3864.
- [101] HEMMERLING F, PIEL J. Strategies to access biosynthetic novelty in bacterial genomes for drug discovery[J]. *Nature Reviews Drug Discovery*, 2022, 21(5): 359-378.
- [102] ZENG T, CHEN Y X X, JIAN Y X, et al. Chemotaxonomic investigation of plant terpenoids with an established database (TeroMOL)[J]. *The New Phytologist*, 2022, 235(2): 662-673.
- [103] CHEN N H, ZHANG R, ZENG T, et al. Developing TeroENZ and TeroMAP modules for the terpenome research platform TeroKit[J]. *Database*, 2023, 2023: baad020.
- [104] 赖奇龙, 姚帅, 查毓国, 等. 微生物组生物合成基因簇发掘方法及应用前景[J]. *合成生物学*, 2023, 4(3): 611-627.
- LAI Q L, YAO S, ZHA Y G, et al. Microbiome-based biosynthetic gene cluster data mining techniques and application potentials[J]. *Synthetic Biology Journal*, 2023, 4(3): 611-627.
- [105] 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.
- [106] REN H Q, SHI C Y, ZHAO H M. Computational tools for

- discovering and engineering natural product biosynthetic pathways[J]. *iScience*, 2020, 23(1): 100795.
- [107] BLIN K, SHAW S, MEDEMA M H, et al. The antiSMASH database version 4: additional genomes and BGCs, new sequence-based searches and more[J]. *Nucleic Acids Research*, 2024, 52(D1): D586-D589.
- [108] KAUTSAR S A, BLIN K, SHAW S, et al. BiG-FAM: the biosynthetic gene cluster families database[J]. *Nucleic Acids Research*, 2021, 49(D1): D490-D497.
- [109] PALANIAPPAN K, CHEN I M A, CHU K, et al. IMG-ABC v.5.0: an update to the IMG/Atlas of Biosynthetic Gene Clusters Knowledgebase[J]. *Nucleic Acids Research*, 2020, 48(D1): D422-D430.
- [110] TERLOUW B R, BLIN K, NAVARRO-MUÑOZ J C, et al. MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters[J]. *Nucleic Acids Research*, 2023, 51(D1): D603-D610.
- [111] BLIN K, SHAW S, AUGUSTIJN H E, et al. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation[J]. *Nucleic Acids Research*, 2023, 51(W1): W46-W50.
- [112] MUNGAN M D, 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.
- [113] 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.
- [114] KAUTSAR S A, VAN DER HOOFT J J J, DE RIDDER D, et al. BiG-SLiCE: a highly scalable tool maps the diversity of 1.2 million biosynthetic gene clusters[J]. *GigaScience*, 2021, 10(1): g1aa154.
- [115] 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.
- [116] HANNIGAN G D, PRIHODA D, PALICKA A, et al. A deep learning genome-mining strategy for biosynthetic gene cluster prediction[J]. *Nucleic Acids Research*, 2019, 47(18): e110.
- [117] LIU M Y, LI Y, LI H Z. Deep learning to predict the biosynthetic gene clusters in bacterial genomes[J]. *Journal of Molecular Biology*, 2022, 434(15): 167597.
- [118] ALMEIDA H, TSANG A, DIALLO A B. Improving candidate Biosynthetic Gene Clusters in fungi through reinforcement learning[J]. *Bioinformatics*, 2022, 38(16): 3984-3991.
- [119] SKINNIDER M A, JOHNSTON C W, GUNABALASINGAM M, et al. Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences[J]. *Nature Communications*, 2020, 11(1): 6058.
- [120] KHALDI N, SEIFUDDIN F T, TURNER G, et al. SMURF: genomic mapping of fungal secondary metabolite clusters[J]. *Fungal Genetics and Biology*, 2010, 47(9): 736-741.
- [121] YU T H, CUI H Y, LI J C, et al. Enzyme function prediction using contrastive learning[J]. *Science*, 2023, 379(6639): 1358-1363.
- [122] SHI Z K, DENG R, YUAN Q Q, et al. Enzyme commission number prediction and benchmarking with hierarchical dual-core multitask learning framework[J]. *Research*, 2023, 6: 153.
- [123] LI W Z, COWLEY A, ULUDAG M, et al. The EMBL-EBI bioinformatics web and programmatic tools framework[J]. *Nucleic Acids Research*, 2015, 43(W1): W580-W584.
- [124] MISTRY J, BATEMAN A, FINN R D. Predicting active site residue annotations in the Pfam database[J]. *BMC Bioinformatics*, 2007, 8: 298.
- [125] SAHAYASHEELA V J, LANKADASARI M B, DAN V M, et al. Artificial intelligence in microbial natural product drug discovery: current and emerging role[J]. *Natural Product Reports*, 2022, 39(12): 2215-2230.
- [126] MULLOWNEY M W, DUNCAN K R, ELSAYED S S, et al. Artificial intelligence for natural product drug discovery[J]. *Nature Reviews Drug Discovery*, 2023, 22(11): 895-916.
- [127] JUMPER J, EVANS R, PRITZEL A, et al. Highly accurate protein structure prediction with AlphaFold[J]. *Nature*, 2021, 596(7873): 583-589.
- [128] SASAKI D, FUJIHASHI M, OKUYAMA N, et al. Crystal structure of heterodimeric hexaprenyl diphosphate synthase from *Micrococcus luteus* B-P 26 reveals that the small subunit is directly involved in the product chain length regulation[J]. *The Journal of Biological Chemistry*, 2011, 286(5): 3729-3740.
- [129] TAO H, LAUTERBACH L, BIAN G K, et al. Discovery of non-squalene triterpenes[J]. *Nature*, 2022, 606(7913): 414-419.
- [130] KANG W, MA T, LIU M, et al. Modular enzyme assembly for enhanced cascade biocatalysis and metabolic flux[J]. *Nature Communications*, 2019, 10(1): 4248.
- [131] SUN X X, YUAN Y J, CHEN Q T, et al. Metabolic pathway assembly using docking domains from type I *cis*-AT polyketide synthases[J]. *Nature Communications*, 2022, 13(1): 5541.
- [132] TIAN P F, WANG J, SHEN X L, et al. Fundamental CRISPR-Cas9 tools and current applications in microbial systems[J]. *Synthetic and Systems Biotechnology*, 2017, 2(3): 219-225.
- [133] LI Z H, LIU M, LYU X M, et al. CRISPR/Cpf1 facilitated large fragment deletion in *Saccharomyces cerevisiae*[J]. *Journal of Basic Microbiology*, 2018, 58(12): 1100-1104.
- [134] LI L, WEI K K, ZHENG G S, et al. CRISPR-Cpf1-assisted multiplex genome editing and transcriptional repression in *Streptomyces*[J]. *Applied and Environmental Microbiology*, 2018, 84(18): e00827-18.

- [135] WU S M, TIAN P F, TAN T W. CRISPR-Cas13 technology portfolio and alliance with other genetic tools[J]. *Biotechnology Advances*, 2022, 61: 108047.
- [136] YANG Y, MAO Y F, WANG R Y, et al. AutoESD: a web tool for automatic editing sequence design for genetic manipulation of microorganisms[J]. *Nucleic Acids Research*, 2022, 50(W1): W75-W82.
- [137] CAESAR L K, MONTASER R, KELLER N P, et al. Metabolomics and genomics in natural products research: complementary tools for targeting new chemical entities[J]. *Natural Product Reports*, 2021, 38(11): 2041-2065.
- [138] DUNCAN K R, CRÜSEMANN M, LECHNER A, et al. Molecular networking and pattern-based genome mining improves discovery of biosynthetic gene clusters and their products from *Salinispora* species[J]. *Chemistry & Biology*, 2015, 22(4): 460-471.
- [139] MAANSSON M, VYNNE N G, KLITGAARD A, et al. An integrated metabolomic and genomic mining workflow to uncover the biosynthetic potential of bacteria[J]. *mSystems*, 2016, 1(3): e00028-15.
- [140] TRYON J H, ROTE J C, CHEN L, et al. Genome mining and metabolomics uncover a rare D-capreomycin containing natural product and its biosynthetic gene cluster[J]. *ACS Chemical Biology*, 2020, 15(11): 3013-3020.
- [141] LI C S, HU Y F, WU X H, et al. Discovery of unusual dimeric piperazyl cyclopeptides encoded by a *Lentzea flaviverrucosa* DSM 44664 biosynthetic supercluster[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2022, 119(17): e2117941119.
- [142] CAESAR L K, BUTUN F A, ROBEY M T, et al. Correlative metabologenomics of 110 fungi reveals metabolite-gene cluster pairs[J]. *Nature Chemical Biology*, 2023, 19(7): 846-854.
- [143] RUDOLF J D, ALSUP T A, XU B F, et al. Bacterial terpenome [J]. *Natural Product Reports*, 2021, 38(5): 905-980.
- [144] IMHOFF J F. Natural products from marine fungi: still an underrepresented resource[J]. *Marine Drugs*, 2016, 14(1): 19.
- [145] YANG Z J, HE J Q, WEI X, et al. Exploration and genome mining of natural products from marine *Streptomyces*[J]. *Applied Microbiology and Biotechnology*, 2020, 104(1): 67-76.



通讯作者: 刘天罡(1979—),男,教授,博士生导师。研究方向为萜类等天然产物的高效合成与创新发现;基于底盘细胞和自动化平台的天然产物基因组挖掘;微生物与人体的代谢互作。

E-mail: liutg@whu.edu.cn



通讯作者: 陶慧(1990—),女,教授,博士生导师。研究方向为复杂微生物来源天然产物的生物合成机制解析与绿色生物制造。

E-mail: thui@whu.edu.cn



第一作者: 雷茹(1998—),女,硕士研究生。研究方向为深海来源真菌及放线菌基因组测序与功能分析。

E-mail: 2017302290014@whu.edu.cn