·大家专版·

潘华奇,博士,中国科学院沈阳应用生态研究所研究员,学科组组长,研究所功能微生物挖掘与高值利用团队首席科学家;兼任生物农药与生物防治产业技术创新战略联盟副理事长、中国微生物学会普通微生物学专业委员会委员、沈阳药科大学和中国医科大学硕士研究生导师;入选中国科学院青年创新促进会会员、辽宁省“兴辽英才计划”青年拔尖人才、辽宁省“百千万人才工程”千层次人才、首批沈阳市中青年科技创新人才等。长期从事特境微生物资源及其活性产物发掘与利用的研究,主要采用多学科交叉的研究方法挖掘特境微生物、研制功能微生物菌剂;同时发现抗肿瘤、抗感染、杀真菌、生物刺激素等生物活性先导分子,通过结构改造和构效关系研究改善成药性,使用合成生物学技术和代谢工程的手段提高产量,并进行产业化开发。截至目前,已承担国家自然科学基金、A 类中国科学院战略性先导科技专项、国家 863 计划等国家级和省部级课题 15 项;发表学术论文 52 篇,其中 SCI 论文 28 篇;授权中国发明专利 13 件,其中已转化或许可专利 3 件;作为核心成员创制了国际上首例甲基营养型芽孢杆菌生物杀菌剂。

后基因组时代微生物天然产物高效发掘体系设计与展望

潘华奇*

(中国科学院 沈阳应用生态研究所,辽宁 沈阳　110016)

摘　要　微生物天然产物具有丰富的化学结构多样性和诱人的生物活性,持续启迪着创新医药和农药的发现。近年来,随着高通量测序技术的快速发展,巨大的微生物基因组数据揭示了多样生物合成和新颖天然产物的潜能远高于以前的认知。然而,如何高效地激活隐性的生物合成基因簇(BGCs) 并识别相应的化合物,以及避免已知代谢产物的重复发现等挑战依然严峻。本文描述了面对这些问题时基因组学、生物信息学、机器学习、代谢组学、基因编辑和合成生物学等新技术在发现药用先导化合物过程中提供的机遇;总结并论述了在潜力菌株优选、BGCs 的生物信息学预测、沉默 BGCs 的高效激活以及目标产物的识别和跟踪方面的新见解;提出了基于潜力菌株选择和多组学挖掘技术从微生物天然产物中高效发现先导结构的系统线路(SPLSD),并讨论了未来天然产物药用先导发现的机遇和挑战。

关键词　潜力菌株;先导化合物;泛生物合成基因簇;沉默基因簇;基因组挖掘;代谢组学;次级代谢产物;合成生物学

中图分类号　Q936　　**文献标识码**　A　　**文章编号**　1005 - 7021(2022) 03 - 0001 - 14

doi: 10.3969/j.issn.1005 - 7021.2022.03.001

Design and Prospects of an Efficient Mining Pipeline for Microbial Natural Products in the Post-Genome Era

PAN Hua-qi*

(*Institute of Applied Ecology*, *Chinese Academy of Science*, *Shenyang* 110016)

基金项目:国家自然科学基金面上项目(31872036,41576136);中国科学院战略性先导科技专项 A 类(XDA28090300);中科院青年创新促进会会员项目(2018229);辽宁省兴辽英才计划项目(XLYC1807268)

作者简介:潘华奇　男,研究员,博士,硕士生导师。主要从事特境微生物与农用微生物菌剂研制、微生物药物与合成生物技术研究。Tel: 024-83970386, E-mail: panhq@iae.ac.cn

* 通讯作者。

收稿日期:2022-05-31

Abstract Microbial natural products possess enormous diversity of chemical structures and fascinating biological activities, which continue to inspire novel discoveries in medicine and pesticide. In recent years, with the rapid development of high-throughput sequencing technology, massive microbial genomic data have revealed a much greater potential for the biosynthesis of diverse and novel natural products than previously appreciated. However, activating cryptic biosynthetic gene clusters (BGCs) and identifying the corresponding compounds, as well as avoiding rediscovery of known metabolites are still challenging. Here, this paper describes the new technologies such as genomics, bioinformatics, machine learning, metabolomics, gene editing and synthetic biology in the discovery of drug lead compounds to address the aforementioned problems. New insights into prioritizing promising strains, bioinformatics prediction of BGCs, efficient activation of silent BGCs, and identification and tracking of target products are summarized and presented. Finally, a systematic pipeline for efficient lead structure discovery from microbial natural products by promising strain selection and multi-omics mining (SPLSD) is established, and future opportunities and challenges in workflows of natural product drug lead discovery are discussed.

Keywords promising strain; lead compounds; pan-biosynthetic gene cluster; silent gene cluster; genome mining; metabolomics; secondary metabolites; synthetic biology

长期的实践表明,不论是新型医药还是农药的创制,天然产物(natural products,NPs)及其启发合成的小分子药物均占据着极其重要的地位。Newman 等统计了过去 39 年的药物,发现小分子药物占所有 1 881 种新批准药物的 77.53%,其中以 NPs 为先导成药的有 916 种,占所有小分子药物 65.71%[1]。同样在植物保护领域,Sparks 和 Bryant 总结了过去 70 年超过 800 种植物保护化合物,发现 NPs 及其启发合成的农药占总数的近一半[2]。可见,化学多样性极其丰富的 NPs 拥有独特的化学空间和药效优势成为创新医药和农药的重要来源与启迪。微生物因具有极其庞大的类群和丰富的代谢产物成为新型医药和农药的重要来源,并且贡献了最多的抗生素[3-4]。自 1928 年 Fleming 发现青霉素开始,微生物来源的药物进入快速发展的黄金时期,并长期成为国际社会的研究热点,尤其是 2015 年著名杀虫剂阿维菌素的发明者卡贝尔和大村智获得诺贝尔奖,又掀起了微生物先导药物的研究热潮[5]。然而随着微生物(尤其土壤微生物)的不断开发和重复筛选(包括菌种和筛选方法的重复),发现新颖药用先导化合物的几率不断下降[6-7](图 1A)。因此,科学家将目光投向包括植物内生菌[8]、海洋微生物[9]、极端环境微生物等尚未充分发掘的特境微生物新资源。进入 21 世纪以来,海洋微生物来源的新 NPs 进入快速增长期(图 1B),尤其是 2008 年以后海洋新 NPs 的增量几乎都是源于海洋微生物。目前,海洋微生物每年贡献约 1 000 个新 NPs,占海洋来源新 NPs 总数的 2/3 以上(图 1B)。可见,特境微生物来源的新 NPs 已成为发现新型活性先导化合物的重要增长点和研究热点[8-10]。

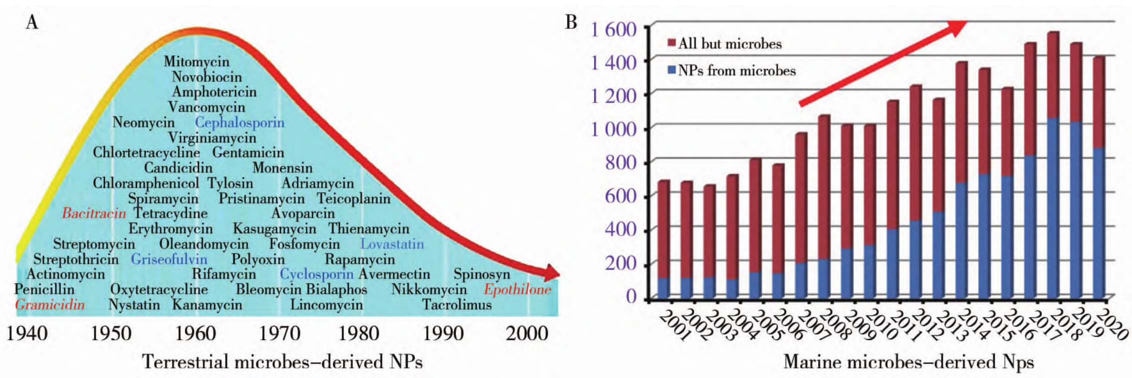

图 1 不同生境微生物来源新 NPs 的发现趋势

Fig. 1 Trend in the discovery of new NPs from microbes in different eco-environment

近十多年来，随着基因组学的迅猛发展，人们认识到微生物具有丰富的编码次级代谢产物的生物合成基因簇(biosynthetic gene clusters，BGCs)，估计高产的放线菌具有 20 ~ 50 个 BGCs[11]，真菌则具有 30 ~ 70 个 BGCs[12-13]，这赋予其生产新结构药用先导化合物的巨大潜能。但是目前发现的 NPs 仅是其冰山一角，绝大部分(大约 80% ~ 90%) 的 BGCs 在常规条件下是沉默的或低表达的，因此许多基因组挖掘的激活策略应运而生[14]。特别是当前合成生物学和机器学习等新技术层出不穷，为微生物 NPs 的发掘带来了新的机遇。在后基因组时代如何实现微生物源药用先导化合物高效发掘? 本文将从富含 BGCs 的“潜力菌种”选择、生物信息学预测、“隐性”BGCs 激活以及激活产物的高效识别等方面进行论述，并提出了在后基因组时代微生物 NPs 高效发掘的集成新线路，泛生物合成基因簇(pan-BGCs) 的概念，以及对发掘 NPs 药用先导化合物过程中一些问题的思考。

1　富含 BGCs“潜力菌种”的选择

从可培养微生物中发掘 NPs 依然是先导药物分子发现的主流，选择菌株则是最为关键的第一步。越来越多的研究发现，产生生物活性代谢产物的微生物在分类学分布上是很不均衡的，放线菌门(Actinobacteria) 、厚壁菌门(Firmicutes) 、变形菌门(Proteobacteria) 、蓝细菌门(Cyanobacteria) 和子囊菌门(Ascomycota) 等类群的基因组中含有丰富的 BGCs(表 1) [15]，特别是这些类群中的链霉菌属、芽孢杆菌属、假单胞菌属和曲霉属等，其成员是过去大多数生物活性物质的贡献者[16]。最新的宏基因组研究发现，在微生物的分类等级中，链霉菌目、分枝杆菌目、伯克氏菌目、根瘤菌目、假单胞菌目、肠杆菌目、杆菌目、放线菌目、蓝藻目和乳酸杆菌目依次是已报道基因组数据中蕴含 BGCs 最为丰富的 Top10 类群(表 1) [17]。因此，利用分类学为导向优先从富含 BGCs 的微生物类群中挖掘药用先导分子是一种行之有效的策略。

表 1　基因组挖掘发现具有丰富 BGCs 谱系的分类学分布

Table 1　The taxonomic distribution of lineages with abundant BGCs discovered by genome

Kingdom	Phylum	Order	Genus
Bacteria	Actinobacteria	Streptomycetaceae	*Streptomyces*、*Kitasatospora*
		Corynebacteriales	*Mycobacterium*、*Mycolicibacterium*、*Nocardia*
		Micromonosporales	*Micromonospora*、*Salinispora*、*Actinoplanes*
		Frankiales	*Frankia*
		Streptosporangiales	*Nonomuraea*
		Pseudonocardiales	*Amycolatopsis*、*Kutzneria*
		Actinomycetales	*Actinomyces*
	Firmicutes	Bacillales	*Bacillus*、*Paenibacillus*
		Eubacteriales	*Clostridium*
		Lactobacillales	*Lactobacillus*
	Bacteroidetes	Flavobacteriales	*Chryseobacterium*
		Bacteroidales	*Bacteroides*
	Proteobacteria	Burkholderiales	*Burkholderia*、*Caballeronia*、*Paraburkholderia*
		Xanthomonadales	*Lysobacter*
		Pseudomonadales	*Pseudomonas*
		Myxococcales	*Myxococcus*、*Cystobacter*、*Stigmatella*
	Cyanobacteria	Nostocale	*Anabaena*、*Nostoc*
		Pleurocapsales	*Pleurocapsa*
Fungi	Ascomycota	Pleosporales	*Bipolaris*、*Cochliobolus*、*Alternaria*
		Aspergillaceae	*Aspergillus*、*Penicillium*
		Hypocreales	*Fusarium*、*Tolypocladium*
		Xylariales	*Pestalotiopsis*
		Mycosphaerellales	*Cercospora*
		Cladosporiales	*Cladosporium*
		Magnaporthales	*Magnaporthe*
		Ostropales	*Cyanodermella*

Note: Data from references [13][15][17]etc.

根据泛基因组学的定义,编码次级代谢途径等菌株生长非必需的相关基因组成了特定微生物种属的非必需基因组[18],它们反映了物种的多样性和进化优势。因此作者认为,富含 BGCs 的微生物类群应具有高度开放的泛基因组,进化速度快,基因得失、倍增、水平转移或基因组重排事件频繁,同时也造就了更加丰富的种及以下分类单元的多样性。这一论断也得到链霉菌属和曲霉属等具有庞大种群数量的支持。截至 2022 年 5 月 28 日,LPSN 数据库公布链霉菌属的有效种为 696 个,也是 MIBiG 数据库中 BGCs 条目最多的属,达到 638 个; 曲霉属超过 750 个种(NCBI 中数据) ,是 MIBiG 数据库中 BGCs 条目第二多的属,为 88 个; 芽孢杆菌属和假单胞菌属的有效种分别为 103 和 299 个,对应 MIBiG 数据库中条目数分别为 56 个和 68 个。总之,上述富含 BGCs 类群往往具有较强的开放型泛基因组,且拥有较大数量的物种组成和多样性[19-20]。

"潜力菌株"所在类群具有高度开放的泛基因组,这是它们适应环境多样性的结果。比如评估链霉菌属泛基因组的流动性值揭示了它们暴露于不同生活方式和栖息地而形成的巨大多样性,它们容易通过水平转移获得遗传物质,从而更好地适应环境[19]。毫无疑问,这也导致了链霉菌较大的基因组和丰富多样的蛋白质编码基因。同样,弗兰克菌属泛基因组的流动性值高达 0.9,说明该群体的个体拥有更多独特的基因[21],鉴于该属是富含 BGCs 的类群,提示该属的 BGCs 分布具有亚种甚至菌株的特异性。

特殊生境即特殊生态环境(special eco-environment) ,指在结构和功能上具有明显的异质性或特殊性,并导致生态元的数量或品质明显不同的生态环境。因此特殊生态环境决定了微生物的物种多样性,也塑造了它们独特的基因组。一个有趣的发现是链霉菌属的两株菌 CNQ-509 和 WAC 06738 属于同一进化分支,但它们分别来自海洋和土壤,这种不同生境来源造就的主要区别是它们基因组中 NRPS(non-ribosomal peptide synthetase) 和 I 型 PKS(polyketide synthase) 基因簇的不同数量[19]。可见,特殊环境对微生物群体中独特 BGCs 的塑造是特境微生物作为新 NPs 重要源泉的根本。

什么样的特殊生境更容易塑造出"天才"或"潜力"微生物呢? 作者认为不应该是极端的生态环境,而应该是生物多样性丰富、化学信号交流频繁,且具有一定空间独立性的微生境。其中植物内生的真菌具有成为高产 NPs 潜力菌株的巨大天然优势。主要原因如下: ①地球上庞大的植物王国种类超过 35 万种,甚至不同地理环境和气候下的同种植物、同种植物的不同组织器官对其内生微生物都是一个独特的微生境,必然塑造出具有个体特异性的内生菌种群。可见,植物内生菌(包括真菌) 是一个极其丰富多样的资源,同样它们也蕴含着数量庞大的 NPs。②在长期的进化过程中内生真菌与其植物宿主逐渐形成了复杂的共生关系,它们参与了植物免疫系统的防御反应[22]。还有一些药用植物内生真菌,它们在生长过程中会产生与宿主相同或相似的生物活性成分[23],提示内生真菌相比其他微生物产生药用 NPs 的比例更高。③随着基因组学的发展,人们认识到,微生物 NPs 的丰富度与其基因组大小成正相关[11],而真菌作为基因组最大的微生物类群之一,蕴含着更加丰富多样的 BGCs[12]。可以预见,植物内生真菌作为自然选择潜力菌株的大本营,将会成为未来药用 NPs 发掘的重要贡献者。

总之,作者建议富含 BGCs"潜力菌株"的选择应优先遵循以下原则: ①根据分类学导向,优先选择公认富含 BGCs 的类群(以属及属内特异分支为佳) ; ②优选微生物的所在类群应广泛分布于不同生境且物种数量较多,具有开放的泛基因组; ③依据系统进化树,利用与目标菌株处于同一或相近分支参考菌株或类群的基因组数据,参照预测目标菌株 BGCs 的潜能(以有可利用的种及种内菌株的参考基因组为佳) ,通过预测结果选择潜力菌株; ④对于同一进化分支,微生物的去重复化非常必要,但根据相应类群的泛基因组学特征应有不同标准,如流动值高的要重视不同来源、不同表型的菌株; ⑤除了表 1 提到的主要类群,富含 BGCs 的小众微生物类群也有不少值得系统开发(如异壁放线菌属、疣孢菌属等) 。尽管本文更关注可培养的微生物类群,但相信随着宏基组学研究的深入,将有更多尚未被培养的蕴含丰富 BGCs 的类群被不断发现。

2 生物信息学预测

对于潜力菌株 BGCs 多样性、新颖性和合成产物结构的精准预测是微生物 NPs 发掘的基础与重点。研究初期,基于关键合成基因序列导向的分析方法,逐渐拉开了微生物 NPs 新结构预测的帷幕。如 Hornung 等通过对含有新型 $FADH_2$ 依赖型卤化酶的系统发育分析,推定相应卤化酶催化产生新型卤代化合物,并使用质谱验证了次级代谢物类别和卤化酶基因序列之间的显著相关性,从而快速发现了新的多环型 xanthone 类 NPs[24]。随后美国 Scripps 海洋研究所 Jensen 研究小组基于 PKS 或 NRPS 缩合结构域 KS(keto-synthase) 或 C(condenstion) 的系统发育树而建立的在线分析方法 NaPDoS (http: //napdos. ucsd. edu/) ,可以预测聚酮或非核糖体肽类化合物的类别与结构[25]。尽管 NaPDoS 已被新的工具替代,但以关键合成基因序列导向的预测方法,在其他新颖同工酶及其催化合成产物的预测中仍持续发挥着重要作用[26]。

近十年来,随着基因组、宏基因组的测序成本不断降低,公共数据库中积累了大量可参考的基因组序列。同时越来越多 NPs 生物合成途径的阐明与新颖酶催化机制的解析,使得利用现有数据库预测 BGCs 成为可能。因此,催生了许多优秀的生物信息学程序包,如能快速鉴定整个微生物基因组中 BGCs 的 antiSMASH(antibiotics & secondary metabolite analysis shell) [27];采用隐马尔科夫模型(hidden markov model,HMM) 搜索策略,识别真菌基因组中 PKS、NRPS、hybrid-PKS/NRPS 和萜类 BGCs 中保守结构域的 SMURF (secondary metabolite unknown regions finder) [28] 等。目前,由诺和诺德基金会生物可持续性中心与丹麦技术大学维护的次级代谢产物生物信息门户网站 SMBP 为大多数 NPs 预测工具提供了一站式目录和链接,为个性化使用挖掘工具与访问数据库提供了便利[29]。SMBP 网站包括了对 PKS/NRPS 结构域进行保守分析和功能预测的 SBSPKS、预测腺苷酰化结构域活性位点结合底物的 NRPSpredictor2、将 PKS/NRPS 基因簇数据与 LC-MS/MS 数据关联的集成平台 PRISM/GNP,以及使用系统发育基因组学来识别 BGCs 的 EvoMining 等。其中,antiSMASH 能够快速鉴定和分析微生物基因组中的 BGCs,并保持其功能的持续扩展与更新[27]。它利用最小信息和存储库 MIBiG 中已知功能 BGCs,通过 KnownClusterBlast 模块比较分析已知 BGCs 来实现预测,并集成与交互连接了 NCBI BLAST + 、HMMer 3 、Muscle 3 、FastTree 、PySVG 和 JQuery SVG 等开源工具,成为当前最受青睐和使用最广泛的 BGCs 预测集成工具[30-31]。

值得注意的是,作者在使用 antiSMASH 挖掘真菌的 NPs 时,发现其对预测真菌 BCGs 中功能基因注释的准确率并不高,建议与 SMURF 预测结果比较确认,必要时还需结合 2ndFind 进行人工注释。2ndFind 作为 FramePlot 4. 0beta 的替代工具由日本国立传染病研究所维护,是利用已出现次级代谢相关蛋白的 Pfam(protein family) 结构域来搜索次级代谢生物合成蛋白的在线注释工具 (https: //biosyn. nih. go. jp/2ndfind/) 。尤其它通过参考已知同源物种的次级代谢相关蛋白数据,并关联了 BLASTP 功能,可以实现基因编码蛋白与 SwissProt 数据库中蛋白序列的快速比对,从而获得更加严谨的功能注释,是人工注释基因功能的有力工具。

为了进一步挖掘微生物中蕴含 BGCs 的巨大潜能,Cimermancic 等提出了基于已知 BGCs 中出现的蛋白质家族 Pfam 结构域频率的新算法 ClusterFinder,使用 HMM 模型可在全基因组范围内预测潜在的 BGCs,为非连锁 BGCs 协同合成未知类型化合物的预测提供可能[32]。他们从 1 154 个原核生物基因组预测得到 33 351 个可能的 BGCs (评估的假阳性率为 5%) ,高置信度的 BGCs 为 10 724 个,其中 7 377 BGCs 是在 antiSMASH 未被预测到的,并发现了尚未被表征的合成芳香多烯羧酸这类最大的 BGC 家族,这极大地提升了基因组挖掘的潜能[32]。而当 antiSMASH 选择加载 ClusterFinder 算法后,Swift 等从肠道新美鞭菌纲的厌氧真菌基因组中相比 SMURF 预测出更多的生物合成基因和 NPs 类别,包括细菌素和脂肪酸类及糖类衍生物等[33]。可以预见,这些新算法的出现将为混合培养协同产生新颖 NPs 的预测提供分析工具。

当前,随着对先导分子药用功能的日趋重视,基因组挖掘技术从追求 NPs 的化学多样性向高

效药理活性转变。2019 年，Yan 等使用抗性基因为导向的基因组挖掘策略从微生物中发现了一种新型天然产物除草剂 aspterric acid，其通过靶向植物支链氨基酸合成途径中二羟酸脱水酶而显著抑制植物的生长[34]。最新的人工智能与基因组挖掘技术相结合，更是为 NPs 结构多样性高效挖掘与功能利用提供了典范。我国王军和陈义华团队整合了多种人工智能领域中自然语言分析(natural language processing) 的神经网络方法，通过深度机器学习构建了准确率超过 90% 的抗菌肽预测模型，利用该模型对 1 万多个微生物组进行了小蛋白预测和抗菌肽挖掘，最后确定 240 余种为潜在新型的抗菌肽候选分子，并合成了其中 216 种; 经实验验证，其中 181 种新型抗菌肽具有抗菌活性，发掘准确率达 83.8%，部分活性接近迄今 NPs 中真核抗菌肽最强活性的表现，并对革兰阴性多重耐药菌具有较强的抑菌能力，在感染肺炎克雷伯菌的动物模型中有 3 个抗菌肽展现出良好的体内治疗作用和安全性[35]。综上可见，利用人工智能赋能微生物基因组挖掘深度拓展了自然界潜在跨物种 BGCs 和 NPs 总库容的认知，为 NPs 结构与生物活性关联提供线索，显著提高了微生物药物发掘的效率，是未来 BGCs 精准预测及其 NPs 定向发掘的发展方向。

3 沉默生物合成基因簇的激活

微生物具有丰富的编码次级代谢产物的 BGCs，这赋予其生产新结构药用先导化合物的巨大潜能[36-37]，但是广泛存在“隐性”BGCs 在常规条件下是不表达或低表达的[38]。当前，高效基因编辑技术和合成生物学的兴起，给基因组挖掘带来了巨大的机遇。沉默 BGCs 激活的策略很多，且被研究者反复总结[13-14,39]。按照它们对激活 BGCs 的靶向性，可以分为靶向性激活和非靶向/多效性激活两大类[39]。这里按照激活作用的层次和位点，将内源性激活策略归纳为细胞水平(菌间的混合培养和菌内的培养调控) 、细胞器/基因簇水平(表观遗传学、核糖体工程和代谢分流等) 、基因水平(全局性调控基因、途径特异性调控基因、生物合成基因和前体供应相关基因等) 、转录元件(启动子工程和转录因子诱饵) 、蛋白水平(突变生物合成和翻译后修饰) ，外源性激活归纳为直接克隆、改造 BGCs 调控及优化宿主表达体系的异源表达，以及无需宿主细胞的体外重构(无细胞代谢工程系统和从头合成的无细胞合成系统[40]) (表 2) 。

上述激活方法均有良好的激活效果且各有特点。比如，内源的靶向激活策略，能特异性激活，但需要全面认识目标 BGCs，以及具备可行的遗传操作体系。表观遗传学和核糖体工程既可采用化学诱导剂，也可采用基因编辑的方式实现，但往往会出现“效应热点”，导致不同菌株的激活效应趋同。培养调控和混合培养随机性较强，但简单易行、使用广泛，且不受 BGCs 序列信息未知的影响，甚至可激活无法通过生物信息学预测的 BGCs; 当然它存在激活的“报酬递减效应”，当培养激活条件达到一定数量后，出现新激活效应将越来越难。BGCs 激活时强化基因功能的策略有基因水平的基因倍增(过表达) 和高活性基因置换(组合生物合成) 、转录水平的强启动子置换、蛋白水平有利于提高酶活性的突变生物合成。在实际应用中这些激活方法经常联合交互使用，如基于遗传改造的激活方式与培养调控策略联合使用[41]，以尽可能多地激活 BGCs，提高微生物 NPs 的多样性，增加发现结构新颖药用先导化合物的几率。

尽管许多 BGCs 激活策略已被大家熟知并广泛应用，但近几年也提出了一些有启发意义的激活策略。如 BGCs 中的关键生物合成蛋白需要翻译后修饰才能发挥活性，在 PKS 和 NRPS 中的酰基或肽基载体蛋白 (carrier protein，CP) 初始为无活性的脱辅基 *apo*-CP 形式，需要通过磷酸泛酰巯基乙胺基转移酶 (phosphopantetheinyl transferase，PPTase) 修饰，将供体辅酶 A 上的 4-磷酸泛酰巯基乙胺转移到 CP 保守的丝氨酸残基侧链羟基上，才能成为有活性的全酶 *holo*-CP 形式。利用这一原理，我国瞿旭东团队首次构建了基于 CP 活性调控的翻译后修饰激活策略，通过在放线菌中过表达 PPTase 激活元件，实现约 70% 研究菌株中代谢物产量的增加[42]。转录因子诱饵 (transcription factor decoy，TFD) 是含有转录因子结合位点的一段核酸序列，其能与抑制型转录因子结合，使阻遏蛋白调控的沉默 BGCs 实现正常转录表达。基

于此，研究人员构建了具有 TFDs 的载体，经结合转移，稳定游离的 TFDs 载体在链霉菌体内能引诱抑制转录的阻遏蛋白分子离开，成功激活了 8 个 PKS 和 NRPS 类型的沉默 BGCs[43]。该研究为大型沉默 BGCs 的转录激活提供了新策略。此外，先进的 CRISPR/Cas9 工具盒，不仅在启动子工程激活等高效基因编辑中广泛使用，而且利用催化活性丧失的 Cas9 突变体（dCas9），将 dCas9 蛋白与转录激活因子和抑制因子融合，发展了激活基因转录的 CRISPRa（CRISPR/dCas-mediated transcriptional activation）和干扰基因转录的 CRISPRi（CRISPR interference）等系统，可调控微生物中次级代谢物生产的转录效率，实现沉默 BGCs 的可控表达[44-45]。

表 2　沉默 BGCs 激活的策略与方法

Table 2　Strategies and methods to activate silent BGCs

Host	Activation levels and sites		Activation approaches	Activation specificity
Endogenous host	Cell level	Inter-strain	Co-cultures	Untargeted
		Intra-strain	Culture regulation using OSMAC approach	Untargeted
	Organelle or gene cluster level	Ribosome, RNA polymerase, histones, signaling pathway related genes or BGCs	Epigenetic modification	Untargeted
			Ribosome engineering	Untargeted
			Altered signal transduction pathways	Untargeted
			Block competing pathways or metabolic shunting	Untargeted
	Gene level	Globally regulated genes	Duplication or replacement of positively regulated genes	Untargeted
			Knockout or silencing of negatively regulated genes	Untargeted
		Pathway-specific regulatory genes	Duplication or replacement of positively regulated genes	Targeted
			Knockout or silencing of negatively regulated genes	Targeted
		Biosynthetic structural genes	Duplication or replacement of biosynthetic genes	Targeted
		Precursor synthesis related genes	Enhancing precursor supply	Untargeted
	Transcriptional level	Promoter element	Promoter engineering	Both
		Transcription factors	Transcription factor decoy	Both
	Protein level	Biosynthesis related proteins	Mutant biosynthesis or directed enzyme evolution	Both
		Carrier protein	Post-translational modification	Untargeted
Heterologous host	Gene cluster level	Intra-strain BGCs	Heterologous expression by direct cloning	Targeted
			Heterologous expression via enhanced transcription	Targeted
	Gene level	Inter-strain BGCs	Synthetic pathway refactoring *in vivo*	Targeted
Free host	Protein level	Biosynthesis related proteins	CFME synthesis or synthetic pathway refactoring *in vivo*	Targeted
	Molecular level	Small organic molecules	*De novo* synthetic system by coupling CFPS and CFME	Targeted

Note: OSMAC, one strain-many compounds; CFME, cell-free metabolic engineering; CFPS, cell-free protein synthesis

事实上，次级代谢产物 BGCs 沉默或低表达的原因具有复杂性和多样性。正常的 BGCs 受到细胞内不同层级调控系统的严格控制，只有在特定的环境或培养条件下表达；同时其表达产物的产量水平（也可理解为产物对宿主细胞的反馈效应）可能取决于宿主细胞的稀有 tRNAs、密码子偏好性、DNA 修饰及各代谢途径的通量，也可能取决于产物稳定性以及其对宿主的毒性等因素。同时，BGCs 中任何关键合成基因沉默或编码蛋白功能失活，也能导致 BGCs 沉默。因此，BGCs 的激活策略也多种多样，不局限于这里提及的方法，凡是能避开现有宿主不同层级的调控系统，强化 BGCs 合成基因和前体供应相关基因的表达，以及编码蛋白活性提高的策略均可应用于沉默或低表达 BGCs 的激活。相信随着人们对 BGCs 沉默机制的认识不断加深，基因高效克隆、组装和编辑等新兴技术的不断推出，在合成生物学和人工智能的助推下，深入挖掘沉默 BGCs 蕴含的“暗物质”将使新 NPs 的发现更上一层楼。

4 激活产物的识别

尽管许多基因组挖掘策略在激活微生物 BGCs 中展现出良好的效果，但是在新颖先导化合物的高效发掘中，仅仅依靠激活目标 BGCs 或增加次级代谢产物的多样性是远远不够的，更重要的是激活 BGCs 及其产生化合物的高效识别。众所周知，从 BGCs 到活性物质的产生，要经历基因的转录、翻译和酶催化生物合成等阶段，因此在转录水平、蛋白水平和代谢物水平的检测，均能表征 BGCs 的激活情况。尤其是转录组、蛋白组、代谢组的发展，加速了 BGCs 表达的识别及其合成 NPs 的发现。比如，通过转录水平的挖掘能高效筛选出 BGCs 激活的方式以及相应的培养条件[46]。在蛋白水平，Gubbens 等基于定量蛋白质组学技术建立了天然产物蛋白质组学发掘策略，该方法将生物合成蛋白水平的变化与次级代谢产物的产生联系起来，还能利用生物活性和蛋白质表达谱之间的相关性，快速发现负责生成目标 NPs 的 BGCs[47]。

编码次级代谢产物的 BGCs 是否表达受到转录、翻译、翻译后修饰、功能蛋白质折叠组装等复杂机制的控制，所以转录或蛋白水平的检测，可以明确沉默 BGCs 激活时在每个阶段的表达状况，但并不能完全体现 BGCs 终端代谢产物的分泌情况。因此，基于代谢水平的高效识别方法对于微生物 NPs 的发现至关重要。目前没有一种化合物的分析方法可以无偏向性地涵盖所有代谢物的识别，基于色谱（LC 和 GC）、紫外吸收光谱（UV 或 DAD）、荧光光谱、红外光谱、质谱（MS）、核磁共振（NMR），甚至特定基团的显色反应都能作为激活产物识别的手段。包括 Thin layer chromatography（TLC）、HPLC-UV、LC-MS、GC-MS、LC-NMR、NMR-PCA[48] 等，但它们在分辨率、灵敏度、通量等方面各有特点。

基于质谱的分析具有灵敏度高、分析速度快、动态范围宽、结构信息丰富、样品用量少和通量高等优势，成为代谢物识别的主流分析平台，也在代谢组学研究中发挥着重要作用。质谱分析发展非常迅速，可根据化合物的理化性质、结构类型和检测目的，选择适宜的离子源（如 EI、ESI、CI、APCI、FI、FAB、MALDI）和质量分析器（如 Magnetic sector、Quadrupole、Ion traps、Time of flight、FTICR）。目前在微生物 NPs 领域使用最广泛、识别激活产物最高效的是色谱-紫外光谱-质谱联用技术（HPLC-DAD-MS），它通过分离粗提物、根据色谱峰的紫外吸收谱特征进行化合物分类、并结合质谱数据识别和发现目标激活化合物。如作者通过基因组挖掘发现芽孢杆菌 SCSIO 05746 具有产生 4 种 NRPS 类物质和 3 种 PKS I 型产物的潜能，然后使用 OSMAC 策略的培养调控进行这些 BGCs 的激活，采用 HPLC-DAD-HRESIMS 成功筛选到同时激活其中六种类型化合物的培养条件，并在正负离子扫描模式下对不同离子峰进行表征，识别了一系列新的 bacillibactin、fengycin、bacillaene 和 macrolactin 类型的衍生物，也为激活目标化合物的跟踪分离与结构鉴定提供指导[49]。

尽管上述色谱-质谱联用技术在产物识别上取得了很大成功，但一方面仅基于一级质谱数据很难获得可靠的化学注释，另一方面对原始数据采用人工识别和分析的方式，非常耗时费力。特别是对某一类群微生物 NPs 进行系统发掘时，不可避免地出现许多 NPs 的重复发现，不同的微生物类群在相同激活条件可能存在趋同效应，那么如何高效地排除不同菌株、不同激活条件下重复

的 NPs,发现新的目标分子? 如何高效评价众多激活策略的效果,筛选更佳的激活方法? 美国加州大学圣地亚哥分校的 Dorrestein 和 Bandeira 团队开发了以 HPLC-HRMS/MS 为基础,结合了质谱碎裂数据分析和数据库搜索功能的分子网络技术(molecular networking,MN) ,较好地解决了上述问题。该方法根据 MS/MS 质谱碎片的相似性建立了可视化分子网络图,直观显示了未知样品中各组成成分之间的关系和化学分类,解读了微生物表型和基因型之间的复杂关联[50]。随后他们倡导建立了 GNPS(global natural products social molecular networking) 开放网络数据库(http: //gnps. ucsd. edu) ,供人们利用已处理或已识别的串联质谱数据(MS/MS) ,快速识别质谱图中已知成分和相关成分,从而促进目标分子的识别和追踪[51]。目前该技术应用十分广泛,已从微生物药用先导化合物发现扩展到整个药物发现和药物代谢等领域[52]。作者应用 MN 技术,从海绵来源链霉菌 18A01 中高效识别并靶向分离一系列具有显著已糖激酶 II 抑制活性的新颖 α-吡喃酮类衍生物[53]。

尽管如此,MN 也很难完整、准确地注释二级质谱中的庞大数据信息,而人工分析每个节点的碎片信息不能满足大数据集进行化学注释的需求。因此,面对如何更好注释代谢物的挑战,科学家开发了新的计算方法和碎片预测方法,极大提高了质谱数据的解析效率,实现了代谢产物的高效识别与解析[54]。如 MS2LDA 一种受文本挖掘启发的方法,将碎片光谱分解为一组保守的碎片和中性丢失特征(称为 Mass2Motifs) ,它能识别同时出现与结构基序相关的离子,并将其分解成一组特征子结构 Mass2Motifs,表示结构家族。目前还提供了用户可以 Web 访问的应用程序 ms2lda.org[55]。DEREPLICATOR 通过比较基于特定计算机碎片规则生成的理论光谱,系统地将肽类天然产物数据库中的结构与质量碎片光谱联系起来,能够注释具有相似氨基酸组成的肽类 NPs。最近推出的 DEREPLICATOR + ,将这种注释策略扩展到聚酮、黄酮、萜烯和其他类别的 NPs[56]。网络注释传播 (network annotation propagation,NAP) 基于对重叠结构指纹最合理的候选结构进行重新排序来改进类似物的注释准确性,通过光谱网络(spectral network) 传播谱库匹配,进行计算机化学注释,实现 NPs 的高效识别与靶向分离[54]。

上述工具增强了从数据集中获取化学结构注释的能力,但是它们都有其自己的输出格式,阻碍了有效数据的匹配组合。因此,MolNetEnhancer 整合 MN、MS2LDA、NAP 等的数据,通过 ClassyFire 对数据集中存在的类别进行自动化学分类,为代谢组学的数据提供更全面的化学概览,同时还能显示每个子结构模式的分子细节[57]。目前,这些先进的计算机注释工具已经被整合到 GNPS 平台,极大地弥补了现有公共存储库参考碎片谱不足的缺陷。使用增强的分子网络技术,Maimone 等研究了立枯丝核菌共同培养诱导放线菌 *Streptomyces lunalinharesii* A54A 产生具有抗真菌活性的次级代谢物,发现了仅在共培养时产生的 158 离子,并注释了去铁胺铁载体和茴香霉素衍生物两类物质,自动化化学分类发现最丰富的离子来自羧酸和衍生物(n = 31) 、异戊二烯醇脂质(n = 15) 等,以及共培养时产生相对数量更多的六种新结构化合物[58]。这一研究也为微生物微量激活产物的高效识别提供了有益启发。

以上详细论述了普适性最强、使用最广泛的非靶向代谢物检测与注释的最新进展,但还有适用于不同场景和目标的其他化学识别技术。比如被称为“分子显微镜”的质谱成像技术(imaging mass spectrometry,IMS) 允许直接从生物样品中二维可视化代谢物、表面脂质、肽和蛋白质的分布,主要包括基质辅助激光解吸电离(matrix assisted laser desorption ionization,MALDI) 质谱成像、解吸电喷雾电离(desorption electrospray ionization,DESI) 质谱成像以及二次离子质谱(secondary ion MS,SIMS) 成像等[59]。其中 MALDI-IMS 由于空间分辨率高和检测分子量范围广,在微生物肽类分子识别中广泛应用[60]。还有基于 NMR 精准预测结构的技术,如小分子精确识别技术 SMART (small molecule accurate recognition technology) 采用卷积神经网络将 HSQC(heteronuclear multiple-quantum coherence) 光谱与实验光谱库进行比较,识别结构相似的分子[61];MADByTE(metabolomics and dereplication by two dimensional experiments) 则是基于 TOCSY(total correlation spectroscopy) 和 HSQC 光谱来识别复杂混合物中的独立自旋体

系，通过匹配样品间的自旋系统特征创建各样品的化学相似性网络，通过匹配已知化合物光谱特征实现化合物去重复，优先识别有生物活性的成分[62]。值得一提的是 TLC-bioautography 和 BioMAP(antibiotic mode of action profile) 筛选模式是将化学特性与生物活性关联的识别方法，有助于活性 NPs 的快速发现。例如，BioMAP 高通量平台通过匹配 15 种临床病原菌的抑制活性，依据抗菌谱类别实现了复杂混合物中潜在抗菌 NPs 的分类与识别，并高效和低成本地发现了结构独特的萘醌抗生素[63]。

总之，基于 MS 的化合物识别技术在不断发展，尤其当前机器学习赋能非靶向代谢组学的化学注释技术，实现了代谢物质谱数据集的高通量处理与化合物精准注释，将在微生物新 NPs 的发掘中发挥越来越重要的作用。当然，其他类型的化合物识别技术也各有特点，研究者应该根据实际需求选择使用。

5 系统学导向微生物产物发掘线路设计与展望

当前，癌症和感染性疾病严重威胁人类健康。据统计，全球每年有近 1 000 万癌症死亡病例[64]，我国更是全球癌症高发率和高死亡率的“重灾区”[65]。在全球十大健康威胁中，和感染密切相关的多种疾病和行为占据了前 10 位中的 6 个[66]。同样，在农业领域，作物病害给农业生产造成了巨大的损失。尤其是植物真菌病害，按最低发病水平计算，可造成世界五大粮食作物每年减少 1.94 亿 t[67]。可见，研制治疗癌症、感染性疾病的新型药物以及防治植物病害的新型绿色农药的需求巨大而迫切。另一方面，独特 NPs 的多样性是新药创制的源泉和基础，尤其在现代基因编辑技术、各类组学技术和人工智能的支持下 NPs 药物展现出巨大的发展潜能。因此，在后基因组时代 NPs 药物必将在需求和新技术的驱动下迎来又一个发展的黄金时期。

那么在新时代如何基于 NPs 的化学多样性，从中高效发掘微生物药用先导化合物是一个重要的基础性科学问题。为此，作者提出了基于潜力菌株优选和多组学导向的高效发掘新颖药用先导化合物的集成研究体系 SPLSD(Systematic Pipeline for efficient Lead Structure Discovery from microbial natural products by promising strain selection and multi-omics mining) (图 2) 。首先选择动植物共附生微生物为优选的研究对象，测定其 16S RNA 或 ITS(internal transcribed spacer) 基因序列并构建系统发育树，初步认识其分类学地位以及基因型的独特性; 然后文献调研或采用 NCBI 基因组数据分析其所在类群的泛基因组特征，依此进行分支菌株的选择和去重复(流动值低的可在种或亚种层次去重复，流动值高的要重视每一个独特来源的基因型菌株) ，并利用相近类群泛基因组数据参考分析选择菌株的 BGCs 丰富度、主要类别和新颖性，优选泛基因组开放程度高、流动值大、参考基因组富含新颖 BGCs、且该进化分支或类群 NPs 研究尚不充分的菌株作为目标潜力菌株“promising strains”; 对于所在类群泛基因组流动值较低且与参考菌株高度同源的，可以使用参考菌株基因组进行目标菌株 BGCs 的预测，对于相应流动值高的则还需要对选择的潜力菌株进行全基因组测序和生物信息学分析，精准认识其合成新颖次级代谢产物的潜能和特异性; 随后对目标潜力菌株采用多种靶向或随机激活的策略(不局限于图 2 中展示的激活策略) ，改变其化学多样性或激活目标产物，并通过不同表达水平检测(转录和代谢水平) 和效应物功能筛选，识别目标药用次级代谢产物的激活条件; 最后使用活性导向和化学追踪的手段对目标产物完成分离、鉴定和活性评价，并通过结构优化不断改善成药性，实现新颖药用先导化合物的高效发掘。

针对 SPLSD 挖掘线路还有几点补充说明。①对于无需菌株本身基因组信息的激活策略，根据实际情况可不测定潜力菌株的基因组序列，但毫无疑问，基因组测序有助于全面了解选择菌株的潜能和特异性，也有助于利用靶向激活策略系统地发掘其蕴含的目标 NPs。② antiSMASH 等基于规则的预测工具对于含有尚未被阐明的生物合成新机制以及非线性催化合成的 BGCs 很难实现准确预测; 同时在实践中往往由于化学结构多样性与其生物合成机制阐明的滞后性，导致生物信息学预测的新颖 BGCs 并不保证能产生新 NPs。因此，基因组序列信

息的挖掘与激活产物的化学识别相结合，不仅能筛选激活培养条件，还能互相印证、促进新 NPs 的发掘及其生物合成途径的阐明。③多样性导向的挖掘方法对于发现新颖 NPs 非常有效，但 NPs 的功能也应被优先考虑，因此针对特定功能和结构类型的 NPs 进行靶向发掘是非常必要的，也是 SPLSD 体系所提倡的。④非靶向 NPs 挖掘或靶向 NPs 识别困难时，对激活或多样性提高的提取物进行充分的活性筛选有助于提早发现 NPs 的功能和潜在应用方向。但对于为数不多的提取物进行多模型活性筛选是不经济的，因此非常有必要建立化合物差异的提取物库，并对其持续进行不同活性模型的高通量筛选，实现高值 NPs 的高效发现。

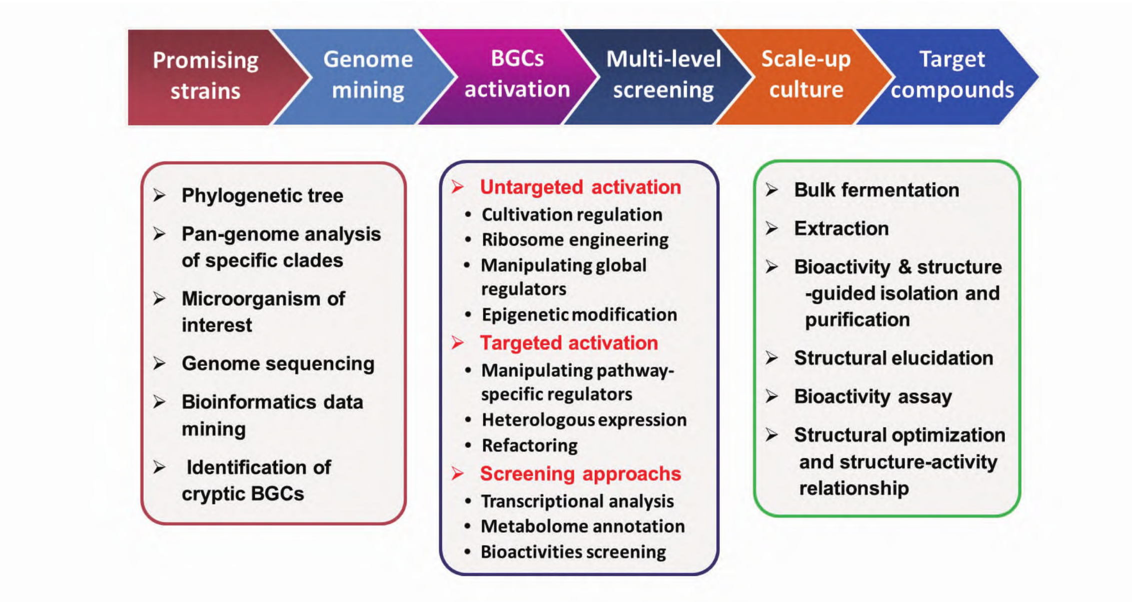

图 2 基于潜力菌株优选和多组学挖掘的高效发现微生物天然产物先导结构的研究系统(SPLSD)

Fig. 2 The systematic pipeline for efficient lead structure discovery from microbial natural products by promising strain selection and multi-omics mining (SPLSD)

体系中的激活策略不仅仅局限于图中列举的方法，任何能避开宿主的调控系统，强化 BGCs 合成基因和前体供应相关基因的表达，以及编码蛋白活性提高的策略均可应用于隐性 BGCs 的激活

The activation approaches are not limited to these methods listed in the figure. Any strategy that can avoid the host´s regulatory system, enhance the expression of biosynthesis genes and precursor supply-related genes, and improve the catalytic activity of encoded proteins can be used to activate cryptic BGCs

在过去的实践中，结构多样性导向的新 NPs 发掘，会产生大量“无活性”NPs。据统计，过去五年海洋来源的 7 500 多种新 NPs 中绝大部分进行了细胞毒性、抗菌、抗真菌或抗炎活性的筛选，但占总数 68% 的新 NPs 不具有上述活性[68]。当前 NPs 生物活性发掘侧重于癌症、微生物感染和炎症，这严重限制了 NPs 化学多样性在新出现代谢、免疫和神经退行性等疾病治疗，以及增强植物免疫、抗逆性和促生等方面的应用。作者倡导的 NPs 系统发掘线路 SPLSD 也强调扩展化学多样性与活性筛选相结合的重要性。小分子 NPs 与生物在自然界中长期共同进化，其生物学功能影响着宿主细胞特定的生理代谢过程，NPs 作为“天生我才必有用”的自然进化产物，应充分进行各类生物活性模型筛选，相信随着通过更多细胞、酶/蛋白等不同功能和水平模型的高通量筛选，甚至虚拟药物筛选，NPs 将会在生命科学研究、医药健康与现代农业生产中发挥越来越重要的作用。

还有，尽管 NPs 因为其结构的多样性和复杂性长期以来都是创制新药的重要来源和启迪，但是 NPs 直接成药却少之又少。据统计，去除 9 个植物源的提取药物，在 1 202 个小分子药物中，NPs 直接成药的 67 个，仅占 5.57%，而 NPs 衍生物药物为 320 个，达到 26.62%[69]。特别在抗感染药(抗菌药、抗真菌药、抗寄生虫药和抗病毒药)中，NPs 衍生物成药和其直接成药的比例高达

83 : 13，提高了约 6.4 倍[70]。农药领域 NPs 直接成药的仅占市场总量的 4%，而以 NPs 为模板或启迪的占到 50%[71]。药用 NPs 表现出“先导强、成药弱”的显著特点。这主要是因为 NPs 来源的药用先导化合物往往存在生物利用度差、代谢不稳定、副作用明显等缺陷。因此，以新颖的药用 NPs 骨架为先导，在保持其结构复杂性和药理活性的基础上，通过持续的结构优化，改善其药学性质，是创制新药的有效途径。而我国作为新 NPs 发现的大国，应加强自主发现 NPs 的化学修饰，通过持续改善其 ADMET 特性(absorption、distribution、metabolism、excretion、toxicity) ，引导实现创新药物的成功开发。

此外，尽管作者强调了可培养微生物中 NPs 的挖掘方法，但是随着宏基因组学、代谢组学、合成生物学和人工智能等的飞速发展，让非培养依赖的 NPs 挖掘技术限制因素越来越少。尤其适合大数据集处理的 ClusterFinder 等新算法和 MolNetEnhancer 等化合物注释工具，揭示了自然界蕴含 BGCs 的无限潜力，使得包含潜在跨物种的“Pan-BGCs”合成的“非自然”NPs 挖掘变成了现实，极大地扩展了自然界总的 NPs 库。同时，基于合成生物学理念 RetroPath、optStoic、RxnFinder 等工具的建立[72]，也为 Pan-BGCs 预测产物的人工合成提供了便利。可以相信，在新技术的推动下非培养依赖的挖掘策略有望成为未来新 NPs 产生的重要途径。

毫无疑问，特境微生物是巨大的资源宝库，其丰富的内涵已表现出了无限潜力。海洋微生物 NPs 经过 20 年的高速发展，现已进入平台期，相信未来动植物共附生微生物，特别是植物内生真菌是挖掘新 NPs 的优良资源和增长点。同时多组学技术和机器学习的不断发展增加了人们对 NPs 产生途径的理解和新 NPs 识别，当前高效的基因克隆与编辑技术、基于合成生物学理念的合成途径重构使得非培养依赖 NPs 的获取变得容易。此外，持续加强 NPs 成药性和可及性研究是更多 NPs 从实验室走向市场的关键。总之，充分利用特境微生物的物种多样性、遗传多样性和化学多样性，整合先进的生物技术、化学技术和计算技术，提高特境微生物先导化合物的发现速度和效率，持续推进 NPs 成药性和可及性的中下游研究，微生物 NPs 必将创造更大的社会效益和经济效益。

参考文献:

[1] Newman DJ, Cragg GM. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019 [J]. Journal of Natural Products, 2020, 83(3): 770-803.

[2] Sparks TC, Bryant RJ. Impact of natural products on discovery of, and innovation in, crop protection compounds [J]. Pest Management Science, 2022, 78(2): 399-408.

[3] Saxena S, Pandey AK. Microbial metabolites as eco-friendly agrochemicals for the next millennium [J]. Applied Microbiology & Biotechnology, 2001, 55(4): 395-403.

[4] Williams PG. Panning for chemical gold: marine bacteria as a source of new therapeutics [J]. Trends in Biotechnology, 2009, 27(1): 45-52.

[5] Shen B. A new golden age of natural products drug discovery [J]. Cell, 2015, 163(6): 1297-1300.

[6] Baltz RH. Marcel faber roundtable: is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration? [J]. Journal of Industrial Microbiology & Biotechnology, 2006, 33(7): 507-513.

[7] Goodfellow M, Fiedler H-P. A guide to successful bioprospecting: informed by actinobacterial systematics [J]. Antonie van Leeuwenhoek International Journal of General and Molecular Microbiology, 2010(98): 119-142.

[8] Gupta S, Chaturvedi P, Kulkarni MG, et al. A critical review on exploiting the pharmaceutical potential of plant endophytic fungi [J]. Biotechnology Advances, 2020(39): 107462.

[9] Carroll AR, Copp BR, Davis RA, et al. Marine natural products [J]. Natural Product Reports, 2019, 36(1): 122-173.

[10] Blunt JW, Copp BR, Keyzers RA, et al. Marine natural products [J]. Natural Product Reports, 2017, 34(3): 235-294.

[11] Baltz RH. Gifted microbes for genome mining and natural product discovery [J]. Journal of Industrial Microbiology & Biotechnology, 2017, 44(4-5): 573-588.

[12] Inglis DO, Binkley J, Skrzypek MS, et al. Comprehensive annotation of secondary metabolite biosynthetic genes and gene clusters of *Aspergillus nidulans*, *A. fumigatus*, *A. niger* and *A. oryzae* [J]. BMC Microbiology, 2013, 13: 23.

[13] Keller NP. Fungal secondary metabolism: regulation, function and drug discovery [J]. Nature Reviews Microbiology, 2019, 17(3): 167-180.

[14] Lyu HN, Liu HW, Keller NP, et al. Harnessing diverse transcriptional regulators for natural product discovery in fungi [J]. Natural Product Reports, 2020, 37(1): 6-16.

[15] Wang H, Sivonen K, Fewer DP. Genomic insights into the distribution, genetic diversity and evolution of polyketide synthases and nonribosomal peptide synthetases [J]. Current Opinion

In Genetics & Development, 2015(35): 79-85.

[16] Bérdy J. Bioactive microbial metabolites[J]. Journal of Antibiotics, 2005, 58(1): 1-26.

[17] Gavriilidou A, Kautsar SA, Zaburannyi N, et al. Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes[J]. Nature Microbiology, 2022, 7(5): 726-735.

[18] Medini D, Donati C, Tettelin H, et al. The microbial pan-genome[J]. Current Opinion in Genetics & Development, 2005, 15(6): 589-594.

[19] Caicedo-Montoya C, Manzo-Ruiz M, Rios-Estepa R. Pan-Genome of the genus *Streptomyces* and prioritization of biosynthetic gene clusters with potential to produce antibiotic compounds [J]. Frontiers in Microbiology, 2021(12): 677558.

[20] Kim Y, Koh I, Lim MY, et al. Pan-genome analysis of Bacillus for microbiome profiling[J]. Scientific Reports, 2017(7): 10984.

[21] Kislyuk AO, Haegeman B, Bergman NH, et al. Genomic fluidity: an integrative view of gene diversity within microbial populations[J]. BMC Genomics, 2011(12): 32.

[22] Yan L, Zhu J, Zhao XX, et al. Beneficial effects of endophytic fungi colonization on plants[J]. Applied Microbiology & Biotechnology, 2019, 103(8): 3327-3340.

[23] Kusari S, Pandey SP, Spiteller M. Untapped mutualistic paradigms linking host plant and endophytic fungal production of similar bioactive secondary metabolites [J]. Phytochemistry, 2013(91): 81-87.

[24] Hornung A, Bertazzo M, Dziarnowski A, et al. A genomic screening approach to the structure-guided identification of drug candidates from natural sources[J]. Chembiochem, 2007, 8(7): 757-766.

[25] 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.

[26] Adamek M, Alanjary M, Ziemert N. Applied evolution: phylogeny-based approaches in natural products research[J]. Natural Product Reports, 2019, 36(9): 1295-1312.

[27] Medema MH, Blin K, Cimermancic P, et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences[J]. Nucleic Acids Research, 2011, 39(suppl 2): W339-W346.

[28] Khaldi N, Seifuddin FT, Turner G, et al. SMURF: Genomic mapping of fungal secondary metabolite clusters [J]. Fungal Genetics & Biology, 2010, 47(9): 736-741.

[29] Weber T, Kim HU. The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production [J]. Synthetic & Systems Biotechnology, 2016, 1(2): 69-79.

[30] Kautsar SA, 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.

[31] Blin K, Shaw S, Kloosterman AM, et al. antiSMASH 6.0: improving cluster detection and comparison capabilities[J]. Nucleic Acids Research, 2021, 49(W1): W29-W35.

[32] Cimermancic P, Medema MH, Claesen J, et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters[J]. Cell, 2014, 158(2): 412-421.

[33] Swift CL, Louie KB, Bowen BP, et al. Anaerobic gut fungi are an untapped reservoir of natural products[J]. Proceedings of the National Academy of Sciences of the United States of America, 2021, 118(18): e2019855118.

[34] Yan Y, Liu Q, Zang X, et al. Resistance-gene-directed discovery of a natural-product herbicide with a new mode of action [J]. Nature, 2018, 559(7714): 415-418.

[35] Ma Y, Guo Z, Xia B, et al. Identification of antimicrobial peptides from the human gut microbiome using deep learning [J]. Nature Biotechnology, 2022,40(6):921-931.

[36] Udwary DW, Zeigler L, Asolkar RN, et al. Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*[J]. Proceedings of the National Academy of Sciences of the United States of America, 2007, 104(25): 10376-10381.

[37] Chaudhry V, Chauhan PS, Mishra A, et al. Insights from the draft genome of *Paenibacillus lentimorbus* NRRL B-30488, a promising plant growth promoting bacterium[J]. Journal of Biotechnology, 2013, 168(4): 737-738.

[38] Park HM, Kim BG, Chang D, et al. Genome-based cryptic gene discovery and functional identification of NRPS siderophore peptide in *Streptomyces peucetius*[J]. Applied Microbiology & Biotechnology, 2013, 97(3): 1213-1222.

[39] Kalkreuter E, Pan G, Cepeda AJ, et al. Targeting bacterial genomes for natural product discovery[J]. Trends in Pharmacological Sciences, 2020, 41(1): 13-26.

[40] Zhuang L, Huang S, Liu W, et al. Total *in vitro* biosynthesis of the nonribosomal macrolactone peptide valinomycin [J]. Metabolic Engineering, 2020(60): 37-44.

[41] Wei Q, Bai J, Yan D, et al. Genome mining combined metabolic shunting and OSMAC strategy of an endophytic fungus leads to the production of diverse natural products[J]. Acta Pharmaceutica Sinica B, 2021, 11(2): 572-587.

[42] Zhang B, Tang W, Wang S, et al. Activation of natural products biosynthetic pathways via a protein modification level regulation[J]. ACS Chemical Biology, 2017, 12(7): 1732-1736.

[43] Wang B, Guo F, Dong SH, et al. Activation of silent biosynthetic gene clusters using transcription factor decoys[J]. Nature Chemical Biology, 2019, 15(2): 111-114.

[44] Peng R, Wang Y, Feng Ww, et al. CRISPR/dCas9-mediated transcriptional improvement of the biosynthetic gene cluster for the epothilone production in *Myxococcus xanthus*[J]. Microbial

Cell Factories, 2018(17): 15.

[45] Zheng Y, Su T, Qi Q. Microbial CRISPRi and CRISPRa systems for metabolic engineering[J]. Biotechnology & Bioprocess Engineering, 2019, 24(4): 579-591.

[46] Qu X, Lei C, Liu W. Transcriptome mining of active biosynthetic pathways and their associated products in *Streptomyces flaveolus*[J]. Angewandte Chemie International Edition, 2011, 50(41): 9651-9654.

[47] Gubbens J, Zhu H, Girard Gv, et al. Natural product proteomining, a quantitative proteomics platform, allows rapid discovery of biosynthetic gene clusters for different classes of natural products[J]. Chemistry & Biology, 2014, 21(6): 707-718.

[48] Liu M, Grkovic T, Liu X, et al. A systems approach using OSMAC, Log P and NMR fingerprinting: An approach to novelty[J]. Synthetic & Systems Biotechnology, 2017, 2(4): 276-286.

[49] Pan H, Tian X, Shao M, et al. Genome mining and metabolic profiling illuminate the chemistry driving diverse biological activities of *Bacillus siamensis* SCSIO 05746[J]. Applied Microbiology & Biotechnology, 2019, 103(10): 4153-4165.

[50] Watrous J, Roach P, Alexandrov T, et al. Mass spectral molecular networking of living microbial colonies[J]. Proceedings of the National Academy of Sciences of the United States of America, 2012, 109(26): E1743-E1752.

[51] Wang M, Carver JJ, Phelan VV, et al. Sharing and community curation of mass spectrometry data with global natural products social molecular networking[J]. Nature Biotechnology, 2016, 34(8): 828-837.

[52] Quinn RA, Nothias LF, Vining O, et al. Molecular networking as a drug discovery, drug metabolism, and precision medicine strategy[J]. Trends in Pharmacological Sciences, 2017, 38(2): 143-154.

[53] Bai Y, Yi P, Zhang Y, et al. Structure-based molecular networking for the target discovery of novel germicidin derivatives from the sponge-associated *Streptomyces* sp. 18A01[J]. Journal of Antibiotics, 2021, 74(11): 799-806.

[54] van Der Hooft JJJ, Mohimani H, Bauermeister A, et al. Linking genomics and metabolomics to chart specialized metabolic diversity[J]. Chemical Society Reviews, 2020, 49(11): 3297-3314.

[55] Wandy J, Zhu Y, van der Hooft JJJ, et al. Ms2lda.org: web-based topic modelling for substructure discovery in mass spectrometry[J]. Bioinformatics, 2018, 34(2): 317-318.

[56] Mohimani H, Gurevich A, Shlemov A, et al. Dereplication of microbial metabolites through database search of mass spectra[J]. Nature Communications, 2018(9): 4035.

[57] Ernst M, Kang KB, Caraballo-Rodriguez AM, et al. MolNetEnhancer: enhanced molecular networks by integrating metabolome mining and annotation tools[J]. Metabolites, 2019, 9(7): 144.

[58] Maimone NM, de Oliveira LFP, Santos SN, et al. Elicitation of *Streptomyces lunalinharesii* secondary metabolism through co-cultivation with *Rhizoctonia solani*[J]. Microbiological Research, 2021(251): 126836.

[59] Watrous JD, Dorrestein PC. Imaging mass spectrometry in microbiology[J]. Nature Reviews Microbiology, 2011, 9(9): 683-694.

[60] Liu WT, Kersten RD, Yang YL, et al. Imaging mass spectrometry and genome mining via short sequence tagging identified the anti-infective agent arylomycin in *Streptomyces roseosporus*[J]. Journal of the American Chemical Society, 2011, 133(45): 18010-18013.

[61] Zhang C, Idelbayev Y, Roberts N, et al. Small molecule accurate recognition technology (SMART) to enhance natural products research[J]. Scientific Reports, 2017, 7(1): 14243.

[62] Egan JM, van Santen JA, Liu DY, et al. Development of an NMR-based platform for the direct structural annotation of complex natural products mixtures[J]. Journal of Natural Products, 2021, 84(4): 1044-1055.

[63] Wong WR, Oliver AG, Linington RG. Development of antibiotic activity profile screening for the classification and discovery of natural product antibiotics[J]. Chemistry & Biology, 2012, 19(11): 1483-1495.

[64] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA-A Cancer Journal for Clinicians, 2018, 68(6): 394-424.

[65] Feng RM, Zong YN, Cao SM, et al. Current cancer situation in China: good or bad news from the 2018 global cancer statistics? [J]. Cancer Communications, 2019(39): 12.

[66] WHO. Ten threats to global health in 2019. World Health Organization 2019; Available from: https://www.who.int/news-room/feature-stories/ten-threats-to-global-health-in-2019.

[67] Fisher MC, Henk DA, Briggs CJ, et al. Emerging fungal threats to animal, plant and ecosystem health[J]. Nature, 2012, 484(7393): 186-194.

[68] Carroll AR, Copp BR, Davis RA, et al. Marine natural products[J]. Natural Product Reports, 2022: 1-50.

[69] Newman DJ, Cragg GM. Natural products as sources of new drugs from 1981 to 2014[J]. Journal of Natural Products, 2016, 79(3): 629-661.

[70] Huffman BJ, Shenvi RA. Natural products in the "marketplace": interfacing synthesis and biology[J]. Journal of the American Chemical Society, 2019, 141(8): 3332-3346.

[71] Gerwick BC, Sparks TC. Natural products for pest control: an analysis of their role, value and future[J]. Pest Management Science, 2014, 70(8): 1169-1185.

[72] Wang L, Dash S, Ng CY. et al. A review of computational tools for design and reconstruction of metabolic pathways[J]. Synthetic & Systems Biotechnology, 2017, 2(4): 243-252.