

人体微生物天然产物研究进展

张雨薇^{1,2}, 刘煜峰^{1,3}, 陈义华^{1,2*}

(1. 中国科学院微生物研究所微生物资源前期开发国家重点实验室, 北京 100101; 2. 中国科学院大学生命科学学院, 北京 100049; 3. 中国科学院大学存济医学院, 北京 100049)

[摘要] 得益于基因测序技术、微生物组学研究和生物信息学分析手段的快速发展, 人体微生物天然产物研究近年来取得了长足的进展。随着研究的深入, 人们发现人体微生物天然产物往往会参与产生菌与宿主或与其他人体微生物之间的相互作用, 可以直接或间接地影响人体健康。相应地, 不少人体微生物天然产物展现出良好的成药性或作为疾病诊疗新型标志物的潜力。因此, 研究人体微生物天然产物不仅能够拓展活性天然产物的来源, 促进药物研发, 而且在疾病诊断和改善人体健康方面具有重要的意义。综述了人体微生物天然产物研究常用的 3 类策略, 并着重介绍了一些具有代表性的化合物的发现方法及其多样的结构与活性, 以期对相关方向的进一步深入研究提供参考。

[关键词] 人体微生物; 活性天然产物; 基因组挖掘

[中图分类号] R78; R914

[文献标志码] A

[文章编号] 1001-5094 (2023) 04-0273-14

DOI: 10.20053/j.issn1001-5094.2023.04.004

Advances in Natural Product Discovery from Human Microbiota

ZHANG Yuwei^{1,2}, LIU Yufeng^{1,3}, CHEN Yihua^{1,2}

(1. State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China; 2. College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China; 3. Savaid Medical School, University of Chinese Academy of Sciences, Beijing 100049, China)

[Abstract] Due to the rapid development of gene sequencing technology, microbiome research, and bioinformatics, the research on natural products from human microbiota has made significant progress in recent years. With more in-depth research, it has been found that the bioactive natural products produced by human microorganisms may frequently participate in the interactions between host-microbe and inter-human microbiota, with direct or indirect influence on human health. Many of these molecules exhibit the potential to become drug candidates or biomarkers for disease diagnosis/treatment. Therefore, researches on natural products from human microbiota may not only expand the source of bioactive natural products for drug discovery, but also be helpful in disease diagnosis and improving human health. This review summarizes three types of strategies for natural product discovery from human microbiota, with particular introduction to the discovery methods of a number of representative molecules as well as their structure and bioactivity diversity, so as to provide reference for further in-depth research in related fields.

[Key words] human microbiota; bioactive natural product; genome mining

上世纪 40 年代以来, 土壤、海洋等环境中的微生物所产生的活性天然产物成为药物的重要来源之一^[1]。在 2 代测序技术发展起来以前, 人们已经认识到人体中存在着大量的微生物, 并通过传统手段分离获得一些它们所产生的活性化合物, 从而拉开了

人体微生物天然产物研究的序幕。近年来, 人类微生物组计划 (Human Microbiome Project, HMP)^[2]、人类肠道宏基因组计划 (Metagenomics of the Human Intestinal Tract, MetaHIT)^[3] 等组学研究的开展, 更新了研究人员对人体内微生物数量和多样性的了解, 也拓展了其对人体微生物天然产物的结构类型和丰富程度的认识。人体微生物作为活性天然产物的潜在重要来源的概念正逐渐被更多的研究者认可^[4-8]。

与环境来源的微生物相比, 人体微生物经历了与人体环境的长期共同进化过程^[9], 适应了人体和人体微生物群落构成的复杂而独特的内部生态系统。

接受日期: 2023-02-12

项目资助: 国家自然科学基金委杰出青年基金 (No. 32025002), 国家重点研发计划 (No. 2018YFA0901901)

* **通信作者:** 陈义华, 研究员;

研究方向: 微生物活性代谢产物, 天然产物生物合成;

Tel: 010-64806121; **E-mail:** chenjihua@im.ac.cn

这决定了人体微生物产生的小分子化合物天然靶向人体或人体微生物组来行使信号传递、免疫调节、种间竞争等功能的活性特征^[5]。与人体健康之间的天然联系暗示着人体微生物天然产物具有良好的成药性。另外, 有些病原菌所产生的天然产物与其致病性密切相关, 对这类天然产物的研究将有助于发展预防和治疗相关疾病的新型诊疗手段^[4, 10]。基于相关认知, 近年来人体微生物天然产物研究备受重视, 得到了快速发展。本文将发现人体微生物活性天然产物的策略为脉络, 介绍近年来报道的代表性天然产物的发现方法、结构特点及其活性特征, 以期为进一步优化人体微生物活性天然产物的挖掘方法、促进相关领域的研究提供思路。

1 基于传统策略发现人体微生物天然产物

如前所述, 在2代基因测序技术发展起来以前, 研究者们利用传统的天然产物发现策略已获得了一些人体微生物产生的活性化合物。这一策略的实现依赖于对菌株的分离和培养, 之后在活性指导下借助高效液相色谱法 (high performance liquid chromatography, HPLC)、质谱 (mass spectrum, MS) 等手段对菌株代谢物进行分析, 并利用液相色谱-质谱法 (liquid chromatography/mass spectrometry, LC/MS)、串联质谱法 (tandem mass spectrometry, MS/MS)、核磁共振 (nuclear magnetic resonance, NMR) 等经典的化学手段对目标化合物进行结构表征 (见图1)。利用传统的天然产物发现策略在人体微生物中获得的活性天然产物大多为核糖体合成和翻译后修饰多肽类 (ribosomally synthesized and post-translationally modified peptides, RiPPs) 化合物。它们通常抑制亲缘关系较近的物种, 具有较窄的抗菌谱, 在协助产生菌占据和维持生态位方面发挥作用。在以往的很多研究中对这类化合物都进行了详细的介绍^[5, 7], 图1简单列举了几种代表性化合物。从表皮葡萄球菌 (*Staphylococcus epidermidis*) 中分离的表皮抗菌肽 (epidermin) 能够有效抑制与表皮葡萄球菌竞争皮肤生态位的痤疮丙酸杆菌 (*Propionibacterium acnes*)^[11]; 由唾液链球菌 (*Streptococcus salivarius*) 产生的 salivaricin B 对唾

液链球菌、化脓链球菌 (*Streptococcus pyogenes*)、轻型链球菌 (*Streptococcus mitis*)、藤黄微球菌 (*Micrococcus luteus*) 等常见人体微生物具有抑制活性^[12]; 大肠埃希菌 (*Escherichia coli*) 能够产生多种相对分子质量较小的小菌素 (microcins), 其中 MccC7 具有腺苷酸化修饰的 C 端, 对肺炎克雷伯菌 (*Klebsiella pneumoniae*)、沙门菌属 (*Salmonella*)、耶尔森菌属 (*Yersinia*) 等微生物都具有抑制活性^[13]。

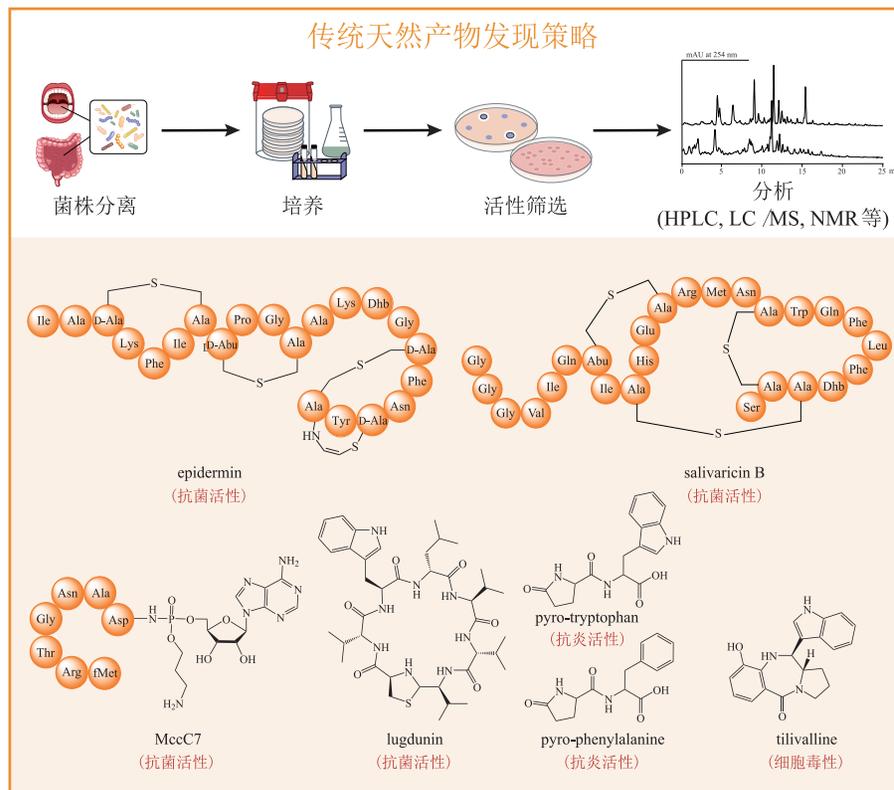
除此之外, 利用传统的天然产物发现策略也从人体微生物中发现了少数其他类型的化合物。与其他生境中的生物一样, 人体微生物之间也存在着激烈的营养和生态位竞争, 它们在长期的共存中发展出了能够相互抑制的“分子武器”。在人体的皮肤或上呼吸道这类营养物质贫乏的环境中, 人体微生物之间的竞争可能尤为激烈^[14-15]。基于这一生态学假设, Zipperer 等^[16] 从一系列鼻腔共生菌中筛选到了可以抑制同一生态位的条件致病菌——金黄色葡萄球菌 (*Staphylococcus aureus*) 生长的路邓葡萄球菌 (*Staphylococcus lugdunensis*) IVK28。随后, 通过非靶向转座子诱变, 作者以抑制金黄色葡萄球菌活性为指示, 对突变株与野生型菌株的代谢谱进行比对分析后, 分离得到了具有良好抗菌活性的非核糖体多肽 (non-ribosomal peptides, NRPs) 类化合物 lugdunin (见图1)。此外, lugdunin 还显示出对耐甲氧西林金黄色葡萄球菌 (methicillin-resistant *S. aureus*, MRSA)、耐万古霉素肠球菌 (vancomycin-resistant *Enterococcus*, VRE) 和单核细胞增生李斯特菌 (*Listeria monocytogenes*) 的良好抑制活性, 并在小鼠金黄色葡萄球菌皮肤感染模型中起到了抑制感染的作用。后续研究中, Bitschar 等^[17] 还发现 lugdunin 能够与人体产生的抗菌肽 DCD-1L 以及 LL-37 协同作用, 发挥对 MRSA 的杀菌作用, 并能够增强人原代角质形成细胞的先天免疫反应。这些结果显示 lugdunin 潜在的应用价值, 也展现了生态学观念的引入在人体微生物活性天然产物挖掘过程中的重要性。

类似地, 研究者们利用活性导向的天然产物发现方法对一些人体益生菌的产物进行了表征。

多项研究表明, 益生菌植物乳杆菌 (*Lactobacillus plantarum*) 能够减少体内炎症^[18], 其发酵液在体外试验中也显示出对核因子 κ B (nuclear factor- κ B, NF- κ B) 信号通路相应的调节功能^[19]。通过主成分分析 (principal component analysis, PCA) 探究植物乳杆菌不同生长阶段代谢产物的差异, Zvanych 等^[20] 从中鉴定出了 4 个具有独特焦谷氨酸环结构的 pyro-dipeptides 类化合物。其中, pyro-phenylalanine 和 pyro-tryptophan (见图 1) 在小鼠腹腔注射实验中能够显著减少脾脏产生的 γ 干扰素 (interferon-gamma, IFN- γ), 显示出这类化合物潜在的抗炎能力。

除了挖掘对人类健康有益的人体微生物天然产

物, 研究者们还利用传统的天然产物发现策略研究了人体病原菌的毒力因子, 其中最典型的例子就是 tilivalline (见图 1) 的发现。2014 年, Schneditz 等^[21] 以对人上皮细胞的细胞毒作用为指示, 利用转座子诱变, 从引发抗生素相关出血性结肠炎 (antibiotic-associated hemorrhagic colitis, AAHC) 的病原菌产酸克雷伯菌 (*Klebsiella oxytoca*) 中鉴定出该菌产生的细胞毒分子 tilivalline, 通过体外试验和动物模型验证了 tilivalline 对肠道黏膜和屏障功能的破坏, 揭示了产酸克雷伯菌引发结肠炎的可能机制之一, 并提供了潜在的诊断标志物和治疗策略。可见, 研究人体病原菌产生的天然产物也能够促进病原菌相关疾病预防和治疗手段的发展。



HPLC: high performance liquid chromatography (高效液相色谱法); LC/MS: liquid chromatography/mass spectrometry (液相色谱-质谱法); NMR: nuclear magnetic resonance (核磁共振)

图 1 基于传统策略获得人体微生物活性天然产物

Figure 1 Obtaining bioactive natural products from human microbiota based on traditional strategies

利用传统策略挖掘人体微生物产生的活性天然产物具有不依赖于序列预测、能够以目标活性或结构特征为导向等优势。然而, 该方法在应用过程中也存在不少限制因素, 例如: 对可培养菌株的依赖

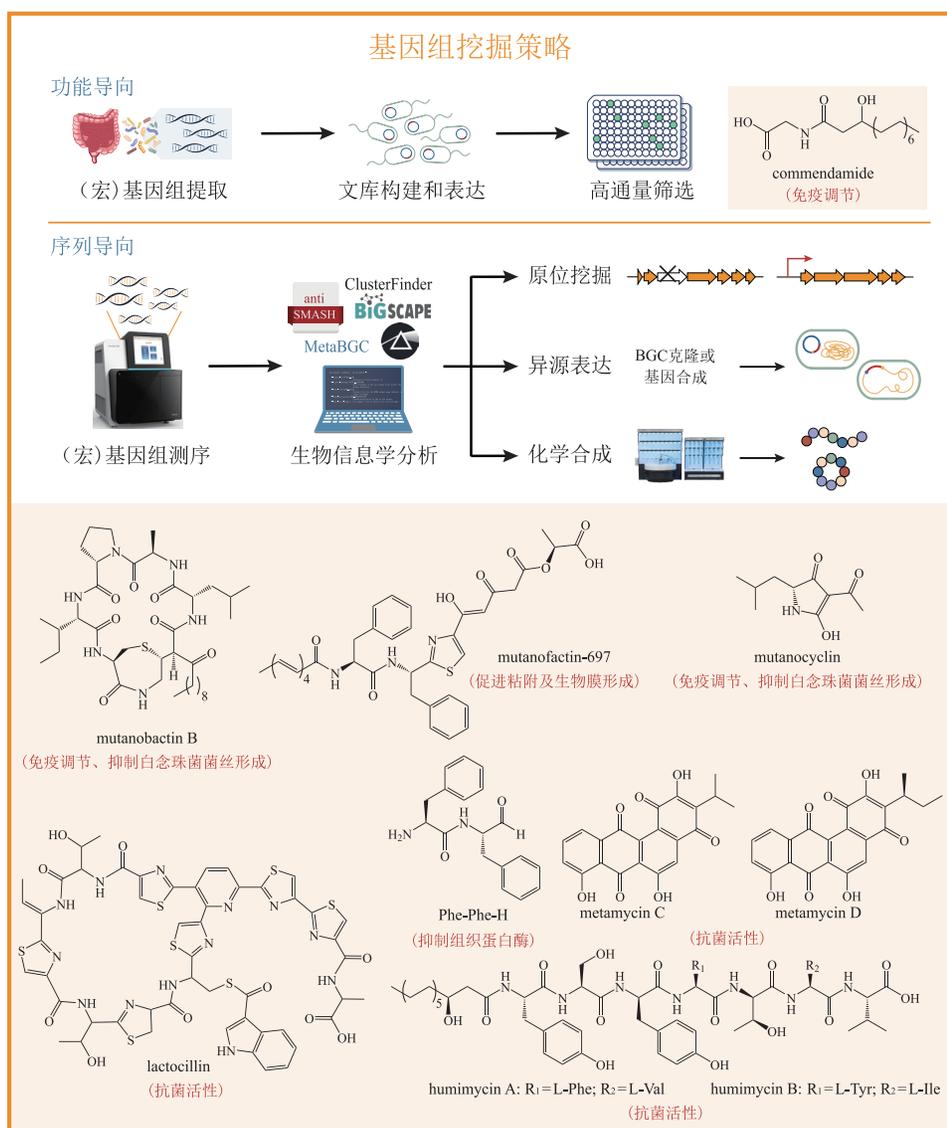
性, 以及难以避免对已知化合物的重复发现。令人鼓舞的是, 培养组学 (culturomics) 的应用和发展使得越来越多的人体微生物能够被培养^[22-23], 该方法首次在多种人体粪便样品中应用就将可培养的人

体肠道细菌数量从原先的 688 种增加到了 1 057 种^[24]; 此外, 基于 LC/MS, MS/MS 和 NMR 数据的去重复方法的进步也赋予了传统天然产物发现方法新的活力^[25], 这一策略依然能够在未来人体微生物活性天然产物发现中提供有力的帮助。

2 基于基因组挖掘方法获得人体微生物天然产物

自 HMP 等计划开展以来, 对于人体微生物领域的持续研究积累了大量的组学数据, 这让研究人

员越来越清晰地认识到人体中蕴藏有丰富的微生物和微生物代谢产物资源^[26-29]。另外, 基因组文库高通量筛选方法的优化、(宏)基因组数据中生物合成基因簇(biosynthetic gene clusters, BGCs)识别算法和分析工具的不断更新, 以及针对不同种类微生物遗传操作方法和异源表达系统的建立, 有力地推动了通过基因组挖掘来获取人体微生物天然产物的方法进步。相关研究根据出发点不同可以分为功能导向和序列导向 2 种策略(见图 2)。



antiSMASH: antibiotics and secondary metabolite analysis shell (抗生素和次级代谢产物分析工具); BiGSCAPE: biosynthetic gene similarity clustering and prospecting engine (生物合成基因簇聚类分析工具); MetaBGC: metagenomic identifier of biosynthetic gene clusters (从宏基因组数据中识别生物合成基因簇的工具); BGC: biosynthetic gene cluster (生物合成基因簇)

图 2 基于基因组挖掘方法获得人体微生物活性天然产物

Figure 2 Obtaining bioactive natural products from human microbiota by genome mining

2.1 功能导向的人体微生物基因组挖掘

功能导向的人体微生物基因组挖掘以功能宏基因组学相关的工作为代表, 该方法首先提取特定环境微生物组的脱氧核糖核酸 (deoxyribonucleic acid, DNA) 来构建宏基因组文库, 随后针对该文库产生的特定活性的化合物进行高通量筛选, 获得具有目标活性的化合物后, 再从对应的克隆中鉴定出负责产生活性天然产物的基因序列。Cohen 等^[30] 利用人体粪便样本提取的 DNA 构建了约含有 75 000 个克隆的宏基因组文库, 随后利用细胞报告系统来筛选具有调节 NF- κ B 活性的克隆, 再通过转座子诱变确认产生相关化合物的基因。这一过程鉴定出 26 个独特的人体共生细菌效应基因 (commensal bacteria effector genes)。其中的 1 个效应基因家族 *Cbeg12* 被注释为 *N*-酰基转移酶, 能够催化产生新颖的 *N*-酰基酰胺类天然产物 commendamide (见图 2)。在随后针对 242 种 G 蛋白偶联受体 (G protein-coupled receptors, GPCRs) 的活性筛选中, 发现 commendamide 可以特异性地激活调节细胞增殖和免疫的关键受体 GPCR132/G2A。此外, 与 *Cbeg12* 高度同源的基因存在于一些人体共生拟杆菌中, 在普通拟杆菌 (*Bacteroides vulgatus*) 的发酵液中也能直接检测到 commendamide, 提示人体共生拟杆菌可能会产生 commendamide, 并通过 GPCR132/

G2A 参与对宿主免疫的影响。Commendamide 的结构与一些人体内源性代谢物类似, 它的发现和功能鉴定显示出人体微生物具有通过模拟宿主内源性信号分子来实现与宿主相互作用的能力。在此基础上, Cohen 等^[31] 在后续工作中系统地分析和研究了 HMP 数据中的 *N*-酰基转移酶, 并发现了人体微生物产生的更多类型的 *N*-酰基酰胺以及它们对其他 GPCRs 的影响。近年来, Piscotta 等^[32] 利用功能宏基因组学对 14 个健康婴儿的粪便样本进行研究, 不仅在针对 NF- κ B 调节活性的筛选中再次验证了 commendamide 的产生及其功能, 还鉴定出一些具有诱导细胞自噬或调节氧化还原电位功能的效应基因。可见, 功能宏基因组学在建立人体微生物组序列和功能的关联方面具有突出优势。

2.2 序列导向的人体微生物基因组挖掘

(宏) 基因组研究伴随着 DNA 测序技术的快速进步, 为人体微生物天然产物挖掘提供了大量可用的序列信息, 同时也对序列和数据的分析提出了更高的要求。负责天然产物产生的基因通常共定位于基因组邻近区域形成 BGC, 而且一般来说具有某些特征序列。基于这一特性, 研究者们开发了多种预测和评估 BGC 的方法。表 1 展示了目前常用的部分 BGC 预测工具。

表 1 常用的生物合成基因簇预测工具

Table 1 Commonly used biosynthetic gene cluster prediction tools

生物信息学工具	预测 BGC 类型	所使用的算法	参考文献
antiSMASH	所有类型	BLAST, HMM, machine learning	[33]
ClustScan	NRPS, PKS	HMM	[34]
NP.searcher	NRPS, PKS	BLAST	[35]
NRPSpredictor2	NRPS	SVM	[36]
NaPDos	NRPS, PKS	HMM	[37]
NaPDos2	NRPS, PKS	BLAST	[38]
ClusterFinder	所有类型	HMM	[39]
RiPPMiner	RiPPs 合成酶	SVM	[40]
RODEO	RiPPs 合成酶	HMM, heuristic scoring, SVM	[41]
BAGEL4	RiPPs 合成酶	HMM	[42]
MetaMiner	RiPPs 合成酶	HMM, heuristic scoring, dereplicator	[43]
EvoMining	所有类型	系统发育信息	[44]
MetaBGC	所有类型	spHMMs	[45]
NeuRiPP	RiPPs 合成酶	NN	[46]
DeepBGC	所有类型	NN, RF	[47]
e-DeepBGC	所有类型	NLP	[48]

续表 1

生物信息学工具	预测 BGC 类型	所使用的算法	参考文献
DeepRiPP	RiPPs 合成酶	NLP	[25]
PRISM 4	NRPS, PKS, RiPPs 合成酶	HMM	[49]
gutSMASH	初级代谢 BGC	pHMMs	[50]
BiG-MAP	所有类型	基于 antiSMASH, gutSMASH, Pfam 等	[51]

BGC: biosynthetic gene cluster (生物合成基因簇); NRPS: nonribosomal peptide synthetase (非核糖体多肽合成酶); PKS: polyketide synthase (聚酮合酶); RiPPs: ribosomally synthesized and post-translationally modified peptides (核糖体合成和翻译后修饰多肽类); BLAST: basic local alignment search tool (基于局部序列比对算法的搜索工具); HMM: Hidden Markov Model (隐马尔可夫模型); SVM: support vector machine (支持向量机); spHMMs: segmented profile Hidden Markov Models (分段轮廓隐马尔可夫模型); NN: neural network (神经网络); RF: random forest (随机森林); NLP: natural language processing (自然语言处理); pHMMs: profile Hidden Markov Models (轮廓隐马尔可夫模型); antiSMASH: antibiotics and secondary metabolite analysis shell (抗生素和次级代谢产物分析工具); gutSMASH: 用于识别肠道微生物中专门的初级代谢途径; Pfam: protein families database (蛋白质家族数据库)

序列导向的人体微生物基因组挖掘研究基于对(宏)基因组序列的深入分析,通过针对性地选取目标 BGC 来进行研究。已开展的工作中研究的 BGCs 有的具有高度的新颖性或突出的代表性;有的来源于高丰度物种或与人体健康、疾病有密切关联的物种;有的则能够被其他组学数据映射,具有潜在的生理学意义。对已经开展的研究进行归纳,根据所采取研究策略的不同可以分为以下 3 类。

2.2.1 在原始菌株中进行原位挖掘 人体肠道和口腔是微生物丰度最高的 2 个部位。与肠道微生物天然产物一样,口腔微生物所蕴含的活性天然产物资源也引发了研究人员广泛的兴趣。在已经开展的口腔微生物天然产物研究中,针对变形链球菌(*Streptococcus mutans*)的研究具有一定的代表性。变形链球菌在人类口腔中普遍存在且长期被认为与龋齿相关^[52]。Ajdici 等^[53]在详细分析 *S. mutans* UA159 的基因组后,从中鉴定出了 TnSmu2 基因组岛。随后对这一区域进行比较分析,发现不同变形链球菌中该基因组岛可能包含了不同类型的 BGCs^[54]。通过敲除 *S. mutans* UA159 该区域中编码聚酮合酶-非核糖体肽合成酶(polyketide synthase-nonribosomal peptide synthetase, PKS-NRPS)杂合酶的 *mub* 基因,研究者们鉴定出了具有特殊的 C—C 成键大环和含氮、硫七元杂环结构的 mutanobactin A^[55],并在随后的几项研究中发现了一系列类似物 mutanobactin B—J^[56-57]。笔者课题组通过研究发现 mutanobactins 类化合物在生物合成过程中通过还原酶结构域释放带有醛基末端的线性脂肽,再历经 3 步非酶催化反应产生 mutanobactins,而这一过程中自发的 C—S 键形成和 C—C 键断裂极

大丰富了 mutanobactins 的结构多样性^[58]。此外,在 mutanobactins 的研究过程中,还利用基于 MS/MS 的算法 iSNAP (informatic search algorithm for natural products) 鉴定了该基因簇的中间产物 mutanamide^[57]。与口腔中变形链球菌需要应对氧胁迫以及该菌与白念珠菌(*Candida albicans*)密切相关等事实相符,*mub* 的存在有利于变形链球菌的抗氧化能力,部分 mutanobactins 类化合物也表现出对白念珠菌菌丝形成和生物膜形成的有效抑制。此外,在利用脂多糖刺激的 RAW264.7 巨噬细胞中,mutanobactin B (见图 2) 能够显著上调促炎细胞因子白细胞介素-6 (interleukin-6, IL-6) 和 IL-12 的表达,并下调单核细胞趋化蛋白-1 (monocyte chemotactic protein-1, MCP-1)、粒细胞集落刺激因子 (granulocyte colony-stimulating factor, G-CSF) 和肿瘤坏死因子- α (tumor necrosis factor- α , TNF- α) 的表达。

最近, Li 等^[59]通过对 *S. mutans* NMT4863 中编码 PKS-NRPS 杂合酶的 *muf* 进行敲除,发现了一系列 mutanofactins 类化合物(见图 2)。Mutanofactins 能够非特异性结合变形链球菌细胞并改变细菌表面的物理化学性质(增加细菌疏水性),还能直接与胞外 DNA 结合并促进胞外 DNA 介导的细胞聚集作用,从而以剂量依赖的方式促进细菌细胞的自聚集、细菌间的粘附,以及随后的生物膜形成。这一特征可能有助于 mutanofactins 产生菌在牙菌斑生物膜形成过程中与周围的变形链球菌或其他微生物结合,促进其在口腔中的定植,并影响生物膜的组成和结构。

此外,人体其他部位的微生物同样具有产生活性天然产物的巨大潜能。2014 年, Cimermanic

等^[39]开发了基于隐马尔可夫模型 (Hidden Markov Model, HMM) 的概率算法 ClusterFinder, 扩大了所能预测的 BGC 类型和范围。Donia 等^[60]利用 ClusterFinder, 从 2 430 个 HMP 参考基因组中预测出 14 000 余个 BGCs, 其中, 3 118 个出现在健康人体微生物组中, 且超过一半 BGCs 分布于肠道和口腔中; 基于编码硫肽类核糖体肽 BGC 在 HMP 数据中体现出的广泛性, 以及已报道硫肽类化合物的良好活性, 作者选取了其中 1 个与抗生素高硫青霉素 (thiocillin) 的 BGC 相似的 *bgc66* 进行了研究。通过比较 1 株分离自阴道的含有 *bgc66* 的格氏乳杆菌 (*Lactobacillus gasseri*) JV-V03 的野生型与基因簇插入突变株的代谢谱, 获得了新颖的天然产物 lactocillin (见图 2)。活性测试发现 lactocillin 可以抑制革兰阳性菌生长, 对病原菌金黄色葡萄球菌和粪肠球菌 (*Enterococcus faecalis*) 都表现出了纳摩尔水平的抗菌活性。

若含有目标 BGC 的菌株易于培养且能够进行遗传操作, 可以针对 BGC 中的关键基因, 构建突变株, 并通过代谢谱比对, 直接建立 BGC 与化合物的关联。这种原位挖掘的方式会较大幅度地反映出天然产物产生的真实状态。然而, 在实际研究中, 很多人体微生物难以分离和培养, 有些虽然可以培养但无法遗传操作, 往往需要通过原位挖掘以外的策略来获取它们产生的天然产物。

2.2.2 通过异源表达进行挖掘 得益于培养组学等方法在人体微生物研究中的应用, 实验室条件下可培养的人体微生物数量持续增加。然而, 一些低丰度物种的分离依然很困难, 针对不同微生物建立稳定的培养和遗传操作方法也需要投入大量时间和精力。即便是在可培养, 并已成功实现遗传操作的种属内, 不同菌株之间的可操作性差异也阻碍了相关方法的广泛应用。因此, 利用异源表达宿主来进行人体微生物天然产物研究得到了广泛的重视。

Guo 等^[61]结合以往研究以及 HMP 公布的最新数据鉴定了 47 个未知功能的 NRPS 类 BGCs, 这些 BGCs 存在于超过 90% 的 HMP 粪便样品中, 且几乎完全来源于肠道微生物中的厚壁菌门 (Firmicutes) 梭菌纲 (Clostridia)。由于梭菌的遗传操作较为困

难, 作者利用 2 种常用的异源表达宿主大肠埃希菌和枯草芽孢杆菌 (*Bacillus subtilis*) 选取了 14 个序列和结构域排布具有代表性的 BGCs 进行异源表达。通过直接克隆或经密码子优化后合成相关 BGCs, 将其转入大肠埃希菌或枯草芽孢杆菌中, 分别置于强启动子 T7 或 hyper-*Pspac* 下进行表达, 成功获得了 32 个 pyrazinones 和 dihydropyrazinones 类化合物。作者根据这些化合物的结构推测其环化由非酶促机制产生, 而未发生环化的二肽醛 (dipeptide aldehydes) 形式前体可能是细菌的真正产物。结合肽醛类化合物在以往研究中表现出的蛋白酶抑制活性, 作者通过体外研究发现其中一些二肽醛确实具有类似的活性, 如: Phe-Phe-H 对于组织蛋白酶表现出显著的选择性 (见图 2)。考虑到组织蛋白酶在免疫监视中的作用, 以及编码产生二肽醛的 BGCs 在肠道中广泛存在, 这一研究结果提示人体微生物可能通过产生具有组织蛋白酶抑制活性的天然产物来影响宿主免疫, 从而促进自身定植。

大肠埃希菌和枯草芽孢杆菌等常用异源表达宿主在表达部分厌氧微生物的 BGCs 时表现出一定的局限性。基于此, 笔者课题组以兼性厌氧菌 *S. mutans* UA159 作为异源表达宿主, 利用其天然感受态建立了适用于低 G+C 含量的厌氧菌的大片段克隆技术——基于自然感受态的大片段克隆技术 (natural competence based large DNA fragment cloning, NabLC)^[62]。NabLC 克隆过程中不需要引入中间载体, 能够避免由于载体不稳定或不兼容引起的克隆失败问题。利用该技术, 课题组对不同来源的低 G+C 含量的厌氧菌 BGCs 成功实现了克隆和表达, 包括来源于表皮葡萄球菌的 pyrazinones 类化合物的 BGC 和 *S. mutans* 35 中编码 PKS-NRPS 杂合酶的 BGC, 后者能够产生 tetramic acids 类化合物 mutanocyclin (见图 2)。活性测试发现 mutanocyclin 具有抑制免疫细胞浸润的活性, 并表现出抑制白念珠菌菌丝形成的能力^[63]。随后, Tang 等^[64]通过对 *S. mutans* B04Sm5 直接进行遗传操作也检测到了 mutanocyclin 的产生, 还发现了 *muc* 基因簇可以产生 reutericyclins 类化合物。其中, reutericyclin A 也曾经从罗伊乳杆菌 (*Limosilactobacillus reuteri*) 中

分离得到^[65]。Reutericyclins 具有广谱抗革兰阳性菌活性, 不仅对乳杆菌、枯草芽孢杆菌、金黄色葡萄球菌等菌种具有抗菌作用, 还能够抑制血链球菌 (*Streptococcus sanguinis*)、戈登链球菌 (*Streptococcus gordonii*)、轻型链球菌等与变形链球菌竞争同一生态位的口腔微生物。据推测, reutericyclins 是 mutanocyclin 生物合成过程中的中间代谢产物, 能够被 *mucF* 编码的脱酰基酶转化为 mutanocyclin, 从而在不同情况下为变形链球菌提供更多生存优势。

已开发的大部分 BGC 预测工具一般适用于基因组信息较为完整的物种, 而在分析一些从临床人体微生物样本获得的微量、碎片化测序结果时可能会遗失很多新颖 BGC 的信息; 此外, 最近的几项宏基因组研究揭示了人体微生物组中还有很多尚未被测序的罕见微生物, 它们的 BGCs 依然有待挖掘。为了突破这一限制, Sugimoto 等^[45]开发了 MetaBGC 算法, 该算法包含 Build, Identify, Quantify 和 Cluster 4 个模块。其中, Build 模块基于轮廓隐马尔可夫模型 (profile Hidden Markov Models, pHMMs) 开发了分段 pHMMs (segmented pHMMs, spHMMs) 算法, 能够直接从读段 (read) 水平 (约 100 bp) 的人体微生物宏基因组数据中识别 BGCs, 在一定程度上避免了对低丰度序列的遗漏。此外, MetaBGC 还具有较高的运算效率, 便于同时分析大量宏基因组样本。利用该方法, 研究者针对 HMP 以及 MetaHIT 的 2 544 个宏基因组样本进行了分析, 从中鉴定出 13 个完整的 II 型 PKS BGCs, 并通过将公开的转录组数据映射到已鉴定的 BGCs 中, 证明其中至少 6 个 BGCs 能够在人体内表达。随后, 作者利用异源表达的方法对 1 个口腔来源的 BGC (*bgc3*) 和 1 个肠道来源的 BGC (*bgc6*) 进行了研究。他们依据链霉菌属的密码子偏好性优化后合成了 *bgc3*, 并整合至白色链霉菌 (*Streptomyces albus*) J1074 中进行表达, 最终鉴定得到了 metamycins A-D; 另外, 还从 *Blautia wexlerae* DSM 19850 中克隆了 *bgc6*, 将其置于大肠埃希菌-枯草芽孢杆菌穿梭载体中的强启动子下游, 整合至基因组平均 G+C 含量与 *Blautia wexlerae* (41.5%) 接近的枯草芽孢杆菌 (*Bacillus subtilis*)

168 *sfp*⁺ (43.5%) 中进行表达, 成功获得了萹环类聚酮化合物 wexrubicin。活性测试发现 wexrubicin, metamycin A 和 metamycin B 对测试细胞和菌株无明显毒性, metamycin C 和 metamycin D (见图 2) 对一些革兰阳性菌, 尤其是链球菌属、奇异菌属 (*Atopobium*)、放线菌属 (*Actinomyces*)、罗斯菌属 (*Rothia*) 以及棒状杆菌属 (*Corynebacterium*) 的口腔分离株具有良好的抑制活性, 提示 *bgc3* 有利于宿主菌竞争口腔生态位, 这一结果也和转录组分析显示的 *bgc3* 在人龈上菌斑样本中早期生物膜形成期间表达的现象一致。

目前, 利用已有的序列分析程序预测 BGCs, 再根据目标 BGCs 特性选择合适的异源宿主进行表达, 已经成功实现了对多种不同来源、不同类型的 BGCs 的表征。随着更多 BGC 克隆方法的建立和完善、DNA 合成成本的下降以及可作为异源表达宿主的菌株列表的扩充, 异源表达策略有望在人体微生物活性天然产物的挖掘中发挥更重要的作用。

2.2.3 化学合成手段助力挖掘 不管是通过原位激活, 还是通过异源表达来获取目标天然产物, 都需要较长的研究周期, 可否获得目标化合物具有不确定性。随着基于序列信息预测天然产物结构的准确性越来越高, 在结构预测基础上直接通过有机合成来制备目标化合物, 减少了微生物培养、BGC 表达、产物分离纯化等过程中可能存在的限制和不确定性。这一方法在挖掘不含修饰或具有简单修饰的 RiPPs 类以及 NRPs 类化合物时得到了成功应用。

例如, Chu 等^[66]发展了将 BGC 预测与化学合成相结合的方法来研究人体微生物中的天然产物, 利用该方法获得的分子称为合成-生物信息学天然产物 (synthetic-bioinformatic natural products, syn-BNPs)^[47]。作者利用抗生素和次级代谢产物分析工具 (antibiotics and secondary metabolite analysis shell, antiSMASH) 对 HMP 和人类口腔微生物组数据库 (human oral microbiome database, HOMD) 中的基因组进行预测, 并选取其中的 NRPS 类 BGCs 进行后续研究。考虑到产物少于 5 个氨基酸的非核糖肽通常被高度修饰, 经过长度筛选后鉴定出 57 个预测产物多于或等于 5 个氨基酸的 NRPS 类 BGCs,

排除其中不完整、包含 PKS 模块或杂环化结构域的 BGCs 后, 利用 NRPSPredictor2 等 3 种 NRPS 预测算法对 25 个 BGCs 的产物进行了预测。根据预测结果, 作者设计了 30 个潜在的 syn-BNPs。对于可能发生 *N*-酰化的 syn-BNPs, 在合成时利用非核糖体肽中常见的 β -羟基肉豆蔻酸进行修饰, 经过 2 轮固相多肽合成后成功获得了其中 25 个 syn-BNPs。随后, 作者利用 1 组常见人体共生菌和致病菌对 syn-BNPs 进行了抗菌活性筛选, 发现来源于马红球菌 (*Rhodococcus equi*) 的 humimycin A 和红串红球菌 (*Rhodococcus erythropolis*) 的 humimycin B (见图 2) 对于厚壁菌门、放线菌门 (Actinomycetes) 的测试菌株具有抗菌活性, 尤其对金黄色葡萄球菌 (包括一些耐甲氧西林金黄色葡萄球菌的临床分离株) 和肺炎链球菌 (*Streptococcus pneumoniae*) 表现出突出的抗菌活性。此外, 作者还尝试通过 LC-MS 检测探究了红球菌属菌株的发酵液中是否能够产生 humimycins, 但未观察到此类化合物的产生, 暗示着在实验室条件下该 BGC 可能不表达, 也印证了 syn-BNPs 方法具有的潜在优势。

3 人工智能辅助人体微生物活性天然产物的发现

近年来, 人体微生物相关的组学数据增长迅速, 对研究者的挖掘效率和准确性提出了更为严格的要求。引入人工智能算法极大地提升了研究人员对海量数据进行分析的能力, 不仅在 BGCs 预测方面突破了以往生物信息学分析工具在预测新颖 BGCs, 尤其是 RiPPs 合成酶类 BGCs 时的局限, 在化合物的活性和作用靶点预测、减少重复发现等方面也表现出了巨大的应用潜力^[67]。

BGCs 预测方面, 表 1 中所列的 antiSMASH 和 RODEO 等工具包及时更新, 在以往主要基于 BLAST 和 HMM 进行 BGCs 预测的基础上, 引入了二元分类器支持向量机 (support vector machine, SVM) 等人工智能算法。基于 HMM 的 ClusterFinder 在以往工作中已经体现出识别新型 BGCs 的强大功能, 但该方法在准确性上存在一定不足。基于自然语言处理 (natural language processing, NLP) 和深度学习 (deep learning) 策

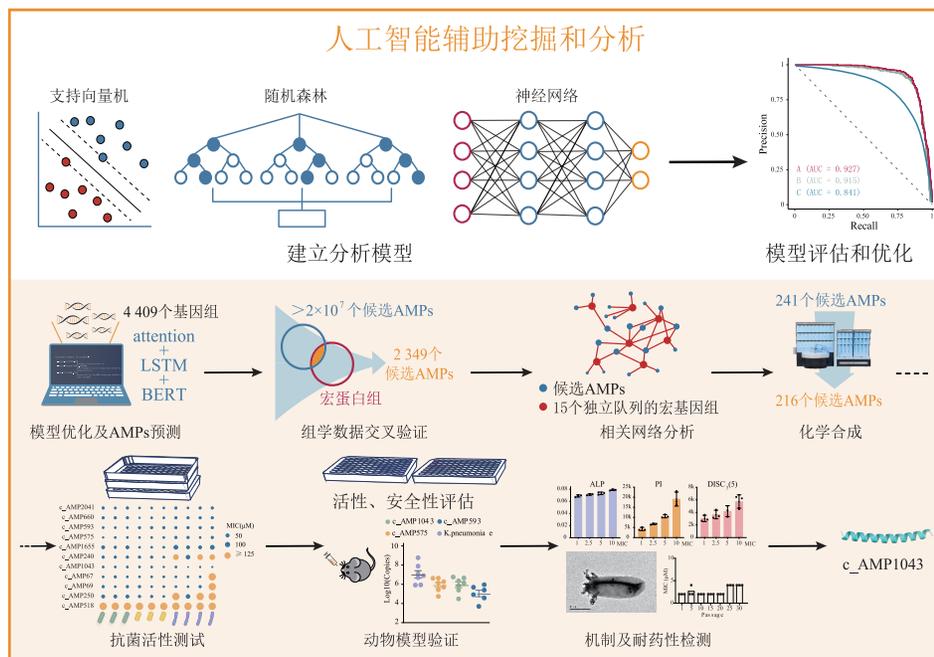
略的 DeepBGC^[47] 在一定程度上弥补了 ClusterFinder 在准确性方面的缺陷, 在针对同一验证集的分析结果中, 该方法相较于 ClusterFinder 显示出更高的识别精度和准确率。最近, Liu 等^[48] 又发布了在此基础上改进的 e-DeepBGC (extension of DeepBGC), 在原有深度学习的基础上引入了新的数据增强 (data augmentation) 步骤, 以识别更多类型的 BGCs。RiPPs 合成酶类 BGCs 序列的高度多样性, 以及相关 BGCs 基因的分散分布性, 很长时间以来都限制了该类 BGCs 的准确预测和新型基因簇的发现。近几年来, 一些针对 RiPPs 合成酶类 BGCs 特点开发的工具, 如: NeuRiPP^[46] 和 DeepRiPP^[25], 通过在预测过程中使用神经网络和自然语言处理在很大程度上突破了传统预测方法的局限。

最近, 笔者课题组与合作者基于 attention、长短期记忆 (long short-term memory, LSTM) 以及基于 transformer 的双向编码器表征 (bidirectional encoder representations from transformers, BERT) 3 种自然语言处理神经网络模型构建了用于抗菌肽 (antimicrobial peptides, AMPs) 挖掘的方法^[68] (见图 3), 预测精确度相比以往的预测工具大幅提高, 达到 91.31%。通过对 4 409 个人体微生物代表性基因组进行挖掘, 笔者课题组与合作者鉴定了 2×10^7 余个候选 AMPs。随后, 经过与宏蛋白组学数据的交叉验证将候选 AMPs 数量精简至 2 349 个。考虑到具有抗菌活性的 AMPs 和其能够抑制的细菌之间存在潜在的负相关性, 进一步使用 15 个独立队列的宏基因组数据构建了 AMP-微生物相关网络, 将候选 AMPs 的列表优化到 241 个。最后, 通过多肽合成的方法成功获得了 216 个 AMPs, 并进行了抗菌活性测试, 阳性率达到 83.8% (181/216)。这些 AMPs 与以往报道的 AMPs 序列一致性均低于 40%, 说明基于自然语言学习的方法可以有效发现新型 AMPs。对活性突出的 11 种 AMPs 进一步测试后显示 10 种 AMPs 对多株临床分离的多重耐药细菌表现出良好抗菌活性。笔者课题组与合作者选择了 3 种在溶血性测试和细胞毒性测试中表现良好的 AMPs 用于治疗感染肺炎克雷伯菌的小鼠模型, 观察到了显著的疗效。进一步对综合活性最好的抗菌

肽 c_AMP1043 的细菌耐药性进行了测试, 结果在 30 天针对大肠埃希菌的抗性诱导实验后, 未观察到明显耐药性的产生。

这一工作更新了从序列信息中挖掘活性天然产

物发现的逻辑, 证明在 AMPs 等类型的化合物发现过程中, 人工智能方法的引入可以避开 BGCs 预测等步骤, 直接发现具有特定生物活性的分子, 大大简化了活性化合物的获取过程。



LSTM: long short-term memory (长短期记忆); BERT: bidirectional encoder representations from transformers (基于 transformer 的双向编码器表征); AMPs: antimicrobial peptides (抗菌肽)

图 3 人工智能辅助的人体微生物活性天然产物挖掘和分析

Figure 3 AI-assisted mining and analysis of bioactive natural products from human microbiota

4 结语与展望

人体内蕴藏着数量庞大的微生物群体, 它们之间的相互作用及其对人体产生的影响与人类的健康息息相关。人体微生物所产生的活性天然产物是人体环境长期筛选的结果, 对相关微生物适应人体环境发挥着重要的作用。相应地, 这就决定了活性天然产物靶向人体或人体微生物组来行使生理功能的活性特征。目前, 人们已经发展了多种获取人体微生物天然产物的研究策略, 成功地获得了一些化合物。对这些化合物的活性研究发现其往往具有免疫调节和抑制其他人体微生物的作用, 但由于人体环境的复杂性和缺乏研究模型的限制, 对其真实生理功能的理解进展缓慢。

除了产生次级代谢产物之外, 人体微生物通过代谢食物或人体分泌物产生的活性分子也能够从多方面影响人体健康。例如: 1) 肠道微生物以膳

食纤维或宿主分泌的黏蛋白为底物进行发酵产生的短链脂肪酸 (short-chain fatty acids, SCFAs) 能够激活 GPCRs 或游离脂肪酸受体 (free fatty acid receptors, FFARs), 从而对维持肠道屏障、肠道运动、激素分泌等过程产生影响, 并与非酒精性脂肪肝等多种代谢疾病相关; 2) 肠道微生物代谢氨基酸产生 5-羟色胺、吲哚等衍生物, 能够调节宿主情绪、睡眠以及免疫反应, 并与神经系统炎症和疾病的发展相关; 3) 人体微生物代谢膳食胆碱可产生三甲胺, 其转化生成的三甲胺 *N*-氧化物 (trimethylamine *N*-oxide, TMAO) 是心血管疾病和其他慢性疾病的潜在风险因子等^[69]。最近, Tintelnot 等^[70]发现肠道微生物产生的色氨酸代谢物吲哚-3-乙酸 (indole-3-acetic acid, 3-IAA) 水平与胰腺导管腺癌 (pancreatic ductal adenocarcinoma, PDAC) 的化疗疗效相关, 这一活性源于中性粒细胞衍生的髓过氧化物酶能够

氧化 3-IAA, 与化疗联用时会引起癌细胞中活性氧的积累和自噬下调, 从而损害癌细胞的代谢适应性。相比次级代谢产物, 目前针对人体微生物产生的初级代谢产物所使用的检测、分析方法和研究流程都更为成熟和完备, 因此, 进一步解析这些活性分子在错综复杂的人体环境中具体的作用靶点和通路无疑将极大推动对人体微生物活性化合物的理解和应用。

从改善健康的视角来看, 人体微生物天然产物的研究蕴藏了巨大的潜力。一方面, 很多来源于人体微生物的天然产物有望成为新型的抗菌化合物或

免疫调节分子; 另一方面, 对相关天然产物进行研究能深入地了解它们的作用机制及在人体中产生的影响, 从而在健康促进和疾病预防中采取更有效、更有针对性的预防或治疗措施。近些年来, 人体微生物研究领域的技术手段和计算工具得到了飞速发展。然而, 在高效整合已有的组学数据、寻找适用于更多微生物及不同类型 BGCs 的宿主菌或遗传操作工具、探究微生物天然产物在宿主中发挥的真实作用等方面还有巨大的进步空间。相信随着这些问题的改善和解决, 人体微生物活性天然产物的发现和利用将会更为高效和精准。

[参考文献]

- [1] Newman D J, Cragg G M. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019[J]. *J Nat Prod*, 2020, 83(3): 770-803.
- [2] Peterson J, Garges S, Giovanni M, et al. The NIH human microbiome project[J]. *Genome Res*, 2009, 19(12): 2317-2323.
- [3] Li J, Jia H, Cai X, et al. An integrated catalog of reference genes in the human gut microbiome[J]. *Nat Biotechnol*, 2014, 32(8): 834-841.
- [4] Donia M S, Fischbach M A. Small molecules from the human microbiota[J/OL]. *Science*, 2015, 349(6246): 1254766[2023-03-01]. <https://www.science.org/doi/10.1126/science.1254766>.
- [5] Mousa W K, Athar B, Merwin N J, et al. Antibiotics and specialized metabolites from the human microbiota[J]. *Nat Prod Rep*, 2017, 34(11): 1302-1331.
- [6] Milshteyn A, Colosimo D A, Brady S F. Accessing bioactive natural products from the human microbiome[J]. *Cell Host Microbe*, 2018, 23(6): 725-736.
- [7] Wang L, Ravichandran V, Yin Y, et al. Natural products from mammalian gut microbiota[J]. *Trends Biotechnol*, 2019, 37(5): 492-504.
- [8] Barber C C, Zhang W. Small molecule natural products in human nasal/oral microbiota[J/OL]. *J Ind Microbiol Biotechnol*, 2021, 48(3/4): kuab010[2023-03-01]. <https://doi.org/10.1093/jimb/kuab010>.
- [9] Dethlefsen L, Mcfall-Ngai M, Relman D A. An ecological and evolutionary perspective on human-microbe mutualism and disease[J]. *Nature*, 2007, 449(7164): 811-818.
- [10] Hu Z, Zhang W. Signaling natural products from human pathogenic bacteria[J]. *ACS Infect Dis*, 2019, 6(1): 25-33.
- [11] Götz F, Perconti S, Popella P, et al. Epidermin and gallidermin: staphylococcal lantibiotics[J]. *Int J Med Microbiol*, 2014, 304(1): 63-71.
- [12] Hyink O, Wescombe P A, Upton M, et al. Salivaricin A2 and the novel lantibiotic salivaricin B are encoded at adjacent loci on a 190-kilobase transmissible megaplasmid in the oral probiotic strain *Streptococcus salivarius* K12[J]. *Appl Environ Microbiol*, 2007, 73(4): 1107-1113.
- [13] Khmel I A, Bondarenko V M, Manokhina I M, et al. Isolation and characterization of *Escherichia coli* strains producing microcins of B and C types[J]. *FEMS Microbiol Lett*, 1993, 111(2/3): 269-274.
- [14] Byrd A L, Belkaid Y, Segre J A. The human skin microbiome[J]. *Nat Rev Microbiol*, 2018, 16(3): 143-155.
- [15] Hardy B L, Merrell D S. Friend or foe: interbacterial competition in the nasal cavity[J/OL]. *J Bacteriol*, 2021, 203(5): e00480-20[2023-03-01]. <https://doi.org/10.1128/JB.00480-20>.
- [16] Zipperer A, Konnerth M C, Laux C, et al. Human commensals producing a novel antibiotic impair pathogen colonization[J]. *Nature*, 2016, 535(7613): 511-516.
- [17] Bitschar K, Sauer B, Focken J, et al. Lugdunin amplifies innate immune responses in the skin in synergy with host-and microbiota-derived factors[J/OL]. *Nat Commun*, 2019, 10(1): 2730[2023-03-01]. <https://doi.org/10.1038/s41467-019-10646-7>.

- [18] Van Baarlen P, Troost F J, Van Hemert S, *et al.* Differential NF- κ B pathways induction by *Lactobacillus plantarum* in the duodenum of healthy humans correlating with immune tolerance[J]. *Proc Nat Acad Sci USA*, 2009, 106(7): 2371–2376.
- [19] Petrof E O, Claud E C, Sun J, *et al.* Bacteria-free solution derived from *Lactobacillus plantarum* inhibits multiple NF- κ B pathways and inhibits proteasome function[J]. *Inflamm Bowel Dis*, 2009, 15(10): 1537–1547.
- [20] Zvanych R, Lukenda N, Kim J J, *et al.* Small molecule immunomodulins from cultures of the human microbiome member *Lactobacillus plantarum*[J]. *J Antibiot*, 2014, 67(1): 85–88.
- [21] Schneditz G, Rentner J, Roier S, *et al.* Enterotoxicity of a nonribosomal peptide causes antibiotic-associated colitis[J]. *Proc Nat Acad Sci USA*, 2014, 111(36): 13181–13186.
- [22] Lagier J C, Dubourg G, Million M, *et al.* Culturing the human microbiota and culturomics[J]. *Nat Rev Microbiol*, 2018, 16(9): 540–550.
- [23] Huang Y, Sheth R U, Zhao S, *et al.* High-throughput microbial culturomics using automation and machine learning[J/OL]. *Nat Biotechnol*, 2023[2023-03-01]. <https://doi.org/10.1038/s41587-023-01674-2>.
- [24] Lagier J C, Khelaifia S, Alou M T, *et al.* Culture of previously uncultured members of the human gut microbiota by culturomics[J/OL]. *Nat Microbiol*, 2016, 1(12): 16203[2023-03-01]. <https://doi.org/10.1038/nmicrobiol.2016.203>.
- [25] 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]. *Proc Nat Acad Sci USA*, 2020, 117(1): 371–380.
- [26] Aleti G, Baker J L, Tang X, *et al.* Identification of the bacterial biosynthetic gene clusters of the oral microbiome illuminates the unexplored social language of bacteria during health and disease[J/OL]. *mBio*, 2019, 10(2): e00321-19[2023-03-01]. <https://doi.org/10.1128/mBio.00321-19>.
- [27] Zhu J, Tian L, Chen P, *et al.* Over 50 000 metagenomically assembled draft genomes for the human oral microbiome reveal new taxa[J]. *Genom Proteom Bioinf*, 2022, 20(2): 246–259.
- [28] Almeida A, Nayfach S, Boland M, *et al.* A unified catalog of 204,938 reference genomes from the human gut microbiome[J]. *Nat Biotechnol*, 2021, 39(1): 105–114.
- [29] Gacesa R, Kurilshikov A, Vich Vila A, *et al.* Environmental factors shaping the gut microbiome in a Dutch population[J]. *Nature*, 2022, 604(7907): 732–739.
- [30] Cohen L J, Kang H S, Chu J, *et al.* Functional metagenomic discovery of bacterial effectors in the human microbiome and isolation of commendamide, a GPCR G2A/132 agonist[J]. *Proc Nat Acad Sci USA*, 2015, 112(35): E4825–E4834.
- [31] Cohen L J, Esterhazy D, Kim S H, *et al.* Commensal bacteria make GPCR ligands that mimic human signalling molecules[J]. *Nature*, 2017, 549(7670): 48–53.
- [32] Piscotta F J, Whitfield S T, Nakashige T G, *et al.* Multiplexed functional metagenomic analysis of the infant microbiome identifies effectors of NF- κ B, autophagy, and cellular redox state[J/OL]. *Cell Rep*, 2021, 36(12): 109746[2023-03-01]. <https://doi.org/10.1016/j.celrep.2021.109746>.
- [33] Blin K, Shaw S, Kloosterman A M, *et al.* antiSMASH 6.0: improving cluster detection and comparison capabilities[J]. *Nucleic Acids Res*, 2021, 49(W1): W29–W35.
- [34] Starcevic A, Zucko J, Simunkovic J, *et al.* ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures[J]. *Nucleic Acids Res*, 2008, 36(21): 6882–6892.
- [35] Li M H, Ung P M, Zajkowski J, *et al.* Automated genome mining for natural products[J]. *BMC Bioinform*, 2009, 10(1): 1–10.
- [36] Röttig M, Medema M H, Blin K, *et al.* NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity[J]. *Nucleic Acids Res*, 2011, 39(suppl_2): W362–W367.
- [37] 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/OL]. *PLoS One*, 2012, 7(3): e34064[2023-03-01]. <https://doi.org/10.1371/journal.pone.0034064>.
- [38] 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/OL]. *J Biol Chem*, 2022, 298(10):

- 102480[2023-03-01]. <https://pubmed.ncbi.nlm.nih.gov/36108739/>. DOI: 10.1016/j.jbc.2022.102480.
- [39] 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.
- [40] Agrawal P, Khater S, Gupta M, *et al.* RiPPMiner: a bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links[J]. *Nucleic Acids Res*, 2017, 45(W1): W80–W88.
- [41] Tietz J I, Schwalen C J, Patel P S, *et al.* A new genome-mining tool redefines the lasso peptide biosynthetic landscape[J]. *Nat Chem Biol*, 2017, 13(5): 470–478.
- [42] van Heel A J, de Jong A, Song C, *et al.* BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins[J]. *Nucleic Acids Res*, 2018, 46(W1): W278–W281.
- [43] Cao L, Gurevich A, Alexander K L, *et al.* MetaMiner: a scalable peptidogenomics approach for discovery of ribosomal peptide natural products with blind modifications from microbial communities[J]. *Cell Syst*, 2019, 9(6): 600–608.
- [44] Sélem-Mojica N, Aguilar C, Gutiérrez-García K, *et al.* EvoMining reveals the origin and fate of natural product biosynthetic enzymes[J/OL]. *Microb Genom*, 2019, 5(12): e000260[2023-03-01]. <https://doi.org/10.1099/mgen.0.000260>.
- [45] Sugimoto Y, Camacho F R, Wang S, *et al.* A metagenomic strategy for harnessing the chemical repertoire of the human microbiome[J/OL]. *Science*, 2019, 366(6471): eaax9176[2023-03-01]. <https://www.science.org/doi/10.1126/science.aax9176>.
- [46] de Los Santos E L C. NeuRiPP: neural network identification of RiPP precursor peptides[J]. *Sci Reps*, 2019, 9(1): 1–9.
- [47] Hannigan G D, Prihoda D, Palicka A, *et al.* A deep learning genome-mining strategy for biosynthetic gene cluster prediction[J/OL]. *Nucleic Acids Res*, 2019, 47(18): e110[2023-03-01]. <https://doi.org/10.1093/nar/gkz654>.
- [48] Liu M Y, Li Y, Li H Z. Deep learning to predict the biosynthetic gene clusters in bacterial genomes[J/OL]. *J Mol Biol*, 2022, 434(15): 167597[2023-03-01]. <https://doi.org/10.1016/j.jmb.2022.167597>.
- [49] Skinnider M A, Johnston C W, Gunabalasingam M, *et al.* Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences[J/OL]. *Nat Commun*, 2020, 11(1): 6058[2023-03-01]. <https://doi.org/10.1038/s41467-020-19986-1>.
- [50] Pascal Andreu V, Roel-Touris J, Dodd D, *et al.* The gutSMASH web server: automated identification of primary metabolic gene clusters from the gut microbiota[J]. *Nucleic Acids Res*, 2021, 49(W1): W263–W270.
- [51] Pascal Andreu V, Augustijn H E, van den Berg K, *et al.* BiG-MAP: an automated pipeline to profile metabolic gene cluster abundance and expression in microbiomes[J/OL]. *Msystems*, 2021, 6(5): e00937-21[2023-03-01]. <https://doi.org/10.1128/mSystems.00937-21>.
- [52] Lemos J A, Palmer S R, Zeng L, *et al.* The biology of *Streptococcus mutans*[J/OL]. *Microbiol Spectr*, 2019, 7(1): [2023-03-01]. <https://doi.org/10.1128/microbiolspec.GPP3-0051-2018>.
- [53] Ajdić D, Meshan W M, McLaughlin R E, *et al.* Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen[J]. *Proc Natl Acad Sci USA*, 2002, 99(22): 14434–14439.
- [54] Wu C, Cichewicz R, Li Y, *et al.* Genomic island TnSmu2 of *Streptococcus mutans* harbors a nonribosomal peptide synthetase-polyketide synthase gene cluster responsible for the biosynthesis of pigments involved in oxygen and H₂O₂ tolerance[J]. *Appl Environ Microbiol*, 2010, 76(17): 5815–5826.
- [55] Joyner P M, Liu J, Zhang Z, *et al.* Mutanobactin A from the human oral pathogen *Streptococcus mutans* is a cross-kingdom regulator of the yeast-mycelium transition[J]. *Org Biomol Chem*, 2010, 8(24): 5486–5489.
- [56] Wang X, Du L, You J, *et al.* Fungal biofilm inhibitors from a human oral microbiome-derived bacterium[J]. *Org Biomol Chem*, 2012, 10(10): 2044–2050.
- [57] Zvanych R, Lukenda N, Li X, *et al.* Systems biosynthesis of secondary metabolic pathways within the oral human microbiome member *Streptococcus mutans*[J]. *Mol Biosyst*, 2015, 11(1): 97–104.
- [58] Wang M, Xie Z J, Tang S B, *et al.* Reductase of mutanobactin synthetase triggers sequential C–C macrocyclization, C–S bond formation, and C–C bond cleavage[J]. *Org Lett*, 2020, 22(3): 960–964.

- [59] Li Z R, Sun J, Du Y, *et al.* Mutanofactin promotes adhesion and biofilm formation of cariogenic *Streptococcus mutans*[J]. *Nat Chem Biol*, 2021, 17(5): 576–584.
- [60] Donia M S, Cimermancic P, Schulze C J, *et al.* A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics[J]. *Cell*, 2014, 158(6): 1402–1414.
- [61] Guo C J, Chang F Y, Wyche T P, *et al.* Discovery of reactive microbiota-derived metabolites that inhibit host proteases[J]. *Cell*, 2017, 168(3): 517–526.
- [62] Hao T T, Xie Z J, Wang M, *et al.* An anaerobic bacterium host system for heterologous expression of natural product biosynthetic gene clusters[J]. *Nat Commun*, 2019, 10(1): 1–13.
- [63] Tao L, Wang M, Guan G B, *et al.* *Streptococcus mutans* suppresses filamentous growth of *Candida albicans* through secreting mutanocyclin, an unacylated tetramic acid[J]. *Virulence*, 2022, 13(1): 542–557.
- [64] Tang X Y, Kudo Y, Baker J L, *et al.* Cariogenic *Streptococcus mutans* produces tetramic acid strain-specific antibiotics that impair commensal colonization[J]. *ACS Infect Dis*, 2020, 6(4): 563–571.
- [65] Ganzle M G, Holtzel A, Walter J, *et al.* Characterization of reutericyclin produced by *Lactobacillus reuteri* LTH2584[J]. *Appl Environ Microbiol*, 2000, 66(10): 4325–4333.
- [66] Chu J, Vila-Farres X, Inoyama D, *et al.* Discovery of MRSA active antibiotics using primary sequence from the human microbiome[J]. *Nat Chem Biol*, 2016, 12(12): 1004–1006.
- [67] Sahayasheela V J, Lankadasari M B, Dan V M, *et al.* Artificial intelligence in microbial natural product drug discovery: current and emerging role[J]. *Nat Prod Rep*, 2022, 39(12): 2215–2230.
- [68] Ma Y, Guo Z, Xia B, *et al.* Identification of antimicrobial peptides from the human gut microbiome using deep learning[J]. *Nat Biotechnol*, 2022, 40(6): 921–931.
- [69] Krautkramer K A, Fan J, Bäckhed F. Gut microbial metabolites as multi-kingdom intermediates[J]. *Nat Rev Microbiol*, 2021, 19(2): 77–94.
- [70] Tintelnot J, Xu Y, Lesker T R, *et al.* Microbiota-derived 3-IAA influences chemotherapy efficacy in pancreatic cancer[J]. *Nature*, 2023, 615(7950): 168–174.



【专家介绍】 陈义华：中国科学院微生物研究所研究员，博士生导师，微生物资源前期开发国家重点实验室常务副主任，中国科学院大学岗位教授。2005年毕业于中国科学院微生物研究所，获理学博士学位。2006年至2011年，在美国威斯康星大学麦迪逊分校（UW-Madison）从事微生物次级代谢产物生物合成博士后研究。2011年通过中国科学院“百人计划”引进到微生物研究所微生物资源前期开发国家重点实验室。

陈义华教授曾先后荣获国家自然科学基金委“优秀青年基金”（2015）和“杰出青年基金”（2020）等项目支持。主要从事微生物天然产物药物生物合成机制和人体微生物活性代谢产物的研究，并在此基础上通过基因组挖掘和合成生物学等手段解析与重构代谢，研发新结构的小分子药物。在 *Nat Biotech*, *PNAS*, *Nat Commun* 等学术刊物上发表研究论文 80 余篇。现任中国微生物学会常务理事、微生物学报副主编及 *mLife* 等学术刊物的编委。