

微生物相关“暗物质”的概述与研究方法进展

朱玉祥^{1,2#}, 林子鹏^{1#}, 王岩^{1,2,3*}

1 中国海洋大学 海洋生命学院, 海洋生物多样性与进化教育部重点实验室, 山东 青岛

2 中国海洋大学, 海洋生物多样性与进化研究所, 山东 青岛

3 崂山实验室, 海洋生态实验室, 山东 青岛

朱玉祥, 林子鹏, 王岩. 微生物相关“暗物质”的概述与研究方法进展[J]. 微生物学报, 2026, 66(4): 1519-1532.

ZHU Yuxiang, LIN Zipeng, WANG Yan. Overview and investigation method development of microbe-associated dark matter[J]. *Acta Microbiologica Sinica*, 2026, 66(4): 1519-1532.

摘要: 微生物的“暗物质”涵盖未被培养的微生物类群及其未知生物学特征, 这一概念已被微生物学界广泛认可。随着微生物研究模式的不断发展, 研究人员相继提出了一系列关于微生物的“暗物质”概念, 且大量有关“暗物质”的研究进展已被报道。然而, 这些概念内涵复杂且相互交叉, 其内在联系与区别尚未得到清晰梳理。本文系统梳理了当前各类微生物相关“暗物质”的概念框架, 并分析了微生物相关“暗物质”的研究方法及发展趋势。这将有助于微生物资源领域的研究者更清晰地认识微生物相关“暗物质”的内涵与研究现状, 推动微生物的“暗物质”资源挖掘工作向更深层次迈进。

关键词: 微生物资源; 暗物质; 微生物暗物质; 基因资源

资助项目: 国家重点研发计划(2024YFC2816000); 国家自然科学基金(42176108, 32370030, 324B2002); 山东省自然科学基金(ZR2024JQ035); 山东省青年泰山学者(tsqn202103029); 崂山实验室科技创新项目(LSKJ202203205); 中央高校基本科研业务费专项资金(202442001)

This work was supported by the National Key Research and Development Program of China (2024YFC2816000), the National Natural Science Foundation of China (42176108, 32370030, 324B2002), the Natural Science Foundation of Shandong Province (ZR2024JQ035), the Young Taishan Scholars Program of Shandong Province (tsqn202103029), the Science and Technology Innovation Project of Laoshan Laboratory (LSKJ202203205), and the Fundamental Research Funds for the Central Universities (202442001).

[#]These authors contributed equally to this work.

*Corresponding author. E-mail: wangy12@ouc.edu.cn

Received: 2025-11-30; Accepted: 2026-02-04; Published online: 2026-02-10

Overview and investigation method development of microbe-associated dark matter

ZHU Yuxiang^{1,2#}, LIN Zipeng^{1#}, WANG Yan^{1,2,3*}

1 MOE Key Laboratory of Evolution and Marine Biodiversity, College of Marine Life Sciences, Ocean University of China, Qingdao, Shandong, China

2 Institute of Evolution & Marine Biodiversity, Ocean University of China, Qingdao, Shandong, China

3 Laboratory for Marine Ecology, Laoshan Laboratory, Qingdao, Shandong, China

Abstract: The dark matter of microbes, encompassing uncultivated microbial taxa and unknown biological features, has been widely accepted by microbiologists. With the advancement of microbial investigation methods, a series of dark matter concepts related to microbes have been successively proposed, and numerous research advances in this field have been well-documented. However, these concepts exhibit complex and overlapping connotations, and their inherent connections and distinctions have not yet been clearly clarified. This review work systematically summarizes the current conceptual frameworks of different dark matter types of microbes and analyzes the relevant investigation methodologies and their development trends. This article will help researchers in the microbial resource field gain a clearer understanding of the conceptual connotation and current research status of microbe-associated dark matter and promote the in-depth development of dark matter resource mining.

Keywords: microbial resources; dark matter; microbial dark matter; gene resources

微生物是全球生态系统的重要组成部分^[1-2], 蕴含着丰富的生物活性资源^[3-4]。挖掘和解析微生物资源是揭示其生态功能^[5-6]、促进社会生产力进步^[7]的重要途径之一。其中, 微生物中的“暗物质”是一类重要的研究对象。“暗物质”是天体物理学的经典概念, 用于描述宇宙中推测存在但难以直接观测的非发光物质^[8]。微生物中的“暗物质”被用于形容和强调微生物的未知成员、生物学特征及功能物质。早在 2005 年, Filée 等^[9]将细菌噬菌体描述为生物界的“暗物质”, 以此凸显它们的未知特性——尽管细菌噬菌体数量庞大且分布广泛, 但其中仅有极小比例已被鉴定, 绝大多数仍处于未知状态。2007 年, Marcy 等^[10]完善了微生物研究中的“暗物质”概念雏形, 指出地球上只有极小部分的微生物能够在实验室条件下培养和研究, 而绝大多数微

生物物种的特征只能通过序列数据进行推测和描述。这些“大量存在”“被间接推测”“难以直接检测”的特征, 与天体物理学中的“暗物质”概念十分相似^[10]。Marcy 等^[10]提出的这一理论框架被后续一系列重要的“暗物质”研究^[11-12]所引用, 为“暗物质”术语在微生物资源领域的蓬勃发展奠定了基础。

无论是微生物的早期“暗物质”理论框架^[10]还是后续的相关研究^[11], 序列分析始终是鉴定和探究各类微生物相关“暗物质”的主要手段, 并持续推动着“暗物质”概念的发展和细化。相较于针对已分离培养菌株的全基因组分析^[13], 宏基因组学手段可直接对自然环境样品中的 DNA 进行测序与分析^[11,14]。通过宏基因组分析, 研究人员发现微生物群落中存在大量功能未知、尚未得到注释和分类的 DNA 序列^[11,15], 并将其

统称为“宏基因组暗物质”(metagenomic dark matter)^[7,16-18](图1)。“宏基因组暗物质”已成为发掘新型微生物类群^[19-21]和生物活性物质^[22-23]的重要宝库。在此基础上,研究人员通过分析宏基因组组装基因组(metagenome-assembled genomes, MAGs)和单细胞基因组(single-cell genomes),得以从基因组尺度解析环境微生物群落特征^[5,11]。这些研究表明,环境中广泛存在尚未表征或难以培养的微生物基因组或物种,它们被命名为“微生物暗物质”(microbial dark matter, MDM)^[24-27](图1)。MDM是微生物中最经典的“暗物质”概念,现已得到微生物资源领域研究人员的广泛认可^[28-32]。

随着微生物资源研究的不断深入,与微生物有关的更多新型“暗物质”概念被相继提出^[33-36]。然而,这些概念之间的联系和区别尚未得到系统性梳理。例如,MDM的范畴并不局限于宏基因组来源的未知基因组,也包括基于16S rRNA基因扩增子发现的假定类群^[37-38];由于MDM和宏基因组分析高度关联,部分研究混用了MDM和“宏基因组暗物质”两个概念^[15]。此外,对于涉及微生物基因资源的众多“暗物质”^[34-36,39]而言,它们中的一部分来源于已培养微生物的基因组^[40],而另一些则依赖于从宏基因组^[35,41]或MDM^[7]中发现。基于上述现状,本文梳理了微生物领域中不同“暗物质”概念之间的关系(图1,表1^[11,18,35,39-46]),并对研究方法的进展和融合前景进行了分析与展望,旨在帮助微生物资源领域的研究者更全面、清晰地认识“暗物质”研究现状。

1 微生物物种资源“暗物质”: 鉴定和培养方法

MDM的核心定义为:基于现有技术手段难以分离培养或充分表征的微生物类群及基因组^[11,47]。Lagkouvardos等^[48]认为,MDM并非真实存在的生物分类单元,因为理论上所有微生物

皆可被分离培养,MDM的“不可培养性”仅源于当前分离技术的局限。这一观点从侧面反映了MDM研究领域的一个共识,即MDM的分离培养极具挑战性。这一挑战推动了组学分析手段的蓬勃发展^[11,49-51],促进了组学技术与新型分离培养方案的逐步融合^[24,52-54],为发现、鉴定和解析MDM提供了有力支持。

1.1 基于生物信息学的MDM分析及其局限性

MAGs是鉴定和分析环境样品中潜在MDM的重要数据来源^[6,55]。然而,由于MAGs是通过宏基因组原始测序数据进行基因组重建(reconstruction)获得的产物^[55-56],其承载的基因组信息本质上是基于序列特征推测得到的^[56]。因此,如何从海量宏基因组数据中获得高质量、高置信度的MAGs以开展对MDM的初步鉴定和分析,仍是本领域的核心难题。这一难题引发了领域内对MDM研究结果的质疑,主要集中在2个方面:(1)DNA污染问题,宏基因组样品中的污染可能严重干扰基因组组装,导致获得的MDM信息与实际不符^[57]; (2)死亡细胞争议,即样本中被识别为MDM的成分可能并非真正不可培养,而是本身已处于死亡状态^[58]。

为降低宏基因组数据中DNA污染对MAGs组装质量的影响,Parks等^[55]制定了宏基因组组装标准,并依据这一标准从宏基因组中获得了完整性更高、污染程度更低的MDM基因组。2020年,该团队进一步提出将代表性基因组作为参考进行物种命名,实现了对MDM基因组的快速批量命名^[59]。这种基于序列信息的MDM命名法(nomenclature),在微生物分类学领域引发了广泛的讨论和关注。长期以来,原核生物命名始终遵循国际原核生物命名法规(international code of nomenclature of prokaryotes, ICNP)的相关标准。然而,ICNP对活体生物材料严格的要求标准,导致大量MDM无法获得有效名称(validly published name)^[60]。为解决这一问

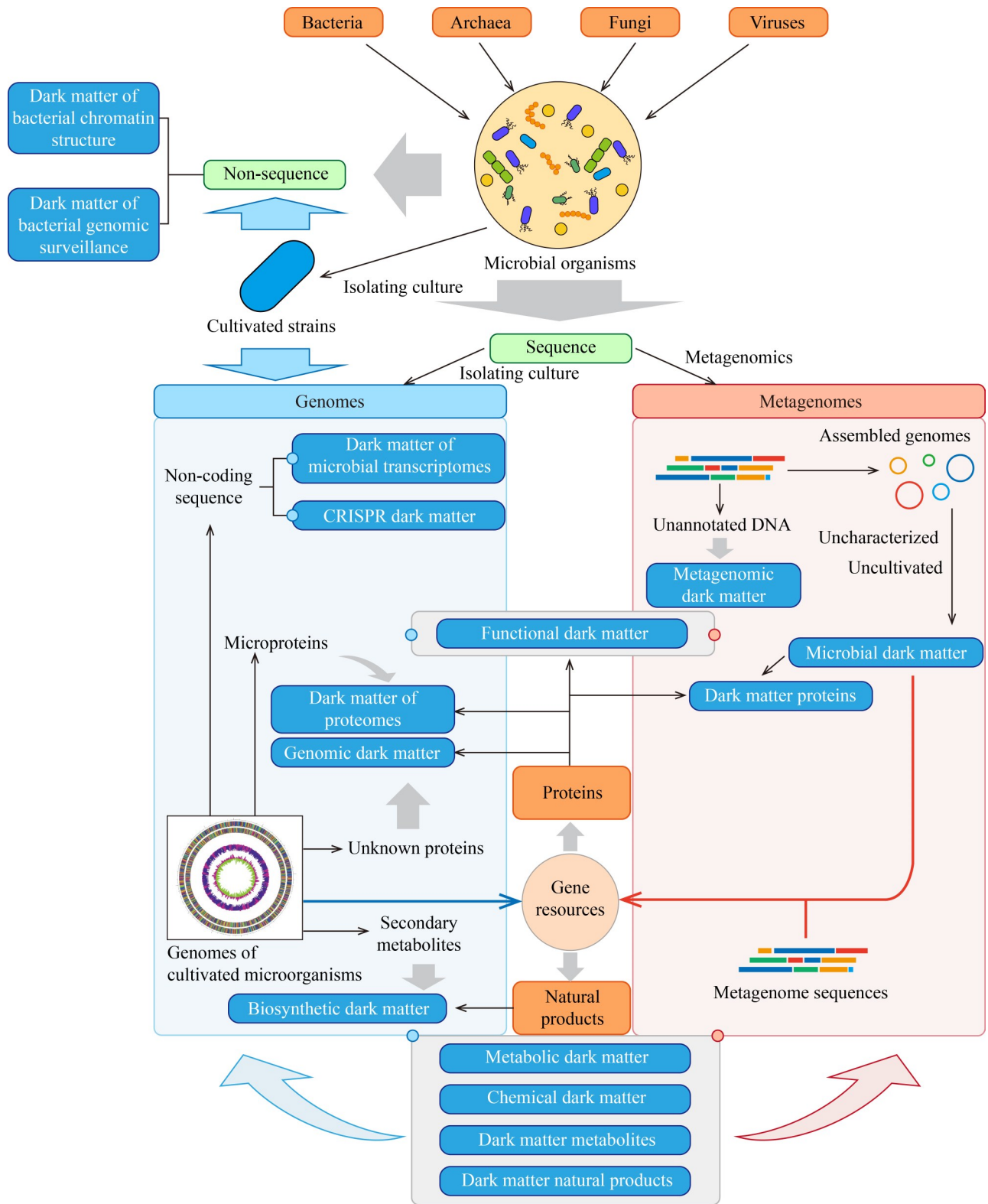


图1 微生物相关“暗物质”的分类示意图

Figure 1 Schematic representation of microbe-associated “dark matter” classification.

表1 微生物序列特征“暗物质”的概念和研究方法

Table 1 Concepts and illumination methods of “dark matter” of microbial sequence features

Concept name	Object		Sources	Illumination methods	References
Metagenomic dark matter	DNA sequence		Metagenomes	Sequence annotation	[18]
MDM	Microbial species, strains, and genomes		Environmental samples and metagenomes	Meta-omic profiling analysis, single-cell genomics, single-cell Raman, culturomics, isolation chip, single-cell-level microfluidics, and cross-feeding	[11]
Genomic dark matter	Proteins (including microproteins)	Proteins and genes	Genomes of cultivated microorganisms	Sequence annotation, genetic manipulation, expression, and purification	[40]
Dark matter proteins	Proteins (including microproteins)	Proteins and genes	MDM	Sequence annotation and heterologous biosynthesis	[35]
Functional dark matter	Proteins (including microproteins)	Proteins and genes	Metagenomes	Sequence annotation and structural clustering	[41]
Dark matter of proteomes	Proteins (including microproteins)	Microproteins	Genomes of cultivated microorganisms	Genetic manipulation, expression, and purification	[42]
Metabolic dark matter	Natural products and BGCs		Genomes of cultivated microorganisms and metagenomes (involving MDM)	Chemical synthesis and heterologous biosynthesis	[43]
Chemical dark matter	Natural products and BGCs		Genomes of cultivated microorganisms and metagenomes (involving MDM)	Chemical synthesis and heterologous biosynthesis	[43]
Dark matter metabolites	Natural products and BGCs		Genomes of cultivated microorganisms and metagenomes (involving MDM)	Chemical synthesis and heterologous biosynthesis	[39]
Dark matter natural products	Natural products and BGCs		Genomes of cultivated microorganisms and metagenomes (involving MDM)	Chemical synthesis and heterologous biosynthesis	[39]
Biosynthetic dark matter	Natural products and BGCs	Secondary metabolites (belonging to natural products) and BGCs	Genomes of cultivated microorganisms	Chemical synthesis, heterologous biosynthesis, and signal-induced expression enhancement	[44]
CRISPR dark matter	Non-coding sequences	Spacer sequences	Genomes of cultivated microorganisms	Sequence annotation	[45]
Dark matter of microbial transcriptomes	Non-coding sequences	Non-coding RNA	Genomes of cultivated microorganisms	Genetic manipulation	[46]

MDM: Microbial dark matter; BGCs: Biosynthetic gene clusters.

题, 本领域的研究者于近几年发布了“从序列数据描述原核生物命名法”(code of nomenclature of prokaryotes described from sequence data, SeqCode)^[60]。SeqCode 独立于 ICNP, 并且将基因组序列纳入有效命名材料, 其不但有助于规范 MDM 的鉴定体系, 更有望进一步促进 MDM 分析技术的进步。值得注意的是, 张道锋等^[60]对 SeqCode 和 ICNP 进行的系统性分析指出, SeqCode 仍面临诸多挑战: 两大命名标准的并行实施可能导致重复命名问题, 而且 SeqCode 对 MDM 命名的有效性仍有待进一步验证。

除序列信息外, 领域内研究者还尝试将其他维度的信息纳入分析框架以进一步提升分析准确性。Zamkovaya 等^[61]将物种互作网络融入宏基因组的重建和分类过程, 这一方法不仅能鉴定环境样本中的潜在 MDM 成员, 还能揭示其在群落中的生态学功能, 即 MDM 在环境群落中承担关键枢纽作用。此外, 第 2 代测序技术(如 Illumina 平台)是获取宏基因组数据集的主要手段, 但作为短读长测序技术, 其局限性导致微生物基因组中的大量信息被“隐藏”, 难以被有效表征和分析^[62]。可实现长读长测序的第 3 代测序技术(如 Oxford Nanopore Technologies 和 PacBio Sequel II 平台)被认为是弥补这一缺陷的有效手段^[62-64]。这一观点已得到了后续研究的验证。Kato 等^[64]在 2022 年将第 3 代长读长测序技术应用于宏基因组分析, 成功获得了更高质量的 MAGs, 为鉴定 MDM 提供了有力支持。

尽管基于宏基因组的 MDM 鉴定方法持续完善, 上述局限性仍未得到完全解决^[55]。这些问题的核心争议在于: 基于现有技术手段, MAGs 本质上是通过混杂序列片段数据集推测获得的结果, 缺乏证明相关生物或细胞真实存在的直接证据。相比之下, 单细胞测序技术虽然要求更高^[11], 但因其能提供来自真实细胞的直接证据, 被视为突破宏基因组局限的关键^[55]。目前, 单细胞基因组和宏基因组的联合分析已经成为 MDM 研究的新兴方向^[65-66]。此外, 拉

曼光谱因能生成单细胞特异性的识别图谱^[67-68], 被应用于单细胞分析^[69]和 MDM 鉴定研究中^[49]。这些融合拉曼光谱的微生物研究方法已与宏基因组分析、单细胞基因组分析等手段相结合^[70-72], 用于 MDM 的鉴定和生理特征分析^[49,72]。Probst 等^[49]先通过宏基因组分析发现一类未被分离培养的古菌, 再利用拉曼光谱显微技术分析其生理特征; Song 等^[72]则先通过单细胞拉曼检测结果筛选出含有胡萝卜素的细胞, 再借助单细胞基因组测序鉴定出这些细胞中的 MDM 成员。值得注意的是, 单细胞拉曼光谱检测技术还可区分细胞的活体或死亡状态^[73], 这为解决上述“死亡细胞争议”提供了新的思路。

1.2 生物信息学技术助力 MDM 分离培养

如前所述, 分离培养是 MDM 研究的核心难题, 其主要障碍在于: MDM 在自然栖息地正常生长所需的化学条件和物理参数, 同实验室培养体系存在显著差异^[74]。一种主流的解决方案是在实验室条件下尽可能还原自然栖息地的营养条件^[74-75]。此外, 部分 MDM 的生长依赖于与其他物种之间的相互作用, 这一特征使其难以被分离纯培养^[74,76-77]。为应对上述挑战, 一系列分离培养方案应运而生, 如分离芯片(isolation Chip, iChip)^[53]、单细胞微流控技术(single-cell-level microfluidics)^[54]、交叉喂养(cross-feeding)^[76]等。其中, Vincy 等^[53]于 2025 年研发的 iChip 技术可基于多通道细胞扩散原理实现对土壤样本中 MDM 细胞的回收、分离和原位培养; 微流控技术可通过微通道实现对单细胞的捕获和分离, 已广泛应用于涉及单细胞分析的 MDM 研究^[47,50,77]; 交叉喂养指不同微生物间通过共享代谢物实现的相互作用^[78], 部分 MDM 对交叉喂养的生长依赖性使其成为基于共培养分离 MDM 的重要研究思路^[76]。培养组学(culturomics)体系作为一种高通量分离培养方案, 已和宏基因组分析技术发生融合^[24,52]。

Li 等^[24]以沙漠土壤样品为研究对象, 提出了“基于培养组学的宏基因组学”(culturomics-based metagenomics, CBM)技术方案, 成功分离培养出了大量属于 MDM 范畴的候选细菌新分类单元。该方案以高通量分离培养为基础, 融合宏基因组分析等生物信息学分析技术^[24], 为 MDM 分离培养提供了新思路。Nedjad 等^[79]提出了一种“逆向设计”思路: 先通过 MDM 基因组注释构建基因组级代谢网络, 再据此逆向设计分离培养的营养条件。上述案例均展现了生物信息技术和分离培养实验相融合的巨大潜力, 是 MDM 分离培养方案的重要发展方向。

1.3 MDM 的原位观察技术助力分离培养

鉴于 MDM 分离培养的高难度, 部分研究者转向了原位观察策略: 在不进行体外分离培养的情况下, 利用荧光原位杂交(fluorescence *in situ* hybridization, FISH)技术对宿主生境依赖型 MDM 进行原位观察^[80-81]。Lin 等^[82]利用该技术揭示了小鼠肠道 MDM 的分类、生长与分裂模式, 证明了 FISH 技术是揭示 MDM 动态变化的强大工具。此外, FISH 技术可与质谱成像相结合^[83]。Lackner 等^[83]以此阐明了海绵宿主内 MDM 的代谢特征, 并提议将此类原位观察结果用于定向设计 MDM 的分离培养方案。这类原位观察方案不但可以提供实验观测证据, 还可与生物信息学分析形成互补, 为破解 MDM “无法培养”的难题提供全新视角。

2 微生物的基因资源“暗物质”: 概念与研究方法

基因资源是微生物资源的重要组成部分^[41,84]。基于宏基因组^[35,41]、MDM^[7]和已培养微生物基因组^[40]中的未知基因, 研究人员提出了一系列“暗物质”相关概念^[34-36,39]。然而, 这些概念在内涵上相互交叉, 彼此间的联系和区别尚未得到清晰梳理。本部分将系统梳理聚焦微

生物基因资源的各类“暗物质”, 并探讨其内在联系与研究方法进展。

2.1 微生物的蛋白相关“暗物质”

在微生物基因资源研究中, “基因组暗物质”是具有代表性的典型概念(图 1, 表 1)。Piao 等^[40]以微生物纤维素酶基因为主要研究对象, 通过传统的序列注释技术, 筛选出了与已知蛋白序列相似性很低的蛋白及其编码基因, 并将这类源于已培养微生物基因组、编码未知功能蛋白的基因序列定义为“基因组暗物质”。微生物基因组编码的蛋白质还包括其他类型, 比如由小开放阅读框(small open reading frames, smORFs)编码的微生物微蛋白(microbial microproteins)^[42,85]。smORFs 指真核生物中短于 100 个密码子^[42,86]、原核生物中短于 50、70 或 100 个密码子^[42]的开放阅读框(open reading frames, ORFs)序列, 其序列规模较小, 导致编码的蛋白质难以被有效注释, 其生物学功能常受到质疑^[42,85]。基于此, Fesenko 等^[42]将这类微蛋白定义为“蛋白质组暗物质”(dark matter of proteomes)(图 1, 表 1)。目前, “蛋白质组暗物质”研究只局限于已培养微生物基因组(图 1), 尚未涉及宏基因组或 MDM 领域。

Pavlopoulos 等^[41]通过分析全球宏基因组数据, 预测了超 11 亿个未被研究的蛋白质序列(无法映射到公共参考基因组蛋白质家族中), 并将它们归类为“功能性暗物质”(functional dark matter)(图 1, 表 1)。尽管该概念最初源于宏基因组研究^[41], 但后续也被用于描述已培养微生物基因组中的未知蛋白^[87]。此外, “暗物质蛋白”(图 1, 表 1)一词目前只出现在 MDM 基因资源的研究中^[35], 其可能专门指代 MDM 基因组中的未知基因与蛋白资源。

2.2 微生物的天然产物相关“暗物质”

除直接编码蛋白的基因外, 已培养微生物基因组中的生物合成基因簇(biosynthetic gene clusters, BGCs)^[33,88]及其对应的天然产物^[43,89],

也是重要的微生物基因资源。针对此类未知的天然产物及其 BGCs, Santos-Aberturas 等^[43]提出了“代谢暗物质”和“化学暗物质”(chemical dark matter)的概念。尽管 Santos-Aberturas 等^[43]将多酮类、萜烯、脂质、多糖类等未知代谢物称为“代谢暗物质”,将未知肽类天然产物(peptidic natural products, PNPs)称为“化学暗物质”,但未对这 2 个定义作出明确界定。此外, Gurevich 等^[89]也将微生物基因组中的 PNPs 称为“代谢暗物质”。这些案例表明,“代谢暗物质”和“化学暗物质”的内涵实质相同,均指代微生物基因组中的未知天然产物及其 BGCs^[43,88-89]。基于相同的核心内涵,Williams 等^[39]将此类物质称为“暗物质代谢物”(dark matter metabolites)或“暗物质天然产物”(dark matter natural products)(图 1,表 1)。值得注意的是,这些“暗物质”定义同样适用于 MDM 或宏基因组来源的未知天然产物资源^[39,43],因此这类概念的内涵趋于一致,均指代未知天然产物及其 BGCs,而且涵盖了已培养微生物基因组、宏基因组和 MDM 三大来源。

在微生物研究中,与天然产物有关的“暗物质”通常包含了次级代谢产物(secondary metabolites, SMs)^[39,43,89]。微生物 SMs 指代不直接参与生长、发育或繁殖,但可能在种间互动中发挥作用的代谢产物^[90]。由于分离培养菌株中的部分 SMs 在常规培养条件下转录水平很低或根本不表达^[90-91],这类“沉默”资源被进一步细分为“生物合成暗物质”(biosynthetic dark matter)^[39,44](图 1,表 1)。

2.3 “暗物质”基因资源研究的方法融合

与已培养微生物的基因组相比^[40],宏基因组和 MDM 中的基因资源“暗物质”难以通过遗传操作直接研究,对生物信息学分析方法的依赖性更强^[84]。虽然传统的分析方法(如结合结构域信息的功能注释、BGCs 预测、MAG 分析等)

仍在使用的^[92-94],但其解析效率已难以满足海量数据的挖掘需求^[84]。针对这一现状,Hoarfrost 等^[84]提出将深度学习技术融入宏基因组基因资源“暗物质”的注释流程中;他们开发了名为 LookingGlass 的分析工具,通过深度学习整合复杂的高维生物变量,不仅能以高置信度注释传统方法无法识别的基因序列,还能捕捉序列间的进化关系及生态信息(如丰度维度、深度趋势等传统注释技术无法反映的信息)。此外,蛋白质结构预测也被纳入微生物基因资源“暗物质”的分析框架^[41,95]。Pavlopoulos 等^[41]在序列分析的基础上,通过蛋白结构聚类分析,成功注释了低序列相似性的基因资源。这一流程后续被 Sussfeld 等^[95]采纳,用于解析海洋宏基因组数据库中的“功能性暗物质”。

基因资源研究的终极目标是获取其编码的生物活性物质。针对 MDM 难以分离培养的瓶颈^[11,47,74],异源生物合成已成为挖掘此类基因资源“暗物质”的主要手段^[7,96]。Santos 等^[7]实现了分析技术与合成研究的融合,构建了从“宏基因组分析与 MDM 鉴定”到“MDM 关键活性基因资源注释”,再到“异源生物合成”的完整流程,为从宏基因组和 MDM 中鉴定、挖掘、表达及利用基因资源“暗物质”提供了优质范例。针对“生物合成暗物质”的转录“沉默”特性,研究者通过化学合成^[97]、异源生物合成^[44]和信号诱导表达增强^[98]等方式开展研究。例如,Zhong 等^[98]通过添加种间信号分子吡啶,激活细菌“沉默”BGCs 的表达,显著提升了目标 SMs 的产量。这类由序列分析引领、合成生物学技术跟进的融合范式,正逐步揭开微生物基因资源“暗物质”的神秘面纱。

3 微生物的非编码序列“暗物质”及其研究前景

除可编码基因序列外,微生物基因组中的非编码序列同样是重要组成成分,并被纳入“暗

物质”的研究范畴。其中,“CRISPR 暗物质”(CRISPR dark matter)^[45]是典型代表(图 1, 表 1)。该概念特指原核生物基因组中来源不明的间隔区(spacers)序列。原核生物通过 CRISPR-Cas 系统获得对噬菌体的适应性免疫,这一过程高度依赖于非编码的间隔区序列^[99-100]。尽管间隔区常因与病毒序列同源^[99]而被 CRISPR 系统识别,但是绝大多数间隔区序列的来源仍然成谜,难以匹配相应的病毒片段^[99]。此外,微生物的非编码“暗物质”概念也延伸至转录水平。Wade 等^[46]将微生物基因组中广泛存在的非编码 RNA (non-coding RNA, ncRNA)转录定义为“微生物转录组暗物质”(dark matter of microbial transcriptomes)(图 1, 表 1)。目前,“微生物转录组暗物质”研究主要局限于已培养微生物的基因组^[45-46],这可能由于宏基因组或 MDM 中 ncRNA 的研究仍面临技术瓶颈,且参考数据集相对匮乏。由于 ncRNA 的表征依赖于实验室分离培养和遗传操作体系^[101],宏基因组或 MDM 来源的 ncRNA 在未来较长时间内可能仍处于未被揭示的“黑暗”状态。

已有研究表明,深度学习模型可通过捕捉复杂的高维生物特征,预测和解析微生物基因组中的基因资源“暗物质”^[84]。因此,整合实验数据和多维度信息训练深度学习模型,或许是预测宏基因组来源“微生物转录组暗物质”的未来发展方向。对“CRISPR 暗物质”而言,当前研究仍聚焦于已培养微生物的基因组,其核心研究方法是序列注释^[99]。Rubio 等^[99]在 2023 年对 6 种致病菌的数千个菌株进行基因组分析,发现“CRISPR 暗物质”在这些菌株中具有独特的分布特征和显著的分布差异。考虑到“CRISPR 暗物质”的分布复杂性^[99]及宏基因组的海量数据^[84],基于多套宏基因组数据集开展的相关研究对硬件算力与时间成本提出了很高要求,这可能是目前尚未出现宏基因组层面“CRISPR 暗物质”研究的主要原因。

4 微生物非序列特征相关的“暗物质”

在微生物研究中,尽管 MDM 可指代微生物物种本身^[24],但目前多数的“暗物质”聚焦于序列层面的特征,如基因组^[55]、基因序列^[40]、蛋白序列^[41]、非编码序列^[46]等。在序列特征之外,一些非序列特征相关的“暗物质”在近年来被相继提出。Racki 等^[102]将细菌染色质中的多聚磷酸盐(polyphosphate, polyP)定义为“细菌染色质结构暗物质”(dark matter of bacterial chromatin structure)(图 1)。此外,由于抗生素耐药性质粒的传播状态难以持续追踪,Sobkowiak 等^[103]将这类未被有效监测的质粒传播状态命名为“细菌基因组监测的暗物质”(dark matter of bacterial genomic surveillance)(图 1)。需要强调的是,这些非序列特征相关的“暗物质”目前仅处于初步概念化阶段,尚缺乏深入的实证讨论。随着“暗物质”研究维度的不断拓展,这些新兴“暗物质”定义在未来的应用中或将面临更多挑战。

5 总结与展望

“暗物质”已成为微生物资源领域广泛认可的重要术语^[104],它们不仅涵盖微生物物种及基因组本身,还深入扩展至不同来源基因组中的基因资源及其功能产物。微生物的“暗物质”研究体系主要根植于“物种挖掘”和“基因资源开发”两大维度,在揭示微生物生态功能、挖掘微生物资源过程中发挥重要作用,并随技术进步持续更新。然而,虽然“暗物质”概念在微生物研究中发展迅速,但各类概念间的内在联系仍不清晰。针对这一现状,本文从宏基因组、微生物物种、基因组、基因资源、非编码序列及非序列特征等多个维度,系统梳理了微生物研究中各类“暗物质”之间的内在联系,并分析了现有研究方法及未来融合方向。

目前绝大多数与微生物有关的“暗物质”(序列特征类暗物质)均高度依赖生物信息学分析

(图 1)。这契合了 Marcy 等^[10]在 2007 年提出的微生物领域“暗物质”基本概念框架,即这一“暗物质”需通过序列特征分析推测其存在。然而,近年来涌现的如“细菌染色质结构暗物质”“细菌基因组监测的暗物质”等非序列特征相关的“暗物质”,则脱离了单纯的序列分析框架,转而用于强调更为广义的“生物学未知性”。这一趋势既体现了“暗物质”概念在微生物领域的高度认可,也引发了对其概念边界过度扩张甚至走向无序化的担忧。

值得注意的是,序列特征类“暗物质”还存在命名冗余的问题。例如,“暗物质代谢物”“暗物质天然产物”“化学暗物质”等概念均用于描述微生物的未知天然产物^[39,43]。此外,微生物研究中的部分“暗物质”的命名源自真核生物研究^[46],不同研究方向研究者的认知差异可能是导致命名冗余的主要原因之一。因此,未来亟需微生物领域专家学者共同制定统一的“暗物质”命名规范,建立清晰、严谨的命名体系。综上所述,本文系统梳理了微生物资源领域的各类“暗物质”概念及研究动向,为本领域研究者提供清晰的概念框架,以期推动微生物界“暗物质”的资源挖掘向更深层次、更具系统性的方向发展。

作者贡献声明

朱玉祥、林子鹏:执行调研,撰写文章;王岩:提出概念,获取基金,项目管理,审阅。

作者利益冲突公开声明

作者声明不存在任何可能会影响本文所报告工作的已知经济利益或个人关系。

参考文献

- [1] Cell Editorial Team. A microbial future[J]. *Cell*, 2024, 187(19): 5119-5120.
- [2] Bahram M, Hildebrand F, Forslund SK, Anderson JL, Soudzilovskaia NA, Bodegom PM, Bengtsson-Palme J, Anslan S, Coelho LP, Harend H, Huerta-Cepas J, Medema MH, Maltz MR, Mundra S, Olsson PA, Pent M, Pölme S, Sunagawa S, Ryberg M, Tedersoo L, et al. Structure and function of the global topsoil microbiome[J]. *Nature*, 2018, 560(7717): 233-237.
- [3] Chen JW, Jia YY, Sun Y, Liu K, Zhou CH, Liu C, Li DH, Liu GL, Zhang CS, Yang T, Huang L, Zhuang YY, Wang DZ, Xu DY, Zhong QL, Guo Y, Li AD, Seim I, Jiang L, Wang LS, et al. Global marine microbial diversity and its potential in bioprospecting[J]. *Nature*, 2024, 633(8029): 371-379.
- [4] Proctor LM, Creasy HH, Fettweis JM, Lloyd-Price J, Mahurkar A, Zhou WY, Buck GA, Snyder MP, Strauss JF, Weinstock GM III, White O, Huttenhower C. The integrative human microbiome project[J]. *Nature*, 2019, 569(7758): 641-648.
- [5] Nayfach S, Roux S, Seshadri R, Udworthy D, Varghese N, Schulz F, Wu DY, Paez-Espino D, Chen IM, Huntemann M, Palaniappan K, Ladau J, Mukherjee S, Reddy TBK, Nielsen T, Kirton E, Faria JP, Edirisinghe JN, Henry CS, Jungbluth SP, et al. A genomic catalog of Earth's microbiomes[J]. *Nature Biotechnology*, 2021, 39(4): 499-509.
- [6] Williams TJ, Allen MA, Berengut JF, Cavicchioli R. Shedding light on microbial “dark matter”: insights into novel *Cloacimonadota* and *Omnitrophota* from an antarctic lake[J]. *Frontiers in Microbiology*, 2021, 12: 741077.
- [7] Santos CA, Morais MAB, Mandelli F, Lima EA, Miyamoto RY, Higasi PMR, Araujo EA, Paixão DAA, Junior JM, Motta ML, Streit RSA, Morão LG, Silva CBC, Wolf LD, Terrasan CRF, Bulka NR, Diogo JA, Fuzita FJ, Colombari FM, Santos CR, et al. A metagenomic ‘dark matter’ enzyme catalyses oxidative cellulose conversion[J]. *Nature*, 2025, 639(8056): 1076-1083.
- [8] Bahcall NA. Dark matter universe[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2015, 112(40): 12243-12245.
- [9] Filée J, Tétart F, Suttle CA, Krisch HM. Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, 102(35): 12471-12476.
- [10] Marcy Y, Ouverney C, Bik EM, Lösekann T, Ivanova N, Martin HG, Szeto E, Platt D, Hugenholtz P, Relman DA, Quake SR. Dissecting biological “dark matter” with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104(29): 11889-11894.
- [11] Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu WT, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, et al. Insights into the phylogeny and coding potential of microbial dark matter[J]. *Nature*, 2013, 499(7459): 431-437.
- [12] Wu DY, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, Hooper SD, Pati A, Lykidis A, Spring S, Anderson IJ, D’haeseleer P, Zemla A, Singer M, Lapidus A, Nolan M, et al. A phylogeny-driven genomic encyclopaedia of

- Bacteria and Archaea[J]. *Nature*, 2009, 462(7276): 1056-1060.
- [13] Alam K, Zhao YM, Lu XF, Gong K, Zhong L, Hao JF, Islam MM, Islam S, Li G, Zhang YM, Li RJ, Li AY. Isolation, complete genome sequencing and *in silico* genome mining of *Burkholderia* for secondary metabolites[J]. *BMC Microbiology*, 2022, 22: 323.
- [14] Handelsman J. Metagenomics: application of genomics to uncultured microorganisms[J]. *Microbiology and Molecular Biology Reviews*, 2004, 68(4): 669-685.
- [15] Osburn ED, McBride SG, Strickland MS. Microbial dark matter could add uncertainties to metagenomic trait estimations[J]. *Nature Microbiology*, 2024, 9(6): 1427-1430.
- [16] Armengaud J. Metaproteomics to understand how microbiota function: the crystal ball predicts a promising future[J]. *Environmental Microbiology*, 2023, 25(1): 115-125.
- [17] Shillitoe EJ. The microbiome of oral cancer[J]. *Critical ReviewsTM in Oncogenesis*, 2018, 23(3/4): 153-160.
- [18] Lobb B, Kurtz DA, Moreno-Hagelsieb G, Doxey AC. Remote homology and the functions of metagenomic dark matter[J]. *Frontiers in Genetics*, 2015, 6: 234.
- [19] Sharma V, Colson P, Giorgi R, Pontarotti P, Raoult D. DNA-dependent RNA polymerase detects hidden giant viruses in published databanks[J]. *Genome Biology and Evolution*, 2014, 6(7): 1603-1610.
- [20] King CH, Desai H, Sylvestry AC, LoTempio J, Ayanyan S, Carrie J, Crandall KA, Fochtman BC, Gasparyan L, Gulzar N, Howell P, Issa N, Krampis K, Mishra L, Morizono H, Pisegna JR, Rao SY, Ren Y, Simonyan V, Smith K, et al. Baseline human gut microbiota profile in healthy people and standard reporting template[J]. *PLoS One*, 2019, 14(9): e0206484.
- [21] Lu X, Dai ZY, Xue JX, Li W, Ni P, Xu J, Zhou CL, Zhang W. Discovery of novel RNA viruses through analysis of fungi-associated next-generation sequencing data[J]. *BMC Genomics*, 2024, 25: 517.
- [22] Zhang Y, Li RZ, Zou G, Guo YT, Wu RW, Zhou Y, Chen HC, Zhou R, Lavigne R, Bergen PJ, Li J, Li JQ. Discovery of antimicrobial lysins from the “dark matter” of uncharacterized phages using artificial intelligence[J]. *Advanced Science*, 2024, 11(32): 2404049.
- [23] Keshri V, Panda A, Levasseur A, Rolain JM, Pontarotti P, Raoult D. Phylogenomic analysis of β -lactamase in archaea and bacteria enables the identification of putative new members[J]. *Genome Biology and Evolution*, 2018, 10(4): 1106-1114.
- [24] Li S, Lian WH, Han JR, Ali M, Lin ZL, Liu YH, Li L, Zhang DY, Jiang XZ, Li WJ, Dong L. Capturing the microbial dark matter in desert soils using culturomics-based metagenomics and high-resolution analysis[J]. *npj Biofilms and Microbiomes*, 2023, 9: 67.
- [25] Dance A. The search for microbial dark matter[J]. *Nature*, 2020, 582(7811): 301-303.
- [26] Gruber-Vodicka HR, Leisch N, Kleiner M, Hinzke T, Liebeke M, McFall-Ngai M, Hadfield MG, Dubilier N. Two intracellular and cell type-specific bacterial symbionts in the placozoan *Trichoplax H₂*[J]. *Nature Microbiology*, 2019, 4(9): 1465-1474.
- [27] Wiegand S, Jogler M, Boedeker C, Pinto D, Vollmers J, Rivas-Marín E, Kohn T, Peeters SH, Heuer A, Rast P, Oberbeckmann S, Bunk B, Jeske O, Meyerdierks A, Storesund JE, Kallscheuer N, Lückner S, Lage OM, Pohl T, Merkel BJ, et al. Cultivation and functional characterization of 79 *Planctomycetes* uncovers their unique biology[J]. *Nature Microbiology*, 2020, 5(1): 126-140.
- [28] Ren Z, Luo W, Li HR, Ding HT, Zhang YL. Decrypting the phylogeny and metabolism of microbial dark matter in green and red Antarctic snow[J]. *ISME Communications*, 2025, 5(1): ycaf003.
- [29] Hawley AK, Nobu MK, Wright JJ, Durno WE, Morgan-Lang C, Sage B, Schwientek P, Swan BK, Rinke C, Torres-Beltrán M, Mewis K, Liu WT, Stepanauskas R, Woyke T, Hallam SJ. Diverse Marinimicrobia bacteria may mediate coupled biogeochemical cycles along ecothermodynamic gradients[J]. *Nature Communications*, 2017, 8: 1507.
- [30] Peng SX, Gao SM, Lin ZL, Luo ZH, Zhang SY, Shu WS, Meng FG, Huang LN. Biogeography and ecological functions of underestimated CPR and DPANN in acid mine drainage sediments[J]. *mBio*, 2025, 16(6): e00705-25.
- [31] Tedersoo L, Magurno F, Alkahtani S, Mikryukov V. Phylogenetic classification of arbuscular mycorrhizal fungi: new species and higher-ranking taxa in *Glomeromycota* and *Mucoromycota* (class *Endogonomycetes*)[J]. *Mycology*, 2024, 107: 273-325.
- [32] Lobato C, Abdelfattah A, Berg G, Cernava T. Defining the cultured and uncultured bacterial fractions in Cannabis seeds[J]. *Environmental Microbiome*, 2025, 20: 68.
- [33] Cruz-Morales P, Kopp JF, Martínez-Guerrero C, Yáñez-Guerra LA, Selem-Mojica N, Ramos-Aboites H, Feldmann J, Barona-Gómez F. Phylogenomic analysis of natural products biosynthetic gene clusters allows discovery of arseno-organic metabolites in model *Streptomyces*[J]. *Genome Biology and Evolution*, 2016, 8(6): 1906-1916.
- [34] Wu SB, Zhou HN, Chen DL, Lu YT, Li YN, Qiao JJ. Multi-omic analysis tools for microbial metabolites prediction[J]. *Briefings in Bioinformatics*, 2024, 25(4): bbae264.
- [35] Danso D, Chow J, Streit WR. Plastics: environmental and biotechnological perspectives on microbial degradation[J]. *Applied and Environmental Microbiology*, 2019, 85(19): e01095-19.
- [36] Erdrich SH, Luthe T, Kever L, Badia Roigé B, Arsova B, Davoudi E, Frunzke J. Expanding the phage galaxy: isolation and characterization of five novel *Streptomyces* Siphoviruses ankus, byblos, DekoNeimoidia, mandalore, and naboo[J]. *Phage*, 2024, 5(3): 153-161.
- [37] Zhu QY, Dupont CL, Jones MB, Pham KM, Jiang ZD, DuPont HL, Highlander SK. Visualization-assisted binning of metagenome assemblies reveals potential new pathogenic profiles in idiopathic travelers’ diarrhea[J]. *Microbiome*, 2018, 6: 201.

- [38] Daisley BA, Reid G. BEEExact: a metataxonomic database tool for high-resolution inference of bee-associated microbial communities[J]. *mSystems*, 2021, 6(2): 10.1128/mSystems.00082-10.1128/mSystems.00021.
- [39] Williams PC, Wernke KM, Tirla A, Herzon SB. Employing chemical synthesis to study the structure and function of colibactin, a “dark matter” metabolite[J]. *Natural Product Reports*, 2020, 37(11): 1532-1548.
- [40] Piao HL, Froula J, Du CB, Kim TW, Hawley ER, Bauer S, Wang Z, Ivanova N, Clark DS, Klenk HP, Hess M. Identification of novel biomass-degrading enzymes from genomic dark matter: populating genomic sequence space with functional annotation[J]. *Biotechnology and Bioengineering*, 2014, 111(8): 1550-1565.
- [41] Pavlopoulos GA, Baltoumas FA, Liu SR, Selvitopi O, Camargo AP, Nayfach S, Azad A, Roux S, Call L, Ivanova NN, Chen IM, Paez-Espino D, Karatzas E, Iliopoulos I, Konstantinidis K, Tiedje JM, Pett-Ridge J, Baker D, Visel A, Ouzounis CA, et al. Unraveling the functional dark matter through global metagenomics[J]. *Nature*, 2023, 622(7983): 594-602.
- [42] Fesenko I, Sahakyan H, Dhyani R, Shabalina SA, Storz G, Koonin EV. The hidden bacterial microproteome[J]. *Molecular Cell*, 2025, 85(5): 1024-1041.e6.
- [43] Santos-Aberturas J, Vior N. Beyond soil-dwelling actinobacteria: fantastic antibiotics and where to find them[J]. *Antibiotics*, 2022, 11(2): 195.
- [44] Xu XR, Feng J, Zhang P, Fan J, Yin WB. A CRISPR/Cas9 cleavage system for capturing fungal secondary metabolite gene clusters[J]. *Journal of Microbiology and Biotechnology*, 2021, 31(1): 8-15.
- [45] Shmakov SA, Sitnik V, Makarova KS, Wolf YI, Severinov KV, Koonin EV. The CRISPR spacer space is dominated by sequences from species-specific mobilomes[J]. *mBio*, 2017, 8(5): e01397-17.
- [46] Wade JT, Grainger DC. Pervasive transcription: illuminating the dark matter of bacterial transcriptomes[J]. *Nature Reviews Microbiology*, 2014, 12(9): 647-653.
- [47] Lok C. Mining the microbial dark matter[J]. *Nature*, 2015, 522(7556): 270-273.
- [48] Lagkouvardos I, Overmann J, Clavel T. Cultured microbes represent a substantial fraction of the human and mouse gut microbiota[J]. *Gut Microbes*, 2017, 8(5): 493-503.
- [49] Probst AJ, Weinmaier T, Raymann K, Perras A, Emerson JB, Rattei T, Wanner G, Klingl A, Berg IA, Yoshinaga M, Viehweger B, Hinrichs KU, Thomas BC, Meck S, Auerbach AK, Heise M, Schintlmeister A, Schmid M, Wagner M, Gribaldo S, et al. Biology of a widespread uncultivated archaeon that contributes to carbon fixation in the subsurface[J]. *Nature Communications*, 2014, 5: 5497.
- [50] Wiegand S, Dam HT, Riba JL, Vollmers J, Kaster AK. Printing microbial dark matter: using single cell dispensing and genomics to investigate the *Patescibacteria*/candidate phyla radiation[J]. *Frontiers in Microbiology*, 2021, 12: 635506.
- [51] Amorin de Hegedüs R, Conesa A, Foster JS. Integration of multi-omics data to elucidate keystone unknown taxa within microbialite-forming ecosystems[J]. *Frontiers in Microbiology*, 2023, 14: 1174685.
- [52] Garcia Mendez DF, Egan S, Wist J, Holmes E, Sanabria J. Meta-analysis of the microbial diversity cultured in bioreactors simulating the gut microbiome[J]. *Microbial Ecology*, 2024, 87: 57.
- [53] Vincy A, Vishakha V, Cohort AP, Shankar M, Biswas I. *In situ* isolation and culturing of recalcitrant soil bacteria using an isolation chip (iChip)[J]. *Journal of Visualized Experiments*, 2025(222): e67692.
- [54] Jia YY, Li XY, Xu FJ, Liu ZF, Fu YL, Xu X, Yang JW, Zhang S, Shen CF. Single-cell-level microfluidics assisted with resuscitation-promoting factor technology (SMART) to isolate novel biphenyl-degrading bacteria from typical soils in Eastern China[J]. *Environmental Pollution*, 2022, 311: 119864.
- [55] Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. Recovery of nearly 8 000 metagenome-assembled genomes substantially expands the tree of life[J]. *Nature Microbiology*, 2017, 2(11): 1533-1542.
- [56] Wang Z, Liang JL, Huang LN, Mengoni A, Shu WS. Metagenomic assembly: reconstructing genomes from metagenomes[M]//*Bacterial Pangenomics: Methods and Protocols*. New York, NY: Springer US, 2021: 139-152.
- [57] Vollmers J, Wiegand S, Lenk F, Kaster AK. How clear is our current view on microbial dark matter (Re-) assessing public MAG & SAG datasets with MDMcleaner[J]. *Nucleic Acids Research*, 2022, 50(13): e76.
- [58] Bellali S, Lagier JC, Million M, Anani H, Haddad G, Francis R, Kuete Yimagou E, Khelaifia S, Levasseur A, Raoult D, Bou Khalil J. Running after ghosts: are dead bacteria the dark matter of the human gut microbiota[J]. *Gut Microbes*, 2021, 13: 1897208.
- [59] Parks DH, Chuvochina M, Chaumeil PA, Rinke C, Mussig AJ, Hugenholtz P. A complete domain-to-species taxonomy for Bacteria and Archaea[J]. *Nature Biotechnology*, 2020, 38(9): 1079-1086.
- [60] 张道锋, 焦建宇, 李文均. ICNP与SeqCode: 组学时代原核生物命名面临的机遇与挑战[J]. *微生物学报*, 2024, 64(7): 2151-2171.
- Zhang DF, Jiao JY, Li WJ. ICNP and SeqCode: opportunities and challenges of prokaryotic nomenclature in the omics era[J]. *Acta Microbiologica Sinica*, 2024, 64(7): 2151-2171 (in Chinese).
- [61] Zamkovaya T, Foster JS, de Crécy-Lagard V, Conesa A. A network approach to elucidate and prioritize microbial dark matter in microbial communities[J]. *The ISME Journal*, 2021, 15(1): 228-244.
- [62] Sedlazeck FJ, Lee H, Darby CA, Schatz MC. Piercing the dark matter: bioinformatics of long-range sequencing and mapping[J]. *Nature Reviews Genetics*, 2018, 19(6): 329-346.
- [63] Schmid M, Frei D, Patrignani A, Schlapbach R, Frey JE, Remus-Emsermann MNP, Ahrens CH. Pushing the limits of *de novo* genome assembly for complex prokaryotic

- genomes harboring very long, near identical repeats[J]. *Nucleic Acids Research*, 2018, 46(17): 8953-8965.
- [64] Kato S, Masuda S, Shibata A, Shirasu K, Ohkuma M. Insights into ecological roles of uncultivated bacteria in Katase hot spring sediment from long-read metagenomics[J]. *Frontiers in Microbiology*, 2022, 13: 1045931.
- [65] Kioukis A, Camargo AP, Pavlidis P, Iliopoulos I, Kyrpides NC, Lagkouvardos I. Global archaeal diversity revealed through massive data integration: uncovering just tip of iceberg[J]. *Microorganisms*, 2025, 13(3): 598.
- [66] Becraft ED, Dodsworth JA, Murugapiran SK, Ohlsson JI, Briggs BR, Kanbar J, De Vlaminck I, Quake SR, Dong HL, Hedlund BP, Swingley WD. Single-cell-genomics-facilitated read binning of candidate phylum EM19 genomes from geothermal spring metagenomes[J]. *Applied and Environmental Microbiology*, 2016, 82(4): 992-1003.
- [67] Liu YC, Gao YS, Niu R, Zhang ZY, Lu GW, Hu HF, Liu TG, Cheng ZZ. Rapid and accurate bacteria identification through deep-learning-based two-dimensional Raman spectroscopy[J]. *Analytica Chimica Acta*, 2024, 1332: 343376.
- [68] Ho CS, Jean N, Hogan CA, Blackmon L, Jeffrey SS, Holodniy M, Banaei N, Saleh AAE, Ermon S, Dionne J. Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning[J]. *Nature Communications*, 2019, 10: 4927.
- [69] Jing XY, Gong YH, Pan HH, Meng Y, Ren YS, Diao ZD, Mu RZ, Xu T, Zhang J, Ji YT, Li YD, Wang C, Qu LY, Cui L, Ma B, Xu J. Single-cell Raman-activated sorting and cultivation (scRACS-Culture) for assessing and mining *in situ* phosphate-solubilizing microbes from nature[J]. *ISME Communications*, 2022, 2: 106.
- [70] Jing XY, Gou HL, Gong YH, Su XL, Xu L, Ji YT, Song YZ, Thompson IP, Xu J, Huang WE. Raman-activated cell sorting and metagenomic sequencing revealing carbon-fixing bacteria in the ocean[J]. *Environmental Microbiology*, 2018, 20(6): 2241-2255.
- [71] Xu L, Yue XL, Li HZ, Jian SL, Shu WS, Cui L, Xu XW. Aerobic anoxygenic phototrophic bacteria in the marine environments revealed by Raman/fluorescence-guided single-cell sorting and targeted metagenomics[J]. *Environmental Science & Technology*, 2024, 58(16): 7087-7098.
- [72] Song YZ, Kaster AK, Vollmers J, Song YQ, Davison PA, Frentrup M, Preston GM, Thompson IP, Murrell JC, Yin HB, Hunter CN, Huang WE. Single-cell genomics based on Raman sorting reveals novel carotenoid-containing bacteria in the Red Sea[J]. *Microbial Biotechnology*, 2017, 10(1): 125-137.
- [73] Zhou HB, Yang DT, Ivleva NP, Mircescu NE, Schubert S, Niessner R, Wieser A, Haisch C. Label-free *in situ* discrimination of live and dead bacteria by surface-enhanced Raman scattering[J]. *Analytical Chemistry*, 2015, 87(13): 6553-6561.
- [74] Li M, Raza M, Song S, Hou LW, Zhang ZF, Gao M, Huang JE, Liu F, Cai L. Application of culturomics in fungal isolation from mangrove sediments[J]. *Microbiome*, 2023, 11: 272.
- [75] Kaeberlein T, Lewis K, Epstein SS. Isolating “uncultivable” microorganisms in pure culture in a simulated natural environment[J]. *Science*, 2002, 296(5570): 1127-1129.
- [76] Tellatin D, Cornet L, Snauwaert V, Compère P, Ongena M, Quinton L, Stulanovic N, Ribeiro Monteiro S, Rigolet A, Burguet P, van Damme P, Carro L, Rigali S. *Melissospora conviva* gen. nov., sp. nov., a novel actinobacterial genus isolated from beehive through cross-feeding interactions[J]. *International Journal of Systematic and Evolutionary Microbiology*, 2025, 75(8): 006868.
- [77] Tan JY, Wang SD, Dick GJ, Young VB, Sherman DH, Burns MA, Lin XN. Co-cultivation of microbial sub-communities in microfluidic droplets facilitates high-resolution genomic dissection of microbial ‘dark matter’[J]. *Integrative Biology*, 2020, 12(11): 263-274.
- [78] Culp EJ, Goodman AL. Cross-feeding in the gut microbiome: ecology and mechanisms[J]. *Cell Host & Microbe*, 2023, 31(4): 485-499.
- [79] Ghassemi Nedjad C, Bolteau M, Bourneuf L, Paulevé L, Frioux C. Seed2LP: seed inference in metabolic networks for reverse ecology applications[J]. *Bioinformatics*, 2025, 41(4): btaf140.
- [80] Lin LY, Song J, Li J, Zuo XL, Wei H, Yang CY, Wang W. Imaging the *in vivo* growth patterns of bacteria in human gut microbiota[J]. *Gut Microbes*, 2021, 13: 1960134.
- [81] Zhou YJ, Lin LY, Wang W. An integrative approach for imaging and quantitative analysis of gut microbiota growth *in vivo* using fluorescent D-amino acid labeling and fluorescence *in situ* hybridization[J]. *Biophysics Reports*, 2025, 11(3): 172.
- [82] Lin LY, Wu QY, Song J, Du YH, Gao J, Song YL, Wang W, Yang CY. Revealing the *in vivo* growth and division patterns of mouse gut bacteria[J]. *Science Advances*, 2020, 6(36): eabb2531.
- [83] Lackner G, Peters EE, Helfrich EJM, Piel J. Insights into the lifestyle of uncultured bacterial natural product factories associated with marine sponges[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2017, 114(3): E347-E356.
- [84] Hoarfrost A, Aptekmann A, Farfaniuk G, Bromberg Y. Deep learning of a bacterial and archaeal universal language of life enables transfer learning and illuminates microbial dark matter[J]. *Nature Communications*, 2022, 13: 2606.
- [85] Wacholder A, Parikh SB, Coelho NC, Acar O, Houghton C, Chou L, Carvunis AR. A vast evolutionarily transient translome contributes to phenotype and fitness[J]. *Cell Systems*, 2023, 14(5): 363-381.e8.
- [86] Jordan B, Weidenbach K, Schmitz RA. The power of the small: the underestimated role of small proteins in bacterial and archaeal physiology[J]. *Current Opinion in Microbiology*, 2023, 76: 102384.

- [87] Cao YZ, Xu MW, Chen Q, Wu DH, Lu J, Cai GL. Potential nutritional and functional matters in yeast culture prepared by soybean meal fermentation[J]. *Journal of the Science of Food and Agriculture*, 2024, 104(14): 8869-8878.
- [88] Xu M, Wang W, Wright GD. Glycopeptide antibiotic discovery in the genomic era[M]//*Chemical Microbiology Part B*. Amsterdam: Elsevier, 2022: 325-346.
- [89] Gurevich A, Mikheenko A, Shlemov A, Korobeynikov A, Mohimani H, Pevzner PA. Increased diversity of peptidic natural products revealed by modification-tolerant database search of mass spectra[J]. *Nature Microbiology*, 2018, 3(3): 319-327.
- [90] Tyc, Song CX, Dickschat JS, Vos M, Garbeva P. The ecological role of volatile and soluble secondary metabolites produced by soil bacteria[J]. *Trends in Microbiology*, 2017, 25(4): 280-292.
- [91] Brakhage AA, Schroeckh V. Fungal secondary metabolites: strategies to activate silent gene clusters[J]. *Fungal Genetics and Biology*, 2011, 48(1): 15-22.
- [92] Rust M, Helfrich EJM, Freeman MF, Nanudorn P, Field CM, Rückert C, Kündig T, Page MJ, Webb VL, Kalinowski J, Sunagawa S, Piel J. A multiproducer microbiome generates chemical diversity in the marine sponge *Mycale hentscheli*[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2020, 117(17): 9508-9518.
- [93] Ziko L, AbdelRaheem O, Nabil M, Aziz RK, Siam R. Bioprospecting the microbiome of Red Sea *Atlantis II* brine pool for peptidases and biosynthetic genes with promising antibacterial activity[J]. *Microbial Cell Factories*, 2022, 21: 109.
- [94] Ma B, Lu CY, Wang YL, Yu JW, Zhao KK, Xue R, Ren H, Lv XF, Pan RH, Zhang JB, Zhu YG, Xu JM. A genomic catalogue of soil microbiomes boosts mining of biodiversity and genetic resources[J]. *Nature Communications*, 2023, 14: 7318.
- [95] Sussfeld D, Lannes R, Corel E, Bernard G, Martin P, Baptiste E, Pelletier E, Lopez P. New groups of highly divergent proteins in families as old as cellular life with important biological functions in the ocean[J]. *Environmental Microbiome*, 2025, 20: 65.
- [96] Bhushan A, Egli PJ, Peters EE, Freeman MF, Piel J. Genome mining- and synthetic biology-enabled production of hypermodified peptides[J]. *Nature Chemistry*, 2019, 11(10): 931-939.
- [97] Kuroda Y, Nicacio KJ, da Silva-Jr IA, Leger PR, Chang S, Gubiani JR, DeFlon VM, Nagashima N, Rode A, Blackford K, Ferreira AG, Sette LD, Williams DE, Andersen RJ, Jancar S, Berlinck RGS, Sarpong R. Isolation, synthesis and bioactivity studies of phomactin terpenoids[J]. *Nature Chemistry*, 2018, 10(9): 938-945.
- [98] Zhong J, Yan XY, Zuo X, Zhao X, Yang JH, Dou Q, Peng LL, Zhu YX, Xiao YC, Bian ZR, He DL, Xu QS, Wright S, Li YY, Du LC, Wang Y, Yuan J. Developing a new treatment for superficial fungal infection using antifungal Collagen-HSAF dressing[J]. *Bioengineering & Translational Medicine*, 2022, 7(3): e10304.
- [99] Rubio A, Sprang M, Garzón A, Moreno-Rodríguez A, Pachón-Ibáñez ME, Pachón J, Andrade-Navarro MA, Pérez-Pulido AJ. Analysis of bacterial pangenomes reduces CRISPR dark matter and reveals strong association between membranome and CRISPR-Cas systems[J]. *Science Advances*, 2023, 9(12): eadd8911.
- [100] Vassallo CN, Doering CR, Littlehale ML, Teodoro GIC, Laub MT. A functional selection reveals previously undetected anti-phage defence systems in the *E. coli* pangenome[J]. *Nature Microbiology*, 2022, 7(10): 1568-1579.
- [101] Dornenburg JE, DeVita AM, Palumbo MJ, Wade JT. Widespread antisense transcription in *Escherichia coli*[J]. *mBio*, 2010, 1: e00024-10.
- [102] Racki LR, Freddolino L. Polyphosphate: the “dark matter” of bacterial chromatin structure[J]. *Molecular Microbiology*, 2025, 123(3): 279-293.
- [103] Sobkowiak A, Schwierzeck V, van Almsick V, Scherff N, Schuler F, Bessonov K, Robertson J, Harmsen D, Mellmann A. The dark matter of bacterial genomic surveillance: antimicrobial resistance plasmid transmissions in the hospital setting[J]. *Journal of Clinical Microbiology*, 2025, 63(6): e00121-25.
- [104] Bernard G, Pathmanathan JS, Lannes R, Lopez P, Baptiste E. Microbial dark matter investigations: how microbial studies transform biological knowledge and empirically sketch a logic of scientific discovery[J]. *Genome Biology and Evolution*, 2018, 10(3): 707-715.