

从盲目拼图到智能复原：AI照亮生命的未知版图

作者：writer 来源：科学网

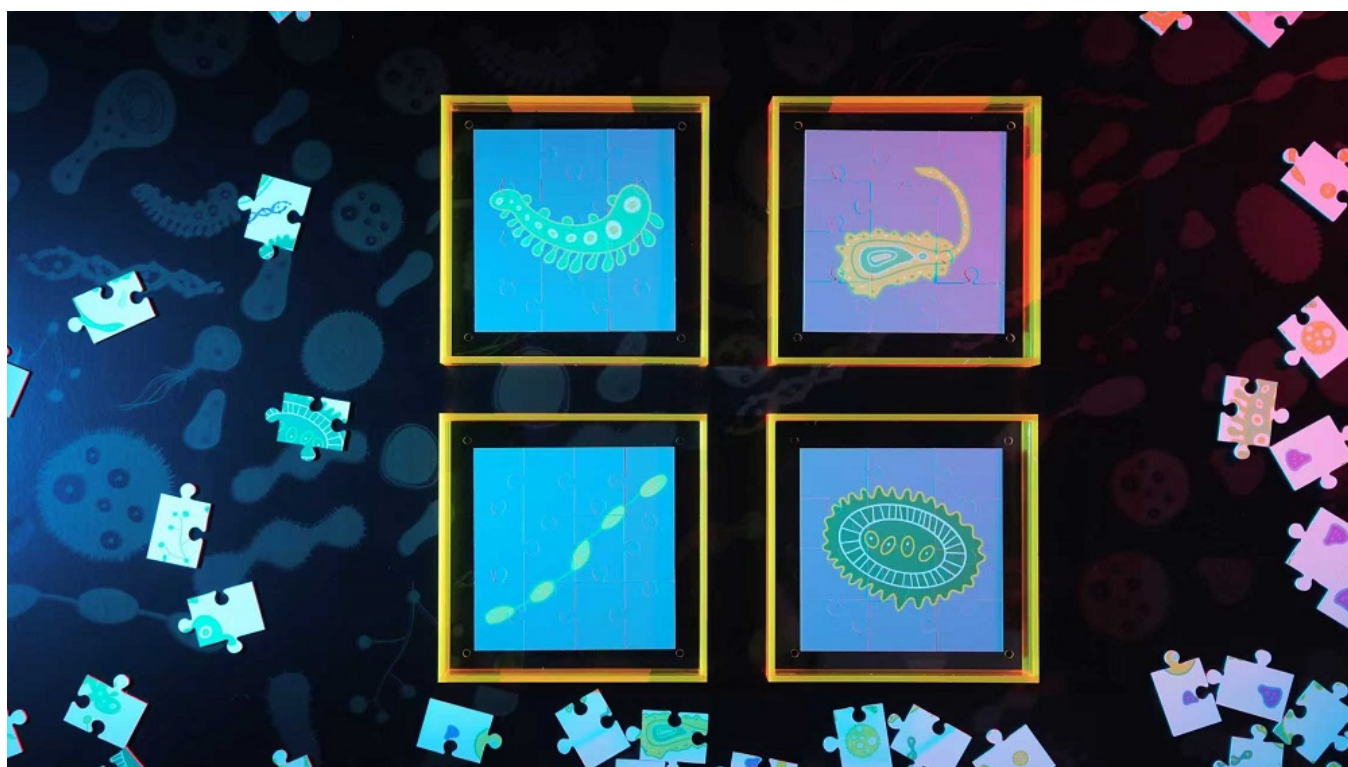
本文原地址：<https://www.iikx.com/news/progress/36831.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

从盲目拼图到智能复原：AI照亮生命的未知版图

。一克土、一片叶、一滴水，均隐藏着人类认知不足1%的微生物世界。

想象一下，如果一种微生物的基因组是一幅拼图，将成千上万种微生物的基因组拼图拆成碎片，并混在一起，而你要在没有任何参考图纸的情况下，将它们准确复原——这就是微生物宏基因组分箱技术面临的终极挑战。



该概念图展示了LorBin如何将宏基因组测序的长读数聚类为高质量的宏基因组组装（MAG）。黑色背景代表复杂的微生物栖息地，框内散布的拼图碎片象征长读数，而框架内由碎片构建出的微生物图案则象征LorBin成功地将这些散落的基因组序列复原为完整的基因组。受访者供图

近日，南京农业大学教授、中国工程院院士沈其荣团队开发的LorBin算法成功解决了三代宏基因组分箱的共性关键技术难题。他们正试图让这些“隐形居民”首次拥有清晰的基因身份证。《自

然—通讯》在线发表了这一成果。

地球生命的未知版图

微生物看不见、摸不着，却与人类生活紧密相连。论文共同通讯作者、南京农业大学教授韦中向记者展示了一份令人震撼的数据：一克土壤中活跃着数以亿计、种类复杂的微生物。

“微生物组被称为地球生命的‘暗物质’，是因为目前仍有超过99%的微生物无法在实验室条件下分离培养。”韦中说。

在医学领域，微生物基因组信息可用于鉴定疾病相关的生物标志物，为疾病早期诊断和精准治疗提供新依据；在农业领域，通过解析作物根际微生物基因组，可以挖掘出能够促进作物生长的有益菌株；在环境修复领域，基因组数据可用于筛选具有高效污染物降解能力、参与生物地球化学循环的功能微生物，为污染治理与生态恢复提供生物技术支撑。

通过破解微生物基因组的信息，人们不仅能够识别环境中已知与未知的微生物种类，还可以直接从基因序列预测其潜在的生理功能、代谢通路，以及微生物与宿主或环境间的相互作用机制。

“微生物研究将从描述性的物种组成分析，迈入功能机制探索的阶段。”韦中说。

宏基因组测序技术成为破解这一难题的关键途径。特别是三代测序技术的发展，让碱基序列的检测长度更长、质量更高。“短读长测序如同将一本书撕成无数碎片，难以拼回原貌；而长读长技术则能直接获取完整的‘段落’甚至‘章节’。”论文共同通讯作者、南京农业大学资源与环境学院副教授江高飞解释道。

然而，技术突破背后隐藏着严峻的瓶颈问题。江高飞指出，三代宏基因组测序分析面临原始数据量庞大、碱基序列错误率高，以及分析算法不完善等挑战。“尤其是在分箱重构高质量单菌基因组这一关键环节，传统工具如同为超级显微镜配了个老花镜，造成了信息资源的巨大浪费。”

论文共同第一作者、南京农业大学智慧农业学院（人工智能学院）副教授薛卫解释道，所谓“宏基因组分箱”的核心目标是从复杂微生物群落中重构出高质量的单菌的基因组。这个过程需要将测序得到的碱基序列进行组装，形成更长的序列片段，随后将这些片段分配到若干个“箱”中。在理想情况下，每个箱对应一个微生物的基因组。

“看似简单，实则挑战巨大。”韦中说，“既要克服复杂微生物群落中物种组成与丰度的不确定性，又要有效区分高度相似近缘物种，甚至菌株水平的遗传变异规律。”

为微生物基因组拼图装上“AI眼睛”

面对这些挑战，资源与环境科学学院LorMe实验室与智慧农业学院（人工智能学院）BioAI实验室展开了跨学科合作。薛卫形象地描述了他们的创新思路：“传统方法就像是在黑暗中盲目拼图，而我们要做的是给这个过程装上‘AI眼睛’。”

LorBin的核心创新在于深度融合了多项前沿AI技术：生成式AI、无监督学习、深度学习和注意力机制等。薛卫进一步解释道：“这项研究中主要有四大技术突破，对应分箱过程中的每个关键难点。这使得LorBin的内核拥有一个能够自主学习和智能决策的‘AI大脑’。”

第一项创新是适配变分自编码器。薛卫打了个比方：“这就像教AI识别带有方言的普通话——它能够自动过滤三代测序数据中的‘噪声’，提取出稳健的、可用于区分不同微生物的深层序列特征。彻底摆脱了对人工设计特征的依赖，实现了对数据本质更智能的解读。”

第二项突破是两阶段多尺度自适应迭代聚类算法。韦中用显微镜对焦来比喻，首先用“低倍镜”快速锁定大目标，分离出高丰度物种；然后再自适应的调整，用“高倍镜”精细挖掘那些数量稀少但重要的稀有微生物。这显著提升了新物种的发现能力。

在质量评估环节，团队引入了单拷贝基因集等信息引导的循环分箱质量评估模型。“这相当于在装配线上设置了质检员，对每个‘产品’进行实时检测，确保基因组不被混淆或错误分裂。”江高飞说。

令人耳目一新的是第四项创新：采用Transformer模型评估高维数据的空间分布特征。“不同微生物的数据在空间中呈现不同形态，有的像苹果般饱满，有的则像月牙般有缺口。理解这种凹凸性直接关系到分类算法的选择。”论文共同第一作者、已毕业研究生刘佐解释说。

“这些技术共同构成一个完整解决方案，使其在重构基因组的完整度、纯度及发现新物种方面均实现显著提升。”江高飞说。

基准测试结果令人振奋。LorBin整体优于以往6种高性能深度学习算法，重构的高质量单菌基因组比现有工具多15-189%，意味着从相同样本中能挖掘出成倍增加的完整、纯净的微生物基因组，极大地扩展了可研究的微生物。

而LorBin识别特有物种数量更是其他工具的2.4~17倍，直接体现了其在探索微生物“暗物质”上的较大优势。这些提升为功能基因挖掘、微生物生态机制解析等领域打开了新窗口，有望推动微生物组学研究进入“看得更多、看得更清、发现更新”的阶段。

薛卫强调：“LorBin具有良好的硬件适配性，只需配备常规GPU的Linux计算机就能顺畅运行，极大降低了使用门槛。”

从实验室“照进”现实

薛卫向《中国科学报》特别强调了这一工具的开源价值：“我们已将LorBin的代码在GitHub平台全面公开，并结合多种工具和模块搭建了智能分析流程。”

这种开源共享的理念，正是团队跨学科合作精神的延伸。回顾研发历程，薛卫感慨道：“LorBin的诞生，源于LorMe实验室与BioAI实验室的‘双向奔赴’。这并非简单的技术嫁接，而是一次生物学与人工智能的深度对话。”

BioAI实验室如同一位“算法建筑师”，手握深度学习工具，致力于构建精妙的计算模型；而LorMe实验室则像是“微生物向导”，凭借对微观世界的深刻洞察，确保每一次模型迭代都直指真实的生物学难题。

合作中最难忘的，并非技术瓶颈，而是跨越那道无形的“语言巴别塔”。对BioAI而言，挑战在于如何将复杂的生命现象“翻译”成机器能理解的数学语言；对LorMe而言，则需要解读每一个算法决策背后所隐藏的生物学意义。

不过，LorBin的突破不仅体现在技术指标上。

在医学领域，LorBin展现出强大的病原体检测能力。团队利用该工具对104个肠道样本进行分析，成功复原了3194个中高质量的微生物基因组，从中识别出477个潜在的病原菌物种。

“例如肺炎克雷伯菌，由于长期暴露于抗生素选择压力下，进化出强大的抗性基因。而像Leclercia adecarboxylata这种新发现的致病菌，虽然本身破坏力不强，却能携带并传播高风险抗性基因。”薛卫表示，这种精准识别能力为临床防控提供了关键线索。

在农业领域，LorBin正在改变传统的微生物组研究范式。韦中指出：“过去我们只知道‘土壤里有好细菌’，而现在能精准锁定那些负责固氮、解磷、促进作物生长的‘明星菌株’。”

这一突破为农业微生物肥料研发提供了新思路。通过解析作物根际微生物基因组，研究人员可以挖掘出能够促进作物生长的有益菌株，实现在不改变作物基因的前提下提高产量。

环境监测是LorBin的另一重要应用场景。该工具能够完整描绘污染物被微生物降解的路径图，并能准确识别环境中耐药基因的携带者及其传播网络。

“这对于评估环境风险、阻断耐药性扩散提供了关键情报。”江高飞说。

LorBin推动了微生物组研究从“观测时代”向“发掘与创造”的转变。“它使我们能够窥见此前被遮蔽的、数量稀少但功能关键的低丰度微生物，这些微生物往往是驱动整个生态系统运转的‘核心引擎’。”韦中说。

相关论文信息：<https://doi.org/10.1038/s41467-025-64916-8>

作者：李晨 来源：中国科学报

更多科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发