

## 特约评述

DOI: 10.12211/2096-8280.2021-074

## 功能性菌群构建的研究进展

黄佳城<sup>1,2</sup>, 张瑗琿<sup>1,2</sup>, 付友思<sup>1,2</sup>, 方柏山<sup>1,2</sup><sup>(1)</sup> 厦门大学化学化工学院, 福建 厦门 361005; <sup>(2)</sup> 厦门市合成生物技术重点实验室, 福建 厦门 361005)

**摘要:** 功能性菌群构建作为一个新兴的研究方向, 随着合成生物学、微生物组学技术的发展, 逐渐成为研究热点。本文将从以下4个方面介绍功能性菌群的研究进展。第一, 功能性菌群研究的初衷及其相对于单一生物体工程的优势和设计难点; 第二, 功能性菌群研究中自下而上 (bottom-up) 和自上而下 (top-down) 的设计策略; 第三, 功能性菌群的分析工具, 包括“宏组学”和“多组学联用”手段以及相关的数据处理流程和软件。第四, 分别从设计策略和分析工具方面出发, 总结了功能性菌群构建过程中的主要挑战, 并展望了未来以“智能设计”为核心的发展方向: ①利用可解释的时空数据模型解析区域范围内功能性菌群的时空变化关系; ②结合神经网络与多模态学习方法建立多组学群落分析流程; ③通过强化学习设计功能性菌群内分布式代谢回路。

**关键词:** 功能性菌群; 自下而上; 自上而下; 宏组学; 多组学; 智能设计

**中图分类号:** Q 816 **文献标志码:** A

## Research progress in construction of functional microbial communities

HUANG Jiacheng<sup>1,2</sup>, ZHANG Aihui<sup>1,2</sup>, FU Yousi<sup>1,2</sup>, FANG Baishan<sup>1,2</sup><sup>(1)</sup> College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, Fujian, China; <sup>(2)</sup> The Key Laboratory for Synthetic Biotechnology of Xiamen City, Xiamen 361005, Fujian, China)

**Abstract:** Study on functional microbial communities has become a hotspot with development of synthetic biology and microbiome. Comparing to single organism, functional microbial communities have several advantages in their robustness when facing environmental interference and for higher yields when producing complex products. This article introduces the research progress in studies on functional microbial communities from aspects: 1) the original motivation of functional microbial community research and the advantages and difficulties in design compared with single organism engineering, 2) the “bottom-up” and “top-down” strategies in the functional microbial community design process, and 3) analysis tools for functional microbial communities, such as “metagenomics”, “multi-omics” and related data processing procedures and software. The main history of environmental microbial analysis is reviewed, and the main concepts of meta-omics is introduced. The major software which is used to process meta-omics data is also commented. In addition, challenges for the functional microbial community construction are highlighted based on

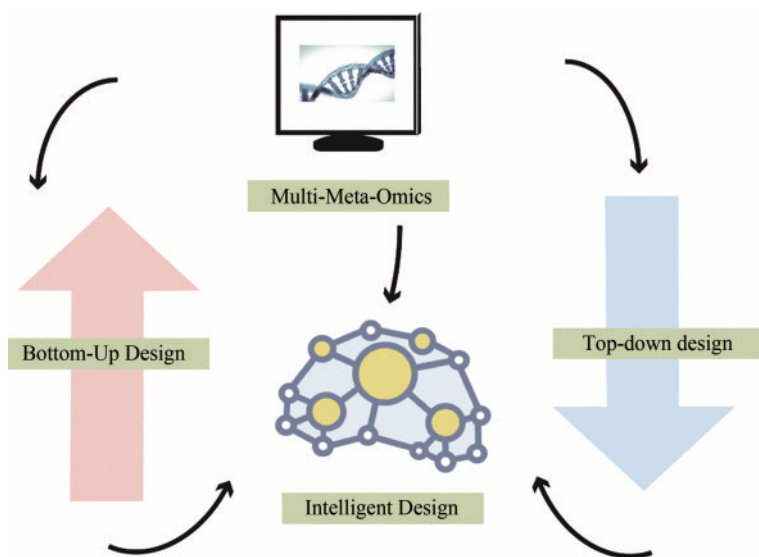
收稿日期: 2021-07-12 修回日期: 2021-11-25

基金项目: 国家自然科学基金 (21978245); 博士后创新人才支持计划 (BX20200197)

引用本文: 黄佳城, 张瑗琿, 付友思, 方柏山. 功能性菌群构建的研究进展[J]. 合成生物学, 2022, 3(1): 155-167

Citation: HUANG Jiacheng, ZHANG Aihui, FU Yousi, FANG Baishan. Research progress in construction of functional microbial communities [J]. Synthetic Biology Journal, 2022, 3(1): 155-167

the design strategy and analysis tools. The “bottom-up” strategy is not suitable for constructing a complicated microbial community, while the biosafety needs to be considered for developing the “top-down” strategy. Finally, the development of microbial communities with “intelligent design” as the core is prospected: first, using an interpretable spatiotemporal data model for temporal-spatial relationship mining of functional microbial communities; second, combining neural network and multi-modal learning methods to establish a multi-omics community analysis process; third, computational design of distributed metabolic circuits in the functional microbial community through reinforcement learning.



**Keywords:** functional microbial community; bottom-up; top-down; meta-omics; multi-omics; intelligent design

## 1 功能性菌群

经历了20年的不断发展，合成生物学逐渐成为了一个综合学科融合之所，一个交叉学科碰撞之处。研究人员创建了形形色色的“微生物细胞工厂”，不仅实现了许许多多化学品的绿色制造，而且开发了生物修复、生物材料合成等多种功能<sup>[1]</sup>。其中，美国加州大学伯克利分校（UCB）化学工程系 Keasling 课题组<sup>[2]</sup>改造大肠杆菌和酿酒酵母合成青蒿素前体是合成生物学最经典的里程碑案例之一。他们历经20年，研究过程涉及载体构建、酶异源表达、底盘细胞开发、启动子与核糖体结合位点高通量筛选、通量平衡分析和转录后调控等，最终实现了青蒿素的半合成。该项研究成果在国际上产生重大影响，被 *Discovery* 杂志评为当年的十大科学进展之一。

然而，单一生物体工程在复杂化合物合成上面临诸多矛盾：第一，复杂目标产物合成所需的大容量外源基因与底盘细胞接收能力的矛盾<sup>[3]</sup>；第二，复杂目标产物合成途径对细胞中资源需求与细胞本体生长代谢的矛盾<sup>[4-5]</sup>；第三，单一生物体目标产物合成过程所需的稳定微环境与环境扰动的矛盾（例如环境变化或产物/副产物胁迫）<sup>[6-7]</sup>。为了解决上述产物合成和生产中单一生物体工程面临的矛盾，研究人员正在开发动态微生物群落中多物种工程改造平台替代单一生物体工程：从最初的菌株代谢通路的模块化构建的“单兵作战”快速发展为合成微生物群落共同协作的“联动共赢”来面对复杂的产物合成难题<sup>[8-10]</sup>。

微生物群落一般由多种微生物组成，相比于单一菌株，这些微生物组成的群体具备多种优势<sup>[8]</sup>：第一，群体中成员的分工协作减轻了外源载体的容量；第二，分工和资源交换减轻群落中

个体的代谢负担；第三，较高的鲁棒性减轻了环境变化的胁迫压力。因此，相对于单一菌株，整个群落成员的协作使其具有更强的代谢能力，同时群落中物种之间的信息交流使其具有更强的稳定性<sup>[9-11]</sup>。正是由于这些特性，微生物群落在复杂基质和污染物的降解<sup>[12-18]</sup>，药物、生物燃料和蛋白质复合物的生物法生产<sup>[13, 19-24]</sup>，功能化生物材料的制备<sup>[18, 25-27]</sup>，分布式逻辑计算/存储器<sup>[28-33]</sup>、生物传感器的构建<sup>[34-35]</sup>等方面具有巨大的潜力。

但是微生物群落平台的发展带来诸多优势的同时，也提升了工作量和操作难度<sup>[8]</sup>。幸运的是，宏组学（meta-omics）的发展为我们带来了微生物群落的大数据<sup>[36]</sup>，日益成熟的人工智能理论和显著提高了其处理和分析大数据的能力<sup>[37]</sup>。通过引入人工智能技术有望从大数据中探究客观规律，并利用这些规律快速寻找特定问题的最佳解决方案。

本文详细介绍了功能性菌群“自下而上(bottom-up)”和“自上而下(top-down)”的两种设计策略，以及功能性菌群的分析工具——“宏组学”的发展现状，总结了功能性菌群构建过程中面临的挑战，并展望了以“智能设计”为核心的发展方向。

## 2 功能性菌群的设计策略

在构建功能性菌群的探索中，研究人员提出两种不同的研究策略：分别是“自下而上”和“自上而下”。这两种研究策略对建立功能性菌群的研究提出不同的见解<sup>[38]</sup>。

### 2.1 “自下而上”的设计策略

“自下而上”的研究策略一般是通过获取微生物组中单个成员的基因组来重建它们的代谢网络，并使用建模或网络分析工具来指导设计微生物群落，使群落具有一定的功能（图1）。这是合成生物学家所关注和努力的方向。该研究策略可以分为以下两个部分：合成微生物群落工具的开发；利用工具促使群落目标功能的实现。

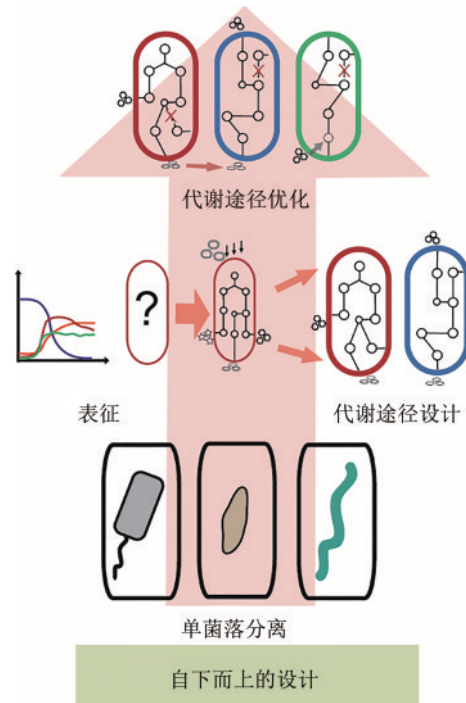


图1 使用“自下而上”策略构建功能菌群

Fig. 1 “Bottom-up” strategy for constructing functional microbial communities

#### 2.1.1 构建合成微生物群落工具的开发

微生物群落具有“群落内部通信”、“感受外界环境变化”以及“交换群落内资源”的功能。为了模拟以上三方面的功能，研究人员建立了“群体感应（quorum sensing, QS）”“诱导性元件”“共营养化”3种方法来开发适用于合成微生物群落的工具。

群体感应<sup>[39]</sup>是一种生物通信系统：在原核物种中，细胞产生的自体诱导物分子能够作为种群感知群落密度的一个信号，随着细胞群的增长，细胞会根据自体诱导剂浓度来调节基因表达以控制群体水平的行为。然而直接利用天然的群体感应系统作为合成微生物群落的工具会存在信号串扰的问题，因此通过对现有QS系统的改造，研究人员建立了一系列具有正交性的QS系统。加州大学旧金山分校（UCSF）的Hasty课题组通过启动子和蛋白质修饰，将Rpa和Tra群体感应系统和Lux和Las系统进行整合，构建具有完全正交性的系统<sup>[40]</sup>，并提出可诱导群体感应（inducible quorum sensing system, iQS）的概念<sup>[41]</sup>。中科院深圳先进技术研究院娄春波课题组<sup>[42]</sup>通过对元件挖掘、理

性设计以及定向进化等手段,开发了10套全新的群体感应系统,形成了一整套具有通用性高、正交性强、可以跨生物界通信的合成生物学工具箱。

诱导性元件使得群落中的成员能够感应环境条件并做出一系列的反应。有的研究通过外源添加分子来改变环境条件进而控制群落中成员的基因表达。麻省理工学院(MIT)的Collins课题组<sup>[43]</sup>在*Lactococcus lactis* NZ9000中构建不同诱导物诱导的回路,通过控制诱导物的添加模拟群落的社交关系;同时环境条件的改变也包括了调整营养浓度和培养条件来改变种群特异性基因表达<sup>[44]</sup>。

群落成员间相互依赖的功能性菌群可以通过共养相互作用的原理进行设计,使得群落中的某一生物体依赖其他群落成员所产生的代谢物,进而构建成员更多、鲁棒性更高的功能性菌群。加州大学圣地亚哥分校(UCSD)的Zengler课题组<sup>[45]</sup>系统性地描述如何通过营养缺陷型策略构建复杂群落的原理。根据这一原理,哥伦比亚大学的Wang课题组<sup>[46]</sup>构建了14种不同缺陷型大肠杆菌群落,它们之间相互依赖、共同生存。

上述这三方面的先驱性的工作可以帮助我们构建正交性更好的“群体感应”系统、调控更为精准的“诱导性元件”和更为复杂的“共营养化”策略,有利于我们建立具有高信噪比的菌种通信,完成高精度的群落内基因表达和构建高鲁棒性的功能性菌群,也为群落行为的实现提供了可靠的工具。

### 2.1.2 利用工具使得群落表现出目标功能

在群落行为的设计中,合成生物学家们利用2.1.1所述的工具驱动微生物群落表现的3种最具代表的行为“菌群的数量控制”、“分布式代谢”和“菌群的空间编程”。

在群落的代谢过程中,某些特定代谢物的相对产量可能会在代谢流层面上影响最终产物的产出或者是对底盘生物造成毒性<sup>[47]</sup>,而通过“控制微生物群体大小水平”则可以改善该问题。Hasty课题组<sup>[48]</sup>将两套正交性的群体感应系统和裂解毒蛋白结合,构建了两个能够维持竞争关系的沙门菌群。

“分布式代谢”则是解决产物合成路线过长的

终极利器。它通过将冗长的代谢回路合理地分割成不同的部分,并导入最佳宿主中,可以越过本来在单一生物工程中由于代谢负担过大而不能生产的难题来合成复杂的天然产物。MIT的Stephanopoulos课题组利用分布式代谢的策略,将生产紫杉醇前体——氧化紫杉烷(oxygenated taxanes)的回路分段导入大肠杆菌和酿酒酵母中,利用共培养技术将氧化紫杉烷的产量提高到33 mg/L,同时证明了该体系也适用于其他氧化类异戊二烯的生产<sup>[19]</sup>。

“菌群的空间编程”则通过在空间上形成有序的组织结构来帮助整个群落应对环境的扰动<sup>[49]</sup>。斯坦福大学的Riedel-Kruse课题组<sup>[50]</sup>通过建立正交、可组合的黏附素库,结合不同的诱导性元件在*E.coli* MG1655和*E.coli* S1间建立了晶格状、相分离和不同的黏附模式,为多细胞的空间编程提供了一个工具箱,研究人员可以利用该工具箱构建各种不同的结构来提升群落的稳定性。

虽然已经能够利用“自下而上”策略建立发挥特定功能的中小型功能性菌群<sup>[16, 46, 48-50]</sup>,同时适用于合成微生物群落的工具的研发在早期合成生物学发展中已经初具雏形,元件的正交性和通用性也已经获得保障<sup>[41-42, 50]</sup>,并且已经完成对群落个体的精准控制目标<sup>[16, 43, 48-50]</sup>。但是由于群落设计原理、模型和计算工具的缺失,导致了群落行为设计的难度大、设计完成的菌群稳定性差等问题,并且合成群落中菌种数量的递增只能通过更复杂的营养化策略,随之带来的是工程菌株中构建的回路大多数是为了维持群落稳定而非产品的工业生产。以上这一系列问题给利用该策略建立大型功能性菌群带来了诸多麻烦。

## 2.2 “自上而下”的设计策略

“自上而下”的研究策略则是微生物学家和组学家的主战场。“自下而上”的研究策略擅长构建中小型功能性菌群,相比之下“自上而下”的设计策略在分析和建立大型、复杂的功能性菌群方面更具优势。这一策略的研究有明显的前期后期分界线。

### 2.2.1 通过驯化获得功能性菌群

该策略前期的研究工作集中在菌群驯化(图2),

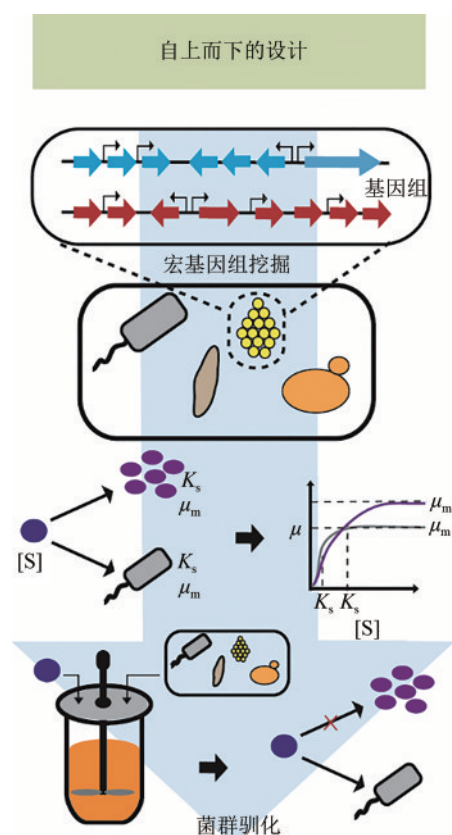


图2 使用“自上而下”策略构建功能菌群构建  
Fig. 2 “Top-Down” strategy for construct functional microbial communities

通过物理化学环境的设计引导现有的微生物组（自然发生或接种）发生生态选择，以执行期望的生物过程。成功的例子包括：来自瓦格宁根大学的 Lettinga 课题组<sup>[51]</sup>通过改变反应器内的流体动力特征来调整硫酸盐废水中硫酸盐还原菌和甲烷产生菌之间的竞争关系；南洋理工大学的 Liu 课题组<sup>[52]</sup>系统地阐述了在生物膜和活性污泥形成过程中，生物反应器内水的剪应力对于微生物群落的形成、结构和代谢过程的重要作用等。同时在这一时期发展了一系列相关的数学模型，用来量化解析系统中化学物质和相关微生物的质量平衡关系，并模拟了化学和生物化学转化率，探究了系统内的微生物群落的空间关系：昆士兰大学的 Batstone 课题组建立了厌氧消化过程的数学模型，推动了传统废水循环中反应器的设计与新兴厌氧过程的发现<sup>[53]</sup>；通过结合物理运输过程模型，来自代尔夫特理工大学的 van Loosdrecht 课题组<sup>[54]</sup>建立了多物种参与的生物膜模型来分析不同环境条

件下产生的生物膜的性质。这一系列与时间、空间相关的模型推动了“自上而下”功能性菌群设计策略的发展，加深了研究者们对于自然环境下微生物群落的形成条件和培养要求的理解，反过来指导了一系列菌群驯化相关的反应器的设计以及培养条件的设置。

### 2.2.2 通过挖掘与原位改造技术打开“环境菌群黑箱”

该策略后期研究目标慢慢转向如何打开环境菌群的“黑箱”，向大自然寻求智慧。这个时期的主要工作集中于两部分：挖掘和原位改造由大自然环境长期进化产生的生态系统（图2）。

利用宏基因组技术进行信息的挖掘。研究者们通过环境微生物的采样，不仅在大自然中发现了新的微生物物种 [如加州大学伯克利分校 (UC Berkeley) 的 Banfield 课题组<sup>[55]</sup>利用宏基因组技术从地下水中发现了一类由于 16S rRNA 的特殊性无法被传统方法检测出的原核生物，并命名其为 CPR]，而且发现了一系列的新生物砖，以及新的化学品合成回路（美国洛克菲勒大学 Brady 团队<sup>[56]</sup>利用 NRPS 基因宏基因组测序和 BAC 文库筛选的方法从 2000 多个土壤样本中发现了一种未知抗生素合成基因簇）。他们还成功发现了新的蛋白质（来自曼彻斯特大学的 Turner 课题组<sup>[57]</sup>在英国海岸的环境样品中发现了 300 多种新的亚胺还原酶）、新的 CRISPR 系统 [诺贝尔奖获得者，加州大学伯克利分校 (UC Berkeley) 的 Doudna 课题组<sup>[58]</sup>从地下水、底泥、酸性尾矿废水生物膜和土壤、婴儿肠道等微生物样品中挖掘出第 1 个来自古菌的 Cas9 蛋白和 2 个新的 CRISPR-Cas 系统] 以及一系列的特殊调控元件（哥伦比亚大学的 Wang 团队<sup>[59]</sup>通过宏基因组挖掘的方式从数据库中发现大量 5' 端调控元件，经高通量实验验证后发现一系列能够实现物种特异性表达的元件）。随着第三代测序（例如 Nanopore 和 PacBio 等）的兴起，测序的读长的增加使得我们能够挖掘到更为完整的群落信息，获得更多大自然进化中的瑰宝。

摒弃在实验室中进行实验的方式，发展原位工程的改造策略。原位工程上的发展给予了研究者们更自由的操作空间和更有希望的应用前景，但相比实验室的非原位工程而言也具有更高的难

度。首先,要保证有方法可以递送质粒进入环境:Wang 课题组<sup>[60]</sup>利用接合型质粒(conjugative plasmids),成功将荧光蛋白基因导入小鼠的肠道菌群中,建立了在哺乳动物肠道微生物中进行原位工程的方式方法;其次,要确保具有群落内精准基因编辑的工具:UCSF的Turnbaugh课题组<sup>[61]</sup>利用噬菌体M13完成了群落中大肠杆菌的精准基因编辑。上述两项工作,分别对标代谢工程中的基因工程改造和基因编辑方法,让我们在原位水平上修改群落属性成为可能。当然也有作为早期工作的延伸,通过改变环境条件进行原位菌落驯化的研究:上海交通大学的张晨虹课题组<sup>[62]</sup>通过高纤维饮食引起整个肠道微生物群落的变化,使得群落发酵产生短链脂肪酸。类似的工作还有最近吸引众多科研者目光的粪便微生物移植技术(FMT)<sup>[63]</sup>。

利用自上而下的策略设计功能性菌群的研究中,早期菌群的驯化已经能够解决很大一部分的生态问题(其中代表性的生态污泥的研究<sup>[51-52]</sup>也已趋近完善,并具有应用价值),后期在打开菌群这个“黑箱”的过程中,元件的挖掘与原位技术也取得了一定的进展<sup>[56-59]</sup>:可以看见自上而下的研究方法为自下而上的元件设计研究提供了不少新思路,两者的界限正在慢慢地被打破,一些新的研究范式正在被催生<sup>[38]</sup>。但是,在后期的研究中,环境样本的收集和元件的验证需要高通量的实验技术,巨大的实验工作量对有限的人力物力造成了极大挑战。除此之外,原位工程技术的发展不仅受到接合型质粒转入范围较小且不稳定的影响,也受限于噬菌体工程<sup>[64]</sup>:是否存在广谱噬菌体或是通过噬菌体改造来改变噬菌体的专一性,进而加强原位工程的精准性。同时该策略虽然在生态治理中具有广泛的应用前景,但是对于工业生产来说可操作性差,难以保障生物安全<sup>[38]</sup>。

### 3 功能菌群的分析手段

“自下而上”和“自上而下”两种微生物组研究策略之所以能够获得迅猛的进步,除了合成生物学技术的发展,还要归功于它们的另一个共性之处——广泛使用组学技术作为群落的分析手段:

“自下而上”的研究方法利用了培养组学技术结合16S扩增子测序技术,解析合成群落的结构以及研究群落动态变化<sup>[46]</sup>,而“自上而下”的研究策略利用宏基因组技术从环境中获得基因组信息<sup>[55-61]</sup>,研究群落的动态变化过程及原因<sup>[62]</sup>。而在此之前,研究人员已经开发了一系列的方法、流程和软件来完成复杂的组学分析问题。

#### 3.1 利用宏组学分析功能性菌群

在群落研究中广泛使用的组学技术包括宏基因组(metagenomics)、宏转录组(metatranscriptomics)、宏蛋白组(metaproteomics)、宏代谢组(metametabolomics)等。其中,扩增子测序<sup>[65]</sup>生成菌群结构谱,用于研究菌群中微生物种类、丰度和比例;宏基因组<sup>[66]</sup>解析了菌群的功能性问题;宏转录组<sup>[65]</sup>、宏蛋白质组<sup>[67]</sup>、宏代谢组<sup>[68]</sup>则揭示了在采集样本时的理化环境条件下,菌群准备发挥什么功能、菌群正在发挥什么功能、菌群已经发挥了什么功能。通过对宏组学数据的处理,我们可以获得整个微生物群落组成、功能、相互作用关系的图谱(图3)。

早期的宏组学作为功能性菌群的分析工具,

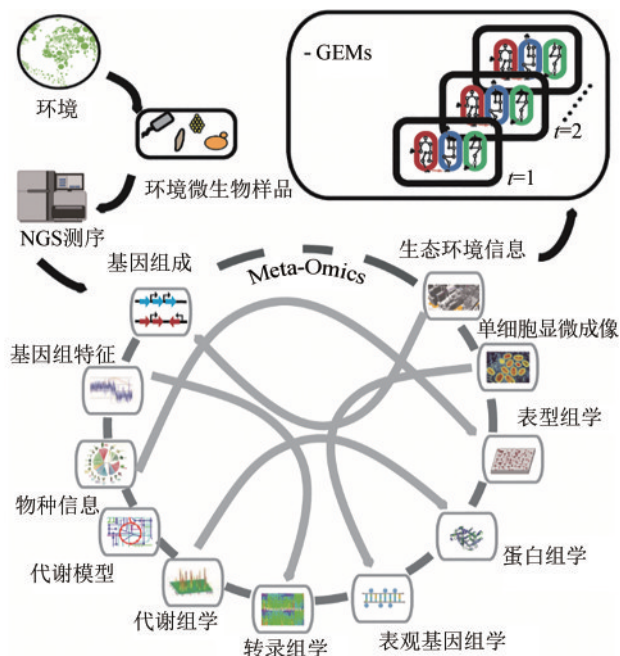


图3 使用“宏组学”分析功能性菌群

Fig. 3 Analyzing functional microbial communities using meta-omics methodology

主要集中于利用单一组学对群落进行分析。科罗拉多大学 Pace<sup>[69]</sup> 在 1985 年最早提出了从环境样品中直接克隆 DNA 的想法，而后扩增子测序和鸟枪测序法的出现推动了这一学科的发展，一系列大型的测序计划接踵而至：由美国国立卫生研究院 (NIH) 发起的人类微生物组计划 (iHMP) 通过扩增子测序 (阶段一) 和宏基因组测序 (阶段二) 的方法加深了人们对于微生物菌群相关疾病的理解<sup>[70]</sup>。同时，研究人员也将目光从地球环境转向生物反应器，如，来自康奈尔大学的 Angenent 课题组<sup>[71]</sup> 通过对 9 个生物反应器长达 1 年的采样并对样品进行扩增子测序，绘制了每个生物反应器中的群落图谱，分析结果表明了每个生物反应器中都具有独特的高稳定性群落结构。近两年来，渐渐有课题组开始利用多组学联用技术对功能菌群的进行分析：卢森堡大学的 Willems 课题组<sup>[72]</sup> 通过宏基因组、宏转录组、宏蛋白质组和宏代谢组学联用的方法，对来自生物污水处理厂的含油微生物种群进行了分析，发现了群落中存在着广泛的表型可塑性和生态位互补性，并利用培养实验揭示了微生物生态系统是如何应对干扰的；香港大学张彤课题组<sup>[73]</sup> 建立了宏基因组、宏转录组、靶向代谢组学联用的分析流程，从具有双酚 A (BPA) 生物降解功能的功能性菌群中识别出了参与种间相互作用的底物，并通过分离培养和共培养的方式证明了基于多组学整合分析方法

构建出的复杂降解途径。相较于单一组学，尤其是在扩增子测序和宏基因组测序方面，多组学联用能够更为精准地刻画功能性菌群的时空景观，但是同时产生的海量数据也为分析带来极高的难度，大幅增加的数据量带来了算力需求提升以及如何进行多组学之间的融合分析<sup>[74]</sup> 等诸多问题。

### 3.2 宏组学数据的主要处理流程与分析软件

宏组学数据的主要处理软件见表 1。

宏组学数据的处理难点主要产生于宏基因组和宏转录组的处理过程中如何从扩增子和鸟枪测序法的数据中解析出整个生物群落的整体概况——主要是针对基因序列的处理。宏蛋白组、宏代谢组两种组学的处理方式和一般蛋白组学和代谢组学方法较为一致，本文此处不再赘述。

目前，处理“宏基因组”数据的主要流程软件包括 QIIME<sup>[75]</sup>、QIIME2<sup>[76]</sup>、bioBakery<sup>[77]</sup> 和 metaWRAP<sup>[78]</sup> 等。其中，由 UCSD 的 Knight 课题组开发的 QIIME2，是 QIIME 的升级版本，由开源软件 Python 和 R 编写，具有便利的 cli、Artifact API、Studio 三种操作模式。其主要对扩增子数据进行分析，功能包括：下机数据的预处理、质量控制、聚类获得操作分类单元 (operating taxonomic units, OTUs)、利用 DADA2<sup>[89]</sup> 降噪获得扩增子序列变

表 1 宏组学数据的主要处理软件

Tab. 1 Major software for processing meta-omics data

软件名	对应处理的组学数据类型	功能
QIIME <sup>[75]</sup> 、QIIME2 <sup>[76]</sup>	扩增子测序	用于质量控制、OTU 处理、物种分类、系统发育重建、可视化的扩增子数据处理流程
bioBakery <sup>[77]</sup>	宏基因组/扩增子测序	基于比对的宏基因组数据处理流程
metaWRAP <sup>[78]</sup>	宏基因组	基于比对的宏基因组数据处理流程
MetaQUBIC <sup>[79]</sup>	宏基因组/宏转录组	基于双聚类的功能基因分类软件
MetaTrans <sup>[80]</sup>	宏转录组	从 RNA-Seq 数据分析微生物群落结构和功能的数据处理流程
SAMSA <sup>[81]</sup>	宏转录组	用于分析肠道微生物组数据的软件，重点关注样本中的生物体特异性活动或功能活动
MG-RAST <sup>[82]</sup>	宏基因组/宏转录组	通过比较蛋白质和核苷酸数据库产生宏基因组序列功能分类的软件
IdentiPy <sup>[83]</sup>	宏蛋白质组	肽识别、搜索、验证、蛋白质推断和量化的数据处理流程
Trans-Proteomic Pipeline <sup>[84]</sup>	宏蛋白质组	大规模可重复的定量 MS 蛋白质组学数据处理软件
compleXView <sup>[85]</sup>	宏蛋白质组	根据蛋白质组数据计算丰度、再现性和特异性的度量以推断 PPI 的网络服务器
Pathos <sup>[86]</sup>	宏代谢组	分析质谱数据并显示代谢物和代谢途径的网络服务器
MetaboAnalyst <sup>[87]</sup>	宏代谢组/多组学整合	用于集成代谢组学数据分析、解释和与其他组学数据集成的网络服务器
Netome <sup>[88]</sup>	宏代谢组/多组学整合	用于处理和分析代谢组学数据以及探索与其他组学数据和元数据关联的工具

体 (amplicon sequence variant, ASV)、特征表的生成、群落多样性指标的计算、群落发育树的建立等, 涵盖了扩增子测序数据处理和分析的方方面面。bioBakery 和 metaWRAP 主要针对宏基因组数据, 分别采用基于比对和基于组装两种不同的分析方式。由哈佛大学 Huttenhower 课题组研开发的 bioBakery 支持多组学的数据整合, 同时支持扩增子数据分析。整体流程整合了 Kneaddata<sup>[90]</sup> 用于对宏基因组测序数据进行质量控制; MetaPhlan<sup>[91]</sup> 用于从宏基因组鸟枪测序数据中分析微生物群落的组成; HUMAnN<sup>[92]</sup> 用于从宏基因组或宏转录组测序数据中准确地分析大量微生物代谢途径和其他分子功能; StrainPhlan<sup>[93]</sup> 用于群落内菌株级别解析度的分析和群落内系统发育树的建立与分析。metaWRAP 通过组装的方式对宏基因组测序数据进行分析: metaWRAP-Read\_qc 模块对原始数据进行预处理和质量控制, MegaHit<sup>[94]</sup>、metaSPAdes<sup>[95]</sup> 或 metaWRAP-Assembly 用于序列组装, Kraken<sup>[96]</sup> 用来对组装的序列进行快速物种分类和丰度分析、MaxBin2<sup>[97]</sup>、metaBAT2<sup>[98]</sup>、CONCOCT<sup>[99]</sup> 3 个软件用于对组装的序列进行进一步分箱。在宏基因组学数据处理的基础上, 基因组尺度代谢模型 (genome-scale metabolic model, GEM) 重建可以由 CarveMe 完成<sup>[100]</sup>。在基因组尺度代谢模型重建的基础上, 目前能够实现直接静态模拟功能 (例如, 通量平衡分析 FBA、动态通量平衡分析 dFBA 等) 的工具只有 metaGEM 这一种<sup>[101]</sup>。群落动态模型可以由广义 Lotka-Volterra 模型实现, 但是尚未有人提出从分析时序性的多组学数据到生成群落含时的动态模型与模拟的标准处理流程, 也尚未有人提出与从分析多组学的空间面板数据到生成群落区域关系的标准处理流程。

在组学算法不完善和样本质量不够稳定的影响下, 利用宏组学研究群落动态存在如下缺点<sup>[65]</sup>: 代谢网络重建不准确和不完整; 许多基因、蛋白质和代谢物的功能未知; 对驱动个体和群落水平表型的进化的原因缺乏理解; 对基因、代谢和生态系统调节体系 (例如, 群体感应信号-响应系统) 的理解有限。而且越来越大的数据量亟需一个高效的信息处理流程。

## 4 挑战与展望

尽管利用上述两种策略 (bottom-up & top-down) 设计功能菌群的研究已经取得较大的成功, 但是由于技术所限, 更多、更广泛的功能菌群的设计与合成存在很大的挑战: 诸如计算能力、数学模型的缺失以及巨大的实验工作量等。而功能性菌群作为一个复杂群体, 单独的“自下而上”和“自上而下”已经不足以用来设计这样复杂的系统。威斯康星大学麦迪逊分校的 McMahon 教授<sup>[38]</sup>通过总结两种策略设计功能性菌群的现有研究, 提出“整合设计 (integrated design)”的概念来解决两种策略的固有缺点: 将基于过程的模型 (top-down 策略) 与利用代谢组学重构的代谢模型 (bottom-up 策略) 合并, 以模拟生态系统过程、质量平衡关系和代谢通量, 并使用基因组及其衍生信息以制定菌群的设计策略; 但是 McMahon 教授并没提出可行的流程。

近些年来人工智能技术<sup>[37, 102]</sup>的兴起, 为解决“宏组学”产生的菌群大数据带来了良兵利器<sup>[103]</sup>。结合了人工智能和整合设计的方针为功能性菌群的建立引入了“智能设计”策略。以下3点问题有望在短时间内获得较好的解决方案:

### (1) 区域范围内功能性菌群的时空变化关系

区域是多尺度的, 从反应器级别到市、省、国家、地球区域, 对于各尺度区域中菌群的时空变化的理解, 有利于研究人员开展微尺度区域下功能性菌群的时空控制与培养和进行大尺度区域下功能性菌群的采集。然而, 对于此方向的研究, 研究人员一般采用解释性强但泛化性差的线性模型<sup>[104]</sup>。近期, 诸如一系列基于深度学习的时空数据模型<sup>[105]</sup>以及神经网络解释性的研究<sup>[106]</sup>的发展, 有助于提高现有模型的泛化性并保持一定的解释性。

### (2) 结合“宏组学”与深度学习的多组学群落分析流程

随着测序价格下降带来的海量组学数据, 需要一个可行、高效的信息处理流程。近年来, 深度学习在解决超高维问题中的优越性使其在生物学问题中具有广泛的应用: 研究人员使用机器学习设计蛋白质序列<sup>[107]</sup>、处理组学数据<sup>[108]</sup>等。机

器学习在“宏组学”数据的处理上,包括OUT聚类、分箱等已经有了许多的应用<sup>[103]</sup>。可以预见的是,接下来的重心应该向“如何对特征表或是其他传统算法输出结果的进行整合”这一方向发展。宏组学数据分析产生了一系列的图数据<sup>[109]</sup>,包括代谢网络、蛋白互作网络等,随着近两年来图神经网络的发展<sup>[110]</sup>,正在逐渐地转化为一个可操作的数据源;多模态学习<sup>[111]</sup>的发展则给予了多组学数据融合问题一个可以接入的框架。

(3) 功能性菌群内分布式代谢回路的计算设计

传统合成生物学通过基于模型化的思想,构建了一系列通量平衡分析模型来对底盘生物的代谢回路进行建模、模拟、预测与改造,发展了基于约束的代谢回路设计的计算方法<sup>[112]</sup>。原位工程的发展也已经给予研究人员工程化复杂群落的能力<sup>[60-64]</sup>,但反观在复杂群落中设计代谢回路,由于群落中的微生物种类和数量、关系复杂性的上升,基于约束的代谢回路设计方式已经不再适用<sup>[38]</sup>。但随着近年来强化学习的发展<sup>[113]</sup>,当把已经构建好的群落动态模型作为一个交互环境,我们仍然有希望在群落级别下的高维优化问题中获得令人满意的结果。

功能性菌群的构建作为一个热门的话题,该方面的研究现如今已经有了令人可喜的成果,但也存在一些问题和挑战。随着组学技术的不断发展和“智能设计”的引入,相信不久的将来这个领域将获得更大的突破。

## 参 考 文 献

- [1] MENG F K, ELLIS T. The second decade of synthetic biology: 2010—2020[J]. *Nature Communications*, 2020, 11: 5174.
- [2] PADDON C J, KEASLING J D. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development[J]. *Nature Reviews Microbiology*, 2014, 12(5): 355-367.
- [3] CALERO P, NIKEL P I. Chasing bacterial chassis for metabolic engineering: a perspective review from classical to non-traditional microorganisms[J]. *Microbial Biotechnology*, 2019, 12(1): 98-124.
- [4] NIELSEN J, KEASLING J D. Engineering cellular metabolism[J]. *Cell*, 2016, 164(6): 1185-1197.
- [5] WU G, YAN Q, JONES J A, et al. Metabolic burden: cornerstones in synthetic biology and metabolic engineering applications[J]. *Trends in Biotechnology*, 2016, 34(8): 652-664.
- [6] SHOU W Y, RAM S, VILAR J M G. Synthetic cooperation in engineered yeast populations[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104(6): 1877-1882.
- [7] BRENNER K, ARNOLD F H. Self-organization, layered structure, and aggregation enhance persistence of a synthetic bio-film consortium[J]. *PLoS One*, 2011, 6(2): e16791.
- [8] MCCARTY N S, LEDESMA-AMARO R. Synthetic biology tools to engineer microbial communities for biotechnology[J]. *Trends in Biotechnology*, 2019, 37(2): 181-197.
- [9] STENUIT B, AGATHOS S N. Deciphering microbial community robustness through synthetic ecology and molecular systems synecology[J]. *Current Opinion in Biotechnology*, 2015, 33: 305-317.
- [10] TSOI R, WU F L, ZHANG C, et al. Metabolic division of labor in microbial systems[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2018, 115(10): 2526-2531.
- [11] BASSLER B L, LOSICK R. Bacterially speaking[J]. *Cell*, 2006, 125(2): 237-246.
- [12] ZHANG H R, PEREIRA B, LI Z J, et al. Engineering *Escherichia coli* coculture systems for the production of biochemical products[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2015, 112(27): 8266-8271.
- [13] XIA T, EITEMAN M A, ALTMAN E. Simultaneous utilization of glucose, xylose and Arabinose in the presence of acetate by a consortium of *Escherichia coli* strains[J]. *Microbial Cell Factories*, 2012, 11: 77.
- [14] VERHOEVEN M D, DE VALK S C, DARAN J M G, et al. Fermentation of glucose-xylose-arabinose mixtures by a synthetic consortium of single-sugar-fermenting *Saccharomyces cerevisiae* strains[J]. *FEMS Yeast Research*, 2018, 18(8): foy075.
- [15] MINTY J J, SINGER M E, SCHOLZ S A, et al. Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2013, 110(36): 14592-14597.
- [16] GILBERT E S, WALKER A W, KEASLING J D. A constructed microbial consortium for biodegradation of the organophosphorus insecticide parathion[J]. *Applied Microbiology and Biotechnology*, 2003, 61(1): 77-81.
- [17] BOKINSKY G, PERALTA-YAHYA P P, GEORGE A, et al. Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered *Escherichia coli*[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2011, 108(50): 19949-19954.
- [18] TSAI S L, GOYAL G, CHEN W. Surface display of a function-

- al minicellulosome by intracellular complementation using a synthetic yeast consortium and its application to cellulose hydrolysis and ethanol production[J]. *Applied and Environmental Microbiology*, 2010, 76(22): 7514-7520.
- [19] ZHOU K, QIAO K J, EDGAR S, et al. Distributing a metabolic pathway among a microbial consortium enhances production of natural products[J]. *Nature Biotechnology*, 2015, 33(4): 377-383.
- [20] MINAMI H, KIM J S, IKEZAWA N, et al. Microbial production of plant benzylisoquinoline alkaloids[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105(21): 7393-7398.
- [21] JONES J A, VERNACCHIO V R, COLLINS S M, et al. Complete biosynthesis of anthocyanins using *E. coli* polycultures[J]. *mBio*, 2017, 8(3): e00621-e00617.
- [22] VILLARREAL F, CONTRERAS-LLANO L E, CHAVEZ M, et al. Synthetic microbial consortia enable rapid assembly of pure translation machinery[J]. *Nature Chemical Biology*, 2018, 14(1): 29-35.
- [23] BAYER T S, WIDMAIER D M, TEMME K, et al. Synthesis of methyl halides from biomass using engineered microbes[J]. *Journal of the American Chemical Society*, 2009, 131(18): 6508-6515.
- [24] WANG E X, DING M Z, MA Q, et al. Reorganization of a synthetic microbial consortium for one-step vitamin C fermentation[J]. *Microbial Cell Factories*, 2016, 15: 21.
- [25] CHEN A Y, DENG Z T, BILLINGS A N, et al. Synthesis and patterning of tunable multiscale materials with engineered cells[J]. *Nature Materials*, 2014, 13(5): 515-523.
- [26] GILBERT C, HOWARTH M, HARWOOD C R, et al. Extracellular self-assembly of functional and tunable protein conjugates from *Bacillus subtilis*[J]. *ACS Synthetic Biology*, 2017, 6(6): 957-967.
- [27] HONG S H, HEGDE M, KIM J, et al. Synthetic quorum-sensing circuit to control consortial biofilm formation and dispersal in a microfluidic device[J]. *Nature Communications*, 2012, 3: 613.
- [28] URRIOS A, MACIA J, MANZONI R, et al. A synthetic multicellular memory device[J]. *ACS Synthetic Biology*, 2016, 5(8): 862-873.
- [29] MACIA J, MANZONI R, CONDE N, et al. Implementation of complex biological logic circuits using spatially distributed multicellular consortia[J]. *PLoS Computational Biology*, 2016, 12(2): e1004685.
- [30] REGOT S, MACIA J, CONDE N, et al. Distributed biological computation with multicellular engineered networks[J]. *Nature*, 2011, 469(7329): 207-211.
- [31] TAMSIR A, TABOR J J, VOIGT C A. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'[J]. *Nature*, 2011, 469(7329): 212-215.
- [32] BOEHM C R, GRANT P K, HASELOFF J. Programmed hierarchical patterning of bacterial populations[J]. *Nature Communications*, 2018, 9: 776.
- [33] BASU S, GERCHMAN Y, COLLINS C H, et al. A synthetic multicellular system for programmed pattern formation[J]. *Nature*, 2005, 434(7037): 1130-1134.
- [34] XIU Y, JANG S, JONES J A, et al. Naringenin-responsive riboswitch-based fluorescent biosensor module for *Escherichia coli* co-cultures[J]. *Biotechnology and Bioengineering*, 2017, 114(10): 2235-2244.
- [35] MEYER A, PELLAUX R, POTOT S, et al. Optimization of a whole-cell biocatalyst by employing genetically encoded product sensors inside nanolitre reactors[J]. *Nature Chemistry*, 2015, 7(8): 673-678.
- [36] HUGENHOLTZ P, TYSON G W. Metagenomics[J]. *Nature*, 2008, 455(7212): 481-483.
- [37] RUSSELL S, NORVIG P. Artificial intelligence: a modern approach[M]. 2<sup>nd</sup> ed. Prentice Hall, 2003.
- [38] LAWSON C E, HARCMBE W R, HATZENPICHLER R, et al. Common principles and best practices for engineering microbiomes[J]. *Nature Reviews Microbiology*, 2019, 17(12): 725-741.
- [39] MILLER M B, BASSLER B L. Quorum sensing in bacteria[J]. *Annual Review of Microbiology*, 2001, 55: 165-199.
- [40] SCOTT S R, HASTY J. Quorum sensing communication modules for microbial consortia[J]. *ACS Synthetic Biology*, 2016, 5(9): 969-977.
- [41] MIANO A, LIAO M J, HASTY J. Inducible cell-to-cell signaling for tunable dynamics in microbial communities[J]. *Nature Communications*, 2020, 11: 1193.
- [42] DU P, ZHAO H W, ZHANG H Q, et al. *De novo* design of an intercellular signaling toolbox for multi-channel cell-cell communication and biological computation[J]. *Nature Communications*, 2020, 11: 4226.
- [43] KONG W T, MELDGIN D R, COLLINS J J, et al. Designing microbial consortia with defined social interactions[J]. *Nature Chemical Biology*, 2018, 14(8): 821-829.
- [44] LÓPEZ-MAURY L, MARGUERAT S, BÄHLER J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation[J]. *Nature Reviews Genetics*, 2008, 9(8): 583-593.
- [45] ZENGLER K, ZARAMELA L S. The social network of microorganisms—how auxotrophies shape complex communities[J]. *Nature Reviews Microbiology*, 2018, 16(6): 383-390.
- [46] MEE M T, COLLINS J J, CHURCH G M, et al. Syntrophic exchange in synthetic microbial communities[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2014, 111(20): E2149-E2156.

- [47] WEGNER A, MEISER J, WEINDL D, et al. How metabolites modulate metabolic flux[J]. *Current Opinion in Biotechnology*, 2015, 34: 16-22.
- [48] SCOTT S R, DIN M O, BITTIHN P, et al. A stabilized microbial ecosystem of self-limiting bacteria using synthetic quorum-regulated lysis[J]. *Nature Microbiology*, 2017, 2: 17083.
- [49] ELIAS S, BANIN E. Multi-species biofilms: living with friendly neighbors[J]. *FEMS Microbiology Reviews*, 2012, 36(5): 990-1004.
- [50] GLASS D S, RIEDEL-KRUSE I H. A synthetic bacterial cell-cell adhesion toolbox for programming multicellular morphologies and patterns[J]. *Cell*, 2018, 174(3): 649-658.e16.
- [51] ALPHENAAR P A, VISSER A, LETTINGA G. The effect of liquid upward velocity and hydraulic retention time on granulation in UASB reactors treating wastewater with a high sulphate content[J]. *Bioresource Technology*, 1993, 43(3): 249-258.
- [52] LIU Y, TAY J H. The essential role of hydrodynamic shear force in the formation of biofilm and granular sludge[J]. *Water Research*, 2002, 36(7): 1653-1665.
- [53] BATSTONE D J, PUYOL D, FLORES-ALSINA X, et al. Mathematical modelling of anaerobic digestion processes: applications and future needs[J]. *Reviews in Environmental Science and Bio/Technology*, 2015, 14(4): 595-613.
- [54] PICIOREANU C, KREFT J U, VAN LOOSDRECHT M C M. Particle-based multidimensional multi species biofilm model[J]. *Applied and Environmental Microbiology*, 2004, 70(5): 3024-3040.
- [55] BROWN C T, HUG L A, THOMAS B C, et al. Unusual biology across a group comprising more than 15% of domain Bacteria[J]. *Nature*, 2015, 523(7559): 208-211.
- [56] HOVER B M, KIM S H, KATZ M, et al. Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens[J]. *Nature Microbiology*, 2018, 3(4): 415-422.
- [57] MARSHALL J R, YAO P Y, MONTGOMERY S L, et al. Screening and characterization of a diverse panel of metagenomic imine reductases for biocatalytic reductive amination[J]. *Nature Chemistry*, 2021, 13(2): 140-148.
- [58] BURSTEIN D, HARRINGTON L B, STRUTT S C, et al. New CRISPR-Cas systems from uncultivated microbes[J]. *Nature*, 2017, 542(7640): 237-241.
- [59] JOHNS N I, GOMES A L C, YIM S S, et al. Metagenomic mining of regulatory elements enables programmable species-selective gene expression[J]. *Nature Methods*, 2018, 15(5): 323-329.
- [60] RONDA C, CHEN S P, CABRAL V, et al. Metagenomic engineering of the mammalian gut microbiome *in situ*[J]. *Nature Methods*, 2019, 16(2): 167-170.
- [61] LAM K N, SPANOGIANNOPOULOS P, SOTO-PEREZ P, et al. Phage-delivered CRISPR-Cas9 for strain-specific depletion and genomic deletions in the gut microbiome[J]. *Cell Reports*, 2021, 37(5): 109930.
- [62] ZHAO L P, ZHANG F, DING X Y, et al. Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes[J]. *Science*, 2018, 359(6380): 1151-1156.
- [63] DREKONJA D, REICH J, GEZAHEGN S, et al. Fecal microbiota transplantation for clostridium difficile infection: a systematic review[J]. *Annals of Internal Medicine*, 2015, 162(9): 630-638.
- [64] SHETH R U, CABRAL V, CHEN S P, et al. Manipulating bacterial communities by *in situ* microbiome engineering[J]. *Trends in Genetics*, 2016, 32(4): 189-200.
- [65] KNIGHT R, VRBANAC A, TAYLOR B C, et al. Best practices for analysing microbiomes[J]. *Nature Reviews Microbiology*, 2018, 16(7): 410-422.
- [66] QUINCE C, WALKER A W, SIMPSON J T, et al. Shotgun metagenomics, from sampling to analysis[J]. *Nature Biotechnology*, 2017, 35(9): 833-844.
- [67] WILMES P, BOND P L. Metaproteomics: studying functional gene expression in microbial ecosystems[J]. *Trends in Microbiology*, 2006, 14(2): 92-97.
- [68] NICHOLSON J K. Global systems biology, personalized medicine and molecular epidemiology[J]. *Molecular Systems Biology*, 2006, 2: 52.
- [69] PACE N R, STAHL D A, LANE D J, et al. The analysis of natural microbial populations by ribosomal RNA sequences[M]// Marshall K C. *Advances in microbial ecology*. Springer, 1986: 1-55.
- [70] GROUP N I H H W, PETERSON J, GARGES S, et al. The NIH human microbiome project[J]. *Genome Research*, 2009, 19(12): 2317-2323.
- [71] WERNER J J, KNIGHTS D, GARCIA M L, et al. Bacterial community structures are unique and resilient in full-scale bioenergy systems[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2011, 108(10): 4158-4163.
- [72] HEROLD M, MARTÍNEZ ARBAS S, NARAYANASAMY S, et al. Integration of time-series meta-omics data reveals how microbial ecosystems respond to disturbance[J]. *Nature Communications*, 2020, 11: 5281.
- [73] YU K, YI S, LI B, et al. An integrated meta-omics approach reveals substrates involved in synergistic interactions in a bisphenol A (BPA)-degrading microbial community[J]. *Microbiome*, 2019, 7(1): 16.
- [74] SEGATA N, BOERNIGEN D, TICKLE T L, et al. Computational meta-omics for microbial community studies[J]. *Molecular Systems Biology*, 2013, 9: 666.
- [75] CAPORASO J G, KUCZYNSKI J, STOMBAUGH J, et al. QI-

- IME allows analysis of high-throughput community sequencing data[J]. *Nature Methods*, 2010, 7(5): 335-336.
- [76] BOLYEN E, RIDEOUT J R, DILLON M R, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2[J]. *Nature Biotechnology*, 2019, 37(8): 852-857.
- [77] MCIVER L J, ABU-ALI G, FRANZOSA E A, et al. bioBakery: a meta'omic analysis environment[J]. *Bioinformatics*, 2017, 34(7): 1235-1237.
- [78] URITSKIY G V, DIRUGGIERO J, TAYLOR J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis[J]. *Microbiome*, 2018, 6(1): 158.
- [79] MA A J, SUN M X, MCDERMAID A, et al. MetaQUBIC: a computational pipeline for gene-level functional profiling of metagenome and metatranscriptome[J]. *Bioinformatics*, 2019, 35(21): 4474-4477.
- [80] MARTINEZ X, POZUELO M, PASCAL V, et al. MetaTrans: an open-source pipeline for metatranscriptomics[J]. *Scientific Reports*, 2016, 6: 26447.
- [81] WESTREICH S T, KORF I, MILLS D A, et al. SAMSa: a comprehensive metatranscriptome analysis pipeline[J]. *BMC Bioinformatics*, 2016, 17(1): 399.
- [82] MEYER F, PAARMANN D, D'SOUZA M, et al. The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes[J]. *BMC Bioinformatics*, 2008, 9: 386.
- [83] LEVITSKY L I, IVANOV M V, LOBAS A A, et al. IdentiPy: an extensible search engine for protein identification in shotgun proteomics[J]. *Journal of Proteome Research*, 2018, 17(7): 2249-2255.
- [84] DEUTSCH E W, MENDOZA L, SHTEYNBERG D, et al. A guided tour of the Trans-Proteomic Pipeline[J]. *Proteomics*, 2010, 10(6): 1150-1159.
- [85] SOLIS-MEZARINO V, HERZOG F. compleXView: a server for the interpretation of protein abundance and connectivity information to identify protein complexes[J]. *Nucleic Acids Research*, 2017, 45(W1): W276-W284.
- [86] LEADER D P, BURGESS K, CREEK D, et al. Pathos: a web facility that uses metabolic maps to display experimental changes in metabolites identified by mass spectrometry[J]. *Rapid Communications in Mass Spectrometry*, 2011, 25(22): 3422-3426.
- [87] CHONG J, SOUFAN O, LI C, et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis[J]. *Nucleic Acids Research*, 2018, 46(W1): W486-W494.
- [88] RAHNAVARD A, HITCHCOCK D, PACHECO J A, et al. Nettle: a computational framework for metabolite profiling and omics network analysis[E/OL]. *BioRxiv*, 2018: 443903. <http://www.biorxiv.org/content/10.1101/443903v1>.
- [89] CALLAHAN B J, MCMURDIE P J, ROSEN M J, et al. DADA2: High-resolution sample inference from Illumina amplicon data[J]. *Nature Methods*, 2016, 13(7): 581-583.
- [90] KneadData. *BioBakery* 2017, <https://bitbucket.org/biobakery/kneaddata>
- [91] BEGHINI F, MCIVER L J, BLANCO-MÍGUEZ A, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3[J]. *eLife*, 2021, 10: e65088.
- [92] FRANZOSA E A, MCIVER L J, RAHNAVARD G, et al. Species-level functional profiling of metagenomes and metatranscriptomes[J]. *Nature Methods*, 2018, 15(11): 962-968.
- [93] TRUONG D T, TETT A, PASOLLI E, et al. Microbial strain-level population structure and genetic diversity from metagenomes[J]. *Genome Research*, 2017, 27(4): 626-638.
- [94] LI D H, LUO R B, LIU C M, et al. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices[J]. *Methods*, 2016, 102: 3-11.
- [95] NURK S, MELESHKO D, KOROBENNIKOV A, et al. metaSPAdes: a new versatile metagenomic assembler[J]. *Genome Research*, 2017, 27(5): 824-834.
- [96] LU J, BREITWIESER F P, THIELEN P, et al. Bracken: estimating species abundance in metagenomics data[J]. *PeerJ Computer Science*, 2017, 3: e104.
- [97] ALNEBERG J, BJARNASON B S, DE BRUIJN I, et al. Binning metagenomic contigs by coverage and composition[J]. *Nature Methods*, 2014, 11(11): 1144-1146.
- [98] WU Y W, SIMMONS B A, SINGER S W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets[J]. *Bioinformatics*, 2015, 32(4): 605-607.
- [99] KANG D D, FROULA J, EGAN R, et al. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities[J]. *PeerJ*, 2015, 3: e1165.
- [100] MACHADO D, ANDREJEV S, TRAMONTANO M, et al. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities[J]. *Nucleic Acids Research*, 2018, 46(15): 7542-7553.
- [101] ZORRILLA F, BURIC F, PATIL K R, et al. metaGEM: reconstruction of genome scale metabolic models directly from metagenomes[J]. *Nucleic Acids Research*, 2021, 49(21): e126.
- [102] GLAZ B, WANG C, HURLEY M, et al. Artificial intelligence in synthetic biology, cyber defense, and aeromechanical design[R]. *CCDC Army Research Laboratory Adelphi United States*, 2020.
- [103] SOUEIDAN H, NIKOLSKI M. Machine learning for metagenomics: methods and tools[E/OL]. *arXiv: 1510.06621*, 2015. <http://arxiv.org/abs/1510.06621>.
- [104] THOMPSON L R, SANDERS J G, MCDONALD D, et al. A communal catalogue reveals Earth's multiscale microbial diver-

- city[J]. *Nature*, 2017, 551(7681): 457-463.
- [105] WANG S Z, CAO J N, YU P. Deep learning for spatio-temporal data mining: a survey[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2020. doi: 10.1109/TKDE.2020.3025580.
- [106] MONTAVON G, SAMEK W, MÜLLER K R. Methods for interpreting and understanding deep neural networks[J]. *Digital Signal Processing*, 2018, 73: 1-15.
- [107] YANG J Y, ANISHCHENKO I, PARK H, et al. Improved protein structure prediction using predicted interresidue orientations[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2020, 117(3): 1496-1503.
- [108] ANDERSON C. Google's AI tool DeepVariant promises significantly fewer genome errors[J]. *Clinical OMICS*, 2018, 5(1): 33.
- [109] LIU Z Q, MA A J, MATHÉ E, et al. Network analyses in microbiome based on high-throughput multi-omics data[J]. *Briefings in Bioinformatics*, 2020, 22(2): 1639-1655.
- [110] ZHOU J, CUI G Q, HU S D, et al. Graph neural networks: a review of methods and applications[J]. *AI Open*, 2020, 1: 57-81.
- [111] RAMACHANDRAM D, TAYLOR G W. Deep multimodal learning: a survey on recent advances and trends[J]. *IEEE Signal Processing Magazine*, 2017, 34(6): 96-108.
- [112] LIN G M, WARDEN-ROTHMAN R, VOIGT C A. Retrosynthetic design of metabolic pathways to chemicals not found in

nature[J]. *Current Opinion in Systems Biology*, 2019, 14: 82-107.

- [113] NGUYEN T T, NGUYEN N D, NAHAVANDI S. Deep reinforcement learning for multiagent systems: a review of challenges, solutions, and applications[J]. *IEEE Transactions on Cybernetics*, 2020, 50(9): 3826-3839.



**通讯作者:** 方柏山(1957—),男,教授。研究方向为合成生物学与生物分子机器;定向进化与生物催化;生物技术过程开发与优化等。

E-mail: fbs@xmu.edu.cn



**第一作者:** 黄佳城(1997—),男,硕士研究生。研究方向为人工智能在合成生物学与宏基因组上的应用等。

E-mail: chonpcaacpnohc@gmail.com