



昆虫肠道的宏基因组学：微生物大数据的新疆界

曹乐, 宁康*

华中科技大学生命科学与技术学院, 湖北 武汉 430074

摘要: 微生物作为自然界中普遍存在的生命体, 通常以“微生物群落”的形式共存。这些物种相互协作适应环境变化的同时, 也对环境产生了长期而深刻的影响。随着人类对于微生物了解的深入, 微生物群落基础研究及其在健康和环境等领域的应用研究日益重要。昆虫肠道内存在种类繁多、数量庞大的微生物, 一方面, 这些肠道微生物种群结构的多样性与昆虫种类、龄期、消化道形式、食物的来源、环境等都息息相关。另一方面, 这些菌群也对宿主的一些生理活动有着一定的影响。随着高通量测序技术、组学技术的发展, 昆虫肠道宏基因组大数据挖掘和应用已经成为研究热点, 极大地推动人类微生物资源利用的能力。本文概述了昆虫肠道微生物宏基因组学的发展现状和发展趋势, 特别是肠道宏基因组学大数据的挖掘工具和应用, 以及现阶段昆虫肠道宏基因组学的研究进展、应用、优势和瓶颈, 并对今后昆虫肠道微生物组大数据研究方向进行展望。

关键词: 微生物组, 昆虫肠道, 高通量测序, 生物大数据

微生物组学(microbiome)的主要研究内容, 包括微生物群落所有遗传物质, 相关环境参数和代谢产物, 以及它们之间的复杂关系。群落微生物组的研究, 是研究微生物遗传信息的重要突破。对微生物群落相关微生物组数据的深入分析, 允许我们从新的层面认识和理解群落对于广义生态环境的响应及反馈, 以及微生物物种之间的相互影响和调控机制, 发掘微生物群落复杂功能和潜在应用途径(健康和环境等领域)。鉴于微生物的重要性及其复杂性, 美国于 2016 年启动了“国家微生物组计划”, 计划投资 1 亿多美元用于微生物研

究, 中国最近也在酝酿启动微生物组研究计划, 并在 2017 年被科技部列为“重大颠覆性技术”之一(参考《中共科学技术部党组关于贯彻落实党的十八届六中全会精神深入实施创新驱动发展战略开启建设世界科技强国新征程的意见》, 国科党组发〔2017〕1号)。

宏基因组(metagenome)研究方法, 又叫微生物环境基因组学、元基因组学, 对特定环境中全部微生物的总 DNA(也称宏基因组, metagenomic)进行克隆, 并通过构建宏基因组文库和筛选等手段获得新的生理活性物质; 或者根据 rRNA 数据库

*通信作者。Tel/Fax: +86-27-87793041; E-mail: ningkang@hust.edu.cn

收稿日期: 2017-09-29; 修回日期: 2018-02-13; 网络出版日期: 2018-05-09

设计引物, 通过系统学分析获得该环境中微生物的遗传多样性和分子生态学信息, 是微生物组研究的重要方法之一。其基于测序技术的宏基因组数据收集与分析克服了传统分离培养方法仅局限于群落中可培养组分的缺陷, 使挖掘、认识与利用不可培养的组分成为可能, 也使得我们能够全面地研究自然状况下微生物群落的结构和组成^[1-2]。尤其是针对现在难以甚至不可在实验室条件下培养的微生物群落, 宏基因组方法是一种必不可少的研究手段。微生物组研究方法通过高通量测序等技术, 一次性获得微生物群落的全部遗传信息并进行生物信息学分析。通过宏基因组研究, 我们可以考察群落中物种与基因的多样性和分布情况, 理解影响群落结构和功能的关键因素, 从而认识分子至细胞乃至种群内部及之间的相互作用机制。宏基因组研究常用的办法是使用细菌、古菌和真菌的特异性引物进行系统发育标记分子(如 16S rRNA)的扩增, 并通过测定其序列来识别微生物群落的物种组分并定量其相对丰度; 微生物群落的结构信息也可以通过对群落的所有 DNA 进行宏基因组测序而得到。全长 DNA 测序(包括二代测序和三代测序)和 16S rRNA 这两种技术手段相结合, 不仅可以获得微生物群落中物种组分的信息, 而且还能够在基因及其功能水平上对群落中的微生物进行解析和比较。正是基于上述这两种技术手段的结合, 使得我们对一些重要的微生物群落的结构和功能的认识迅速取得了重大突破^[3-4]。而在微生物组应用研究方面, 目前主要聚焦于与人类生产和生活密切相关的健康^[5]、环境^[6]和能源^[7]领域。例如, 通过稳定和恢复寄生在人体的微生物群来提供相关的健康服务; 对微生物进行筛选优化来找出能够处理生活废水的菌株; 利用微生物的氧化还原反应提取能量等。

目前, 微生物定殖最多的器官是胃肠道^[8], 因为肠黏膜具有较大的表面积, 可以容纳大量微生物群落在其表面定殖^[9], 此外, 胃肠道含有大量微生物可以利用的营养物质, 是微生物首选的定殖点。而且肠道微生物与宿主处于共生的关系, 其参与宿主营养吸收、代谢作用、肠道和免疫系统发育等重要生理过程, 所以极具研究价值。

但是由于大多数肠道微生物不能被人工培养, 所以较以往的体外分离培养的传统研究方法, 免培养分析法在认识微生物群落结构方面更具优势。采用研究人类及其微生物联合的基因结构——宏基因组学, 可提供大量的、丰富的基因信息。宏基因组学方法有明显的优势, 综合分析了复杂微生物群落的系统、物理和功能特性, 充分了解了微生物的动态变化, 可以用来研究系统发生的多样性, 提供了肠道微生物潜在功能的库存, 从而成为菌群多样性和功能相结合研究的工具。Relman 实验室和美国基因组研究所成员于 2005 年第一次全面开展人类肠道宏基因组学的研究以完善我们对胃肠道微生物群落间多样性的认识^[10]。微生物群落结构的改变可能引起肠道疾病也可能增强人体抵抗力等, 其对宿主的代谢具有一定影响, 这就需要将代谢活动和肠道细菌多样性结合分析。肠道菌群活动紊乱与肠道、肝脏、胰腺甚至脑部特定疾病都存在密切联系。例如肝硬化患者的肠道菌群功能多样性降低, 对药物、必需氨基酸、丙酸盐等的代谢能力及炎症反应显著增强, 但对胆汁酸等的代谢能力以及细胞周期相关的功能显著降低, 与临床表现——胃肠蠕动减慢, 肠道通透性增加, 肝功能损害, 对药物氨基酸等的清除功能下降, 而且胆汁、胆酸分泌减少相关联^[11]。

本文主要对昆虫肠道微生物进行研究和分析,

昆虫肠道微生物本身较于其他肠道微生物有其特殊性,这与昆虫及其肠道本身的一些特性有关。

昆虫的特点:昆虫是全球最多样化、数量最多、进化历史最悠久、分布最广泛的动物之一,同时也是生态习性最丰富的动物类群^[12]。而且它们具有多样化的取食特性和行为,因此几乎陆地上所有的食物资源都能被昆虫所消耗^[13]。

昆虫的发育特点:昆虫的发育过程有两种形式,完全变态和不完全变态。不完全变态的发育过程比较缓慢,昆虫经卵孵化到达幼虫阶段(也被称为若虫),并在之后的生长过程中幼虫会经历几次蜕变,最后完全发育成成虫。在整个完全变态的发育过程中,昆虫会经历卵、幼虫、蛹和成虫这4个阶段,但是幼虫与成虫的形态差异非常大,因此幼虫必须要经历蛹的阶段才能成为成虫,并且在变态过程中,幼虫特有的组织器官会逐渐退化,成虫的特有器官会逐渐形成并达到性成熟。一般来说,幼虫从蜕皮发育到成蛹前,其生理功能和生活环境并不会发生很大的改变,并且性器官不成熟,不能进行繁殖;但是在变态发育后他的生理功能和生活环境一般会发生巨大的改变。例如,许多昆虫尤其是农林业害虫,他们在幼虫期时一般以植物为专一食物,而发育成成虫后,食性可能会发生完全改变,变成杂食性。

昆虫肠道的发育特点:大部分昆虫肠道主要由前肠、中肠和后肠组成^[14]。前肠一般具有用于临时贮存食物的嗉囊;中肠是很多昆虫用来消化食物、吸收养分的主要部位;后肠(包括回肠和直肠)存有一些含氮废物和食物残渣,为昆虫肠道微生物提供营养环境^[15],其中回肠的主要功能是将中肠已消化吸收后残余的食物废渣和马氏管在血淋巴中收集的代谢废物送入直肠,而直肠的功能

是在贮存过程中重吸收粪便中的水分^[15]并排出体外。昆虫肠道因为昆虫种类的不同而差异很大,这也是昆虫为了适应各种特殊生态环境和饮食习惯长期协同进化的结果,这种协同进化会逐渐导致在昆虫特定肠道部位定居特定的肠道微生物的现象^[15]。对于经过完全变态发育的昆虫来说,幼虫、蛹和成虫等几个阶段肠道的区分十分明显,因为在这一变态过程中肠和其他器官会发生彻底变化,如成虫阶段,幼虫的肠道会完全消失,并且在蛹期时肠道会被围食膜包裹等^[14-15]。但是,有些肠道微生物可以保留下来,因为许多昆虫的肠道内具有特化的隐窝(crypts)结构,它可以促进微生物的留存^[14-15]。此外,在昆虫达到成虫阶段前的多次蜕皮过程中,前肠或后肠壁会逐渐形成适于细菌定殖的表面结构^[15]。因此,昆虫从幼虫到成虫时期形态会发生一定变化,消化道的形态和物理化学性质差异也很大,并且食物摄取也会发生相应的变化,会在一定程度上影响微生物群落的结构。

昆虫的肠道营养:昆虫对基本的营养要求包括蛋白质、碳水化合物、维生素、水、无机盐、特殊的微量物质。但是由于昆虫的种类及其发育的阶段不同,对食物的营养要求差别通常较大。此外,摄入的营养质量常影响到昆虫的生长速度,营养条件不良或者半饥饿状态还会使幼虫期延长、蜕皮次数增加或者进入特殊时期。

1 微生物组大数据的研究现状、趋势和瓶颈

1.1 微生物组大数据的研究现状

目前一些常用的宏基因组生物信息学分析软

件已经应用于各项微生物群落科学研究之中, 并由此建立了一些数据库, 其中包括 Greengenes^[16]、RDP^[17]和 SILVA^[18]等大型特征序列数据库, 还有一些微生物群落研究相关的数据库包括 MG_RAST^[19]、CAMERA^[20]、MMCD 等专业数据库, 以及 NCBI

和 EBI^[21]等通用数据库(表 1)。

有关宏基因组的生物信息学分析软件主要有: 整合所有分析步骤的宏基因组分析流程的工具, 它集合质量控制、序列归类、分析预测等功能于一体, 如 Phyloshop^[22]、QIIME^[23]等分析平台,

表 1. 代表性的生物信息学分析平台和数据库资源

Table 1. Representative bioinformatics analysis platform and database resources

Software (Platform)	Database	Analysis object	Analysis strategy	Analysis results	References
MEGAN	NCBI	16S rRNA	Sequence comparative analysis	Classification of species, structures, abundance and functions, species-species comparisons	[46]
ConStrains	Integrated database	Metagenome	Sequence comparative analysis and sequence components analysis	Species structure and abundance	[33]
MetaPhlan	Integrated database	Metagenome	Sequence comparative analysis and sequence components analysis	Species structure and abundance	[35–36]
PICRUSt	Integrated database	Metagenome, 16S rRNA	Sequence comparative analysis and sequence components analysis	Classification of species, structures and functions	[47]
antiSMASH	Integrated database	Metagenome	Sequence comparative analysis and sequence components analysis	Analysis of biosynthesis gene cluster	[38,40]
CARMA	Pfam	16S rRNA	Sequence comparative analysis	Classification of species, structures and functions	[27]
Sort-ITEMS	NCBI	16S rRNA	Sequence comparative analysis	Classification of species, structures and functions	[29]
Phyloshop	Greengenes	Whole genome, 16S rRNA	Sequence comparative analysis	Classification of species, structures and functions	[22]
UniFrac	NCBI	16S rRNA	Sequence comparative analysis	Classification of species, structures, abundance and functions, species-species comparisons	[28]
QIIME	–	16S rRNA	Sequence comparative analysis	Classification of species, structures, abundance and functions	[23,48]
PhyloPythia	NCBI	16S rRNA	Sequence comparative analysis	Classification of species, structures, abundance and functions	[32]
MG-RAST	Integrated database	Whole genome, 16S rRNA	Sequence comparative analysis and sequence components analysis	Classification of species, structures, abundance and functions, species-species comparison	[30]
CAMERA	Integrated database	Whole genome, 16S rRNA	Sequence comparative analysis and sequence components analysis	Classification of species, structures, abundance and functions, species-species comparisons	[20]
Galaxy	Integrated database	Whole genome, 16S rRNA	Sequence comparative analysis and sequence components analysis	Classification of species, structures, abundance and functions, species-species comparisons	[31]

和 MG-RAST^[24]、CAMERA^[20]等, 为微生物宏基因组研究和成果 的分享提供一站式服务的网站。除此之外, 还有功能较有针对性的软件, 如负责 对 Solexa 或者 Roche454 测序仪所生成的短序列进行质量控制的软件包有 Mothur^[25]和 QC-Chain^[26] 等。对测序数据进行序列比对分析和序列成分分析的分析平台有 CARMA^[27]、UniFrac^[28]、Sort-ITEMS^[29]、MG-RAST^[30]、Galaxy^[31]、PhyloPythia^[32]等。其中涉及到对群落中的亚种鉴定和分析^[33]的部分, 有 MIDAS^[34]、ConStrains^[33]、MetaPhlan^[35-36]等软件包。

有关功能划分的微生物组研究只是刚起步阶段^[37], 特别是在生物合成基因簇(BGC)^[38-40]和抗生素抗性基因簇(ARG)^[41]等基因功能团的分析方面。在生物合成基因簇研究方面, 有代表性的 antiSMASH 3.0 和 NaPDoS^[42]分析平台^[38,40]以及 DoBISCUIT^[43]、ClusterMine360^[44]、IMG-ABC^[45]等数据库, 但在抗生素抗性基因簇分析方面仍缺乏优化的、可以全面整合各类信息和方法的研究平台, 使得 ARG 的研究实际上比 BGC 的研究更为滞后。

这一系列宏基因组生物信息学分析软件正支撑着人类利用新一代测序技术对广袤的微生物世界的探索。平台化运行的宏基因组数据收集与分析已经支撑了包括肠道^[49-50]、口腔^[51-52]、土壤^[53]和食品等诸多重要生态环境与过程的机理研究。而微生物组大数据更高级的数据挖掘就会涉及到微生物群落生态和进化方面的机制和规律的解析, 目前的重点研究领域^[54-56]是宿主表型组和基因组以及微生物群落的数据整合与数据挖掘^[54,57], 以及微生物群落内关于各个物种的 SNP 差异分析^[58]和群落内部的共进化^[59]问题。同时, 已经开发了

一些先进的微生物群落物种的组装方法, 为更加详细准确的功能基因注释提供了方法基础^[60]。可以基于高精度的微生物群落基因功能注释进行更准确的富集分析, 进一步加强了对微生物群落功能的深入理解^[36], 以及对重要代谢通路的解析^[61]和对菌群宿主间代谢调控机制的解析^[61-62]。并由此提出了针对群落微生态网络较为完整的分析方法^[63-64], 并有实例表明已成功地在大规模的海洋微生态网络研究中^[65-66]应用。而且, 在最新的网路分析方法中将时序样本考虑在内, 可以交互式地展示出动态网络的分析结果^[67], 进一步加强对于微生态的深入理解。

本课题组最近几年来也在微生物组大数据分析方法和应用方面做了一些工作。其中在方法开发方面, 主要包括: (1) 针对群落物种和功能解析的 Parallel-Meta 系列方法^[68], 不但提高了定性和定量分析精度, 而且将群落结构和功能解析效率提高了 1-2 个数量级。(2) 针对海量群落样本比较和搜索的 Meta-Storms 系列方法^[69], 实现了以极高的效率和准确率完成上千个群落样本比对和搜索。(3) 初步建立了元基因组数据整合数据模型、样本聚类分析策略、生物标记识别策略和可视化策略, 构建多来源、异质性的元基因组数据整合方法和生物标记识别方法^[70]。而在应用方面, 已经完成了包括水体环境^[71]、环境微生物群落^[72]和人体微生物群落^[73]在内的多项微生物群落相关工作(相关工作请参考: <http://www.microbioinformatics.org/publications.html>)。

1.2 微生物组大数据的研究趋势

从微生物群落相关研究的技术要求来讲, 微生物群落大数据分析和数据挖掘的任务数量有指数型增长的趋势。而且在这些微生物群落研究项

目中, 每个微生物群落大数据分析项目的数据量也在增加, 并且在不同角度的数据整合与分析也越来越受到重视^[74]。但相应地, 只是基于单一类型数据(宏基因组、单细胞或代谢物组数据)的数据挖掘就越来越无法满足微生物群落相关研究的需求, 需要对除宏基因组外的单细胞和代谢物组的数据进行整合分析。近年来, 以 16S rRNA 生物标记为基础的分析技术, 例如 MOTHUR^[25]、QIIME^[23]、Parallel-META^[75]等, 大大拓展了微生物群落结构的研究范围, 使得人们能够进一步地认识微生物群落的物种多样性, 但其高保守性和多拷贝性使其应用范围(如定量分析等)受到限制, 需要采用更加优化的生物标记策略或者全基因组测序手段。而且在功能结构方面, 对 DNA 组装、基因预测、基因注释以及代谢通路分析等方法的开发和技术完备的需求也更加迫切。

从微生物组的研究深度和因果性关系的建立角度来讲, 目前国际上正处在从相关性研究到因果性研究的重要阶段^[76]。其中, 肠道中某些微生物的存在, 能够对 5-羟色胺的合成过程产生深远影响。在以小鼠为模型进行的实验中, 研究人员发现, 肠道内“无菌”的小鼠血液中 5-羟色胺的含量要比正常小鼠低了大约 60%^[77]。另外, 基于小鼠试验表明, 饮食磷脂酰胆碱所致的 TMAO 产生取决于肠道微生物的代谢, 而 TMAO 水平增高与主要心脑血管偶发不良事件危险增加相关^[78]。然而, 还有其他很多器官的功能均和人体肠道菌群有关, 但是相关机制却还不是很清晰^[79]。

从微生物组及其相关研究的应用范围来讲, 宏基因组研究手段已经渗透到环境生物监测与治理^[61,80-92]和极端环境^[93]、营养与健康等以利用或克服复杂微生物群落及其产物为目的的科学领

域。在医学领域, 微生物群落对于维护人类健康有着很大的作用, 尤其是在人体口腔环境^[51-52]、肠道及其消化机制^[49-50]、皮肤敏感度^[94]等方面, 而了解人体微生物群落结构与功能的变化有助于把握人类相关健康动态。在生物能源领域, 复杂的生物能源过程如纤维素乙醇的转化与发酵^[95]、沼气的生成^[96]等, 都是依赖于微生物群落的作用而完成。通过宏基因组手段, 能够实现对多种生物过程要素的调控。同时, 在环境治理方面, 宏基因组学技术有助于检测污染物对生态系统的影响以及被污染环境的恢复程度的评估, 并通过生物添加物与生物刺激提高环境恢复的概率, 改进污染清理的策略。另外, 探索土壤中的微生物群落与植物之间的交互作用, 改进氮循环, 从而改善农作物的生长, 也是微生物组技术在农业领域中的重要应用。

1.3 微生物组大数据的研究瓶颈

由于微生物组数据本身的特性, 如群落的结构具有不均衡性、数据的来源具有多样性、尚未形成成熟的质控标准、数据的极端高通量等, 群落全基因信息的获取和生物信息学分析已成为制约宏基因组解析的核心瓶颈。目前宏基因组研究的问题主要反映在以下方面。

目前包括宏基因组在内的组学领域正在进行一场由 Roche454、Solexa 和 Solid 等新一代测序技术所带来的革命^[97-98]。新一代测序技术能够较经济地对基因组进行高倍率的覆盖, 虽然读长较短, 但数据量更多。更重要的是新一代测序技术没有基因组序列读取上的倾向性, 同时碱基也具有较高的正确率^[99]。但是目前基于新一代测序技术分析群落的宏基因组仍然需要较长的时间和较高的费用。

微生物群落的宏基因组数据本身的特性(如其极端高通量、多来源性、多形式存在性等)以及有效利用这些数据的方式(如新生成数据与历史性数据比较、分布式产生而集中性比较、数据的多角度多层次整合分析、信息几乎无限可扩展化等)迫切需要强大的生物信息学分析平台。例如,目前宏基因组的数据来自于 Illumina 和 Solexa 等不同仪器平台。不同来源的数据其处理参数通常差别很大,却都在不同程度上代表了其微生物群落的结构和功能,而不同来源的数据之间往往需要进行相互比较,这就需要通过信息化手段融合处理多个数据源的数据。对算法有全新要求,新一代序列通常较短,由于宏基因组序列包含了多个物种的信息,而且不同物种的丰度存在很大差异,因此,现存的序列分析、功能注释和物种组分分析方法在宏基因组中的物种数量比较大时均不能满足分析要求^[100],需要通过新型数据处理算法的设计来实现高通量宏基因组数据的精确分析。微生物组研究中数据量的爆炸式增长,也伴随着研究者对相关数据进行深入挖掘的渴望^[37],而机器学习或者更为高级的深度学习技术可以非常好地适应这种大数据挖掘研究^[101-102]。然而,由于微生物组和深度学习均是较新的研究领域,微生物组大数据的深入挖掘目前还处在起步阶段。其中机器学习在微生物组学研究中的应用较为成熟,也在血糖水平预测^[103]、菌群与肥胖^[104]、死亡微生物组学^[105]等方面有所应用。更为高阶的深度学习能从非常庞大的数据中挖掘到通常方法不易识别的信息,因此将在微生物组大数据分析中大显身手,解析群落微生物之间的相互关系和对宿主机体(细胞)调控的影响,以及由此引发的对疾病发展的影响。

2 昆虫肠道的宏基因组学

2.1 昆虫肠道菌群宏基因组研究的特点

当前,随着高通量测序技术的成熟和微生物组研究的深入,肠道菌群在昆虫肠道发育和营养研究中的重要性得到了越来越清晰的认识和理解。昆虫肠道微生物组随之受到越来越多的关注:昆虫的物种多样性、环境多样性、食物多样性等特点,造就了昆虫肠道菌群的多样性,以及极为丰富的功能基因库。昆虫肠道菌群是所有栖息于昆虫消化道内的微生物的统称^[106],也是昆虫体内微生物相互作用最集中的群体^[107]。一般而言,肠道微生物主要是由细菌和古菌组成的,其中细菌的种类较为丰富和多样,而且昆虫肠道内微生物也主要以细菌为主。昆虫肠道系统是伴随取食、消化、排泄等活动而多变的环境,其中定殖着大量的微生物,这些微生物和人体所寄生的微生物相似,都被普遍认为是与宿主昆虫的正常生命活动密不可分的^[108]。在昆虫的肠道中含有多种酶系统,它们在维生素合成、脂肪和碳水化合物的吸收和利用中是必不可少的,并且在抵御外来菌种的侵入和定殖、促进免疫系统的功能中也起着重要的作用^[15]。

可以看出,昆虫肠道宏基因组学研究数据样本及其相关科学问题较为丰富,这是由昆虫肠道微生物组样本的丰富及异质性决定的:昆虫多样化和进化的成功,在某种程度上部分依赖于各种有益微生物的贡献^[15],这些微生物与宿主有着密切的关系并参与到宿主生活的很多方面,如昆虫的生理和进化^[109],改善宿主的营养,帮助消化食物成分,还能防止捕食者、寄生虫和病原体入侵,协助种间种内通讯等。因此,相关的菌群宏基因组数据也变得极为丰富和具有重要功能挖掘价

值^[110-113]。对昆虫相关肠道的宏基因组进行深入的数据挖掘,将会形成微生物大数据研究的新疆界。昆虫肠道菌群通常与农业、生态以及医药等多个领域的学科相关^[113]。昆虫作为一个很好的实验模型,通过昆虫建立细菌感染模型来研究人类疾病^[112];此外,还可以利用这种模型来研究共生微生物影响媒介昆虫的疾病传播效率^[111]或疾病的发生规律^[110];同时,还可借助这一模型开展肠道共生菌与宿主相互作用的相关研究,不但有利于昆虫资源的开发利用和对害虫的管制,而且有利于从昆虫肠道这一特殊环境中获得具有特殊功能的细菌资源,用于植物病害的防治,控制动物疫病的传播等,进而促进人类对微生物与其宿主间互惠共生关系的了解^[108]。

除了样本的丰富及异质性,昆虫肠道微生物组本身较于其他肠道微生物组学具有其特殊性。研究发现昆虫的肠道微生物和宿主之间存在着广泛的非致病关系,其中包括互利的关系,甚至有些肠道共生菌群是宿主生长发育、繁殖过程中所必需的^[114-115]。昆虫的肠道系统是一个多变的环境,会伴随取食、消化、排泄等活动而发生改变,使其肠道微生物的结构处于动态变化中。因此,昆虫的肠道微生物组研究也具其特殊的科学问题:昆虫的多样性使得其肠道结构和环境也具有丰富的多样性,而肠道微生物的多样性往往和昆虫肠道特殊的结构和环境因素相关(如等翅目白蚁^[116]),如肠道的pH值、肠道的氧化还原情况、消化酶的存在情况等,其中食料对昆虫肠道微生物也存在一定复杂的影响(例如白蚁^[117]、蟑螂^[118]等)。此外昆虫的变态发育会致使取食发生变化,也会对肠道微生物种类产生影响。肠道微生物也会参与昆虫与宿主的某些复杂的相互关系^[119],例

如,蝗虫肠道中的几种微生物可以利用植物代谢物质合成一些小分子酚类,这些物质是蝗虫聚集信息素的前体,这就意味着蝗虫合成的有些信息素受肠道细菌调节^[120]。

深入地研究昆虫肠道微生物组,将会促进我们对于昆虫肠道菌群更深入的理解。昆虫肠道菌群与宿主之前存在长期协同进化的关系,与宿主的种系关系有关,而且不同代次的宿主之间可以传播微生物。此外,昆虫肠道菌群还具有一些重要的功能,如在宿主营养供给、消化及吸收上起到至关重要的作用。在生理上的影响,果蝇中肠中发现定殖的菌群与上皮细胞的更新速率有关^[121-122]。还可以影响昆虫的寿命,研究表明,果蝇成虫在不同生活阶段时,其肠道共生群的存在可以改变宿主的寿命^[114]。对昆虫的生长发育也有一定影响,共生菌能够通过影响生长率和个体大小的方式影响果蝇的系统发育^[123],研究表明,无菌的果蝇幼虫较正常幼虫表现出增长减慢和发育变缓等现象。在昆虫的生命周期中如果受到寄生虫和致病菌威胁时,还能通过共生菌群的屏蔽作用与疾病进行对抗或参与某些代谢对机体进行保护^[124]。

2.2 昆虫肠道菌群宏基因组研究的主要问题

昆虫肠道菌群宏基因组研究的问题主要有以下几个方面:

2.2.1 肠道菌群的环境适应性:昆虫肠道菌群与宿主共同进化来适应相对较稳定的肠道环境^[125],因此,随着宿主进化以适应不断变化的环境,其肠道中的微生物也参与甚至指导其进化过程,并发挥重要的作用。据证明,蚜虫依靠细菌来提供自身不能制造而且正常情况下无法从饮食获得的必需氨基酸,并且蚜虫和肠道细菌存在共进化关系,在进化过程中新蚜虫物种的出现是反映在其共生

体中物种的形成事件^[126]。因为每一个物种都具有一些必不可少的共同进化的共生体, 具有一些庞大的微生物来适应不断变化的环境。由于环境适应能力有差异, 导致物种分散、隔离从而产生了新的物种。即使是不那么重要的共生体, 也依赖他们的宿主, 帮助并和宿主一起在不断变化的环境中适应新的生态位^[127]。

2.2.2 肠道菌群的动态变化: 昆虫的肠道在不停地蠕动, 食物不断地被消化、吸收与排泄, 肠道菌群也随之被排除体外, 又因为菌群自身的增殖从而维持在一个动态平衡中。但是, 菌群还是会发生一些变化, 如成长发育、季节变化、环境改变等不同因素都会导致菌种一定比例的差异。这些都是正常的生理改变所导致的, 与机体仍然保持着平衡关系, 但是如果是病理因素导致的菌群改变, 就会引起较大幅度的变化, 打破平衡关系。

2.2.3 菌群-宿主关键调控机制: 菌群主要是对宿主的免疫和代谢相关基因有重要的影响。例如, 肠道菌群可以影响肠道分泌蛋白的表达, 从而参与外周器官脂质代谢的调节。此外肠道菌群还是致炎分子的来源之一, 而致炎分子与炎症发生和巨噬细胞的招募都有关系。而肠道菌群的组成很可能受宿主饮食影响^[128-130]。

2.3 昆虫肠道菌群的应用研究

2.3.1 果蝇肠道微生物的相关研究: 果蝇由于生活史短、易饲养、繁殖快、突变型多、遗传资料较丰富并且肠道中的共生菌群简单只有 5-20 种, 大多数都可以在体外培养, 使其成为研究宿主与共生微生物之间相互作用的理想遗传模型^[123,131-132]。果蝇主要以腐烂水果为食, 其肠道含有丰富的发酵微生物, 例如乳酸杆菌和醋酸杆菌, 他们同时也是黑腹果蝇肠道微生物中数量最多的^[123,131-132], 这表

明这些细菌菌株在进化上是适应果蝇提供的特定肠道环境的。并且由于宿主的某些生理活动与其共生菌株相关联, 宿主的发育会受到其较强的影响^[133-134]。例如, 在某些营养限制条件下, 醋酸杆菌数目较多来维持幼虫的生长速率和成虫的体型大小, 这表明特定共生微生物的菌株可以通过影响宿主特定影响条件下的生长速率和体型来影响宿主的系统发育和存活率。我们也因此可以仔细挖掘其中的作用机制, 肠道微生物主要靠其代谢物来影响宿主复杂的生理活动。有关的实例是对斑点果蝇(一种杂食性的入侵昆虫)在入侵新的地区时, 肠道菌群多样性与其丰度之间的研究^[135]。入侵昆虫被广泛认为是对环境质量和人类福利造成重大损害的源头之一。此外, 他们的共生体可能对其入侵行为存在一定影响^[136] (其共生体对宿主的生存和繁殖起作用)。对不同地区的斑点果蝇采样然后进行 DNA 提取、扩增和测序并对其肠道微生物的多样性进行分析得到, 醋酸杆菌科和肠杆菌科在果蝇的成虫和幼虫时期的结构组成存在显著差异, 可以从其入手, 通过改变环境因子来发展病虫害防治策略^[137-138]。

由于果蝇模式动物的特点, 相关基因组学和微生物组学实验手段和操作系统较为成熟, 因此果蝇被越来越多的用于微生物组研究和干预治疗的应用研究中。

(1) 通过对果蝇的整个转录网络进行分析, 来探究微生物菌群对果蝇基因共表达的影响。对所有共表达基因的转录模块进行分析, 找出带有限定菌群和无菌条件下培养的两种果蝇中转录网络的强度存在显著性差异的模块, 并用后续实验证明在依赖微生物模块的基因共表达比定菌果蝇中不依赖微生物的模块和无菌果蝇中的所有模块的

基因共表达强度都要强。这表明微生物的存在会增强基因共表达,从而构建了动物宿主的转录网络^[139]。这一发现为更进一步了解宿主及其常驻微生物群落的共同进化机制有深远的影响。

(2) 了解黑腹果蝇模型中宿主-微生物的相互作用。以肠道定殖菌群种类、性别和年龄三个因素为相互独立的自变量,对黑腹果蝇进行 16S rRNA 的焦磷酸测序来分析比较差异,发现即使使用相同的营养条件饲养,某些菌属种类的相对比例在肠道定殖不同菌种、不同性别、不同年龄的宿主中都存在显著差异。进一步分析得到,宿主肠道中的某些菌属与性别之间存在偏向性,肠道微生物群落随宿主年龄改变的模式与宿主肠道定殖的微生物、宿主的性别有关^[140]。

(3) 有关菌群改造的应用可以利用果蝇模型进行验证。有研究者在埃及伊蚊体内鉴定出一种细胞膜蛋白 Mesh,在肠道共生菌群调控过程中起重要作用,它主要通过 Arresting 介导的 MAPK JNK/ERK 的磷酸化级联反应,引起双氧化酶 Duox 的表达发生变化,导致生成的 ROS 水平变化,最终对肠道微生物群落的增殖进行调控^[141]。研究者之后在果蝇模型中对发现的分子机制进行了全面验证,揭示了重要病媒昆虫肠道共生菌群稳态的调控机制。

(4) 利用果蝇模型对肠道稳定性和微生物定殖概率进行探究。将自然共生体喂养给无菌果蝇个体,虽然细菌剂量很高但仍有果蝇保持肠道无菌的状态,这说明即使在低复杂度的宿主-微生物相互作用模型中,微生物的定殖状态仍然是不唯一的,微生物在肠道中的定殖是有概率的。除此之外,还发现已经定殖的菌群会减少之后其他菌群定殖的概率,促进肠道稳定性,并且肠道多样性的稳定受一些外界因素随机调控,这对于对抗传

染病和在肠道中稳定建立益生菌具有重要意义^[142]。

而利用果蝇肠道微生物群落的宏基因组进行大数据挖掘研究,也将有可能建立更有价值的果蝇肠道菌群模式,并挖掘相关菌群利用途径。

2.3.2 其他昆虫肠道微生物的相关研究: 在应用方面,昆虫的肠道发育和营养深刻地影响着农业生产。传统农业是利用植物和动物资源组成的“二维结构”,将这种二维结构调整调整为植物、动物和微生物的“三维结构”是新型农业的重要战略调整(表2)。在现代农业中,由于化学用品过度施用,造成了种植环境恶化、作物产量下降、营养流失等严重问题,影响到了人类的可持续发展。在农业中引入微生物技术可以有效改善这一弊端,实现经济和环境双赢,微生物不仅可以作为肥料,也可以作为微生物农药、微生物食品,甚至在处理农作物废料(如秸秆等)时,发挥着转化清洁能源的作用。新型农业应当充分利用农业微生物的潜在功能来调整当前的农业结构,改善农业作物的生长环境,充分利用潜在的生物资源^[143-144]。当然,以微生物改造土壤、杀灭害虫、作为肥料等都应当建立在对微生物有充分认识的情况下,而宏基因组学正是我们充分认识、改造、利用微生物的关键点。一旦宏基因组学发展完善,对农业生产中的微生物和菌群有了充分的解析后,这些微生物菌群就可以补充到新型农业的生态位中去,充当肥料,改善土壤。微生物杀虫剂不存在环境污染问题,也不会造成虫害抗性增加,宏基因组学充分保证了生物杀虫剂中菌群的结构和功能,这样才不会使得生物杀虫剂危害人类自身安危^[145-146]。故而微生物组学研究将会成为新型农业的“指导者”,首先充分挖掘有益菌群信息,对菌群实现控制和改造,最后引入农业种植中,形成一种环境友好型的农作物产量提高方式。

表 2. 其他昆虫肠道微生物研究

Table 2. Study of intestinal microflora in other insects

Research object	Purpose	Conclusions	References
Mosquito	Investigate the synergistic anti-malaria mechanism of mosquitoes and their intestinal microflora, for purpose of cutting the malarial transmission from their source	Find new symbiotic bacteria that can sustain continuous intergenerational transmission in Anopheles, thereby effectively driving anti-malaria effector molecules to spread rapidly throughout the mosquito population to break the propagation of malaria from the source	[147]
	Study the cytoplasmic inherited symbiotic microbes that regulate the intestinal reproductivity of mosquitoes to prevent pests by controlling their reproduction	Transplant the newly discovered symbiotic microorganisms <i>Wolbachia</i> into <i>Aedes aegypti</i> , to effectively inhibit <i>Dengue</i> virus infection and shorten the lifespan of female mosquitoes	[148]
		Significantly inhibit the development of malaria parasites in the mid-gut of mosquitoes with the presence of the strains <i>Serratia</i> and <i>Enterococcus</i>	[149]
Termite	Research on the intestinal symbiotic microbiota of the termites, along with attempts to determine the activity of its specific cellulose-decomposition functionality and decipher its genomes	It was found that the functions of reductive acetic acid production and nitrogen fixation has been achieved when the termite intestinal bacteria metabolized cellulose, which was confirmed to have connection with the high expression of spirochete genes.	[150]
Bee	Study the relationship between the presence of the intestinal symbiotic microbiome and the host immune system	Antibiotics treatment leads to maladjustment of the intestinal microflora, and reduces the survival rate of the bees. Moreover, accordingly with the risk of pathogenic bacterial infection in antibiotic bee	[151]
		The presence of certain socially -transmitted gut microbes (eg, beta-proteobacteria) may reduce the <i>Crithidia bombi</i> infection of <i>Bumblebee</i>	[152]
<i>Spodoptera litura</i>	Investigation on the bacterial diversity and function of the fourth instar larvae of <i>Spodoptera litura</i> to find out the relationship between the diversity of intestinal bacteria and dietary patterns, providing new ideas about biological control of the <i>Spodoptera litura</i>	Changes on the diversity of the <i>Spodoptera litura</i> intestinal microbial microflora have influences on many indicators such as relative growth rate, relative food intake, food utilization rate, food conversion rate, and approximate consumption rate	[153]
Silkworm	Explore the effects of fluoride on the silkworm intestinal microflora to develop potential probiotics and increase the resistance of the silkworm to fluoride	Fluoride can change the silkworm intestinal microbial floral diversity and destroy the balance of the intestinal microecology, reducing the activity of digestive enzymes in the intestine and the physiological activity of the silkworm. Besides, studies have shown that certain strains, such as <i>Bacillus</i> and <i>Enterococcus</i> have potential anti-fluoride efficacy	[154]
<i>Bacillus thuringiensis</i> (Bt)	Research on the insecticidal mechanism of <i>Bacillus thuringiensis</i> (mainly refer to Lepidoptera and Coleoptera) and the role of intestinal microbes in this process	<i>Bt</i> infecting insects causes damage to the host's intestinal epithelial cells, through which certain intestinal commensal bacteria (especially <i>Serratia</i> and <i>Clostridium</i>) enter the blood cavity and massively proliferate, thereby weakening the host's immune response and accelerating host's death.	[155]

随着高通量基因组学数据的指数级增长,生物学和医学研究已经进入大数据时代。随之而来的就是微生物科学的革命性进展,开拓了微生物大数据的新疆界,全面了解宿主与微生物以及环境之间的信息交互,为人类医疗、生活等方面都带来巨大的便利。

2.4 昆虫肠道微生物组研究的优势

昆虫种类繁多,生存环境复杂,因此是研究肠道菌群适应性的一类重要载体。与哺乳动物相比,大部分昆虫肠道微生物种类较少,并且一些昆虫肠道会寄生大量特定的微生物,因此可以以昆虫肠道菌群为载体研究其对环境的适应能力和对宿主的影响。以摩门教蟋蟀为例,其肠道微生物主要以乳酸菌科、肠杆菌科和链球菌科为主,但是在不同的肠道区域相对丰度方面存在显著差异,乳酸菌在前肠较为常见而肠杆菌更倾向于出现在后肠。它们都是通过改善营养,来增强对病原体的抵抗和调节社会行为。并且利用 PICRUSt^[47]来预测 16S rRNA 序列中的基因含量,可以发现这些菌群能合成参与碳水化合物代谢和防御病原体的酶,这些也表明了宿主的肠道系统是营养和病原体防御比较重要的场所^[156]。

昆虫具有独特的生长过程和形态,因此是研究肠道菌群动态变化的良好对象。由于一些昆虫在生长发育过程中会发生变态发育,形态会发生较大的变化,随之生活环境和饮食习惯也会发生较大变化,这些都会引起肠道菌群较大程度的改变。家蚕是常见的变态发育昆虫,其幼虫在发育中会经历多次蜕皮,每次蜕皮过程中前后肠覆盖的外骨骼以及中肠的围食膜也会一起蜕下,其中定殖的微生物也一并被去除,然后肠道微生物会重新形成,群落结构会发生巨大改变^[15]。然后基

于 16S rRNA 序列进行分析,发现家蚕肠道中肠球菌属为优势菌群,并且其多样性的降低会容易引起血液型脓病^[157]的爆发,可以通过改变食料喂养条件来减少蚕病发生^[158],提高经济效益。

2.5 昆虫肠道微生物组研究的瓶颈

数据的整合与数据库的缺乏表现在昆虫分布广泛,取食方式和食物种类多样,其肠道内含有数以万计的微生物,数量庞大,种类丰富,关系复杂。此外,某些昆虫由于完全变态发育会使生活习性和食物来源发生变化,导致肠道共生菌群结构发生变化,这些都是他们复杂多样化的肠道微生物的形成原因,都为数据的整合带来困难。而且现阶段仍缺乏更加准确、完善的测序技术、分析技术等。

在进化和生态研究方面,由于微生物生长周期较短,遗传物质的交换频繁,微生物群落表现出一定的遗传灵活性和可塑性,易产生突变导致进化,形成了一种快速进化的模式。在不同的环境选择压力下,产生新功能基因的菌群会迅速扩散,趋向于形成一个新的分布模式,这就使得整个群落的遗传稳定性相对较差,这同样也是当前微生物分类系统还不完善的原因之一。微生物为了适应不同的生态环境,会采取相应的较为灵活的代谢方式和适应策略,这同样也增加了微生物代谢的多样性和复杂性,给微生物研究带来了一些困难。

在具体研究的操作层面,昆虫肠道微生物种类丰富,多样性高,使得数据整合的工作量相对较大,而且由于分离培养技术、测序技术等局限性,物种的鉴定受到阻碍,不利于数据的整合。此外,由于受到环境等因素的影响,同一物种不同时期肠道微生物的结构会发生一定的改变,这

使数据的复杂度大大增加,也对数据库的建立造成了一定的阻碍。

另一方面,数据分析与挖掘方法缺乏,迄今为止,人们对共生微生物之间复杂的相互关系和其在昆虫生理活动中所起的具体作用和机制仍停留在初级阶段。这一方面是由于涉及到大量的代谢、免疫途径,数据庞大;另一方面,是因为对数据的分析不够全面,缺乏能够从庞大数据集中抽提出有用信息的方法。

3 总结与展望

综上所述,昆虫肠道宏基因组的研究,是生物资源利用方面的宝库。针对昆虫肠道宏基因组进行广泛而深入的研究,将会极大地促进人类对于昆虫肠道菌群的理解,进而促进相关资源的利用和生态环境保护。而宏基因组生物大数据的挖掘,是昆虫肠道菌群研究中十分重要的一个部分:深入的肠道菌群大数据挖掘,将会揭示基于小数据难以发掘的深刻生态和进化规律,进而有助于理性设计相关应用。

同时,现有昆虫肠道宏基因组研究还处于初级阶段。有关昆虫肠道微生物多样性及功能的研究已有相关报道,如家蚕、白蚁、蜜蜂、蟑螂等,但是对共生微生物彼此复杂的相互关系以及倡导菌群对宿主生理活动所起的具体作用和机制尚不十分确切。有关研究应主要从以下几个方面着手:首先是通过昆虫肠道菌群的研究来探索新的害虫防治方法,可以通过了解昆虫与其共生菌群之间的复杂关系——在昆虫生理、发育、进化中起到的菌体作用,以及相应的机制然后研制、应用微生态制剂,影响宿主生理活动来对微生物资源进行发掘和利用。其次可以着眼于经济类昆虫的

产量提高,如家蚕、虫药等,通过强化某些菌种使其更好地在昆虫体内定殖,来提高昆虫对饲料的消化利用率或对环境的适应性,促进其生长,降低养殖成本。简言之,应用方向就是理解其互作关系来抑制有害昆虫带来的消极影响,以及促进经济类昆虫生长,研制、应用微生态制剂。

昆虫肠道微生物组的研究为微生物组大数据研究提供了机遇,同时也提出了挑战。肠道菌群的研究仍停留在初级阶段,群落全集因信息的获取和生物信息学分析仍是核心瓶颈,需要更高效低成本、读长更长且准确率相对较高的测序仪(如三代测序),更加通用的交互式分析平台,在物种数量较大时仍能满足功能注释和物种组分分析要求的算法。昆虫肠道微生物组研究的逐步发展会伴随着数据量爆炸式的增长,这就需要更高级的深度学习从庞大的数据集中挖掘出传统方法无法识别的信息,解析群落微生物与宿主之间的复杂关系以及对宿主机体的调控和影响,由此研究发现更多未知的功能,挖掘更多微生物资源的潜在价值,促进微生物产业的开发,为人类创造更宝贵的财富。

参考文献

- [1] Suau A, Bonnet R, Sutren M, Godon JJ, Gibson GR, Collins MD, Doré J. Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. *Applied and Environmental Microbiology*, 1999, 65(11): 4799–4807.
- [2] Furrer E. A molecular revolution in the study of intestinal microflora. *Gut*, 2006, 55(2): 141–143.
- [3] Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N, Sorek R, Tringe SG, Podar M, Martin HG, Kunin V, Dalevi D, Madejska J, Kirton E, Platt D, Szeto E, Salamov A, Barry K, Mikhailova N, Kyrpidis NC, Matson EG, Ottesen EA, Zhang XN, Hernández M, Murillo C, Acosta LG, Rigoutsos I, Tamayo G, Green BD, Chang C,

- Rubin EM, Mathur EJ, Robertson DE, Hugenholtz P, Leadbetter JR. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature*, 2007, 450(7169): 560–565.
- [4] Mou XZ, Sun SL, Edwards RA, Hodson RE, Moran MA. Bacterial carbon processing by generalist species in the coastal ocean. *Nature*, 2008, 451(7179): 708–711.
- [5] Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman DA. The application of ecological theory toward an understanding of the human microbiome. *Science*, 2012, 336(6086): 1255–1262.
- [6] Chen YC, Lin CJ, Jones G, Fu SY, Zhan HY. Application of statistical design for the optimization of microbial community of synthetic domestic wastewater. *Biodegradation*, 2011, 22(1): 205–213.
- [7] Price PB. Microbial life in glacial ice and implications for a cold origin of life. *FEMS Microbiology Ecology*, 2007, 59(2): 217–231.
- [8] Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences of the United States of America*, 1998, 95(12): 6578–6583.
- [9] Sekirov I, Russell SL, Antunes LC, Finlay BB. Gut microbiota in health and disease. *Physiological Reviews*, 2010, 90(3): 859–904.
- [10] Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA. Diversity of the human intestinal microbial flora. *Science*, 2005, 308(5728): 1635–1638.
- [11] Schreiber S, Rosenstiel P, Albrecht M, Hampe J, Krawczak M. Genetics of Crohn disease, an archetypal inflammatory barrier disease. *Nature Reviews Genetics*, 2005, 6(5): 376–388.
- [12] Basset Y, Cizek L, Cuénoud P, Didham RK, Guilhaumon F, Missa O, Novotny V, Ødegaard F, Roslin T, Schmidl J, Tishechkin AK, Winchester NN, Roubik DW, Aberlenc HP, Bail J, Barrios H, Bridle JR, Castaño-Meneses G, Corbara B, Curletti G, Duarte da Rocha W, De Bakker D, Delabie JHC, Dejean A, Fagan LL, Floren A, Kitching RL, Medianero E, Miller SE, Gama de Oliveira E, Orivel J, Pollet M, Rapp M, Ribeiro SP, Roisin Y, Schmidt JB, Sørensen L, Leponce M. Arthropod diversity in a tropical forest. *Science*, 2012, 338(6113): 1481–1484.
- [13] Shi WB, Syrenne R, Sun JZ, Yuan JS. Molecular approaches to study the insect gut symbiotic microbiota at the ‘omics’ age. *Insect Science*, 2010, 17(3): 199–219.
- [14] Chapman RF, Simpson SJ, Douglas AE. The insects: structure and function. 5th ed. Cambridge: Cambridge University Press, 929.
- [15] Engel P, Moran NA. The gut microbiota of insects - diversity in structure and function. *FEMS Microbiology Reviews*, 2013, 37(5): 699–735.
- [16] DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, 2006, 72(7): 5069–5072.
- [17] Cole JR, Chai B, Marsh TL, Farris RJ, Wang Q, Kulam SA, Chandra S, McGarrell DM, Schmidt TM, Garrity GM, Tiedje JM. The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Research*, 2003, 31(1): 442–443.
- [18] Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig WG, Peplies J, Glöckner FO. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, 2007, 35(21): 7188–7196.
- [19] Ye Y, Doak TG. A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLoS Computational Biology*, 2009, 5(8): e1000465.
- [20] Seshadri R, Kravitz SA, Smarr L, Gilna P, Frazier M. CAMERA: a community resource for metagenomics. *PLoS Biology*, 2007, 5(3): e75.
- [21] Hunter S, Corbett M, Denise H, Fraser M, Gonzalez-Beltran A, Hunter C, Jones P, Leinonen R, McAnulla C, Maguire E, Maslen J, Mitchell A, Nuka G, Oisel A, Pesseat S, Radhakrishnan R, Rocca-Serra P, Scheremetjew M, Sterk P, Vaughan D, Cochrane G, Field D, Sansone SA. EBI metagenomics - a new resource for the analysis and archiving of metagenomic data. *Nucleic Acids Research*, 2014, 42(D1): D600–D606.
- [22] Shah N, Tang H, Doak TG, Ye Y. Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics. *Pacific Symposium on Biocomputing*, 2005, 16(6): 165–176.
- [23] Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenkov T, Zaneveld J, Knight R. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 2010, 7(5): 335–336.
- [24] Meyer F, Paarmann D, D’Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA. The metagenomics RAST server -

- a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, 2008, 9: 386.
- [25] Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, van Horn DJ, Weber CF. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, 2009, 75(23): 7537–7541.
- [26] Zhou Q, Su XQ, Jing GC, Ning K. Meta-QC-Chain: comprehensive and fast quality control method for metagenomic data. *Genomics, Proteomics & Bioinformatics*, 2014, 12(1): 52–56.
- [27] Krause L, Diaz NN, Goesmann A, Kelley S, Nattkemper TW, Rohwer F, Edwards RA, Stoye J. Phylogenetic classification of short environmental DNA fragments. *Nucleic Acids Research*, 2008, 36(7): 2230–2239.
- [28] Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, 2005, 71(12): 8228–8235.
- [29] Monzoorul Haque M, Ghosh TS, Komanduri D, Mande SS. SOrt-ITEMS: sequence orthology based approach for improved taxonomic estimation of metagenomic sequences. *Bioinformatics*, 2009, 25(14): 1722–1730.
- [30] Glass EM, Wilkening J, Wilke A, Antonopoulos D, Meyer F. Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harbor Protocols*, 2010, 2010(1): pdb.prot5368.
- [31] Merelli I, Viti F, Milanesi L. IBDsite: a galaxy-interacting, integrative database for supporting inflammatory bowel disease high throughput data analysis. *BMC Bioinformatics*, 2012, 13 Suppl 14: S5.
- [32] Chen T, Yu WH, Izard J, Baranova OV, Lakshmanan A, Dewhirst FE. The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database (Oxford)*, 2010, 2010: baq013.
- [33] Luo CW, Knight R, Siljander H, Knip M, Xavier RJ, Gevers D. ConStrains identifies microbial strains in metagenomic datasets. *Nature Biotechnology*, 2015, 33(10): 1045–1052.
- [34] Nayfach S, Rodriguez-Mueller B, Garud N, Pollard KS. An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. *Genome Research*, 2016, 26(11): 1612–1625.
- [35] Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods*, 2012, 9(8): 811–814.
- [36] Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, Tett A, Huttenhower C, Segata N. MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nature Methods*, 2015, 12(10): 902–903.
- [37] Hunter CI, Mitchell A, Jones P, McAnulla C, Pesseat S, Scheremetjew M, Hunter S. Metagenomic analysis: the challenge of the data bonanza. *Briefings in Bioinformatics*, 2012, 13(6): 743–746.
- [38] Blin K, Medema MH, Kottmann R, Lee SY, Weber T. The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters. *Nucleic Acids Research*, 2017, 45(D1): D555–D559.
- [39] Medema MH, Cimermančič P, Sali A, Takano E, Fischbach MA. A systematic computational analysis of biosynthetic gene cluster evolution: lessons for engineering biosynthesis. *PLoS Computational Biology*, 2014, 10(12): e1004016.
- [40] Weber T, Blin K, Duddela S, Krug D, Kim HU, Brucoleri R, Lee SY, Fischbach MA, Müller R, Wohlleben W, Breitling R, Takano E, Medema MH. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research*, 2015, 43(W1): W237–W243.
- [41] Hu YF, Yang X, Qin JJ, Lu N, Cheng G, Wu N, Pan YL, Li J, Zhu LY, Wang X, Meng ZQ, Zhao FQ, Liu D, Ma JC, Qin N, Xiang CS, Xiao YH, Li LJ, Yang HM, Wang J, Yang RF, Gao GF, Wang J, Zhu BL. Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. *Nature Communications*, 2013, 4: 2151.
- [42] Ziemert N, Podell S, Penn K, Badger JH, Allen E, Jensen PR. The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS One*, 2012, 7(3): e34064.
- [43] Ichikawa N, Sasagawa M, Yamamoto M, Komaki H, Yoshida Y, Yamazaki S, Fujita N. DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters. *Nucleic Acids Research*, 2013, 41(Database issue): D408–D414.
- [44] Conway KR, Boddy CN. ClusterMine360: a database of microbial PKS/NRPS biosynthesis. *Nucleic Acids Research*, 2013, 41(Database issue): D402–D407.
- [45] Hadjithomas M, Chen IMA, Chu K, Ratner A, Palaniappan K, Szeto E, Huang JH, Reddy TBK, Cimermančič P, Fischbach MA, Ivanova NN, Markowitz VM, Kyrpidis NC, Pati A. IMG-ABC: a knowledge base to fuel discovery of biosynthetic gene clusters and novel secondary metabolites. *mBio*, 2015, 6(4): e00932–15.
- [46] Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. *Genome Research*, 2007, 17(3): 377–386.
- [47] Langille MGI, Zaneveld J, Caporaso JG, McDonald D,

- Knights D, Reyes JA, Clemente JC, Burkpile DE, Thurber RLV, Knight R, Beiko RG, Huttenhower C. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology*, 2013, 31(9): 814–821.
- [48] Lawley B, Tannock GW. Analysis of 16S rRNA gene amplicon sequences using the QIIME software package. *Methods in Molecular Biology (Clifton, N.J.)*, 2017, 1537: 153–163.
- [49] Jumpertz R, Le DS, Turnbaugh PJ, Trinidad C, Bogardus C, Gordon JI, Krakoff J. Energy-balance studies reveal associations between gut microbes, caloric load, and nutrient absorption in humans. *The American Journal of Clinical Nutrition*, 2011, 94(1): 58–65.
- [50] Lozupone CA, Hamady M, Cantarel BL, Coutinho PM, Henrissat B, Gordon JI, Knight R. The convergence of carbohydrate active gene repertoires in human gut microbes. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105(39): 15076–15081.
- [51] Cephas KD, Kim J, Mathai RA, Barry KA, Dowd SE, Meline BS, Swanson KS. Comparative analysis of salivary bacterial microbiome diversity in edentulous infants and their mothers or primary care givers using pyrosequencing. *PLoS One*, 2011, 6(8): e23503.
- [52] Yang F, Zeng XW, Ning K, Liu KL, Lo CC, Wang W, Chen J, Wang DM, Huang RR, Chang XZ, Chain PS, Xie G, Ling JQ, JianXu. Saliva microbiomes distinguish caries-active from healthy human populations. *The ISME Journal*, 2012, 6(1): 1–10.
- [53] Xia WW, Zhang CX, Zeng XW, Feng YZ, Weng JH, Lin XG, Zhu JG, Xiong ZQ, Xu J, Cai ZC, Jia ZJ. Autotrophic growth of nitrifying community in an agricultural soil. *The ISME Journal*, 2011, 5(7): 1226–1236.
- [54] Shapira M. Gut microbiotas and host evolution: scaling up symbiosis. *Trends in Ecology & Evolution*, 2016, 31(7): 539–549.
- [55] Moeller AH, Caro-Quintero A, Mjungu D, Georgiev AV, Lonsdorf EV, Muller MN, Pusey AE, Peeters M, Hahn BH, Ochman H. Cospeciation of gut microbiota with hominids. *Science*, 2016, 353(6297): 380–382.
- [56] Segre JA, Salafsky N. EVOLUTION. Hominid superorganisms. *Science*, 2016, 353(6297): 350–351.
- [57] Filyk HA, Osborne LC. The multibiome: the intestinal ecosystem's influence on immune homeostasis, health, and disease. *Ebiomedicine*, 2016, 13: 46–54.
- [58] Chen YW, Li ZC, Hu SF, Zhang J, Wu JQ, Shao NS, Bo XC, Ni M, Ying XM. Gut metagenomes of type 2 diabetic patients have characteristic single-nucleotide polymorphism distribution in *Bacteroides coprocola*. *Microbiome*, 2017, 5(1): 15.
- [59] Chacon JM, Harcombe WR. Antimicrobials: constraints on microbial warfare. *Nature Microbiology*, 2016, 1: 16225.
- [60] Ji PF, Zhang YM, Wang JF, Zhao FQ. MetaSort untangles metagenome assembly by reducing microbial community complexity. *Nature Communications*, 2017, 8: 14306.
- [61] Schirmer M, Smeekens SP, Vlamakis H, Jaeger M, Oosting M, Franzosa EA, ter Horst R, Jansen T, Jacobs L, Bonder MJ, Kurilshikov A, Fu JY, Joosten LAB, Zhernakova A, Huttenhower C, Wijmenga C, Netea MG, Xavier RJ. Linking the human gut microbiome to inflammatory cytokine production capacity. *Cell*, 2016, 167(4): 1125–1136.e8.
- [62] Guo CJ, Chang FY, Wyche TP, Backus KM, Acker TM, Funabashi M, Taketani M, Donia MS, Nayfach S, Pollard KS, Craik CS, Cravatt BF, Clardy J, Voigt CA, Fischbach MA. Discovery of reactive microbiota-derived metabolites that inhibit host proteases. *Cell*, 2017, 168(3): 517–526.e18.
- [63] Layeghifard M, Hwang DM, Guttman DS. Disentangling interactions in the microbiome: a network perspective. *Trends in Microbiology*, 2017, 25(3): 217–228.
- [64] Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, Huttenhower C. Microbial Co-occurrence relationships in the human microbiome. *PLoS Computational Biology*, 2012, 8(7): e1002606.
- [65] Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, Cornejo-Castillo FM, Costea PI, Cruaud C, d'Ovidio F, Engelen S, Ferrera I, Gasol JM, Guidi L, Hildebrand F, Kokoszka F, Lepoivre C, Lima-Mendez G, Poulain J, Poulos BT, Royo-Llonch M, Sarmiento H, Vieira-Silva S, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Tara Oceans Coordinators, Bowler C, de Vargas C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Jaillon O, Not F, Ogata H, Pesant S, Speich S, Stemmann L, Sullivan MB, Weissenbach J, Wincker P, Karsenti E, Raes J, Acinas SG, Bork P. Ocean plankton. Structure and function of the global ocean microbiome. *Science*, 2015, 348(6237): 1261359.
- [66] Lima-Mendez G, Faust K, Henry N, Decelle J, Colin S, Carcillo F, Chaffron S, Ignacio-Espinosa JC, Roux S, Vincent F, Bittner L, Darzi Y, Wang J, Audic S, Berline L, Bontempi G, Cabello AM, Coppola L, Cornejo-Castillo FM, d'Ovidio F, De Meester L, Ferrera I, Garet-Delmas MJ, Guidi L, Lara E, Pesant S, Royo-Llonch M, Salazar G, Sánchez P, Sebastian M, Souffreau C, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Tara Oceans Coordinators, Gorsky G, Not F, Ogata H, Speich S, Stemmann L, Weissenbach J, Wincker P, Acinas SG, Sunagawa S, Bork P, Sullivan MB, Karsenti E, Bowler C, de

- Vargas C, Raes J. Ocean plankton. Determinants of community structure in the global plankton interactome. *Science*, 2015, 348(6237): 1262073.
- [67] Vázquez-Baeza Y, Gonzalez A, Smarr L, McDonald D, Morton JT, Navas-Molina JA, Knight R. Bringing the dynamic microbiome to life with animations. *Cell Host & Microbe*, 2017, 21(1): 7–10.
- [68] Su X, Xu J, Ning K. Parallel-META: efficient metagenomic data analysis based on high-performance computation. *BMC Systems Biology*, 2012, 6 Suppl 1: S16.
- [69] Su XQ, Xu J, Ning K. Meta-Storms: efficient search for similar microbial communities based on a novel indexing scheme and similarity score for metagenomic data. *Bioinformatics*, 2012, 28(19): 2493–2501.
- [70] Su XQ, Hu JQ, Huang S, Ning K. Rapid comparison and correlation analysis among massive number of microbial community samples based on MDV data model. *Scientific Reports*, 2014, 4: 6393.
- [71] Han MZ, Gong YH, Zhou CY, Zhang JQ, Wang Z, Ning K. Comparison and interpretation of taxonomical structure of bacterial communities in two types of lakes on Yun-Gui plateau of China. *Scientific Reports*, 2016, 6: 30616.
- [72] Su XQ, Song BX, Wang XT, Ma XL, Xu J, Ning K. Meta-Mesh: metagenomic data analysis system. *Chinese Journal of Biotechnology*, 2014, 30(1): 6–17. (in Chinese)
苏晓泉, 宋宝兴, 王雪涛, 马新乐, 徐健, 宁康. Meta-Mesh——元基因组数据分析系统. *生物工程学报*, 2014, 30(1): 6–17.
- [73] Su XQ, Wang XJ, Jing GC, Huang S, Xu J, Ning K. Application of Meta-Mesh on the analysis of microbial communities from human associated-habitats. *Quantitative Biology*, 2015, 3(1): 4–18.
- [74] Yang B. The genome war: how Craig Venter tried to capture the code of life and save the world. *Discovery Medicine*, 2004, 4(21):84.
- [75] Su XQ, Xu J, Ning K. Parallel-META: efficient metagenomic data analysis based on high-performance computation. *BMC Systems Biology*, 2012, 6 Suppl 1: S16.
- [76] Fritz JV, Desai MS, Shah P, Schneider JG, Wilmes P. From meta-omics to causality: experimental models for human microbiome research. *Microbiome*, 2013, 1(1): 14.
- [77] Yano JM, Yu K, Donaldson GP, Shastri GG, Ann P, Ma L, Nagler CR, Ismagilov RF, Mazmanian SK, Hsiao EY. Indigenous bacteria from the gut microbiota regulate host serotonin biosynthesis. *Cell*, 2015, 161(2): 264–276.
- [78] Zhu WF, Gregory JC, Org E, Buffa JA, Gupta N, Wang ZN, Li L, Fu XM, Wu YP, Mehrabian M, Sartor RB, McIntyre TM, Silverstein RL, Tang WHW, DiDonato JA, Brown JM, Lulis AJ, Hazen SL. Gut microbial metabolite TMAO enhances platelet hyperreactivity and thrombosis risk. *Cell*, 2016, 165(1): 111–124.
- [79] Schroeder B, Backhed F. Signals from the gut microbiota to distant organs in physiology and disease. *Nature Medicine*, 2016, 22(10): 1079–1089.
- [80] Planer JD, Peng YQ, Kau AL, Blanton LV, Ndao IM, Tarr PI, Warner BB, Gordon JI. Development of the gut microbiota and mucosal IgA responses in twins and gnotobiotic mice. *Nature*, 2016, 534(7606): 263–266.
- [81] Ganai SC, Sanos SL, Kallfass C, Oberle K, Johner C, Kirschning C, Lienenklaus S, Weiss S, Staeheli P, Aichele P, Diefenbach A. Priming of natural killer cells by nonmucosal mononuclear phagocytes requires instructive signals from commensal microbiota. *Immunity*, 2012, 37(1): 171–186.
- [82] Clemente JC, Ursell LK, Parfrey LW, Knight R. The impact of the gut microbiota on Human health: an integrative view. *Cell*, 2012, 148(6): 1258–1270.
- [83] Johnson CH, Spilker ME, Goetz L, Peterson SN, Siuzdak G. Metabolite and microbiome interplay in cancer immunotherapy. *Cancer Research*, 2016, 76(21): 6146–6152.
- [84] Tringe SG, Zhang T, Liu XG, Yu YT, Lee WH, Yap J, Yao F, Suan ST, Ing SK, Haynes M, Rohwer F, Wei CL, Tan P, Bristow J, Rubin EM, Ruan YJ. The airborne metagenome in an indoor urban environment. *PLoS One*, 2008, 3(4): e1862.
- [85] Song BK, Palleroni NJ, Kerkhof LJ, Häggblom MM. Characterization of halobenzoate-degrading, denitrifying *Azoarcus* and *Thauera* isolates and description of *Thauera chlorobenzoica* sp. nov.. *International Journal of Systematic and Evolutionary Microbiology*, 2001, 51(2): 589–602.
- [86] Song BK, Palleroni NJ, Häggblom MM. Isolation and characterization of diverse halobenzoate-degrading denitrifying bacteria from soils and sediments. *Applied and Environmental Microbiology*, 2000, 66(8): 3446–3453.
- [87] Rabus R. Functional genomics of an anaerobic aromatic-degrading denitrifying bacterium, strain EbN1. *Applied Microbiology and Biotechnology*, 2005, 68(5): 580–587.
- [88] Narihiro T, Sekiguchi Y. Microbial communities in anaerobic digestion processes for waste and wastewater treatment: a microbiological update. *Current Opinion in Biotechnology*, 2007, 18(3): 273–278.
- [89] Harms G, Rabus R, Widdel F. Anaerobic oxidation of the aromatic plant hydrocarbon p-cymene by newly isolated denitrifying bacteria. *Archives of Microbiology*, 1999, 172(5): 303–312.
- [90] Etchebehere C, Errazquin I, Barrandeguy E, Dabert P, Moletta R, Muxí L. Evaluation of the denitrifying microbiota of

- anoxic reactors. *FEMS Microbiology Ecology*, 2001, 35(3): 259–265.
- [91] Etchebehere C, Cabezas A, Dabert P, Muxi L. Evolution of the bacterial community during granules formation in denitrifying reactors followed by molecular, culture-independent techniques. *Water Science and Technology*, 2003, 48(6): 75–79.
- [92] Rabus R, Widdel F. Anaerobic degradation of ethylbenzene and other aromatic hydrocarbons by new denitrifying bacteria. *Archives of Microbiology*, 1995, 163(2): 96–103.
- [93] Gupta R, Beg QK, Lorenz P. Bacterial alkaline proteases: molecular approaches and industrial applications. *Applied Microbiology and Biotechnology*, 2002, 59(1): 15–32.
- [94] Kong HH. Skin microbiome: genomics-based insights into the diversity and role of skin microbes. *Trends in Molecular Medicine*, 2011, 17(6): 320–328.
- [95] Vasudevan D, Richter H, Angenent LT. Upgrading dilute ethanol from syngas fermentation to *n*-caproate with reactor microbiomes. *Bioresource Technology*, 2014, 151: 378–382.
- [96] Shi WB, Moon CD, Leahy SC, Kang DW, Froula J, Kittelmann S, Fan C, Deutsch S, Gagic D, Seedorf H, Kelly WJ, Atua R, Sang C, Soni P, Li D, Pinares-Patino CS, McEwan JC, Janssen PH, Chen F, Visel A, Wang Z, Attwood GT, Rubin EM. Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. *Genome Research*, 2014, 24(9): 1517–1525.
- [97] Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen ZT, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu PG, Begley RF, Rothberg JM. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 2005, 437(7057): 376–380.
- [98] Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, 2008, 24(3): 133–141.
- [99] Warren RL, Sutton GG, Jones SJM, Holt RA. Assembling millions of short DNA sequences using SSAKE. *Bioinformatics*, 2007, 23(4): 500–501.
- [100] Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, 2006, 444(7122): 1027–1031.
- [101] Mamoshina P, Vieira A, Putin E, Zhavoronkov A. Applications of deep learning in biomedicine. *Molecular Pharmaceutics*, 2016, 13(5): 1445–1454.
- [102] Angermueller C, Parnamaa T, Parts L, Stegle O. Deep learning for computational biology. *Molecular Systems Biology*, 2016, 12(7): 878.
- [103] Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, Ben-Yacov O, Lador D, Avnit-Sagi T, Lotan-Pompan M, Suez J, Mahdi JA, Matot E, Malka G, Kosower N, Rein M, Zilberman-Schapira G, Dohnalová L, Pevsner-Fischer M, Bikovsky R, Halpern Z, Elinav E, Segal E. Personalized nutrition by prediction of glycemic responses. *Cell*, 2015, 163(5): 1079–1094.
- [104] Sze MA, Schloss PD. Looking for a signal in the noise: revisiting obesity and the microbiome. *mBio*, 2016, 7(4): e01018–16.
- [105] Johnson HR, Trinidad DD, Guzman S, Khan Z, Parziale JV, DeBruyn JM, Lents NH. A Machine learning approach for using the postmortem skin microbiome to estimate the postmortem interval. *PLoS One*, 2016, 11(12): e0167370.
- [106] Rangberg A, Diep DB, Rudi K, Amdam GV. Paratransgenesis: an approach to improve colony health and molecular insight in honey bees (*Apis mellifera*)? *Integrative and Comparative Biology*, 2012, 52(1): 89–99.
- [107] Cummings JH, Macfarlane GT. Role of intestinal bacteria in nutrient metabolism. *Clinical Nutrition*, 1997, 16(1): 3–11.
- [108] Lemaitre B, Hoffmann J. The host defense of *Drosophila melanogaster*. *Annual Review of Immunology*, 2007, 25: 697–743.
- [109] Crotti E, Balloi A, Hamdi C, Sansonno L, Marzorati M, Gonella E, Favia G, Cherif A, Bandi C, Alma A, Daffonchio D. Microbial symbionts: a resource for the management of insect-related problems. *Microbial Biotechnology*, 2012, 5(3): 307–317.
- [110] Chouaia B, Rossi P, Epis S, Mosca M, Ricci I, Damiani C, Ulissi U, Crotti E, Daffonchio D, Bandi C, Favia G. Delayed larval development in *Anopheles* mosquitoes deprived of *Asaia* bacterial symbionts. *BMC Microbiology*, 2012, 12(S1): S2.
- [111] Ricci I, Valzano M, Ulissi U, Epis S, Cappelli A, Favia G. Symbiotic control of mosquito borne disease. *Pathogens and Global Health*, 2012, 106(7): 380–385.
- [112] Ishii K, Hamamoto H, Sekimizu K. Establishment of a bacterial infection model using the european honeybee, *Apis mellifera* L. *PLoS One*, 2014, 9(2): e89917.
- [113] Engel P, Moran NA. Functional and evolutionary insights into the simple yet specific gut microbiota of the honey bee from metagenomic analysis. *Gut Microbes*, 2013, 4(1): 60–65.

- [114] Brummel T, Ching A, Seroude L, Simon AF, Benzer S. *Drosophila* lifespan enhancement by exogenous bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101(35): 12974–12979.
- [115] Dillon R, Charnley K. Mutualism between the desert locust *Schistocerca gregaria* and its gut microbiota. *Research in Microbiology*, 2002, 153(8): 503–509.
- [116] Schmitt-Wagner D, Friedrich MW, Wagner B, Brune A. Phylogenetic diversity, abundance, and axial distribution of bacteria in the intestinal tract of two soil-feeding termites (*Cubitermes* spp.). *Applied and Environmental Microbiology*, 2003, 69(10): 6007–6017.
- [117] Tanaka H, Aoyagi H, Shiina S, Doudou Y, Yoshimura T, Nakamura R, Uchiyama H. Influence of the diet components on the symbiotic microorganisms community in hindgut of *Coptotermes formosanus* Shiraki. *Applied Microbiology and Biotechnology*, 2006, 71(6): 970–970.
- [118] Santo Domingo JW, Berry CJ, Summer M, Fliermans CB. Microbiology of spent nuclear fuel storage basins. *Current Microbiology*, 1998, 37(6): 387–394.
- [119] Molyneux RJ, Campbell BC, Dreyer DL. Honeydew analysis for detecting phloem transport of plant natural products: implications for host-plant resistance to sap-sucking insects. *Journal of Chemical Ecology*, 1990, 16(6): 1899–1909.
- [120] Dillon RJ, Vennard CT, Charnley AK. A note: gut bacteria produce components of a locust cohesion pheromone. *Journal of Applied Microbiology*, 2002, 92(4): 759–763.
- [121] Buchon N, Broderick NA, Poidevin M, Pradervand S, Lemaitre B. *Drosophila* intestinal response to bacterial infection: activation of host defense and stem cell proliferation. *Cell Host & Microbe*, 2009, 5(2): 200–211.
- [122] Casali A, Battle E. Intestinal stem cells in mammals and *Drosophila*. *Cell Stem Cell*, 2009, 4(2): 124–127.
- [123] Shin SC, Kim SH, You H, Kim B, Kim AC, Lee KA, Yoon JH, Ryu JH, Lee WJ. *Drosophila* microbiome modulates host developmental and metabolic homeostasis via insulin signaling. *Science*, 2011, 334(6056): 670–674.
- [124] Li JL, Qin HR, Wu J, Sadd BM, Wang XH, Evans JD, Peng WJ, Chen YP. The prevalence of parasites and pathogens in Asian honeybees *Apis cerana* in China. *PLoS One*, 2012, 7(11): e47955.
- [125] Ohkuma M, Sato T, Noda S, Ui S, Kudo T, Hongoh Y. The candidate phylum ‘Termite Group 1’ of bacteria: phylogenetic diversity, distribution, and endosymbiont members of various gut flagellated protists. *FEMS Microbiology Ecology*, 2007, 60(3): 467–476.
- [126] De Clerck C, Fujiwara A, Joncour P, Léonard S, Félix ML, Francis F, Jijakli MH, Tsuchida T, Massart S. A metagenomic approach from aphid’s hemolymph sheds light on the potential roles of co-existing endosymbionts. *Microbiome*, 2015, 3: 63.
- [127] Baudouin-Cornu P, Surdin-Kerjan Y, Marlière P, and Thomas D, Environmental disruption of host-microbe co-adaptation as a potential driving force in evolution. *Front Genet*, 2014, 5(5528): 168.
- [128] Hildebrandt MA, Hoffmann C, Sherrill-Mix SA, Keilbaugh SA, Hamady M, Chen YY, Knight R, Ahima RS, Bushman F, Wu GD. High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology*, 2009, 137(5): 1716–1724.e1-2.
- [129] Walker AW, Ince J, Duncan SH, Webster LM, Holtrop G, Ze XL, Brown D, Stares MD, Scott P, Bergerat A, Louis P, McIntosh F, Johnstone AM, Lobley GE, Parkhill J, Flint HJ. Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *The ISME Journal*, 2011, 5(2): 220–230.
- [130] Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Science Translational Medicine*, 2009, 1(6): 6ra14.
- [131] Storelli G, Defaye A, Erkosar B, Hols P, Royet J, Leulier F. *Lactobacillus plantarum* promotes *Drosophila* systemic growth by modulating hormonal signals through TOR-dependent nutrient sensing. *Cell Metabolism*, 2011, 14(3): 403–414.
- [132] Ryu JH, Kim SH, Lee HY, Bai JY, Nam YD, Bae JW, Lee DG, Shin SC, Ha EM, Lee WJ. Innate immune homeostasis by the homeobox gene *Caudal* and commensal-gut mutualism in *Drosophila*. *Science*, 2008, 319(5864): 777–782.
- [133] Douglas AE. Is the regulation of insulin signaling multi-organismal? *Science Signaling*, 2011, 4(203): pe46.
- [134] Tower J. *Lactobacillus plantarum* gives *Drosophila* the grow signal. *Cell Metabolism*, 2011, 14(3): 283–284.
- [135] Martínez-Sañudo I, Simonato M, Squartini A, Mori N, Marri L, Mazzon L. Metagenomic analysis reveals changes of the *Drosophila suzukii* microbiota in the newly colonised regions. *Insect Science*, 2017, doi: 10.1111/1744-7917.12458.
- [136] Mota MM, Takemoto S, Takeuchi Y, Hara N, Futai K. Comparative studies between Portuguese and Japanese isolates of the pinewood nematode, *Bursaphelenchus xylophilus*. *The Journal of Nematology*, 2006, 38(4): 429–433.
- [137] Chandler JA, Lang JM, Bhatnagar S, Eisen JA, Kopp A. Bacterial communities of diverse *Drosophila* species: ecological context of a host-microbe model system. *PLoS Genetics*, 2011, 7(9): e1002272.
- [138] Mazzetto F, Gonella E, Crotti E, Vacchini V, Syrpas M, Pontini M, Mangelinckx S, Daffonchio D, Alma A. Olfactory

- attraction of *Drosophila suzukii* by symbiotic acetic acid bacteria. *Journal of Pest Science*, 2016, 89(3): 783–792.
- [139] Dobson AJ, Chaston JM, Douglas AE. The *Drosophila* transcriptional network is structured by microbiota. *BMC Genomics*, 2016, 17(1): 975.
- [140] Han G, Lee HJ, Jeong SE, Jeon CO, Hyun S. Comparative analysis of *Drosophila melanogaster* gut microbiota with respect to host strain, sex, and age. *Microbial Ecology*, 2017, 74(1): 207–216.
- [141] Xiao XP, Yang LJ, Pang XJ, Zhang RD, Zhu YB, Wang PH, Gao GJ, Cheng G. A Mesh-Duox pathway regulates homeostasis in the insect gut. *Nature Microbiology*, 2017, 2: 17020.
- [142] Obadia B, Guvener ZT, Zhang VV, Ceja-Navarro JA, Brodie EL, Ja WW, Ludington WB. Probabilistic invasion underlies natural gut microbiome stability. *Current Biology*, 2017, 27(13): 1999–2006.e8.
- [143] Bakker MG, Manter DK, Sheflin AM, Weir TL, Vivanco JM. Harnessing the rhizosphere microbiome through plant breeding and agricultural management. *Plant and Soil*, 2012, 360(1/2): 1–13.
- [144] Carbonetto B, Rascovan N, Álvarez R, Mentaberry A, Vázquez MP. Structure, composition and metagenomic profile of soil microbiomes associated to agricultural land use and tillage systems in Argentine Pampas. *PLoS One*, 2014, 9(6): e99949.
- [145] Schäfer RB, Caquet T, Siimes K, Mueller R, Lagadic L, Liess M. Effects of pesticides on community structure and ecosystem functions in agricultural streams of three biogeographical regions in Europe. *Science of the Total Environment*, 2007, 382(2/3): 272–285.
- [146] Widenfalk A, Bertilsson S, Sundh I, Goedkoop W. Effects of pesticides on community composition and activity of sediment microbes—responses at various levels of microbial community organization. *Environmental Pollution*, 2008, 152(3): 576–584.
- [147] Wang S, Dos-Santos ALA, Huang W, Liu KC, Oshaghi MA, Wei G, Agre P, Jacobs-Lorena M. Driving mosquito refractoriness to *Plasmodium falciparum* with engineered symbiotic bacteria. *Science*, 2017, 357(6358): 1399–1402.
- [148] Bian GW, Joshi D, Dong YM, Lu P, Zhou GL, Pan XL, Xu Y, Dimopoulos G, Xi ZY. *Wolbachia* invades *Anopheles stephensi* populations and induces refractoriness to *Plasmodium* infection. *Science*, 2013, 340(6133): 748–751.
- [149] Bando H, Okado K, Guelbeogo WM, Badolo A, Aonuma H, Nelson B, Fukumoto S, Xuan XN, Sagnon N, Kanuka H. Intra-specific diversity of *Serratia marcescens* in *Anopheles* mosquito midgut defines *Plasmodium* transmission capacity. *Scientific Reports*, 2013, 3: 1641.
- [150] Ohkuma M, Noda S, Hattori S, Iida T, Yuki M, Starns D, Inoue J, Darby AC, Hongoh Y. Acetogenesis from H₂ plus CO₂ and nitrogen fixation by an endosymbiotic spirochete of a termite-gut cellulolytic protist. *Proceedings of the National Academy of Sciences of the United States of America*, 2015, 112(33): 10224–10230.
- [151] Raymann K, Shaffer Z, Moran NA. Antibiotic exposure perturbs the gut microbiota and elevates mortality in honeybees. *PLoS Biology*, 2017, 15(3): e2001861.
- [152] Koch H, Schmid-Hempel P. Socially transmitted gut microbiota protect bumble bees against an intestinal parasite. *Proceedings of the National Academy of Sciences of the United States of America*, 2011, 108(48): 19288–19292.
- [153] Thakur A, Dhammi P, Saini HS, Kaur S. Effect of antibiotic on survival and development of *Spodoptera litura* (Lepidoptera: Noctuidae) and its gut microbial diversity. *Bulletin of Entomological Research*, 2016, 106(3): 387–394.
- [154] Li GN, Xia XJ, Tang WC, Zhu Y. Intestinal microecology associated with fluoride resistance capability of the silkworm (*Bombyx mori* L.). *Applied Microbiology and Biotechnology*, 2016, 100(15): 6715–6724.
- [155] Caccia S, Di Lelio I, La Stora A, Marinelli A, Varricchio P, Franzetti E, Banyuls N, Tettamanti G, Casartelli M, Giordana B, Ferré J, Gigliotti S, Ercolini D, Pennacchio F. Midgut microbiota and host immunocompetence underlie *Bacillus thuringiensis* killing mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, 2016, 113(34): 9486–9491.
- [156] Smith CC, Srygley RB, Healy F, Swaminath K, Mueller UG. Spatial structure of the mormon cricket gut microbiome and its predicted contribution to nutrition and immune function. *Frontiers in Microbiology*, 2017, 8: 801.
- [157] Haloi K, Kalita MK, Nath R, Devi D. Characterization and pathogenicity assessment of gut-associated microbes of muga silkworm *Antheraea assamensis* Helfer (Lepidoptera: Saturniidae). *Journal of Invertebrate Pathology*, 2016, 138: 73–85.
- [158] Liang X, Fu YM, Tong L, Liu H. Microbial shifts of the silkworm larval gut in response to lettuce leaf feeding. *Applied Microbiology and Biotechnology*, 2014, 98(8): 3769–3776.

Metagenomics of insect gut: new borders of microbial big data

Le Cao, Kang Ning*

College of Life Science & Technology, Huazhong University of Science and Technology, Wuhan 430074, Hubei Province, China

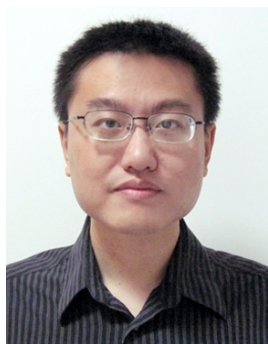
Abstract: Microbes, the ubiquitous organisms in nature, usually live in the "microbial community". A microbial community usually contains tens to thousands of different microorganisms, where they cooperate with each other to adapt to changes in the environment to complete the evolution of life, while their life activities also have a long and profound impact on the environment. With a better understanding of microorganisms, basic research on microbial community and its application in the field of health and the environment is increasingly important. There are a large number and a wide variety of microbes in the intestine of insects. On the one hand, the diversity of these intestinal microbial populations is closely related to insect species, age, digestive tract form, food source and environment. On the other hand, these microflora also have some effect on the host's physiological activities. With the development of high-throughput sequencing technology and omics technology, the use of molecular biology technology to quickly, qualitatively and quantitatively study insect intestinal microbes has become a popular field. Data mining and application on insect intestine metagenome will greatly promote our ability to use microbial resources. This paper summarizes the development of metagenomics of insect intestine, especially the excavation tools and applications of large data of intestinal genome and the research progress, application, advantages, as well as bottlenecks of metagenomics in insect intestine. The future research direction of insect microbial population is also forecasted.

Keywords: metagenomics, insect gut, high throughput sequencing, biological data

(本文责编: 李磊)

*Corresponding author. Tel/Fax: +86-27-87793041; E-mail: ningkang@hust.edu.cn

Received: 29 September 2017; Revised: 13 February 2018; Published online: 9 May 2018



宁康, 华中科技大学生命科学与技术学院教授, 博士生导师, 生物信息与系统生物学系主任, 湖北省楚天学者特聘教授。2003年本科毕业于中国科学技术大学计算机科学技术专业, 2008年博士毕业于新加坡国立大学生物信息学专业, 2010年于美国密歇根大学医学院完成博士后研究工作回国。先后在中国科学院青岛生物能源与过程研究所和华中科技大学从事微生物相关生物信息学研究工作, 研究重点方向为微生物信息学中的生物大数据挖掘及其在健康与环境等领域的应用。在微生物组学领域从事科研工作10余年, 发表高水平学术论文50余篇, 文章总引用超过1500次, 并是10余项发明专利和软件著作权的拥有者或主要作者。